

# **Advanced Blind Source Separation Methods for Multivariate Data Modeling and Clustering**

**Ali Al-gumaei**

**A Thesis**

**in**

**The Department**

**of**

**Concordia Institute for Information Systems Engineering (CIISE)**

**Presented in Partial Fulfillment of the Requirements**

**for the Degree of**

**Doctor of Philosophy (Information and Systems Engineering) at**

**Concordia University**

**Montréal, Québec, Canada**

**March 2025**

**© Ali Al-gumaei, 2025**

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Ali Al-gumaei**

Entitled: **Advanced Blind Source Separation Methods for Multivariate Data  
Modeling and Clustering**

and submitted in partial fulfillment of the requirements for the degree of

**Doctor of Philosophy (Information and Systems Engineering)**

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

\_\_\_\_\_ Chair  
*Dr. Mohsen Ghafouri*

\_\_\_\_\_ External Examiner  
*Dr. Belkacem Chikhaoui*

\_\_\_\_\_ Examiner  
*Dr. Abdessamad Ben Hamza*

\_\_\_\_\_ Examiner  
*Dr. Lyes Kadem*

\_\_\_\_\_ Examiner  
*Dr. Jamal Bentahar*

\_\_\_\_\_ Supervisor  
*Dr. Nizar Bouguila*

Approved by

\_\_\_\_\_  
Dr. Chung Wang, Chair  
Department of Concordia Institute for Information Systems Engineering  
(CIISE)

February 27, 2025

\_\_\_\_\_  
Dr. Mourad Debbabi, Dean  
Gina Cody School of Engineering and Computer Science

# Abstract

## Advanced Blind Source Separation Methods for Multivariate Data Modeling and Clustering

Ali Al-gumaei, Ph.D.

Concordia University, 2025

As the amount of data being generated keeps growing, there is an increasing demand for adaptable approaches that can effectively extract the overall trends while also maintaining subject-specific information from large-scale datasets. Modeling complex and high-dimensional data presents significant challenges across healthcare, human action recognition, and speech recognition.

In this dissertation, we develop a bounded multivariate generalized Gaussian mixture model (BMGGMM) integrated with independent component analysis (ICA) to effectively capture the correlated features in multivariate data. While independent vector analysis (IVA) extends ICA to handle multiple datasets by leveraging inter-dataset dependencies and preserving their correlation structures, it suffers from limitations when dealing with complex datasets. To overcome this, we propose a novel blind source separation (BSS) method that combines IVA with the BMGGMM framework, enabling robust modeling of complex data distributions with varying shapes and dimensions. Second, we introduced the integration of the ICA-BMGGMM and IVA-BMGGMM to the hidden Markov model (HMM) to boost their performance in terms of source separation. The performance of IVA deteriorates as the number of datasets and sources increases. To address this limitation, we propose the adaptive constrained IVA (acIVAMGGMM) and bounded acIVAMGGMM techniques. These methods integrate multiple reference signals into the IVA function and adaptively control the reference-estimated source relations. Finally, we introduce a new approach, ICA and IVA for common subspace analysis (ICABMGGMM-CS) and IVABMGGMM-CS, designed for the subspace analysis of multi-subject fMRI data. These methods leverage the strengths of both ICA and IVA while effectively addressing the challenges posed by high dimensions.

# Contribution of Authors

This dissertation consists of nine manuscripts. Six manuscripts have been published; one manuscript has been accepted, and the rest have been submitted for publication in refereed academic journals. Each Chapter comprises the content of a manuscript that has been meticulously reformatted and systematically reorganized to adhere to the guidelines established by the School of Graduate Studies.

- Manuscript 1 (Chapter 2): Ali Algumaei, Muhammad Azam, Fatma Najar, and Nizar Bouguila. "Bounded multivariate generalized Gaussian mixture model using ICA and IVA." *Pattern Analysis and Applications* 26, no. 3 (2023): 1223-1252.
- Manuscript 2 (Chapter 3): Ali Algumaei, Muhammad Azam, Manar Amayri, and Nizar Bouguila. "ICA and IVA bounded multivariate generalized Gaussian mixture based hidden Markov models." *Engineering Applications of Artificial Intelligence* 123 (2023): 106345.
- Manuscripts 3, 4 (Chapter 4): Ali Algumaei, Muhammad Azam, and Nizar Bouguila. "Adaptive Constrained IVAMGGMM: Application to Mental Disorders Detection." *IEEE Transactions on Emerging Topics in Computational Intelligence* (2024).  
Ali Algumaei, Muhammad Azam, Manar Amayri, and Nizar Bouguila. "Adaptive Constrained ICAMGGMM: An Improvement Over ICA." In *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 1-7. IEEE, 2023.
- Manuscripts 5, 6, 7 (Chapter 5): Ali Algumaei, Muhammad Azam, and Nizar Bouguila. "Novel approach for ECG separation using adaptive constrained IVABMGGMM." *Digital Signal Processing* 149 (2024): 104476.  
Ali Algumaei, Muhammad Azam, and Nizar Bouguila. "Mental Disorder Detection using Non-Orthogonal Adaptive Constrained IVABMGGMM." In *2024 IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP)*, vol. 1, pp. 65-70. IEEE, 2024.  
Ali Algumaei, Muhammad Azam, Manar Amayri, and Nizar Bouguila. "Adaptive Constrained ICABMGGMM: application to ECG blind source separation." *The 31st International Conference on Neural Information Processing (ICONIP 2024)*.

- Manuscript 8 (Chapter 6): Ali Algumaei, Muhammad Azam, and Nizar Bouguila. "Identifying the common subspaces regions using independent component analysis and graph-theoretical analysis" Pattern Recognition, submitted.
- Manuscript 9 (Chapter 7): Ali Algumaei, Muhammad Azam, and Nizar Bouguila. "Homogeneous Subgroup Identification in Resting State fMRI using Common Subspace Independent Vectors Analysis." ACM Transactions on Computing for Healthcare, submitted.

# Acknowledgments

First, I wish to express my deepest and most heartfelt gratitude to my exceptional supervisor, Prof. Nizar Bouguila. Your unwavering guidance, encouragement, and unconditional availability have been a constant source of inspiration. I am profoundly thankful for your trust in me, for everything I have learned from you, and for every single opportunity you have so generously provided. Thank you for your patience, caring, understanding, and continuous support.

I would also like to thank my committee members for their invaluable time and effort in reviewing my work throughout each milestone of my PhD. Their thoughtful, constructive, and insightful feedback has been instrumental in enhancing the quality of my research and broadening my academic perspective. Your guidance has not only improved this work but has also inspired me to explore new dimensions in my field. Thank you for your dedication and support. A special thanks goes to Dr. Muhammad Azam for his invaluable suggestions, insightful discussions, and unwavering support throughout my research. Your contributions have significantly enriched my work, and I deeply appreciate your generosity and encouragement. Thanks Dr. Manar Amayri and Dr. Fatma Najjar, for their contribution to my work. I would also like to thank my labmates for the wonderful moments and unforgettable memories we have shared.

Special thanks to my beloved mother (Halya Alorimi) and my sister (Hyam Al-gumaei), whose unwavering love, support, and belief in me have been the bedrock of my journey. Your resilience, selflessness, and prayers have been my greatest strength and inspiration. Thanks to my wife, Muna Al-gumaei, for her unwavering support during this journey. Her patience and understanding have been a constant source of strength and motivation. Thanks to my loved kids, Mazen, Yazan, and Karim, for being the greatest source of happiness and inspiration in my family.

# Contents

<b>List of Figures</b>	<b>xii</b>
<b>List of Tables</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Problem statement and motivations . . . . .	1
1.2 Thesis Contributions . . . . .	3
1.3 Thesis Outline . . . . .	6
<b>2 Bounded Multivariate Generalized Gaussian Mixture Model using ICA and IVA</b>	<b>7</b>
2.1 Introduction . . . . .	8
2.2 Bounded multivariate generalized Gaussian mixture model . . . . .	11
2.2.1 Model definition . . . . .	11
2.2.2 Parameter estimation of the bounded multivariate generalized Gaussian mixture model . . . . .	13
2.3 Model selection using MML criterion for BMGGMM . . . . .	16
2.3.1 Derivation of the prior distribution . . . . .	17
2.4 ICA mixture of bounded multivariate generalized Gaussian distributions . . . . .	19
2.4.1 Parameter estimation using ICA and gradient ascent . . . . .	20
2.5 IVA mixture of multivariate generalized Gaussian distributions . . . . .	21
2.6 Experimental results . . . . .	23
2.6.1 Experimental evaluation of BMGGMM . . . . .	24

2.6.2	Experiments and results for clustering using BMGGMM-MML . . . . .	28
2.6.3	Experimental results for BMGGMM with ICA . . . . .	29
2.6.4	Experimental results for IVA-BMGGMM . . . . .	36
2.7	Conclusion and discussions . . . . .	46
<b>3</b>	<b>ICA and IVA bounded multivariate generalized Gaussian mixture based hidden Markov models</b>	<b>48</b>
3.1	Introduction . . . . .	49
3.2	Integration of ICA and IVA into the HMM framework . . . . .	52
3.2.1	ICA mixture of bounded multivariate generalized Gaussian distributions . .	52
3.2.2	IVA mixture of bounded multivariate generalized Gaussian distributions . .	52
3.2.3	Hidden Markov model . . . . .	53
3.2.4	ICA-BMGGMM and IVA-BMGGMM integration into the HMM framework	55
3.3	Experimental results . . . . .	58
3.3.1	Human action recognition . . . . .	59
3.3.2	Speech recognition . . . . .	64
3.3.3	Energy disaggregation . . . . .	69
3.3.4	Epileptic Seizure clustering using EEG signals . . . . .	72
3.4	Conclusion . . . . .	74
<b>4</b>	<b>Adaptive constrained IVAMGGMM: Application to neurological disorders detection</b>	<b>76</b>
4.1	Introduction . . . . .	77
4.2	Related Work . . . . .	81
4.2.1	Constrained ICA models . . . . .	81
4.2.2	Constrained IVA models . . . . .	82
4.3	Adaptive constrained ICA models . . . . .	83
4.3.1	ICA based on multivariate generalized Gaussian mixture model . . . . .	83
4.3.2	Constrained ICA-MGGMM . . . . .	84
4.3.3	Adaptive Constrained ICA-MGGMM . . . . .	85
4.4	Adaptive constrained IVA models . . . . .	86

4.4.1	IVA based on multivariate generalized Gaussian mixture model . . . . .	86
4.4.2	Constrained IVA-MGGMM . . . . .	87
4.4.3	Adaptive constrained IVA-MGGMM . . . . .	88
4.5	Results and discussion . . . . .	89
4.5.1	Simulation results . . . . .	90
4.5.2	Alzheimer's detection using cICA-MGGMM and acICA-MGGMM . . . . .	93
4.5.3	Alzheimer's detection using cIVA-MGGMM and acIVA-MGGMM . . . . .	98
4.5.4	Schizophrenia detection using cICA-MGGMM and acICA-MGGMM . . . . .	103
4.5.5	Schizophrenia detection using cIVA-MGGMM and acIVA-MGGMM . . . . .	106
4.5.6	Functional connectivity in resting state networks of ADHD patients re- vealed by acICA-MGGMM and acIVA-MGGMM . . . . .	110
4.5.7	Epileptic Seizure prediction . . . . .	113
4.5.8	Comparison to similar BSS approaches in Schizophrenic detection . . . . .	116
4.6	More Experiments for cICA-MGGMM and acICA-MGGMM . . . . .	124
4.6.1	Speech signal separation . . . . .	125
4.6.2	EEG signal separation . . . . .	127
4.7	Conclusion . . . . .	131
<b>5</b>	<b>Novel Approach for ECG Separation Using Adaptive Constrained ICABMGGMM and IVABMGGMM</b>	<b>133</b>
5.1	Introduction . . . . .	134
5.2	ICA-based approaches for blind source separation . . . . .	138
5.2.1	Bounded multivariate generalized Gaussian mixture model using ICA . . . . .	138
5.2.2	Constrained bounded ICAMGGMM (cICABMGGMM) . . . . .	138
5.2.3	Adaptive constrained bounded ICAMGGMM (acICABMGGMM) . . . . .	139
5.3	IVA-based approaches for BSS . . . . .	140
5.3.1	Bounded multivariate generalized Gaussian mixture model using IVA . . . . .	140
5.3.2	Constrained IVABMGGMM (cIVABMGGMM) . . . . .	141
5.3.3	Adaptive constrained bounded IVAMGGMM (acIVABMGGMM) . . . . .	142

5.4	Experimental Results for cICABMGGMM and acICABMGGMM . . . . .	144
5.4.1	Simulation experiment . . . . .	144
5.4.2	Fetal ECG signal extraction . . . . .	146
5.4.3	Arrhythmia detection . . . . .	149
5.4.4	Robustness to noise . . . . .	152
5.5	Experiments and Results for cIVABMGGMM and acIVABMGGMM . . . . .	153
5.5.1	Heartbeat separation . . . . .	153
5.5.2	Fetal ECG signal separation . . . . .	157
5.5.3	Arrhythmia detection . . . . .	160
5.6	Conclusion . . . . .	164
<b>6</b>	<b>Identifying the common subspaces regions using independent component analysis and graph-theoretical analysis</b>	<b>166</b>
6.1	Introduction . . . . .	167
6.2	Background . . . . .	169
6.2.1	Independent component analysis for the bounded multivariate generalized Gaussian mixture model . . . . .	169
6.3	Proposed model: ICABMGGMM with the common subspaces framework . . . . .	169
6.3.1	ICABMGGMM-CS: subset analysis . . . . .	170
6.3.2	ICABMGGMM-CS: common subspace identification . . . . .	171
6.4	Results and discussion . . . . .	172
6.4.1	Schizophrenia detection using the common subspaces ICABMGGMM . . . . .	173
6.4.2	Alzheimer diagnosis using the common subspaces ICABMGGMM . . . . .	181
6.5	Conclusion . . . . .	186
<b>7</b>	<b>Homogeneous Subgroup Identification in Resting State fMRI using Common Subspace Independent Vectors Analysis</b>	<b>189</b>
7.1	Introduction . . . . .	190
7.2	Related Work . . . . .	193
7.3	Background . . . . .	194

7.3.1	Independent vector analysis for the bounded multivariate generalized Gaussian mixture model . . . . .	194
7.4	Proposed IVA mixture for common subspace analysis . . . . .	194
7.4.1	IVABMGGMM-CSS: subset analysis . . . . .	195
7.4.2	IVABMGGMM-CSS: common subspace identification . . . . .	197
7.5	Results and discussion . . . . .	197
7.5.1	Application to simulated data . . . . .	197
7.5.2	Schizophrenic diagnosis using the common subspace IVAMGGMM and IV-ABMGGMM . . . . .	200
7.5.3	Autism detection using the common subspace IVAMGGMM and IVAB-MGGMM . . . . .	209
7.6	Conclusion . . . . .	219
<b>8</b>	<b>Conclusion and future work</b>	<b>220</b>
8.1	Conclusions . . . . .	220
8.2	Future Directions . . . . .	221
	<b>Appendix A</b>	<b>223</b>
A.1	Mean estimation . . . . .	223
A.2	Covariance estimation . . . . .	225
A.3	Shape parameter estimation . . . . .	227
	<b>Appendix B</b>	<b>230</b>
B.1	Estimation of $\Lambda$ . . . . .	230
	<b>Bibliography</b>	<b>232</b>

# List of Figures

Figure 2.1	General framework of the IVA model. The $n$ th source component vectors (SCV) is formed by grouping the corresponding $m$ th source from each source vector together. . . . .	24
Figure 2.2	BSS for heartbeat signals using ICA. The left acoustic signal represents the fetal heartbeat sound, while the right acoustic signal represents the mother heartbeat sound. . . . .	30
Figure 2.3	BSS for heartbeat signals using ICA-BMGGMM. The fetal heartbeat sound is represented by the left acoustic signal, while the mother heartbeat sound is represented by the right acoustic signal. . . . .	31
Figure 2.4	Original ECG channels. These three channels were extracted from a pregnant women dataset. . . . .	31
Figure 2.5	ICA-BMGGMM ECG channels. On the left are the extracted ICA-BMGGMM ECG signals for mother. On the right is the fetal ICA-BMGGMM ECG signal. . .	32
Figure 2.6	Speech signals and their corresponding spectrograms from source 1. The first row represents both the raw speech signal and its spectrogram. The second row illustrates the speech mixed signal and its spectrogram. The third row represents the extracted speech signal using ICA-BMGGMM and its corresponding spectrogram. . . . .	34
Figure 2.7	Speech signals and their corresponding spectrograms from source 2. The first row represents both the raw speech signal and its spectrogram. The second row illustrates the speech mixed signal and its spectrogram. The third row represents the extracted speech signal using ICA-BMGGMM and its corresponding spectrogram. . . . .	35

Figure 2.8	Fp1 channel before and after blink removal. On the upper, the EEG signal is contaminated by eye blinking signals. After the eye blink signals have been removed, the cleaning EEG signal is shown at the bottom. . . . .	36
Figure 2.9	EEG block diagram. Illustration of epoching on continuous EEG data to build a dataset for IVA algorithms to obtain estimated components for EEG artifact removal. SCVs formed from estimates are also depicted. . . . .	39
Figure 2.10	EEG channels before eye blink removal. EEG signal contaminated by the artifacts of eye blinking. . . . .	39
Figure 2.11	EEG channels after eye blink removal. The obtained EEG signals after removing eye blink artifacts with IVA-BMGGMM. . . . .	40
Figure 2.12	Fp1 channel before and after BSS. The blue EEG signal represents EEG signals contaminated by eye-blinking artifacts. The yellow one represents the clean EEG signals obtained after IVA-BMGGMM was used to eliminate eye blinking. . . . .	41
Figure 2.13	Block diagram of the proposed schizophrenia detection system. . . . .	43
Figure 2.14	The defected regions for schizophrenic patients. On the left are the defected regions for frontal networks. The activated regions in the DMN are shown on the right. . . . .	44
Figure 2.15	Original ECG channels. These three channels were extracted from a pregnant women dataset. . . . .	44
Figure 2.16	IVA-BMGGMM ECG channels. On the left are the extracted IVA-BMGGMM ECG signals for mother. On the right is the fetal ICA-BMGGMM ECG signal. . . . .	45
Figure 3.1	Graphical representation for the HMM model. . . . .	55
Figure 3.2	Training procedures for HMM. . . . .	55
Figure 3.3	Integration of the proposed models into the HMM framework. . . . .	56
Figure 3.4	A sample frame for each action in the Weizmann action dataset: (top row) running, side jumping, skip jumping, jumping, and jumping in place; (bottom row) bending, jumping jack, walking, one hand waving, and two hands waving. . . . .	60
Figure 3.5	Confusion matrix for the IVABMGGMM-HMM model for Weizmann data. . . . .	61

Figure 3.6	Examples of frames from the KTH dataset showing various human actions in different scenarios. d1: outdoors, d2: outdoors with scale variations, d3: outdoors with different clothes, d4: indoors. . . . .	62
Figure 3.7	Example for the dense optical flow features for boxing, handclapping, and jogging. The first row represents the human action, and the second row represents the extracted features corresponding to each action. . . . .	63
Figure 3.8	Confusion matrix for the IVABMGGMM-HMM model for KTH data. . . .	64
Figure 3.9	The accuracy differences between the proposed models on TIMIT data with 7 and 5 classes. . . . .	68
Figure 3.10	Confusion matrix for the IVBMGGMM-HMM for spoken digit data. . . .	69
Figure 3.11	Typical appliance signatures for Ref, DW, WD and KO. . . . .	71
Figure 3.12	The accuracy differences for the proposed models over the four appliances. . . . .	72
Figure 3.13	ROC curves for refrigerator using the four models IVABMGGMM-HMM, IVA-HMM, ICABMGGMM-HMM and ICA-HMM. . . . .	73
Figure 4.1	ICA and IVA models for Alzheimer’s and Schizophrenia detection based on resting state fMRI data. . . . .	91
Figure 4.2	Number of iterations needed for acICA-MGGMM to converge. . . . .	94
Figure 4.3	Number of acIVA-MGGMM iterations. . . . .	94
Figure 4.4	Dissimilarity factor for (a) acICA-MGGMM vs cICA-MGGMM, (b) acIVA-MGGMM vs cIVA-MGGMM. . . . .	95
Figure 4.5	Spatial correlation for the proposed cICA-MGGMM and acICA-MGGMM concerning the number of subjects. . . . .	97
Figure 4.6	Performance of the proposed cIVA-MGGMM and acIVA-MGGMM compared to IVA, IVA-G, cIVA, and acIVA in terms of the number of subjects using the Joint-ISI metric. . . . .	99
Figure 4.7	Spatial correlation for IVA, IVA-G, cIVA, acIVA, cIVA-MGGMM, and acIVA-MGGMM in terms of number of subjects. . . . .	100
Figure 4.8	The joint-ISI for the four proposed models; cICA-MMGGMM, acICA-MMGGMM, cIVA-MGGMM, and acIVA-MGGMM in terms of the number of subjects. . . . .	100

Figure 4.9	Localization of the areas in the AAL atlas. The areas exhibiting higher discriminability between the healthy and Alzheimer’s groups are depicted by blue circles, while unaffected areas are represented by green circles. . . . .	102
Figure 4.10	Spatial correlation for ICA, Non-ICA, cICA, acICA, cICA-MGGMM, and acICA-MGGMM in terms of number of subjects for Schizophrenic patients detection.	105
Figure 4.11	Spatial correlation for IVA, IVA-G, cIVA, acIVA, cIVA-MGGMM and acIVA-MGGMM in terms of number of subjects for Schizophrenia detection. . . . .	107
Figure 4.12	The joint-ISI for the four proposed and base models in terms of number of subjects. . . . .	107
Figure 4.13	Localization of the areas in the AAL atlas. The areas with higher discriminability between healthy and Schizophrenic groups are marked by red circles, while the unaffected areas are represented by yellow circles. . . . .	109
Figure 4.14	The activation regions in the motor area for the non-ADHD subject using acICA-MGGMM and acIVA-MGGMM. The first row depicts the activation areas using acIVA-MGGMM, whereas the second row illustrates the activation regions using acICA-MGGMM. . . . .	111
Figure 4.15	Activation regions using acIVA-MGGMM in both control and ADHD. The activation regions in control and ADHD subjects are represented by the first and second rows, respectively. . . . .	112
Figure 4.16	Correlation matrix using acIVA-MGGMM for both healthy and ADHD subjects. (a) Control correlation. (b) Patient correlation. . . . .	113
Figure 4.17	The activation regions in the visual area for healthy controls and ADHD patients using the proposed acIVA-MGGMM. The activation regions in control and ADHD subjects are illustrated by the top and bottom rows, respectively. . . . .	114
Figure 4.18	ADHD and control subjects connectomes. The first row represents control subject connectomes while the second row indicates ADHD patient connectomes.	115
Figure 4.19	Joint-ISI for both acICA-MGGMM and acIVA-MGGMM for the EEG data.	115
Figure 4.20	Schizophrenia detection using the recent BSS approaches. . . . .	116

Figure 4.21	Comparison of Schizophrenia detection performance using recent BSS models, evaluated with the Joint-ISI metric. The left figure illustrates results from ICA-based models, while the right figure presents those from IVA-based models. . . . .	117
Figure 4.22	Spatial correlation performance of BSS approaches in diagnosing Schizophrenia. The left figure shows the performance of ICA-based approaches, while the right figure highlights results from IVA-based models. . . . .	119
Figure 4.23	The most discriminating IVA components for Schizophrenic patients using proposed acIVA-MGGMM. The number in parentheses demonstrates the number of IVA components for each brain network. . . . .	120
Figure 4.24	The most discriminative ICA components for Schizophrenic patients using acICA-MGGMM. . . . .	120
Figure 4.25	Correlation matrix for healthy subjects (a) and Schizophrenic patients (b) using the acIVA-MGGMM components. . . . .	123
Figure 4.26	Performance of acICA-MGGMM and base models using ISI metric with a variation of mixtures. . . . .	127
Figure 4.27	PESQ measure for acICA-MGGMM and other models with different mixtures.	128
Figure 4.28	Speech signal separation using acICA-MGGMM. The first row represents the original speech signal. The second row illustrates the the mixed speech signals. The third row demonstrates the speech signal estimation via acICA-MGGMM. . .	129
Figure 4.29	Comparison of the performance of acICA-MGGMM and the basic models in relation to the ISI metric as a function of the number of subjects. . . . .	130
Figure 4.30	Evaluation of acICA-MGGMM's performance in relation to spatial correlation as it varies with the number of subjects. . . . .	131
Figure 4.31	The estimated EEG signals using acICA-MGGMM. The first row represents the normal estimated EEG signal and the second row indicates the Seizure one. . .	132
Figure 5.1	The convergence process and robustness of our model where (a) illustrates the iterations number required for convergence of acICABMGGMM, while (b) represents dissimilarity factor for acICABMGGMM vs cICABMGGMM. . . . .	146

Figure 5.2	Real ECG recordings. The first row represents the abdominal recording, whereas the second to fourth rows represent thoracic recordings. . . . .	147
Figure 5.3	FECG signal estimation employing acICABMGGMM, coupled with R-peak identification using orange color. . . . .	149
Figure 5.4	Our proposed model performance in terms of Joint-ISI and spatial correlation where (a) illustrates the model’s performance based on Joint-ISI, while (b) assesses its effectiveness in terms of spatial correlation. . . . .	151
Figure 5.5	Our proposed model performance in terms of different noise conditions and spatial correlation where (a) illustrates the model’s performance based on the fetal ECG signal extraction experiment, while (b) assesses its performance based on the arrhythmia detection experiment. . . . .	153
Figure 5.6	Heartbeat source signals. The first row indicates the normal heartbeat alongside its corresponding spectrograms and MFCCs features. The second row represents the abnormal heartbeat and its corresponding spectrogram and MFCCs features.	154
Figure 5.7	The estimated heartbeats using acIVABMGGMM, along with their corresponding spectrogram. The first row represents the normal estimated heartbeats, and the second row indicates the abnormal one. . . . .	156
Figure 5.8	The effectiveness of our proposed models in terms of Joint-ISI when compared with conventional methods. . . . .	157
Figure 5.9	FECG signal estimation employing acIVABMGGMM, coupled with R-peak identification. . . . .	159
Figure 5.10	Joint-ISI for both base and proposed models. . . . .	159
Figure 5.11	ECG data processing steps: The first row represents the original ECG signals. The second row indicates the filtered ECG signal (BPF). The third and fourth rows illustrate its derivative (DES) and moving window-integrated version (MOV).	160
Figure 5.12	The joint-ISI for the four proposed models; cIVABMGGMM, acIVABMGGMM, and other compared models in terms of the number of subjects. . . . .	162
Figure 5.13	ECG signal estimate using acIVABMGGMM, along with the detection of R-peaks. . . . .	163

Figure 5.14	Our proposed model performance in terms of Joint-ISI and spatial correlation. The left graph illustrates the model's performance based on Joint-ISI, while the right graph assesses its effectiveness in terms of spatial correlation. . . . .	163
Figure 6.1	ICABMGGMM-CS framework. The first step is to apply the ICABMGGMM on the multi-subjects data. Then, the identification for the common subgroups is conducted. . . . .	171
Figure 6.2	The general block diagram for the ICA common subspaces approaches. . .	173
Figure 6.3	The proposed ICABMGGMM-CS and comparing models performance based on the Joint-ISI measure. . . . .	174
Figure 6.4	The performance of proposed ICABMGGMM-CS and comparing models based on Spatial correlation. . . . .	175
Figure 6.5	The spatial maps of the common subspace ICA components for Schizophrenic patients. The ICAMGGMM-CS components are displayed in the left column, while the ICABMGGMM-CS components are illustrated in the right column. . . . .	177
Figure 6.6	The connectogram map of the common subspace ICA components for Schizophrenic patients. . . . .	178
Figure 6.7	The graph analysis for the separated ICABMGGMM-CS components. . . .	180
Figure 6.8	Graph analysis for the ICABMGGMM approach conducted on COBRE data using path length and betweenness centrality. . . . .	180
Figure 6.9	ROC curves for ICABMGGMM-CS and compared models based on the resultant from two-sample t-test. . . . .	182
Figure 6.10	Joint ISI metric for proposed and compared models. . . . .	183
Figure 6.11	Spatial correlation measure for proposed and compared models. . . . .	184
Figure 6.12	The spatial maps of the common subspace ICA components for Alzheimer patients. . . . .	185
Figure 6.13	Brain regions localization. The blue color represents the affected regions due to Alzheimer whereas the green color indicates the normal regions. . . . .	186
Figure 6.14	Graph analysis for our proposed ICABMGGMM-CS. . . . .	187

Figure 6.15	Graph analysis for the ICABMGGMM approach conducted on ADNI data using path length and betweenness centrality. . . . .	187
Figure 6.16	ROC curves for ICABMGGMM-CS and compared models based on the resultant from two-sample t-test using the ADNI data. . . . .	188
Figure 7.1	IVABMGGMM-CSS framework. The first step is to apply the IVABMGGMM on the multi-subjects data. Then, the identification for the common subgroups is conducted. . . . .	196
Figure 7.2	The joint-ISI metric for proposed IVAMGGMM-CSS and IVABMGGMM-CSS algorithms in terms of sample number. . . . .	199
Figure 7.3	Normalized ISR for our proposed models and base models as a function of number of samples. . . . .	200
Figure 7.4	The performance of IVA methods based on the normalized average ISR as a function of shape parameter $\beta_j$ . Each point represents the result of 100 separate runs. . . . .	201
Figure 7.5	Common discriminative components correlation. . . . .	201
Figure 7.6	Participant demographics in CNP data. . . . .	202
Figure 7.7	The general framework illustrates the comprehensive proposed pipeline for fMRI data analysis. Our approach is conduct over the SCVs to identify the common and non-common subgroups. . . . .	203
Figure 7.8	Heatmap for the IVA conducted models across a range of combination settings. Each cell represents the performance score, with low performance (cool colors) and high performance (warm colors). . . . .	206
Figure 7.9	The performance of conducted models using SAR measure. . . . .	206
Figure 7.10	The IVABMGGMM-CSS performance based on the Joint-ISI metric. . . . .	207
Figure 7.11	The spatial maps of the IVABMGGMM-CSS components for Schizophrenia. . . . .	209
Figure 7.12	The IVABMGGMM-CSS performance in separating the Autism subjects using SIR measurement. . . . .	212
Figure 7.13	SAR measurements for our proposed models compared to base models. . . . .	213
Figure 7.14	The performance of IVABMGGMM-CSS in terms of Joint-ISI. . . . .	215

Figure 7.15 The most discriminating IVA components for Autism patients. . . . . 216

# List of Tables

Table 2.1	Performance on Stroke data based on different metrics. . . . .	25
Table 2.2	Performance on skin cancer data based on different metrics. . . . .	25
Table 2.3	Performance on Alzheimer data based on different metrics. . . . .	26
Table 2.4	Performance on Diabetes data based on different metrics. . . . .	27
Table 2.5	Performance on human resources data based on different metrics. . . . .	28
Table 2.6	Number of clusters determined by MML. . . . .	28
Table 2.7	Objective measure for separation of heartbeat signals. . . . .	30
Table 2.8	Objective measure for separation of 2 and 3 ECG signals. . . . .	32
Table 2.9	Objective measure for separation of 4 and 5 ECG signals. . . . .	33
Table 2.10	Objective measure for separation of 2 and 3 speech signals. . . . .	33
Table 2.11	Objective measure for separation of 4 and 5 speech signals. . . . .	34
Table 2.12	Objective measures for EEG experiments. . . . .	36
Table 2.13	Objective measure for separation of 2 and 3 speech signals. . . . .	37
Table 2.14	Objective measure for separation of 4 and 5 speech signals. . . . .	37
Table 2.15	Objective measure for separation of EEG signals. . . . .	41
Table 2.16	Objective measure for fMRI signals. . . . .	43
Table 2.17	Objective measure for separation of 2 and 3 ECG signals. . . . .	46
Table 2.18	Objective measure for separation of 4 and 5 ECG signals. . . . .	46
Table 3.1	Performance on Weizmann data based on different metrics. . . . .	60
Table 3.2	Performance on KTH data based on different metrics. . . . .	63
Table 3.3	Classes of phones in TIMIT data used in this experiment. . . . .	66

Table 3.4	Performance on TIMIT data based on different metrics. . . . .	66
Table 3.5	5 classes of phones in TIMIT data used in this experiment [1]. . . . .	67
Table 3.6	Performance on TIMIT data with 5 classes based on different metrics. . . . .	67
Table 3.7	Performance on spoken digit data based on different metrics. . . . .	69
Table 3.8	Appliances and houses used. . . . .	70
Table 3.9	Performance on REDD data based on different metrics for all appliances. . .	71
Table 3.10	Performance on EEG Seizure data based on different metrics. . . . .	74
Table 4.1	Mixture parameters estimation using proposed acIVA-MGGMM for gener- ated data with 2 clusters. . . . .	92
Table 4.2	Mixture parameters estimation using proposed acIVA-MGGMM for gener- ated data with 3 clusters. . . . .	93
Table 4.3	Performance of the proposed cICA-MGGMM and acICA-MGGMM in terms of Joint-ISI with respect to the number of sources. . . . .	97
Table 4.4	Performance of the proposed cICA-MGGMM and acICA-MGGMM in terms of Joint-ISI with respect to the number of sources for Schizophrenic data. . . . .	105
Table 4.5	Performance of the proposed cIVA-MGGMM and acIVA-MGGMM in terms of Joint-ISI with respect to the number of sources for Schizophrenic data. . . . .	106
Table 4.6	Brain networks with their corresponding IVA Components using proposed acIVA-MGGMM model. . . . .	121
Table 4.7	Objective metrics for separation of 2 and 3 speech sources using acICA- MGGMM and compared models . . . . .	125
Table 4.8	Objective metrics for separation of 4 and 5 speech sources using acICA- MGGMM and compared models. . . . .	125
Table 4.9	SDR and SIR measures for separation of EEG signals using acICA-MGGMM and compared models. . . . .	128
Table 5.1	Performance of the proposed cICABMGGMM and acICABMGGMM in terms of presented performance metrics. . . . .	148
Table 5.2	SDR and SIR separation metrics for cICABMGGMM and acICABMGGMM over source combinations ranging from 2 to 48. . . . .	150

Table 5.3	The evaluation of cICABMGGMM and acICABMGGMM performance over source combinations from 2 to 48, assessed using SAR and SNR separation measurements. . . . .	151
Table 5.4	Objective measures for separation of 2 and 3 heartbeat signals using our proposed models acIVABMGGMM and cIVABMGGMM. . . . .	155
Table 5.5	Objective measure for separation of 4 and 5 heartbeat signals using acIVABMGGMM and cIVABMGGMM proposed models. . . . .	155
Table 5.6	Performance of the proposed cIVABMGGMM and acIVABMGGMM in terms of presented performance metrics. . . . .	158
Table 5.7	SDR and SIR separation measurements for cIVABMGGMM and acIVABMGGMM across source combinations spanning from 2 to 48. . . . .	161
Table 5.8	The effectiveness of cIVABMGGMM and acIVABMGGMM across source combinations from 2 to 48 evaluated through SAR and SNR separation criteria. . . . .	161
Table 6.1	ICA components and their coordinates for different brain networks. . . . .	179
Table 6.2	Brain network components and associated anatomical regions. . . . .	185
Table 7.1	Performance of IVAMGGMM-CSS and IVABMGGMM-CSS using SDR separation measure. . . . .	205
Table 7.2	The most discriminate IVA components and peak activation in the t-maps of the resting-state networks. . . . .	210
Table 7.3	Performance of IVAMGGMM-CSS and IVABMGGMM-CSS using SDR separation measure. . . . .	212
Table 7.4	Peak activation in the t-maps of the resting-state networks that show group differences. . . . .	217

# LIST OF Acronyms

AAL	automated anatomical labeling
ABIDE	Autism brain imaging data exchange
acIVAMGGMM	adaptive constrained independent vector analysis with multivariate generalized Gaussian mixture model
acICA-MGGMM	adaptive constrained independent component analysis with multivariate generalized Gaussian mixture model
ACG	anterior cingulate and paracingulate gyri
AD	Alzheimer's disease
ADHD	Attention-deficit/hyperactivity disorder
ADNI	Alzheimer's disease neuroimaging initiative
AG	angular gyrus
AI	artificial intelligence
AIC	Akaike information criterion
AN	Auditory network
ASD	Autism spectrum disorder
AUC	area under the curve
BG	Basal Ganglia
BGGMM	bounded generalized Gaussian mixture model
BIC	Bayesian inference criterion
BMGGMM	bounded multivariate generalized Gaussian mixture model
BOLD	blood-oxygen-level-dependent
BoVW	bags-of-visual-words
BSS	blind source separation
CAL	calcarine fissure and surrounding cortex
CNP	neuropsychiatric phenomics
COBRE	Center of Biomedical Research Excellence
CT	computed tomography

CTG	Cardiotocography
DMN	default mode network
DPABI	data processing and analysis for brain imaging
EEG	electroencephalography
EM	expectation maximization
EPI	echo planner imaging
EOG	Electrocardiogram
fALFF	fractional amplitude of low-frequency fluctuations
FDR	false discovery rate
FFG	fusiform gyrus
FHR	fetal heart rate
fMRI	functional magnetic resonance imaging
FNR	false-negative rate
FPR	false-positive rate
FWHM	full-width half-maximum
GMM	Gaussian mixture model
HC	healthy controls
HMM	hidden Markov model
HOS	higher order statistics
ICA	independent component analysis
IVA	independent vector analysis
ICAMGGMM-CS	ICA multivariate generalized Gaussian mixture for common subspace
ICA-BMGGMM	ICA bounded multivariate generalized Gaussian mixture model
ICABMGGMM-CS	ICA bounded multivariate generalized Gaussian mixture for common subspace
IFG	inferior frontal gyrus
IFG-operc	opercular part of IFG
INS	insula
ISI	intersymbol interference

ISR	interferenceto-signal ratio
IVA-GGD	IVA generalized Gaussian distribution
IVAMGGMM-CSS	IVA multivariate generalized Gaussian mixture model for common subspaces
JBSS	Joint blind source separation
LEC	Laplace empirical criterion
LING	lingual gyrus
MCC	Matthews' correlation coefficient
MDL	minimum description length
MEMPR	Prepared Rapid Acquisition Gradient Echo
MFCCs	Mel-frequency cepstral coefficients
MFG	Middle frontal gyrus
MGGMM	multivariate generalized Gaussian mixture
MGGD	multivariate generalized Gaussian distribution
ML	maximum likelihood
MML	minimum message length
MNI	Neurological Institute space
MOG	middle occipital gyrus
mPFC	medial prefrontal cortex
MTG	middle temporal gyrus
NC	normal controls
NPV	negative predictive value
PCA	principal component analysis
PCC	posterior cingulate cortex
PESQ	perceptual evaluation of speech quality
PHG	parahippocampal gyrus
PoCG	Postcentral
PreCG	precentral gyrus
ReHo	regional homogeneity

ROIs	regions of interest signals
ROC	receiver operating characteristics
SAR	signal-to-artefact ratio
SCVs	source component vectors
SDR	signal-to-distortion ratio
SFG	superior frontal gyrus, dorsolateral
SIFT	scale-invariant feature transform
SIR	signal-to-interference ratio
sMRI	structural MRI
SN	Sensorimotor network
SNR	signal-to-noise ratio
SPN	Saliency processing network
SOS	second order statistics
STG	superior temporal gyrus
SZ	Schizophrenic
TE	Echo time
THA	thalamus
TN	true negative
TP	true positive
TR	repetition time
VN	Visual network
WMN	Working memory network

# Chapter 1

## Introduction

### 1.1 Problem statement and motivations

The growing accumulation of large databases has transformed data analysis and modeling into powerful tools capable of revolutionizing various fields of science and engineering. These advancements present significant opportunities across diverse areas, including e-commerce, industry, healthcare, and social media. The widespread applications of data analysis in these domains have driven an increasing demand for the development of sophisticated data mining techniques. Advancements in this field are essential to enhance information retrieval, facilitate knowledge discovery, and enable learning from data patterns to support smart and intelligent decision-making. In recent years, the application of machine learning has grown exponentially, demonstrating its potential to extract valuable information and uncover patterns from complex datasets. The development of machine learning algorithms and techniques has emerged as a highly active area of research over the past few decades, driven by the rapidly increasing demand for artificial intelligence (AI) applications across diverse domains. The exponential growth of machine learning applications and advancements in data mining techniques have unlocked new opportunities for extracting valuable insights from complex datasets across various domains. These developments are particularly relevant in fields like healthcare, where sophisticated methods for information retrieval and knowledge discovery are crucial for interpreting intricate patterns.

One prominent application of data-driven techniques lies in the medical field, where advanced

technologies are increasingly utilized to investigate and understand complex biological systems. However, the analysis of medical data, whether from imaging modalities such as functional magnetic resonance imaging (fMRI), computed tomography (CT), or electroencephalography (EEG), or from non-imaging sources like electronic health records and genetic data, presents significant challenges. These challenges include issues like low signal-to-noise ratios, biological variability, and the sheer volume and complexity of data, which underscore the need for advanced machine learning algorithms to extract meaningful and interpretable patterns. This intersection of data science and medical research holds immense potential to revolutionize healthcare, enabling breakthroughs in diagnostics, treatment planning, and personalized medicine. For instance, machine learning algorithms can identify early markers of diseases such as cancer or Alzheimer's, predict patient responses to therapies, and optimize treatment strategies. Furthermore, these approaches support real-time decision-making in critical care scenarios, enhance drug discovery processes, and enable precision medicine by tailoring interventions to individual patient profiles. By addressing the complexities of medical data, data-driven techniques have the potential to transform healthcare practices and significantly improve patient outcomes.

Beyond healthcare, machine learning models have also demonstrated significant potential in domains such as human action recognition and energy disaggregation, both of which face unique challenges that necessitate innovative solutions. In human action recognition, machine learning techniques are employed to analyze data from video, motion sensors, or wearable devices to identify and classify human activities. Despite its broad applications—including improved safety in industrial settings, enhanced human-computer interaction, and advancements in smart surveillance and assistive technologies for the elderly or disabled—this field presents considerable challenges. These include dealing with variations in lighting, occlusion of subjects, differences in perspectives, and the diversity of human behaviors and movement patterns. Additionally, sensor noise and the need for real-time processing add further complexity to the development of robust and accurate models.

Similarly, energy disaggregation—analyzing aggregate energy consumption data to disaggregate it into individual appliance-level usage—offers transformative benefits but comes with its own

set of challenges. These include the high variability of energy consumption patterns across households, appliances with overlapping consumption profiles, and the lack of labeled training data for many real-world scenarios. Moreover, achieving accurate disaggregation often requires balancing model complexity with computational efficiency to ensure scalability for widespread adoption. Despite these hurdles, successful energy disaggregation enables consumers to monitor and optimize their energy usage, supports energy efficiency initiatives, and helps utilities improve grid management and demand forecasting.

Given these complexities, there is a critical need for robust, data-driven methods capable of effectively analyzing diverse and complex datasets while remaining resilient to the inherent confounding factors. In healthcare, such methods are essential for extracting meaningful patterns from noisy and heterogeneous data, enabling breakthroughs in diagnostics, treatment planning, and personalized medicine. Similarly, in human action recognition, advanced data-driven approaches are vital to address challenges like sensor noise, variations in lighting and perspective, and the diversity of human behaviours, thereby facilitating applications in safety, assistive technologies, and smart surveillance. In energy disaggregation, these methods play a pivotal role in distinguishing overlapping appliance-level usage patterns from aggregate energy data, overcoming issues such as variability across households and limited labeled datasets. By addressing these unique challenges, data-driven approaches can enhance our understanding of the underlying structures within these datasets and drive advancements across fields, ultimately enabling smarter, more efficient, and impactful solutions.

## 1.2 Thesis Contributions

In this dissertation, we address the aforementioned challenges through the following contributions:

- ✍ **Bounded multivariate generalized Gaussian mixture model using ICA and IVA.**

This work proposes a bounded multivariate generalized Gaussian mixture model (BMGGMM) integrated with independent component analysis (ICA), relaxing strict source independence for more flexible modeling. Extending to multi-source datasets like fMRI and electroencephalogram (EEG), this work introduces IVA-BMGGMM, a generalized framework that enhances ICA-based techniques, improving adaptability and performance in complex scenarios.

This work is published in *Pattern Analysis and Applications Journal* [2].

✍ **ICA and IVA bounded multivariate generalized Gaussian mixture-based hidden Markov models.**

In this work, we integrate ICA and independent vector analysis (IVA) with a bounded multivariate generalized Gaussian mixture model (ICA-BMGGMM) into the hidden Markov model (HMM) framework. To address the limitation of ICA's assumption of source independence, we combine IVA and IVA-BMGGMM with HMM to enhance modeling capabilities. We validate the proposed models through applications in human action recognition, speech recognition, and energy disaggregation.

This work is published in *Engineering Applications of Artificial Intelligence* [3].

✍ **Adaptive constrained IVAMGGMM: Application to mental disorders detection.**

In this work, we propose the constrained ICA-MGGMM, an ICA-based multivariate Gaussian mixture model that relaxes ICA's independence assumption. Additionally, we introduce the adaptive constrained ICA-MGGMM (acICA-MGGMM) to manage the relationship between reference signals and estimated sources. While IVA can capture patterns from fMRI data, its performance declines with increasing datasets and weak correlations. To address this, we present cIVA-MGGMM, incorporating reference signals for high-dimensional data, and its adaptive version (acIVA-MGGMM) using a full covariance matrix. These models integrate prior information into ICA/IVA to overcome their limitations in high-dimensional data.

This work is published in *IEEE Transactions on Emerging Topics in Computational Intelligence* [4].

✍ **Novel approach for ECG separation using adaptive constrained IVABMGGMM.**

In this work, we introduce constrained independent vector analysis integrated with the bounded multivariate generalized Gaussian mixture model (cIVABMGGMM) to address the limitations of ICA for multivariate data, along with its adaptive version, acIVABMGGMM, designed to reduce constraints. The acIVABMGGMM uses a full covariance matrix to account for feature correlations, effectively overcoming the challenges of ICA and IVA in multivariate data analysis. This adaptive framework combines data-driven adaptability with the ability to handle noise and artifacts in model-based approaches.

This work is published in *Digital Signal Processing* [5]. Part of this work is published in the *International Conference on Advanced Technologies for Signal and Image Processing (AT-SIP)* [6] and presented in part at the *31st International Conference on Neural Information Processing (ICONIP 2024)* [7] [accepted].

#### ✍ **Identifying the common subspaces regions using independent component analysis and graph-theoretical analysis.**

This work introduces ICAMGGMM-CS, a model for common subspace estimation based on the multivariate generalized Gaussian mixture distribution, enabling the identification of common patterns and enhancing diagnostic precision and scalability. To further improve robustness in source separation, the framework incorporates a bounded indicator function, forming the ICABMGGMM-CS model, which combines data-driven adaptability with model-based stability to manage noise and artifacts. Additionally, a multivariate graph-theoretical framework summarizes components into meaningful, interpretable features.

This work is submitted to *IEEE transaction on Artificial Intelligence*.

#### ✍ **Homogeneous Subgroup Identification in Resting State fMRI using Common Subspace Independent Vectors Analysis.**

In this work, we propose the IVABMGGMM-CSS model, a novel common subspace IVA framework based on the bounded multivariate generalized Gaussian mixture distribution, to address the limitations of joint blind source separation in capturing latent structures across multiple datasets, specifically its inefficiency with high-dimensional data. The model captures common structures shared across datasets while preserving subject-specific variability

to manage noise and artifacts in high-dimensional data.

This work is submitted to *IEEE Transactions on Signal Processing*.

### 1.3 Thesis Outline

The rest of the thesis is organized as follows:

- ❑ Chapter 2 introduces the Bounded multivariate generalized Gaussian mixture model using ICA and IVA.
- ❑ In Chapter 3, we present ICA and IVA bounded multivariate generalized Gaussian mixture based hidden Markov models.
- ❑ Chapter 4 displays the adaptive constrained IVAMGGMM for application to mental disorders detection.
- ❑ In Chapter 5, we discuss the novel approach for ECG separation using adaptive constrained IVABMGGMM and ICABMGGMM.
- ❑ Chapter 6 presents the identification of the common subspaces regions using independent component analysis and graph-theoretical analysis.
- ❑ In Chapter 7, we discuss the homogeneous subgroup identification in resting state fMRI using common subspace independent vectors analysis.
- ❑ Chapter 8 concludes the thesis with discussion remarks and future insights.

## Chapter 2

# Bounded Multivariate Generalized Gaussian Mixture Model using ICA and IVA

A bounded multivariate generalized Gaussian mixture model with a full covariance matrix is proposed for modeling data in a bounded support region. For model selection, we propose the minimum message length criterion. Furthermore, we propose a bounded multivariate generalized Gaussian mixture model with independent component analysis. By employing the mixture model with independent component analysis, the assumed independence of the sources can be relaxed. For data with multiple sources such as functional magnetic resonance imaging and electroencephalogram databases, we propose the bounded multivariate generalized Gaussian mixture model with independent vector analysis as a generalized technique for the independent component analysis-based one. For a more insightful model analysis, we validate the proposed mixture model in data clustering through a variety of medical applications. We also propose the application of the independent component analysis-based model in speech (Romanian read-speech corpus), Electrocardiogram, and Electroencephalogram databases. For validation of the independent vector analysis-based model performance, different medical and speech databases are used. The results presented demonstrate the effectiveness of the proposed approaches for modeling different types of data.

## 2.1 Introduction

Finite mixtures are flexible and powerful probabilistic data modeling tools which are often employed for signal and image processing [8], computer vision [9], and machine learning [10]. The three primary difficulties in mixture modeling are determining the probability density function, estimating the parameters, and determining the number of clusters. The Gaussian mixture model (GMM) is the most popular technique for statistical data modeling, used in various applications. However, the Gaussian distribution has several limitations, such as the rigidity of its shape. For that, the generalized Gaussian mixture model (GGMM) was proposed to improve data modeling, given that it can provide more flexibility to fit data better than the conventional Gaussian [11]. The GGMM comprises the Laplacian, and Gaussian distributions as special examples, as well as the uniform distribution as a limiting case [12], and has been used to solve a variety of challenging problems [8, 11, 13, 14]. Several studies have shown that the GGMM can be a good alternative to the GMM due to the flexibility of its shape, which allows modeling a wide range of non-Gaussian signals [15, 16, 17]. For more flexibility, the multivariate generalized Gaussian mixture model (MG-GMM) was proposed to model non-Gaussian data with correlated features using a full covariance matrix [18, 19]. Despite the outstanding clustering results obtained with the aforementioned models, the observed data in many real applications fall within bounded support regions [20]. To overcome the issues associated with unbounded mixture models, the bounded generalized Gaussian mixture model (BGGMM) has been proposed in [20] and [21]. BGGMM has proven to be successful in many speech processing applications [22, 23] for its flexibility to fit different shapes of data. However, the BGGMM is not able to deal with correlated features as the covariance matrix is taken as diagonal. Motivated by the aforementioned observations, we propose a new bounded model, which is an extension to the MGGMM to deal with bounded support regions. The BMGGMM takes into account correlated features using a full covariance matrix and has the flexibility to fit a large variety of non-Gaussian bounded support data shapes.

In terms of parameter estimation, the most widely used strategy for mixture model is based on maximizing the likelihood function using the expectation maximization (EM) framework [24]. To apply the EM method, an appropriate number of clusters must be predefined; otherwise, the

EM algorithm would produce unsatisfactory results. Therefore, determining the number of optimal components that best characterize the data is crucial for the mixture model. Different techniques for automatically determining the number of clusters have been proposed in literature [25, 26] such as Akaike information criterion (AIC) [27], minimum description length (MDL) [28], and Laplace empirical criterion (LEC) [10]. However, the number of clusters may be overestimated or underestimated using these criteria. Furthermore, the minimum message length (MML) criterion was introduced in [29] and [30] as a technique to find the optimal number of mixture components. Therefore, in order to accurately find the number of clusters, we propose an MML criterion for choosing the best number of components that can describe the data using BMGGMM.

In our research work, we are interested in signal processing applications. One of the most challenges in these applications is separating the signal from the noise. This problem can be easily solved with blind source separation (BSS). BSS is one of the most active research areas due to its applicability to a wide range of problems, especially when little information is known about the observed data. BSS techniques are used in a variety of applications including audio separation [31], target detection in video sequences [32], and medical imaging data analysis [33]. ICA was effectively applied to solve the BSS problem instantaneously under the assumption of statistical independence of the latent sources. The main assumption of ICA is that the observations are linear, statistically independent mixtures of the sources, which will be separated by maximizing output independence. By employing a mixture model, this assumption can be relaxed; this is known as ICA mixture model [34]. The mixture model of the ICA is supposed to classify the observed data into mutually exclusive classes and provide a linear combination of independent sources with components of the model [35]. Several improvements have been proposed for ICA mixture models applied to various applications, including segmentation, image enhancement, and BSS [36]. ICA extended by BGGMM has been proposed to overcome unbounded mixture model limitations [37, 38]. In this work, we consider the statistical independence and the sparsity to propose the integration of BMGGMM into the ICA framework, addressing the limitations of traditional ICA. ICA-BMGGMM allows for the modelling of classes with non-Gaussian structure and the identification of statistically significant structures in data. The algorithm for learning the model's parameters employs gradient ascent to maximize the log likelihood function.

ICA was limited by permutation and scaling issues, which are detailed in [39]. Hence, the independent vector analysis (IVA) approach is an extension of the ICA that was proposed to reduce the permutation problem that occurs in most of the BSS algorithms [40]. The IVA technique is based on a dependence model that preserves inter-frequency dependencies within each source vector. In a frequency domain, dependent sources are organized together as a multivariate variable where each source's inter-frequency dependencies depend on the prior source signal. In the ICA technique, the independence is calculated separately for each frequency component in each frequency bin. In ICA conventional algorithms, sources were assumed independent. In contrast, the IVA technique formulates the problem by not considering the independence of the frequency bins but by accepting that dependencies exist between frequency bins. Therefore, it measures dependency over the entire multivariate sources and preserves a higher order of frequency inter-dependence and structure. The IVA method employs the score function, and its shape is critical to the algorithm's performance. All sources were modeled in the original IVA algorithm through the identical Laplacian distributions [40]. Different sources can have distinct statistical features, and the best solution may not be to model all the sources with the same distribution. Among the IVA solutions, IVA-Gaussian (IVA-G) takes advantage of linear dependencies but ignores higher-order statistics [41]. IVA with generalized Gaussian distribution (IVA-GGD) is a more general IVA implementation that considers both second and higher-order statistics. However, IVA-GGD employs a fixed set of shape parameter values [42]. We propose IVA-BMGGMM for multivariate data such as fMRI and EEG. IVA-BMGGMM is an algorithm that accounts for both higher order statistics (HOS) and second order statistics (SOS) and provides a more suitable model for multivariate non-Gaussian data. Furthermore, the IVA-BMGGMM assumes second-order correlation within support component vectors, which provides more flexibility for extracting knowledge from the data.

In a nutshell, the main purpose of this research work is to improve the clustering capability of the existing models such as GMM, GGMM, MGGMM, and BGGMM. The BMGGMM model can model multivariate, non-Gaussian data in bounded regions. We also propose to integrate the MML with the BMGGMM to automate the determination of the number of clusters which will improve the ability to model data in different real-world applications. According to the literature, several mixture models were used to improve the ICA model's performance in the separation of mixture

sources. To that end, we propose integrating BMGGMM into the ICA model to improve its separation capability. To address the ICA limitation in multivariate data, we propose incorporating BMGGMM into the IVA model to improve further its ability to separate mixture sources.

The BMGGMM will be explored in several clustering applications, including medical diagnosis of skin cancer, Alzheimer, diabetes, and stroke diseases. In this work, the MML criterion for model selection is developed for BMGGMM and verified through different medical datasets that are challenging to cluster and model selection methodologies. Similar experiments are carried out using different techniques chosen from the literature to validate the proposed model selection criterion (MML). We consider the presented ICA-BMGGMM model in data clustering applications including speech, EEG, and ECG databases to insight into the analysis of the model. For heartbeat sound separation, SUFHSDB dataset is chosen. Due to the limitation of ICA in multivariate datasets, we also propose the IVA-BMGGMM to overcome these limitations. We evaluate the performance of the proposed model (IVA-BMGGMM) using different applications such as speech signal, EEG, ECG, and schizophrenic fMRI data.

The rest of this Chapter is organized as: In Section 2.2, we describe the BMGGMM and the parameters estimation. Section 2.3 introduces the model selection using the MML criterion adopted to the BMGGMM. Next, we explain the extended BMGGMM using ICA and IVA algorithms in Section 3.2.1 and Section 3.2.2, respectively. In Section 2.6, we report and discuss the results of the proposed models (BMGGMM, ICA-BMGGMM, and IVA-BMGGMM). Finally, we conclude our Chapter in Section 3.4.

## **2.2 Bounded multivariate generalized Gaussian mixture model**

### **2.2.1 Model definition**

Bounded support mixtures were proposed to tackle the problems posed by unbounded distributions when the data are defined in a bounded support range. In this Section, we present a bounded multivariate generalized Gaussian distribution (BMGGMM) and the estimation of its parameters using a maximum likelihood estimation via the EM algorithm. Consider a  $D$ -dimensional random variable  $\mathbf{x}_n = (\mathbf{x}_1, \dots, \mathbf{x}_D)$ , the probability distribution function for the finite mixture model can be

written as:

$$p(\mathbf{x}|\Theta) = \sum_{j=1}^M p(\mathbf{x}|\xi_j)p_j \quad (2.1)$$

where  $p_j$  is the prior probability for each cluster with the constraint that  $\forall_j, p_j \geq 0; \sum_{j=1}^M p_j = 1$ . Here  $\Theta$  is the complete set of parameters,  $\xi_j$  is the mixture model parameters of  $j$ th component, and  $M$  is the number of components in the mixture model. Given the data  $\mathcal{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ , the likelihood is written as:

$$p(\mathcal{X}|\Theta) = \sum_{n=1}^N \sum_{j=1}^M p(\mathbf{x}_n|\xi_j)p_j \quad (2.2)$$

To develop the BMGGMM, an indicator function defining the boundary conditions must be presented, which is defined as:

$$H(\mathbf{x}_n|j) = \begin{cases} 1 & \text{if } \mathbf{x}_n \in \Omega_j \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

where  $\Omega_j$  represents the bounded support region. Using the indicator function, we present the bounded distribution given as:

$$p(\mathbf{x}_n|\xi_j) = \frac{f(\mathbf{x}_n|\xi_j)H(\mathbf{x}_n|j)}{\int_{\Omega_j} f(\mathbf{u}|\xi_j)du} \quad (2.4)$$

In this work, we consider the term  $f(\mathbf{x}_n|\xi_j)$  to be the multivariate generalized Gaussian distribution (MGGD) for each component of the mixture model defined by:

$$f(\mathbf{x}|\Theta) = \frac{\Gamma\left(\frac{D}{2}\right)}{\pi^{\frac{D}{2}} \Gamma\left(\frac{D}{2\beta_j}\right) 2^{2\beta_j} (\mathbf{c}_j)^{\frac{D}{2}} |\Sigma_j|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2(\mathbf{c}_j)^{\beta_j}} \left[ (\mathbf{x} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j) \right]^{\beta_j} \right\} \quad (2.5)$$

where  $\mathbf{x}_n \in \mathbb{R}^D$ ,  $\boldsymbol{\mu}_j$ ,  $\Sigma_j$  and  $\beta_j$  are mean, covariance matrix, and shape parameter, respectively, and  $\mathbf{c}_j$  is the scale parameter defined as:

$$\mathbf{c}_j = \left( \frac{\beta_j}{DN} \sum_{n=1}^N (\mathbf{x}_n^T \Sigma_j^{-1} \mathbf{x}_n)^{\beta_j} \right)^{\frac{1}{\beta_j}} \quad (2.6)$$

## 2.2.2 Parameter estimation of the bounded multivariate generalized Gaussian mixture model

We apply the maximum likelihood (ML) approach via EM algorithm for estimating the BMG-GMM parameters. The complete log-likelihood function is given by:

$$\mathcal{L}(\mathcal{X}, Z, \Theta) = \sum_{n=1}^N \sum_{j=1}^M \log \left( p(\mathbf{x}_n | \xi_j) p_j \right)^{Z_{nj}} \quad (2.7)$$

where  $Z = \{Z_1, \dots, Z_N\}$ ,  $Z_{nj}$  is a membership indicator that encodes the membership of each observation to its belonging mixture component. By replacing each  $Z_{nj}$  by its expectation, the posterior probability can be written as:

$$\hat{Z}_{nj} = \frac{p(\mathbf{x}_n | \xi_j) p_j}{\sum_{j=1}^M p(\mathbf{x}_n | \xi_j) p_j} \quad (2.8)$$

Equation (2.7) can be maximized by taking the gradient of the log-likelihood with respect to  $\boldsymbol{\mu}_j$ ,  $\Sigma_j$  and  $\beta_j$ .

### Mixing parameter estimation

While estimating  $p_j$ , a Lagrange multiplier is used to guarantee that the constraints  $p_j \geq 0$  and  $\sum_{j=1}^M p_j = 1$  are met. Then, the complete log-likelihood is written as:

$$\mathcal{L}(\Theta, Z, \mathcal{X}) = \sum_{n=1}^N \sum_{j=1}^M \hat{Z}_{nj} \log \left( p(\mathbf{x}_n | \xi_j) p_j \right) + \Lambda \left( 1 - \sum_{j=1}^M p_j \right) \quad (2.9)$$

where  $\Lambda$  is the Lagrangian multiplier. Taking the gradient of the log-likelihood function with respect to  $p_j$ , we get the estimated value of  $p_j$  as:

$$\hat{p}_j = \frac{1}{N} \sum_{n=1}^N p(j|\mathbf{x}_n) \quad (2.10)$$

where  $N$  is the number of observations.

### Mean estimation

Maximizing the log-likelihood given in Equation (2.7) with respect to  $\boldsymbol{\mu}_j$ , we estimate the update of  $\boldsymbol{\mu}_j$ . Derivation details are given in Appendix A; by using these derivations,  $\boldsymbol{\mu}_j$  can be estimated as follows:

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_{n=1}^N \hat{Z}_{nj} \left[ \mathbf{x}_n - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j)(\mathbf{u} - \boldsymbol{\mu}_j) du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \right]}{\sum_{n=1}^N \hat{Z}_{nj}} \quad (2.11)$$

The expectation term  $\int_{\Omega_j} f(\mathbf{u}|\xi_j)(\mathbf{u} - \boldsymbol{\mu}_j) du$  of function  $(\mathbf{u} - \boldsymbol{\mu}_j)$  can be approximated as:

$$\int_{\Omega_j} f(\mathbf{u}|\xi_j)(\mathbf{u} - \boldsymbol{\mu}_j) du \approx \frac{1}{V} \sum_{v=1}^V (\mathbf{s}_{jv} - \boldsymbol{\mu}_j) H(\mathbf{s}_{jv}|j) \quad (2.12)$$

where  $\mathbf{s}_{jv}$  represents the random variable that is derived from the generalized Gaussian distribution for the  $j$ th component of the mixture model.  $V$  is the number of random variables  $\mathbf{s}_{jv}$ . Similarly, the term  $\int_{\Omega_j} f(\mathbf{u}|\xi_j) du$  can be approximated as:

$$\int_{\Omega_j} f(\mathbf{u}|\xi_j) du \approx \frac{1}{V} \sum_{v=1}^V H(\mathbf{s}_{jv}|j) \quad (2.13)$$

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_{n=1}^N \hat{Z}_{nj} \left[ \mathbf{x}_n - \frac{\sum_{v=1}^V (\mathbf{s}_{jv} - \boldsymbol{\mu}_j) H(\mathbf{s}_{jv}|j)}{\sum_{v=1}^V H(\mathbf{s}_{jv}|j)} \right]}{\sum_{n=1}^N \hat{Z}_{nj}} \quad (2.14)$$

### Covariance estimation

The new value of the covariance matrix  $\Sigma_j$  can be estimated by maximizing the log-likelihood function given in Equation (2.7) with respect to  $\Sigma_j$ . The computation derivative of log-likelihood with respect to  $\Sigma_j$  is given in Appendix A.

$$\hat{\Sigma}_j = \frac{1}{\sum_{n=1}^N \hat{Z}_{nj}} \sum_{n=1}^N \hat{Z}_{nj} \left\{ \frac{\beta_j}{(\mathbf{c}_j)^{\beta_j}} \sum_{n=1}^N \mathbf{y}^{\beta_j-1}(\mathbf{x}_n - \boldsymbol{\mu}_j)(\mathbf{x}_n - \boldsymbol{\mu}_j)^T - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left\{ \frac{\beta_j}{(\mathbf{c}_j)^{\beta_j}} \sum_{n=1}^N \mathbf{y}^{\beta_j-1}(\mathbf{u}_n - \boldsymbol{\mu}_j)(\mathbf{u}_n - \boldsymbol{\mu}_j)^T - \Sigma_j \right\} du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \right\} \quad (2.15)$$

where  $\mathbf{y} = (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j)$ . The term  $\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left[ \frac{\beta_j}{(\mathbf{c}_j)^{\beta_j}} \sum_{n=1}^N \mathbf{y}^{\beta_j-1}(\mathbf{u} - \boldsymbol{\mu}_j)(\mathbf{u} - \boldsymbol{\mu}_j)^T - \Sigma_j \right] du$  in Equation (2.15) can be approximated as:

$$\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left[ \frac{\beta_j}{(\mathbf{c}_j)^{\beta_j}} \sum_{n=1}^N \mathbf{y}^{\beta_j-1}(\mathbf{u}_n - \boldsymbol{\mu}_j)(\mathbf{u}_n - \boldsymbol{\mu}_j)^T - \Sigma_j \right] du \approx \sum_{v=1}^V \left[ \frac{\beta_j}{(\mathbf{c}_j)^{\beta_j}} \sum_{n=1}^N \mathbf{y}^{\beta_j-1}(\mathbf{s}_{jv} - \boldsymbol{\mu}_j)(\mathbf{s}_{jv} - \boldsymbol{\mu}_j)^T - \Sigma_j \right] H(\mathbf{s}_{jv}|j) \quad (2.16)$$

$$\hat{\Sigma}_j = \frac{1}{\sum_{n=1}^N \hat{Z}_{nj}} \sum_{n=1}^N \hat{Z}_{nj} \left\{ \frac{\beta_j}{(\mathbf{c}_j)^{\beta_j}} \sum_{n=1}^N \mathbf{y}^{\beta_j-1}(\mathbf{x}_n - \boldsymbol{\mu}_j)(\mathbf{x}_n - \boldsymbol{\mu}_j)^T - \frac{\sum_{v=1}^V \left[ \frac{\beta_j}{(\mathbf{c}_j)^{\beta_j}} \sum_{n=1}^N \mathbf{y}^{\beta_j-1}(\mathbf{s}_{jv} - \boldsymbol{\mu}_j)(\mathbf{s}_{jv} - \boldsymbol{\mu}_j)^T - \Sigma_j \right] H(\mathbf{s}_{jv}|j)}{\sum_{v=1}^V H(\mathbf{s}_{jv}|j)} \right\} \quad (2.17)$$

### Shape parameter estimation

The new value of the shape parameter  $\beta_j$  can be estimated by maximizing the log-likelihood function given in Equation (2.7) with respect to  $\beta_j$ . The computation derivative of log-likelihood

with respect to  $\beta_j$  is given in Appendix A, and by following these derivations, the  $\beta_j$  can be estimated as:

$$\hat{\beta}_{m+1} = \hat{\beta}_j - \frac{\alpha(\hat{\beta}_j)}{\hat{\alpha}(\hat{\beta}_j)} \quad (2.18)$$

$$\alpha(\hat{\beta}_j) = \sum_{n=1}^N \hat{Z}_{nj} \left\{ Q - \frac{y^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left[ \log(\mathbf{y}) - \log(\mathbf{c}_j) \right] - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left\{ Q - \frac{y_u^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left[ \log(\mathbf{y}_u) - \log(\mathbf{c}_j) \right] \right\} du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \right\} \quad (2.19)$$

where  $Q = \frac{1}{\beta_j} + \frac{D}{2\beta_j} \psi\left(\frac{D}{2\beta_j}\right) + \frac{D}{2\beta_j^2} \log(2)$ ,  $\mathbf{y} = (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j)$ ,  $\mathbf{y}_u = (\mathbf{u}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{u}_n - \boldsymbol{\mu}_j)$  and  $\psi$  is the digamma function. The above Equation (2.19) can be approximated as:

$$\alpha(\hat{\beta}_j) = \sum_{n=1}^N \hat{Z}_{nj} \left\{ Q - \frac{\mathbf{y}^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left[ \log(\mathbf{y}) - \log(\mathbf{c}_j) \right] - \frac{\sum_{v=1}^V \left\{ Q - \frac{\mathbf{y}_s^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left[ \log(\mathbf{y}_s) - \log(\mathbf{c}_j) \right] \right\} H(\mathbf{s}_{j_v}|j)}{\sum_{v=1}^V H(\mathbf{s}_{j_v}|j)} \right\} \quad (2.20)$$

where  $\mathbf{y}_s = (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j)$ . The complete learning procedure for the BMGGMM model is illustrated in Algorithm 1.

## 2.3 Model selection using MML criterion for BMGGMM

Several methods of model selection have been presented in order to estimate the number of components of a mixture model [43]. We propose the minimum message length (MML) criterion for model selection in BMGGMM, which is based on information theory. The optimal number of

---

**Algorithm 1** BMGGMM

---

- 1: Dataset  $\mathcal{X}$
  - 2: Output:  $\Theta$ .
  - 3: Initialization Algorithm
  - 4: Apply K-means to obtain mean and covariance for each component. Set  $\beta_j = 2$ .
  - 5: **if** change in likelihood  $>$  threshold **then**
  - 6:     **E-step:**
  - 7:     Compute posterior probabilities using Equation (2.8)
  - 8: **else**
  - 9:     **M-step:**
  - 10:     Update the mixing parameter  $\hat{p}_j$  using Equation (2.10)
  - 11:     Update  $\mu_j$  using EM algorithm: Equation (2.14)
  - 12:     Update  $\Sigma_j$  using Equation (2.17)
  - 13:     Update  $\beta_j$  using Newton-Raphson method: Equation (2.18), (2.19) and (2.20).
  - 14: **end if** when the likelihood is converged
  - 15: Return the model's parameters  $\hat{\mu}_j, \hat{\beta}_j$  and  $\hat{\Sigma}_j$ .
- 

components in the mixture is determined by applying MML using the following formula:

$$MML(M) \approx -\log \left[ p(\Theta_M) \right] - \mathcal{L}(\Theta_M, Z, \mathcal{X}) + \frac{1}{2} \log |F(\Theta_M)| + \frac{N_p}{2} + \frac{N_p}{2} \log(M_{N_p}) \quad (2.21)$$

where  $p(\Theta_M)$  is the prior probability,  $|F(\Theta_M)|$  is the determinant of the Fisher information matrix,  $N_p = M(3D + 1)$  is the number of free parameters, and  $M_{N_p}$  is the optimal quantization lattice constant, which is equal to  $1/12$  for  $N_p = 1$  [26]. To use Equation (2.21), we derive the  $p(\Theta_M)$  and  $F(\Theta_M)$  as a first step.

### 2.3.1 Derivation of the prior distribution

In the mixture model, the parameters of different classes are assumed independent. Thus, we present the prior distribution as follows:

$$p(\Theta) = p(\boldsymbol{\pi})p(\boldsymbol{\mu})p(\boldsymbol{\beta})p(\boldsymbol{\Sigma}) \quad (2.22)$$

where  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_M)$ ,  $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_M)$ ,  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_M)$  and  $\boldsymbol{\pi} = \{(\boldsymbol{p}_1, \dots, \boldsymbol{p}_M) : \sum_{j=1}^M p_j = 1\}$ . Using a uniform prior, the prior of the mixing weights is represented as:

$$p(\boldsymbol{\pi}) = (M - 1)! \quad (2.23)$$

For the parameter  $\boldsymbol{\mu}$  and  $\Sigma$ , we adopt the prior distribution as described in [44]. Thus, the joint prior distribution for the mean  $\boldsymbol{\mu}$  and the covariance  $\Sigma$  is as follows:

$$p(\boldsymbol{\mu}, \Sigma) \propto \prod_{j=1}^M |\Sigma|^{-M \frac{D+1}{2}} \quad (2.24)$$

We use a uniform distribution  $U[0, h]$  for the shape parameter  $\boldsymbol{\beta}$ , where  $h$  is the maximum value. The prior for  $\boldsymbol{\beta}$  is then given by:

$$p(\boldsymbol{\beta}) = \frac{1}{h^M} \quad (2.25)$$

By substituting Equations (2.23) and (2.25) in Equation (2.22), we get:

$$p(\Theta) \propto \frac{|\Sigma|^{(-M \frac{D+1}{2})} (M-1)!}{h^M} \quad (2.26)$$

### Derivation of the Fisher information matrix

The Fisher information is approximated by the determinant of the Hessian matrix. Fisher information for a mixture leads to a complex analytical form of MML. We approximate the Fisher information as follows:

$$|F(\Theta)| = |F(\boldsymbol{\pi})| \prod_{j=1}^M |F(\boldsymbol{\mu}_j)| |F(\boldsymbol{\beta}_j)| |F(\Sigma_j)| \quad (2.27)$$

where  $|F(\boldsymbol{\pi})|$ ,  $|F(\boldsymbol{\mu}_j)|$ ,  $|F(\boldsymbol{\beta}_j)|$  and  $|F(\Sigma_j)|$  are determinants of Fisher information matrices for  $\boldsymbol{\pi}$ ,  $\boldsymbol{\mu}_j$ ,  $\Sigma_j$ , and  $\boldsymbol{\beta}_j$ , respectively. Following [11, 45], the Fisher information for the mixing weights is given by:

$$|F(\boldsymbol{\pi})| = \frac{N^{M-1}}{\sum_{j=1}^M p_j} \quad (2.28)$$

where  $N$  is the number of observations. The Hessian matrices for parameters  $\boldsymbol{\mu}_j$ ,  $\Sigma_j$ , and  $\boldsymbol{\beta}_j$  are given by:

$$F(\boldsymbol{\mu}_j) = -\frac{\partial^2 \mathcal{L}(\Theta, Z, \mathcal{X})}{\partial \mu_{jm_1} \partial \mu_{jm_2}} \quad (2.29)$$

$$F(\Sigma_j) = -\frac{\partial^2 \mathcal{L}(\Theta, Z, \mathcal{X})}{\partial \Sigma_{jm_1} \partial \Sigma_{jm_2}} \quad (2.30)$$

$$F(\boldsymbol{\beta}_j) = -\frac{\partial^2 \mathcal{L}(\Theta, Z, \mathcal{X})}{\partial \beta_j^2} \quad (2.31)$$

where  $m_1, m_2 \in \{1, \dots, D\}$ . Derivatives are calculated with regard to these parameters, which gives the required Fisher information matrices detailed in Appendix A.

## 2.4 ICA mixture of bounded multivariate generalized Gaussian distributions

In the ICA mixture model, observed data is supposed to come from a mixture model and to be categorized in mutually exclusive classes [46], where each data vector  $\mathbf{x}_n$  can be represented as:

$$\mathbf{x}_n = \mathbf{A}_j \mathbf{s}_j + \mathbf{b}_j \quad (2.32)$$

where  $\mathbf{A}_j$  is a mixing matrix,  $\mathbf{s}_j$  is the source vector, and  $\mathbf{b}_j$  is the bias vector for each mixture component  $j$  [47, 48, 36]. ICA-BMGGMM can be defined by adopting Equation (2.32) in Equation (2.4).

### 2.4.1 Parameter estimation using ICA and gradient ascent

The essential assumption of parameter estimation using ICA is zero mean and unit variance. For each class, the gradient of complete data log-likelihood for parameters is given as:

$$\nabla_{\mathbf{A}_j} \mathcal{L}(\Theta, Z, \mathcal{X}) = \nabla_{\mathbf{A}_j} \sum_{n=1}^N \sum_{j=1}^M \hat{Z}_{nj} \log \left( p(\mathbf{x}_n | \xi_j) p_j \right) \quad (2.33)$$

where  $\nabla_{\mathbf{A}_j}$  represents gradient with respect to mixing weight, mean, covariance matrix, shape parameter, basis function, and bias vector, respectively.

#### Basis functions estimation

The standard ICA model for the log-likelihood is given by:

$$\log \left( p(\mathbf{x}_n | \xi_j) \right) = \log \frac{p(\mathbf{s}_j)}{|\det \mathbf{A}_j|} \quad (2.34)$$

For each ICA mixture component, the adaptation of basis functions is obtained by maximizing the log-likelihood with respect to basis functions  $\mathbf{A}_j$ , for each mixture model component:

$$\nabla_{\mathbf{A}_j} \mathcal{L}(\Theta, Z, \mathcal{X}) = \nabla_{\mathbf{A}_j} \sum_{n=1}^N \hat{Z}_{nj} \log \left( p(\mathbf{x}_n | \xi_j) \right) \quad (2.35)$$

where  $\nabla_{\mathbf{A}_j}$  represents gradient with respect to basis functions. The parameter  $A_j$  is estimated using gradient ascent as follows:

$$\hat{\mathbf{A}}_j = \mathbf{A}_j + \alpha \left( \hat{Z}_{nj} \mathbf{A}_j \left[ I - 2 \tanh(\mathbf{s}_j) \mathbf{s}_j^T \right] \right) \quad (2.36)$$

where  $\alpha$  represents step size and  $\mathbf{s}_j$  given by:

$$\mathbf{s}_j = \mathbf{A}_j^{-1} \left( \mathbf{x}_n - \mathbf{b}_j \right) \quad (2.37)$$

The complete computation to estimate basis functions is given in Appendix B.

### Bias vector estimation

An estimate of bias vector using gradient ascent is as follows:

$$\hat{\mathbf{b}}_j = \frac{\sum_{n=1}^N \mathbf{x}_n \hat{Z}_{nj}}{\sum_{n=1}^N \hat{Z}_{nj}} \quad (2.38)$$

Estimation steps for mean, covariance, and shape parameters are given in detail in Section 2.2. The

---

#### Algorithm 2 ICA-BMGGMM

---

- 1: Dataset  $\mathcal{X}$
  - 2: Output:  $\Theta$ .
  - 3: Initialization Algorithm
  - 4: Apply K-means to obtain mean and covariance for each component. Set  $\beta_j=2$ .
  - 5: **if** Change in likelihood > threshold **then**
  - 6:     **E-step:**
  - 7:     Compute posterior probabilities using Equation (2.8)
  - 8: **else**
  - 9:     **M-step:**
  - 10:     **Start ICA algorithm**
  - 11:     Update the basis functions  $\mathbf{A}_j$  using Equation (2.36)
  - 12:     Update the bias vector  $\mathbf{b}_j$  using Equation (2.38)
  - 13:     **End ICA algorithm**
  - 14:     Update the mixing parameter  $\hat{p}_j$  using Equation (2.10)
  - 15:     Update  $\mu_j$  using EM algorithm: Equation (2.14)
  - 16:     Update  $\Sigma_j$  using Equation (2.17)
  - 17:     Update  $\beta_j$  using Newton-Raphson method: Equation (2.18), (2.19) and (2.20)
  - 18: **end if** when the likelihood is converged
  - 19: Return the model's parameters  $\hat{\mathbf{A}}_j, \hat{\mathbf{b}}_j, \hat{\mu}_j, \hat{\beta}_j$  and  $\hat{\Sigma}$ .
- 

complete procedure for the ICA mixture model is shown in Algorithm 2.

## 2.5 IVA mixture of multivariate generalized Gaussian distributions

In several applications, it is necessary to jointly analyze not just a single but several sources, with dependence among them. Examples include processing the medical data, such as multi-subject fMRI and EEG, and multiple frequency bins while addressing the problem of a convolutive ICA in the frequency domain. IVA is a natural extension of ICA for the simultaneous extraction of independent components from multiple sources, which creates a framework to avoid the blind source separation permutation problem [49].

In IVA, components within a source are independent, while corresponding components extracted from different mixture sources are maximally dependent. Thus, corresponding component dependence across multiple sources and component independence within a single source are maximized simultaneously as described in Fig. 2.1. This cannot be achieved by running separate ICAs on each source [50].

In IVA, each mixture source  $\mathbf{x}_n^{[m]}$ ,  $m = 1, \dots, M$  is a linear mixture of  $N$  independent sources.

$$\mathbf{x}_n^{[m]} = \mathbf{A}_j^{[m]} \mathbf{s}_j^{[m]}, \quad m = 1, \dots, M \quad (2.39)$$

where  $\mathbf{A}_j^{[m]}$  is an invertible mixing matrix and  $\mathbf{s}_j^{[m]}$  is a vector of sources for  $m$ th mixture sources. The source vector  $\mathbf{s}_j^{[m]}$  is defined by concatenating  $n$ th sources from each  $M$  mixture source. The goal of the IVA algorithm is to estimate  $M$  mixing matrices that yield source estimates  $\mathbf{y}^{[m]} = \mathbf{W}^{[m]} \mathbf{x}^{[m]}$ . This is achieved by minimizing the mutual information cost function, which is equivalent to maximizing the likelihood function as:

$$J_{IVA} = \sum_{n=1}^N H(\mathbf{y}_n) - \sum_{m=1}^M \log|\det(\mathbf{W}^{[m]})| - H(\mathbf{x}) \quad (2.40)$$

where  $H(\mathbf{y}_n)$  is the differential entropy of  $n$ th source vectors and  $H(\mathbf{x})$  is considered as constant. By definition, the term  $H(\mathbf{y}_n)$  is given by  $\sum_{m=1}^M H(\mathbf{y}_n^{[m]}) - I(\mathbf{y}_n)$ , where  $I(\mathbf{y}_n)$  denotes the mutual information within the  $t$ th source component vector (SCV). The optimization of the cost function takes into account both the independence within the dataset (via the entropy term) and the dependence across datasets (via the mutual information term). The gradient of the cost function is given by:

$$\frac{\partial J_{IVA}}{\partial \mathbf{W}^{[m]}} = - \sum_{n=1}^N E \left\{ \frac{\partial \log(p(\mathbf{y}_n))}{\partial \mathbf{y}_n^{[m]} \frac{\partial \mathbf{y}_n^{[m]}}{\partial \mathbf{W}^{[m]}}} \right\} - \left( \mathbf{W}^{[m]} \right)^{-T} \quad (2.41)$$

where  $p(\mathbf{y}_n)$  is a probability density function, which plays an important role in the development of the IVA algorithm. In our case, we use  $p(\mathbf{y}_n)$  of BMGGMM to improve the IVA model. Using

Equations (2.4) and (2.40), the cost function for IVA mixture can be expressed as:

$$J_{IVA} = -\log(\phi_j) + \frac{D}{2} \log(\mathbf{c}_j) + \frac{1}{2} \log |\Sigma_j| + \frac{1}{2(\mathbf{c}_j)^{\beta_j}} \mathbf{y}^{\beta_j} - \log |h_n^T \mathbf{W}_n^{[m]}|$$

$$- \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left[ -\log(\phi_j) + \frac{D}{2} \log(\mathbf{c}_j) + \frac{1}{2} \log |\Sigma_j| + \frac{1}{2(\mathbf{c}_j)^{\beta_j}} (\mathbf{y}_u)^{\beta_j} \right] du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \quad (2.42)$$

where  $\phi_j = \left[ \frac{\beta_j \Gamma(\frac{D}{2})}{\pi^{\frac{D}{2}} \Gamma(\frac{D}{2\beta_j}) 2^{2\beta_j}} \right]$ ,  $\mathbf{y}$  is defined as:

$\mathbf{y} = (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j)$ ,  $\mathbf{y}_u = (\mathbf{u} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{u} - \boldsymbol{\mu}_j)$  and the gradient of the cost function of IVA-BMGGMM is given by:

$$\frac{\partial J_{IVA}}{\partial \mathbf{W}^{[m]}} = E \left[ \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} (\mathbf{y})^{\beta_j - 1} \Sigma_j^{-1} (\mathbf{x}_n^{[m]} - \boldsymbol{\mu}_j) \mathbf{x}_n^{[m]} \right] - \frac{h_n^{[m]}}{(h_n^{[m]})^T \mathbf{W}_n^{[m]}}$$

$$- \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left[ \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} (\mathbf{y}_u)^{\beta_j - 1} \Sigma_j^{-1} (\mathbf{x}_n^{[m]} - \boldsymbol{\mu}_j) \mathbf{x}_n^{[m]} \right] du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \quad (2.43)$$

where the term  $\frac{h_n^{[m]}}{(h_n^{[m]})^T \mathbf{W}_n^{[m]}}$  is introduced to solve the issue of poor convergence that is raised due to the inverse of  $\mathbf{W}$  at each iteration. The  $M$  mixing matrices are updated as follows:

$$(\mathbf{W}_n^{[m]})^{new} = (\mathbf{W}_n^{[m]})^{old} - \gamma \frac{\partial J_{IVA}}{\partial \mathbf{W}^{[m]}} \quad (2.44)$$

where  $\gamma$  is a scalar step size and can be fixed to a small number.

## 2.6 Experimental results

This Section will discuss and demonstrate the validation of the proposed models, namely BMGGMM, ICA-BMGGMM, and IVA-BMGGMM, as described in the following subsections.

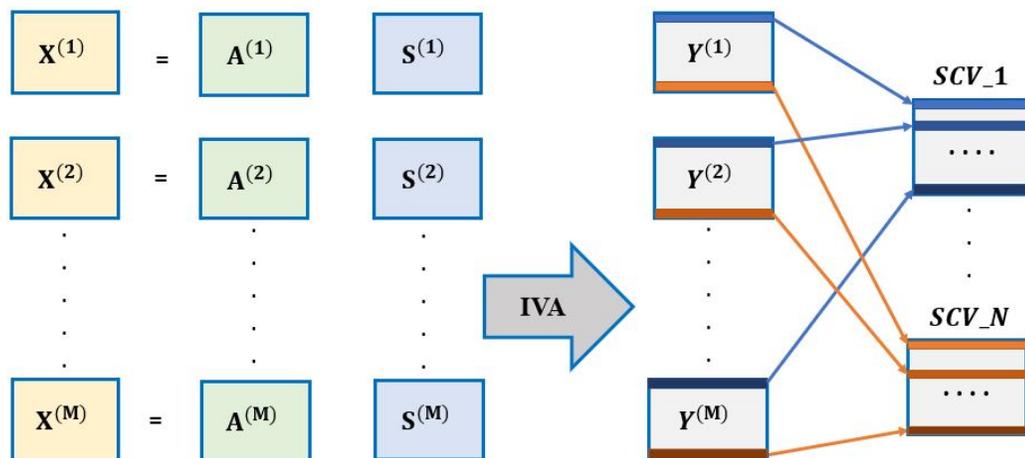


Figure 2.1: General framework of the IVA model. The  $n$ th source component vectors (SCV) is formed by grouping the corresponding  $m$ th source from each source vector together.

### 2.6.1 Experimental evaluation of BMGGMM

We consider medical applications, including stroke, skin cancer, alzheimer, diabetes, and human resources analysis datasets to validate the performance of BMGGMM.

In terms of performance metrics, we compared the performance of the employed models using both objective and subjective measures. We selected widely used measures namely accuracy, recall, precision, F1 score, false-positive rate (FPR), Jaccard score and Log loss. Recall is used to validate the model's ability to recognize positive labels. Precision is calculated by dividing the percentage of corrected positive labels by the total number of positive labels. The F1 score combines precision and sensitivity into a single metric. The F1 score will range between 0 and 1. FPR is a rate calculated from the confusion matrix, where the FPR is the ratio of false positive labels to negative labels. The log-loss value represents how close the prediction probability is to the true value (0 or 1 in the case of binary classification). The greater the difference between predicted and actual probability, the greater the log-loss value. For more details, the reader is referred to [51, 52, 53].

#### Stroke detection

Stroke is one of the leading causes of adult mortality worldwide, affecting 6.2 million people each year [54]. As a result, significant research has been conducted in order to facilitate better and more accurate stroke detection [55, 56]. The stroke dataset is composed of 4900 observations

recorded for 4692 healthy subjects and 208 patients, and it has ten attributes such as gender, age, various diseases, and smoking status. This dataset was created to classify healthy persons and patients with stroke. We apply the BMGGMM on the stroke data to validate its performance. The results of the BMGGMM are compared to GMM, GGMM, BGGMM, and MGGMM. From Table 2.1, it can be observed that accuracy (95.77%), precision (95.64%) and all other performance metrics indicate that the proposed model (BMGGMM) outperforms the compared models (GMM, GGMM, BGGMM, MGGMM) in clustering stroke datasets to its relevant classes. Thus, we can conclude that BMGGMM can be used in medical diagnosis process.

Table 2.1: Performance on Stroke data based on different metrics.

Model	Accuracy	Precision	Recall	F1 score	FPR	Jaccard score	Log loss
<b>BMGGMM</b>	<b>95.77</b>	<b>95.64</b>	<b>99.21</b>	<b>97.43</b>	<b>16.61</b>	<b>95.54</b>	<b>4.09</b>
MGGMM	80.26	95.34	83.25	89.64	35.52	80.44	6.81
BGGMM	79.77	94.77	81.32	77.84	34.12	75.77	11.45
GGMM	64.02	94.23	64.72	76.65	35.33	61.24	12.11
GMM	62.02	94.52	64.53	76.46	31.35	62.12	13.11

### Skin cancer detection

Skin cancer is one of the most common cancers in the world. To reduce skin cancer death rates, early detection of this form of the disease is vital. Machine learning has recently proven an effective way of identifying tumors of the skin as either malignant or benign [57]. The skin dataset consists of 3272 observations from 1776 benign subjects and 1496 malignant subjects. The objective of using this data in this experiment is to categorize benign and malignant subjects through clustering. To efficiently extract the essential features of an image, we used the scale-invariant

Table 2.2: Performance on skin cancer data based on different metrics.

Model	Accuracy	Precision	Recall	F1 score	FPR	Jaccard score	Log loss
<b>BMGGMM</b>	<b>74.08</b>	<b>73.43</b>	<b>69.52</b>	<b>71.61</b>	<b>13.45</b>	<b>75.63</b>	<b>8.95</b>
MGGMM	66.98	73.22	64.13	65.43	13.82	66.91	11.40
BGGMM	60.81	67.26	58.39	59.50	13.60	63.26	13.53
GGMM	55.06	52.09	52.18	53.44	16.39	52.56	15.52
GMM	54.48	52.32	54.24	58.55	19.23	54.95	15.72

feature transform (SIFT) technique. We used bags-of-visual-words (BoVW) to represent data as D-dimensional feature vectors [58]. In this experiment, we applied the proposed model (BMGGMM) to distinguish between benign and malignant subjects. BMGGMM’s clustering performance is compared to GMM, GGMM, and MGGMM. FPR (13.45) and Log loss (8.95) from Table 2.2 indicate lower rates, demonstrating that our proposed model outperforms MGGMM, BGGMM, GGMM, and GMM. The proposed algorithm demonstrates its effectiveness in clustering skin cancer from non-skin cancer classes, implying that BMGGMM may be used in cancer detection in the early stage.

### Alzheimer disease diagnosis

Alzheimer’s disease (AD) is one of the most common causes of dementia. According to the Alzheimer report [59], approximately 50 million people were affected by Alzheimer’s disease in 2018, with the number expected to triple by 2050. The dataset was collected using longitudinal MRI data from 5069 subjects aged from 60 to 96. Each subject was scanned at least once. Throughout this data, 2560 subjects were classified as non-dementia and 2509 as dementia. We used the BoVW approach to represent the features of data. Next, we applied our proposed model (BMGGMM) to differentiate between Alzheimer patients and healthy subjects. We observe from Table 2.3 that accuracy (79.83%), precision (81.54%), F1 score (76.22%) and other metrics indicate that our proposed model performed better than GMM, GGMM, BGGMM, and MGGMM in distinguishing between Alzheimer and healthy subjects.

Table 2.3: Performance on Alzheimer data based on different metrics.

Model	Accuracy	Precision	Recall	F1 score	FPR	Jaccard score	Log loss
<b>BMGGMM</b>	<b>79.83</b>	<b>81.54</b>	<b>65.23</b>	<b>76.22</b>	<b>16.56</b>	<b>72.84</b>	<b>6.97</b>
MGGMM	74.93	78.92	70.83	66.72	18.56	69.84	8.66
BGGMM	75.53	79.23	71.32	72.34	15.27	70.34	7.31
GGMM	74.79	80.93	65.65	72.49	20.75	68.34	8.70
GMM	51.85	51.76	50.38	66.26	21.42	49.16	16.63

## Diabetes experiment

Diabetes mellitus is one of the most common diseases among the elderly. According to the International Diabetes federation, 451 million people worldwide had diabetes in 2017 (World Health Organization)<sup>1</sup>. Early detection and management of DM also prevents complications and helps to reduce the chances of severe health problems [60]. We consider a dataset containing 390 observations with 16 attributes categorized into diabetes and healthy subjects for this application.

From results in Table 2.4, it is observed that BMGGMM successfully distinguishes between diabetes and non-diabetes categories. Accuracy (84.61%), recall (96.47%), F1 score (91.96%) and other performance measures indicate the effectiveness of BMGGMM as compared to other models.

Table 2.4: Performance on Diabetes data based on different metrics.

Model	Accuracy	Precision	Recall	F1 score	FPR	Jaccard score	Log loss
<b>BMGGMM</b>	<b>84.61</b>	<b>84.32</b>	<b>96.47</b>	<b>91.96</b>	<b>14.28</b>	<b>84.17</b>	<b>5.31</b>
MGGMM	74.10	84.53	85.41	84.94	18.89	73.37	8.94
BGGMM	66.15	80.90	66.66	76.92	15.33	62.50	11.69
GGMM	56.41	85.74	58.65	69.83	21.19	51.27	15.05
GMM	54.41	84.53	56.62	66.48	26.93	49.21	17.05

## Human resources analysis experiment

The development of the global workforce and the increasing importance of business analysis as a strategic organizational capacity today have a significant impact on human resources management [61]. The Human resources dataset <sup>2</sup> is composed of 14,999 employees, and each observation has ten features such as satisfaction level, number of projects, and salary. The dataset is categorized into two groups where the employee stays or leaves a specific job. BMGGMM is conducted to examine the level of satisfaction and then classify employees into two classes. Table 2.5 provides the results for our proposed model compared to GMM, GGMM, BGGMM, and MGGMM. FPR (17.57) and Log loss (8.22) show lower error rates for our proposed model compared to MGGMM, BGGMM, GGMM, and GMM. These results demonstrate the efficacy of BMGGMM performance

<sup>1</sup><https://www.who.int/news-room/fact-sheets/detail/diabetes>

<sup>2</sup><https://www.kaggle.com/jacksonchou/hr-data-for-analytics>

in clustering the level of employer satisfaction in certain jobs.

As shown in Table 2.5, the improvement in the proposed BMGGMM model is clearly not significant when compared to other models, which can be explained by the nature, complexity, and dimensionality of the data. As shown in the Table, the high dimensionality of this data degrades the performance of both our proposed model and the compared models. A dimensionality reduction technique such as PCA or a feature selection algorithm to extract the most effective features is highly recommended to improve the classification capability of our proposed model.

Table 2.5: Performance on human resources data based on different metrics.

Model	Accuracy	Precision	Recall	F1 score	FPR	Jaccard score	Log loss
<b>BMGGMM</b>	<b>76.19</b>	<b>76.45</b>	<b>94.95</b>	<b>86.88</b>	<b>17.57</b>	<b>76.38</b>	<b>8.22</b>
MGGMM	75.69	76.74	94.95	86.39	20.22	75.17	8.39
BGGMM	73.45	73.07	85.35	83.15	24.64	67.67	12.62
GGMM	63.98	73.56	82.29	77.48	21.91	63.85	12.43
GMM	61.95	72.74	79.93	76.82	23.19	61.37	13.14

## 2.6.2 Experiments and results for clustering using BMGGMM-MML

Model selection through MML is validated using the same experiments that were used to validate the BMGGMM. The proposed MML technique for model selection is compared to several deterministic model selection criteria including Akaike’s information criterion (AIC) [27], Bayesian inference criterion (BIC) [62], Consistent AIC (CAIC) [63], LEC [64], and MDL [65]. MML re-

Table 2.6: Number of clusters determined by MML.

Data	D	N	K*	MML	AIC	BIC	CAIC	LEC	MDL
Stroke	10	4900	2	<b>2</b>	3	2	2	5	2
Skin Cancer	16	3272	2	<b>2</b>	3	4	4	1	4
Alzheimer	16	5068	2	<b>2</b>	2	2	2	3	2
Diabetes	15	390	2	<b>2</b>	5	3	3	2	3
employee	8	14998	2	<b>2</b>	2	2	2	5	2

\* Actual number of clusters in the data.

sults are compared to other model selection criteria to assess the performance of our technique, and noteworthy findings are drawn from these results. Each dataset is tested to evaluate the performance

and viability of MML. Model selection criteria were employed to detect the exact number of clusters in each dataset. The results of each method against each dataset are summarized in Table 2.6, where it is observed that MML has successfully determined the number of categories in all datasets.

### 2.6.3 Experimental results for BMGGMM with ICA

In this subsection, we evaluate the proposed model (ICA-BMGGMM) using a variety of applications, including speech, electrocardiogram (EOG), and electroencephalogram databases, as described in the following subsections.

#### Separating heartbeat sounds

The pregnancy period is critical in the diagnosis and prevention of a variety of fetal abnormalities. Pregnant women should be monitored regularly in most developed countries. Monitoring fetal heart rate (FHR) usually begins around the 24th week, using doppler ultrasonographic cardiotocography [66]. In this experiment, we used a Shiraz University fetal heart sounds database (SUFHSDB)<sup>3</sup>. The SUFHSDB dataset is collected from mothers aged 16 to 47 years with a digital stethoscope placed on the lower maternal abdomen. It consists of 99 subjects who had one signal recorded, three subjects who had two signals recorded, and seven cases of twins recorded separately, totaling 109 recordings. The sampling rate was generally 16,000 Hz with 16-bit quantization, and a few recordings were made at 44,100 Hz.

Firstly, the ICA mixture model is used to estimate basis functions, which are then used to separate mixed signals. In this experiment, we estimated the basis functions of  $2 \times 2$  to compute two heartbeat sound sources. We have used linear mixtures of 2 sources to recover the heartbeat sound sources and executed blind source separation by employing ICA-BMGGMM. Subjective and objective measures on recovered sources were performed to assess the quality of recovered heartbeat signals and the viability of the ICA mixture model in BSS. The subjective analysis includes heartbeat signals before and after source separation. The objective analysis includes signal-to-distortion ratio (SDR), signal-to-interference ratio (SIR), signal-to-noise ratio (SNR), signal-to-artefact ratio (SAR), and intersymbol interference (ISI) [67]. Figs. 2.2 and 2.3 compare the performance of the

---

<sup>3</sup><https://physionet.org/content/sufhsdb/1.0.1/>

Table 2.7: Objective measure for separation of heartbeat signals.

Model	BSS of ECG				
	SDR	SIR	SAR	SNR	ISI
<b>ICA-BMGGMM</b>	<b>77.93</b>	<b>78.99</b>	<b>84.57</b>	<b>77.73</b>	<b>0.49</b>
ICA-MGGMM	45.91	38.72	62.56	55.64	0.51
ICA-BGGMM	28.30	28.32	30.41	29.29	0.49
ICA	12.45	13.51	19.28	25.26	0.53

proposed model to ICA visually. The fetal and maternal acoustic ECG signals had similar morphology, but the fetal heartbeat frequency was higher than the maternal heartbeat frequency. Table 2.7 compares the performance of ICA-BMGGMM to that of ICA, ICA-BGGMM, and ICA-MGGMM using SDR, SIR, SAR, SNR, and ISI performance metrics. It is clearly shown that ICA-BMGGMM outperforms ICA, ICA-BGGMM, and ICA-MGGMM for all four measures: SDR (77.93), SIR (78.99), SAR (84.57), and SNR(77.73). Furthermore, ISI (0.49) for ICA-BMGGMM is lower than that for ICA, ICA-BGGMM, and ICA-MGGMM models.

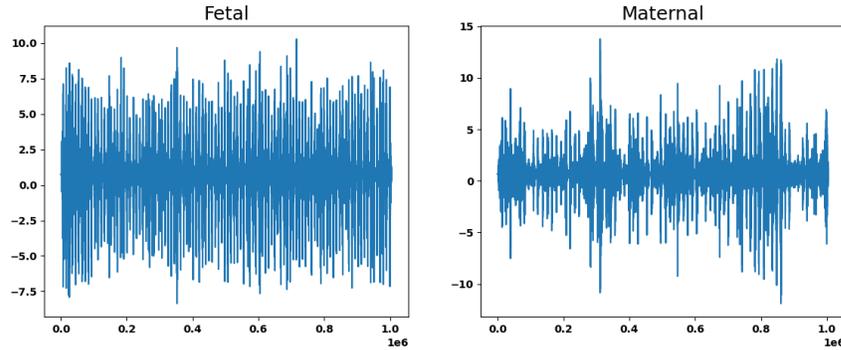


Figure 2.2: BSS for heartbeat signals using ICA. The left acoustic signal represents the fetal heartbeat sound, while the right acoustic signal represents the mother heartbeat sound.

### Fetal ECG extraction

Cardiotocography (CTG) measures the relationship between FHR and uterine contractions during pregnancy [68]. Intravenously, CTG measures the FHR and uterine pressure to ensure the optimal health of the fetus. Machine learning techniques are being developed to assist in predicting the fatal cardiac disease. Pregnant woman dataset<sup>4</sup> contains nine channels; we only keep the first

<sup>4</sup><https://homes.esat.kuleuven.be/tokka/daisydata.html>

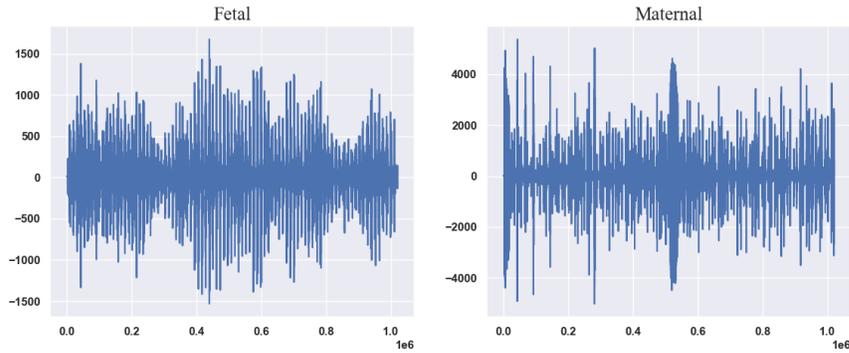


Figure 2.3: BSS for heartbeat signals using ICA-BMGGMM. The fetal heartbeat sound is represented by the left acoustic signal, while the mother heartbeat sound is represented by the right acoustic signal.

three recordings, as shown in Fig. 2.4, captured using electrodes placed on the patient’s abdomen. There are 2500 measurements in each captor signal. The ICA mixture model is used to estimate basis functions  $2 \times 2$ ,  $3 \times 3$ ,  $4 \times 4$ , and  $5 \times 5$  in separated experiments to compute 2, 3, 4, and 5 sources. In BSS, only ECG signals after linear mixing are visible. The basis functions that will be used during the source separation are unknown. ICA mixture models were evaluated using subjective and objective measures. In the subjective analysis, ECG signals obtained before and after source separation are used. SDR, SIR, SAR, SNR, and ISI are among the objective metrics that are evaluated. We conducted four experiments using this BSS framework to compute 2, 3, 4, and

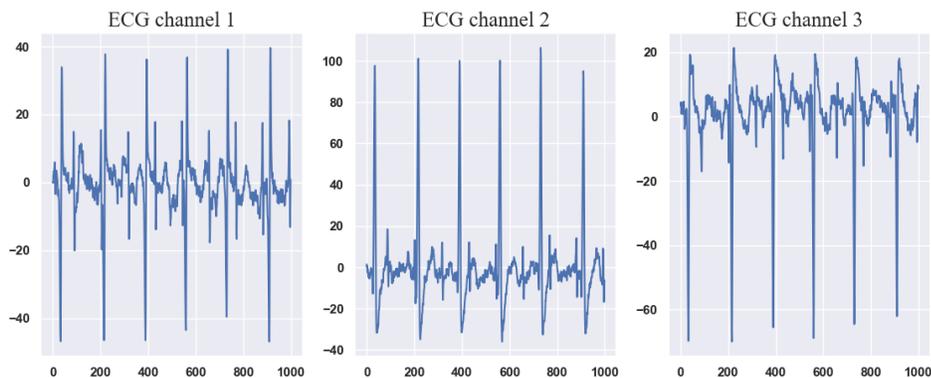


Figure 2.4: Original ECG channels. These three channels were extracted from a pregnant women dataset.

5 ECG sources. We used a linear mixture of 2, 3, 4, and 5 ECG sources from the database and

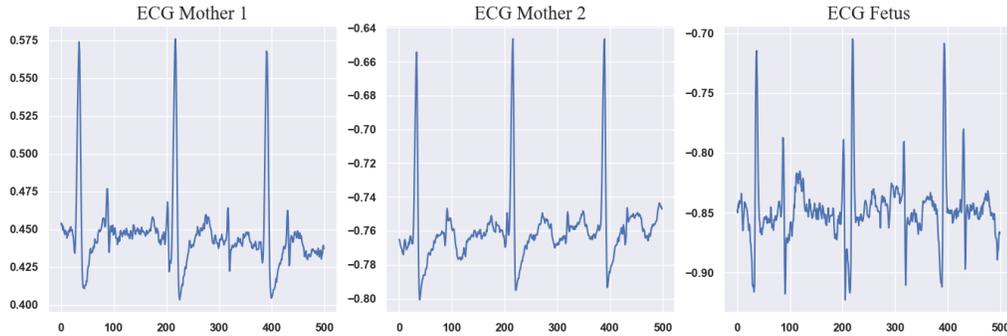


Figure 2.5: ICA-BMGGMM ECG channels. On the left are the extracted ICA-BMGGMM ECG signals for mother. On the right is the fetal ICA-BMGGMM ECG signal.

conducted BSS using three models (ICA, ICA-BGGMM, ICA-MGGMM, and ICA-BMGGMM) to recover 2, 3, 4, and 5 ECG sources, respectively. Objective analysis is performed on recovered sources to determine the quality of recovered ECG signals and the viability of ICA mixture models in BSS. This experiment is repeated ten times with different linear ECG mixtures from the original database, and then the objective measures are averaged across these ten separated experiments. The objective measures obtained after recovery of ECG source signals are shown in Tables 2.8 and 2.9 for the majority of performance measures. In case of BSS for 2 ECG signals, we find that the ICA-BMGGMM model outperforms ICA, ICA-BGGMM, and ICA-MGGMM models for SDR (28.46), SIR (27.09), SAR (261.57), and also the other metrics. Figs. 2.4 and 2.5 show a comparison of ECG signals before and after BSS. The morphology of the fetal and maternal ECG signals was similar, but the fetal heartbeat frequency was higher than the maternal heartbeat frequency. The above BSS

Table 2.8: Objective measure for separation of 2 and 3 ECG signals.

Model	BSS for of 2 ECG signals					BSS for of 3 ECG signals				
	SDR	SIR	SAR	SNR	ISI	SDR	SIR	SAR	SNR	ISI
<b>ICA-BMGGMM</b>	<b>28.46</b>	<b>27.09</b>	<b>261.57</b>	<b>32.40</b>	<b>0.45</b>	<b>19.01</b>	<b>19.02</b>	<b>256.79</b>	<b>19.34</b>	<b>0.45</b>
ICA-MGGMM	9.50	9.50	255.58	16.80	0.46	9.91	9.91	251.63	16.22	0.47
ICA-BGGMM	9.20	9.21	253.32	20.31	0.45	14.63	14.53	235.45	10.22	0.45
ICA	8.76	8.76	246.20	15.74	0.47	8.45	8.46	248.85	15.88	0.49

experiments demonstrate the effectiveness of ICA-BMGGMM in comparison to commonly used ICA, ICA-BGGMM, and ICA-MGGMM models. It is also observed that the rate of improvement slows as the number of linear mixtures in source separation increases. These findings show that

Table 2.9: Objective measure for separation of 4 and 5 ECG signals.

Model	BSS for of 4 ECG signals					BSS for of 5 ECG signals				
	SDR	SIR	SAR	SNR	ISI	SDR	SIR	SAR	SNR	ISI
<b>ICA-BMGGMM</b>	<b>14.01</b>	<b>14.02</b>	<b>256.26</b>	<b>14.26</b>	<b>0.44</b>	<b>14.08</b>	<b>14.87</b>	<b>251.79</b>	<b>19.26</b>	<b>0.43</b>
ICA-MGGMM	10.44	10.45	251.86	15.03	0.47	12.16	10.16	249.68	17.78	0.48
ICA-BGGMM	9.48	9.58	252.80	14.29	0.52	10.59	10.56	251.76	17.12	0.44
ICA	7.36	7.36	248.66	13.22	0.48	10.10	8.10	245.37	16.43	0.49

the ability of ICA-BMGGMM to recover the sources decreases as the mixing complexity increases.

This is a limitation of the ICA-BMGGMM that will be addressed in future work.

### Blind source separation of speech signals

We evaluate our proposed technique with random signals obtained from Romanian read-speech corpus (RoDigits) dataset<sup>5</sup>. This corpus contains 37.5 hours of connected spoken digits from 154 people between 20 and 45 years old. In the final corpus, each speaker recorded 100 clips of 12 Romanian-generated digits at random, with a sampling frequency of 16 kHz, in a total of 15,389 sound files. In BSS, the ICA mixture model estimates basis functions, which are then separated from mixed signals. To compute 2, 3, 4, and 5 sources, we estimated basis functions 2 x 2, 3 x 3, 4 x 4, and 5 x 5 in separated experiments. Only speech signals after linear mixing are observed when using BSS. There is no prior knowledge of the basis functions that will be used in the source separation. The BSS framework is evaluated using both subjective and objective measures. Speech signals obtained before and after source separation are used in the subjective analysis. The major components of objective analysis are SDR, SIR, SAR, SNR, and perceptual evaluation of speech quality (PESQ) [67]. In order to validate the BSS on ICA mixture models, RoDigits speech corpora

Table 2.10: Objective measure for separation of 2 and 3 speech signals.

Model	BSS of 2 speech signals					BSS of 3 speech signals						
	SDR	SIR	SAR	SNR	PESQ	ISI	SDR	SIR	SAR	SNR	PESQ	ISI
<b>ICA-BMGGMM</b>	<b>28.27</b>	<b>28.83</b>	<b>59.97</b>	<b>26.28</b>	<b>1.25</b>	<b>0.46</b>	<b>9.91</b>	<b>9.98</b>	<b>58.18</b>	<b>18.62</b>	<b>1.91</b>	<b>0.36</b>
ICA-MGGMM	21.22	22.24	49.42	16.78	1.23	0.47	8.72	8.73	46.98	17.79	1.71	0.39
ICA-BGGMM	26.99	26.67	56.23	20.53	1.24	0.5	9.88	9.88	54.60	17.19	1.07	0.73
ICA	11.77	13.28	36.03	9.57	1.09	0.51	8.23	8.24	40.49	16.46	1.65	0.53

are employed. We conducted four experiments with the BSS framework to compute two, three, four,

<sup>5</sup><https://speed.pub.ro/downloads/speech-datasets/>

Table 2.11: Objective measure for separation of 4 and 5 speech signals.

Model	BSS of 4 speech signals						BSS of 5 speech signals					
	SDR	SIR	SAR	SNR	PESQ	ISI	SDR	SIR	SAR	SNR	PESQ	ISI
<b>ICA-BMGGMM</b>	<b>7.93</b>	<b>7.93</b>	<b>49.54</b>	<b>17.27</b>	<b>1.45</b>	<b>0.37</b>	<b>7.34</b>	<b>7.37</b>	<b>45.95</b>	<b>17.68</b>	<b>1.39</b>	<b>0.35</b>
ICA-MGGMM	7.35	6.37	45.03	16.57	1.39	0.38	6.31	6.48	44.58	15.29	1.33	0.36
ICA-BGGMM	6.08	6.08	47.82	15.94	1.03	0.76	6.10	6.10	45.85	16.16	1.01	0.79
ICA	6.86	6.19	41.22	15.17	1.35	0.49	6.15	6.39	42.96	14.76	1.29	0.52

and five speech sources. We took a linear mixture of these speech sources from the database, and we employed three models (ICA, ICA-BGGMM, ICA-MGGMM, and ICA-BMGGMM) to recover them. Once the source is recovered, objective analysis will be performed to determine the quality of speech signals recovered and the viability of ICA mixture models in BSS. The experiment is performed ten times with objective measures averaged over ten separated experiments using different linear speech mixtures of 2, 3, 4, and 5 from the original database sources.

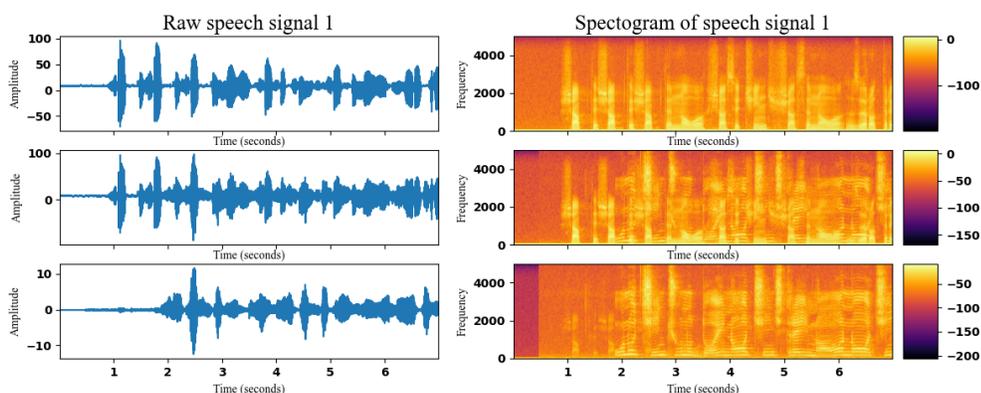


Figure 2.6: Speech signals and their corresponding spectrograms from source 1. The first row represents both the raw speech signal and its spectrogram. The second row illustrates the speech mixed signal and its spectrogram. The third row represents the extracted speech signal using ICA-BMGGMM and its corresponding spectrogram.

The objective measures obtained after the recovery of speech source signals are shown in Tables 2.10 and 2.11. For the majority of performance metrics such as SDR (28.27) and SIR (28.83) for BSS of 2 speech signal, we find that the proposed model (ICA-BMGGMM) outperforms ICA, ICA-BGGMM, and ICA-MGGMM models. It is also worth noting that when the number of linear mixtures in source separation increases, improvement decreases as shown in Tables 2.10 and 2.11. Figs. 2.6 and 2.7 present speech signals and their corresponding spectrograms before mixing,

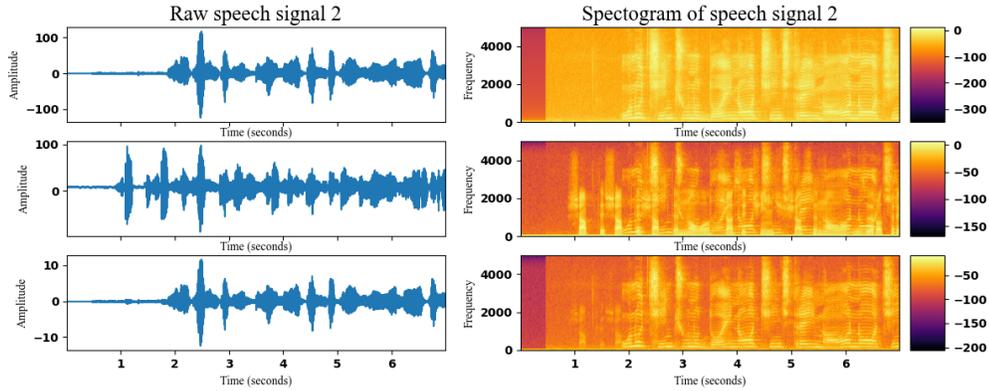


Figure 2.7: Speech signals and their corresponding spectrograms from source 2. The first row represents both the raw speech signal and its spectrogram. The second row illustrates the speech mixed signal and its spectrogram. The third row represents the extracted speech signal using ICA-BMGGMM and its corresponding spectrogram.

after mixing, and after BSS for two different speech sources to provide a clear comparison to the reader. As demonstrated by these two figures, The ICA-BMGGMM is capable of recovering speech resources signals from mixed speech signals.

### EEG eye blink removal

In the electroencephalogram analysis, eye movements and blink artifacts can cause various issues. EEG contaminated by ocular artifacts can be manually removed. An electrocardiogram uses electrodes placed above and around the eyes to measure eye movement activity directly. EEG data<sup>6</sup> from 14 males and 13 females were used to evaluate the proposed algorithm for removing EOG artifacts. This EEG data were collected at a sampling rate of 200 Hz with a bandpass filter set to 0.5-40 Hz on 19 channels (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2).

After pre-processing the actual continuous EEG signals, EEG data dimensions are  $M = 19$  channels and  $N = 5601$  observations. ICA, ICA-BGGMM, ICA-MGGMM, and ICA-BMGGMM models were used to assess the performance in ten separate experiments, each with its own set of EEG data. Artifact rejection can be implemented after the artifactual sources have been separated

<sup>6</sup><https://data.mendeley.com/datasets/wb6yvr725d/1>

using ICA-BMGGMM by automatically detecting and removing the artifactual sources. The objective measures (SDR, SIR, SAR, SNR, and ISI) are used to compare the performance.

Table 2.12 shows objective measures after recovery of EEG signals. We find that ICA-BMGGMM exceeds those of ICA, ICA-BGGMM, and ICA-MGGMM models for SAR (254.35), SNR (29.43), and other measures. We have taken the Fp1 channel as an example to clarify ICA-BMGGMM performance in removing EEG artifacts, as shown in Fig. 2.8. The obtained Fp1 EEG signal demonstrates ICA-BMGGMM’s ability to eliminate eye blinking caused by Fp1 EEG electrodes placed around the eye. This BSS experiment showed that the ICA-BMGGMM model outperforms ICA, ICA-BGGMM, and ICA-MGGMM models.

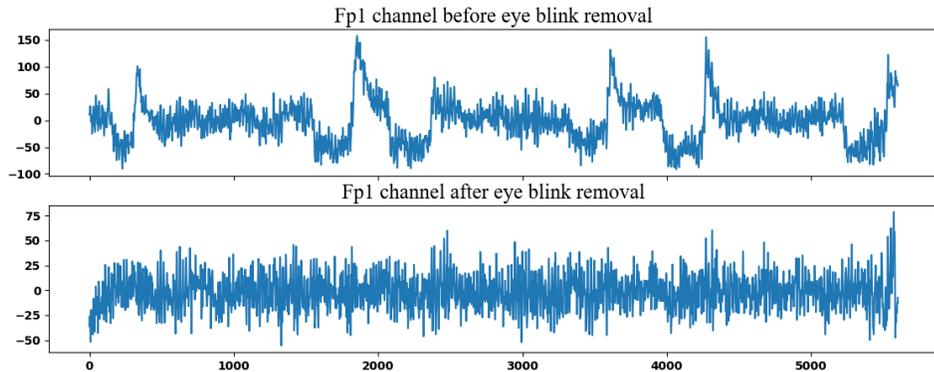


Figure 2.8: Fp1 channel before and after blink removal. On the upper, the EEG signal is contaminated by eye blinking signals. After the eye blink signals have been removed, the cleaning EEG signal is shown at the bottom.

Table 2.12: Objective measures for EEG experiments.

Model	BSS of EEG				
	SDR	SIR	SAR	SNR	ISI
<b>ICA-BMGGMM</b>	<b>46.13</b>	<b>43.74</b>	<b>254.35</b>	<b>29.43</b>	<b>0.45</b>
ICA-MGGMM	45.02	42.76	249.08	26.77	0.52
ICA-BGGMM	45.16	42.98	252.17	24.55	0.53
ICA	44.26	41.90	245.20	22.65	0.63

## 2.6.4 Experimental results for IVA-BMGGMM

In this subsection, we evaluated the performance of the proposed model (IVA-BMGGMM) using different multivariate applications, including speech, ECG, fMRI, and EEG databases, as

explained in the following subsections.

### Speech separation using IVA-BMGMM

The speech data used in this subsection is described in Subsection 2.6.3. The IVA mixture model estimates basis functions in BSS, which is then used to separate mixed signals. We estimated basis functions  $2 \times 2$ ,  $3 \times 3$ ,  $4 \times 4$ , and  $5 \times 5$  in separate experiments to compute 2, 3, 4, and 5 sources. When using BSS, only speech signals after linear mixing are observed. There is no previous knowledge of the basis functions that will be employed during the source separation. Both subjective and objective measures are used to evaluate the BSS framework. The subjective analysis comprised of speech signals is obtained before and after source separation. SDR, SIR, SAR, SNR, ISI, and PESQ are major components of objective analysis [69, 67]. RoDigits speech corpora are used to validate blind source separation based on IVA mixture models. We executed four experiments using this BSS framework to compute 2, 3, 4, and 5 speech sources.

Table 2.13: Objective measure for separation of 2 and 3 speech signals.

Model	BSS of 2 speech signals						BSS of 3 speech signals					
	SDR	SIR	SAR	SNR	PESQ	ISI	SDR	SIR	SAR	SNR	PESQ	ISI
<b>IVA-BMGMM</b>	<b>30.78</b>	<b>30.92</b>	<b>87.26</b>	<b>29.67</b>	<b>2.38</b>	<b>0.38</b>	<b>10.94</b>	<b>10.95</b>	<b>81.96</b>	<b>28.55</b>	<b>2.81</b>	<b>0.33</b>
<b>ICA-BMGMM</b>	<b>28.27</b>	<b>28.83</b>	<b>59.97</b>	<b>26.28</b>	<b>1.25</b>	<b>0.46</b>	<b>9.91</b>	<b>9.98</b>	<b>58.18</b>	<b>18.62</b>	<b>1.91</b>	<b>0.36</b>
IVA-MGGMM	25.45	25.53	76.27	19.89	2.28	0.39	9.13	10.14	79.82	24.17	2.76	0.36
ICA-MGGMM	21.22	22.24	49.42	16.78	1.23	0.47	8.72	8.73	46.98	17.79	1.71	0.39
IVA-BGGMM	29.25	29.38	62.35	27.91	1.25	0.49	8.72	8.76	61.17	24.60	1.09	0.49
ICA-BGGMM	26.99	26.67	56.23	20.53	1.24	0.50	9.88	9.88	54.60	17.19	1.07	0.73
IVA	20.60	20.71	61.87	13.06	2.10	0.48	8.95	9.50	64.99	19.27	2.61	0.41
ICA	11.77	13.28	36.03	9.57	1.09	0.51	8.23	8.24	40.49	16.46	1.65	0.53

Table 2.14: Objective measure for separation of 4 and 5 speech signals.

Model	BSS of 4 speech signals						BSS of 5 speech signals					
	SDR	SIR	SAR	SNR	PESQ	ISI	SDR	SIR	SAR	SNR	PESQ	ISI
<b>IVA-BMGMM</b>	<b>8.28</b>	<b>8.29</b>	<b>64.00</b>	<b>22.59</b>	<b>2.74</b>	<b>0.32</b>	<b>7.69</b>	<b>7.69</b>	<b>71.55</b>	<b>22.73</b>	<b>2.58</b>	<b>0.31</b>
<b>ICA-BMGMM</b>	<b>7.93</b>	<b>7.93</b>	<b>49.54</b>	<b>17.27</b>	<b>1.45</b>	<b>0.37</b>	<b>7.34</b>	<b>7.37</b>	<b>45.95</b>	<b>17.68</b>	<b>1.39</b>	<b>0.35</b>
IVA-MGGMM	7.65	7.66	63.30	22.30	2.66	0.35	6.95	6.96	69.39	22.62	2.48	0.33
ICA-MGGMM	7.35	6.37	45.03	16.57	1.39	0.38	6.31	6.48	44.58	15.29	1.33	0.36
IVA-BGGMM	7.23	7.26	62.17	21.57	1.09	0.51	5.05	5.07	63.94	15.52	1.03	0.52
ICA-BGGMM	6.08	6.08	47.82	15.94	1.03	0.76	6.10	6.10	45.85	16.16	1.01	0.79
IVA	6.95	6.96	62.56	20.36	2.53	0.40	6.48	6.48	68.01	19.83	2.31	0.38
ICA	6.86	6.19	41.22	15.17	1.35	0.49	6.15	6.39	42.96	14.76	1.29	0.52

We considered a linear mixture of 2, 3, 4, and 5 speech sources from the RoDigits database

and performed the BSS using six different models (IVA, IVA-BGGMM, IVA-MGGMM, IVA-BMGGMM, ICA, ICA-BGGMM, ICA-MGGMM, and ICA-BMGGMM) to recover the speech sources. Once sources have been recovered, they are subjected to objective analysis to determine the quality of the recovered speech signals and the viability of IVA mixture models in BSS. Using different linear speech mixtures of 2, 3, 4, and 5 sources from the original database, the experiment is repeated ten times with the objective measures averaged over these ten separated experiments. These results show that the ability of IVA-BMGGMM to recover the sources decreases as the mixing complexity increases.

Tables 2.13 and 2.14 indicate the objective measures obtained after the recovery of speech source signals. For 2, 3, 4, and 5 sources, we find that the proposed model (IVA-BMGGMM) outperforms compared models includes IVA-MGGMM, IVA-BGGMM, IVA, ICA-BMGGMM, ICA-MGGMM, ICA-BGGMM, and ICA for most performance metrics. For BSS of 2 speech signal, we observed a significant improvement for IVA-BMGGMM (SDR = 30.78) compared to ICA-BMGGMM (SDR = 28.27), IVA-MGGMM (SDR = 25.45), ICA-MGGMM (SDR = 21.22), IVA-BGGMM (SDR = 29.25), ICA-BGGMM (SDR = 26.99), IVA (SDR = 20.60), and ICA (SDR = 11.77). However, when source signals increase as in BSS of 5 speech signals, we noticed low improvements. According to this BSS experiments, IVA-BMGGMM effectiveness performance is better than IVA-MGGMM, IVA-BGGMM, IVA, ICA-BMGGMM, ICA-MGGMM, ICA-BGGMM, and ICA models. Also, it has been noticed that when the number of linear mixtures in source separation increases, the rates of improvement decrease.

### **EEG eye blink removal experiment**

An EEG dataset is used to evaluate the proposed algorithm. This database was described in 2.6.3 in detail. This continuous EEG data were segmented in a 10-second EEG window, where each window will be referred to as an epoch. After epoching, 200 epochs are extracted from one recording per channel. Then the data was reshaped to be in a source vector form, then used as input for IVA models. Then the IVA models are implemented to remove EOG artifacts as shown in Fig. 2.9. After preprocessing actual continuous EEG signals, EEG data dimensions are  $M = 19$  channels,  $T = 10$  time points, and  $N = 200$  epochs. After that, 200 epochs are used to run multiple

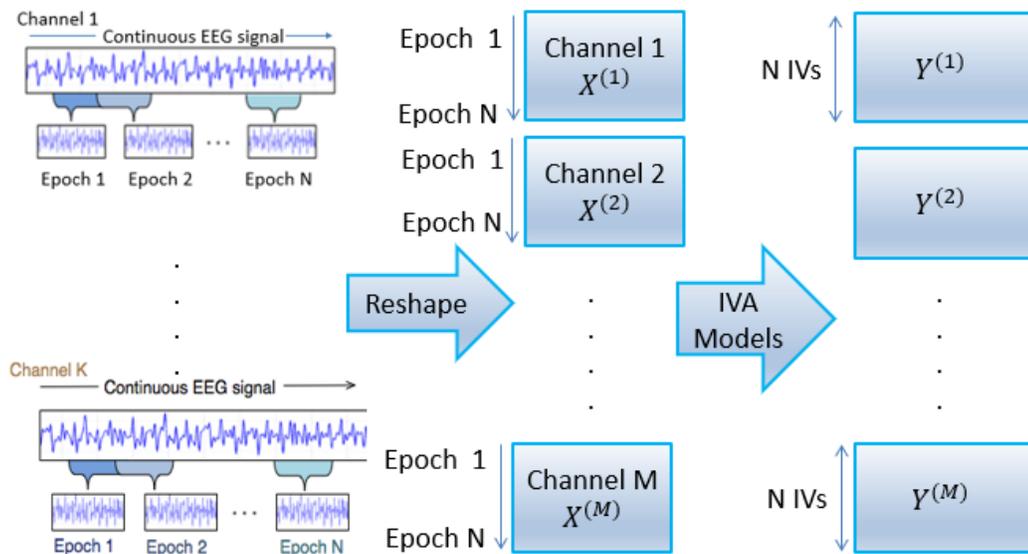


Figure 2.9: EEG block diagram. Illustration of epoching on continuous EEG data to build a dataset for IVA algorithms to obtain estimated components for EEG artifact removal. SCVs formed from estimates are also depicted.

IVA algorithms for ten repeated experiments. When IVA has separated artifactual sources, artifact rejection can be implemented by automatically detecting and removing the sources corresponding to the artifact. IVA, IVA-BGGMM, IVA-MGGMM, IVA-BMGGMM, ICA, ICA-BGGMM, ICA-MGGMM, and ICA-BMGGMM are evaluated on ten separate experiments, each with its own set of EEG data. The objective measures averaged out over these ten separated experiments. The objective

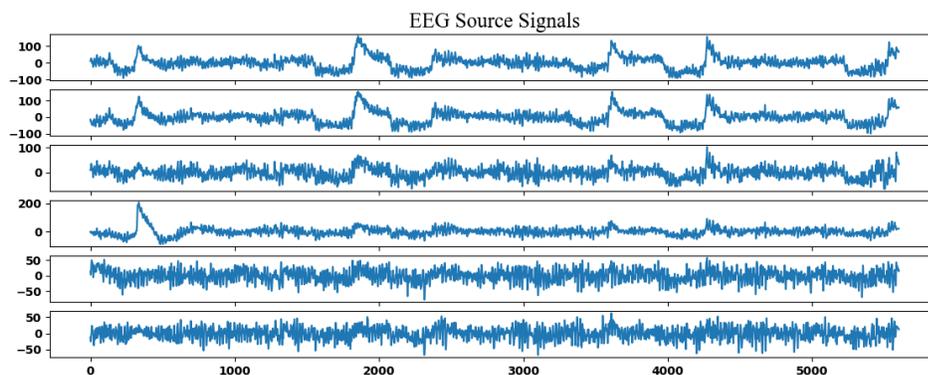


Figure 2.10: EEG channels before eye blink removal. EEG signal contaminated by the artifacts of eye blinking.

measures (SDR, SIR, SAR, SNR, and ISI) obtained after the recovery of EEG signals are shown in

Table 2.15. We find that the IVA-BMGGMM model exceeds the compared models, including IVA-MGGMM, IVA-BGGMM, IVA, ICA-BMGGMM, ICA-MGGMM, ICA-BGGMM, and ICA for the majority of performance metrics. SDR (51.89), SIR (59.02), SAR (259.97), and SNR (34.34), for the proposed model IVA-BMGGMM, are highly significant compared to other models.

Figs. 2.10 and 2.11 indicate EEG signals prior and after artifact removal using BSS, and it is observed that EOG artifacts were removed correctly. For visualization purposes, we only show the first 6 EEG signals out of a total of 19. The first six EEG channels' results demonstrate that IVA-BMGGMM can eliminate eye blinking caused by Fp1 and Fp2 EEG electrodes placed around the eye. For more clarification, we took the FP1 channel as an example to clarify the performance of IVA-BMGGMM to remove EEG artifacts as shown in Fig. 2.12. BSS experiments show that the IVA-BMGGMM model has better performance than compared models namely: IVA-MGGMM, IVA-BGGMM, IVA, ICA-BMGGMM, ICA-MGGMM, ICA-BGGMM, and ICA. This success in IVA-BMGGMM can be used to improve the performance of the existing BSS system that uses EEG artifact removal.

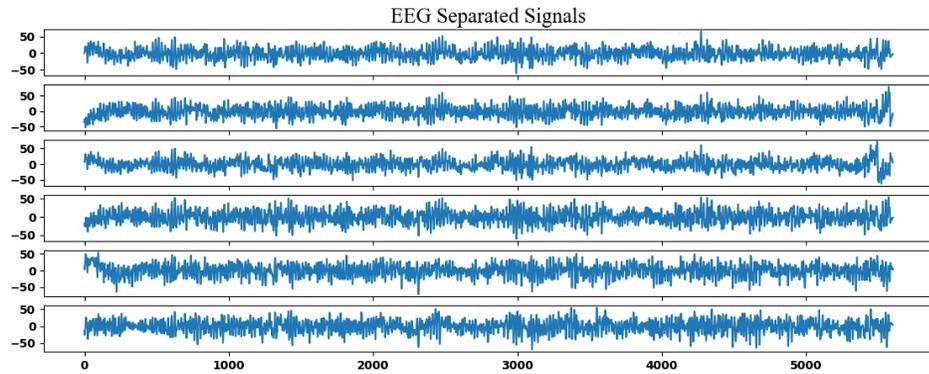


Figure 2.11: EEG channels after eye blink removal. The obtained EEG signals after removing eye blink artifacts with IVA-BMGGMM.

### Schizophrenia detection

Early detection of mental disorders is still a challenging task. Schizophrenia is a common psychiatric brain disorder that affects about 1% of the global population [70]. This disorder affects many brain functions and is distinguished by hallucinations, disorganized thinking, delusions, and

Table 2.15: Objective measure for separation of EEG signals.

Model	BSS of EEG				
	SDR	SIR	SAR	SNR	ISI
<b>IVA-BMGGMM</b>	<b>51.89</b>	<b>59.02</b>	<b>259.97</b>	<b>34.34</b>	<b>0.33</b>
<b>ICA-BMGGMM</b>	<b>46.13</b>	<b>43.74</b>	<b>254.35</b>	<b>29.43</b>	<b>0.45</b>
IVA-MGGMM	47.51	44.98	251.82	32.03	0.37
ICA-MGGMM	45.02	42.76	249.08	26.77	0.52
IVA-BGGMM	49.03	46.33	253.77	27.50	0.60
ICA-BGGMM	45.16	42.98	252.17	24.55	0.53
IVA	44.48	42.93	247.51	30.64	0.41
ICA	44.26	41.90	245.20	22.65	0.63

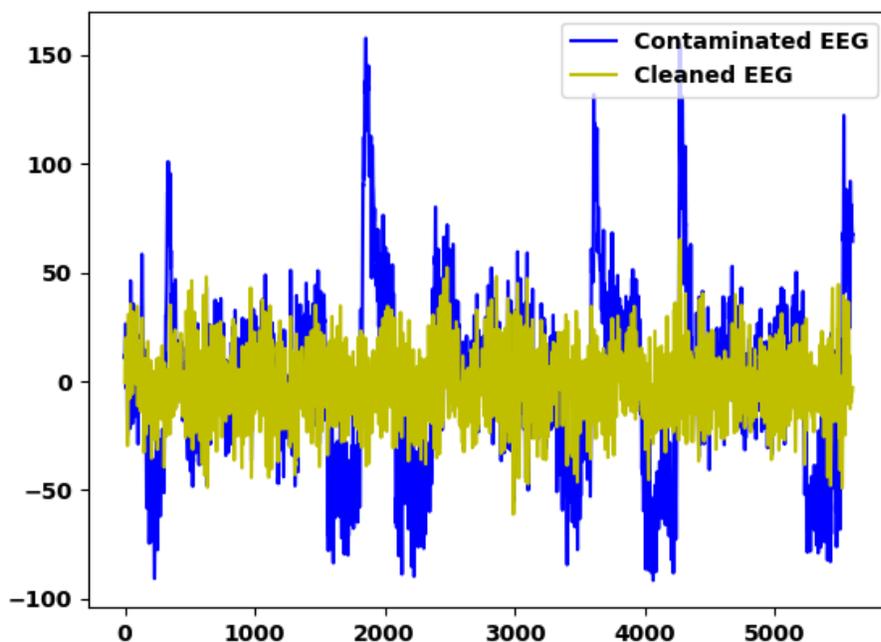


Figure 2.12: Fp1 channel before and after BSS. The blue EEG signal represents EEG signals contaminated by eye-blinking artifacts. The yellow one represents the clean EEG signals obtained after IVA-BMGGMM was used to eliminate eye blinking.

disintegration of reality perception [71]. We used the Center of Biomedical Research Excellence (COBRE) dataset<sup>7</sup> to classify schizophrenic patients and healthy subjects. Preprocessing was performed on the data to correct motion correction and spatial normalization into standard Montreal

<sup>7</sup><https://fcon1000.projects.nitrc.org/indi/retro/cobre.html>

Neurological Institute template (MNI). The preprocessed data was parcellated into 116 regions using automated anatomical labeling (AAL) to extract the regions of interest signals (ROIs). IVA, IVA-BGGMM, IVA-MGGMM, IVA-BMGGMM, ICA, ICA-BGGMM, ICA-MGGMM, and ICA-BMGGMM are applied. To get the activated regions, we reconstruct the original dimensions. All these steps are illustrated in Fig. 2.13.

In MATLAB software, the data processing and analysis for brain imaging (DPABI) tool [72] was used to execute all preprocessing steps. For each participant, slice time correction was applied for interleaved acquisition. Head motion correction based on Friston's 24-parameter motion model [73] was performed. Co-registration of structural and functional images to map the functional information to the anatomical space was executed. Then, the images were spatially normalized to the MNI standard using the DARTEL template [74] and resampled to  $3 \times 3 \times 3 \text{ mm}^3$ . The generated images were spatially smoothed with a 4 mm full-width half-maximum (FWHM) Gaussian kernel. Additionally, a band-pass filter with a frequency range of 0.01 – 0.1 Hz was employed to linearly detrend and temporally filter the images, effectively reducing low-frequency drifts and eliminating physiological high-frequency noise. Then, we applied brain parcellation using AAL atlas on the preprocessed data. IVA, IVA-BGGMM, IVA-MGGMM, IVA-BMGGMM, ICA, ICA-BGGMM, ICA-MGGMM, and ICA-BMGGMM algorithms are executed to evaluate performance on ten separated experiments, each with separate fMRI data. SDR, SIR, SAR, SNR, and ISI measuring parameters are used to compare performance. These measures are averaged over ten times for fMRI data. In this experiment, it is observed that IVA-BMGGMM outperforms IVA-MGGMM, IVA-BGGMM, IVA, ICA-BMGGMM, ICA-MGGMM, ICA-BGGMM, and ICA models for the majority of performance metrics as described in Table 2.16. Although the improvement is not highly significant for IVA-BMGGMM (SDR = 65.58 and SIR = 59.87) compared to ICA (SDR = 61.15 and SIR = 55.37), these findings open the door to do more modifications on IVA models to be used instead of ICA for detecting the brain activation regions in schizophrenic patients. The activated regions for schizophrenic patients are illustrated in Fig. 2.14. We picked (5, 7, 18) IVA components to see the different defected regions in schizophrenic patients. It is observed from Fig. 2.14 that the frontal network and default mode network (DMN) is highly activated in schizophrenic patients. The

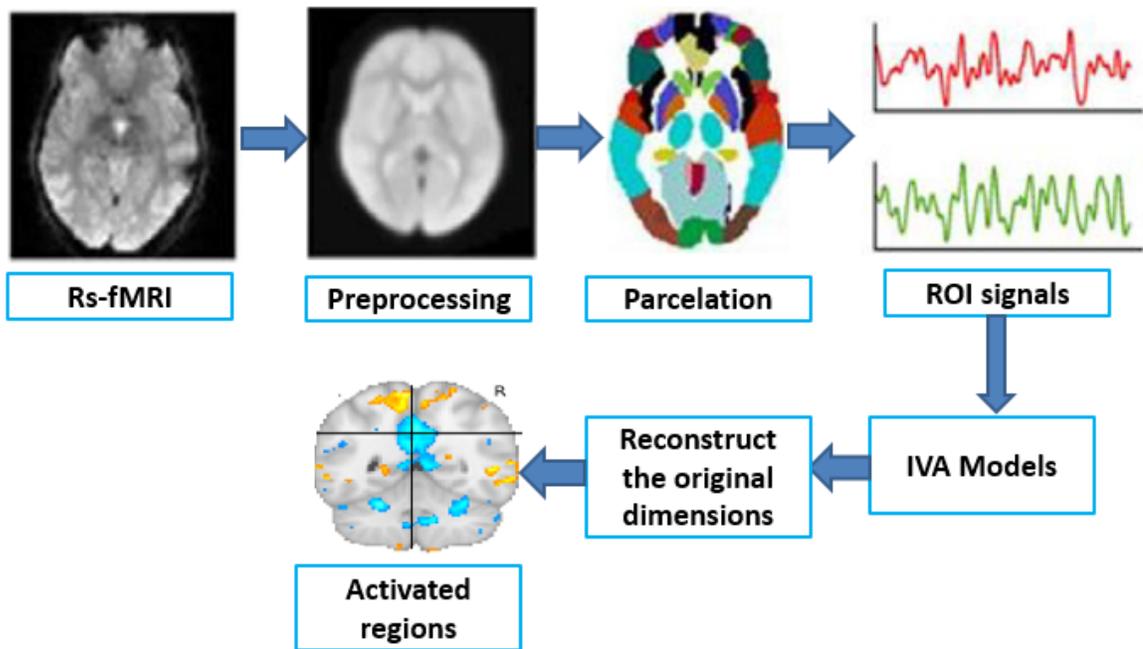


Figure 2.13: Block diagram of the proposed schizophrenia detection system.

DMN is responsible for remembering and thinking about future whereas the frontal network controlling decision making, planing and organizing. As we can see, these two networks are defective in schizophrenic patients.

Table 2.16: Objective measure for fMRI signals.

Model	BSS of fMRI				
	SDR	SIR	SAR	SNR	ISI
<b>IVA-BMGGMM</b>	<b>65.58</b>	<b>59.87</b>	<b>135.92</b>	<b>73.39</b>	<b>0.32</b>
<b>ICA-BMGGMM</b>	<b>63.23</b>	<b>57.61</b>	<b>135.39</b>	<b>33.89</b>	<b>0.37</b>
IVA-MGGMM	65.17	59.65	135.76	65.34	0.31
ICA-MGGMM	61.69	56.25	133.91	31.20	0.39
IVA-BGGMM	63.02	55.79	133.93	29.94	0.38
ICA-BGGMM	61.64	54.11	132.09	26.29	0.36
IVA	65.01	59.38	135.65	62.41	0.36
ICA	61.15	55.37	133.87	29.94	0.41

### Fetal ECG extraction

The previous Section explained the context for this experiment 2.6.3. Traditional maternal techniques that focused primarily on fetal heartbeat and uterine contractions were used to ensure the

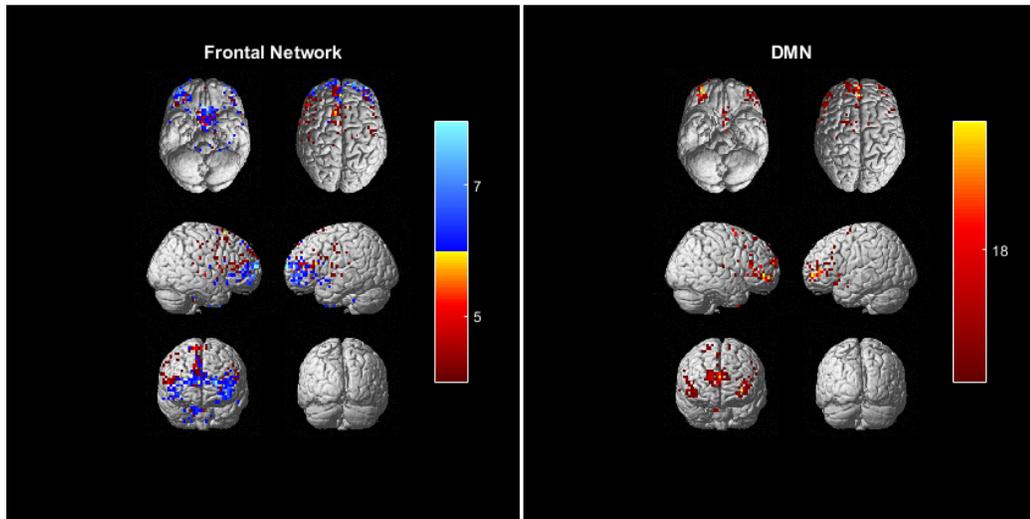


Figure 2.14: The defected regions for schizophrenic patients. On the left are the defected regions for frontal networks. The activated regions in the DMN are shown on the right.

well-being of the fetus [75]. However, due to the involuntary movements of the fetus and the small size of the heart, assessing the fetal heart is still challenging. Therefore, new technologies such as machine learning that improve initial obtained images, help extract measures, or aid in cardiac problems detection are critical for optimal fetal heart assessment. For this experiment, the ECG dataset<sup>8</sup> contains nine channels; we only keep the first three recordings as shown in Fig. 2.15, that were captured using electrodes placed on the abdomen of a pregnant woman. Each captor signal contains 2500 measurements. IVA mixture model is used to estimate basis functions  $2 \times 2$ ,  $3 \times 3$ ,

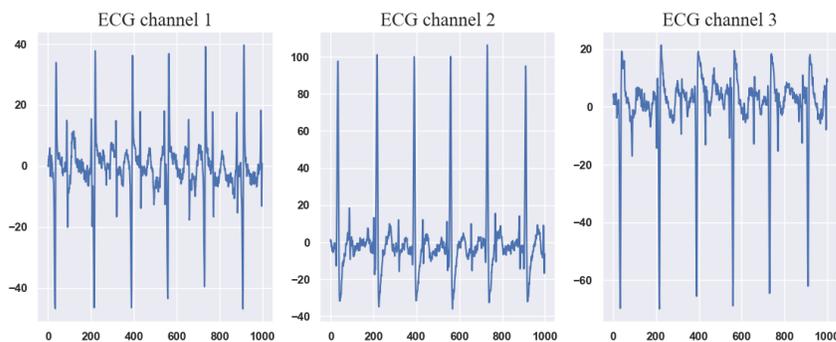


Figure 2.15: Original ECG channels. These three channels were extracted from a pregnant women dataset.

$4 \times 4$ , and  $5 \times 5$  in separate experiments to compute 2, 3, 4, and 5 sources. Only ECG signals after

<sup>8</sup><https://homes.esat.kuleuven.be/tokka/daisydata.html>

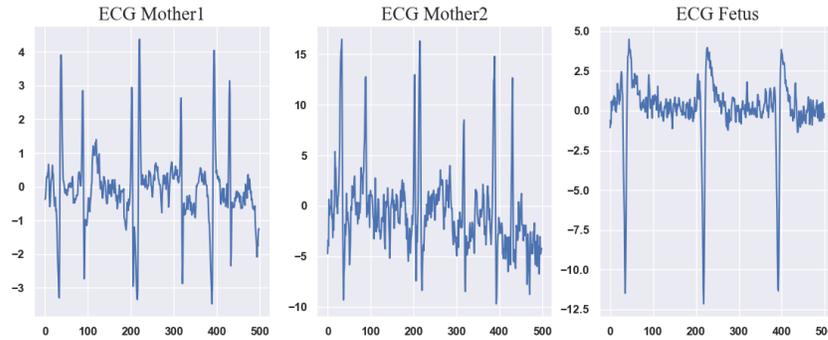


Figure 2.16: IVA-BMGGMM ECG channels. On the left are the extracted IVA-BMGGMM ECG signals for mother. On the right is the fetal ICA-BMGGMM ECG signal.

linear mixing are visible in BSS. There is no prior knowledge of the basis functions that will be used during the source separation. Subjective and objective measures were used to evaluate the IVA mixture models. ECG signals obtained before and after source separation are used in the subjective analysis. The objective analysis includes SDR, SIR, SAR, SNR, and ISI.

We conducted four experiments using this BSS framework to compute 2, 3, 4, and 5 ECG sources. We used a linear mixture of 2, 3, 4, and 5 ECG sources from the database and conducted BSS using six models (IVA, IVA-BGGMM, IVA-MGGMM, IVA-BMGGMM, ICA, ICA-BGGMM, ICA-MGGMM, and ICA-BMGGMM) to recover 2, 3, 4, and 5 ECG sources, respectively. The recovered sources are subjected to objective analysis to determine the quality of recovered ECG signals and the viability of IVA and ICA mixture models in BSS. The experiment is repeated ten times using different linear ECG mixtures of 2, 3, 4, and 5 sources from the original database, with the objective measures averaged out over these ten separated experiments.

Tables 2.17 and 2.18 show objective measures obtained after the recovery of ECG source signals. We find that the IVA-BMGGMM model outperforms IVA-MGGMM, IVA-BGGMM, IVA, ICA-BMGGMM, ICA-MGGMM, ICA-BGGMM, and ICA models for 2, 3, 4, and 5 sources. In BSS of 2 ECG signals, we observed a highly significant improvement for both IVA-BMGGMM and ICA-BMGGMM compared to other models. SDR (37.26), SIR (36.59), SAR (269.65), and SNR (39.09) for IVA-BMGGMM are the best measures among all performance metrics. A comparison of ECG signals before and after BSS can be seen in Figs. 2.15 and 2.16. Fig. 2.16 shows the ability of IVA-BMGGMM to recover the fetal and mother ECG signals from the mixed ECG signal. The

fetal ECG signal has a higher frequency than the mother’s ECG signal. The above BSS experiments show the effectiveness of IVA-BMGGMM compared with widely used different models such as IVA-MGGMM, IVA-BGGMM, IVA, ICA-BMGGMM, ICA-MGGMM, ICA-BGGMM, and ICA. It is also observed that as the number of linear mixtures in source separation increases, the improvement rate decreases. These findings show that as the mixing complexity increases, the ability of IVA-BMGGMM to recover the sources decreases. This is an IVA-BMGGMM limitation that can be addressed in future work.

Table 2.17: Objective measure for separation of 2 and 3 ECG signals.

Model	BSS of 2 ECG signals					BSS of 3 ECG signals				
	SDR	SIR	SAR	SNR	ISI	SDR	SIR	SAR	SNR	ISI
<b>IVA-BMGGMM</b>	<b>37.26</b>	<b>36.59</b>	<b>269.65</b>	<b>39.09</b>	<b>0.28</b>	<b>33.50</b>	<b>32.94</b>	<b>259.22</b>	<b>29.49</b>	<b>0.31</b>
<b>ICA-BMGGMM</b>	<b>28.46</b>	<b>27.09</b>	<b>261.57</b>	<b>32.40</b>	<b>0.45</b>	<b>19.01</b>	<b>19.02</b>	<b>256.79</b>	<b>19.34</b>	<b>0.45</b>
IVA-MGGMM	31.69	31.25	263.08	37.92	0.32	26.89	26.46	256.62	26.65	0.33
ICA-MGGMM	9.50	9.50	255.58	16.80	0.46	9.91	9.91	251.63	16.22	0.47
IVA-BGGMM	28.84	28.84	259.48	25.85	0.49	26.99	25.99	245.85	20.95	0.51
ICA-BGGMM	9.20	9.21	253.32	20.31	0.45	14.63	14.53	235.45	20.22	0.45
IVA	26.26	25.75	258.31	27.26	0.40	21.84	21.41	252.39	23.38	0.42
ICA	8.76	8.76	246.20	15.74	0.47	8.45	8.46	248.85	15.88	0.49

Table 2.18: Objective measure for separation of 4 and 5 ECG signals.

Model	BSS of 4 ECG signals					BSS of 5 ECG signals				
	SDR	SIR	SAR	SNR	ISI	SDR	SIR	SAR	SNR	ISI
<b>IVA-BMGGMM</b>	<b>31.73</b>	<b>31.17</b>	<b>259.72</b>	<b>26.29</b>	<b>0.30</b>	<b>33.41</b>	<b>32.56</b>	<b>258.25</b>	<b>25.98</b>	<b>0.37</b>
<b>ICA-BMGGMM</b>	<b>14.01</b>	<b>14.02</b>	<b>256.26</b>	<b>14.26</b>	<b>0.44</b>	<b>14.08</b>	<b>14.87</b>	<b>251.79</b>	<b>19.26</b>	<b>0.43</b>
IVA-MGGMM	29.72	28.25	256.66	24.29	0.36	29.67	29.81	256.04	24.81	0.38
ICA-MGGMM	10.44	10.45	251.86	15.03	0.47	12.16	10.16	249.68	17.78	0.48
IVA-BGGMM	29.91	29.71	253.34	20.35	0.52	19.74	20.13	253.16	24.40	0.51
ICA-BGGMM	9.48	9.58	252.80	14.29	0.52	10.59	10.56	251.76	17.12	0.44
IVA	25.05	24.53	251.64	21.58	0.42	26.04	25.17	252.26	22.49	0.43
ICA	7.36	7.36	248.66	13.22	0.48	10.10	8.10	245.37	16.43	0.49

## 2.7 Conclusion and discussions

In this Chapter, we proposed the BMGGMM to address the issues associated with unbounded mixture models. Furthermore, we proposed BMGGMM for data clustering in various applications, including medical diagnosis for diseases such as stroke, Alzheimer, skin cancer, and diabetes, as well as data analysis for human resources. Using different metrics, we verified the effectiveness

of BMGGMM, which outperforms GMM, GGMM, BGGMM, and MGGMM in clustering several data sets. The number of optimal components that best characterize the data is critical for the mixture model. As a result, we proposed MML as a model selection criterion using BMGGMM to accurately determine the number of clusters in a dataset to enhance unsupervised learning of the mixture model. Moreover, we validated MML over the same applications.

This work proposed BMGGMM with ICA as a model for statistical learning, which is further proposed in BSS to validate the algorithm's effectiveness using speech, ECG, and EEG databases. This algorithm overcomes the limitations of ICA, which assumes that sources are independent of one another. The presented applications have demonstrated the proposed model's success. According to the results of the BSS experiments, ICA-BMGGMM outperforms ICA, ICA-BGGMM, and ICA-MGGMM models. It is also observed that the rate of improvement slows as the number of linear mixtures in source separation increases.

In this work, we also proposed BMGGMM with independent vector analysis as a generalization for the ICA-BMGGMM model. The proposed model relaxed the ambiguity issue related to the ICA model, and it is more applicable for data with multiple sources such as fMRI and EEG databases. IVA-BMGGMM was validated through different medical datasets, including ECG, EEG, fMRI, and speech signals, to assess the success of the proposed model compared to ICA and IVA different models. The results of the proposed model (IVA-BMGGMM) are compared with (IVA-MGGMM, IVA-BGGMM, IVA, ICA-BMGGMM, ICA-MGGMM, ICA-BGGMM, and ICA) models for speech and medical datasets, where IVA-BMGGMM demonstrated its effectiveness.

## Chapter 3

# ICA and IVA bounded multivariate generalized Gaussian mixture based hidden Markov models

Machine learning, a branch of artificial intelligence, is an area of computational science that is concerned with the analysis and interpretation of patterns and structures in data to enable learning and decision-making without the participation of a human. Hidden Markov models (HMMs), which have been acknowledged for decades but have recently made a significant revival in machine learning, are one of the most impressively powerful probabilistic models. HMMs are frequently employed in machine learning to model heterogeneous time series data. In this Chapter, we integrate traditional ICA and IVA with a bounded multivariate generalized Gaussian mixture model (ICA-BMGGMM) into the HMM approach. One limitation of ICA is that it assumes the sources to be independent from each other. This assumption can be relaxed by combining standard IVA and BMGGMM (IVA-BMGGMM) into the HMM approach to improve their modeling capability. We validate our proposed models using a variety of applications, such as human action recognition, speech recognition, and energy disaggregation. The results presented in the work demonstrate the effectiveness of the proposed approaches for modeling different types of data. These data include KTH and Weizmann datasets for human action recognition, TIMIT and SDR for

speech recognition, REDD dataset for energy disaggregation and EEG dataset for elliptic Seizure classification. For all conducted experiments, our proposed models outperform other comparing models for all performance metrics such as accuracy, sensitivity, and precision. The best detection results were found using the IVABMGGMM-HMM for the reported experiments.

### 3.1 Introduction

Numerous approaches, including ICA, principal component analysis (PCA), and IVA, have been developed to address the BSS problem [76]. ICA effectively solves BSS under the assumption that observations are linear, statistically independent mixtures of sources, separated by maximizing output independence [77, 78]. This assumption is relaxed in the ICA mixture model, which classifies data into exclusive classes and combines independent sources using model components [34, 36]. To address unbounded mixture model limitations, ICA with a bounded multivariate generalized Gaussian mixture model (BMGGMM) has been proposed [37, 79, 80, 38]. In previous work, we incorporated BMGGMM into ICA by leveraging statistical independence and sparsity [2].

As mentioned in [81], ICA is limited by permutation and scaling issues. To tackle these constraints, the IVA was proposed to reduce the permutation problem that occurs in most of the BSS algorithms [40]. IVA decomposes each dataset into a linear combination of statistically independent variables that take into account cross-dataset correlations. It has been successfully applied to a variety of applications where joint source separation is required, and statistical information from multiple datasets must be fully leveraged, such as joint analysis of electroencephalographic datasets [82], fMRI applications [83], and multi-channel audio array processing [84]. To develop IVA, a variety of algorithms have been proposed that take into account different kinds of statistical features such as second-order statistics (SOS) and higher-order statistics (HOS). IVA-Gaussian (IVA-G) [49] takes full use of SOS, but IVA-Laplacian (IVA-L) [85], which assumes a Laplacian distribution as source prior, uses only statistics greater than two. Multivariate generalized Gaussian distribution over the IVA model (IVA-GGD) [42] assumes a multivariate generalized Gaussian distribution (MGGD) as the source prior. This model includes a wide range of unimodal distributions

such as sub-Gaussian, super-Gaussian, and normal distributions and thus exploits SOS and HOS between and within datasets. When estimating the scatter matrix, IVA-GGD assumes that the samples are independent and identically distributed (i.i.d.) and uses a fixed set of shape parameter values. For problems with unbounded distributions for data that fall within bounded regions, the IVA model is enhanced using the BMGGMM. The bounded regions are included in IVA-BMGGMM, which is an improvement over IVA-MGGD.

A hidden Markov model (HMM) is a dynamic probabilistic approach that is employed in several domains, including speech processing [86], object and gesture classification [87], [88], and unusual event detection [89],[90]. HMM has received a lot of attention from researchers and has been used successfully in a variety of fields as it has unique properties that other traditional learning methods don't have [91, 92]. HMM, a dual stochastic process, has hidden Markov chains with a fixed number of states and a set of observable random functions. However, each function is significant to a state of the chains, and the hidden process may be represented using the visible process's sequences. HMM is well-suited to modeling dynamic time series and has good pattern classification capabilities, particularly for signals with a large amount of information.

According to the literature, several studies have employed ICA and IVA in speech recognition [93, 94, 95, 96], object detection [97, 98], artifact removal for EEG data [99, 100] and estimating the time-varying representations for large-scale resting-state fMRI data [101]. To tackle the ICA limitations, the integration of mixture models and ICA has been proposed in several studies [38, 37, 80, 102]. Moreover, ICA and HMM have been combined in [103] and [104] for fatigue detection and fault recognition respectively. We integrate the IVA technique into the HMM framework, inspired by the integration of ICA and HMM. To the best of our knowledge, this is the first study of the IVA approach's integration with the HMM scheme. Furthermore, we propose the ICA mixture combined with HMM to relax the issues related to ICA. Finally, we propose the integration of the IVA-BMGGMM and the HMM framework to improve the IVA-HMM scheme. The IVABMGGMM-HMM is a more general IVA-HMM implementation that uses the shape parameter and covariance matrix to account for both second and higher-order statistics. In addition, this model is more suitable for data that fall in bounded regions.

To validate these three proposed models, we also considered several applications. To start,

we employed our models for human action recognition. In computer vision and pattern recognition, human activity modeling has been an important research area [105]. Robotics, health care, video surveillance systems, and human-to-computer interfaces are applications that use activity recognition. The challenge for action recognition systems is accurately classifying incoming data (video/images). Feature extraction and action learning are the two most important processes in human activity recognition. We picked the KTH dataset [106] to validate the proposed models. We also used the Weizmann data to test our models' ability to identify different actions. Furthermore, we applied the proposed algorithms for speech recognition. In this case, the TIMIT dataset with eight dialects is used to validate our models. For more than two decades, TIMIT's phone recognition has been the subject of extensive research, and its performance has steadily improved over time. When it comes to evaluation, there are several methods available, but they focus on three areas: phone segmentation [107], phone classification [108], and phone recognition [109, 110]. In our case, we used TIMIT to validate our models capability for phone recognition applications. Using the spoken digit dataset, we extended the validity of our models in the field of speech recognition. We tested our models' ability to recognize the digit classes in this data. Moreover, we applied our models to energy disaggregation, a process also known as non-intrusive load monitoring (NILM). NILM plays a crucial role in modern smart grid applications, offering significant potential in areas such as demand-side management, load forecasting, and energy conservation. The goal of energy disaggregation is to use aggregated meter data to estimate the energy usage of each appliance on a customer site. Various approaches for NILM have been proposed, each employing a different signal processing and machine learning technique (reviews can be found in [111, 112]). The HMMs and their variants (see, for example, [113, 114]), signal processing methods such as dynamic time warping [115, 116], single-channel source separation [117], graph signal processing [118, 119], and support vector machines with K-means [120] are among the approaches proposed. In our case, we proposed integrating the IVA, IVA-BMMGGMM, and ICABMGGMM into the HMM framework. These models are able to model both Gaussian and non-Gaussian data. Furthermore, the IVABMGGMM-HMM and the ICABMGGMM-HMM can handle correlated features in bounded support regions. For this application, we used the REDD dataset to test the capability of our proposed models to categorize the active state of the selected appliances which plays an important role

in energy saving.

The rest of this Chapter is organized as follows: The proposed models (ICA-BMGGMM, IVA, IVA-BMGGMM) integrated with HMM are discussed in Section 3.2. Section 3.3 evaluates the proposed method's performance using different experiments. Section 3.4 concludes our Chapter.

## 3.2 Integration of ICA and IVA into the HMM framework

### 3.2.1 ICA mixture of bounded multivariate generalized Gaussian distributions

The observed data in the ICA mixture model is assumed to come from a mixture model and be classified into mutually exclusive classes [46], where each data vector  $\mathbf{x}_n$  can be represented as:

$$\mathbf{x}_n = A_j \mathbf{s}_j + \mathbf{b}_j \quad (3.1)$$

ICA-BMGGMM is defined by substituting Equation (3.1) in Equation (2.4). The ICA mixture model is described in Chapter 2.

### 3.2.2 IVA mixture of bounded multivariate generalized Gaussian distributions

IVA considers components within a source to be independent, and components taken from different mixture sources to be maximally dependent. As shown in Fig. 2.1, equivalent component dependency across many sources and component independence within a single source is maximized simultaneously. This cannot be accomplished by performing individual ICAs on each source [50].

In IVA, each mixture source  $\mathbf{x}_n^{[m]}$ ,  $m = 1, \dots, M$  is a linear mixture of  $N$  independent sources.

$$\mathbf{x}_n^{[m]} = A_j^{[m]} \mathbf{s}_j^{[m]}, \quad m = 1, \dots, M \quad (3.2)$$

The mathematical Equations for the IVA mixture are given in detail in Chapter 2.

### 3.2.3 Hidden Markov model

The hidden Markov model is a probabilistic model used to describe random process statistical features. It is divided into two parts: the hidden Markov chain (known as the hidden layer) and the observed quantity (known as the observation layer). A collection of parameters governs a hidden Markov model, such as the number of state transitions and the emission probability. HMM-based modeling has three main tasks: the first is to optimize the model's parameters given training data; the second is scoring, which calculates the joint probability of a sequence given the model; and the third is decoding, which finds the optimal series of hidden states as shown in Fig. 3.1.

It is worth mentioning that the emission probability for continuous observations is generally taken as a Gaussian mixture [86, 87, 121]. For any applications of HMM, we always have to solve three major problems:

- The evaluation problem: This is mainly considered by computing the probability of a particular sequence of observations. This problem can be solved by the forward and backward algorithms [122, 123].
- The decoding problem: It can be tackled by finding the most likely sequence of hidden states which could have generated a given sequence of observations. This problem can be solved by the Viterbi algorithm [124, 125].
- The learning problem: This is also considered as the selection process among all models to determine which model best matches the observations. We use the Baum–Welch algorithm [121] to find the best set of state transitions and emission probabilities to estimate the parameters that maximise the probability of a given set of observations (see Fig. 3.2).

HMM can be defined by the following parameters: according to [86], given time-series observations  $\mathcal{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$  generated by hidden states  $Z = [z_1, z_2, \dots, z_N]$ ;  $z_k \in [1, K]$ , where  $K$  is the number of the hidden states, a transition probabilities matrix:  $B_{jk} = p(z_n = j | z_{n-1} = j')$  and the emission probabilities matrix:  $p(\mathbf{x}_n | \Lambda)$  are defined. The initial probability is defined as  $\pi_j$ , which is the probability of starting the observation sequence from state  $j$ . The joint probability

distribution for both hidden states and observable variables is written as:

$$p(\mathcal{X}, Z|\Theta) = p(Z_1|\pi) \left( \prod_{n=2}^N p(z_n|z_{n-1}, B) \right) \prod_{n=1}^N p(\mathbf{x}_n|\Lambda) \quad (3.3)$$

where  $\Theta = \{\pi, B, \Lambda\}$  defines the set of parameters of HMM. There are numerous options for emission distribution, including Gaussian distribution and mixture models such as the GMM. In our case, we consider IVA, IVA-BMGGMM, and ICA-BMGGMM to be used as emission distributions.

### Maximum likelihood estimation

We used the maximum log-likelihood technique via the EM algorithm [126], which identifies the best matches to the data. The EM framework begins with initial parameters. Then, in the E step, we use the forward-backwards algorithm to determine the state's posterior distribution. We update parameters  $\Theta$  using this posterior distribution by maximizing the full data likelihood concerning each parameter in the M step as shown in Fig. 3.2. By denoting  $Z$  as hidden variables and  $\mathcal{X}$  as the data, we can express the data likelihood as:

$$\mathcal{L}(\Theta, \Theta^{old}) = \sum_Z p(Z|\mathcal{X}, \Theta^{old}) \log p(\mathcal{X}, Z|\Theta) \quad (3.4)$$

By introducing the Lagrange multiplier  $\lambda$ , the marginal posterior distribution and the joint posterior distribution are given by:

$$\lambda(Z_{nk}) = p(Z_{nk}|\mathcal{X}, \Theta) \quad (3.5)$$

$$\xi(Z_{n-1,j}, Z_{nk}) = p(Z_{n-1,j}, Z_{nk}|\mathcal{X}, \Theta) \quad (3.6)$$

Here,  $\lambda(Z_{nk})$  represents the conditional probability where  $Z_{nk} = 1$  if  $\mathbf{x}_n$  is emitted from the  $k$ th

state and 0 otherwise. Then, the complete-data log-likelihood is expressed as:

$$\begin{aligned} \log \mathcal{L}(\mathcal{X}, Z|\Theta) &= \sum_{k=1}^K \lambda(Z_{1k}) \log \pi_k + \sum_{n=2}^N \sum_{j=1}^M \sum_{k=1}^K \xi(Z_{n-1,j}, Z_{nk}) \log B_{jk} \\ &+ \sum_{n=1}^N \sum_{k=1}^K \lambda(Z_{nk}) \log p(\mathbf{x}_n|\Lambda_{nk}) \end{aligned} \quad (3.7)$$

where  $\lambda(Z_{1k})$  is the conditional probability,  $\xi(Z_{n-1,j}, Z_{nk})$  is the posterior distribution and  $p(\mathbf{x}_n|\Lambda_{nk})$  is the emission distribution.

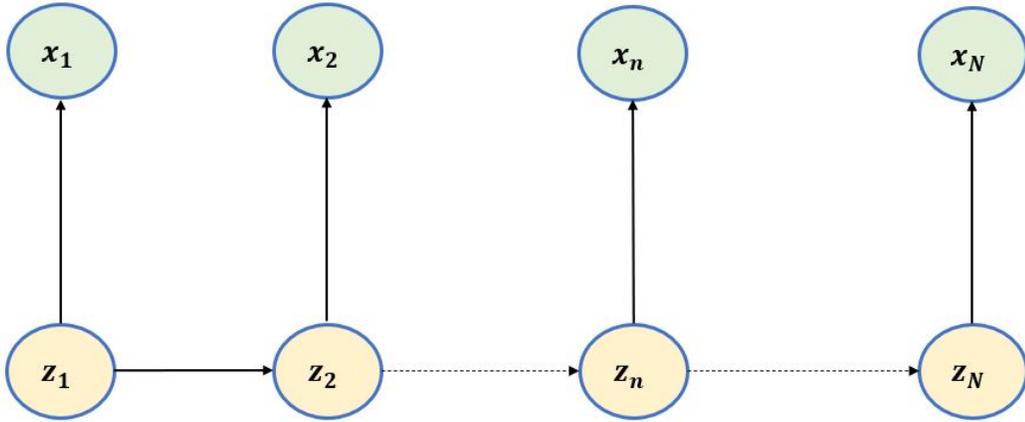


Figure 3.1: Graphical representation for the HMM model.

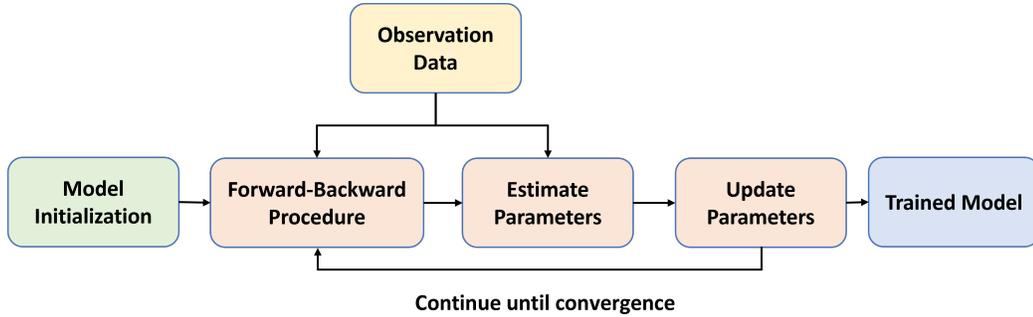


Figure 3.2: Training procedures for HMM.

### 3.2.4 ICA-BMGGMM and IVA-BMGGMM integration into the HMM framework

As demonstrated in the previous subsection, the emission distribution  $p(\mathbf{x}_n|\Lambda_{nk})$  is commonly used as the GMM for the majority of continuous observation cases. In this Chapter, we propose to

integrate IVA, IVA-BMGGMM, and ICA-BMGGMM into the HMM scheme. The primary motivation behind the choice of adopting the ICA-BMGGMM, IVA and IVA-BMGGMM distributions as emission probabilities is the capability of these distributions to model both Gaussian and non-Gaussian data. This integration can be done by substituting the  $p(\mathbf{x}_n|\Lambda_{nk})$  in Equation (3.7) with Equations (3.2) and (2.4) for IVA-BMGGMM. Similarly, to integrate ICA-BMGGMM to the HMM model, we substitute the  $p(\mathbf{x}_n|\Lambda_{nk})$  in Equation (3.7) by Equations (3.1) and (2.4). A schematic of this integration is shown in Fig. 3.3. To estimate the parameters, we maximize the expectation of the complete-data log-likelihood in Equation (3.7) with respect to  $\pi, B, \Lambda$  as:

### Estimation of $\pi$ and $B$

The maximization of  $\pi_k$  and  $B_{jk}$  using the Lagrange multiplier are given as:

$$\pi_k = \frac{\lambda(Z_{1k})}{\sum_{j=1}^M \lambda(Z_{1j})} \quad (3.8)$$

$$B_{jk} = \frac{\sum_{n=1}^N \xi(Z_{n-1,j}, Z_{nk})}{\sum_{k=1}^K \sum_{n=2}^N \xi(Z_{n-1,j}, Z_{nk})} \quad (3.9)$$

where  $\pi_k$  and  $B_{jk}$  follow these constrains:  $\sum_{k=1}^K \pi_k = 1$  and  $\sum_{k=1}^K B_{jk} = 1$  respectively. The estimation of  $\Lambda$  is detailed in Appendix B.

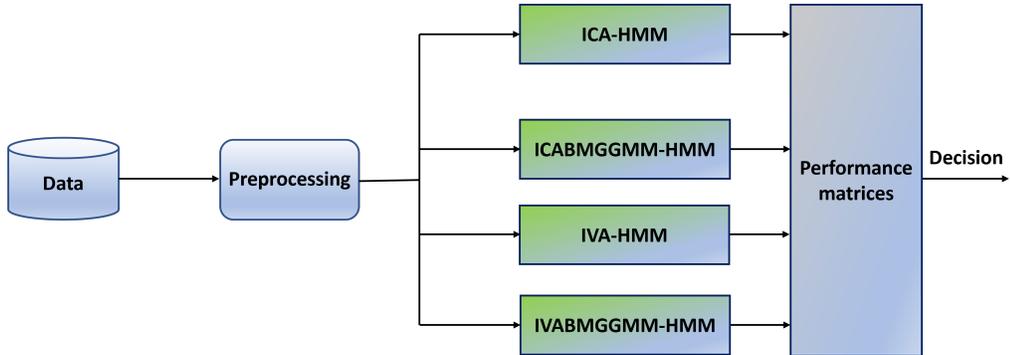


Figure 3.3: Integration of the proposed models into the HMM framework.

The complete learning of ICABMGGMM-HMM is given in Algorithm 3. The goal of this algorithm is to find the optimal parameters of the ICABMGGMM-HMM. The gradient descent method [127, 128] is used in the training process. The optimal estimation of the parameter  $\Theta$  is obtained as follows:

$$\Theta^* = \arg \max_{\Theta} \mathcal{L}(\mathcal{X}, Z|\Theta) \quad (3.10)$$

where  $\mathcal{L}(\mathcal{X}, Z|\Theta)$  is the complete-data likelihood. During the testing process, log-likelihood feature functions are calculated using the parameters from ICABMGGMM-HMM and IVABMGGMM-HMM learned during the training stage. Therefore, the most probable class label  $\mathbf{y}^*$  of the testing sequence is inferred as follows:

$$\mathbf{y}^* = \arg \max_{\mathbf{y} \in Y} \mathcal{L}(\mathbf{y}|\mathcal{X}, \Theta^*) \quad (3.11)$$

---

**Algorithm 3** ICABMGGMM-HMM

---

- 1: Dataset  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ .
  - 2: Output:  $\{\pi, B, \Lambda\}$ .
  - 3: Initialization for  $\Theta = [\pi, B, \Lambda]$ .
  - 4: **while** iterations  $\leq$  Max iterations or parameters do not converge **do**
  - 5:     **E-step:**
  - 6:     **for do** $[\mathbf{x}_1, \dots, \mathbf{x}_n]$
  - 7:         Compute  $\lambda(Z_{nk})$  and  $\xi(Z_{n-1,j}, Z_{nk})$  use the forward-backward algorithm.
  - 8:         Compute the posterior distribution of the state using Equation (3.4).
  - 9:     **end for**
  - 10:    **M-step:**
  - 11:    **for**  $1 \leq j \leq M$  **do**
  - 12:        Update  $\pi_k$  and  $B_{jk}$  using Equations (3.8) and (3.9).
  - 13:        Update the mixing parameter  $\hat{p}_{kj}$  using Equation (2.10).
  - 14:        Update  $\boldsymbol{\mu}_{kj}$  using Equation (2.14).
  - 15:        Update  $\Sigma_{kj}$  using Equation (2.17).
  - 16:        Update  $\boldsymbol{\beta}_{kj}$  using the Newton-Raphson method: Equations (2.18) and (2.20).
  - 17:        Update the basis functions  $A_{kj}$  using Equation (2.36).
  - 18:        Update the bias vector  $\mathbf{b}_{kj}$  using Equation (2.38).
  - 19:    **end for**
  - 20: **end while** when the likelihood becomes stable.
-

### 3.3 Experimental results

We propose validating our proposed models (ICABMGGMM-HMM, IVA-HMM, and IVABMGGMM-HMM) by categorizing video and speech datasets. We used the KTH and Weizmann datasets for our human action recognition experiments. The KTH dataset includes six action classes: boxing, hand-clapping, hand-waving, running, walking, and jogging. In the Weizmann dataset, the actions are: bending, jumping jack, jumping, jumping in place, side jumping, running, skip jumping, walking, waving one hand, and waving two hands. Many speech processing applications use speech data clustering, such as speaker verification, classification, and speech recognition. We used the TIMIT and spoken digits datasets for the speech experiments. Both male and female speakers are represented in the TIMIT dataset. The spoken digits dataset contains ten types of data from various speakers. We extended the validation of our models to include an application related to energy consumption. We chose the REDD dataset for this experiment to validate our models' ability to detect the state of appliances. Furthermore, we tested the capability of our proposed models to distinguish between Seizure patients and healthy subjects using EEG signals.

In terms of performance metrics, we compared the performance of the employed models using both objective and subjective measures. We select the most important measures to demonstrate the differences in each model's performance for the objective metrics. Accuracy, sensitivity, specificity, precision, F1 score, negative predictive value (NPV), Matthews' correlation coefficient (MCC), FPR, false-negative rate (FNR), false discovery rate (FDR), and Log loss are the performance metrics. Accuracy, sensitivity, specificity, precision, F1 score, and Log loss are discussed in Chapter 2. The percentage of true negative labels to total negative labels is defined as the NPV. MCC measures the overall quality of the model by taking into account the total number of true labels as well as the false positive and negative labels for both classes. FPR and FNR are rates calculated from the confusion matrix, where the FPR is the ratio of false positive labels to negative labels and the FNR is the ratio of false-negative labels to positive labels. FDR is the ratio of false positive labels to total positive labels.

For the subjective measures, we used a confusion matrix, receiver operating characteristics (ROC) curves, and the area under the curve (AUC). A confusion matrix is a matrix that compares

the number of correct versus incorrect predictions for each class. Confusion matrix compares the actual target values to those predicted by the machine learning model. This provides us with a comprehensive picture of how well our classification model is performing and what types of errors it is making. For a binary classification problem, we would have a  $2 \times 2$  matrix with four values: True negative (TN), true positive (TP), false positive (FP), and false negative (FN). The ROC curve is a plot of the model’s performance (a plot of the true positive rate and the false positive rate) at all classification thresholds [129]. The AUC is used to measure the entire two-dimensional area under the curve and thus measuring the model’s performance at all possible classification thresholds.

### 3.3.1 Human action recognition

#### Weizmann data

Human action recognition is an important topic in computer vision and has various real-world applications. Human activity recognition seeks to assess and understand ongoing activities in video sequences. The Weizmann dataset includes 90 low-resolution videos of 10 natural actions (bending, jumping jack, jumping, jumping in place, side jumping, running, skip jumping, walking, waving one hand, waving two hands). These actions are shown in Fig. 3.4. To begin, we preprocessed the data by employing a fixed-size box around the actors in each video to maximize classification accuracy while minimizing background noise. Then, the dense optical flow features from the video sequences were extracted. Optical flow features are effective and accurate motion information representation in video sequences [130]. These dense optical flow features were computed between two consecutive frames  $t$  and  $t + 1$ . This results in  $2D$  vectors each representing the horizontal and vertical movement of a pixel at position  $(x, y)$ , respectively. Then we set one action video for testing and other action videos for training. We trained and tested our proposed models over the extracted dense optical flow features. Here, we set the number of states  $K = 10$ . From Table 3.1, we notice that the proposed models (ICABMGGMM-HMM, IVA-HMM, IVABMGGMM-HMM) outperform the compared model (ICA-HMM) for several performance metrics. The accuracy (93.80%, 84.32%, 83.36%) , the sensitivity (94.95%, 87.57%, 86.28%), and the F beta (93.74%, 83.94%, 83.33%) , for the IVABMGGMM-HMM, the IVA-HMM and the ICABMGGMM-HMM, respectively, are

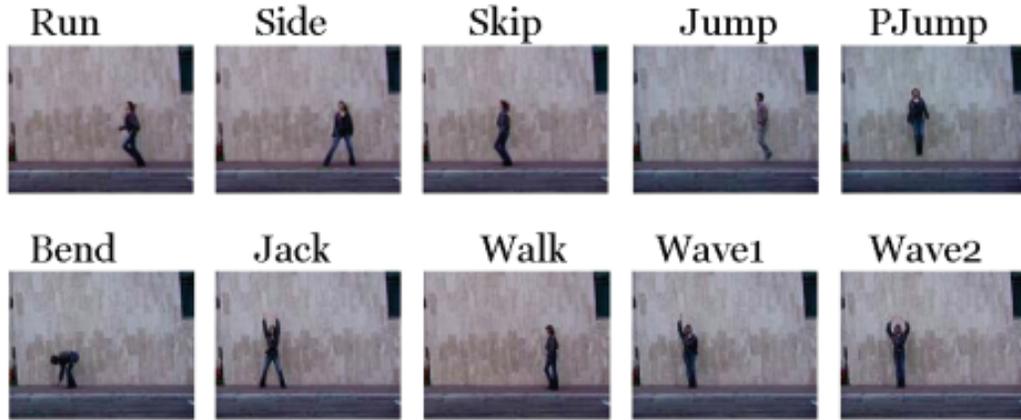


Figure 3.4: A sample frame for each action in the Weizmann action dataset: (top row) running, side jumping, skip jumping, jumping, and jumping in place; (bottom row) bending, jumping jack, walking, one hand waving, and two hands waving.

better than those for the ICA-HMM. In terms of recognizing precision, we clearly see that the proposed models (96.13%, 88.44%, 86.27%) are significantly better than the base model (76.38%).

For the low rate metrics, FPR (0.71%, 1.79%, 1.90%) and FNR (5.04%, 12.42%, 13.71 %) for

Table 3.1: Performance on Weizmann data based on different metrics.

Performance metrics / Model	IVABMGGMM-HMM	IVA-HMM	ICABMGGMM-HMM	ICA-HMM
Accuracy	<b>93.80</b>	84.32	83.36	75.58
Sensitivity	<b>94.95</b>	87.57	86.28	73.35
Specificity	<b>99.28</b>	98.20	98.09	76.13
Precision	<b>96.13</b>	88.44	86.27	76.38
F1 score	<b>95.54</b>	88.00	86.28	79.69
F beta	<b>93.74</b>	83.94	83.33	74.66
NPV	<b>99.30</b>	98.24	98.09	76.14
Jaccard	<b>89.39</b>	75.70	73.52	51.31
MCC	<b>93.14</b>	82.59	81.19	71.76
FNR	<b>5.04</b>	12.42	13.71	26.64
FPR	<b>0.71</b>	1.79	1.90	3.86
FDR	<b>3.86</b>	11.55	13.72	33.61

the proposed models are lower than the FPR (3.86%) and FNR (26.64%) of the base model. Most performance metrics show that the IVABMGGMM-HMM is the best model in our system. That reflects the capability of this model to classify the different actions in Weizmann data. This success is explained by the fact that this model considers the full covariance matrix and relaxes the limitations of the ICA models. The confusion matrix for this model is shown in Fig. 3.5, along with the classification accuracy for each class. It is clear from the confusion matrix that the main source of

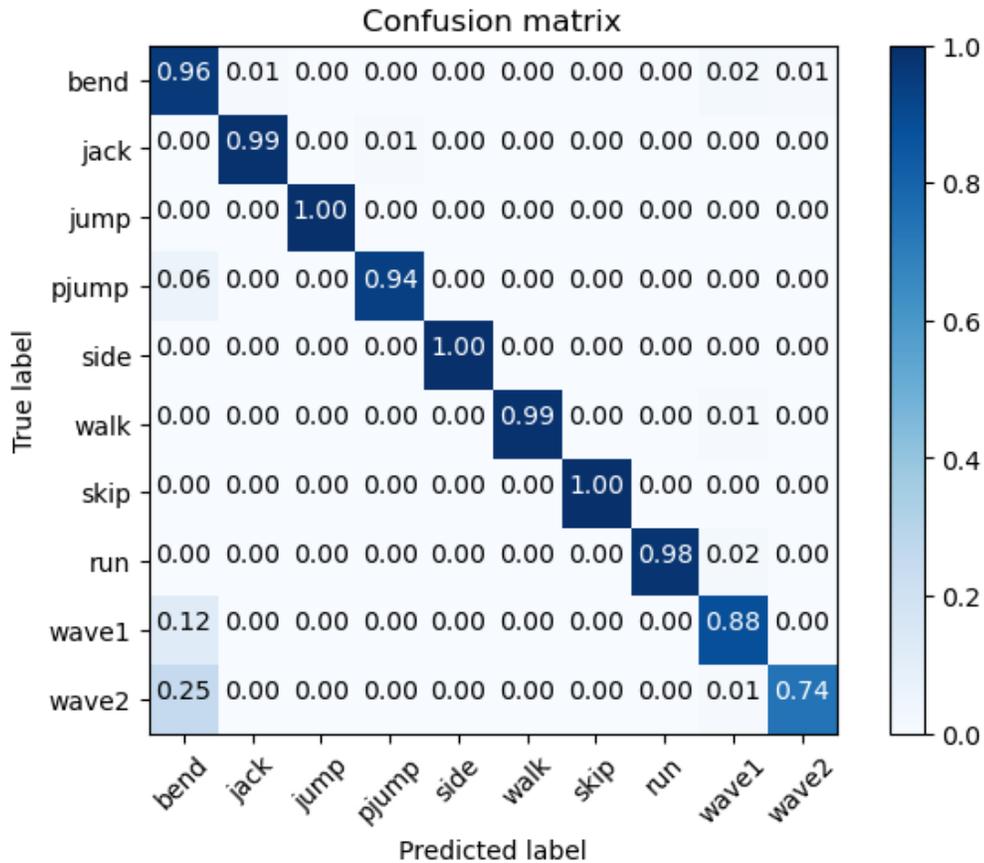


Figure 3.5: Confusion matrix for the IVABMGGMM-HMM model for Weizmann data.

misclassification happens in "wave1" and "wave2," in which 12% and 25% of the test videos were wrongly classified and received prediction labels of "bend" instead. The proposed models have thus successfully demonstrated their effectiveness in clustering video data into specific activities, which opens the door to be used in real-world scenarios for recognition and AI-assisted surveillance in an unsupervised manner.

### **KTH data**

In this application, the KTH human action dataset [106] is used to evaluate the performance of our system. There are 600 action videos (2391 video sequences) in this dataset that describe various human activities such as running, hand-waving, boxing, walking, jogging, and hand-clapping. The dataset contains videos of 25 people where each subject repeats the six actions in four different scenarios: outdoors, outdoors with zooming in and out, outdoors with different clothes, and

indoors. Fig. 3.6 shows some illustrations for this dataset for different actions in various scenarios. First step, we preprocessed the data by drawing a fixed-size box around the actors in each video to improve classification accuracy while reducing background noise. Then, the dense optical flow features (optical flow vector) are estimated by the optical flow model between two consecutive frames  $t$  and  $t + 1$ . In our case, the intuition is that this will help to describe the motions of the limbs in the videos. These features corresponding to each human action are illustrated in Fig. 3.7. After the feature extraction step, we trained and tested the ICABMGMM-HMM, IVA-HMM, and IVABMGMM-HMM models, as well as the ICA-HMM over the dense optical flow extracted features. In the training process, for each action, we set 10 subject’s videos for the testing and others videos for the training. In total, we have 360 videos for training and 140 videos for testing. Here, we set the number of states  $K = 6$ . The performance metrics for our models and

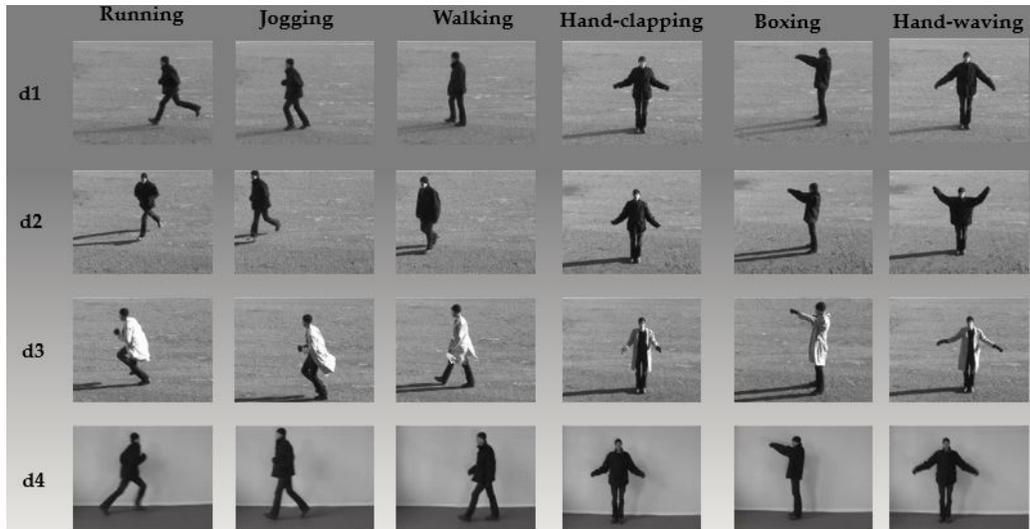


Figure 3.6: Examples of frames from the KTH dataset showing various human actions in different scenarios. d1: outdoors, d2: outdoors with scale variations, d3: outdoors with different clothes, d4: indoors.

the ICA-HMM compared model are shown in Table 3.2. In terms of the high rate metrics, accuracy (93.34%, 86.77%, 80.78%), precision (93.18%, 87.70%, 78.62%), F1 score (93.12%, 86.86%, 78.96%), and other metrics demonstrate that the proposed models outperform the compared model with accuracy (78.5%), precision (76.26%), and F1 score (79.34%). In terms of low rate measures,

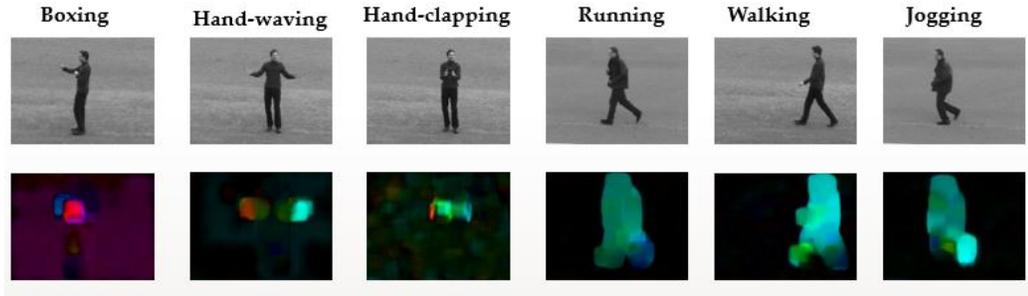


Figure 3.7: Example for the dense optical flow features for boxing, handclapping, and jogging. The first row represents the human action, and the second row represents the extracted features corresponding to each action.

FPR (1.33%, 2.71%, 4.41%) and other performance metrics show the (ICABMGGMM-HMM, IVA-HMM, IVABMGGMM-HMM) are more effective than the ICA-HMM (FPR = 8.34%). Also, we see clearly that the IVABMGGMM-HMM obtains the highest accuracy (93.34%) among other models. In other words, better results for the IVABMGGMM-HMM are attained by taking into account the bounded region and the full covariance matrix. The confusion matrix for the best performance is shown in Fig. 3.8, along with the classification accuracy for each class. As we can notice from the confusion matrix that the main source of misclassification was between the “jogging”, “running” and “walking”, which are highly similar actions. These promising results demonstrate our models’ ability in human action recognition and related applications such as object recognition.

Table 3.2: Performance on KTH data based on different metrics.

Performance metrics / Model	IVABMGGMM-HMM	IVA-HMM	ICABMGGMM-HMM	ICA-HMM
Accuracy	<b>93.34</b>	86.77	80.78	78.54
Sensitivity	<b>93.06</b>	86.04	77.22	77.92
Specificity	<b>98.66</b>	97.28	95.58	83.53
Precision	<b>93.18</b>	87.70	78.62	76.26
F1 score	<b>93.12</b>	86.86	78.96	79.34
NPV	<b>98.66</b>	97.36	95.75	86.46
MCC	<b>91.96</b>	84.03	74.24	75.66
FPR	<b>1.33</b>	2.71	4.41	8.34
FNR	<b>6.81</b>	13.95	20.95	22.25
FDR	<b>7.55</b>	12.29	19.21	22.64

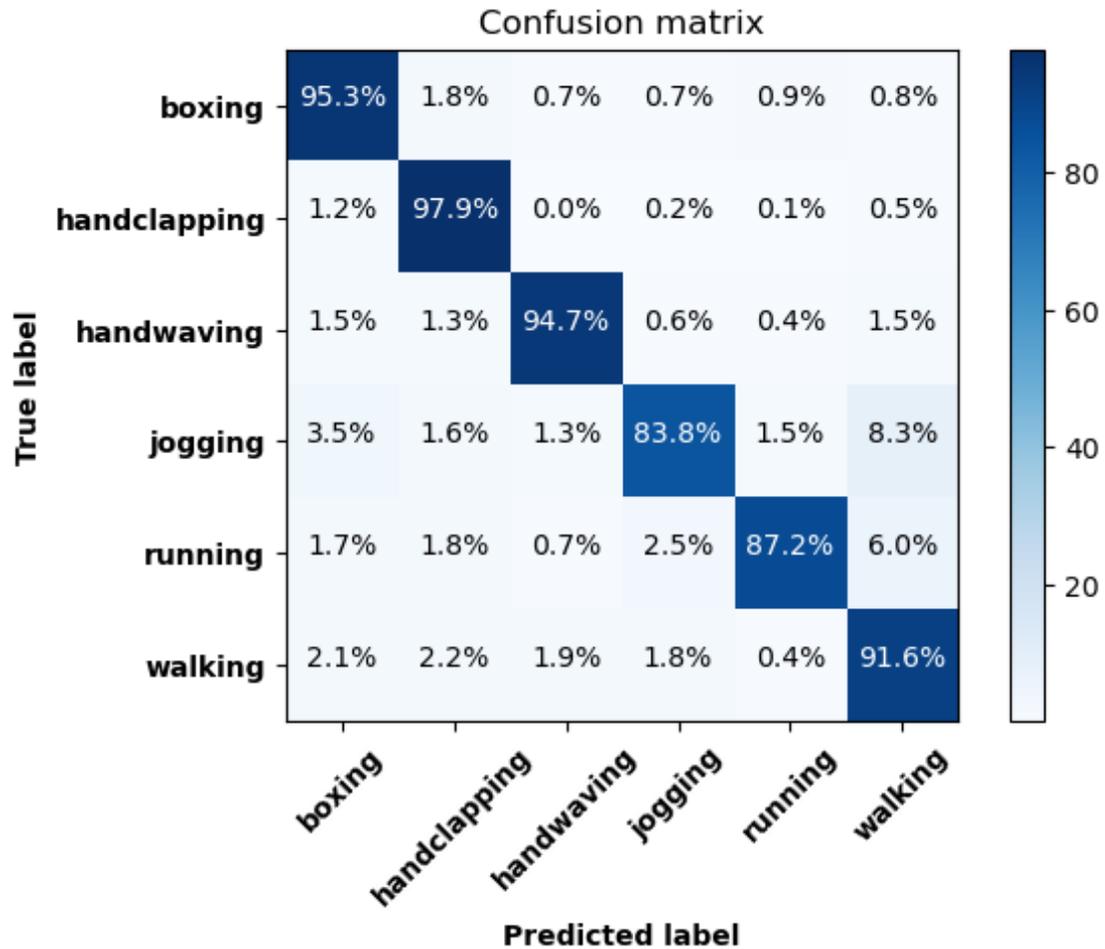


Figure 3.8: Confusion matrix for the IVABMGMM-HMM model for KTH data.

### 3.3.2 Speech recognition

#### TIMIT data

Phone-based speech recognition has been the subject of intensive research because it is free of vocabulary limitations. The quality of the phone recognizer determines the performance of large vocabulary automatic speech recognition (LVASR) systems. That is why research teams are constantly working to improve phone recognizers' performance. Phone recognition is, in fact, a recurring issue in the speech recognition community. Phone recognition is used in a variety of applications. In addition to standard LVASR systems [131, 1], it can be found in applications related to language recognition [132, 133], keyword detection [133], music identification and translation [134, 135], and applications of speaker identification [136]. Phone recognition in TIMIT has been the subject

of extensive research for more than two decades, and its performance has steadily improved over time. There are numerous methods available, but in terms of evaluation, they focus on three areas: phone segmentation, phone categorization, and phone recognition.

The DARPA TIMIT acoustic-phonetic continuous speech corpus (TIMIT - Texas instruments (TI) and Massachusetts institute of technology (MIT)), described in [137], contains phonetically balanced prompted english speech recordings. It was recorded at 16 kHz with a Sennheiser close-talking microphone and a sample resolution of 16 bits. TIMIT contains 6300 sentences (5.4 hours) spoken by 630 speakers from eight major dialect regions of the United States. At the phone level, all sentences were manually segmented. The training set contains 4620 utterances, but usually only SI and SX sentences are used, resulting in 3696 sentences from 462 speakers. The test set contains 1344 utterances from 168 speakers. The data was divided into 7 classes, as shown in Table 3.3. These classes are namely Stops, Affricates, Fricatives, Nasals, Vowels, Semi-vowels and Silences. For the feature extraction step, the 13th-order Mel-frequency cepstral coefficients (MFCCs) were obtained; their delta and acceleration features were combined to generate 39-dimensional feature vectors. Next, we trained and tested the ICABMGGMM-HMM, the IVA-HMM, and the IVABMGGMM-HMM, as well as the ICA-HMM over the obtained features. In this experiment, we set the number of states  $K = 7$ .

Table 3.4 provides the performance metrics for our proposed models (ICABMGGMM-HMM, IVA-HMM, IVABMGGMM-HMM) compared to the ICA-HMM model. For the high rate measures, our proposed models perform well in terms of accuracy (72.63%, 69.82%, 67.05%), sensitivity (72.60%, 69.81%, 67.12%), and specificity (94.86%, 94.97%, 92.97%) when compared to the ICA-HMM, which has accuracy (65.34%), sensitivity (65.32%), and specificity (91.23%). For the low rate metrics, FPR (5.02%, 5.13%, 7.02%) and Log loss (3.21%, 6.98%, 8.57%) show the lower error rates for our proposed models compared to the ICA-HMM with FPR (7.54%) and Log loss (12.53%). The results clearly show that the IVABMGGMM-HMM has demonstrated its success in clustering speech features, which can be used to improve the speech recognition systems. This success (7.29% accuracy improvement for the IVABMGGMM-HMM compared to ICA-HMM) is ascribed to the effect of the full covariance matrix and the viability of the IVABMGGMM to relax the limitations of ICA models.

Table 3.3: Classes of phones in TIMIT data used in this experiment.

Phone class	# TIMIT labels	TIMIT labels
Stops	14	b d g p t k dx q dcl gcl pcl tcl kcl dcl bcl
Affricates	2	jh ch
Fricatives	8	s sh z zh f th v dh
Nasals	7	m n ng em en eng nx
Semi-vowels	7	l r w y hh hv el
Vowels	20	iy ih eh ey ae aa aw ay ah ao oy ow uh uw ux er ax ix axr ax-h
Silences	3	pau epi h#

Table 3.4: Performance on TIMIT data based on different metrics.

Performance metrics / Model	IVABMGMM-HMM	IVA-HMM	ICABMGMM-HMM	ICA-HMM
Accuracy	<b>72.63</b>	69.82	67.05	65.34
Sensitivity	<b>72.60</b>	69.81	67.12	65.32
Specificity	<b>94.86</b>	94.97	92.97	91.23
Precision	<b>63.78</b>	63.01	60.53	57.42
F1 score	<b>67.03</b>	66.24	61.86	59.81
NPV	<b>94.46</b>	93.11	94.55	92.33
MCC	<b>63.49</b>	63.25	63.04	52.44
FPR	<b>5.02</b>	5.13	7.02	7.54
Log loss	<b>3.21</b>	6.98	8.57	12.53

### TIMIT data with 5 classes

To investigate the behaviour of the classes, we extended the TIMIT experiment following the suggested class organization in [1]. They have proposed combining the classes of Vowels and Semi-vowels to be one class (Vowels). They also merged the Affricates and Fricatives classes to be in one class (Fricatives) as shown in Table 3.5. Here, the number of states is equal to  $K = 5$ . Table 3.6 shows that when dealing with TIMIT classes using the suggested criteria (5 classes), our models outperformed the models in the previously suggested method in Table 3.3 (7 classes). It is worth mentioning that the three proposed models outperform the baseline model (ICA-HMM) for most of the performance metrics as shown in Table 3.6. The performance outcomes of our models for TIMIT with 5 and 7 classes are visualized in Fig. 3.9. The enhancement for the ICABMGMM-HMM and IVA-HMM models in TIMIT with five classes is slightly better than those in the same data with seven classes. In the five classes case, the IVABMGMM-HMM model outperforms the same model in TIMIT with 7 classes by 7.27%.

From this experiment, we notice that the preparation of the data and the behaviour of the classes

play an important role in the clustering performance. This success for our models is explained by the fact that these models consider the full covariance matrix and tackle the ICA limitations. Also, that reflects the ability of our models to detect features in bounded Gaussian regions. We also see from the TIMIT experiment that the IVABMGGMM-HMM has effectively demonstrated its viability in clustering speech features to improve the representation of speech data.

Table 3.5: 5 classes of phones in TIMIT data used in this experiment [1].

Phone class	# TIMIT labels	TIMIT labels
Vowels	25	aa, ae, ah, ao, ax, ax-h, axr, ay, aw, eh, el, er, ey, ih, ix, iy, l, ow, oy, r, uh, uw, ux, w, y
Stops	8	p, t, k, b, d, g, jh, ch
Fricatives	10	s, sh, z, zh, f, th, v, dh, hh, hv
Nasals	7	m, em, n, nx, ng, eng, en
Silences	11	h#, epi, pau, bcl, dcl, gcl, pcl, tcl, kcl, q, dx

Table 3.6: Performance on TIMIT data with 5 classes based on different metrics.

Performance metrics / Model	IVABMGGMM-HMM	IVA-HMM	ICABMGGMM-HMM	ICA-HMM
Accuracy	<b>79.90</b>	68.28	67.52	63.58
Sensitivity	<b>68.70</b>	64.71	62.52	60.23
Specificity	<b>94.96</b>	92.95	91.74	90.40
Precision	<b>73.18</b>	66.92	66.94	65.15
F1 score	<b>70.87</b>	65.80	63.70	60.66
NPV	<b>95.45</b>	92.18	92.01	90.31
MCC	<b>74.1</b>	64.36	62.19	60.58
FPR	<b>0.05</b>	0.08	0.08	0.09
Log loss	<b>2.81</b>	3.07	5.05	7.84

### English spoken digit data

In this subsection, we used the spoken-digit dataset to validate the robustness of our proposed models. This data contains recordings of spoken digits ranging from 0 to 9 at 8 kHz. The recordings have been trimmed such that there is almost no silence at the beginning and end. It consists of 3000 utterances from 6 English speakers<sup>1</sup>. The MFCCs were obtained in the feature extraction step. The ICABMGGMM-HMM, the IVA-HMM, and the IVABMGGMM-HMM, as well as the ICA-HMM, are then trained and tested over the retrieved MFCCs features. Here, we set the number of states  $K = 10$ .

Table 3.7 compares the performance metrics of our proposed models (ICABMGGMM-HMM,

<sup>1</sup><https://github.com/Jakobovski/free-spoken-digit-dataset>

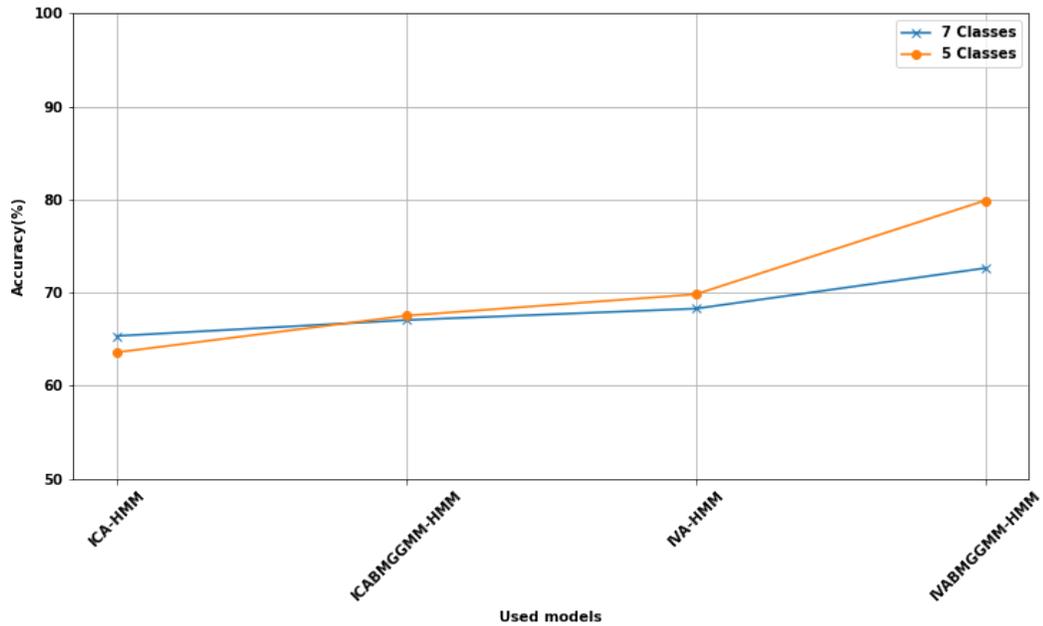


Figure 3.9: The accuracy differences between the proposed models on TIMIT data with 7 and 5 classes.

IVA-HMM, and IVABMGGMM-HMM) to the ICA-HMM. In terms of high rate, our models outperformed the ICA-HMM model. Accuracy, precision, and F1 score are higher than those for the ICA-HMM. In terms of error rates, such as FPR, FDR, and FNR are lower than the comparable metrics for the ICA-HMM. The enhancement of our proposed models is 11.74% as shown in terms of accuracy of IVBMGGMM-HMM (96.97%) compared to ICA-HMM (85.23%). The classification accuracy for each class and the confusion matrix for the best performance is shown in Fig. 3.10. We see clearly from the confusion matrix that the main source of misclassification happens in the “6” digit, in which 29% of the test utterances were wrongly classified, and received prediction labels of “3” or “8” instead. From this experiment, we can notice the robustness of our models compared to the TIMIT experiments. This enhanced performance can be ascribed to the nature of the data and the behaviour of the classes. Furthermore, the IVABMGGMM-HMM model includes the full covariance matrix and can tackle the ICA limitations. These three speech experiments infer that our proposed models can be used in speech recognition systems and their applications.

Table 3.7: Performance on spoken digit data based on different metrics.

Performance metrics / Model	IVABMGMM-HMM	IVA-HMM	ICABMGMM-HMM	ICA-HMM
Accuracy	<b>96.97</b>	95.29	92.97	85.23
Sensitivity	<b>96.01</b>	95.37	93.07	85.17
Specificity	<b>99.55</b>	99.47	99.22	87.02
Precision	<b>96.35</b>	95.98	93.66	83.03
F1 score	<b>96.18</b>	95.67	93.37	82.09
NPV	<b>99.56</b>	99.48	99.23	89.06
MCC	<b>95.60</b>	94.85	92.31	80.56
Jaccard	<b>92.26</b>	91.02	86.87	83.88
FPR	<b>0.04</b>	0.05	0.07	0.97
FNR	<b>3.98</b>	4.62	6.92	8.82
FDR	<b>3.64</b>	4.01	6.33	6.96

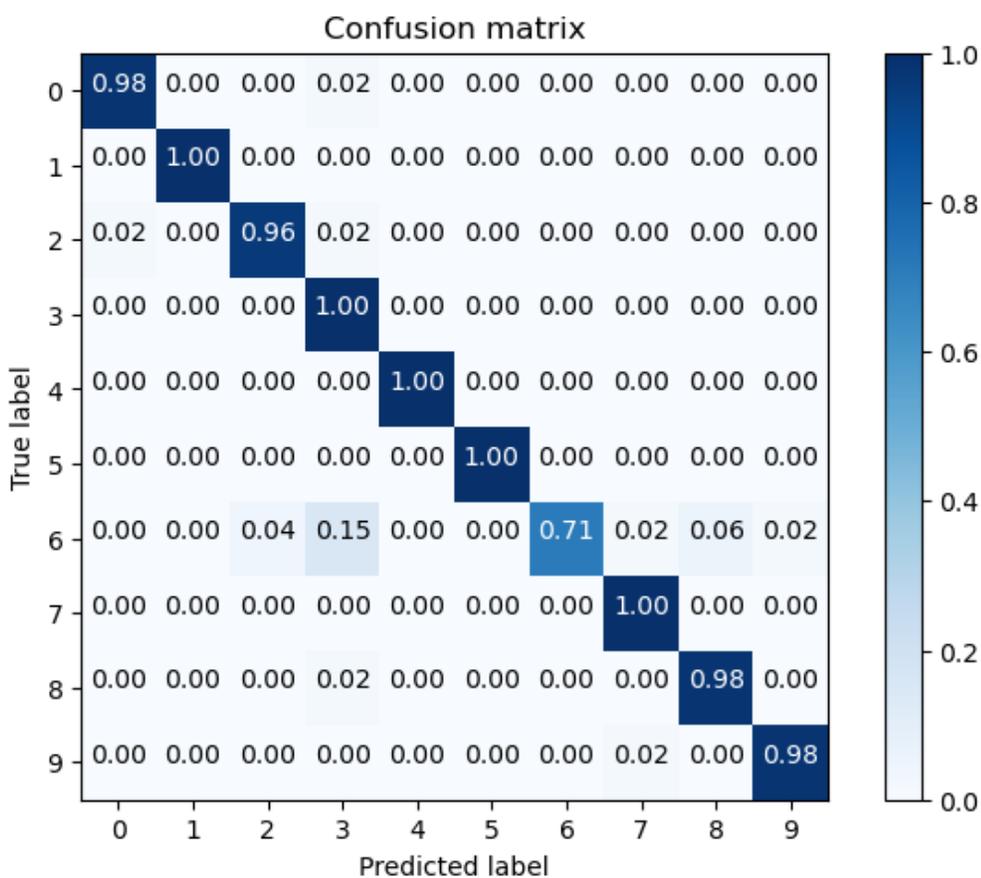


Figure 3.10: Confusion matrix for the IVBMGGMM-HMM for spoken digit data.

### 3.3.3 Energy disaggregation

Energy disaggregation of appliances using non-intrusive load monitoring (NILM) is a set of information processing methods used to extract appliance-level information from a meter's total or

aggregate load. The world is currently focusing on reducing our use of electricity while also being concerned about other resources such as natural gas, water, and grey water. To address these issues, numerous technologies are being developed. One of the most effective cost-cutting techniques is NILM. NILM algorithms can determine which appliances are running in a home-based on power line analysis. Solutions to NILM inform about consumption issues or present conservation solutions to homeowners that would allow them to understand their home’s consumption and take action to reduce electricity consumption. For this application, we used the REDD dataset [117] to validate our proposed models. It includes actual power measurements for six houses. Two aggregate signals of phases A and B with sampling frequencies of 1 Hz and sub-metered power signals of individual loads with sampling frequencies of 1/3 Hz are recorded in each house. In this work, data from houses 1 – 5 is used for training while data from house 6 is reserved for testing. We picked four appliances: Refrigerator (Ref), Washer-dryer (WD), Dishwasher (DW), and Kitchen outlets (KO) over the consumption period of one year. Table 3.8 presents the training houses, testing houses, and the setting for each appliance to be active. We trained our models, including ICABMGGMM-HMM, IVA-HMM, and IVABMGGMM-HMM and compare them with ICA-HMM using each appliance individually. We repeated these steps for the four different appliances. We validated the performance of our proposed models using several performance metrics. For this experiment, we chose the number of states to be  $K = 2$ . This real-world test was used to show the ability of our models to handle completely unknown appliances, duty cycles, and consumption—see, for example, Fig. 3.11. Table 3.9 compares the performance of the suggested models over the Refrigerator,

Table 3.8: Appliances and houses used.

Appliance	Training houses	Testing houses	On state (Watts)	
Refrigerator	1,2,3,5	6	>80	No Refrigerator in house 4
Dishwasher	1,2,3,5	6	> 25	No Dishwasher in house 4
Washer-dryer	1,2,3,4,5	6	>25	All houses are included
Kitchen outlets	1,2,3,4,5	6	>25	All houses are included

Dishwasher, Washer-dryer and Kitchen outlets. In this Table, when comparing state estimation and consumption estimation performance of the proposed models and the base model across all results, we observe that our proposed models outperformed the comparable model. For the Refrigerator

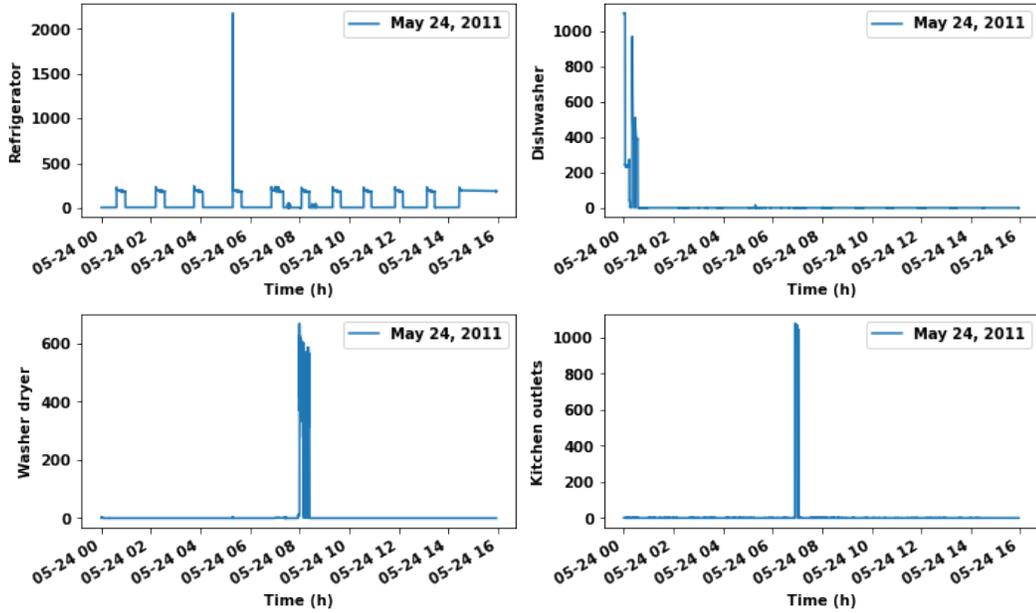


Figure 3.11: Typical appliance signatures for Ref, DW, WD and KO.

Table 3.9: Performance on REDD data based on different metrics for all appliances.

Performance metrics	IVABMGGMM-HMM				IVA-HMM				ICABMGGMM-HMM				ICA-HMM			
	Ref	DW	WD	KO	Ref	DW	WD	KO	Ref	DW	WD	KO	Ref	DW	WD	KO
Accuracy	93.21	94.18	96.01	97.98	88.34	88.28	88.78	87.12	66.25	71.28	67.18	81.03	62.35	62.43	61.25	70.63
Sensitivity	93.98	96.40	97.98	98.03	90.69	93.14	83.36	88.35	66.25	71.28	59.22	87.32	60.43	60.32	55.32	71.73
Specificity	93.96	96.36	97.92	98.31	90.62	93.24	83.46	88.45	66.35	71.38	59.32	87.42	59.64	65.34	54.53	71.43
Precision	73.37	73.84	83.25	92.31	66.51	65.49	78.54	89.74	69.64	80.08	72.74	56.61	64.67	69.44	69.57	52.62
F1 score	82.41	83.63	74.63	88.40	76.74	76.91	72.73	69.87	67.91	75.43	63.68	67.00	61.32	70.53	58.43	58.92
Jaccard	87.29	89.00	92.33	96.05	79.11	79.02	79.83	77.14	59.54	55.38	60.58	68.11	52.64	50.31	52.57	60.75
FPR	00.06	00.03	00.02	00.01	00.09	00.06	00.16	00.12	00.33	00.28	00.40	00.11	00.52	00.42	00.92	01.10
Log loss	02.34	02.00	01.37	00.69	04.02	04.04	11.33	02.93	11.65	09.91	03.87	06.55	14.43	11.62	06.42	10.42

appliance, the proposed models (IVABMGGMM-HMM, IVA-HMM, ICABMGGMM-HMM) performed better than the comparable model for all selected measures. For the Dishwasher; accuracy, sensitivity, and Jaccard for our models are higher than those for the ICA-HMM. In terms of the lower rate metrics, FPR and Log loss for our proposed models are lower than FPR and Log loss for the base model. Similarly, we see that our models outperformed the comparable model for all selected metrics in the case of Washer-dryer and Kitchen outlets. When comparing state estimation and consumption estimation performance of the proposed models, we notice that the IVABMGGMM-HMM is the best performing model compared to the performance of other models. We present the performance outcomes for our models in terms of accuracy in Fig. 3.12. We see a clear performance improvement for the IVABMGGMM-HMM (blue colour) compared to the base model

(green colour). In this experiment, we used the ROC curves to describe the AUC for the four models. Fig. 3.13 demonstrates that the IVABMGGMM-HMM is the best model compared to other models. We noticed that the AUC for the IVABMGGMM-HMM (0.97) is the largest area compared to the ICA-HMM area (0.74). To summarize, the AUC for the proposed models is greater than the AUC for the compared model. This experiment demonstrates the ability of our presented models to categorize the active state appliance correctly. This success reflects the ability of our models to detect features in bounded Gaussian regions. These promising results can be used as guidance to researchers in the field of energy consumption applications.

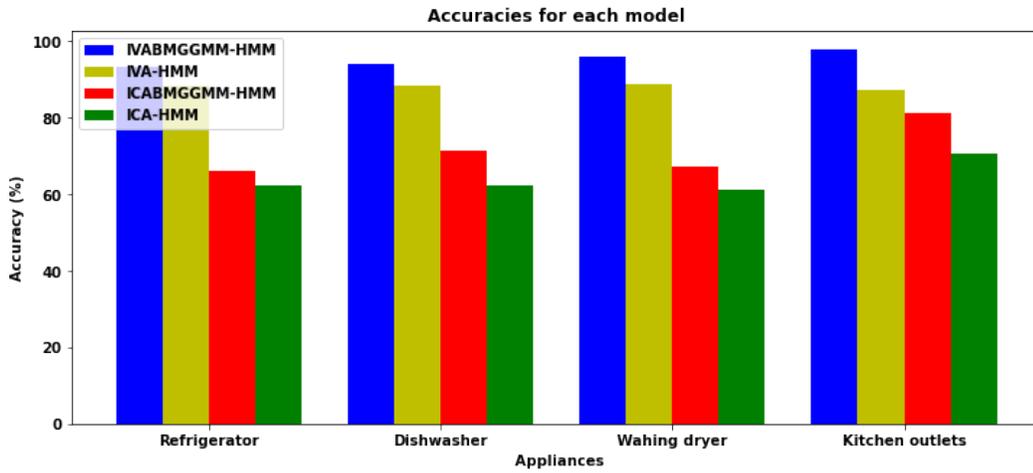


Figure 3.12: The accuracy differences for the proposed models over the four appliances.

### 3.3.4 Epileptic Seizure clustering using EEG signals

The proposed methods are validated using the EEG dataset described in [138]. The same 128-channel amplifier system is used to record all EEG signals, with an average common reference. The dataset is divided into five subsets (labelled  $Z$ ,  $O$ ,  $N$ ,  $F$ , and  $S$ ), each with 100 single-channel EEG segments lasting 23.6 seconds. Subsets  $Z$  and  $O$  contain EEG data from five healthy volunteers with their eyes open and closed. The  $F$  and  $N$  subsets are made up of EEG data from five Seizure patients collected during the Seizure-free period;  $F$  is collected from the epileptic zone, while  $N$  is collected from the hippocampal formation in the opposite hemisphere of the brain. Subset  $S$  records EEG data from patients during Seizure activity.

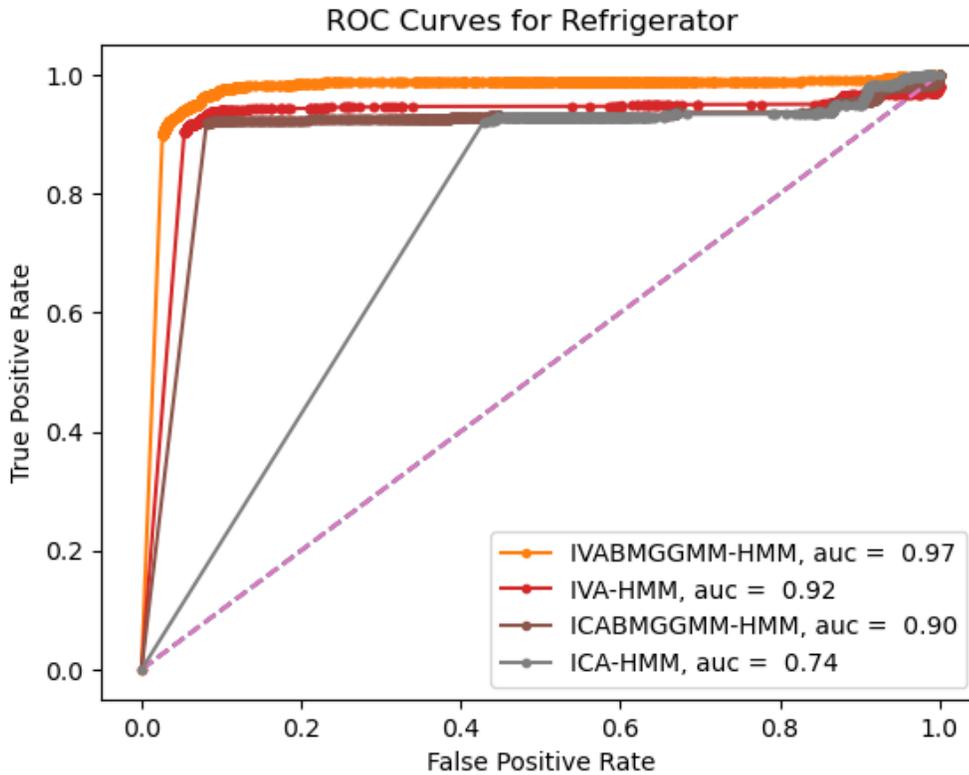


Figure 3.13: ROC curves for refrigerator using the four models IVABMGGMM-HMM, IVA-HMM, ICABMGGMM-HMM and ICA-HMM.

The proposed EEG-based epileptic Seizure clustering approach comprises many primary processes, including pre-processing, time-frequency analysis, feature extraction, and clustering. For both time and frequency domain smoothing, we used a 64-point Hamming window. For the feature extraction step, 13 cepstral coefficients were obtained for each window in the EEG dataset from a filter-bank consisting of nine filters. The input EEG signal is first transformed into the frequency domain using the fast Fourier transform and then applied to a bank of triangular filters to compute a weighted sum of filter spectral components approximating a Mel scale [139]. The MFCCs are obtained by converting the log Mel spectrum into the time domain using the discrete cosine transform. The performance is evaluated over ten iterations, with a random selection of training and testing datasets of EEG signals at each step. In this experiment, our goal is to distinguish between the Seizure and non-Seizure EEG signals, we chose the number of states  $K = 2$ . At each iteration, we selected 80% of the EEG data as a training dataset, with an equal number of data points in each class, and the remaining EEG segments are used for testing. Then our proposed models are used to train

and test the extracted features. As shown in Table 3.10 the proposed models outperform the ICA-HMM. Sensitivity (90.22%, 89.47%, 89.47%), specificity (97.47%, 95.04%, 94.14%), precision (96.97%, 95.88%, 96.78%), and other performance metrics indicate the effectiveness of the proposed models as compared to the ICA-HMM with sensitivity of 87.64%, specificity of 90.69%, and precision of 90.33%. In terms of accuracy, the IVABMGGMM-HMM (95.54%) has improved by 7.11% when compared to the base model (ICA-HMM) (88.43%). The improved IVABMGGMM-HMM performance can be ascribed to the model’s ability to detect features in bounded Gaussian regions. Furthermore, as compared to ICA models, the IVA models perform better with multivariate data such as EEG.

Table 3.10: Performance on EEG Seizure data based on different metrics.

Performance metrics / Model	<b>IVABMGGMM-HMM</b>	IVA-HMM	ICABMGGMM-HMM	ICA-HMM
Accuracy	<b>95.54</b>	94.58	93.54	88.43
Sensitivity	<b>90.22</b>	89.47	89.47	87.64
Specificity	<b>97.47</b>	95.04	94.14	90.69
Precision	<b>96.97</b>	95.88	96.78	90.33
F1 score	<b>95.00</b>	93.93	92.93	91.42
NPV	<b>95.33</b>	91.22	91.05	81.55
FNR	<b>9.44</b>	10.52	10.53	14.32
FPR	<b>1.95</b>	2.85	9.85	11.63
FDR	<b>1.11</b>	3.21	9.52	9.67

### 3.4 Conclusion

We investigated the integration of ICA, IVA, ICA-BMGGMM and IVA-BMGGMM into HMM to improve their capability for modeling data. Initially, we tested the combination of ICA and HMM as the base model. Then, we proposed the ICA-BMGGMM to be integrated into the HMM framework to improve the ICA-HMM modeling capability. Next, we tackled the ICA limitations for multivariate data by investigating the integration of IVA into HMM. Finally, we explored the combination of IVA-BMGGMM and HMM to improve the modeling ability of IVA-HMM. We validated our proposed models using different applications. We used the KTH and Weizmann datasets to train our models to recognise various human actions. For the speech recognition, we tested our models’ ability to detect different types of speech using TIMIT and english spoken digit datasets. We extended our validation process by using our proposed models to distinguish the active state of

the selected appliances for energy disaggregation. Furthermore, we used the EEG dataset to train our models to identify Seizure patients from healthy subjects. We improved the modeling capability for ICA and IVA to model Gaussian and non-Gaussian data. The presented experiments demonstrated the capability of our models to distinguish between the classes when compared to the base model. Specifically, the IVABMGMM-HMM revealed its ability to detect the classes using the presented experiments as compared to other proposed models.

## Chapter 4

# **Adaptive constrained IVAMGGMM: Application to neurological disorders detection**

There is an increasing demand for adaptable approaches to examine extensive fMRI data, aiming to capture overall patterns that represent a population while retaining the unique characteristics of individuals. The use of ICA is on the rise as a valuable tool for examining the concealed spatio-temporal patterns present in brain imaging data. ICA has limitations in separating the correlated sources in multivariate data such as fMRI. For that, we propose an ICA-based multivariate generalized Gaussian mixture model combined with the constrained ICA to form the cICA-MGGMM. This model relaxes the independence assumption of ICA. Also, we propose the adaptive constrained ICA-MGGMM (acICA-MGGMM) to adaptively control the association between reference signals and estimated sources. IVA is a data-driven technique that concurrently calculates global spatial and temporal patterns from fMRI data involving multiple subjects, effectively maintaining individual variability. Nonetheless, as our findings indicate, the performance of IVA is compromised when the quantity of datasets and components increases, particularly in cases of weak correlations among

these components across the datasets. In this Chapter, we investigate this problem and propose an efficient approach to tackle it by integrating reference signals of the estimation patterns into the formulation, offering guidance in high-dimensional situations. For that, we propose the cIVA-MGGMM to relax the ICA limitations for multivariate data. cIVA-MGGMM provides an effective framework for including references, but its effectiveness relies on user-defined constraint parameters. These parameters enforces a predetermined level of association between the reference signals and estimation patterns. To tackle these limitations, we introduce the adaptive cIVA-MGGMM (acIVA-MGGMM) to adapt and separate the activated brain sources. This model employs a full covariance matrix, which takes into account the feature correlation. ICA and IVA models face restrictions when used with multivariate high-dimensional data, such as resting state fMRI data. For that, our four constrained methods incorporate prior information about the sources into the ICA and IVA models to address the limitations of ICA and IVA in high-dimensional data. We validate our proposed models using simulation, Alzheimer's, Schizophrenia, EEG, and Attention-deficit/hyperactivity disorder datasets. Based on the presented performance metrics, our proposed models outperform the base models on the conducted experiments.

## 4.1 Introduction

As the amount of data being generated keeps growing, there is an increasing demand for adaptable approaches that can effectively extract the overall trends while also maintaining subject-specific information from large-scale datasets. Joint blind source separation (JBSS) is a method used in signal processing and machine learning to separate multiple signals that have been mixed together into a single signal without any prior knowledge about the mixing process or the original sources. The goal of JBSS is to recover the original sources from the observed mixture. Two popular methods for BSS are ICA [140] and IVA [42].

ICA has proven to be effective in various applications of BSS, such as speech separation [141], image processing, and biomedical signal analysis [142]. ICA's assumption of source independence is one of its limitations. However, a method for addressing this limitation by employing mixture models is proposed in [143]. In a mixture model, the observed data is categorized into distinct and

mutually exclusive classes [144, 145, 146, 147, 148]. This model is known as the ICA mixture model.

Constrained ICA (cICA) has emerged as a highly effective semi-blind technique. The concept of cICA, as introduced in [149], involves incorporating prior information into the ICA cost function through the utilization of equality and inequality constraints within a Lagrangian framework. By doing so, cICA algorithms have demonstrated enhanced performance in source estimation, leveraging the advantages of model-driven methodologies, such as noise robustness, while preserving the adaptable data-driven nature of ICA [150]. Several studies on c-ICA have demonstrated that incorporating reliable prior information enhances its overall performance in source separation for resting-state fMRI data [151, 152, 153]. In this work, we propose cICA-MGGMM by combining the ICA multivariate generalized Gaussian mixture model (ICA-MMGGMM) and the constrained model. The objective of cICA-MGGMM is to offer a structured and adaptable approach for integrating additional assumptions and prior information, when accessible, into the cost function. By doing so, it transforms an initially ill-posed ICA problem into a better-posed one, enabling a wider range of applications. The cICA-MGGMM is based on user-defined parameters but selecting the right constraint parameters can be difficult because prior information accuracy is frequently unknown. To that end, we propose adaptive cICA-MGGMM, which allows us to control the association between estimated and reference sources adaptively. The adaptive cICA-MGGMM method guarantees that the prior information is employed to guide the solution without imposing inaccurate constraints.

IVA is a more recent BSS method that extends ICA by incorporating mixing matrix information into the estimation process. Unlike ICA, which assumes an unknown mixing matrix, IVA assumes a known mixing matrix up to scaling and permutation. IVA attempts to estimate the original sources by determining a linear transformation that maximizes the non-Gaussianity of the transformed components. IVA has been demonstrated to be effective in situations where the sources are highly dependent, and it has applications in blind speech separation [154] and blind source separation of fMRI data [155].

IVA was initially designed to solve the convoluted ICA problem in the frequency domain by employing multiple frequency bins [156]. As a result, the IVA-Laplacian (IVA-L) algorithm [49] was created, which considers only HOS and assumes a Laplacian distribution for the underlying

source component vectors. IVA-Gaussian (IVA-G) [49] takes advantage of linear dependencies but ignores HOS. Finally, multivariate generalized Gaussian IVA is a more general IVA that takes into account both second and higher-order statistics [42, 157, 158].

Constrained IVA (cIVA), a method highlighted in [159], optimizes the IVA cost function by integrating prior knowledge regarding the sources or columns of the mixing matrix. It relaxes the assumption of independence by permitting a favorable combination of data-driven and model-driven techniques while also ensuring model matching through the utilization of precise constraints. Several studies have provided evidence of cIVA's effectiveness in discriminating between individuals with mental illnesses and those who are in a healthy state [159, 101, 160]. In this Chapter, we propose a general mathematical formulation called cIVA-MGGMM, which provides a general framework for cIVA by utilizing a multivariate generalized Gaussian mixture (MGGMM). This formulation enables the inclusion of diverse types of prior information concerning the sources or mixing vectors. The new cIVA-MGGMM framework combines the flexibility of data-driven methods with the robustness to noise and other artifacts of model-based methods, as well as exploiting dependence across datasets, making it an effective source separation tool.

The cIVA-MGGMM, on the other hand, uses user-defined constraint parameters to control the degree of correspondence between the reference signal and the estimated component. In real-world scenarios, choosing the right constraint parameters can be difficult because the accuracy of the prior information is frequently unknown. This task becomes even more difficult when prior knowledge about multiple signals is used. To that end, we introduce adaptive cIVA-MGGMM (acIVA-MGGMM) to enable adaptive control over the association between estimation patterns and reference signals. This method makes it easier to incorporate multiple reference signals into the IVA framework. Rather than imposing inaccurate constraints, the acIVA-MGGMM technique uses prior knowledge to guide the solution effectively.

To validate the effectiveness of our four models, we employ cICA-MGGMM, acICA-MGGMM, cIVA-MGGMM, and acIVA-MGGMM on simulation data, four real-world applications for diagnosing mental health conditions, Attention-deficit/hyperactivity disorder (ADHD), and Seizure prediction. The global concern over mental illnesses, such as Alzheimer's and Schizophrenia, has grown. To address these disorders effectively, it is crucial to conduct appropriate and timely assessments for

accurate differentiation between these diseases and their corresponding healthy subjects. Machine learning, a sub-field of artificial intelligence, focuses on classification, regression, and clustering problems. By leveraging data and algorithms, ML mimics human learning processes and continually enhances accuracy across various tasks [161]. With the increasing availability of extensive fMRI datasets, there is a growing need for a methodology capable of accurately estimating meaningful global functional networks while simultaneously capturing individual-specific characteristics. Achieving this objective is crucial in advancing precision medicine, as it requires thorough consideration of individual variability. cIVA-MGGMM, as an efficient approach, integrates prior knowledge regarding the sources or columns of the mixing matrix into the IVA cost function. By incorporating a balance between data-driven and model-driven approaches, it relaxes the assumption of independence and ensures a desirable level of flexibility. Furthermore, it maintains a strong model match by employing precise constraints. In contrast, cIVA-MGGMM utilizes user-defined constraint parameters to regulate the level of correspondence between the estimated component and the reference signal.

The effectiveness of cIVA-MGGMM relies on both the prior information and the user-defined constraint parameters. In cases where the prior information is accurate, it is necessary to increase the constraint parameters and conversely, decrease them when the prior information is unreliable. However, selecting appropriate constraint parameters becomes challenging in most practical applications due to the uncertainty regarding the accuracy of the prior information. This task becomes even more complex when incorporating prior information for multiple signals. For that, we employ acIVA-MGGMM on these datasets with high dimensionality, which dynamically adjusts the relationship between the reference signals and the estimated patterns. This adaptation allows for the effective integration of multiple reference signals into the IVA framework. We assess the performance of our models using a simulated dataset. We also use the Alzheimer’s disease neuroimaging initiative (ADNI) dataset to evaluate the ability of our models to differentiate between individuals with and without Alzheimer’s disease. The obtained acIVA-MGGMM components are used to detect the particular brain areas that are affected by Alzheimer’s disease. To further evaluate the effectiveness of our models, we employ them on a dataset provided by the COBRE database to identify the early stages of Schizophrenia and differentiate between individuals with Schizophrenia

and those who are healthy. We also use the resulting acIVA-MGGMM components to identify the defective brain regions in Schizophrenic patients. To enhance validation, we implemented our proposed models for predicting Seizures and ADHD. Overall, our models have demonstrated potential usefulness in diagnosing various mental health conditions.

The Chapter is organized as follows: Section 4.2 introduces the related work and motivation. Section 4.3 describes the adaptive constrained ICA models. In Section 4.4, we introduce the general framework for incorporating adaptive constrained parameters into the IVA objective function. Section 4.5 describes our experimental results followed by the conclusion in Section 4.7.

## **4.2 Related Work**

### **4.2.1 Constrained ICA models**

cICA, as introduced by Lu and Rajapakse in 2005, offers the capability to automatically extract desired components in a predefined sequence and reduce computational overhead by incorporating constraints into the traditional ICA [144]. cICA employs a specific type of prior knowledge to create constraints that can be applied to either the source matrix or the mixing matrix [149]. In terms of fMRI data analysis, cICA was employed to separate temporally independent components associated with the task, utilizing temporal constraints on the source matrix, all without the need to separate all sources [162, 163, 164]. Long (2019) [164] applied cICA to estimate temporarily independent components of interest from fMRI data, employing temporal constraints on the mixing matrix. To enhance the convergence of cICA, Wang (2011) [148] introduced learning-rate-free cICA algorithms, which were then used to separate spatially independent components from fMRI data while applying temporal constraints to the mixing matrix. Furthermore, cICA was employed to isolate the spatially independent components of interest from fMRI data by applying spatial constraints to the source matrix [165]. Rasheed [166] proposed a constrained spatiotemporal ICA, employing two successive cICA stages, one for each domain, to identify the most independent yet desired sources in both spatial and temporal domains. Prior information available in the spatial/temporal domain was input into the first cICA stage, and the output of this stage was subsequently utilized as a constraint in the second cICA stage.

## 4.2.2 Constrained IVA models

IVA is a data-driven method that extends the capabilities of ICA to handle multiple datasets [77]. IVA has demonstrated its effectiveness in capturing the variability across these datasets when compared to group ICA [167, 168]. This makes it an appealing approach for studying brain dynamics, particularly when dealing with spatial variations [169, 170]. However, it's important to note that IVA faces a challenge related to dimensionality when the number of datasets and components increases beyond a certain threshold, especially with a fixed number of samples in each observation [171]. cIVA is a highly efficient approach that integrates prior knowledge about either the sources or the columns of the mixing matrix into the IVA cost function [159, 172]. It alleviates the strict independence assumption by allowing for a flexible equilibrium between data-driven and model-driven methodologies, offering an improved model fit through the utilization of precise constraints. The study conducted in [169] delved into the variations present in fMRI data across both the temporal and spatial dimensions using the cIVA approach. However, it's important to note that cIVA relies on user-defined constraint parameters that govern the level of agreement between the reference signal and the estimated component. Nevertheless, in many real-world scenarios, choosing appropriate constraint parameters poses a challenge because it's often unclear whether the provided prior information is precise. This challenge becomes even more intricate when dealing with the integration of prior information related to multiple signals. As the availability of large-scale fMRI datasets continues to grow, there is a pressing need for a method that can effectively estimate meaningful global functional networks while also capturing individual-specific characteristics. This is a critical step in enabling precision medicine, which aims to fully account for individual differences. In this study, we propose an effective approach to address the high dimensionality challenge of independent ICA and IVA by integrating reference information about functional networks into the ICA and IVA framework respectively. This integration serves as a guide in scenarios with high-dimensional data. cICA-MGGMM and cIVA-MGGMM are valuable methods that incorporate prior information about the sources or columns of the mixing matrix into the ICA and IVA cost functions. They strike a balance between data-driven and model-driven techniques by relaxing the independence assumption

and enforcing accurate constraints. However, cICA-MGGMM and cIVA-MGGMM rely on user-defined constraint parameters, which determine the level of correspondence between the reference signal and the estimated component. The success of these two methods depends on selecting the appropriate prior information and constraint parameters. For instance, when the prior information is inaccurate, lower constraint parameters should be used to avoid imposing incorrect constraints on the decomposition. Conversely, when the prior information is correct, a higher constraint parameter is necessary to minimize the influence of noise and artifacts on the components. Nevertheless, in practical applications, selecting an optimal constraint parameter can be challenging because it is often unclear whether the prior information is accurate. This complexity is further compounded when incorporating prior information related to multiple signals. For that, we introduce adaptive versions for both cICA-MGGMM and cIVA-MGGMM. These adaptive models can regulate the relationship between the reference signals and the estimated patterns, facilitating the seamless integration of multiple reference signals into the ICA and IVA frameworks. The acICA-MGGMM and acIVA-MGGMM methods guarantee that prior information is leveraged to guide the solution rather than imposing incorrect constraints.

### 4.3 Adaptive constrained ICA models

#### 4.3.1 ICA based on multivariate generalized Gaussian mixture model

In the ICA mixture model, it is assumed that the observed data originates from a mixture model and be classified into mutually exclusive classes [46]. In ICA mixture each data vector  $\mathbf{x}_n$  can be represented as:

$$\mathbf{x}_n = \mathbf{A}_j \mathbf{s}_{j,n} + \mathbf{b}_j \quad (4.1)$$

To establish the framework for a multivariate generalized Gaussian mixture model (MGGMM), we assume that each component of this mixture is given as:

$$p(\mathbf{x}|\xi_j) = \frac{\Gamma(\frac{D}{2})}{\pi^{\frac{D}{2}} \Gamma\left(\frac{D}{2\beta_j}\right)} \frac{\beta_j}{2^{2\beta_j} \mathbf{c}_j^{\frac{D}{2}} |\Sigma_j|^{\frac{D}{2}}} \exp\left[\frac{-1}{2\mathbf{c}_j^{\beta_j}} (\mathbf{y}_j)^{\beta_j}\right] \quad (4.2)$$

where  $\mathbf{x} \in \mathbb{R}^D$ ,  $\boldsymbol{\mu}_j$ ,  $\Sigma_j$  and  $\beta_j$  are mean, covariance matrix, and shape parameter respectively,  $\mathbf{y}_j = (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j)$ , and  $\mathbf{c}_j$  is the scale parameter defined as:

$$\mathbf{c}_j = \left( \frac{\beta_j}{DN} \sum_{n=1}^N (\mathbf{x}_n^T \Sigma_j^{-1} \mathbf{x}_n)^{\beta_j} \right)^{\frac{1}{\beta_j}} \quad (4.3)$$

To develop the ICA mixture model (ICA-MGGMM), we adopt Equation (4.2) in Equation (4.1). We use the EM algorithm to estimate the parameters of the MGGMM within the ICA mixture framework through the maximum likelihood approach as discussed in Chapter 2.

### 4.3.2 Constrained ICA-MGGMM

In this subsection, we introduce the cICA-MGGMM, which is considered an improved version of ICA-MGGMM. To define the constraint parameters within the objective function of ICA mixture (ICA-MGGMM), we combine the maximum likelihood of ICA-MGGMM with constraint parameters  $(\gamma_l, \lambda_l)$ , represented in Equation (4.6). The goal is to minimize this objective function while adhering to inequality constraints, as specified in Equation (4.4). In this context, a regularization parameter  $\lambda_l$  is employed to prevent over-fitting and guide the optimization algorithm towards solutions that align more closely with the desired characteristics. The penalty parameter  $(\gamma_l)$  plays a crucial role in controlling the trade-off between maximizing independence and penalizing complexity. The likelihood objective function of the ICA-MGGMM is optimized subject to these constraints, resulting in the formulation of the cICA-MGGMM cost function:

$$h_l(\mathbf{y}_l, \mathbf{r}_l) = \rho_l - \epsilon(\mathbf{y}_l, \mathbf{r}_l) \leq 0 \quad (4.4)$$

where inequality constraint function denoted as  $h_l(\mathbf{y}_l, \mathbf{r}_l)$ , a distance measure represented by  $\epsilon$ ,

a reference value for source  $\mathbf{y}_l$  indicated as  $\mathbf{r}_l$ , and an inequality constraint threshold denoted as  $\rho_l$ . The distance measure is typically defined using correlation distance and mean square error. However, we utilize the absolute value of Pearson's correlation coefficient to define the distance measure as:

$$\epsilon(\mathbf{y}_l, \mathbf{r}_l) = \left| \text{corr}(\mathbf{y}_l, \mathbf{r}_l) \right| \quad (4.5)$$

where the condition  $0 \leq \epsilon(\mathbf{y}_l, \mathbf{r}_l) \leq 1$  is met. The optimization function for cICA-MGGMM is given by:

$$\mathcal{J}_{\text{cICA-MGGMM}} = E \left[ \log p(\mathbf{y}_l) \right] + \log |h_l^T \mathbf{y}_l| + \frac{1}{2\gamma_l} \left( \left( \max\{0, \gamma_l h_l(\mathbf{y}_l, \mathbf{r}_l) + \lambda_l\} \right)^2 - \lambda_l^2 \right) \quad (4.6)$$

where  $\lambda_l$  is a regularization parameter,  $\gamma_l$  is the penalty parameter and  $p(\mathbf{y}_l)$  indicates the ICA-MGGMM objective function.

### 4.3.3 Adaptive Constrained ICA-MGGMM

Incorporating reference signals offers a valuable approach to relax the assumption of independence and enhance the optimization process toward an improved solution. To achieve this, we introduce the acICA-MGGMM method, which effectively regulates the level of correspondence between the estimated source and the reference signal. In our approach, we denote  $L$  as the total number of constraints, where each constraint is associated with a constraint parameter  $\rho_l$ . Additionally,  $\mathbf{r}_l$  represents the reference signal for the  $l$ th constraint. To determine the appropriate constraint parameter, the acIVA-MGGMM method selects a value for  $\rho_l$  from a predefined set of possible values denoted as  $\mathcal{P}$ . Learning steps for acICA-MGGMM are provided in Algorithm 4. The gradient update function for this algorithm is given as:

$$\frac{\partial \mathcal{J}_{\text{acICA-MGGMM}}}{\partial \mathbf{W}_l} = E \left[ \frac{\partial}{\partial \mathbf{W}_l} \log p(\mathbf{y}_l) \right] + \frac{h_l^T}{h_l^T \mathbf{y}_l} + h_l'(\mathbf{y}_l, \mathbf{r}_l) \lambda_l \mathbf{r}_l^T \quad (4.7)$$

where  $h'_l$  represents the derivative of  $h_l$  w.r.t  $(\mathbf{y}_l, \mathbf{r}_l)$ . In each iteration of the algorithm, we update the Lagrange multiplier using gradient ascent:

$$\lambda_l \leftarrow \max \left\{ 0, \gamma_l h_l(\mathbf{y}_l, \mathbf{r}_l) + \lambda_l \right\} \quad (4.8)$$

The constraint parameter is given as:

$$\hat{\rho}_l = \arg \min_{\rho_l \in \mathcal{P}} \left\{ \min \left[ \left| \rho_l - |corr(\mathbf{y}_l, \mathbf{r}_l)| \right| \right] \right\} \quad (4.9)$$

---

**Algorithm 4** Adaptive constrained ICA-MGGMM

---

1. Define a set of  $\rho_l$  values for  $\mathcal{P}$  sorted in ascending order.
  2. Randomly initialize  $W_l$  and set  $\lambda_l = 0, \gamma_l$  as a positive value.
  3. **For**  $n = 1, \dots, N$  **do**  
    Compute  $\mathbf{y} = \mathbf{W}\mathbf{x}$ .
  5. **IF** Reference  $\mathbf{r}_l$  is available **then**  
    Compute  $\lambda_l$  using Equation (4.8).  
    Compute  $\rho_l$  using Equation (4.9).  
    Compute  $\frac{\partial J_{acICA-MGGMM}}{\partial \mathbf{W}_l}$  using Equation (4.7).  
     $\mathbf{W}_l = \mathbf{W}_l + \frac{\partial J_{acICA-MGGMM}}{\partial \mathbf{W}_l}$
  10. **else**  
    Let  $\frac{\partial J_{acICA-MGGMM}}{\partial \mathbf{W}_n} = E \left[ \frac{\partial}{\partial \mathbf{W}_n} \log p(\mathbf{y}_n) \right] + \frac{h_n^T}{h_n^T \mathbf{W}_n}$   
     $\mathbf{W}_n = \mathbf{W}_n + \frac{\partial J_{acICA-MGGMM}}{\partial \mathbf{W}_n}$
  13. **end if**
  14. **End For**
  15. Repeat step 3 to 14 until convergence.
- 

## 4.4 Adaptive constrained IVA models

### 4.4.1 IVA based on multivariate generalized Gaussian mixture model

IVA method assumes that components within a single source are independent while components from different sources are as dependent as possible. Fig. 2.1 demonstrates that IVA simultaneously achieves maximum independence within each source and maximum dependence across multiple sources. Unlike performing separate ICAs on each source, IVA achieves this joint optimization

[50]. The traditional IVA model is given in detail in Chapter 2.

The gradient of the cost function for the IVA mixture can be determined by incorporating Equations (4.2) and (2.40) as:

$$\frac{\partial J_{IVA-MGGMM}}{\partial W^{[m]}} = E \left[ \frac{\beta_j}{2\mathbf{c}_j} (\mathbf{y}_j)^{\beta_j-1} \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j) \mathbf{x}_n^{[m]} \right] - \frac{h_n^{[m]}}{(h_m^{[m]})^T \mathbf{W}_n^{[m]}} \quad (4.10)$$

where  $\mathbf{y}_j = (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j)$ .

#### 4.4.2 Constrained IVA-MGGMM

By combining the advantages of both techniques, cIVA effectively captures the unique variability exhibited by the features extracted from a group of subjects. The cIVA approach incorporates prior knowledge about the sources into the IVA model, constraining the solution space and effectively addressing the challenges posed by high-dimensionality. By introducing a constraint term, the cost function for the constrained IVA-MGGMM can be expressed as:

$$\begin{aligned} J_{cIVA-MGGMM} &= J_{IVA-MGGMM} - \sum_{l=1}^L \frac{1}{2\gamma_l} \sum_{m=1}^m \left\{ \left[ \max \left\{ 0, \lambda_l^{[m]} + \gamma_l^{[m]} g(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \right\} \right]^2 \right. \\ &\quad \left. - \left( \lambda_l^{[m]} \right)^2 \right\} \end{aligned} \quad (4.11)$$

where  $\lambda_l^{[m]}$  is the regularization parameter,  $\gamma_l^{[m]}$  is the penalty parameter and  $g(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]})$  is the inequality constraint function given as:

$$g(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) = \rho_l - \epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \leq 0 \quad (4.12)$$

where  $\mathbf{y}_l^{[m]} = (\mathbf{W}_l^{[m]}) \mathbf{x}^{[m]}$  is the estimated constraint component,  $\mathbf{r}_l^{[m]}$  is the reference vector for each corresponding constraint component of the SCVs,  $\epsilon(\cdot, \cdot)$  is a function to measure the similarity between the estimated SCVs and the reference signal and  $\rho_l$  is the constraint parameter. The constraint function, defined in Equation (4.12), provides flexibility in choosing dissimilarity functions, including the inner product, mean square error, mutual information, and correlation. In this work,

we utilize the absolute value of Pearson's correlation coefficient as follows:

$$\epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) = \left| \text{corr}(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \right| \quad (4.13)$$

The absolute value of Pearson's correlation coefficient as a similarity measure restricts  $\epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]})$  to be between 0 and 1, steering  $\rho_l \leq 1$ .

### 4.4.3 Adaptive constrained IVA-MGGMM

We propose the acIVA-MGGMM method, which allows for regulating the level of correspondence between the estimated source and the reference signal. The acIVA-MGGMM algorithm selects a value of  $\rho_l$  from a set of possible values denoted as  $\mathcal{P}$ . The demixing matrices, denoted as  $\mathbf{W}^{[m]}$ , are randomly initialized, and we set  $\lambda_l^{[m]} = 0$  and  $\gamma_l^{[m]}$  to a positive scalar value. During each iteration, we estimate the sources,  $\mathbf{y}_l^{[m]}$ , where  $m$  ranges from 1 to  $M$  and  $l$  ranges from 1 to  $L$ . The first SCV is selected as the constrained component, followed by the estimation of  $\lambda_l^{[m]}$  and  $\rho_l$  using the following Equations, respectively:

$$\lambda_l^{[m]} = \max \left[ 0, \gamma_l^{[m]} g(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) + \lambda_l^{[m]} \right] \quad (4.14)$$

$$\hat{\rho}_l = \arg \min_{\rho_l \in \mathcal{P}} \left\{ \min \left[ \left| \rho_l - \left| \text{corr}(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \right| \right| \right] \right\} \quad (4.15)$$

The acIVA-MGGMM gradient is computed as:

$$\begin{aligned} \frac{\partial J_{\text{acIVA-MGGMM}}}{\partial \mathbf{W}_l^{[m]}} &= \frac{\partial J_{\text{IVA-MGGMM}}}{\partial \mathbf{W}_l^{[m]}} - \frac{1}{\gamma_l^{[m]}} \left\{ \left[ \max \left( 0, \gamma_l^{[m]} \left( \hat{\rho}_l - \epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \right) \right) \right] \right. \\ &\quad \left. + \left[ \lambda_l^{[m]} \right]^2 - \left( \lambda_l^{[m]} \right)^2 \right\} \end{aligned} \quad (4.16)$$

To update this parameter, we search for the highest value of  $\rho_l$  from the set  $\mathcal{P}$  that satisfies the condition specified in Equation (4.12) for all  $M$  subjects. Subsequently, we compute the distance  $\epsilon(., .)$  of the estimated correlation for all possible values of  $\rho_l$  from the set  $\mathcal{P}$  across all subjects.

---

**Algorithm 5** Adaptive constrained IVA-MGGMM

---

- 1: Define a set of  $\rho_l$  values for  $\mathcal{P}$  sorted in ascending order.
- 2: Randomly initialize  $M$  de-mixing matrices  $\mathbf{W}^{[m]}$ .
- 3: Set  $\lambda_l^{[m]} = 0, \gamma_l^{[m]}$  as a positive value.
- 4: **for**  $n=1, \dots, N$  **do**
- 5:     **for**  $m=1, \dots, M$  **do**
- 6:         Compute  $\mathbf{y}^{[m]} = \mathbf{W}^{[m]}\mathbf{x}^{[m]}$ ,  $m=1, \dots, M$ .
- 7:         **if**  $n \in \{1, \dots, L\}$  **then**
- 8:              $i=1$
- 9:             **repeat**
- 10:                  $i \leftarrow i + 1$
- 11:                 **until**  $\rho_{l_i} - |\epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]})| < 0 < \rho_{l_{i+1}} - |\epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]})|$
- 12:                 Compute  $\lambda_l^{[m]}$  using Equation (4.14).
- 13:                 Compute  $\hat{\rho}_l$  using Equation (4.15).
- 14:                 Compute  $\frac{\partial J_{acIVA-MGGMM}}{\partial \mathbf{W}_l^{[m]}}$  using Equation (4.16).
- 15:                 Update  $\mathbf{W}_l^{[m]} \leftarrow \mathbf{W}_l^{[m]} + \frac{\partial J_{acIVA-MGGMM}}{\partial \mathbf{W}_l^{[m]}}$ .
- 16:             **else**
- 17:                 Compute  $\frac{\partial J_{IVA-MGGMM}}{\partial \mathbf{W}_n^{[m]}}$ .
- 18:                 Update  $\mathbf{W}_n^{[m]} \leftarrow \mathbf{W}_n^{[m]} + \frac{\partial J_{IVA-MGGMM}}{\partial \mathbf{W}_n^{[m]}}$ .
- 19:             **end if**
- 20:         **end for**
- 21:     **end for**
- 22: Repeat until convergence.

---

The value of  $\rho_l$  associated with the smallest distance is selected. This new value of the constraint parameter is utilized to calculate the gradient in line 14 of Algorithm 5 and update the demixing matrix. The process is then repeated, obtaining a new estimation of the sources, until convergence is achieved.

## 4.5 Results and discussion

We apply our proposed models cICA-MGGMM, acICA-MGGMM, cIVA-MGGMM, and acIVA-MGGMM to detect mental disorders in their early stages. Initially, we evaluate the effectiveness of our models by applying them to simulated data generated randomly from the multivariate generalized Gaussian distribution. In this experiment, we examine how well our models can estimate the mixture models and evaluate their convergence. To differentiate the performance between the cIVA-MGGMM and acIVA-MGGMM, we assess their performance based on the dissimilarity factor. In real-world applications, we employ these models to distinguish between healthy subjects

and Alzheimer’s patients using the ADNI public dataset. For this, we collected 120 subjects from the ADNI website with Alzheimer’s and healthy controls. We also use our models to detect the most defective brain regions for Alzheimer’s patients. Then, we test our proposed models over the Schizophrenic dataset. We chose the COBRE data which includes 140 subjects, 70 healthy controls, and 70 Schizophrenic patients. We used the adaptive constrained IVA-MGMM components to localize the affected brain regions for Schizophrenic patients. To further validate our models, we apply them to predict subjects with Seizures and those without Seizures using EEG data. Additionally, we expand the application of our models to identify individuals with ADHD in comparison to healthy controls. To compare the performance of the proposed models and the base models, we used joint inter-symbol interference (Joint-ISI) and spatial correlation. ISI finds the average performance over several runs and is given by:

$$ISI(\mathbf{G}) = \frac{1}{2N(N-1)} \left[ \sum_{n=1}^N \left( \sum_{j=1}^M \frac{|g_{nj}|}{\max_j |g_{nj}|} - 1 \right) + \sum_{j=1}^M \left( \sum_{n=1}^N \frac{|g_{nj}|}{\max_j |g_{nj}|} - 1 \right) \right] \quad (4.17)$$

Here,  $\mathbf{G} = \mathbf{WA}$  represents the correlation matrix between the original signals and the separated signals, where  $g_{nj}$  denotes the correlation coefficient between the  $n$ th source and the  $m$ th separated signals. The ISI index is utilized to assess the quality of a decomposition by quantifying the deviation of  $\mathbf{G}$  from the identity matrix. In the context of decomposing multiple datasets, the ISI can be extended to a joint ISI (Joint-ISI) for IVA as:

$$Joint-ISI(\mathbf{G}) = ISI \left( \sum_{m=1}^M |\mathbf{G}^{[m]}| \right) \quad (4.18)$$

#### 4.5.1 Simulation results

To assess the robustness of our proposed models, we evaluate their performance on simulated data generated from the multivariate generalized Gaussian distribution. Here, we generated two different datasets with 2 and 3 clusters respectively. Initially, we employ the K-means algorithm to set the mixture parameters, encompassing mean, covariance, and shape parameters. The mixing

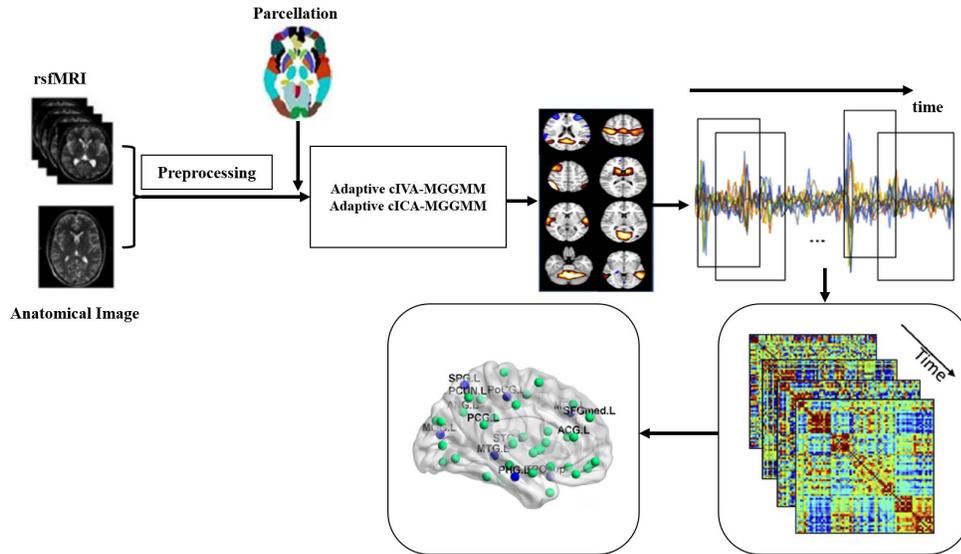


Figure 4.1: ICA and IVA models for Alzheimer’s and Schizophrenia detection based on resting state fMRI data.

parameter is initialized at  $1/M$ , ensuring a sum of 1. Subsequently, we utilize the EM algorithm to compute the posterior probability in the E-step and update the mixture parameters in the M-step. To gauge model convergence, a small tolerance value is specified. We iterate, updating the mixture parameters until our proposed model achieves convergence, utilizing the optimal parameters for source separation. Tables 4.1 and 4.2 present the mixture parameters along with their corresponding estimated values. In this context, we utilize the acIVA-MGGMM to derive estimations for mixing, mean, shape, and covariance parameters. Tables 4.1 and 4.2 indicate that our proposed model accurately estimates the mixture parameters. We set a maximum of 1020 iterations for each model, with a convergence condition defined by a small tolerance of 0.001. This iterative process yields a new de-mixing matrix denoted as  $\mathbf{W}$ , which is employed to estimate the sources. The convergence process is visualized in Figs. 4.2 and 4.3. Notably, our proposed models, acICA-MGGMM and acIVA-MGGMM, exhibit significantly improved convergence compared to the base models ICA and IVA. Specifically, our proposed models converge at 250 iterations, whereas the base models require more than 500 iterations to converge. We evaluated the effectiveness of the acICA-MGGMM and acIVA-MGGMM methods across a range of  $\gamma_l$  values spanning from 1 to 1000. No significant variation in their performance was observed. Consequently, we opted for a small positive value, setting  $\gamma_l$  to 3. Regarding the cICA-MGGMM and cIVA-MGGMM methods with fixed constraint

Table 4.1: Mixture parameters estimation using proposed acIVA-MGGMM for generated data with 2 clusters.

Generated data	Mixture parameters			
	$\pi$	$\mu$	$\beta$	$\Sigma$
$n_j = 300$	0.5	2	[0.01, 0.02, 0.02]	$\begin{bmatrix} 1 & 0.5 & 0.2 \\ 0.5 & 1 & 0.5 \\ 0.2 & 0.5 & 1 \end{bmatrix}$
$n_j = 300$	0.5	1.5	[0.04, 0.01, 0.3]	$\begin{bmatrix} 1 & 0.4 & 0.25 \\ 0.4 & 1 & 0.4 \\ 0.25 & 0.4 & 1 \end{bmatrix}$
Generated data	Estimated parameters			
	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\Sigma}$
$n_j = 300$	0.49	1.8	[0.02, 0.04, 0.03]	$\begin{bmatrix} 0.98 & 0.49 & 0.19 \\ 0.49 & 0.92 & 0.48 \\ 0.19 & 0.48 & 0.99 \end{bmatrix}$
$n_j = 300$	0.51	1.3	[0.03, 0.02, 0.27]	$\begin{bmatrix} 1.12 & 0.35 & 0.22 \\ 0.43 & 0.97 & 0.43 \\ 0.23 & 0.38 & 0.93 \end{bmatrix}$

parameters, we varied the values of  $\rho_l$  within the range of 0.001 to 0.9. A  $\rho_l$  value of 0.001 represents weaker constraints, while a value of 0.9 indicates stronger constraints. In Figs. 4.4 (a) and (b), we assessed the performance of acICA-MGGMM against cICA-MGGMM and acIVA-MGGMM against cIVA-MGGMM concerning the dissimilarity factor. Each box in the figures displays the median as a horizontal orange line, with the upper and lower edges corresponding to the 75th and 25th percentiles, respectively. A higher dissimilarity factor suggests poorer source estimation. With the cICA-MGGMM method, employing a higher constraint parameter leads to a deterioration in the estimation quality of the constrained component, while the acICA-MGGMM method maintains a low dissimilarity factor. A similar pattern is observed with cIVA-MGGMM and acIVA-MGGMM when using a higher constraint parameter; the dissimilarity factor of the constrained vector remains low with acIVA-MGGMM but increases with the regular cIVA-MGGMM.

Table 4.2: Mixture parameters estimation using proposed acIVA-MGGMM for generated data with 3 clusters.

Generated data	Mixture parameters			
	$\pi$	$\mu$	$\beta$	$\Sigma$
$n_j = 300$	0.33	2	[0.01, 0.02, 0.02]	$\begin{bmatrix} 1 & 0.5 & 0.2 \\ 0.5 & 1 & 0.5 \\ 0.2 & 0.5 & 1 \end{bmatrix}$
$n_j = 300$	0.33	1.5	[0.04, 0.01, 0.3]	$\begin{bmatrix} 1 & 0.4 & 0.25 \\ 0.4 & 1 & 0.4 \\ 0.25 & 0.4 & 1 \end{bmatrix}$
$n_j = 300$	0.33	1	[1.34, 2.12, 1.22]	$\begin{bmatrix} 1 & 0.8 & 0.7 \\ 0.8 & 1 & 0.8 \\ 0.7 & 0.8 & 1 \end{bmatrix}$
Generated data	Estimated parameters			
	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\Sigma}$
$n_j = 300$	0.42	1.7	[0.03, 0.04, 0.01]	$\begin{bmatrix} 0.96 & 0.44 & 0.23 \\ 0.45 & 1.1 & 0.45 \\ 0.23 & 0.52 & 1.2 \end{bmatrix}$
$n_j = 300$	0.30	1.2	[0.05, 0.03, 0.25]	$\begin{bmatrix} 1.1 & 0.43 & 0.26 \\ 0.34 & 1.2 & 0.42 \\ 0.26 & 0.42 & 0.95 \end{bmatrix}$
$n_j = 300$	0.28	0.75	[0.93, 2.25, 1.45]	$\begin{bmatrix} 1.3 & 0.73 & 0.62 \\ 0.83 & 0.93 & 0.72 \\ 0.65 & 0.82 & 1.1 \end{bmatrix}$

#### 4.5.2 Alzheimer's detection using cICA-MGGMM and acICA-MGGMM

Alzheimer's disease is the most common neurodegenerative disease in the elderly [173]. There is a significant delay between the start of AD pathology and the clinical diagnosis of AD dementia, which can only be confirmed by autopsy [173, 174]. As a result, it is extremely difficult to detect AD early and accurately [174] and there is a need for intelligent methods to assist clinicians in the personalized diagnosis of this disease [175]. Early and accurate detection of Alzheimer's disease is beneficial for disease management [176]. Magnetic resonance imaging (MRI) is an imaging

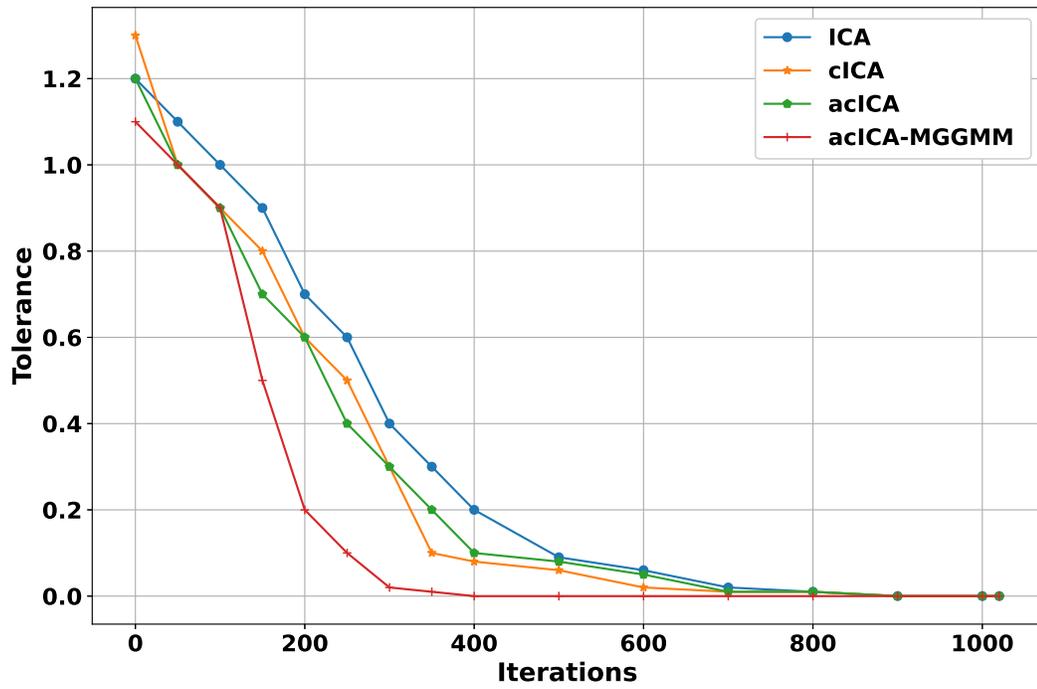


Figure 4.2: Number of iterations needed for acICA-MGGMM to converge.

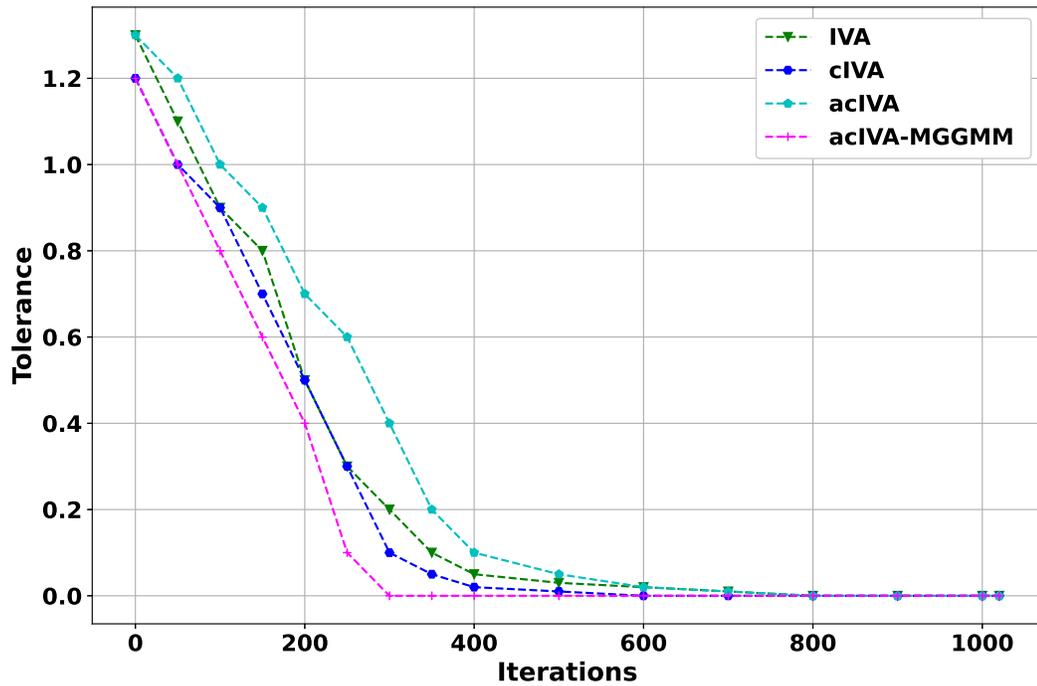


Figure 4.3: Number of acIVA-MGGMM iterations.

technique that produces high-quality images of the anatomical structures of the human body, particularly the brain and provides valuable information for clinical diagnosis and biomedical research

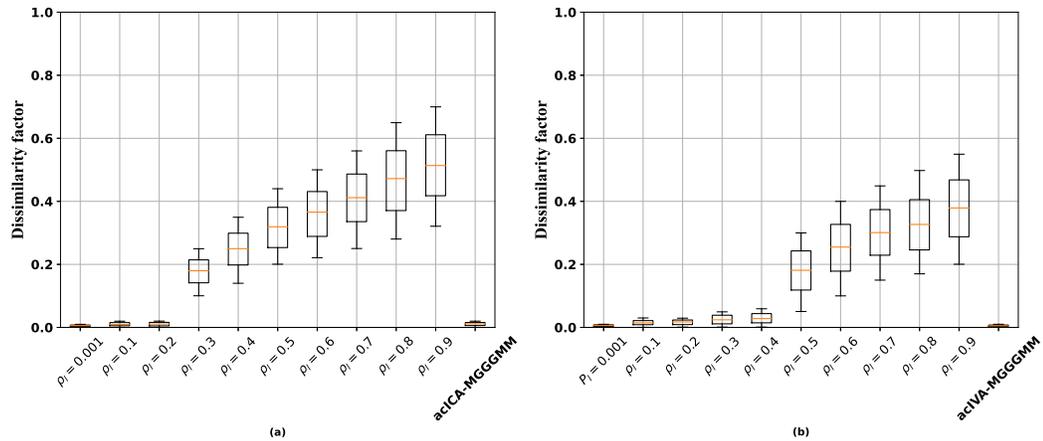


Figure 4.4: Dissimilarity factor for (a) acICA-MGGMM vs cICA-MGGMM, (b) acIVA-MGGMM vs cIVA-MGGMM.

[177]. The automated and accurate classification of MR images improves the diagnostic value of MRI [178, 179].

## Dataset

The data used in this study the ADNI database. All ADNI studies adhere to the Helsinki Declaration and U.S. regulations including 21 CFR Part 50 (Protection of Human Subjects), and Part 56 (Institutional Review Boards), as well as good clinical practice guidelines. The ADNI protocol received approval from the Institutional Review Boards of all participating institutions. In this study, 120 subjects from ADNI were used with 60 female and 60 male subjects divided into normal controls (NC) and AD patients. The rs-fMRI images were collected using Philips Achieva scanners operating at 3.0 Tesla. The scan protocol employed echo planner imaging (EPI) with the following parameters: Echo time (TE) of 30 ms, repetition time (TR) of 3000 ms, flip angle of 80 degrees, pixel size of  $3.3 \times 3.3$  mm, acquisition matrix size of  $64 \times 64$ , and slice thickness of 3.3 mm. The scanning process included the acquisition of 48 slices and 140 volumes. A complete description of ADNI is available publicly <sup>1</sup>. Moreover, the data access requests are to be sent to the ADNI website <sup>2</sup>.

<sup>1</sup><https://adni.loni.usc.edu/>

<sup>2</sup><http://adni.loni.usc.edu/data-samples/access-data/>

## Data preprocessing

The rs-fMRI data underwent standard preprocessing using the statistical parametric mapping (SPM12) software package [180]. The preprocessing steps included discarding the first ten time-point volumes to achieve magnetization equilibrium, performing slice time correction for interleaved acquisition, correcting subject motion using realignment, and co-registering functional and structural images. The images were then normalized to the SPM12 EPI template and subsequently smoothed using a Gaussian kernel with a full-width half-maximum of 4 mm. Then, based on the AAL atlas, each volume was parcellated into 116 ROIs [181]. Finally, for each ROI, the average intensity time series was extracted and then subjected to a band-pass filter ranging from 0.01 Hz to 0.08 Hz. This filtering process aims to diminish non-neuronal influences on the fluctuations of the blood-oxygen-level-dependent (BOLD) signal.

## Experimental Results for cICA-MGGMM and acICA-MGGMM

We employ the proposed cICA-MGGMM and acICA-MGGMM on Alzheimer’s resting state fMRI data to differentiate Alzheimer’s patients from healthy subjects. In this experiment, we use 55 uncorrelated components for each subject. These components are the most informative, resulting from applying PCA on the parcellated data. We apply the cICA-MGGMM and acICA-MGGMM to the reduced data to distinguish between healthy subjects and Alzheimer’s patients. For cICA-MGGMM, we use the following values for the inequality constraint threshold and the fixed constrained learning parameter: we vary  $\rho_l$  from 0.001 to 0.9, where the constraint parameter with the smallest distance  $\epsilon$  is chosen as the optimal parameter for the separation process, and  $\gamma_l$  to 3. The acICA-MGGMM method uses the range of constraints parameter  $\rho_l$  from the sets  $\mathcal{P}$  and  $\gamma_l$ . The set  $\mathcal{P}$  is defined as 0.001, ..., 0.9, and we set  $\gamma_l$  to 3. We randomly initialize the parameters for this model and run it five times with different initializations. The Joint-ISI metric is used to select the best of these five runs. The demixing matrix for the best run is used to estimate the sources. To validate the separation quality, we use the Joint-ISI to measure the difference between the reference signal and the estimated one. All these steps are shown in Fig. 4.1. Table 4.3 compares ICA, cICA, Non-linear ICA (Non-ICA), acICA, cICA-MGGMM, and acICA-MGGMM performance in

Table 4.3: Performance of the proposed cICA-MGGMM and acICA-MGGMM in terms of Joint-ISI with respect to the number of sources.

# of combinations	acICA-MGGMM	cICA-MGGMM	acICA	cICA	Non-ICA	ICA
2	0.0356	0.0476	0.0452	0.0423	0.0441	0.0421
5	0.0369	0.0561	0.0582	0.0462	0.0487	0.0521
10	0.0352	0.0650	0.0692	0.0673	0.0690	0.0533
20	0.0327	0.0672	0.0745	0.0748	0.0774	0.0622
40	0.0347	0.0701	0.0761	0.0842	0.0896	0.0634
60	0.0701	0.0728	0.0783	0.0881	0.0793	0.2351
80	<b>0.2756</b>	<b>0.3021</b>	0.3180	0.3291	0.3195	0.3428
100	<b>0.3052</b>	<b>0.3322</b>	0.3492	0.3653	0.3524	0.3941
120	<b>0.3442</b>	<b>0.3792</b>	0.3816	0.3973	0.3912	0.4336

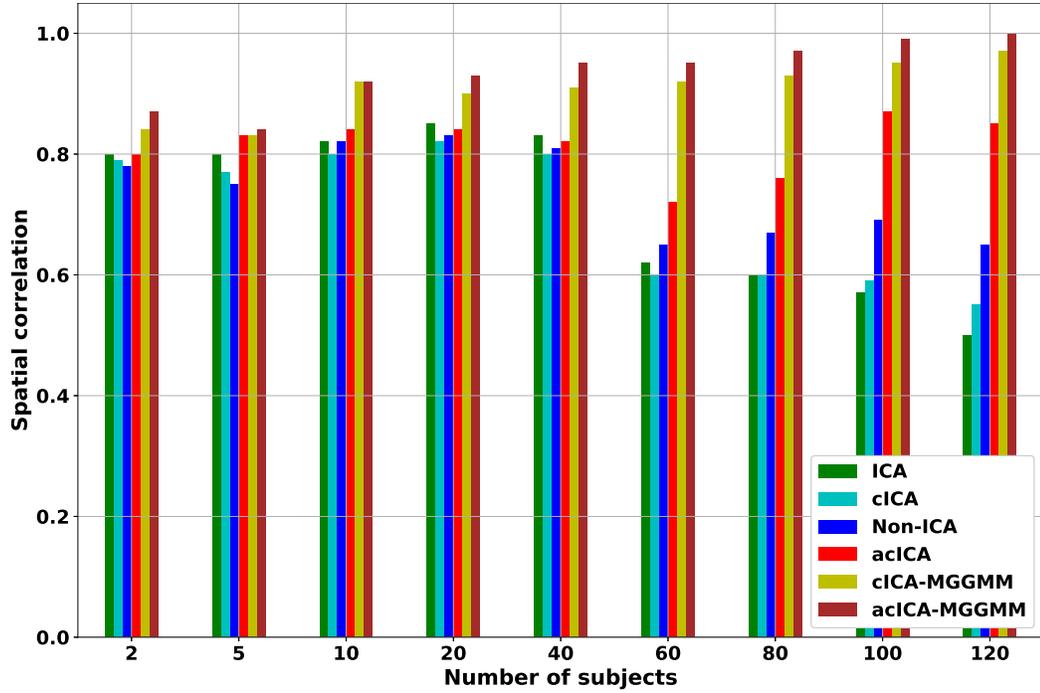


Figure 4.5: Spatial correlation for the proposed cICA-MGGMM and acICA-MGGMM concerning the number of subjects.

terms of Joint-ISI. This Table displays that our proposed models exhibit slightly better performance compared to the basic models across a variety of lower combinations, which vary from 2 to 60. It is clear that as the number of subjects increases, the ICA-compared models' performance degrades. We see significant improvements in performance for cICA-MGGMM and acICA-MGGMM starting with 80 subjects when compared to ICA base models. It is also clear that the performance of

cICA-MGGMM is slightly less than that of acICA-MGGMM, depending on the constraint parameters used. The acICA-MGGMM model addresses this limitation in high-dimensional data using the prior information.

We use the spatial correlation metric to assess the robustness of our proposed model as shown in Fig. 4.5. The cICA-MGGMM and acICA-MGGMM clearly outperform the base models in terms of preserving the correlation between the reference signal and estimated sources, especially when a large number of subjects is used. The spatial correlation between the estimated constrained sources and the corresponding reference sources demonstrates a similar pattern to the findings observed in the Joint-ISI analysis. The performance degradation for ICA, Non-ICA, cICA, and acICA in high-dimensional data is shown for data with 80, 100, and 120 combinations of subjects.

### 4.5.3 Alzheimer’s detection using cIVA-MGGMM and acIVA-MGGMM

After completing the preprocessing step of the fMRI data, the AAL atlas is employed for the parcellation procedure to create 116 brain regions. The mean fMRI signal for each region is then calculated. We employ acIVA-MGGMM on both patients and healthy groups, using the group ICA components of these classes as reference signals. To select the most informative components, we use PCA to reduce the dimensions of each subject’s data. From this, we obtain a dimension-reduced dataset,  $\mathbf{x}^{[m]} \in R^{55 \times 58600}$ , consisting of  $N = 55$  uncorrelated components per subject.

To implement cIVA-MGGMM and acIVA-MGGMM models, we divide the data of each subject into  $T = 15$  windows with a length of  $L = 14$  and a 50% overlap. This resulted in a total of  $TK = 1800$  windows. By employing cIVA-MGGMM and acIVA-MGGMM on the data of each subject, we reduce the dimensionality of the SCV from 1800 to 15. The first SCV is required to be correlated with one of the 15 group components. We apply cIVA and acIVA methods, which use the IVA-MGGMM algorithm, to the windowed datasets of each subject. In the case of cIVA-MGGMM, we set  $\mathcal{P}$  and  $\gamma_l$  to  $(0.001 - 0.9)$  and 3 respectively. The set  $\mathcal{P}$  and  $\gamma_l^{[m]}$  were defined as in subsection 4.5.2 for acIVA-MGGMM. We run both cIVA-MGGMM and acIVA-MGGMM five times with different initializations. We then choose the best run using the Joint-ISI index, which extends the ISI approach to multiple subjects. The estimated demixing matrices of the chosen run are used to compute the estimated sources. We denote the estimated source corresponding to the  $l$ th

constraint for the  $m$ th subject at the  $t$ th window from the consistent run as  $\mathbf{y}_l^{[t,m]}$ .

According to Fig. 4.6, the efficiency of IVA increases as the number of subjects increases but only up to a particular limit for various numbers of sources. This range can be regarded as the optimal range where IVA exhibits reliable estimation of the underlying parameters. Within this range, the application of cIVA-MGGMM and acIVA-MGGMM does not yield a significant performance improvement compared to IVA. This suggests that the performance achieved by IVA in this range is the best that can be obtained. However, when dealing with high-dimensional data, employing cIVA-MGGMM and acIVA-MGGMM demonstrates a remarkable enhancement in performance. This indicates that incorporating prior information serves as a valuable reference for finding superior solutions in such scenarios.

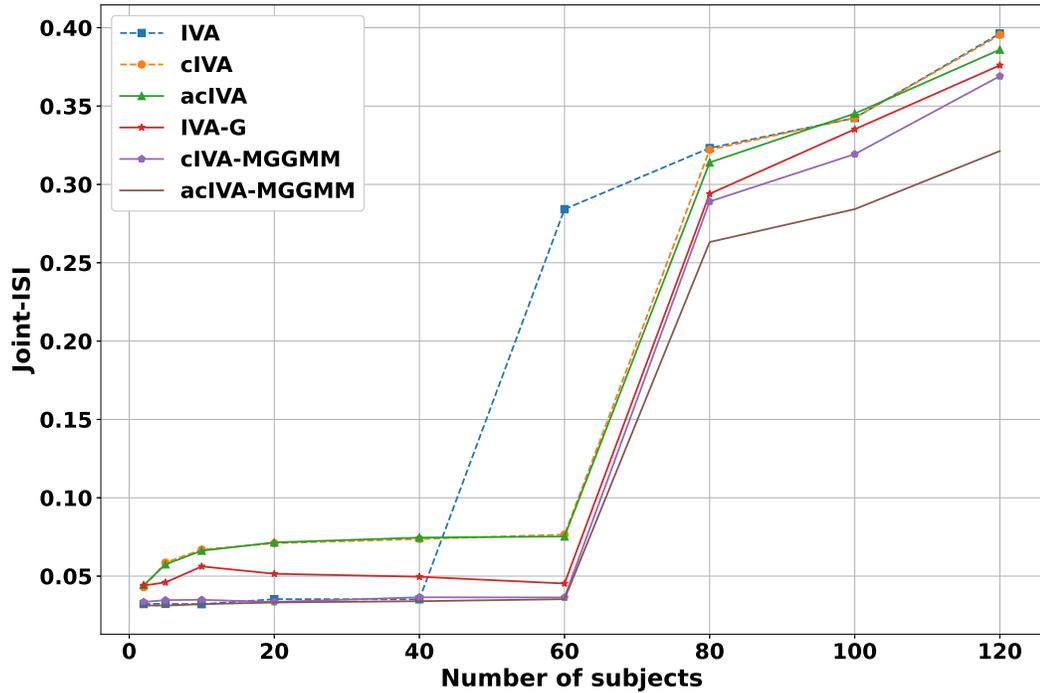


Figure 4.6: Performance of the proposed cIVA-MGGMM and acIVA-MGGMM compared to IVA, IVA-G, cIVA, and acIVA in terms of the number of subjects using the Joint-ISI metric.

The spatial correlation between the estimated constrained sources and the corresponding ground truth exhibits a comparable pattern to the findings depicted in Fig. 4.7. A high spatial correlation indicates that the acIVA-MGGMM method does not impose excessive constraints on the decomposition process, thereby successfully estimating spatial components consistently across subjects.

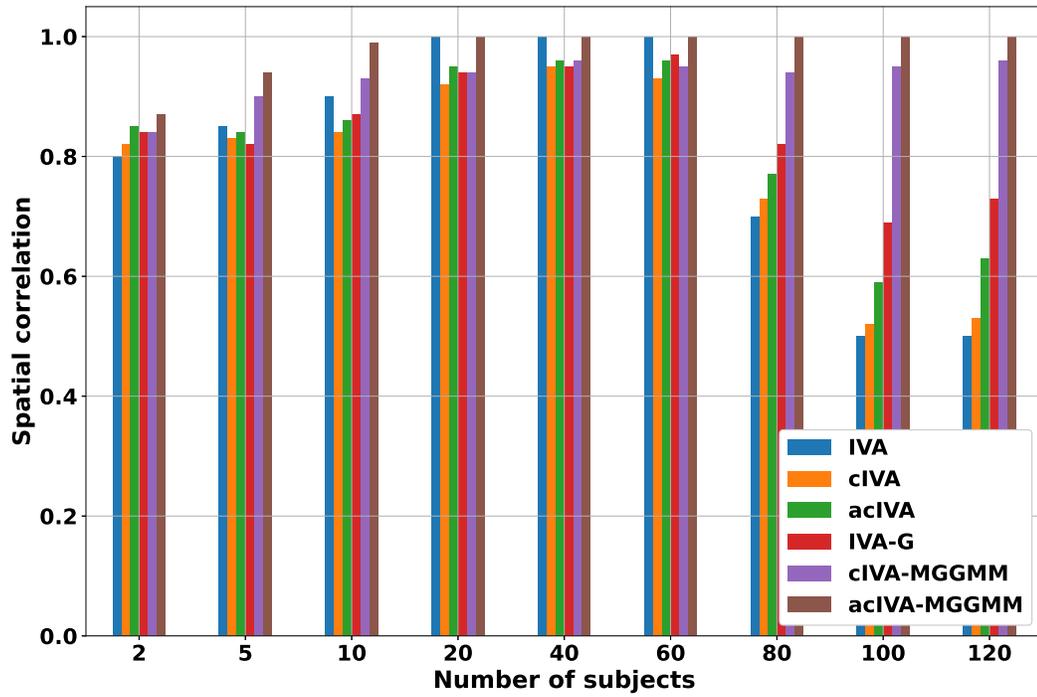


Figure 4.7: Spatial correlation for IVA, IVA-G, cIVA, acIVA, cIVA-MGGMM, and acIVA-MGGMM in terms of number of subjects.

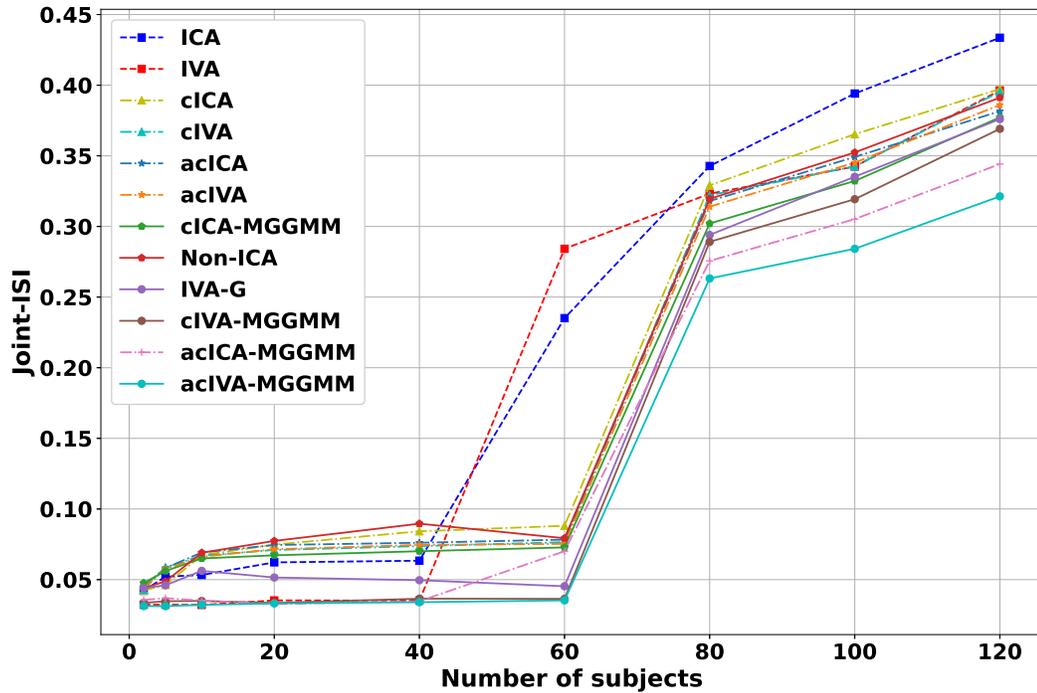


Figure 4.8: The joint-ISI for the four proposed models; cICA-MMGGMM, acICA-MMGGMM, cIVA-MGGMM, and acIVA-MGGMM in terms of the number of subjects.

Consequently, in high-dimensional scenarios, the acIVA-MGGMM method surpasses the standard IVA approach by accurately estimating the underlying sources. It is also clear that cIVA-MGGMM outperforms IVA, cIVA, IVA-G, and acIVA, particularly for high-dimensional data with 80, 100 and 120 subjects' combinations.

Finally, we use the Joint-ISI metric with the same setup for all experiments to validate the four proposed models cICA-MGGMM, acICA-MGGMM, cIVA-MGGMM and acIVA-MGGMM with the base models ICA, Non-ICA, cICA, acICA, IVA, IVA-G, cIVA and acIVA on Alzheimer's fMRI data. Fig. 4.8 shows that the proposed models outperform the base models when the number of subjects is large. These findings demonstrate the ICA and IVA limitation in high-dimensional data. Based on the findings presented in the figure, it is evident that our models exhibit superior capability in recovering the constrained sources in high-dimensional scenarios compared to the compared models, which performed poorly. This observation highlights the valuable role of prior information as a reference for seeking improved solutions in high-dimensional scenarios.

### **Brain networks identification**

The acIVA-MGGMM model improves functional network detection by using reference signals. This is due to its adaptive nature which allows for more flexible interaction among functionally connected regions. The DMN is a complex network in the brain consisting of interconnected regions that act as both hubs and subsystems. It is frequently associated with activation during rest when an individual is contemplating their own thoughts, other people, and events from both the past and future. As a result, DMN has been extensively studied in a variety of disorders. The highly activated regions using the acIVA-MGGMM method: frontal superior medial left and right (SFGmed), anterior cingulate and paracingulate gyri (ACG), parahippocampal gyrus (PHG), and precentral gyrus (PreCG). The significant regions that are shown in Fig. 4.9 are in agreement with the previous findings. Several studies have found that patients with Alzheimer's disease have a significant reduction in the volume of grey and white matter in the superior temporal gyrus (STG) [182, 183, 184]. Consistent with these findings, we observe that the functional connectivity of the STG in the early stages of AD was disrupted. The volume of the PHG is found to be significantly smaller in Alzheimer's disease patients than in healthy control subjects and atrophy of this region in

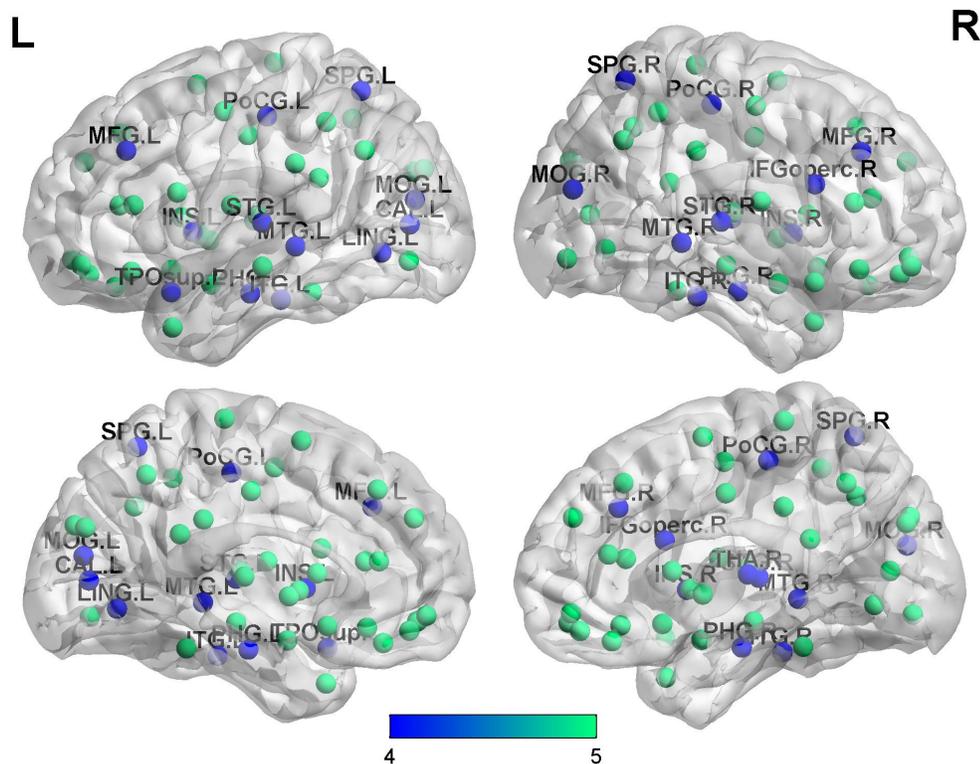


Figure 4.9: Localization of the areas in the AAL atlas. The areas exhibiting higher discriminability between the healthy and Alzheimer's groups are depicted by blue circles, while unaffected areas are represented by green circles.

Alzheimer's disease was linked to specific aspects of the patient's memory impairments [184]. Our findings are consistent with previous research studies which found that the PHG is lower in both the left and right brains of Alzheimer's patients. Several previous studies have shown inferior frontal gyrus (IFG) and middle temporal gyrus (MTG) involvement in AD [185], though the role of these regions in the early stages of AD remains unknown. Our findings revealed that the opercular part of IFG (IFG-operc) in the right hemisphere differed significantly between AD and healthy controls. The MTG has been shown to be involved in accessing lexical and semantic information [186]. Our findings indicate that this region is involved in the disruption of these functions in the early stages of AD. We identified another brain region that differed significantly between AD and healthy subjects. Middle frontal gyrus (MFG), left and right insula (INS), Left calcarine fissure and surrounding cortex (CAL) and lingual gyrus (LING) [184] are among these regions. These brain regions previously have been found to be highly discriminative for separating patients with AD from healthy controls

[187]. These brain regions are also found to be significantly different in AD and healthy subjects: middle occipital gyrus (MOG), Postcentral (PoCG), and right thalamus (THA).

#### **4.5.4 Schizophrenia detection using cICA-MGGMM and acICA-MGGMM**

Schizophrenia is a serious mental disorder that affects approximately 1% of the general population and is characterized by abnormal sensory perceptions, cognitive impairments, concrete thinking, and limited emotional range [188]. The use of functional MRI (fMRI) and structural MRI (sMRI) is becoming more widespread as these techniques show promise in diagnosing neurological disorders including Schizophrenia. Additionally, machine learning can analyze images to extract relevant information and create models that accurately predict the likelihood of disease onset, outperforming traditional methods. This technique has been used successfully in the diagnosis of Schizophrenia, as evidenced by research studies.

#### **Dataset**

The dataset used in this study is sourced from the COBRE database. The dataset comprises functional and anatomical MR data samples obtained from a total of 70 healthy subjects and 70 patients diagnosed with Schizophrenia. The following parameters were used for anatomical imaging: TR/TE/TI = 2530/[1.64, 3.5, 5.36, 7.22, 9.08]/900 ms, flip angle = 7 degrees, slab thickness = 176 mm, FOV = 256 × 256 mm, data matrix = 256 × 256 × 176, number of echoes = 5, voxel size = 1 × 1 × 1 mm, pixel bandwidth = 650 Hz, and a total scan time of 6 minutes. The TR, TI, and time to encode partitions for the multi-echo magnetization prepared rapid acquisition gradient echo (MEMPR) with 5 echoes are similar to those of a conventional MPRAGE, thereby resulting in a comparable contrast between gray matter, white matter, and cerebrospinal fluid. For rs-fMRI data collection, a single-shot full k-space EPI sequence was employed with ramp sampling correction, using the inter-commissural line (AC-PC) as a reference. The acquisition parameters for rs-fMRI were as follows: TR = 2 seconds, TE = 29 ms, matrix size = 64 × 64, 32 slices, and voxel size = 3 × 3 × 4 mm<sup>3</sup>. In addition to rs-fMRI and anatomical MRI data, phenotypic data including age, handedness, gender, and diagnostic information were recorded for each subject.

## Data preprocessing

The preprocessing steps are discussed in subsection 2.6.4.

## Experimental Results for cICA-MGGMM and acICA-MGGMM

After the preprocessing steps, we parcellate the brain regions into 90 brain regions using the AAL atlas. The PCA reduction is then used to extract the most informative uncorrelated components  $\mathbf{x}^{[m]} \in R^{20 \times 52420}$ . The data for each subject is partitioned and then subjected to a windowing process based on the criteria specified in the subsection 4.5.3. Our proposed models, namely acICA-MGGMM, cICA-MGGMM, and the base models are utilized to analyze the reduced data and distinguish between the components associated with Schizophrenia and those belonging to the healthy control group. To establish the experimental setup, for cICA-MGGMM, we set  $\mathcal{P}_l$  and  $\gamma_l$  to  $(0.001 - 0.9)$  and 3 respectively. For acICA-MGGMM, we configure  $\mathcal{P}$  and  $\gamma_l$  according to the Alzheimer’s experimental setting described in subsection 4.5.2. After running our models five times with different initializations, we utilize the Joint-ISI metric to select the optimal run. Subsequently, we employ the demixing matrix obtained from the best run to differentiate between the sources associated with Schizophrenia and those related to the healthy control class.

Table 4.4 presents a comparison of the performance between our proposed models and the base models. On one hand, the Joint-ISI metric shows similar results across all models for combinations ranging from 2 to 60 subjects’ combinations. This period is considered the optimal performance range for base models such as ICA and cICA. On the other hand, we observe a noticeable decline in the performance of ICA, Non-ICA, cICA, and acICA when applied to high-dimensional data. As the number of subjects increases, particularly in combinations between 80 and 140, we notice a significant improvement in the performance of our cICA-MGGMM and acICA-MGGMM models compared to the base models. This improvement can be attributed to the utilization of prior information incorporated in our proposed models which addresses the limitations of base models in high-dimensional scenarios. Fig. 4.10 illustrates the spatial correlation among all ICA-based models. We observe a consistent trend in the performance of these models, similar to what was observed with the Joint-ISI metric. Specifically, cICA-MGGMM and acICA-MGGMM exhibit higher spatial

Table 4.4: Performance of the proposed cICA-MGGMM and acICA-MGGMM in terms of Joint-ISI with respect to the number of sources for Schizophrenic data.

# of combinations	acICA-MGGMM	cICA-MGGMM	acICA	cICA	Non-ICA	ICA
2	0.035	0.036	0.034	0.036	0.035	0.036
5	0.034	0.035	0.035	0.034	0.033	0.035
10	0.035	0.035	0.036	0.035	0.036	0.034
20	0.036	0.036	0.036	0.034	0.037	0.036
40	0.035	0.036	0.034	0.035	0.036	0.039
60	0.034	0.034	0.033	0.034	0.033	0.042
80	<b>0.351</b>	<b>0.361</b>	0.378	0.385	0.375	0.463
100	<b>0.366</b>	<b>0.372</b>	0.383	0.432	0.443	0.485
120	<b>0.369</b>	<b>0.377</b>	0.389	0.432	0.429	0.521
140	<b>0.374</b>	<b>0.384</b>	0.421	0.450	0.463	0.537

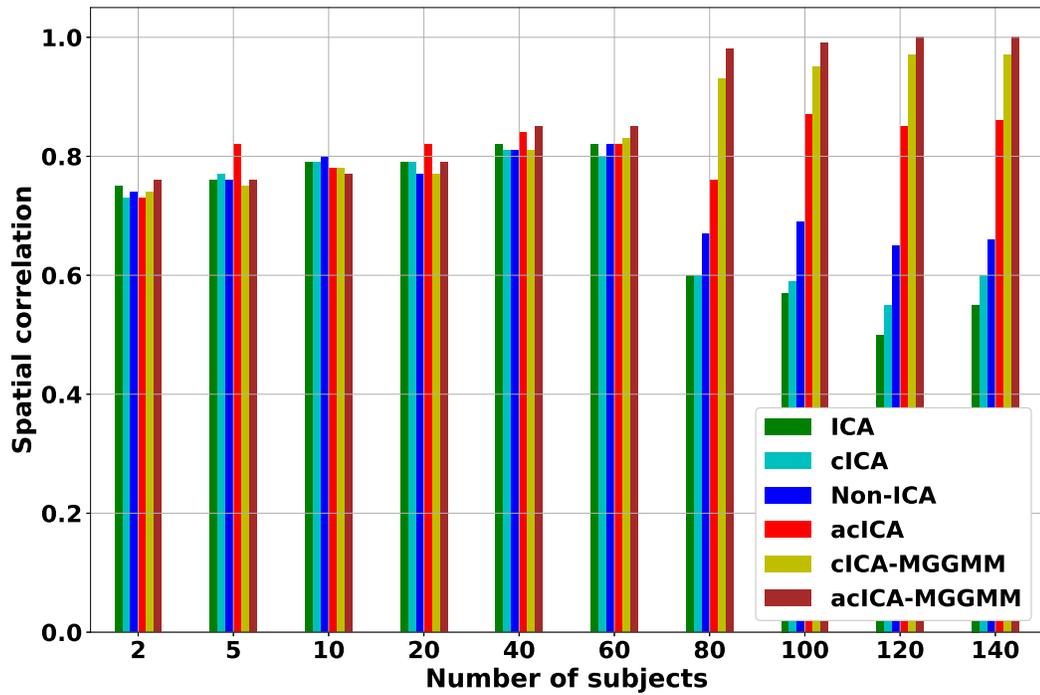


Figure 4.10: Spatial correlation for ICA, Non-ICA, cICA, acICA, cICA-MGGMM, and acICA-MGGMM in terms of number of subjects for Schizophrenic patients detection.

correlation in high-dimensional data compared to the base models. This indicates the enhanced capability of these proposed models in accurately estimating the sources, surpassing the performance of the base models.

#### 4.5.5 Schizophrenia detection using cIVA-MGGMM and acIVA-MGGMM

After preprocessing and partitioning the COBRE data, we utilize our models to discriminate between Schizophrenic patients and healthy subjects. Initially, we employ PCA to reduce the dimensionality of the subject data and extract the most significant components. Subsequently, we configure all parameters for our proposed models following the approach described in the previous subsections. We apply IVA, IVA-G, cIVA, acIVA, cIVA-MGGMM, and acIVA-MGGMM on the reduced data to distinguish Schizophrenic patients from healthy controls.

Table 4.5: Performance of the proposed cIVA-MGGMM and acIVA-MGGMM in terms of Joint-ISI with respect to the number of sources for Schizophrenic data.

# of combinations	acIVA-MGGMM	cIVA-MGGMM	IVA-G	acIVA	cIVA	IVA
2	0.0241	0.0325	0.0344	0.0374	0.0466	0.0425
5	0.0430	0.0241	0.0563	0.0579	0.0567	0.0560
10	0.0354	0.0320	0.0678	0.0654	0.0642	0.0658
20	0.0421	0.0451	0.0765	0.0718	0.0708	0.0726
40	0.0532	0.0582	0.0793	0.0779	0.0764	0.0779
60	0.0517	0.0572	0.0815	0.0797	0.0780	0.0803
80	<b>0.2322</b>	<b>0.3212</b>	0.4936	0.4848	0.4902	0.4875
100	<b>0.2435</b>	<b>0.3464</b>	0.4967	0.4873	0.4915	0.4837
120	<b>0.2653</b>	<b>0.3785</b>	0.4987	0.4932	0.4952	0.4853
140	<b>0.2757</b>	<b>0.3950</b>	0.4989	0.4940	0.4891	0.4908

Table 4.5 presents the Joint-ISI index of several models including IVA, IVA-G, cIVA, acIVA, cIVA-MGGMM, and acIVA-MGGMM. The basic models show unsatisfactory performance in handling high-dimensional data. On the other hand, our proposed models perform better, particularly when dealing with a large number of subjects. During the period when lower combinations were used, our models demonstrate a slight superiority over the base models, which is regarded as the optimal phase for IVA models. In terms of high-dimensional data, there is a considerable enhancement in the performance of our models compared to the base models.

The acIVA-MGGMM is the best-performing model compared to other implemented models. This model can effectively separate the sources of Schizophrenic and healthy individuals. This improvement is attributed to the constrained parameters that we incorporate into our proposed models.

Fig. 4.11 presents a comparison of the performance of our models in terms of spatial correlation. It is evident that the base models exhibit lower performance in high-dimensional data while our

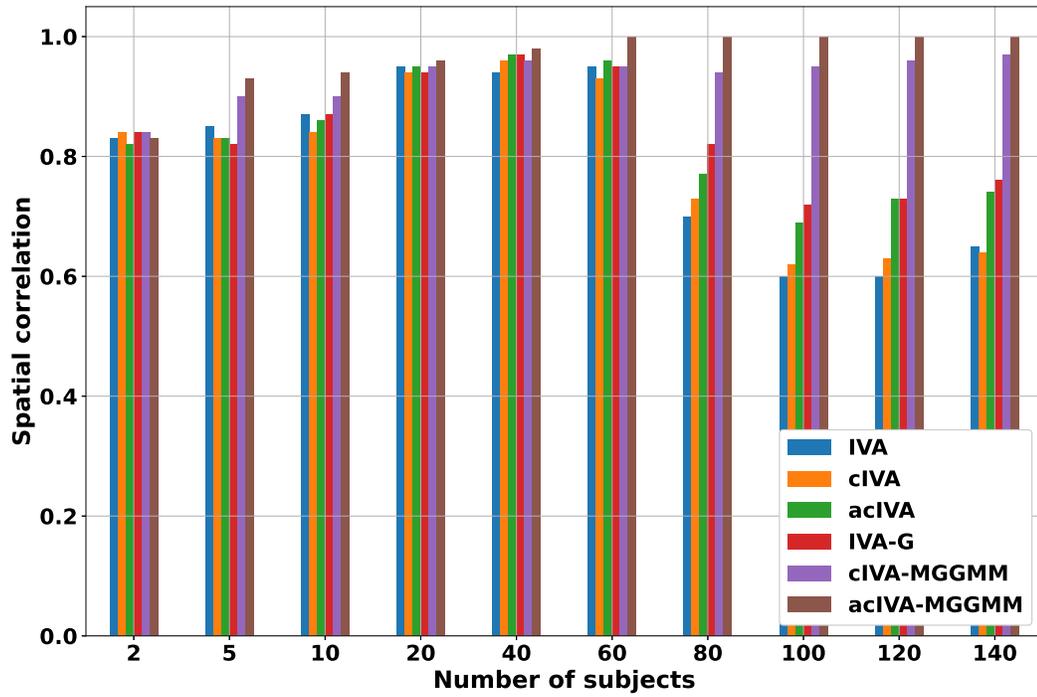


Figure 4.11: Spatial correlation for IVA, IVA-G, cIVA, acIVA, cIVA-MGGMM and acIVA-MGGMM in terms of number of subjects for Schizophrenia detection.

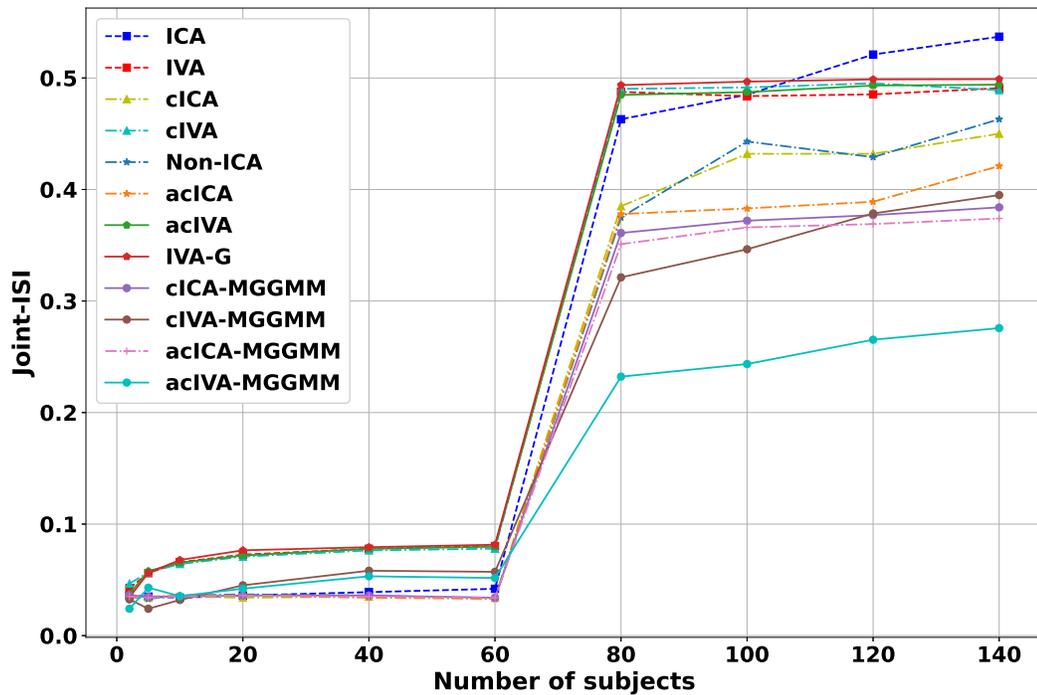


Figure 4.12: The joint-ISI for the four proposed and base models in terms of number of subjects.

models demonstrate superior performance. In particular, the acIVA-MGGMM preserves the high correlation between the reference signal and the estimated sources. This can be attributed to the capability of our models to accurately estimate the underlying sources by leveraging the constraints parameters and prior information.

In summary, we evaluate the effectiveness of our four proposed models in detecting patients with Schizophrenia and the performance results are presented in Fig. 4.12. It is evident that our models exhibit superior performance compared to the base models as demonstrated by the Joint-ISI index. The degradation curve of the base models is also noticeable for high-dimensional data. Among all the implemented models, the acIVA-MGGMM model demonstrated the best performance. This improvement can be attributed to the constraints parameters that were incorporated adaptively into the cost function for this model.

### **Brain networks identification**

The acIVA-MGGMM model enhances the identification of functional networks by incorporating reference signals. Its adaptive characteristics enable more versatile communication between regions that are functionally linked, leading to improved performance. The acIVA-MGGMM components are used to extract the brain regions that are most indicative of the presence of Schizophrenia. Fig. 4.13 demonstrates the localization of AAL areas in both Schizophrenic and healthy subjects. The red areas show the discriminant regions between the two classes whereas the yellow areas represent unaffected brain regions. These regions are consistent with the findings of several studies. For instance, PHG, MTG, amygdala, ITG, thalamus, and hippocampus show lower volume in Schizophrenic brains than in healthy brains. These findings are consistent with previous MRI studies demonstrating atrophy in these regions in patients with Schizophrenia [189, 190, 191]. Research investigating the functional and structural brain networks of individuals with Schizophrenia disease has revealed that cognitive impairments in these patients could be linked to irregularities in the connectivity between various brain regions within the temporal lobe [192, 193]. Previous studies reported that the ITG is affected during the early stage of Schizophrenia [194, 195]. The ITG has a significant impact on verbal fluency, which is a cognitive function that is affected during the early

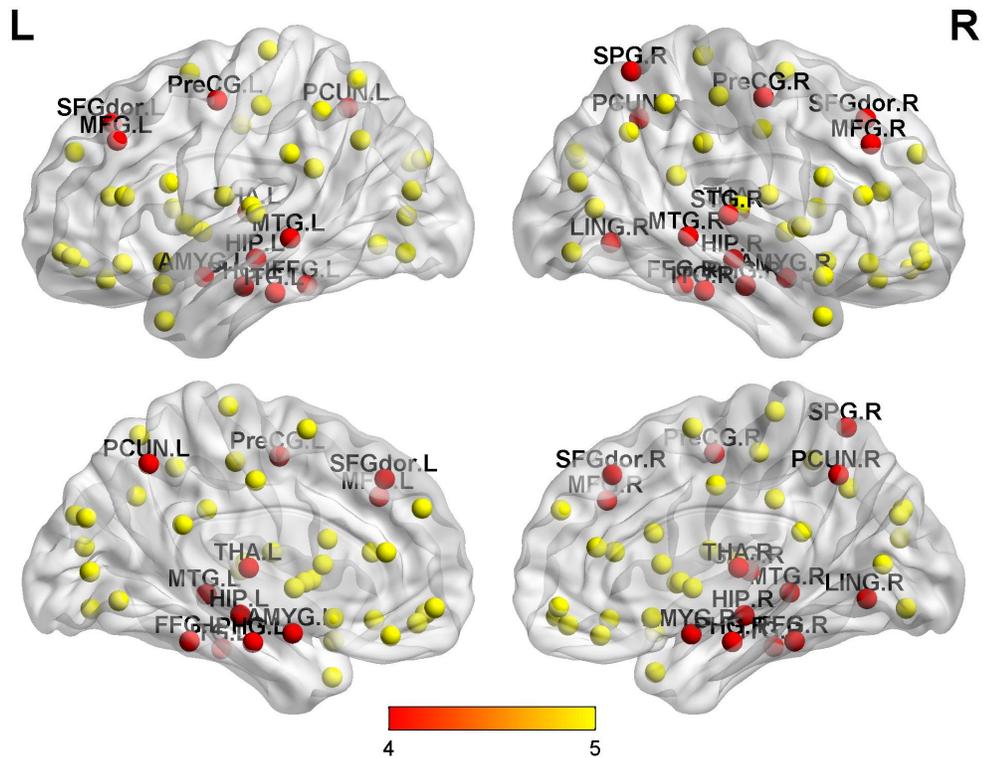


Figure 4.13: Localization of the areas in the AAL atlas. The areas with higher discriminability between healthy and Schizophrenic groups are marked by red circles, while the unaffected areas are represented by yellow circles.

stages of Schizophrenia. Dysfunction in this particular brain region may contribute to certain clinical dysfunctions associated with early-stage Schizophrenia. Our findings indicate that the left ITG is significantly reduced in individuals with Schizophrenia compared to healthy individuals. MFG, fusiform gyrus (FFG), and SFG demonstrate disruption in brain functions for Schizophrenic subjects. These results are in agreement with the defective regions identified in the study conducted by Shi et al. [196], which indicates the involvement of these brain regions in Schizophrenia. Consistent with these studies, our findings suggest that the left lingual gyrus could potentially be impacted during the initial phase of Schizophrenia. Several research studies have indicated notable abnormalities in the LING, the PHG, the PreCG, and the Precuneus (PCUN) among individuals with Schizophrenia [190, 191]. In our work, the right LING, PreCG, the right PCUN, and the right PHG were found to be significantly smaller in Schizophrenia patients than in healthy control subjects.

#### **4.5.6 Functional connectivity in resting state networks of ADHD patients revealed by acICA-MGGMM and acIVA-MGGMM**

ADHD is increasingly recognized as a prevalent factor in neurodevelopmental conditions, along with common psychological and behavioral issues experienced by children [197]. Globally, ADHD affects approximately 5.29% of the population, with a substantial number of adolescents affected, making the management and treatment of ADHD a crucial concern. This condition is frequently linked to learning challenges and behavioral problems [198], which can significantly impact the social and academic performance of those affected. Numerous studies have highlighted the difficulties in distinguishing between normal behavior and the level of ADHD symptoms requiring intervention due to the subjective nature of diagnosis [199]. Therefore, research on the objective diagnosis of ADHD has gained significant importance, making ADHD a prominent focus in the fields of medicine and psychology in recent years.

Prior research extensively employed rs-fMRI to investigate the brains of individuals with ADHD [200]. rs-fMRI has emerged as a prominent research tool, demonstrating significant utility in various domains, including neuroscience, cognitive science, life sciences, and statistical analysis. It has proven to be beneficial in the context of diagnosing and treating ADHD [201]. ICA is a widely used method for statistical analysis of fMRI data [202].

We obtained publicly available fMRI data from the ADHD-200 global competition website<sup>3</sup>. The data were exclusively selected from the New York university (NYU) child study center. Following the guidelines of the health insurance portability and accountability act (HIPAA) and the protocols of the 1000 Functional Connectomes Project, all datasets were anonymized, ensuring the absence of any protected health information. The inclusion criteria were applied to both subjects with ADHD and typically developing (TD) subjects. The original fMRI dataset underwent preprocessing using a publicly available toolbox called DPABI [72]. The preprocessing steps included: (1) discarding the initial 10 volumes to ensure the stability of the BOLD signal; (2) addressing differences in acquisition times between slices through slice timing correction; (3) correcting head motion; (4) normalization, involving registration of the data to the EPI standard template and

---

<sup>3</sup><http://fcon1000.projects.nitrc.org/indi/adhd200/index.html>

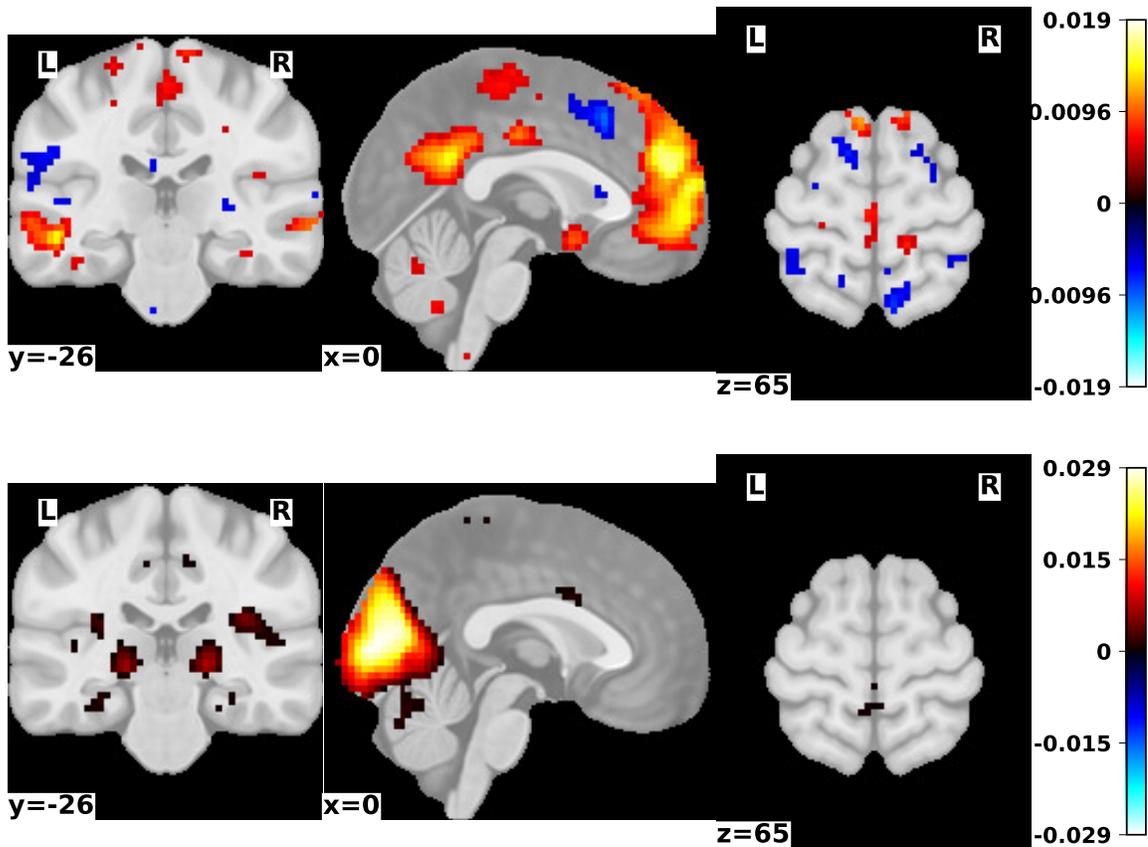


Figure 4.14: The activation regions in the motor area for the non-ADHD subject using acICA-MGGMM and acIVA-MGGMM. The first row depicts the activation areas using acIVA-MGGMM, whereas the second row illustrates the activation regions using acICA-MGGMM.

resampling to  $3.0 \times 3.0 \times 3.0 \text{ mm}^3$ ; and (5) applying spatial smoothing with a 6-mm FWHM Gaussian kernel. Individuals with head movement exceeding 2.0 mm were excluded from the analysis. We applied the proposed acICA-MGGMM and acIVA-MGGMM models to the preprocessed ADHD data to capture functional connectivity in both ADHD patients and control subjects. Initially, we highlight the performance distinction between the two proposed models, acICA-MGGMM and acIVA-MGGMM, in extracting functional connections within the motor region. Fig. 4.14 illustrates this performance difference specifically in terms of extracting activation regions in the motor area of non-ADHD subjects. In this figure, the first row depicts the activation regions obtained using acIVA-MGGMM, while the second row shows the extracted regions using acICA-MGGMM.

We employ the acIVA-MGGMM method to identify activation regions and predict the affected areas in individuals with ADHD, as illustrated in Fig. 4.15. acIVA-MGGMM components exhibit

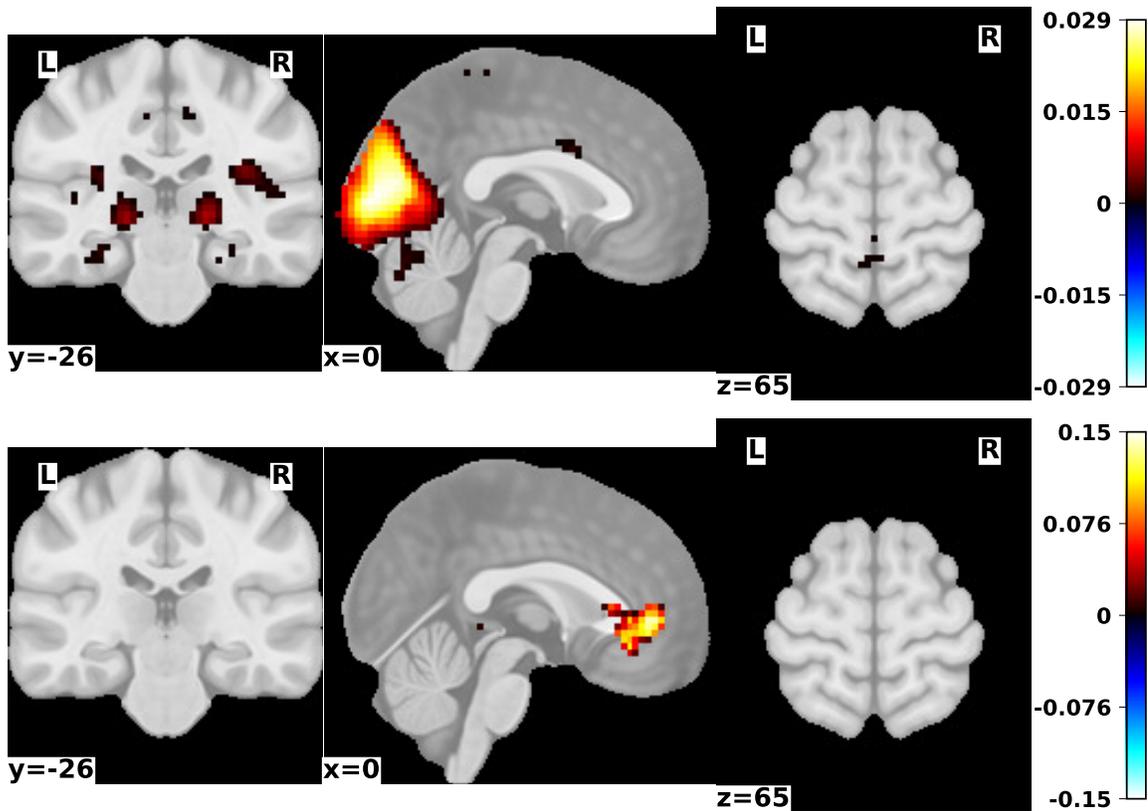


Figure 4.15: Activation regions using acIVA-MGGMM in both control and ADHD. The activation regions in control and ADHD subjects are represented by the first and second rows, respectively.

diminished functional connectivity in the motor area of the ADHD brain, contrasting with the higher functional connectivity observed in normal subjects. We additionally assess the components of acIVA-MGGMM in the visual area for both control subjects and individuals with ADHD. Illustrated in Fig. 4.17, we observe reduced activation intensity in ADHD patients (the bottom row) in contrast to the heightened activation intensity in normal brain regions (the top row).

We evaluate the robustness of our proposed models by constructing a correlation matrix to illustrate the connectivity strength among different brain regions. Fig. 4.16 presents the correlation matrix for both healthy subjects and those with ADHD. In the case of healthy controls, depicted in Fig. 4.16 (a), there is a high correlation strength between the brain regions. In contrast, for ADHD patients, as depicted in Fig. 4.16 (b), the correlation among brain regions is notably lower. Fig. 4.18 depicts the increased connectivity among edges in the healthy brain in contrast to the reduced edge

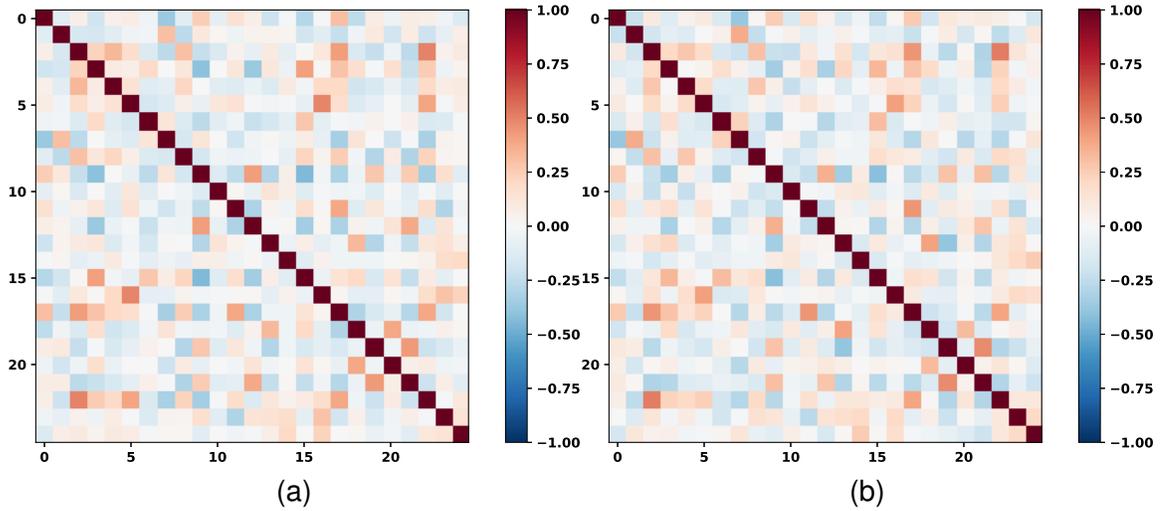


Figure 4.16: Correlation matrix using acIVA-MGGMM for both healthy and ADHD subjects. (a) Control correlation. (b) Patient correlation.

connectivity observed in the ADHD brain.

#### 4.5.7 Epileptic Seizure prediction

Epilepsy is a persistent neurological condition affecting the human brain, marked by Seizures caused by sudden and temporary electrical disruptions in the brain. The study of human brain activity relies significantly on EEG, a crucial tool that furnishes essential information. This study utilizes the Freiburg EEG database 2007 [203] to assess the effectiveness of the proposed method. The database comprises invasive EEG recordings from 21 patients with medically intractable focal Epilepsy. These recordings were obtained at the Epilepsy center of the university hospital of Freiburg and consist of EEG data collected on six channels at a sampling rate of 256 Hz. Each patient’s data includes two datasets: ”ictal” and ”interictal.” The ”ictal” dataset contains files with epileptic Seizures, along with a minimum of 54 min of pre-ictal data. On the other hand, the ”interictal” dataset encompasses approximately 24 hours of EEG recordings without any Seizure activity. Therefore, in this study, we consider the EEG signal smaller than 54 min.

After the preprocessing stage, we employed a 64-point Hamming window for smoothing in both the time and frequency domains. During the feature extraction phase, we derived 13 cepstral coefficients for each window within the EEG dataset. Initially, the input EEG signal was transformed into

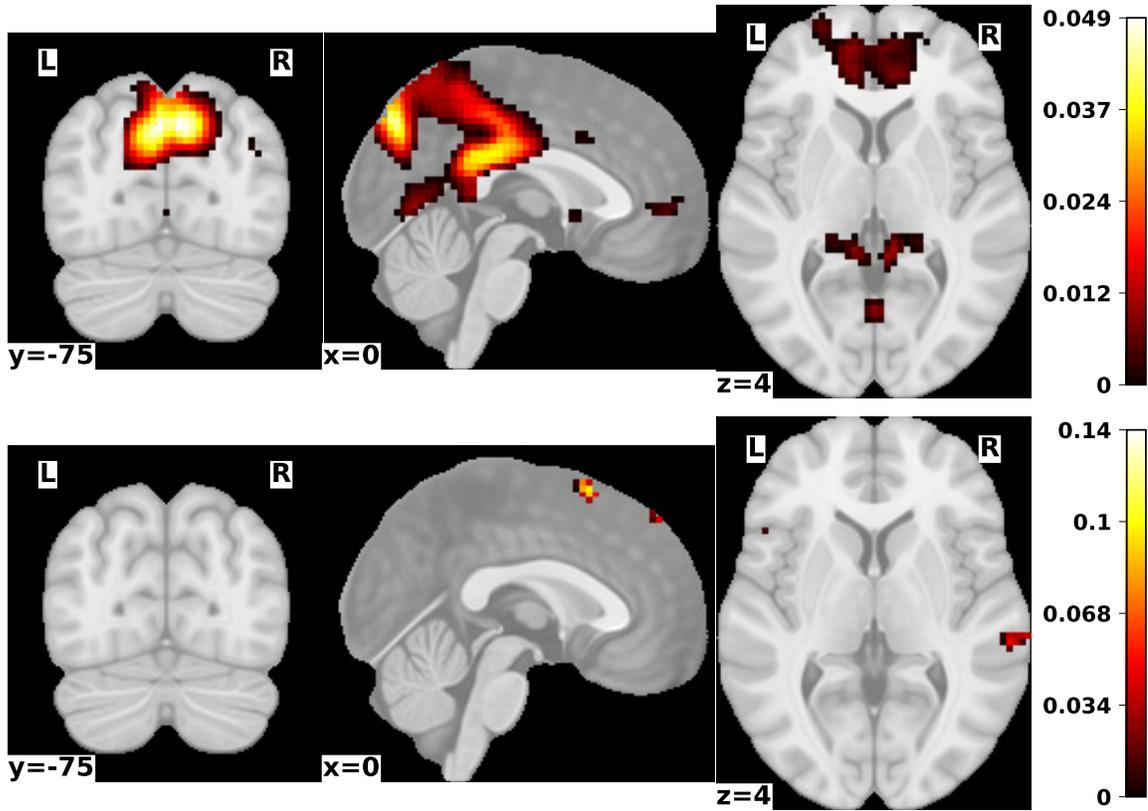


Figure 4.17: The activation regions in the visual area for healthy controls and ADHD patients using the proposed acIVA-MGGMM. The activation regions in control and ADHD subjects are illustrated by the top and bottom rows, respectively.

the frequency domain using the fast Fourier transform. Subsequently, the signal was subjected to a bank of triangular filters to calculate a weighted sum of filter spectral components, approximating a Mel scale as described in [139]. The resulting Mel frequency cepstral coefficients (MFCCs) were obtained by converting the logarithm of the Mel spectrum into the time domain through the discrete cosine transform. We applied the BSS framework to the extracted MFCCs features using models based on ICA and IVA. We utilized the acICA-MGGMM and acIVA-MGGMM to distinguish between Seizure and non-Seizure classes. In Fig. 4.19, the left side illustrates the performance of acICA-MGGMM compared to other ICA-based models in terms of Joint-ISI, while the right side shows the Joint-ISI for acIVA-MGGMM and other IVA-based models. Both acICA-MGGMM and acIVA-MGGMM outperform other models, demonstrating lower Joint-ISI values, indicating reduced overlap between the estimated signal and the reference signal. This enhancement can be

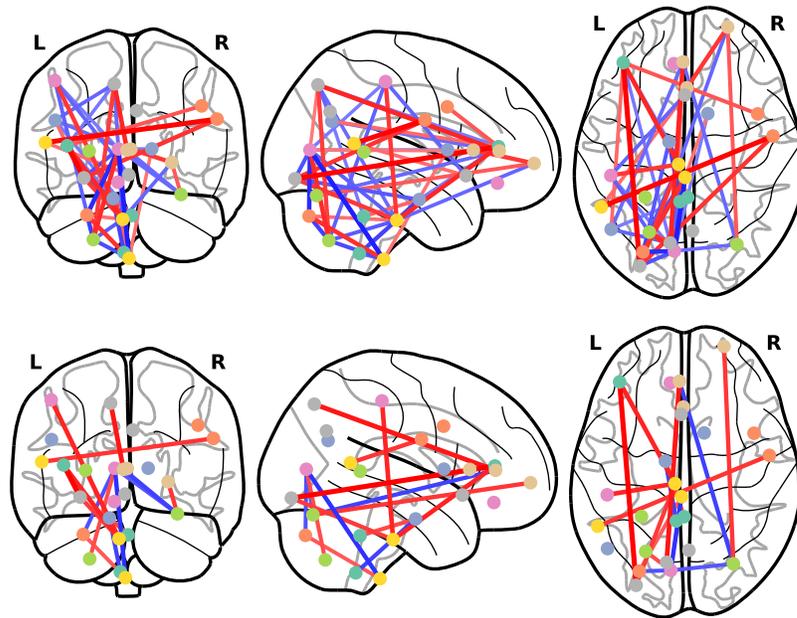


Figure 4.18: ADHD and control subjects connectomes. The first row represents control subject connectomes while the second row indicates ADHD patient connectomes.

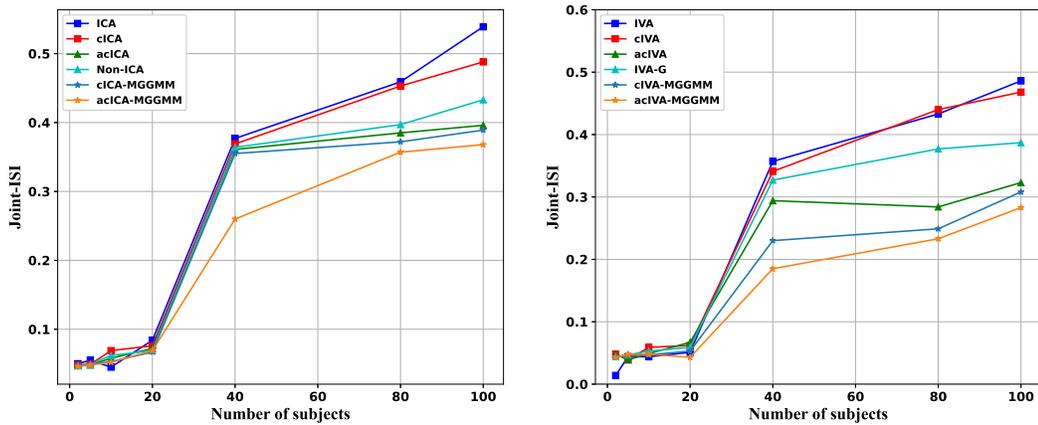


Figure 4.19: Joint-ISI for both acICA-MGGMM and acIVA-MGGMM for the EEG data.

attributed to the incorporation of prior knowledge in an adaptive manner and the consideration of correlated features using the full covariance matrix.

## 4.5.8 Comparison to similar BSS approaches in Schizophrenic detection

### Experiment setup

Using the COBRE dataset, which comprises 70 Schizophrenic patients and 70 healthy controls, we conducted this experiment following the pipeline illustrated in Fig. 4.20. The preprocessed brain volumes were parcellated into 116 regions based on the AAL atlas, and we then extracted the crucial features associated with Schizophrenia, specifically the fractional amplitude of low-frequency fluctuations (fALFF) and regional homogeneity (ReHo). After fusing the fALFF and ReHo features, we applied our BSS methods with the same parameter settings detailed in subsection 4.5.2. The fALFF method, as introduced by Zou et al. [204], analyzes resting-state fMRI data by calculating the power of low-frequency components (0.01-0.08 Hz) in BOLD signals on a voxel-by-voxel basis. Similarly, the ReHo method, described by Zang et al. [205], calculates Kendall's coefficient of concordance among neighboring voxels to assess the similarity of BOLD signals, thereby identifying patterns of local synchronization and variations across different brain regions. We then applied recent BSS techniques to the fused features to extract the most informative ICA and IVA components. Finally, a two-sample t-test was conducted on the extracted ICA and IVA components to identify significant differences between the Schizophrenic patients and healthy controls.

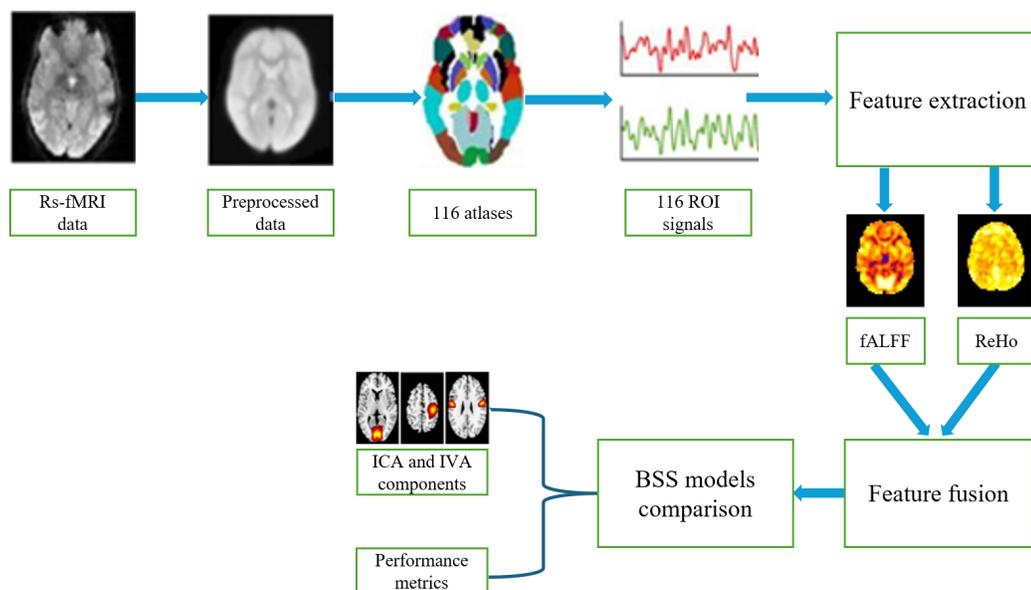


Figure 4.20: Schizophrenia detection using the recent BSS approaches.

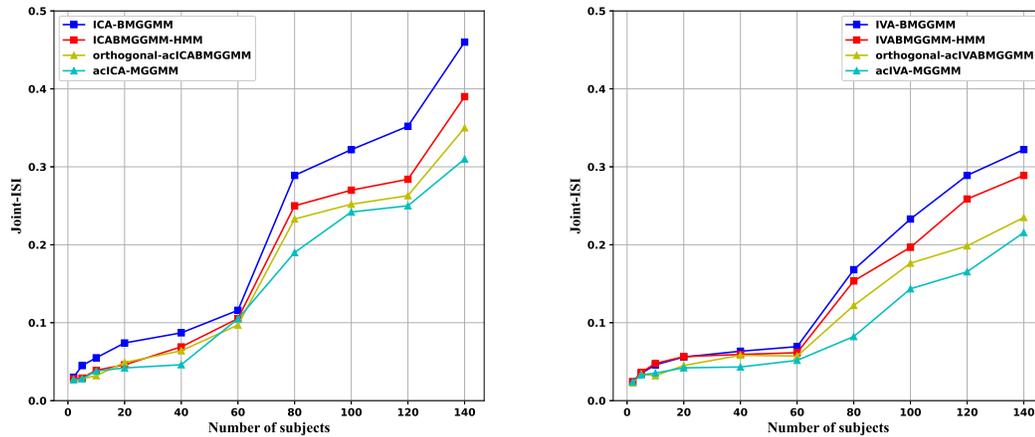


Figure 4.21: Comparison of Schizophrenia detection performance using recent BSS models, evaluated with the Joint-ISI metric. The left figure illustrates results from ICA-based models, while the right figure presents those from IVA-based models.

## Experimental results

To evaluate the robustness and stability of our proposed acICA-MGGMM and acIVA-MGGMM models, we applied them to distinguish Schizophrenic patients from healthy subjects. In this experiment, we compared the performance of our approaches with recent BSS methods, including ICA-BMGGMM [2], IVA-BMGGMM [2], ICABMGGMM-HMM [3], IVABMGGMM-HMM [3], orthogonal acICAMGGMM [206], and orthogonal acIVABMGGMM [5]. We assessed the models' performance using metrics such as Joint-ISI, spatial correlation, and their effectiveness in identifying brain regions affected by Schizophrenia.

We applied both proposed and comparative methods to the fused fALFF and ReHo feature matrix. The extracted ICA and IVA components were then evaluated using Joint-ISI metric and spatial correlation. Fig. 4.21 presents a comparison of the performance of the proposed acICA-MGGMM and acIVA-MGGMM models against other BSS methods, utilizing the Joint-ISI metric to assess the effectiveness in separating Schizophrenic time-courses from those of healthy subjects. Lower Joint-ISI values indicate reduced interference in the estimated IVA components, as demonstrated by the acIVA-MGGMM and orthogonal acIVABMGGMM models in the right panel of Fig. 4.21. While the IVA-BMGGMM and IVABMGGMM-HMM models perform well with a smaller sample size (5 to 60 subjects), their performance deteriorates as the number of subjects increases, particularly between 60 and 140, highlighting their limitations in handling large-scale data. Although

the orthogonal acIVABMGGMM model maintains relatively strong performance even with a larger number of subjects, it slightly under-performs compared to our proposed acIVA-MGGMM. This difference can be attributed to the orthogonal mixing matrix assumption in the orthogonal acIV-ABMGGMM method, which limits its ability to handle more complex scenarios when separating fMRI sources. We attribute this notable improvement to the constrained parameters of our proposed models, acIVA-MGGMM. By leveraging prior knowledge, these constraints effectively guide the models towards more accurate and robust solutions, ensuring better performance and reliability in complex scenarios.

Notably, the proposed acICA-MGGMM outperforms other BSS methods in the quality of separating Schizophrenic signals, as illustrated in the left panel of Fig. 4.21. A similar trend is observed in the ICA-based versions of ICA-BMGGMM, ICABMGGMM-HMM, and orthogonal acICAMGGMM, which also experience a decline in performance when applied to large-scale datasets. This decline can be attributed to several factors. The lack of constrained parameters in ICA-BMGGMM and ICABMGGMM-HMM limits their ability to effectively capture the complex relationships within high-dimensional data, leading to suboptimal performance. Without these constraints, the models struggle to differentiate between Schizophrenic and normal time-courses, resulting in less accurate predictions. Additionally, the orthogonal components in orthogonal acICAMGGMM, while simplifying the model and reducing computational complexity, impose rigid constraints that may not align with the data's intrinsic structure. This misalignment limits the model's ability to fully exploit underlying patterns and correlations, further contributing to performance degradation. In contrast, our proposed acICA-MGGMM models address these issues by incorporating constrained parameters and leveraging prior knowledge. These enhancements enable acICA-MGGMM to manage high-dimensional data effectively, improving both accuracy and robustness. The constraints guide the model towards more meaningful solutions, while the integration of prior knowledge helps discern relevant features from noise, thereby overcoming the limitations faced by the ICA-based models.

Fig. 4.22 illustrates the spatial correlation performance of various BSS approaches in diagnosing Schizophrenia. The left panel compares the performance of ICA-based methods, while the right panel highlights IVA-based models. Across both panels, it is evident that acICA-MGGMM

and acIVA-MGGMM consistently achieve higher spatial correlation as the number of subjects increases, indicating superior ability in preserving spatial structures within the data. In contrast, ICA-BMGGMM, ICABMGGMM-HMM, orthogonal acICAMGGMM and their IVA versions, display lower spatial correlation, especially when the number of subjects exceeds 60, reflecting a potential limitation in capturing the complex spatial relationships in large-scale fMRI data. This can be attributed to the absence of constrained parameters in ICA-BMGGMM and ICABMGGMM-HMM, as well as the orthogonal assumption in both orthogonal acICAMGGMM and orthogonal acIVABMGGMM. The acICA-MGGMM and acIVA-MGGMM models, in particular, demonstrate robust performance across all subject numbers, with spatial correlation nearing 1.0 as the number of subjects approaches 140.

The performance of acICA-MGGMM and acIVA-MGGMM improve steadily as the number of subjects increases, highlighting their scalability. This ability to scale effectively with larger datasets enhances the robustness and generalizability of acICA-MGGMM and acIVA-MGGMM, making them valuable tools in neuroimaging research and clinical applications. The consistent performance improvement with additional subjects demonstrates that acICA-MGGMM and acIVA-MGGMM can leverage extensive datasets to provide precise and meaningful insights. Furthermore, acICA-MGGMM and acIVA-MGGMM offer significant advantages for researchers and clinicians, enabling the extraction of reliable patterns and information from large-scale studies, thereby advancing our understanding and treatment of neurological conditions.

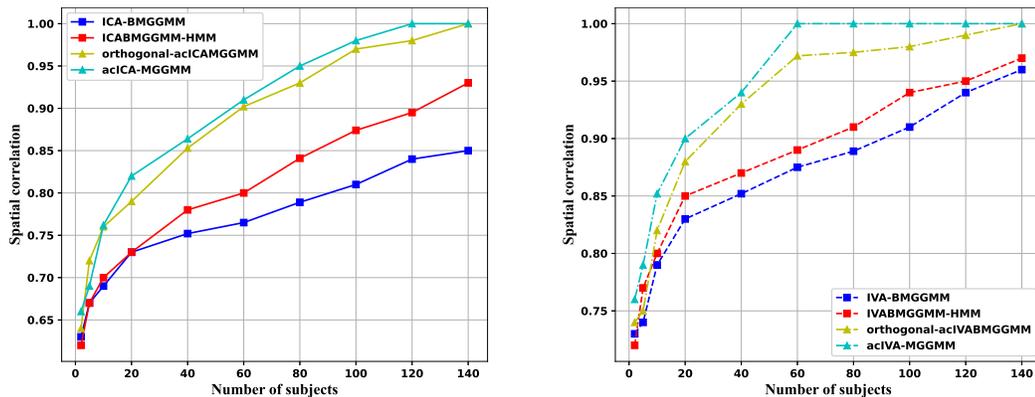


Figure 4.22: Spatial correlation performance of BSS approaches in diagnosing Schizophrenia. The left figure shows the performance of ICA-based approaches, while the right figure highlights results from IVA-based models.

## Brain network identification and statistical analysis

In our previous analysis, we demonstrated that acICA-MGGMM and acIVA-MGGMM are the most effective models among the evaluated BSS methods, as indicated by their superior performance on Joint-ISI and spatial correlation measures. These findings establish the utility of acICA-MGGMM and acIVA-MGGMM components in distinguishing Schizophrenic patients from healthy controls. Fig. 4.23 specifically highlights the most discriminating acIVA-MGGMM components, which effectively differentiate between Schizophrenic and control subjects. Here, a two-sample t-test was employed to create discriminative connectivity maps, which confirmed the differences between healthy controls and individuals with Schizophrenia. To account for multiple comparisons, the FDR correction was applied, resulting in a significance threshold of  $p < 0.05$  ( a minimum  $z > 3.2$  and cluster significance set at  $p < 0.05$ ). This rigorous statistical approach ensures the reliability of the identified neural distinctions between the two groups. The most informative IVA

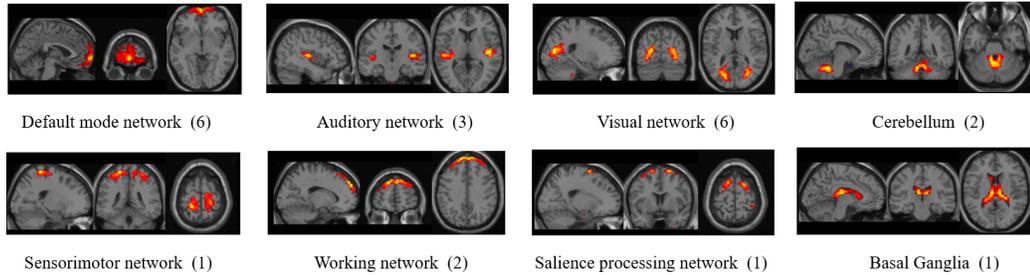


Figure 4.23: The most discriminating IVA components for Schizophrenic patients using proposed acIVA-MGGMM. The number in parentheses demonstrates the number of IVA components for each brain network.

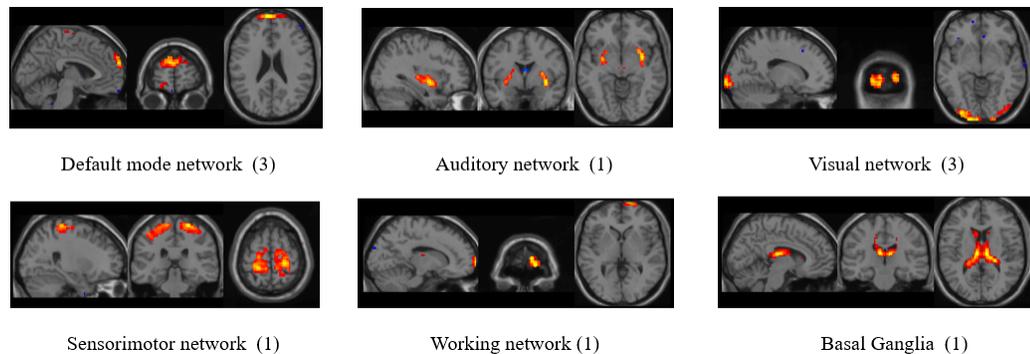


Figure 4.24: The most discriminative ICA components for Schizophrenic patients using acICA-MGGMM.

Table 4.6: Brain networks with their corresponding IVA Components using proposed acIVA-MGGMM model.

Brain network	IVA component	Anatomical description	Abbreviation	Hemisphere	Coordinates (X, Y, Z)	Peak (X, Y, Z)
Default mode network	3	Middle cingulate & paracingulate gyri	MCC	R	8, -48, 36	6, -42, 36
	12	Superior frontal gyrus, medial	SFGmedial	R	8, 69, 14	-12, 72, 14
				L	-9, 69, 14	
	13	Posterior cingulate gyrus	PCC	L	-1, -36, 32	-16, -36, 30
	21	Superior frontal gyrus, medial orbital	PFCventmed	L	-5, 69, -4	-4, 66, -4
				R	8, 70, -4	
50	Superior frontal gyrus, medial orbital	PFCventmed	R	20, 66, -10	20, 66, -10	
63	Calcarine fissure and surrounding cortex	CAL	L	-22, -66, 9	-22, -66, 6	
			R	21, -66, 9		
Auditory network	11	Postcentral gyrus	PoCG	R	66, -12, 16	66, -18, 18
	15	Superior temporal gyrus	STG	R	43, -23, 7	42, -18, 8
	32	Postcentral gyrus	PoCG	L	-61, -12, 19	58, -12, 20
Visual network	16	Lingual gyrus	LING	L	-17, -96, -16	-12, -102, -16
				R	22, -95, -16	
	30	Calcarine fissure and surrounding cortex	CAL	R	5, -58, 12	6, -54, 6
				L	-5, -58, 12	
	35	Calcarine fissure and surrounding cortex	CAL	L	-24, -68, 6	-24, -76, 12
				R	29, -65, 6	
	39	Inferior temporal gyrus	ITG	R	54, -65, -9	54, -72, -16
L				-65, -45, -17		
55	Precuneus	PCUN	L	-4, -78, 50	-4, -78, 50	
			R	7, -78, 50		
57	Cuneus	CUN	R	10, -88, 22	8, -88, 36	
Cerebellum	17	Lobule IV, V of cerebellar hemisphere	CER4.5	L	-6, -54, -5	-4, -54, -6
				R	10, -54, -13	
	47	Lobule IX of vermis	VER9	-	-2, -54, -3	
				-	0, -53, -28	-10, -46, -28
				-	1, -57, -28	
Lobule VIII of vermis	VER8	-	7, -58, -31			
		R	-10, -46, -28			
Sensorimotor network	22	Precuneus	PCUN	L	-16, -40, 57	-22, -40, 68
				R	12, -40, 59	
Working memory network	45	Superior frontal gyrus, dorsolateral	SFG	L	-17, 54, 38	-18, 54, 38
				R	6, 54, 38	
	Middle frontal gyrus	MFG	L	-7, 59, 38		
			R	8, 57, 38		
54	Superior frontal gyrus, medial orbital	PFCventmed	R	42, 54, -4	44, 54, -4	
Salience processing network	51	Superior frontal gyrus, dorsolateral	SFG	R	18, 6, 68	18, 6, 68
				L	-22, -6, 72	
	Supplementary motor area	SMA	R	10, 6, 70		
L			-9, 3, 72			
Basal Ganglia	67	Thalamus	THA	R	10, -18, 20	-10, -22, 14
				L	-12, -15, 20	
		Caudate nucleus	CAU	L	-10, -3, 19	
R	12, -3, 20					

components include DMN, Auditory network (AN), Visual network (VN), Cerebellum, Sensorimotor network (SN), Working memory network (WMN), Salience processing network (SPN), and Basal Ganglia (BG). The DMN is widely recognized for its activation during rest, when individuals engage in self-referential thoughts, contemplation of others, or reflections on the past and future. Due to its involvement in such introspective and cognitive processes, the DMN has been extensively studied across various neurological and psychiatric disorders. The network's primary functional hubs include the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), precuneus, and

angular gyrus (AG), all of which play critical roles in maintaining the DMN's core functions. Understanding the alterations in these regions can provide valuable insights into the pathophysiology of disorders like Schizophrenia, where disruptions in self-referential processing are common.

As shown in Fig. 4.24 and Table 4.6, the DMN exhibits a significant number of IVA components (six), highlighting regions such as the middle cingulate and paracingulate gyri, posterior cingulate gyrus, and superior frontal gyrus. Studies have reported the involvement of these brain regions in self-referential thought and cognitive control, functions often disrupted in Schizophrenia [196, 191, 207]. Similarly, the Visual network also shows a high number of discriminative components (16, 30, 35, 39, 55, 57), particularly in regions such as the calcarine fissure and surrounding cortex, inferior temporal gyrus, lingual gyrus, and precuneus, indicating potential alterations in visual processing pathways in patients. These findings align with the defective regions identified by several studies [194, 189, 191]. In addition to the DMN and visual network, other networks such as the Auditory network, Cerebellum, Sensorimotor network, Working memory network, Salience processing network, and Basal Ganglia display the critical IVA components that further delineate the neural underpinnings of Schizophrenia. For instance, the Salience processing network, which includes the superior frontal gyrus and supplementary motor area, reflects disrupted salience attribution, an indicator of the disorder. The cerebellar regions, specifically the lobules of vermis and cerebellar hemispheres, also play a crucial role in the coordination of motor and cognitive functions, which are frequently impaired in Schizophrenia. These brain regions are in agreement with finding in previous studies [194, 190, 189, 207].

Previous researches [207, 208, 209] have shown that the lingual gyrus, calcarine fissure and surrounding cortex, inferior temporal gyrus, precuneus, and cuneus are the main affected areas of Schizophrenia. These findings are consistent with the findings of this study as shown in Visual network. The Auditory network, which includes regions such as the postcentral gyrus and superior temporal gyrus, reflects auditory processing and perception disorders that are common in Schizophrenia patients. The Working memory network, which contains substantial components in the superior and middle frontal gyrus, demonstrates the impairments in working memory and executive function that are frequent in Schizophrenia. Numerous studies have reported significant abnormalities in postcentral gyrus, superior temporal gyrus, superior frontal gyrus, and middle frontal

gyrus [190, 194, 189, 208]. The Sensorimotor network, represented by components in the pre-cuneus, shows potential issues with motor coordination and sensory processing. Such deficits can limit a patient’s ability to successfully interact with their surroundings. The Basal Ganglia, which contains the thalamus and caudate nucleus, stresses the significance of these subcortical structures in motor control and cognitive abilities that are typically impaired in Schizophrenia. Recent studies [191, 194, 189, 207, 210] have similarly identified the thalamus and caudate regions as popular regions affected in Schizophrenia. For further analysis, we utilized the acICA-MGGMM components

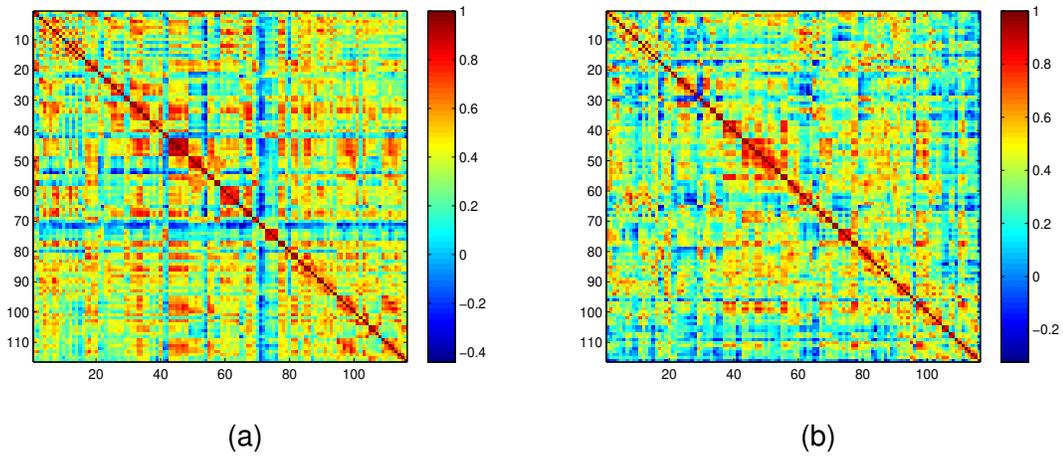


Figure 4.25: Correlation matrix for healthy subjects (a) and Schizophrenic patients (b) using the acIVA-MGGMM components.

to extract the brain networks affected by Schizophrenia, as illustrated in Fig. 4.24. The identified networks include DMN, AN, VN, SN, WMN, and Basal Ganglia. These networks align with those extracted using the proposed acIVA-MGGMM method. Notably, our analysis revealed lower activation levels in the DMN and VN when using the acICA-MGGMM method compared to those extracted using the acIVA-MGGMM method. This difference can be attributed to the lower number of ICA components used in the acICA-MGGMM method for network extraction. Consequently, this underscores the enhanced capability of the acIVA-MGGMM method to detect variability in the data, demonstrating its superior sensitivity in identifying the affected brain networks in Schizophrenic patients compared to the acICA-MGGMM method. Additionally, Fig. 4.23 identifies components in the Cerebellum and Salience processing network, which are absent in Fig. 4.24. This suggests

that the acIVA-MGGMM method may capture a more comprehensive set of discriminating features across a wider range of brain networks compared to the acICA-MGGMM method.

These differences highlight the variability in the sensitivity of the two proposed methods in identifying critical brain network components associated with Schizophrenia. The identification of specific IVA components across multiple networks underscores the robustness of the acIVA-MGGMM model in capturing the complex and multifaceted nature of Schizophrenia. This provides valuable insights for potential diagnostic markers and therapeutic targets. By detailing the specific brain regions and networks involved, this approach offers a comprehensive understanding of the neural disruptions associated with Schizophrenia, paving the way for improved diagnostic and therapeutic strategies.

Using the acIVA-MGGMM components, we illustrate the differences in brain connectivity between Schizophrenic patients and healthy subjects by examining the correlation matrices of brain regions. Fig. 4.25 highlights significant differences in brain connectivity between the two groups. The correlation matrix for healthy individuals demonstrates a well-organized and efficient network, characterized by strong intra-regional connectivity along the diagonal and diverse inter-regional interactions. In contrast, the correlation matrix for Schizophrenic patients shows more dispersed and less structured connectivity patterns. This indicates potential disruptions in Schizophrenic patients brain network organization.

## **4.6 More Experiments for cICA-MGGMM and acICA-MGGMM**

To evaluate the robustness and scalability of our proposed models, we extend the application of cICA-MGGMM and acICA-MGGMM to additional domains, including EEG and speech datasets. These extensions allow us to test the models under varying conditions and data structures, providing valuable insights into their adaptability and performance across diverse types of data.

In this section, we consider two different datasets, speech and EEG databases to validate the proposed model. For the speech experiment, the Romanian read-speech corpus (RoDigit) is used to validate the model. The proposed model and the base models are employed to separate the mixed speech signals. For the EEG experiment, the proposed model is executed to distinguish Seizure

patients from healthy subjects. The conducted models are evaluated using variations of separation metrics, including SDR, SIR, SAR, PESQ, and ISI.

#### 4.6.1 Speech signal separation

We introduce an efficient BSS method called acICA-MGGMM to separate the mixed speech signals. The proposed technique is evaluated using random signals obtained from the RoDigits dataset which is discussed in detail in subsection 2.6.3.

Within the BSS framework, we conducted four experiments aimed at estimating two, three, four, and five speech sources. We obtained a linear combination of these speech sources from the database and applied three models (ICA, ICAMGGMM, and acICA-MGGMM) to retrieve them. After successfully recovering the sources, we conducted an objective analysis including SDR, SIR, SAR, SNR, and PESQ to evaluate the quality of the separated speech signals and assess the suitability of ICA mixture models in the context of BSS. The experiments were repeated ten times, and objective metrics were averaged across ten separate runs using various linear mixtures of 2, 3, 4, and 5 original speech signals. Table 4.7 illustrates the separation indices obtained after recovering

Table 4.7: Objective metrics for separation of 2 and 3 speech sources using acICA-MGGMM and compared models

Model	BSS of 2 speech sources			BSS of 3 speech sources		
	SDR	SIR	SAR	SDR	SIR	SAR
<b>acICA-MGGMM</b>	<b>25.32</b>	<b>25.43</b>	<b>55.97</b>	<b>19.43</b>	<b>19.32</b>	<b>52.18</b>
ICAMGGMM	21.31	22.43	49.53	8.42	8.65	46.74
ICA	11.52	13.32	36.33	8.13	8.15	40.19

Table 4.8: Objective metrics for separation of 4 and 5 speech sources using acICA-MGGMM and compared models.

Model	BSS of 4 speech sources			BSS of 5 speech sources		
	SDR	SIR	SAR	SDR	SIR	SAR
<b>acICA-MGGMM</b>	<b>17.32</b>	<b>17.52</b>	<b>49.34</b>	<b>17.54</b>	<b>17.65</b>	<b>49.32</b>
ICAMGGMM	7.35	6.37	45.03	6.31	6.48	44.58
ICA	6.62	6.34	41.21	6.43	6.32	42.55

speech source signals for the proposed model and other models. We use these models to separate

mixture speech signals into their corresponding speech sources using different mixtures of 2, 3, 4, and 5. For 2 and 3 mixtures, acICA-MGGMM slightly outperforms the other compared models for most of the separation metrics. However, as the number of mixtures increased, the performance of the base models degraded. In this range with 4 mixture, our proposed model significantly outperforms ICA and ICAMGGMM models using SDR (17.32 dB), SIR (17.52 dB), and SAR (49.34 dB), as shown in Table 4.8. As the complexity of the mixture increases, our model maintains a consistent level of performance. This trend is evident across 5 different mixtures, with a notable improvement in our model's separation efficiency, as indicated by the SDR of acICA-MGGMM (17.65 dB) compared to the SDR of ICA (6.32 dB). This indicates the effectiveness of acICA-MGGMM in separating speech-mixed signals into their original sources. It is worth noting that the adaptive separation approach and the consideration of prior information in the acICA-MGGMM cost function make this model more flexible in separating this level of mixtures.

To confirm the efficacy of our proposed model, we evaluate its separation performance using the ISI metric, a measure applied consistently across all models employed in this experiment. As illustrated in Fig. 4.26, it's evident that the performance curve of acICA-MGGMM demonstrates greater stability when compared to the other models. This enhancement becomes notably pronounced when dealing with mixtures of more than 3 speech mixtures signals. This demonstrates the valuable role of incorporating prior information as a valuable reference for achieving superior solutions in such complex scenarios.

To evaluate the effectiveness of our proposed model, we use the PESQ index to highlight the quality of separation. Fig. 4.27 shows that the separation quality of our proposed model is significantly better than the compared models. In cases where the number of mixtures is low, our model slightly outperforms ICA and ICAMGGMM in terms of separation quality. However, when the number of mixtures increases to 3, 4, and 5 the separation quality of the base models is poor compared to the high separation quality of our acICA-MGGMM. This improvement in terms of estimated quality can be attributed to the consideration of prior knowledge and second and higher-order statistics in the acICA-MGGMM cost function. Moreover, the acICA-MGGMM model takes into account the inter-correlated features in the data by incorporating the complete covariance matrix. To highlight the effectiveness of our proposed model, we employ subjective metrics as performance

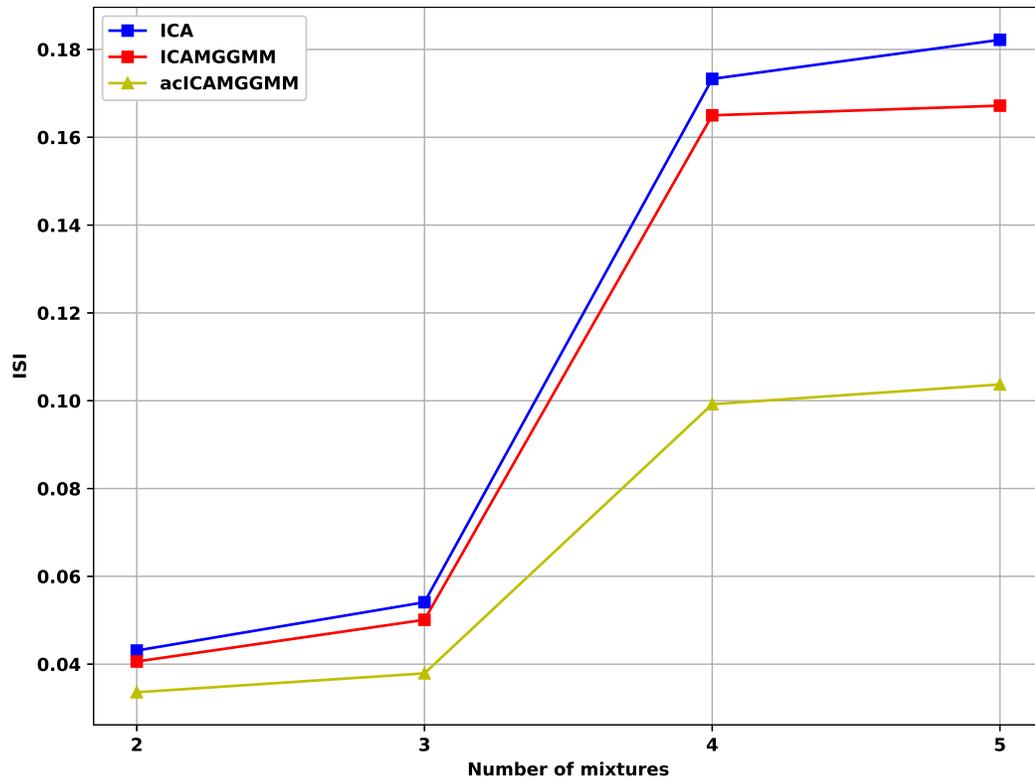


Figure 4.26: Performance of acICA-MGGMM and base models using ISI metric with a variation of mixtures.

indicators to assess the quality of speech separation. The subjective evaluation encompasses both the speech signals before and after the source separation. Fig. 4.28 visually presents the speech signals in their original state, after mixing, and subsequent to separation using acICA-MGGMM. It is evident that our proposed model achieves a high-quality separation in the case of two mixed signals. This accomplishment expands the potential of applying acICA-MGGMM in diverse applications, including tasks like fetal ECG separation.

#### 4.6.2 EEG signal separation

The proposed model's validation uses the EEG dataset detailed in [138] and in subsection 3.3.4. Table 4.9 demonstrates the separation metrics of Seizure EEG signals from healthy ones using various models (acICA-MGGMM, ICAMGGMM, and ICA). These models are evaluated across different subject combinations to assess their performance with high-dimensional data. These findings indicate that acICA-MGGMM outperforms the other models in terms of SDR and SIR metrics.

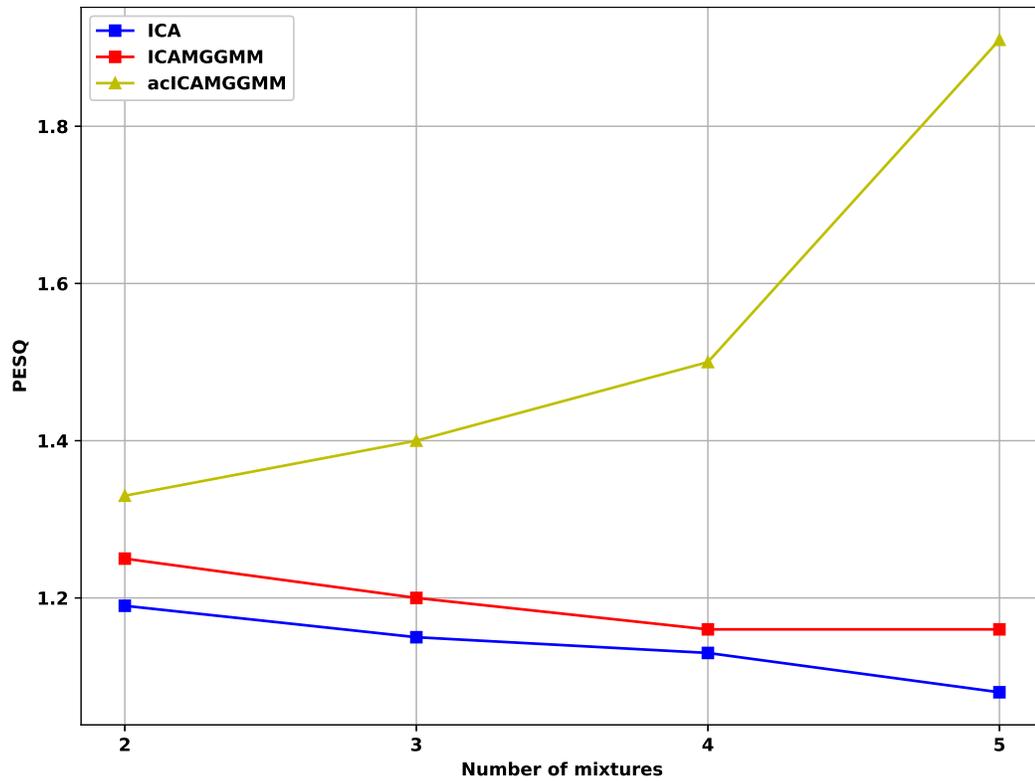


Figure 4.27: PESQ measure for acICA-MGGMM and other models with different mixtures.

Table 4.9: SDR and SIR measures for separation of EEG signals using acICA-MGGMM and compared models.

Model	SDR with different combination					SIR with different combination				
	2	5	10	20	40	2	5	10	20	40
<b>acICA-MGGMM</b>	<b>41.30</b>	<b>40.61</b>	<b>38.62</b>	<b>35.28</b>	<b>33.25</b>	<b>38.71</b>	<b>37.42</b>	<b>35.65</b>	<b>32.58</b>	<b>30.41</b>
ICAMGGMM	40.31	34.76	25.54	20.23	17.53	38.64	31.41	24.04	20.09	15.91
ICA	29.34	23.53	20.54	18.63	15.83	28.68	22.83	18.53	15.60	12.72

While this improvement is most prominent in smaller combinations, such as 2, it is not statistically significant. This implies that ICA’s performance is at its peak within this range. However, when dealing with high-dimensional data involving subject combinations ranging from 5 to 40, employing acICA-MGGMM leads to a substantial improvement in performance compared to the relatively poor performance of the base models. To emphasize these results, as described in Table 4.9, it is shown that the acICA-MGGMM outperforms other models. Specifically, for the 40 mixtures, our model demonstrates its effectiveness with a high SDR metric of 33.25, while ICAMGGMM and ICA exhibit poorer performance with SDR values of 17.53 and 15.83, respectively. We also observe

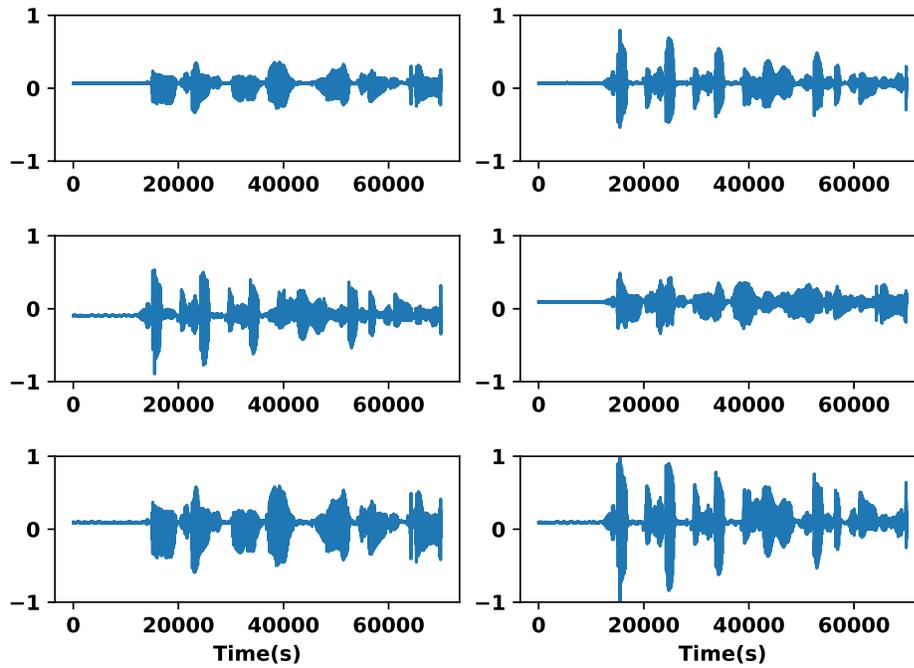


Figure 4.28: Speech signal separation using acICA-MGGMM. The first row represents the original speech signal. The second row illustrates the mixed speech signals. The third row demonstrates the speech signal estimation via acICA-MGGMM.

a similar pattern in our model’s performance when separating more complex cases, as indicated by the SIR index. This suggests that incorporating prior information and utilizing an adaptive approach for EEG signal separation is advantageous in such scenarios. To evaluate the performance of the models, we used the ISI metric to assess their effectiveness in separating the sources. Fig. 4.29 illustrates how the three models (acICA-MGGMM, ICAMGGMM, and ICA) perform in terms of ISI as the number of subjects increases, thereby evaluating their efficacy in handling high-dimensional data. These results indicate that the acICA-MGGMM model demonstrates greater performance stability compared to other models. When utilizing the acICA-MGGMM model, the level of interference in the separated EEG signal is noticeably reduced compared to the ICA and ICAMGGMM models. This superior quality of separation is particularly notable in high-dimensional scenarios involving 20 and 40 subjects. This improvement can be attributed to the enhanced cost function of the acICA-MGGMM model, which can effectively account for correlated features within the multivariate data. Furthermore, the acICA-MGGMM approach ensures that the solution is guided by appropriate prior knowledge, instead of imposing incorrect constraints. Spatial correlation serves as

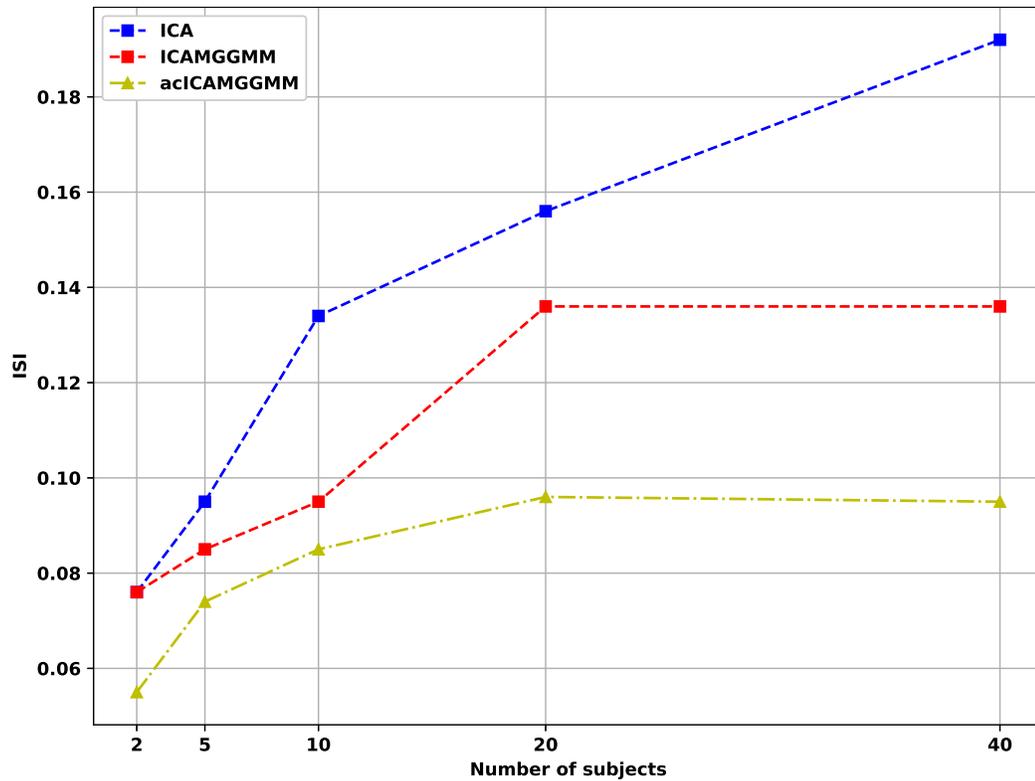


Figure 4.29: Comparison of the performance of acICA-MGGMM and the basic models in relation to the ISI metric as a function of the number of subjects.

a key indicator for assessing the performance of the applied models. The variations in the separation performance of the three BSS algorithms applied in the EEG separation experiment are depicted in Fig. 4.30. The outcomes reveal that, especially in cases involving 2 and 5 subjects, the spatial correlation results for the conducted models are quite similar, notably for acICA-MGGMM and ICAMGGMM. This similarity arises because the base models can effectively account for feature correlations in scenarios where there are fewer mixed EEG sources. However, when dealing with high-dimensional data comprising 10, 20, and 40 subjects, acICA-MGGMM demonstrates its superior ability to distinguish Seizure patients from healthy subjects compared to other models. This is because the proposed model takes into consideration correlated features within the multivariate data by utilizing the full covariance matrix. This experiment highlights acICA-MGGMM's capability to separate EEG signals at high-dimensional levels, leveraging its enhanced cost function to select the most appropriate prior information during the estimation process.

To emphasize the effectiveness of Seizure EEG separation, we subject our model to a visual

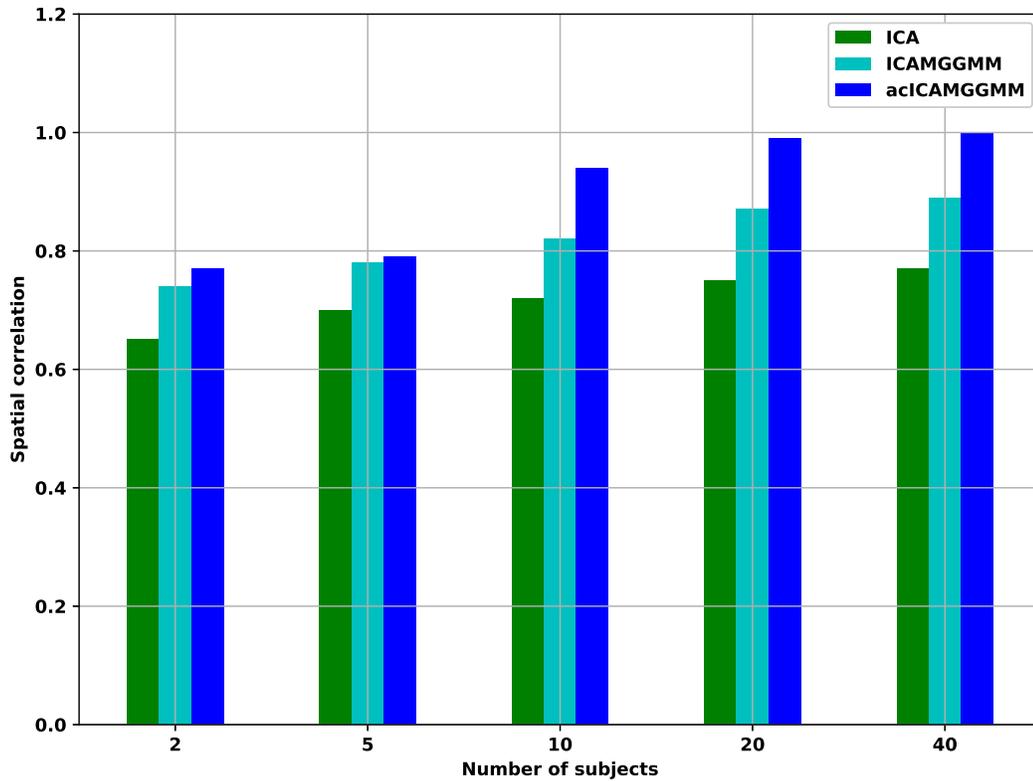


Figure 4.30: Evaluation of acICA-MGGMM’s performance in relation to spatial correlation as it varies with the number of subjects.

assessment. Fig. 4.31 presents the separated Seizure and non-Seizure EEG signals achieved with acICA-MGGMM. This visual representation highlights the proficiency of acICA-MGGMM in separating EEG signals. This accomplishment can be attributed to the incorporation of prior information and the use of the full covariance matrix, which accounts for correlated features, especially in multivariate data like the EEG Seizure database.

## 4.7 Conclusion

In this work, we proposed the cICA-MGGMM to separate the mixing sources. This model was able to separate the sources using the full covariance matrix. We also proposed the acICA-MGGMM to relax the ICA independence assumption by detecting the correlated features within the data. Moreover, we introduced the constrained IVA-MGGMM to be applied over the multivariate data. We also proposed the acIVA-MGGMM to improve IVA performance in the high-dimensional

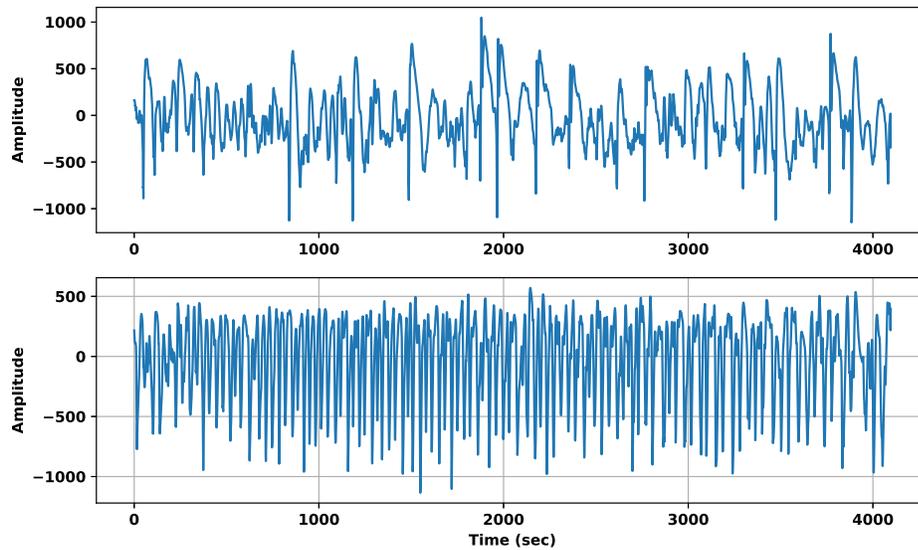


Figure 4.31: The estimated EEG signals using acICA-MGGMM. The first row represents the normal estimated EEG signal and the second row indicates the Seizure one.

data using the prior information. We validated the proposed model using simulation data, resting state fMRI data, EEG data, and ADHD data. For that, we used the ADNI Alzheimer’s data to test our models’ capability to distinguish Alzheimer’s patients from healthy subjects. We also employed our models on the Schizophrenic data to differentiate the Schizophrenic patients from the healthy controls. Using the adaptive constrained IVA-MGGMM, we additionally identified the affected brain regions associated with both Alzheimer’s disease and Schizophrenia. By employing Joint-ISI and spatial correlation metrics, the conducted experiment demonstrated the effectiveness of our proposed models in distinguishing between the various classes in comparison to the base models. Notably, the acIVA-MGGMM model exhibited its capability in identifying the classes within the given experiment, outperforming both the other proposed models and the base models.

## **Chapter 5**

# **Novel Approach for ECG Separation**

## **Using Adaptive Constrained**

## **ICABMGGMM and IVABMGGMM**

An electrocardiogram records the heart's electrical activity, providing essential insights into cardiac health. Traditional visual analysis of ECGs is time-consuming and limits continuous monitoring, highlighting the need for automated signal processing. This Chapter introduces a new blind source separation method, cICABMGGMM, combining a bounded multivariate generalized Gaussian mixture model with constrained independent component analysis, addressing the typical independence assumption of ICA. We further propose the adaptive version, acICABMGGMM, which dynamically adjusts the association between estimated and reference sources using prior information without enforcing incorrect constraints. Moreover, we introduce the constrained independent vector analysis integrated with the bounded multivariate generalized Gaussian mixture model (cIVABMGGMM) to tackle the limitations of ICA when applied to multivariate data, accompanied by its adaptive version, the acIVABMGGMM, aimed at alleviating associated constraints. The acIVABMGGMM employs a full covariance matrix that considers feature correlations, effectively addressing the challenges posed by ICA and IVA models when analyzing multivariate data. The innovative

acIVABMGGMM framework merges the adaptability inherent in data-driven methods with the capability to manage noise and other artifacts often encountered in model-based approaches. This technique effectively employs prior knowledge to guide the solution, avoiding the imposition of inaccurate constraints. To overcome these challenges, our set of four constrained methods incorporates prior source information into the ICA and IVA model, effectively mitigating its limitations in data with a high number of sources. We assess the efficacy of our proposed models through three distinct ECG separation experiments, which include heartbeat separation, fetal ECG extraction, and arrhythmia detection. Notably, the performance metrics demonstrate the superiority of our models over the baseline approaches in the conducted experiments.

## 5.1 Introduction

Constrained independent component analysis (cICA) [150, 211, 212] is a variation of ICA that adds additional constraints to the separation process. The cICA imposes specific constraints that guide the separation process, in contrast to the classic ICA, which tries to blindly separate source signals based on the assumption of statistical independence. These constraints can be generated from a variety of sources, such as prior knowledge of the properties of the source signals or the application area. The cICA aims to increase the accuracy and reliability of the separation findings by integrating these constraints. In recent years, the cICA method has been successfully applied to a variety of applications [213, 206].

In order to address the challenges associated with unbounded mixture models, a previous study introduced the ICA bounded generalized Gaussian mixture model. The ICABGGMM has demonstrated its effectiveness in various applications due to its ability to adapt to different data distributions [37, 22, 38, 214, 3]. However, the ICABGGMM has limitations in handling correlated features because it assumes a diagonal covariance matrix. In a previous works [2], the investigation of statistical independence and sparsity was conducted to integrate the bounded multivariate generalized Gaussian mixture model (BMGGMM) into the ICA model. The BMGGMM was chosen due to its ability to effectively capture the inherent boundaries of real-world data, which often exhibits natural constraints. By imposing these bounds within the mixture model, the BMGGMM offers

a more accurate and flexible representation of the data, enhancing the separation performance in ICA applications. This integration allows for improved modeling of sparse sources under statistical independence, which is crucial in scenarios where traditional Gaussian assumptions may fail to adequately describe the observed data distributions.

To improve the clustering capability of the ICABMGGMM, we introduce new approaches to address the issues related to unbounded models and constraint parameters. For that, we introduce a method called constrained ICABMGGMM (cICABMGGMM), which combines ICABMGGMM with a constrained model. The goal of cICABMGGMM is to provide a structured and flexible approach for incorporating additional assumptions and prior information into the cost function, whenever available.

The cICABMGGMM method relies on user-defined parameters, but selecting the appropriate constraint parameter can be challenging due to the frequent lack of accurate prior information. To address this issue, we propose an adaptive version of cICABMGGMM (acICABMGGMM), which enables us to control the association between estimated and reference sources adaptively. The acICBMGGMM method ensures that prior information is utilized to guide the solution without imposing inaccurate constraints. Overall, our approach offers a way to enhance the performance and applicability of ICA by incorporating additional assumptions and prior information. The new acICABMGGMM model further improves the flexibility and adaptability of the method by allowing for dynamic control of the association between estimated and reference sources.

The new framework called constrained IVA (cIVA) is capable of integrating any type of previous knowledge regarding the sources or the mixing vectors [215, 216, 217, 218]. Unlike traditional methods that rely on orthogonal demixing matrices, cIVA does not have this requirement. It combines the advantages of data-driven techniques, such as flexibility, with the strengths of model-based methods, such as resilience to noise and other imperfections. Additionally, it effectively utilizes dependencies among datasets, thus making it a valuable tool for separating sources. To address the challenges associated with applying unbounded mixture models to bounded data, we propose a new extension to the multivariate generalized Gaussian mixture (MGGMM) called the constrained bounded MGGMM model based on IVA (cIVABMGGMM). The cIVABMGGMM takes into consideration correlated features by incorporating a full covariance matrix and offering the flexibility

to accurately model a wide range of non-Gaussian bounded support data shapes.

cIVABMGGMM relies on a user-defined constraint parameter to regulate the level of agreement between the reference signal and the estimated component. The effective performance of cIVABMGGMM depends heavily on the accurate selection of prior information and the appropriate constraint parameter set by the user. When the prior information is inaccurate, a lower constraint parameter should be employed to avoid enforcing the inaccurate prior on the decomposition [101, 219, 159]. Conversely, when the prior information is accurate, a higher constraint parameter should be used to prevent the components from being influenced by noise and artifacts. However, in practical scenarios, determining the optimal constraint parameter poses challenges, as it is difficult to ascertain the accuracy of the prior information. This difficulty is further compounded when dealing with multiple signals and their respective prior information that needs to be incorporated into the process.

To address these limitations, we also introduce adaptive cIVABMGGMM (acIVABMGGMM) to enable adaptive control over the association between estimation patterns and reference signals. It is a method that makes it easier to incorporate multiple reference signals into the IVA framework. This formulation allows for the incorporation of various types of prior information related to the sources or mixing vectors. The new acIVABMGGMM framework combines the flexibility of data-driven methods with the ability to handle noise and other artifacts commonly found in model-based methods. The acIVABMGGMM technique uses prior knowledge to guide the solution effectively instead of imposing inaccurate constraints. Additionally, it leverages the interdependence between datasets, making it a powerful tool for separating sources effectively. Furthermore, the acIVABMGGMM assumes second-order correlation within bounded support component vectors, which provides more flexibility for extracting knowledge from the data.

To detect abnormal ECG signals, the proposed models will be applied to several clustering applications, including heartbeat separation, arrhythmia detection, and fetal ECG separation databases, to gain insight into the analysis of these models. We employ our proposed models to distinguish between the normal ECG signals and the abnormal ones. The performance of our proposed model is evaluated using the same applications such as heartbeat signal separation, fetal separation, and arrhythmia detection. Overall, our models have demonstrated potential usefulness in diagnosing

various ECG applications. The contributions of this study are summarized as follows:

- We propose the cICABMGGMM model to improve the estimation of underlying sources. This model has some limitations with the user-defined constraint parameters.
- We propose acICABMGGMM as a solution to overcome these limitations. The acICABMGGMM model can adaptively control the relationship between the reference signal and the estimated one without imposing inaccurate constraints. This approach allows for the modeling of classes with non-Gaussian structures and the identification of statistically significant structures in the data.
- We introduce the cIVABMGGMM to be more adaptable in terms of separating the underlying sources in bounded multivariate data.
- We also propose the acIVABMGGMM which provides more flexibility in controlling the relationship between the reference signal and the estimated signal. It also enables the integration of multiple reference signals into the IVA framework. The acIVABMGGMM assumes a second-order correlation within bounded support component vectors, which allows for greater flexibility in extracting knowledge from the data.
- We tested our proposed models on various ECG applications, including heartbeat separation, fetal ECG extraction, and arrhythmia detection. We used acICABMGGMM and acIVABMGGMM components to identify the estimated ECG signal and compared them with the original ECG sources.
- We incorporate our models into the Pan-Tompkins QRS detection algorithm for the purpose of identifying the R-peak within QRS signals.

The Chapter is organized as follows: In Sections 5.2 and 5.3, we introduce the general framework for incorporating adaptive constrained parameters into the ICA and IVA bounded multivariate objective functions. Section 5.4 describes our experimental results, followed by the conclusion in Section 5.6.

## 5.2 ICA-based approaches for blind source separation

### 5.2.1 Bounded multivariate generalized Gaussian mixture model using ICA

In the ICA mixture model, the observed data is assumed to originate from a combination of different components, each belonging to distinct and non-overlapping categories. The bounded ICA mixture Equations are discussed in detail in Chapter 2.

### 5.2.2 Constrained bounded ICAMGGMM (cICABMGGMM)

To estimate independent components, it is essential to use a suitable measure of independence in the calculation of the demixing matrix. Maximum likelihood offers an inherent objective function for ICA, as follows:

$$\mathcal{J}_{ICA}(\mathbf{W}) = E \left[ \sum_{n=1}^N \log p(\mathbf{y}_n) \right] + \log |\det(\mathbf{W})| \quad (5.1)$$

The choice of distribution for  $p(\mathbf{y}_n)$  distinguishes between different ICA algorithms. In our case,  $p(\mathbf{y}_n)$  is modeled as a bounded multivariate generalized Gaussian distribution as in Equation (2.4). Incorporating reference signals presents a valuable approach that relaxes the assumption of independence and improves the optimization process. To accomplish this, we combine Equations (5.1), (2.4), and (2.7), and then apply constraint parameters to this combination formulation. Then, the gradient of cICABMGGMM is given by:

$$\frac{\partial \mathcal{J}_{cICABMGGMM}}{\partial \mathbf{W}_l} = \frac{\partial \mathcal{J}_{ICABMGGMM}}{\partial \mathbf{W}_l} - \mathbf{h}'_l(\mathbf{y}_l, \mathbf{r}_l) \lambda_l \mathbf{r}_l \quad (5.2)$$

where  $\mathbf{h}'_l(\mathbf{y}_l, \mathbf{r}_l)$  denotes the derivative of  $\mathbf{h}_l(\mathbf{y}_l, \mathbf{r}_l)$ . Within the cICABMGGMM framework, the objective function is optimized while adhering to the constraint imposed by the inequality constraint function  $\mathbf{h}_l(\mathbf{y}_l, \mathbf{r}_l)$ , which helps in guiding the learning process and ensuring the model adheres to certain properties.

$$\mathbf{h}_l(\mathbf{y}_l, \mathbf{r}_l) = \rho_l - \epsilon(\mathbf{y}_l, \mathbf{r}_l) \leq 0 \quad (5.3)$$

This distance measure  $\epsilon$  is compared against a reference value  $\mathbf{r}_l$  associated with bounded source  $\mathbf{y}_l$ , and the inequality constraint threshold labeled as  $\rho_l$ . In this Chapter, we utilize the absolute value of Pearson's correlation coefficient to define the distance measure, such as  $\epsilon(\hat{\mathbf{y}}_l, \mathbf{r}_l) = |\text{corr}(\mathbf{y}_l, \mathbf{r}_l)|$ . The condition  $0 \leq \epsilon(\hat{\mathbf{y}}_l, \mathbf{r}_l) \leq 1$  is met. So, increasing the  $\rho_l$  value ensures that the estimated source closely matches the reference signal, preventing any variation in the reference component across datasets. Conversely, reducing  $\rho_l$  causes the estimated component to diverge from the reference signal, making it susceptible to noise and other imperfections. Therefore, the choice of  $\rho_l$  significantly impacts the performance of the cICABMGGMM algorithm. In this work, we introduce an adaptive approach for selecting the constraint parameter called acICABMGGMM, which will be discussed in the next subsection.

### 5.2.3 Adaptive constrained bounded ICAMGGMM (acICABMGGMM)

In this section, we introduce the acICABMGGMM, an enhanced version of ICABMGGMM. By introducing a constraint term, the cost function for the acICABMGGMM approach can be expressed as:

$$\begin{aligned} \mathcal{J}_{acICABMGGMM} = & E \left[ \log p(\mathbf{y}_l) \right] + \log \left| \mathbf{h}_l^T \mathbf{y}_l \right| \\ & + \frac{1}{2\gamma_l} \left( \left( \max \left\{ 0, \gamma_l \mathbf{h}_l(\mathbf{y}_l, \mathbf{r}_l) + \lambda_l \right\} \right)^2 - \lambda_l^2 \right) \end{aligned} \quad (5.4)$$

where  $p(\mathbf{y}_l)$  represents the ICABMGGMM cost function, derived from the combination of Equations (2.4) and (5.1),  $\lambda_l$  is a Lagrange multiplier, and  $\gamma_l$  is the penalty parameter.

In the acICABMGGMM approach, we use  $L$  to represent the total number of constraints, where each constraint is associated with a constraint parameter  $\rho_l$ . To determine the appropriate constraint parameter, the acICABMGGMM method selects a value for  $\rho_l$  from a predefined set of possible values denoted as  $\mathcal{P}$ . For acICABMGGMM learning steps, we initiate the demixing matrices, denoted as  $\mathbf{W}$ , with random values and set  $\lambda_l$  to 0, along with assigning a positive scalar value to  $\gamma_l$ . We also initialize the mean and covariance using the K-means algorithm. The shape parameter is set to 2. In each iteration, we calculate the posterior probability using the EM method. In the E-step, we calculate the posterior probability of each data point belonging to each cluster. In the M-step,

we update the mixing, mean, covariance, and shape parameters using the posterior probabilities calculated in the E-step. We calculate an estimate for the sources, denoted as  $\mathbf{y}_l$ . We then update the Lagrange multiplier  $\lambda_l$ . The gradient update function for this algorithm is as follows:

$$\Delta \mathbf{w}_l = E \left[ \frac{\partial}{\partial \mathbf{w}_l} \log p(\mathbf{y}_l) \right] + \frac{\mathbf{h}_l^T}{\mathbf{h}_l^T \mathbf{y}_l} + \mathbf{h}'_l(\mathbf{y}_l, \mathbf{r}_l) \lambda_l \mathbf{r}_l^T \quad (5.5)$$

where  $\mathbf{h}'_l$  represents the derivative of  $\mathbf{h}_l$  w.r.t  $(\mathbf{y}_l, \mathbf{r}_l)$ . In each iteration of the algorithm, we employ gradient ascent to update the Lagrange multiplier:

$$\hat{\lambda}_l \leftarrow \max \left\{ 0, \gamma_l \mathbf{h}_l(\mathbf{y}_l, \mathbf{r}_l) + \lambda_l \right\} \quad (5.6)$$

The constraint parameter is given as:

$$\hat{\rho}_l = \arg \min_{\rho_l \in \mathcal{P}} \left\{ \min \left( \left| \rho_l - \left| \text{corr}(\mathbf{y}_l, \mathbf{r}_l) \right| \right| \right) \right\} \quad (5.7)$$

## 5.3 IVA-based approaches for BSS

### 5.3.1 Bounded multivariate generalized Gaussian mixture model using IVA

In IVA mixture, components within a source are independent, while corresponding components extracted from different mixture sources are maximally dependent. This means that corresponding component dependence across multiple sources and component independence within a single source are maximized simultaneously. The indicator function (5.8) is used to develop the IVA mixture as described in Chapter 2. The indicator function is defined as:

$$H(\mathbf{x}|j) = \begin{cases} 1 & \text{if } \mathbf{x} \in \Omega_j \\ 0 & \text{otherwise} \end{cases} \quad (5.8)$$

where the bounded support region is denoted by  $\Omega_j$ . By utilizing the indicator function, we describe the bounded distribution as follows:

$$p(\mathbf{x}|\xi_j) = \frac{f(\mathbf{x}|\xi_j)H(\mathbf{x}|j)}{\int_{\Omega_j} f(\mathbf{u}|\xi_j)du} \quad (5.9)$$

The derivative of the cost function in IVA is expressed as:

$$\frac{\partial \mathcal{J}_{IVA}}{\partial \mathbf{W}_n^{[m]}} = -\sum_{n=1}^N E \left\{ \frac{\partial \log(p(\mathbf{y}_n))}{\partial \mathbf{y}_n^{[m]} \frac{\partial \mathbf{y}_n^{[m]}}{\partial \mathbf{w}_n^{[m]}}} \right\} - \left( \mathbf{w}_n^{[m]} \right)^{-T} \quad (5.10)$$

### 5.3.2 Constrained IVABMGMM (cIVABMGMM)

cIVA successfully captures the distinct variations found in the features extracted from a set of subjects. This approach integrates prior information about the sources into the IVA model, which limits the potential solutions and effectively handles the difficulties posed by high dimensionality. By introducing the constraint term, the cost function for the constrained IVABMGMM can be formulated as:

$$\begin{aligned} \mathcal{J}_{cIVABMGMM}(\mathbf{W}_l^{[m]}) &= \mathcal{J}_{IVABMGMM} - \frac{1}{2\gamma_l^{[m]}} \left\{ \left[ \max \{0, \lambda_l^{[m]} + \gamma_l^{[m]} \mathbf{h}_l(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]})\} \right]^2 \right. \\ &\quad \left. - (\lambda_l^{[m]})^2 \right\} \end{aligned} \quad (5.11)$$

Here,  $\lambda_l^{[m]}$  represents the Lagrangian multiplier and  $\gamma_l^{[m]}$  (where  $\gamma_l^{[m]} > 0$ ) is a learning parameter. By employing the vector gradient descent method, we can express the gradient update function of Equation (5.11) as:

$$\frac{\partial \mathcal{J}_{cIVABMGMM}(\mathbf{W}_l^{[m]})}{\partial \mathbf{W}_l^{[m]}} = \frac{\partial \mathcal{J}_{IVABMGMM}}{\partial \mathbf{W}^{[m]}} - \mathbf{h}'_l(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \hat{\lambda}_l^{[m]} (\mathbf{r}_l^{[m]})^T \quad (5.12)$$

where  $\mathbf{h}'_l$  denotes the derivative of  $\mathbf{h}_l$  with respect to  $(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]})$ , and  $\phi^{[m]}(\mathbf{y}_l) = -\partial \log p_1(\mathbf{y}_1) / \partial \mathbf{y}_1^{[m]}$ . In our case,  $\phi^{[m]}(\mathbf{y}_l)$  represent by the BMGGMM Equation (2.4). The Lagrange multiplier is updated in each iteration using the gradient ascent method:

$$\hat{\lambda}_l^{[m]} \leftarrow \max \left[ 0, \gamma_l^{[m]} \mathbf{h}_l(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) + \lambda_l^{[m]} \right] \quad (5.13)$$

Within this framework, the mutual information objective function is optimized while adhering to the constraint imposed by the inequality constraint function  $\mathbf{h}_l$ .

$$\mathbf{h}_l(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) = \rho_l - \epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \leq 0 \quad (5.14)$$

where  $\mathbf{r}_l^{[m]}$  represents the reference vector for  $\mathbf{y}_l^{[m]}$ . The cIVABMGGMM framework introduced in this study offers the ability to define the distance measure  $\epsilon$  in a flexible manner. In this work, Pearson correlation is employed as the chosen distance measure:

$$\epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) = \left| \text{corr}(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \right| \quad (5.15)$$

By employing Pearson correlation as the distance measure, we impose a constraint on  $\epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]})$  to fall within the range of 0 and 1, effectively limiting the value of  $\rho_l^{[m]}$  to be less than or equal to 1. As a result, a higher value of  $\rho_l^{[m]}$  imposes a more stringent constraint on the decomposition process, while a lower value of  $\rho_l^{[m]}$  weakens the influence of the constraint on the decomposition.

The performance of the cIVABMGGMM algorithm is highly dependent on the selection of the constraint parameter  $\rho_l^{[m]}$ . For that, an adaptive technique called acIVABMGGMM is proposed to select the constraint parameter. The details of this technique are explained in the next subsection.

### 5.3.3 Adaptive constrained bounded IVAMGGMM (acIVABMGGMM)

The use of reference signals is an effective way to improve the performance of optimization searches and leverage the benefits of both model-driven and data-driven techniques. However, using a fixed value for the constraint parameter may not allow the model to efficiently estimate the local patterns for each dataset. To address this issue, we introduce the acIVABMGGMM method,

which allows for controlling the degree of similarity between the estimated source and the reference signal. This technique automatically adjusts the constraint parameter during the algorithm's execution to optimize its performance. By doing so, it is expected to improve the performance of the cIVABMGMM algorithm.

In the acIVABMGMM approach, we choose a value  $\rho_l$  from a set of potential values represented by  $\mathcal{P}$ . The demixing matrices, denoted as  $\mathbf{W}^{[m]}$ , are initialized randomly, and we set  $\lambda_l^{[m]} = 0$  and  $\gamma_l^{[m]}$  to a positive scalar value. The parameter estimation process is given in Subsection 5.2.3. During each iteration, we estimate the sources, denoted as  $\mathbf{y}_l^{[m]}$ , where  $m$  varies from 1 to  $M$ , and  $l$  varies from 1 to  $L$ . The first SCV is selected as the constrained component. The acIVABMGMM gradient is computed as:

$$\begin{aligned} \frac{\partial \mathcal{J}_{acIVABMGMM}}{\partial \mathbf{W}_l^{[m]}} &= \frac{\partial \mathcal{J}_{cIVABMGMM}}{\partial \mathbf{W}_l^{[m]}} - \frac{1}{\gamma_l^{[m]}} \left\{ \left[ \max \left( 0, \gamma_l^{[m]} \left( \hat{\rho}_l - \epsilon(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \right) \right) \right] \right. \\ &\quad \left. + \left[ \lambda_l^{[m]} \right]^2 - (\lambda_l^{[m]})^2 \right\} \end{aligned} \quad (5.16)$$

To update this parameter, we conduct a search to find the highest value of  $\rho_l^{[m]}$  from the set  $\mathcal{P}$  that satisfies the condition specified in Equation (5.14) for all  $M$  subjects. Afterward, we calculate the distance  $\epsilon(\cdot, \cdot)$  of the estimated correlation for all possible values of  $\rho_l^{[m]}$  from the set  $\mathcal{P}$  across all subjects. Subsequently, we estimate  $\lambda_l^{[m]}$  and  $\rho_l^{[m]}$  using the following Equations, respectively:

$$\hat{\lambda}_l^{[m]} = \max \left\{ 0, \gamma_l^{[m]} \mathbf{h}_l(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) + \lambda_l^{[m]} \right\} \quad (5.17)$$

$$\hat{\rho}_l^{[m]} = \arg \min_{\rho_l^{[m]} \in \mathcal{P}} \left\{ \min \left[ \left| \rho_l - \text{corr}(\mathbf{y}_l^{[m]}, \mathbf{r}_l^{[m]}) \right| \right] \right\} \quad (5.18)$$

The value of  $\rho_l^{[m]}$  that results in the smallest distance is selected. This new value of the constraint parameter is then used to compute the gradient and update the demixing matrix. The process is iterated until convergence is achieved. As a result, the constraint parameter gradually approaches the actual correlation between the source and the reference signal with each iteration, without imposing

the reference signal directly onto the decomposition process.

## 5.4 Experimental Results for cICABMGGMM and acICABMGGMM

We propose to validate the effectiveness of our proposed models, cICABMGGMM and acICABMGGMM, through the process of ECG signal separation. To achieve this, we conducted a simulation experiment and two real applications: fetal ECG extraction and arrhythmia identification. In the first ECG experiment, our models are utilized to predict the fetal ECG signal from the maternal signal. In the second ECG experiment, our proposed models are used to differentiate between a normal ECG signal and one exhibiting arrhythmia. We incorporate state-of-the-art models for comparison with our proposed models in both experiments. In terms of evaluating performance, both subjective and objective metrics are employed to assess the proposed models. The subjective evaluation involves examining ECG signals obtained both before and after source separation, while the objective assessment encompasses metrics such as SDR, SIR, SAR, SNR, and Joint-ISI. We also propose the incorporation of our models into the Pan-Tompkins algorithm to enhance the detection of R-peaks. The Pan-Tompkins method is one of the most popular and commonly used algorithms for QRS detection [220]. It has been widely explored and has demonstrated good performance in identifying QRS complexes in ECG data. The Pan-Tompkins algorithm uses a series of filters to emphasize the frequency content of the QRS complex and reduce background noise. It then squares the signal to magnify the QRS contribution, making it simpler to recognize the QRS complex. Finally, adaptive thresholds are used to detect the peaks in the filtered estimated ECG signal.

### 5.4.1 Simulation experiment

To evaluate the reliability of our proposed models, we examine their performance using simulated data generated from the multivariate generalized Gaussian distribution. We generate a multivariate dataset, denoted as  $\mathbf{x} \in \mathbb{R}^N$ , such that  $\mathbf{x} = \mathbf{A}\mathbf{s}$ . The elements of the mixing matrix  $\mathbf{A}$  are drawn randomly from a uniform distribution. The  $N$  source components are produced using a multivariate generalized Gaussian distribution. To control the correlation within an IC component, we generate the corresponding matrix with elements randomly sampled from a normal distribution

$N(0, 1)$ . Once the ICs components are obtained, the sources for this dataset are  $s = [s_1, \dots, s_N]$ . These ICs are formed from  $V = 1000$  samples. The acICABMGGMM is performed on these ICs, with 100 runs. The performance is evaluated using the Joint-ISI metric, which quantifies the ability of the methods to separate the sources. This approach is particularly suitable for multichannel ECG analysis, where the sources typically exhibit Gaussian distributions and correlation across subjects. The average Joint-ISI across all converged runs for all models is depicted in Fig. 5.1.

Initially, we utilize the K-means algorithm to establish the mixture parameters, which include mean, covariance, and shape parameters. The mixing parameter is initialized at  $1/M$  to ensure a total sum of 1. Subsequently, we employ the EM algorithm to calculate the posterior probability in the E-step and update the mixture parameters in the M-step. To determine model convergence, we specify a small tolerance value. We iterate through updating the mixture parameters until our proposed model achieves convergence, utilizing the optimal parameters for source separation. We set a maximum of 1020 iterations for each model, with a convergence threshold set at 0.001. This results in the derivation of a new de-mixing matrix, denoted as  $\mathbf{W}$ , which we use to estimate the sources  $\mathbf{y}_n = \mathbf{W} \times \mathbf{x}_n$ . The convergence process is illustrated in Fig. 5.1 (a). It shows that our proposed model, acICABMGGMM, demonstrates significantly improved convergence compared to the base models, ICA, Adaptive cICA, and Fixed-point cICA. Our proposed model converges within 250 iterations, whereas the base models require more than 500 iterations to converge.

We evaluated the efficacy of the acICABMGGMM method across a range of  $\gamma_l$  values from 1 to 1000, observing no significant change in performance. Consequently, we opted for a small positive value, setting  $\gamma_l$  to 3. For the cICABMGGMM method with fixed constraint parameters, we varied the values of  $\rho_l$  from 0.001 to 0.9. A  $\rho_l$  value of 0.001 indicates weaker constraints, while a value of 0.9 signifies stronger constraints. In order to verify if the proposed method accurately estimates the constrained source, we measure the dissimilarity factor,  $\alpha$ , between the constrained estimated source,  $\mathbf{y}_l$ , and corresponding ground truth,  $\mathbf{r}_l$ , computed as  $\alpha = 1 - |\text{corr}(\mathbf{y}_l, \mathbf{r}_l)|$ . Fig. 5.1 (b) assesses the performance of acICABMGGMM versus cICABMGGMM in terms of the dissimilarity factor. In each box plot, the median is represented by a horizontal orange line, with the upper and lower edges corresponding to the 75th and 25th percentiles, respectively. A higher value of this metric indicates a weaker influence of the constraints on the source. With the cICABMGGMM

method, the estimation quality of the constrained component deteriorates when a higher constraint parameter is employed, while the acICABMGGMM method maintains a low dissimilarity factor. The proposed acICABMGGMM demonstrates better performance than regular cICABMGGMM for lower values  $\rho_l = 0.001$ , which is equivalent to performing unconstrained ICA.

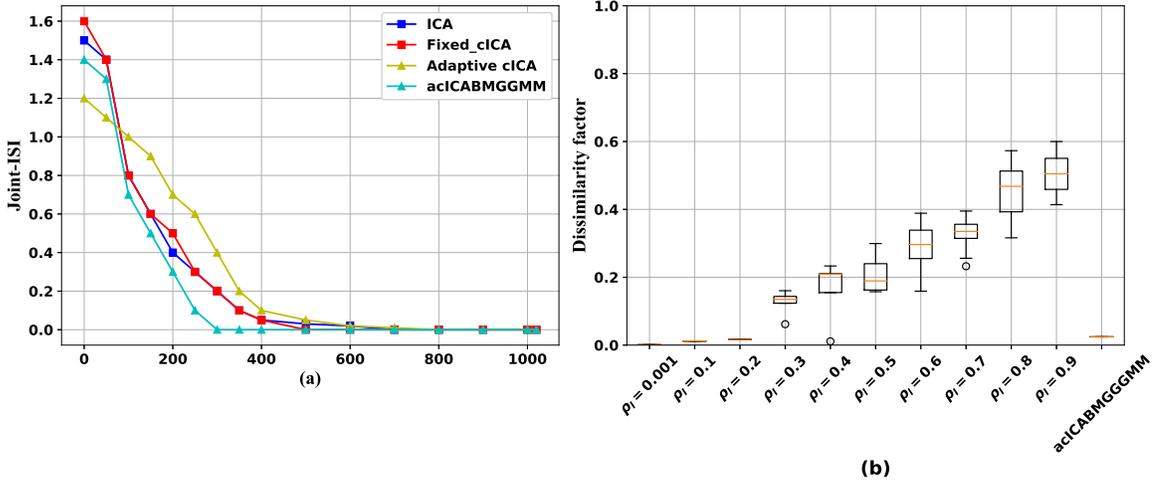


Figure 5.1: The convergence process and robustness of our model where (a) illustrates the iterations number required for convergence of acICABMGGMM, while (b) represents dissimilarity factor for acICABMGGMM vs cICABMGGMM.

## 5.4.2 Fetal ECG signal extraction

Obtaining fetal ECG signals from the ECG signals of the mother’s abdomen poses a challenge due to the interference from artifacts caused by muscle contraction, breathing, and noise. Recently, reference-based methods are employed for fetal ECG extraction, but their efficacy is not consistently high. In response to this challenge, researchers are investigating the application of BSS to extract fetal ECG. The experiment utilized real electrode recordings retrieved from the DAISY database [221]. These recordings were acquired from a pregnant woman’s body using eight skin electrodes at a sampling frequency of 500 Hz. In our investigation, we utilized the first abdominal recording in conjunction with three thoracic recordings (as illustrated in Fig. 5.2). The first recording depicts a mixture of fetal electrocardiogram (FECG), maternal electrocardiogram (MECG), and noise. The second signal was collected from the maternal thoracic region, with each sensor signal consisting of 2500 measurements.

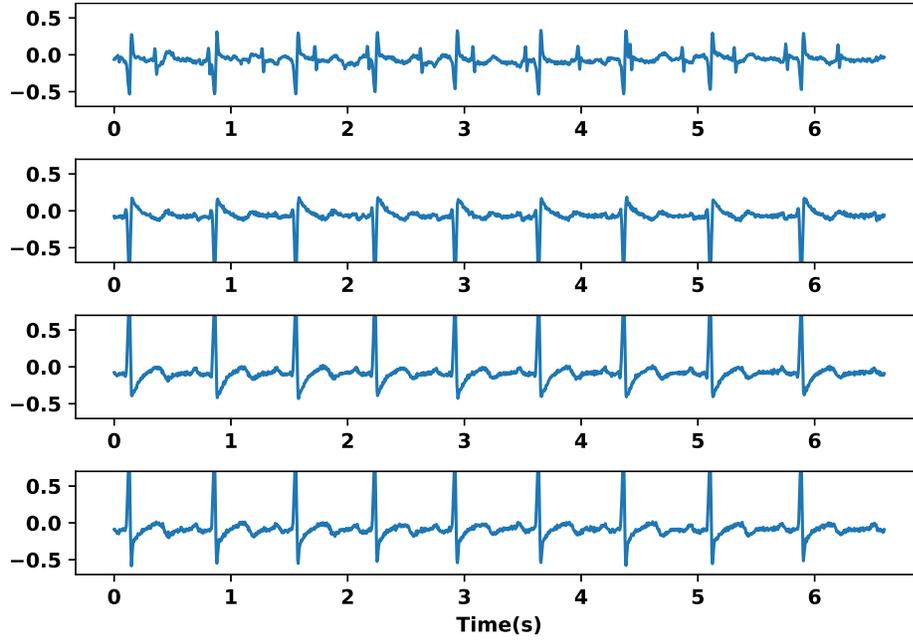


Figure 5.2: Real ECG recordings. The first row represents the abdominal recording, whereas the second to fourth rows represent thoracic recordings.

Following the filtration of the original ECG signals, our models were applied to estimate  $2 \times 2$  basis functions, resulting in the derivation of 2 estimated sources. In the cICABMGGMM approach, we utilized specific values for the threshold of the inequality constraint ( $\rho_l = 0.001$  to  $0.9$ ) where the constraint parameter with the smallest distance  $\epsilon$  is chosen as the optimal parameter. We evaluated the effectiveness of the cICABMGGMM and acICABMGGMM methods across a range of  $\gamma_l$  values spanning from 1 to 1000. No significant variation in their performance was observed. Consequently, we opted for a small positive value, setting  $\gamma_l$  to 3. Conversely, for the acICABMGGMM method, a range of constraint parameters  $\rho_l$  was employed from the set  $\mathcal{P} = 0.001, \dots, 0.9$  while maintaining a constant value of  $\gamma_l = 3$  for the constrained learning. To ensure robustness, we initialized the model's parameters randomly and executed the process with varying initial setups. The optimal run was determined through evaluation using the Joint-ISI metric. Here, the implicit objective is to opt for the most suitable initialization parameters, while the stability of each model is confirmed by subjecting it to 1020 iterations against a low threshold, ensuring convergence is examined for each model.

In Table 5.1, we present objective evaluations conducted on recovered ECG signals. Our analysis reveals that acICABMGGMM outperforms ICA, Non-orthogonal cICA, Adaptive ICA, ICAMGGMM, Fixed-point cICA, acICAMGGMM, and ICABMGGMM models across various measures, including SAR with a value of (229.43) and SNR with a value of (21.57). Notably, both the acICABMGGMM and cICABMGGMM models demonstrate superior performance compared to other ICA-based models. The metric Joint-ISI clearly shows the effectiveness of our proposed models in separating the FECG signal from the MECG signal. It is worth mentioning that the cICABMGGMM model shows a slight performance lag compared to the acICABMGGMM model, and this difference might be attributed to the selection of constraint parameters. Fig. 5.3 illustrates the fetal

Table 5.1: Performance of the proposed cICABMGGMM and acICABMGGMM in terms of presented performance metrics.

Models	SDR	SIR	SAR	SNR	Joint-ISI
<b>acICABMGGMM</b>	<b>18.25</b>	<b>18.78</b>	<b>229.43</b>	<b>21.57</b>	<b>0.074</b>
<b>cICABMGGMM</b>	<b>14.61</b>	<b>15.75</b>	<b>226.82</b>	<b>17.52</b>	<b>0.125</b>
ICABMGGMM	10.52	13.14	224.98	13.04	0.185
acICAMGGMM	10.09	12.69	224.45	12.52	0.186
Fixed-point cICA	9.20	12.08	214.88	11.15	0.216
ICAMGGMM	9.64	11.64	225.53	13.95	0.289
Adaptive cICA	9.23	11.11	220.07	13.54	0.292
Non-orthogonal cICA	8.95	10.96	119.89	12.61	0.391
ICA	9.01	11.42	224.56	14.91	0.365

extracted ECG signals from the maternal ECG signals. This illustration showcases the capability of acICABMGGMM to extract fetal and maternal ECG signals from a mixture ECG signal, particularly when the fetal ECG signal exhibits a higher frequency than the maternal ECG signal. Moreover, we utilized the Pan-Tompkins algorithm on the estimated FECG to detect the R-peaks related to complex QRS patterns. This method effectively identifies these R-peaks within our estimated FECG signals, as indicated by the orange colour. These experiments show that our models exhibit superior performance compared to commonly used models like cICAMGGMM, ICAMGGMM, Adaptive cICA, Fixed-point cICA, and ICA. The degradation performance of the base models is also noticeable. Among all the implemented models, the acICABMGGMM model demonstrated the best performance. This improvement can be attributed to the constraints parameters that were incorporated adaptively into the cost function for this model. Furthermore, this approach considers

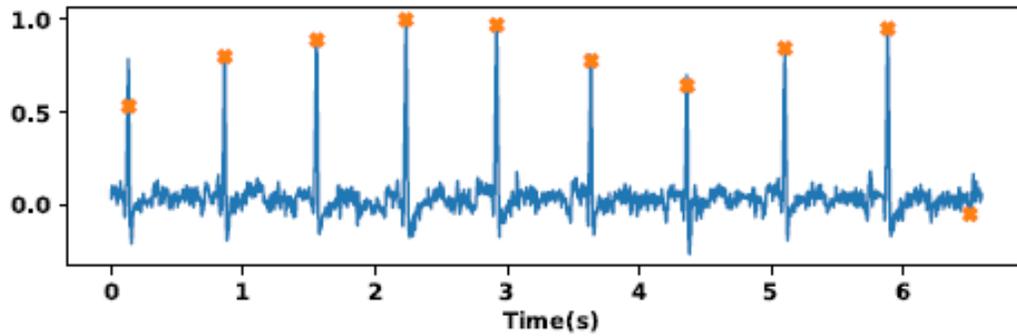


Figure 5.3: FECG signal estimation employing acICABMGGMM, coupled with R-peak identification using orange color.

correlated features by integrating a full covariance matrix, thus providing the flexibility required for precise extraction of the FECG signals from the MEGC signals. Overall, this marked enhancement in the quality of FECG extraction can be attributed to the incorporation of prior information in our proposed models.

### 5.4.3 Arrhythmia detection

This study utilizes the publicly accessible MIT-BIH dataset, accessible through PhysioNet [222]. The dataset consists of 48 records of 2-channel ECG signals digitized at a rate of 360 Hz, with a database bandwidth ranging from 0.1 Hz to 40 Hz. It includes 25 male subjects aged between 32 and 89 years and 22 female subjects aged between 23 and 89 years. Among the subjects, 60% were healthy. Each ECG signal underwent several preprocessing steps, including high-pass filtering (0.5 Hz, 5th-order Butterworth), low-pass filtering (40 Hz, 5th-order Butterworth), subsequent normalization (subtracting the mean and dividing by the standard deviation), and a sign-change to ensure consistent positive peaks for heartbeats.

We have employed the proposed cICABMGGMM and acICABMGGMM techniques on the preprocessed MIT-BIH data to differentiate between normal subjects and those with arrhythmia disease. The configuration settings for the cICABMGGMM and acICABMGGMM approaches are derived directly from those utilized in the previous experiment. To ensure robustness, we initiated the model's parameters randomly and conducted separate runs with different initializations. The best run was determined using the Joint-ISI metric.

The performance of various methods, including ICA, Non-orthogonal cICA, Adaptive ICA, ICAMGGMM, Fixed-point cICA, acICAMGGMM, ICABMGGMM, and acICABMGGMM was evaluated in terms of Joint-ISI, SDR, SIR, SAR, and SNR. Tables 5.2 and 5.3 demonstrate that our proposed models consistently show better performance compared to the basic models across a range of lower component combinations, spanning from 2 to 6. Notably, as the number of subjects increases, the performance of ICA-based models declines. With more than 24 subjects, both cICABMGGMM and acICABMGGMM exhibit significant performance improvements over the ICA-based models. The performance of cICABMGGMM is marginally inferior to acICABMGGMM, depending on the specific constraint parameters used. The acICABMGGMM model addresses the issue in data with high number of mixtures by incorporating prior information and dynamically managing the correlation between the reference signal and the estimated signal. As shown in Table 5.2: SDR and SIR separation metrics for cICABMGGMM and acICABMGGMM over source combinations ranging from 2 to 48.

Model	SDR					SIR				
	# subject combinations									
	2	6	12	24	48	2	6	12	24	48
<b>acICABMGGMM</b>	<b>33.72</b>	<b>9.86</b>	<b>8.19</b>	<b>7.19</b>	<b>5.03</b>	<b>34.54</b>	<b>12.25</b>	<b>8.89</b>	<b>5.28</b>	<b>4.54</b>
<b>cICABMGGMM</b>	<b>24.10</b>	<b>7.15</b>	<b>6.06</b>	<b>6.05</b>	<b>4.69</b>	<b>25.11</b>	<b>8.06</b>	<b>5.92</b>	<b>4.37</b>	<b>3.85</b>
ICABMGGMM	22.77	5.00	4.97	3.39	2.47	24.71	6.65	4.90	2.87	2.40
acICAMGGMM	14.42	4.88	4.40	3.69	2.76	18.30	6.28	4.64	2.95	2.49
Fixed-point cICA	15.43	4.01	3.64	2.91	2.72	13.21	5.84	4.16	3.66	2.02
ICAMGGMM	12.48	4.72	4.91	3.88	2.48	15.63	6.04	4.39	2.96	2.58
Adaptive cICA	10.26	4.29	4.03	3.36	2.31	14.86	4.83	4.52	3.73	2.92
Non-orthogonal cICA	10.25	3.98	3.82	3.29	2.58	13.73	5.09	4.65	3.05	2.67
ICA	11.16	4.35	4.04	4.15	2.98	14.02	5.28	4.25	2.68	2.21

bles 5.2 and 5.3, acICABMGGMM and cICABMGGMM models show strong SDR, SIR, SAR, and SNR performance with lower subject combinations (2, 6, 12, 24). However, performance significantly declines as the number of subjects increases to 48, with acICABMGGMM's SDR dropping from 33.72 to 7.19 and SIR from 34.54 to 4.54 dB. A similar trend is observed for SAR and SNR, as shown in Table 5.3. This reduction in performance is likely due to the increased complexity and signal overlap with more subjects, reducing the models' separation effectiveness. Despite this, both proposed models still outperform other methods for all subject combinations.

Fig. 5.4 illustrates the improved separation performance of our models in comparison to conventional methods. For Joint-ISI (depicted in the left graph), the proposed ICA models (cICABMG-GMM and acICABMGMM) demonstrate slightly superior performance compared to traditional methods, particularly in the case of lower mixtures ranging from 2 to 12. A significant enhancement becomes evident with a higher number of mixtures, such as 24 and 48, where our models outperform traditional methods based on Joint-ISI metrics. Similar trends are observed in the spatial correla-

Table 5.3: The evaluation of cICABMGMM and acICABMGMM performance over source combinations from 2 to 48, assessed using SAR and SNR separation measurements.

Model	SAR					SNR				
	# subject combinations					# subject combinations				
	2	6	12	24	48	2	6	12	24	48
<b>acICABMGMM</b>	<b>254.30</b>	<b>136.08</b>	<b>94.38</b>	<b>61.52</b>	<b>52.09</b>	<b>36.40</b>	<b>24.66</b>	<b>14.23</b>	<b>8.45</b>	<b>5.75</b>
<b>cICABMGMM</b>	<b>244.68</b>	<b>125.94</b>	<b>83.94</b>	<b>52.00</b>	<b>42.16</b>	<b>25.89</b>	<b>14.69</b>	<b>8.11</b>	<b>6.88</b>	<b>4.53</b>
ICABMGMM	242.06	116.37	73.64	42.20	32.05	24.15	9.81	6.58	4.92	3.32
acICAMGGMM	241.77	115.94	68.79	39.07	22.07	18.64	8.65	6.28	4.13	2.97
IVA	241.75	115.43	67.05	35.74	24.83	13.73	6.16	5.67	3.88	2.98
Fixed-point cICA	240.07	116.60	65.33	32.53	21.65	12.96	6.05	5.42	3.06	2.93
ICAMGGMM	241.79	115.71	66.36	37.77	22.01	15.09	6.43	4.49	3.90	2.78
Adaptive cICA	240.35	116.72	63.91	32.67	21.42	13.52	4.90	4.29	3.72	2.53
Non-orthogonal cICA	239.41	117.63	64.05	34.37	23.80	13.04	5.36	4.91	3.82	2.61
ICA	241.16	119.05	64.79	32.91	22.73	14.06	4.94	4.34	3.92	2.39

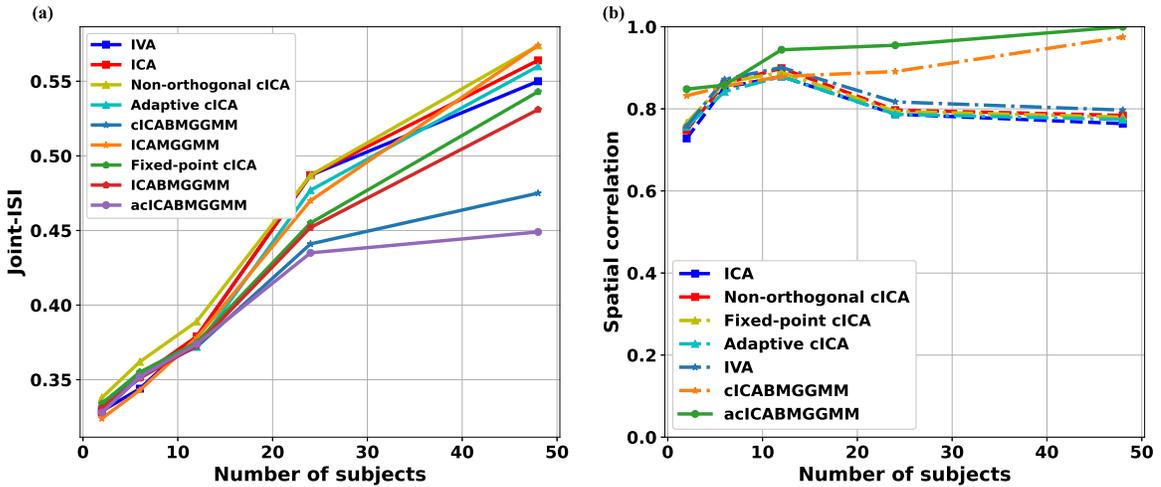


Figure 5.4: Our proposed model performance in terms of Joint-ISI and spatial correlation where (a) illustrates the model’s performance based on Joint-ISI, while (b) assesses its effectiveness in terms of spatial correlation.

tion analysis, as seen in the Joint-ISI evaluation. For lower mixture numbers, the spatial correlation

for our proposed models is slightly superior to that of the base models. However, as the number of mixtures increases, the performance of ICA traditional methods declines. Notably, we observe substantial performance improvements in our models, starting at 24 mixtures when compared to the base models. Additionally, the performance of cICABMGGMM is slightly below that of acICABMGGMM, depending on the constraint parameters used. The results shown in this figure make it clear that our models demonstrate a superior ability to recover constrained sources in scenarios with a high number of mixtures compared to traditional models, which displayed poor performance. This can be attributed to the effectiveness of our models in accurately estimating the underlying sources, leveraging the constraint parameters, and utilizing prior information.

#### **5.4.4 Robustness to noise**

We conducted additional experiments to evaluate the robustness of the proposed method under varying noise conditions. Specifically, rician noise was introduced at six distinct noise levels to assess the performance of our models on both fetal and arrhythmia applications. The results depicted in Fig. 5.5, highlight the impact of noise on the proposed methods. For the most effective separation model, acICABMGGMM, a performance reduction of 0.19 and 0.13 was observed at the highest noise levels (-6 dB and 0 dB, respectively). Similarly, for the arrhythmia dataset, acICABMGGMM exhibited a performance decrease of 0.18 and 0.13 under the same high noise conditions. Reducing the noise level leads to a noticeable increase in the spatial correlation of our models, demonstrating their robustness under noisy conditions. This resilience is particularly evident in scenarios with 6 dB and 12 dB noise levels. Despite the challenging conditions, acICABMGGMM demonstrated superior robustness, achieving the highest spatial correlation in noise-free scenarios (18 dB and 24 dB) and exhibiting the smallest performance declines (0.19 for fetal data and 0.18 for arrhythmia data) under high noise conditions at -6 dB. This highlights acICABMGGMM's exceptional ability to maintain performance stability across varying noise levels in both fetal and arrhythmia experiments.

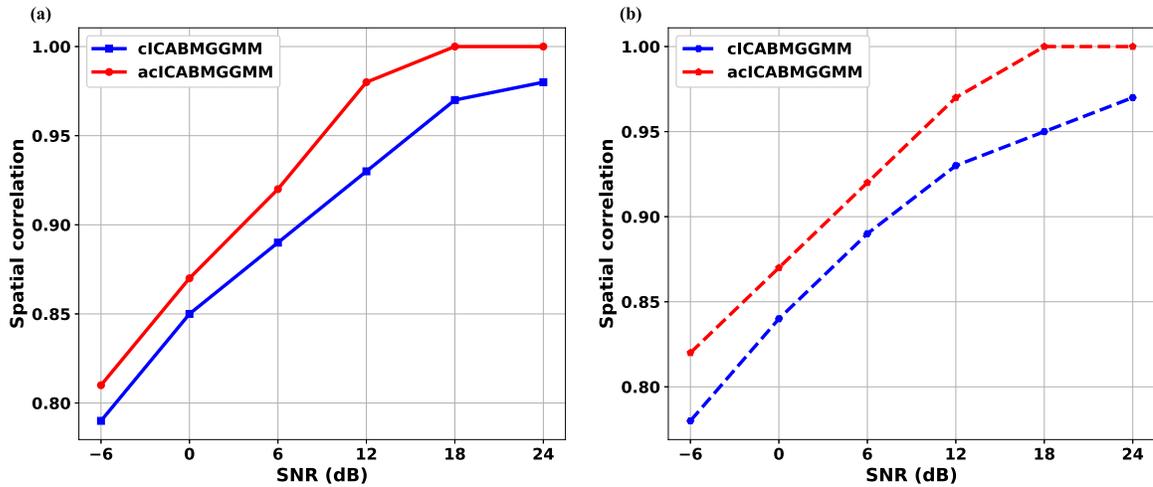


Figure 5.5: Our proposed model performance in terms of different noise conditions and spatial correlation where (a) illustrates the model's performance based on the fetal ECG signal extraction experiment, while (b) assesses its performance based on the arrhythmia detection experiment.

## 5.5 Experiments and Results for cIVABMGGMM and acIVABMGGMM

We propose validating the efficacy of our proposed models cIVABMGGMM, and acIVABMGGMM through ECG signal separation. To achieve this, we conduct three distinct experiments: heartbeat separation, fetal ECG extraction, and arrhythmia detection. In the heartbeat separation experiment, our models are employed to distinguish between four different classes, namely normal, murmur, extrahls, and extrasystoles. In the second experiment, we utilized our models to estimate the fetal ECG signal from the maternal signal. In the third experiment, we apply our proposed models to differentiate between a normal ECG signal and an ECG signal with arrhythmia.

### 5.5.1 Heartbeat separation

The dataset used for the classification of heartbeats was originally created for a machine learning challenge by Peter Bentley [223]. The dataset contains four classes of heartbeats: normal, murmur, extrahls, and extrasystoles. Normal represents healthy heartbeats; murmur represents extra sounds that occur when there is turbulence in blood flow that causes extra vibrations that can be heard; extrahls represent heartbeats with an additional sound; and extrasystoles represent additional

heartbeats that occur outside the physiological heart rhythm and can cause unpleasant symptoms [224, 225, 226]. Samples of this data are given in Fig. 5.6. Here, we show the two classes of heartbeat signals and their corresponding spectrograms, and the MFCCs features. Following the

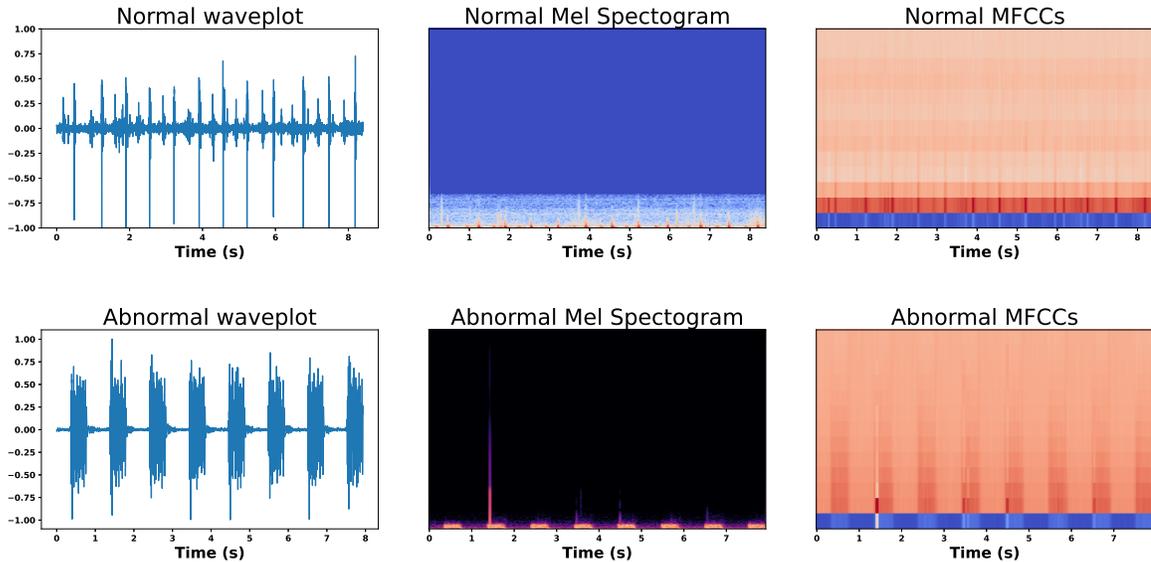


Figure 5.6: Heartbeat source signals. The first row indicates the normal heartbeat alongside its corresponding spectrograms and MFCCs features. The second row represents the abnormal heartbeat and its corresponding spectrogram and MFCCs features.

initial dataset processing, we extracted the MFCCs. During the feature extraction process, we computed the 13th-order MFCCs features and then combined their delta and acceleration characteristics to create 39-dimensional feature vectors. Then, the IVA mixture models were utilized to estimate basis functions, which were then used to separate mixed signals in different experiments. In this experiment, the basis functions of  $2 \times 2$ ,  $3 \times 3$ ,  $4 \times 4$ , and  $5 \times 5$  in separate experiments to compute 2, 3, 4, and 5 heartbeat sources. Linear mixtures of 2 sources were used to recover the heartbeat sources, and BSS was executed by employing cIVABMGGMM and acIVABMGGMM. The configurations for both cIVABMGGMM and acIVABMGGMM are given in Subsection 5.4.2. To ensure robustness, we initialized the model's parameters randomly and executed the process five times with varying initial setups. The optimal run was determined through evaluation using the Joint-ISI metric. Here, this step is implied to choose the best initialization parameters whereas the stability of each model is validated using 1020 iterations against the small value of the threshold to check the convergence for each model. To assess the quality of the recovered heartbeat signals and the

viability of the IVA mixture model in BSS, subjective and objective measures were performed on the recovered sources.

Table 5.4: Objective measures for separation of 2 and 3 heartbeat signals using our proposed models acIVABMGGMM and cIVABMGGMM.

Model	BSS of 2 heartbeat signals					BSS of 3 heartbeat signals				
	SDR (dB)	SIR (dB)	SAR (dB)	SNR (dB)	Joint-ISI	SDR (dB)	SIR (dB)	SAR (dB)	SNR (dB)	Joint-ISI
<b>acIVABMGGMM</b>	<b>35.32</b>	<b>25.21</b>	<b>200.21</b>	<b>332.20</b>	<b>0.031</b>	<b>32.51</b>	<b>23.65</b>	<b>190.28</b>	<b>330.03</b>	<b>0.042</b>
<b>cIVABMGGMM</b>	<b>30.91</b>	<b>21.35</b>	<b>198.32</b>	<b>330.68</b>	<b>0.034</b>	<b>30.31</b>	<b>20.65</b>	<b>186.74</b>	<b>328.86</b>	<b>0.048</b>
IVABMGGMM	29.87	19.35	196.82	329.96	0.035	29.01	18.43	183.30	322.44	0.052
acIVAMGGMM	23.87	19.01	193.50	327.04	0.039	21.07	17.95	182.06	323.56	0.063
CIVAMGGMM	25.71	18.19	191.81	324.59	0.056	22.35	14.15	181.21	322.54	0.138
IVAMGGMM	23.58	19.88	190.59	325.89	0.152	20.84	17.77	179.75	317.81	0.168
cIVA	24.23	19.53	192.61	320.67	0.248	23.41	16.92	178.31	316.61	0.157
IVA-G	20.68	17.30	180.84	315.09	0.055	22.02	12.89	171.95	314.28	0.158
IVA	21.54	19.47	189.25	320.24	0.055	20.66	16.93	171.79	315.05	0.168

Table 5.5: Objective measure for separation of 4 and 5 heartbeat signals using acIVABMGGMM and cIVABMGGMM proposed models.

Model	BSS of 4 heartbeat signals					BSS of 5 heartbeat signals				
	SDR (dB)	SIR (dB)	SAR (dB)	SNR (dB)	Joint-ISI	SDR (dB)	SIR (dB)	SAR (dB)	SNR (dB)	Joint-ISI
<b>acIVABMGGMM</b>	<b>30.21</b>	<b>20.31</b>	<b>182.33</b>	<b>325.53</b>	<b>0.152</b>	<b>20.34</b>	<b>18.24</b>	<b>171.42</b>	<b>318.21</b>	<b>0.231</b>
<b>cIVABMGGMM</b>	<b>28.75</b>	<b>18.52</b>	<b>179.05</b>	<b>320.02</b>	<b>0.157</b>	<b>18.15</b>	<b>16.93</b>	<b>166.83</b>	<b>312.64</b>	<b>0.265</b>
IVABMGGMM	26.38	16.81	176.13	315.92	0.162	15.01	14.33	161.35	295.98	0.345
acIVAMGGMM	24.54	15.03	174.71	312.84	0.174	14.30	13.70	154.24	286.18	0.421
CIVAMGGMM	19.56	15.03	174.68	312.05	0.539	9.76	14.14	146.89	285.15	0.692
IVAMGGMM	17.04	15.57	171.91	304.78	0.454	9.13	12.98	152.51	280.33	0.645
cIVA	19.23	14.22	170.48	302.71	0.732	8.35	10.93	152.21	280.11	0.820
IVA-G	18.73	13.33	165.24	299.60	0.729	5.03	8.08	136.73	272.33	0.922
IVA	17.89	14.86	164.24	304.44	0.739	6.97	10.41	149.86	279.81	0.942

Table 5.4 presents a comparison of the performance of IVA, IVA-G, cIVA, IVAMGGMM, cIVAMGGMM, acIVAMGGMM, IVABMGGMM, cIVABMGGMM, and acIVABMGGMM using various performance metrics such as SDR, SIR, SAR, SNR, and Joint-ISI. The results clearly demonstrate that acIVABMGGMM outperforms the other models in all four measures: SDR (35.32 dB), SIR (25.21 dB), SAR (200.21 dB), and SNR (332.20 dB) for the case of 2 mixtures. Additionally, the Joint-ISI value (0.031) for acIVABMGGMM is lower compared to the other conducted models. As the number of mixtures increased, a decline in model performance was observed, especially for the base models IVA and cIVA as shown in Table 5.5. Also, it clearly shows that the proposed models perform better than the base model for all performance metrics, indicating their ability to separate the mixing ECG signals. The ability of acIVABMGGMM to separate mixing heartbeats is observed, as shown in Fig. 5.7. This demonstrates that the proposed acIVABMGGMM algorithm effectively categorizes mixed ECG signals into their respective classes. The top

row corresponds to a normal heartbeat, while the bottom row represents an abnormal one. In summary, the experiments demonstrate that acIVABMGGMM performs better than cIVABMGGMM and other compared models. As the number of linear mixtures in source separation increases, the rate of improvement decreases. However, a notable enhancement is observable in the effectiveness of both our cIVABMGGMM and acIVABMGGMM models when compared with the base models. This enhancement can be ascribed to the integration of prior information within our proposed models, effectively mitigating the constraints of the base models in these scenarios. To empha-

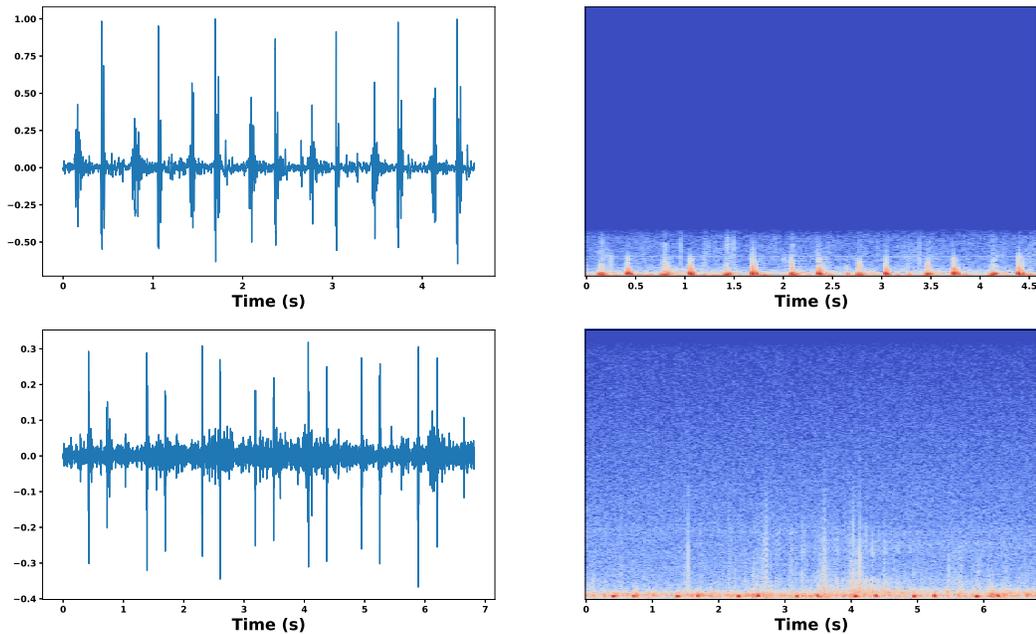


Figure 5.7: The estimated heartbeats using acIVABMGGMM, along with their corresponding spectrogram. The first row represents the normal estimated heartbeats, and the second row indicates the abnormal one.

size the superior performance of our proposed models in contrast to the traditional approaches, we assess their effectiveness using the Joint-ISI metric. As depicted in Fig. 5.8, our proposed models consistently outperform the basic models, particularly when dealing with more than two mixtures. Notably, both cIVABMGGMM and acIVABMGGMM exhibit substantial improvements compared to both the basic models including IVA, IVA-G, and cIVAMGGMM. This measure signifies diminished interference in the estimated ECG signals when considering our proposed models. In total, our proposed models demonstrate their effectiveness in separating the normal heartbeat signal from the abnormal ones accurately. These models illustrate enhanced flexibility in efficiently modeling

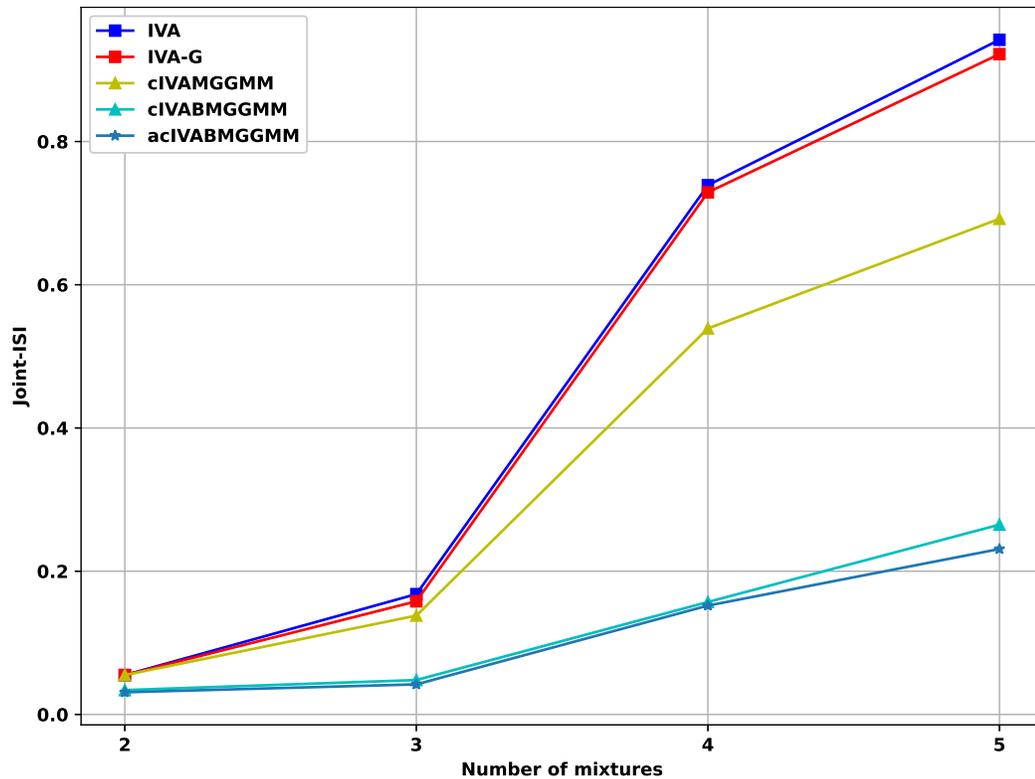


Figure 5.8: The effectiveness of our proposed models in terms of Joint-ISI when compared with conventional methods.

the data, attributed to their cost function, which is constructed using the full covariance matrix and incorporates prior information.

### 5.5.2 Fetal ECG signal separation

The experiment utilized real electrode recordings obtained from the DAISY database [221] as discussed in subsection 5.4.2. There is no a priori knowledge of the basis functions utilized during the source separation process. A linear combination of 2 ECG sources from the dataset was mixed, and BSS was performed using our proposed models cIVABMGGMM and acIVABMGGMM to recover the 2 ECG sources. We established the constraint parameters for these models following the configuration used in the previous heartbeat experiment. The resultant extracted sources underwent objective and subjective analysis to validate the quality of the recovered ECG signals and assess the effectiveness of the IVA mixture models within the BSS framework. This entire process was repeated ten times using distinct linear mixtures of 2 sources from the original dataset, with the

Table 5.6: Performance of the proposed cIVABMGGMM and acIVABMGGMM in terms of presented performance metrics.

Models	SDR	SIR	SAR	SNR	Joint-ISI
<b>acIVABMGGMM</b>	<b>28.96</b>	<b>39.39</b>	<b>284.51</b>	<b>63.67</b>	<b>0.032</b>
<b>cIVABMGGMM</b>	<b>19.41</b>	<b>34.73</b>	<b>280.22</b>	<b>59.51</b>	<b>0.062</b>
IVABMGGMM	16.97	29.19	276.83	56.18	0.092
acIVAMGGMM	15.68	27.41	275.28	49.88	0.160
cIVAMGGMM	15.72	27.51	269.77	48.02	0.320
IVAMGGMM	15.54	25.34	272.98	49.03	0.260
cIVA	13.12	26.26	272.60	45.29	0.480
IVA-G	13.62	12.74	267.23	41.59	0.391
IVA	13.14	10.60	270.59	41.58	0.416

objective metrics averaged across these ten separate experimental runs.

Table 5.6 illustrates the objective measures obtained after recovering ECG source signals. The acIVABMGGMM model outperforms other models such as cIVABMGGMM, IVABMGGMM, acIVAMGGMM, cIVAMGGMM, IVAMGGMM, cIVA, IVA-G, and IVA for the case of 2 mixture sources. In this experiment, both acIVABMGGMM and cIVABMGGMM show a significant improvement compared to other models. The performance metrics SDR (28.96), SIR (39.39), SAR (284.51), and SNR (63.67) indicate that acIVABMGGMM performs the best. Fig. 5.9 illustrates the fetal extracted ECG signals from the maternal ECG signals. This figure demonstrates the ability of acIVABMGGMM to recover fetal and mother ECG signals from the mixed ECG signal, where the fetal ECG signal has a higher frequency than the mother’s ECG signal. We additionally employed the Pan-Tompkins algorithm on the estimated FECG for identifying the R-peaks associated with complex QRS patterns. This approach possesses the capability to successfully identify these R-peaks within our estimated FECG signals.

To validate the effectiveness of our proposed models, cIVABMGGMM, and acIVABMGGMM, we used the Joint-ISI metric with the same experimental setup for all tests. Fig. 5.10 demonstrates that the proposed models outperform the base models for the mixture of two ECG signals. This finding highlights the limitations of the IVA basic model in this type of separation. The proposed models exhibit superior capability in recovering the constrained sources in fetal ECG separation

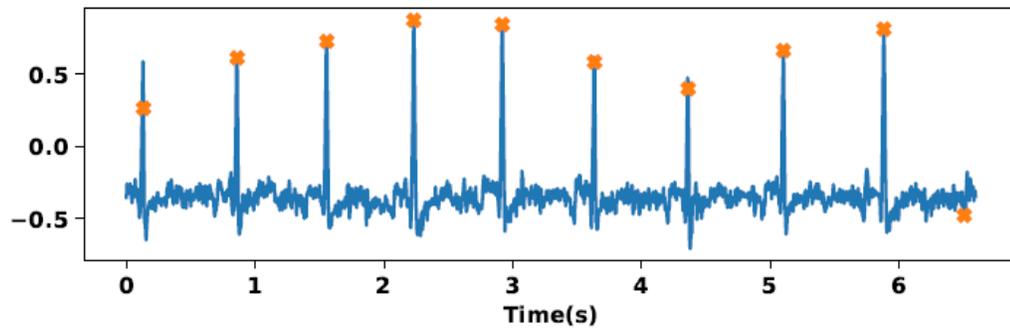


Figure 5.9: FCGG signal estimation employing acIVABMGMM, coupled with R-peak identification.

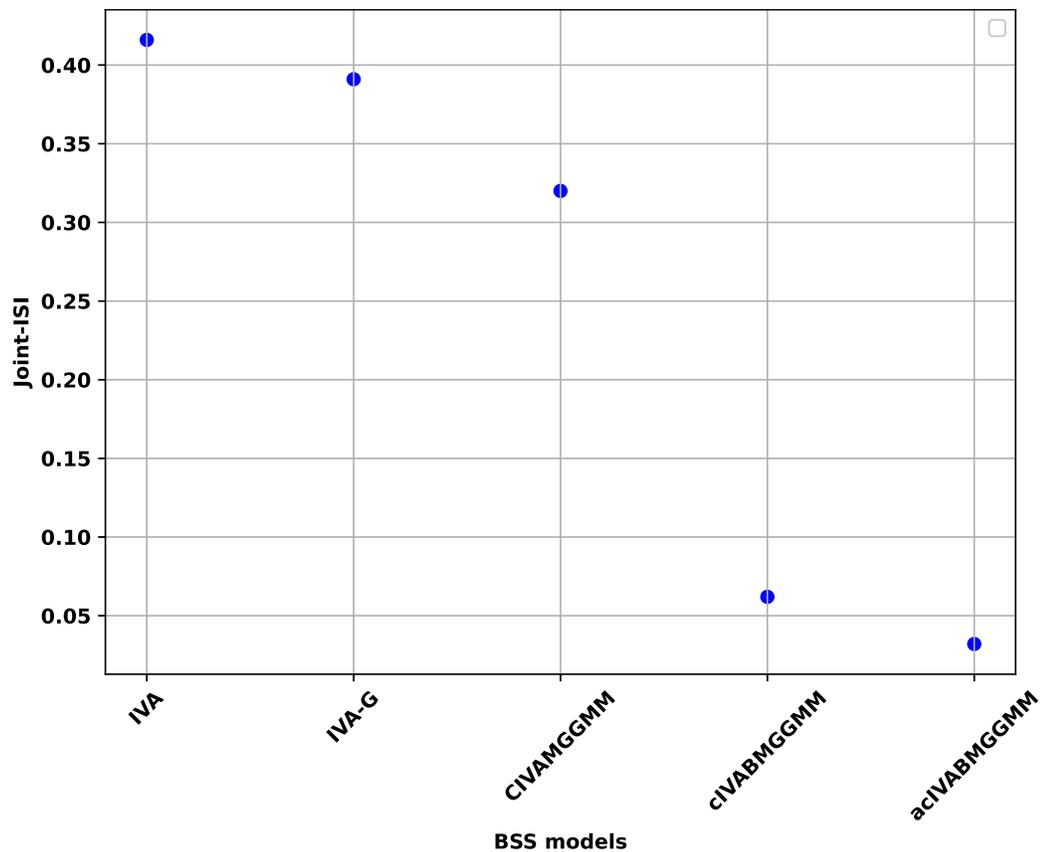


Figure 5.10: Joint-ISI for both base and proposed models.

compared to the base models, which performed poorly. This observation emphasizes the valuable role of prior information as a reference for seeking improved solutions in ECG signal separation.

These experiments show that our models exhibit superior performance compared to commonly

used models like IVABMGGMM, acIVAMGGMM, cIVAMGGMM, IVAMGGMM, cIVA, IVA-G, and IVA. The degradation performance of the base models is also noticeable. Among all the implemented models, the acIVABMGGMM model demonstrated the best performance. This improvement can be attributed to the constraints parameters that were incorporated adaptively into the cost function for this model. Furthermore, this approach considers correlated features by integrating a full covariance matrix, thus providing the flexibility required for precise extraction of the FECG signals from the MEEG signals.

### 5.5.3 Arrhythmia detection

The MIT-BIH dataset, which is publicly accessible through PhysioNet [222], is utilized in this study. This dataset was described in subsection 5.4.3. As shown in Fig. 5.11 each ECG signal underwent a series of preprocessing steps, including high-pass filtering (0.5 Hz, 5th-order Butterworth), low-pass filtering (40 Hz, 5th-order Butterworth), subsequent normalization (subtracting the mean and dividing by the standard deviation), and, a sign-change to ensure uniform positive peaks for heartbeats.

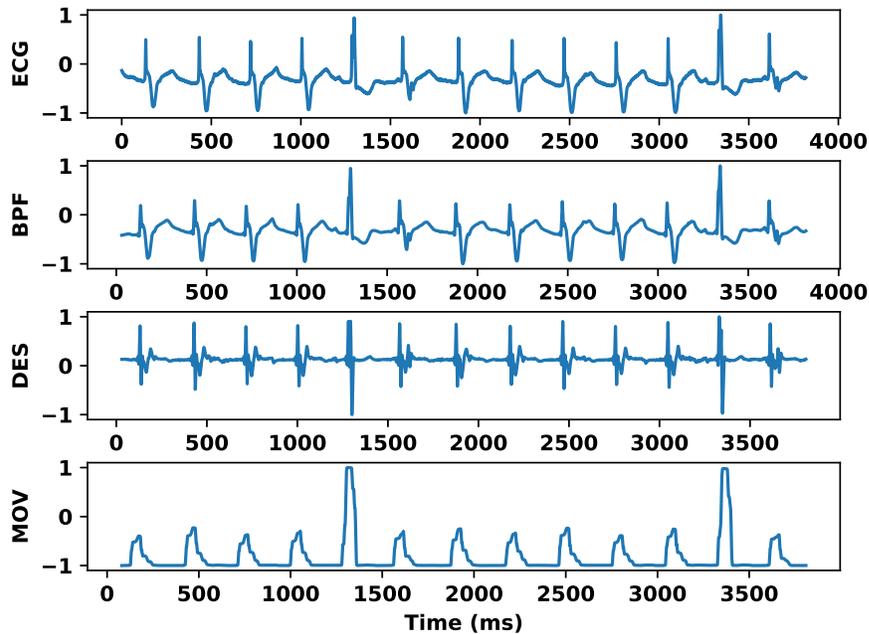


Figure 5.11: ECG data processing steps: The first row represents the original ECG signals. The second row indicates the filtered ECG signal (BPF). The third and fourth rows illustrate its derivative (DES) and moving window-integrated version (MOV).

We employ the cIVA and acIVA techniques, utilizing the IVABMGGMM approach, on the pre-processed MIT-BIH ECG data. The constraint parameters for these models were established based on the configuration employed in the previous experiments. To ensure robustness, we initiated the model’s parameters randomly and conducted five separate runs with different initializations. The best run was determined using the Joint-ISI metric. To validate the quality of separation, we compared the estimated signal with the reference signal using the Joint-ISI metric, SDR, SIR, SAR, and SDR. The demixing matrices from the chosen run are used to compute the estimated ECG sources.

Tables 5.7 and 5.8 present a comparison of the performance between our proposed models and the

Table 5.7: SDR and SIR separation measurements for cIVABMGGMM and acIVABMGGMM across source combinations spanning from 2 to 48.

Model	SDR					SIR				
	# subject combinations					# subject combinations				
	2	6	12	24	48	2	6	12	24	48
<b>acIVABMGGMM</b>	<b>41.30</b>	<b>14.45</b>	<b>11.39</b>	<b>8.20</b>	<b>6.63</b>	<b>46.72</b>	<b>33.54</b>	<b>23.25</b>	<b>13.19</b>	<b>9.53</b>
<b>cIVABMGGMM</b>	<b>29.84</b>	<b>9.91</b>	<b>8.65</b>	<b>6.26</b>	<b>5.21</b>	<b>34.32</b>	<b>23.41</b>	<b>12.43</b>	<b>7.89</b>	<b>6.63</b>
IVABMGGMM	28.15	8.27	6.89	5.00	3.94	22.72	10.10	8.10	5.38	4.21
acIVAMGGMM	17.90	7.94	6.98	4.98	3.67	22.61	8.91	6.97	4.68	3.79
cIVAMGGMM	16.32	7.85	5.73	4.61	3.70	20.53	8.51	5.05	4.41	3.06
IVAMGGMM	17.87	7.66	6.53	4.25	3.56	19.42	8.23	6.61	4.57	3.82
cIVA	17.70	7.46	5.49	4.33	3.42	19.02	7.98	6.64	4.12	3.75
IVA-G	15.39	5.61	4.74	4.53	3.26	14.74	6.23	5.28	4.49	3.11
IVA	16.36	7.51	5.15	4.84	3.54	17.32	7.23	6.09	4.31	3.42

Table 5.8: The effectiveness of cIVABMGGMM and acIVABMGGMM across source combinations from 2 to 48 evaluated through SAR and SNR separation criteria.

Model	SAR					SNR				
	# subject combinations					# subject combinations				
	2	6	12	24	48	2	6	12	24	48
<b>acIVABMGGMM</b>	<b>282.23</b>	<b>210.22</b>	<b>133.32</b>	<b>92.72</b>	<b>62.82</b>	<b>42.03</b>	<b>35.59</b>	<b>32.85</b>	<b>22.22</b>	<b>11.73</b>
<b>cIVABMGGMM</b>	<b>277.97</b>	<b>195.85</b>	<b>125.80</b>	<b>82.32</b>	<b>54.10</b>	<b>39.90</b>	<b>25.48</b>	<b>22.77</b>	<b>12.06</b>	<b>9.32</b>
IVABMGGMM	271.92	187.78	112.30	77.72	44.36	31.37	15.79	13.19	8.03	6.33
acIVAMGGMM	267.81	186.99	95.94	74.65	41.53	21.78	15.58	13.08	7.02	5.95
cIVAMGGMM	263.31	183.09	93.43	72.56	39.04	21.08	13.18	12.38	7.42	4.85
IVAMGGMM	264.80	173.69	92.81	63.11	39.64	21.18	15.58	13.59	7.39	5.89
cIVA	261.98	164.41	82.58	53.65	36.85	19.88	15.61	13.53	7.06	5.74
IVA-G	256.16	153.13	81.75	42.20	32.83	16.21	12.06	10.26	8.94	4.10
IVA	262.50	154.46	82.92	52.58	36.43	19.15	15.05	13.16	7.22	5.53

base models. Our proposed models demonstrate a better enhancement in performance for combinations spanning from 2 to 6 combinations of subjects. When applied to a high number of ECG

sources, we observe a noticeable decline in the performance of IVA, IVA-G cIVA, and cIVAMGGMM. As the number of subjects increases, particularly in combinations between 12 and 48, we notice a significant improvement in the performance of our cIVABMGGMM and acIVABMGGMM models compared to the base models. This improvement can be attributed to the utilization of prior information incorporated in our proposed models, which addresses the limitations of base models in scenarios with a high number of subjects. Fig. 5.12 presents a comparison of the performance of

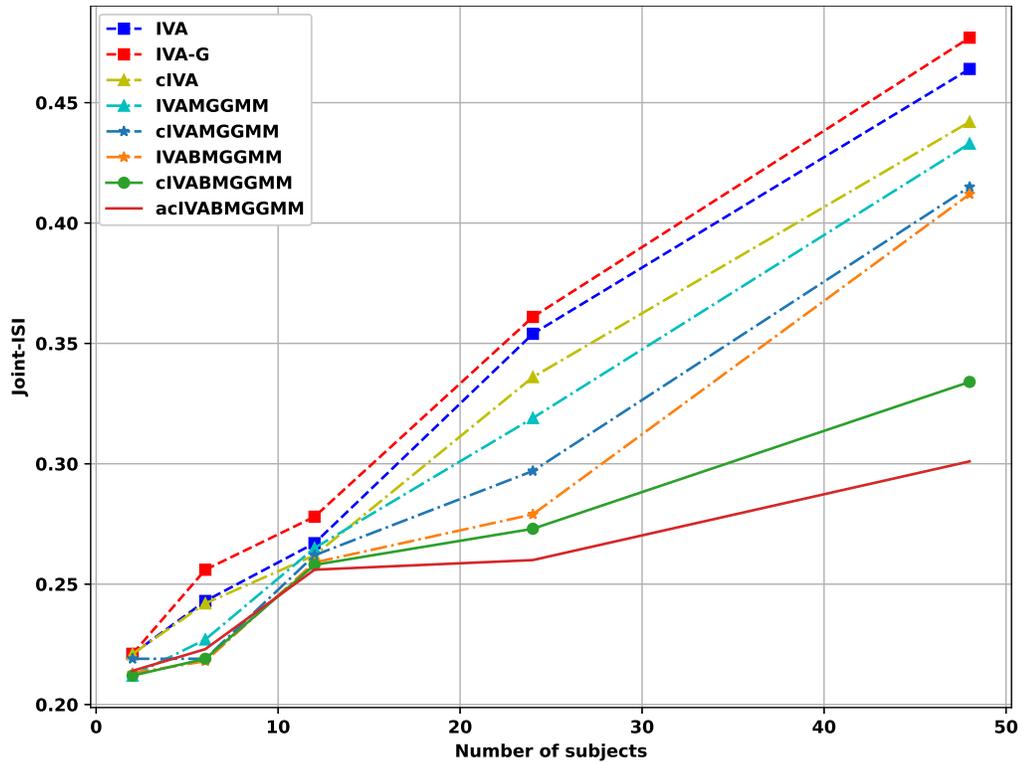


Figure 5.12: The joint-ISI for the four proposed models; cIVABMGGMM, acIVABMGGMM, and other compared models in terms of the number of subjects.

our models in terms of Joint-ISI. It is evident that the effectiveness of basic IVA improves with an increase in the number of subjects, but only up to a specific threshold for varying source numbers. This particular range can be considered the optimal zone where IVA reliably estimates the underlying parameters. Within this interval, the use of cIVABMGGMM and acIVABMGGMM does not yield a significant performance boost compared to IVA. This implies that the performance achieved by IVA within this range represents the highest attainable. However, it demonstrates that the base models show reduced effectiveness when dealing with a high number of ECG sources, whereas our

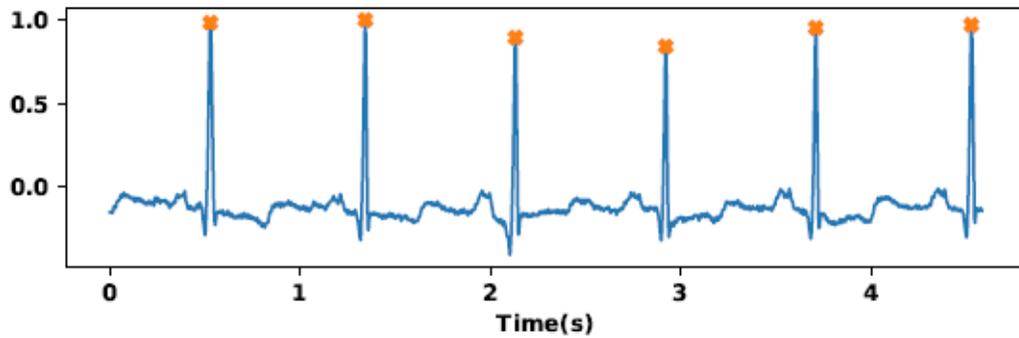


Figure 5.13: ECG signal estimate using acIVABMGMM, along with the detection of R-peaks.

models display enhanced performance. Notably, acIVABMGMM is the best-performing model compared to other implemented models. This model can effectively separate the sources of arrhythmias and healthy individuals. This outcome can be attributed to the competence of our models in precisely estimating the underlying sources by utilizing constraint parameters and prior information.

Fig. 5.13 demonstrates that our proposed model acIVABMGMM can accurately extract the arrhythmia ECG signals. The Pan-Topkins algorithm is utilized to identify the R-peaks of the QRS complex in the estimated ECG signal. The effectiveness of the Pan-Topkins model in detecting the R-peaks values of the QRS complex is demonstrated. These R-peaks are marked as orange crosses.

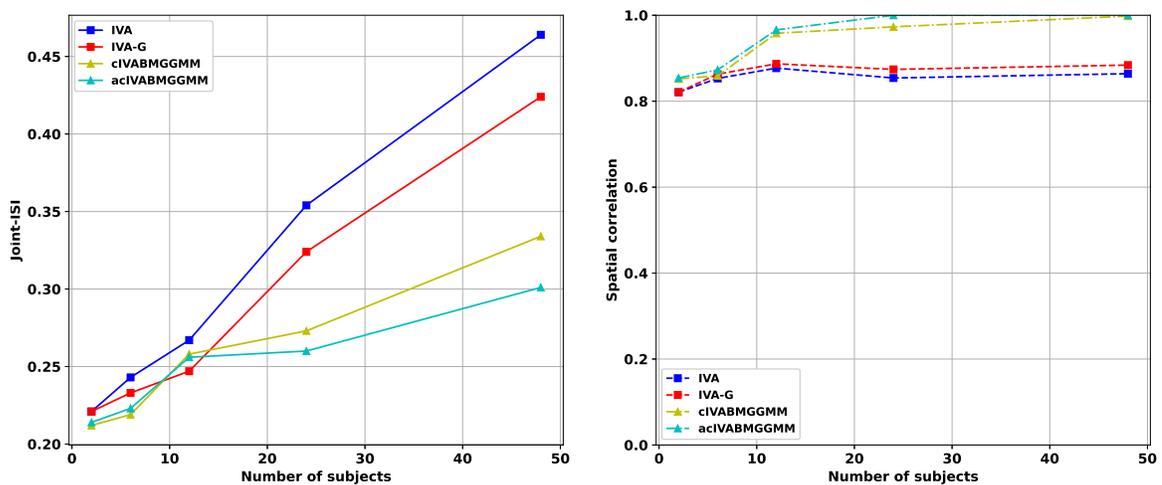


Figure 5.14: Our proposed model performance in terms of Joint-ISI and spatial correlation. The left graph illustrates the model's performance based on Joint-ISI, while the right graph assesses its effectiveness in terms of spatial correlation.

Figs. 5.14 highlight the enhanced separation performance of our models compared to traditional methods. In terms of Joint-ISI (depicted in the left graph), the proposed IVA models (cIVABMGMM and acIVABMGGMM) exhibit slightly better performance than traditional methods in the case of lower mixtures, ranging from 2 to 12. The substantial improvement becomes evident when dealing with a higher number of mixtures, such as 24 and 48, where our models outperform traditional methods in terms of Joint-ISI metrics, indicating lower interference in the estimated ECG signals. Similar trends are observed in the spatial correlation analysis as in the Joint-ISI assessment. For lower numbers of mixtures, the spatial correlation for our proposed models is slightly better than the base models. However, as the number of mixtures increases, the performance of traditional IVA methods deteriorates. Notably, we observe significant performance enhancements in our models, starting at 24 mixtures when compared to the base models. Furthermore, the performance of cIVABMGGMM is slightly inferior to that of acIVABMGGMM, depending on the constraint parameters employed. Based on the findings presented in this figure, it is evident that our models exhibit superior capability in recovering the constrained sources in a high number of mixtures compared to the traditional models, which performed poorly. This can be attributed to the capability of our models to accurately estimate the underlying sources by leveraging constraint parameters and prior information.

## 5.6 Conclusion

In this Chapter, we introduced the cICABMGGMM for the purpose of source separation. This model effectively separated the sources by utilizing the full covariance matrix. However, its performance relies heavily on the choice of constraint parameters, as demonstrated in our experiments. To address these limitations, we presented acICABMGGMM, designed to relax the independence assumption inherent in ICA and leverage the full covariance matrix to identify correlated features in the data. Furthermore, we introduced the cIVABMGGMM for the purpose of source separation. This model effectively separated the sources by utilizing the full covariance matrix. However, its performance relies heavily on the choice of constraint parameters, as demonstrated in our experiments. To address these limitations, we presented the acIVABMGGMM, designed to relax the

limitations inherent in IVA and leverage the full covariance matrix to identify correlated features in the data. In this model, constraint parameters adaptively control the relationship between reference signals and estimated ones. We additionally presented the integration of our proposed models with the Pan-Tompkins algorithm, aimed at identifying the R-peaks within the estimated ECG signals. We validated our models using heartbeat separation, fetal ECG extraction, and arrhythmia detection. The conducted experiment demonstrated the ability of our proposed models to accurately estimate the ECG signal.

## Chapter 6

# Identifying the common subspaces regions using independent component analysis and graph-theoretical analysis

Blind source separation presents a promising approach for early detection of mental disorders. However, traditional BSS techniques struggle with computational efficiency when processing high-dimensional data, thereby limiting the number of datasets that can be effectively analyzed. This Chapter proposes ICAMGGMM-CS, an advanced common subspace independent component analysis model built on the multivariate generalized Gaussian mixture distribution. Leveraging this distribution enables the model to capture intricate relationships between source signals and estimated components, facilitating the identification of shared component patterns. This approach enhances diagnostic accuracy and scalability, making the model more effective across varied applications. To enhance the resilience and accuracy of source separation, this model integrates a bounded indicator function, establishing the ICABMGGMM-CS framework. This approach combines the adaptability of data-driven methods with the robustness of model-based techniques to effectively manage noise and artifacts. Furthermore, we integrate a multivariate graph-theoretical framework with our proposed models, enabling the summarization of components into meaningful features. Our results highlight the model's capability to consistently identify frequent patterns in resting-state functional

magnetic resonance imaging data, accurately localizing affected brain regions. These findings underscore the model’s potential as a powerful tool for early mental disorder detection and improved diagnosis.

## 6.1 Introduction

Identifying components as common or distinct across the datasets allows for a more comprehensive and realistic representation of the underlying relationships among them [227, 228]. Several fusion methods based on ICA have specifically addressed this aspect, either utilizing only the common components or incorporating both distinct and common components to describe the underlying relationships across multiple datasets. An example of such a method is joint ICA (jICA), which was presented by Calhoun et al. [229]. This approach has been applied in the fusion of various medical imaging modalities, including EEG and fMRI [230, 231], as well as in the integration of fMRI, structural MRI (sMRI), and EEG [232]. Furthermore, it has been employed in the fusion of fMRI data from multiple tasks [233], among other applications. The jICA approach interprets all estimated components as common and aligned across datasets, assuming a globally common mixing matrix between datasets.

Another approach, known as MCCA-jICA [234, 235], expands on canonical correlation analysis (CCA) for multiple datasets by incorporating a prior step known as multiset canonical correlation analysis (MCCA). MCCA aims to maximize the linear correlation between datasets before the jICA step. For the subsequent jICA stage, MCCA-jICA provides a better model match by aligning components that are more likely to be common across datasets. MCCA-jICA has demonstrated success in fusing fMRI, diffusion tensor imaging, and sMRI [236], as well as in the fusion of fMRI, EEG, and sMRI data [237], and the fusion of MRI data [238]. Conversely, both jICA and MCCA-jICA operate under the fundamental assumption that the majority of underlying sources in the decompositions are common across all datasets. Consequently, these methods perform poorly when certain components are not common to all datasets. To that end, a flexible method is required to consider these challenges.

In this work, we introduce the ICA multivariate generalized Gaussian mixture for common

subspace (ICAMGGMM-CS) model, designed to effectively capture and recover the correlation structure of latent variables across a large number of datasets. Unlike traditional approaches, ICAMGGMM-CS explicitly accounts for inter-dataset correlations, providing robust identification guarantees that ensure accurate estimation of these correlations. Leveraging the strengths of ICAMGGMM-CS as a framework for subgroup identification, this model excels at identifying common components that reveal the underlying structure of the datasets. Specifically, to identify this optimal number for a single ICA decomposition, the model order is fixed, and the number of datasets (subjects) is explored. The data is then divided into subsets of subjects, and ICAMGGMM-CS is performed on each subset to identify a common subspace. This forms the first stage of ICAMGGMM-CS, known as subset analysis. The common components from each subset are further compared to identify consistent components across all subjects and within subgroups, which constitutes the second stage of ICAMGGMM-CS: common subspace identification.

To overcome the limitations of unbounded BSS methods, particularly in estimating data shapes constrained within bounded support regions, we also propose a bounded version of the model: ICABMGGMM-CS. This algorithm addresses critical challenges associated with ICA, particularly in terms of computational complexity and performance within common subspace separation. ICABMGGMM-CS allows for the precise identification of common subgroups based on individual components, enabling a deeper analysis of each component's contribution to define the subgroups, thereby improving interpretability. After identifying common components within the covariance matrices, the model separates these common elements from the non-common ones. The ICABMGGMM-CS algorithm then independently processes the common and non-common subspaces, enhancing the accuracy of common subspace estimation while significantly reducing data dimensionality at the ICA stage. Furthermore, the ICABMGGMM-CS framework retains the strengths of data-driven methods, particularly their ability to handle noise and artifacts typically encountered in model-based approaches, and it efficiently leverages inter-dataset dependencies to enhance source separation performance. By incorporating second-order correlations within bounded components, this method enhances flexibility in extracting meaningful insights, providing a robust tool for identifying and analyzing complex patterns within the data.

Finally, graph theory has been applied to brain networks to examine topographical differences

between healthy individuals and patients, using features such as clustering and centrality. These features rely on similarity measures between components to assess overall efficiency and the importance of components in facilitating information transfer within functional brain networks. Since these features are calculated for each subject, variability in graph theory (GT) features reflects the variability captured in individual subject results. To this end, we integrate a multivariate GT framework into our proposed models that summarizes components into features, allowing for effective comparison of the performance of JBSS methods in capturing subject variability for group analysis of real fMRI data.

The rest of this Chapter is organized as: Section 7.3 provides the mathematical foundation and details of the proposed method. In Section 7.5, we validate our model through different experiments. Section 7.6 summarizes our findings and concludes Chapter.

## 6.2 Background

### 6.2.1 Independent component analysis for the bounded multivariate generalized Gaussian mixture model

The complete discussion for the ICABMGGMM is given in Chapter 2.

### 6.3 Proposed model: ICABMGGMM with the common subspaces framework

We assume that the source space underlying the observed data consists of three sets of source components. We express the ICA function as:

$$\begin{aligned}
 \mathbf{x} &= \mathbf{A}_j \mathbf{s}_j \\
 &= \mathbf{A}_{j_C} \mathbf{s}_{j_C} + \mathbf{A}_{j_G} \mathbf{s}_{j_G} + \mathbf{A}_{j_D} \mathbf{s}_{j_D} \\
 &= \begin{bmatrix} \mathbf{A}_{j_C} & \mathbf{A}_{j_G} & \mathbf{s}_{j_D} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{j_C} \\ \mathbf{s}_{j_G} \\ \mathbf{s}_{j_D} \end{bmatrix}
 \end{aligned} \tag{6.1}$$

where  $\mathbf{A}_{j_C}$ ,  $\mathbf{A}_{j_G}$ , and  $\mathbf{A}_{j_D}$  denote the respective mixing matrices for common sources, group-specific sources, and distinct sources. To develop the (ICABMGGMM-CS), we adopt Equation (2.4) in Equation (6.1). This advanced method comprises two main steps: subset analysis and common subspace identification. To address the challenge of high dimensionality, the subset analysis step involves performing multiple independent ICABMGGMM decompositions on randomly selected subsets of subjects from the overall population. During the common subspace identification step, both group-specific and distinct source components are identified, along with a common subspace that remains consistent across the entire population. These steps are illustrated in Fig. 6.1. Further details on the ICABMGGMM-CS method are provided in the following subsections.

### 6.3.1 ICABMGGMM-CS: subset analysis

The general schematic illustrating the ICABMGGMM-CS process is shown in Fig. 6.1. During the subset analysis phase, groups of  $M$  subjects are randomly sampled from the entire population, and a  $M \times M$  correlation matrix is computed for each source component, yielding  $M(M - 1)/2$  unique correlation values. These values are used to evaluate the similarity of components, allowing us to identify whether a source component is common across the  $M$  datasets. For source components representing a common component, we anticipate consistently high correlation values. The degree of "commonality" for an source component is quantified by the ratio of correlation values exceeding an empirically determined threshold,  $\gamma$ , to the total number of correlation values for that source component.

$$Q = \frac{N_\gamma}{M(M - 1)/2} \quad (6.2)$$

where  $N_\gamma$  denotes the number of correlation values in the source component's correlation matrix that exceed the threshold  $\gamma$ . An source component corresponding to a common component is expected to exhibit high commonality, with  $Q$  values approaching 1. Within each subset, source components are sorted in descending order based on their mean correlation values, ensuring that source components with high source correlations are prioritized at the top, while those with lower correlations are positioned at the bottom.

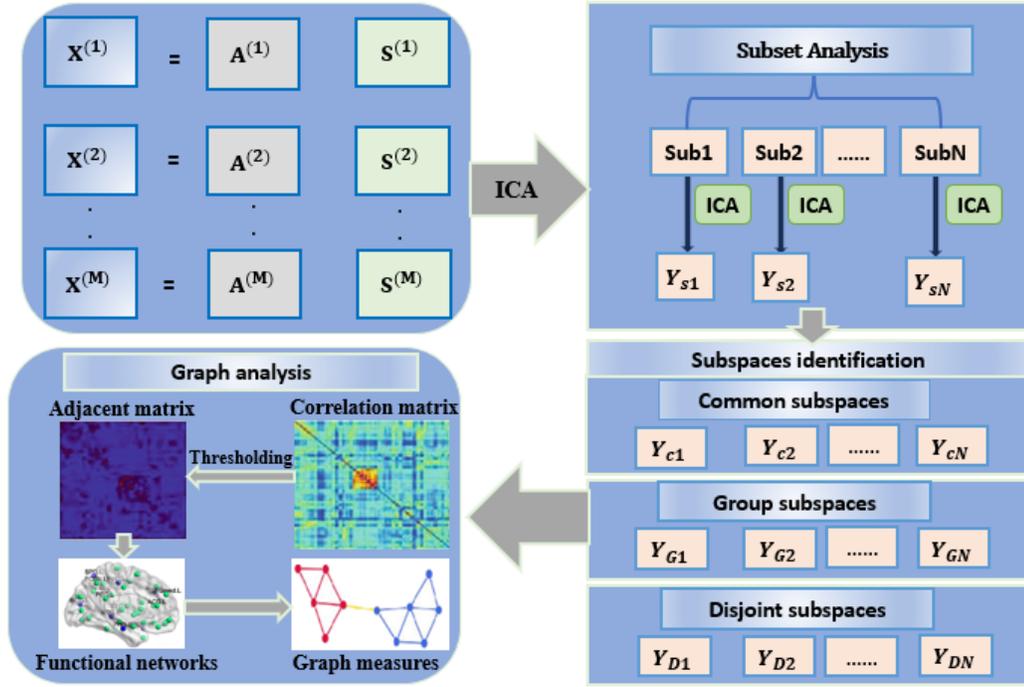


Figure 6.1: ICABMGMM-CS framework. The first step is to apply the ICABMGMM on the multi-subjects data. Then, the identification for the common subgroups is conducted.

### 6.3.2 ICABMGMM-CS: common subspace identification

The second phase aims to identify a consistent common subspace, denoted as  $Y_C$ , shared by all subjects within a group. To enable consistent comparison across  $S$  subsets, we normalize the number of components in each subspace  $Y_{sl}$  to a common value,  $C$ , across all subsets. These  $C$  source components are then considered as potential candidates for constructing the common subspace, where the specific value of  $C$  is set as the maximum among the  $C_s$  values to ensure the inclusion of the highest number of candidates. For each subset  $s$ , the  $M$  components are averaged to obtain a mean component  $\mathbf{y}_{cs}$ . The cross-correlation for each mean component is subsequently calculated by taking the average correlation between that component and its corresponding components across the remaining  $S - 1$  subsets.

$$p_{cs} = \frac{1}{S-1} \sum_{l \neq s} |p_{csl}| \quad (6.3)$$

The Pearson correlation coefficient between the mean components  $\mathbf{y}_{cs}$  and  $\mathbf{y}_{cl}$  is represented as  $p_{csl}$ , where  $1 \leq l \leq S$  and  $l \neq s$ . The cross-correlation of each subset is obtained by averaging the correlations of its  $C$  mean components.

$$p_s = \frac{1}{C} \sum_c p_{cs} \quad (6.4)$$

Subset  $s_0$ , which exhibits the highest cross-correlation  $p_s$  and thus the most consistent components, is selected to define the common subspace across all subjects. The complete learning procedure for the ICABMGMM-CS model is detailed in Algorithm 6.

---

**Algorithm 6** The common subspace ICABMGMM algorithm

---

- 1: **Input:** Dataset  $\mathcal{X}$
  - 2: Apply K-means clustering to initialize the means and covariances for each component.
  - 3: Set initial value of  $\beta_j = 0.5$  for each component.
  - 4: Randomly initialize de-mixing matrix  $\mathbf{W}$
  - 5: **Repeat until convergence:**
  - 6: **for**  $n=1, \dots, N$  **do**
  - 7:     Compute:  $\mathbf{y}_n = \mathbf{W}\mathbf{x}_n$
  - 8:     **Expectation Maximization:** Repeat until the change in likelihood  $>$  threshold
  - 9:     **E-step:** Compute posterior probabilities using Equation (2.8)
  - 10:     **M-step:**
  - 11:         Update the mixing parameter  $\mathbf{p}_j$  using Equation (2.10)
  - 12:         Update mean  $\boldsymbol{\mu}_j$  using Equation (2.14)
  - 13:         Update covariance  $\Sigma_j$  using Equation (2.17)
  - 14:         Update shape parameter  $\beta_j$  using Equations (2.18), (2.19)
  - 15:         Update the basis functions  $A_j$  using Equation (2.36)
  - 16:         Update the bias vector  $b_j$  using Equation (2.38)
  - 17:     Compute:  $\mathbf{y}_n = \mathbf{W}\mathbf{x}_n$
  - 18:     **if**  $Q > \gamma$  **then**
  - 19:         Assign  $\mathbf{y}_n$  to the common subspace
  - 20:     **else**
  - 21:         Assign  $\mathbf{y}_n$  to a distinct subspace
  - 22:     **end if**
  - 23: **end for**
  - 24: **Return:** The model parameters  $\hat{\boldsymbol{\mu}}_j, \hat{\beta}_j, \hat{\Sigma}_j$ , and common components  $Y_C$
- 

## 6.4 Results and discussion

In this work, we apply our proposed models into two distinct neuroscience applications. Firstly, we utilize the COBRE Schizophrenia dataset to assess the effectiveness of our model in identifying

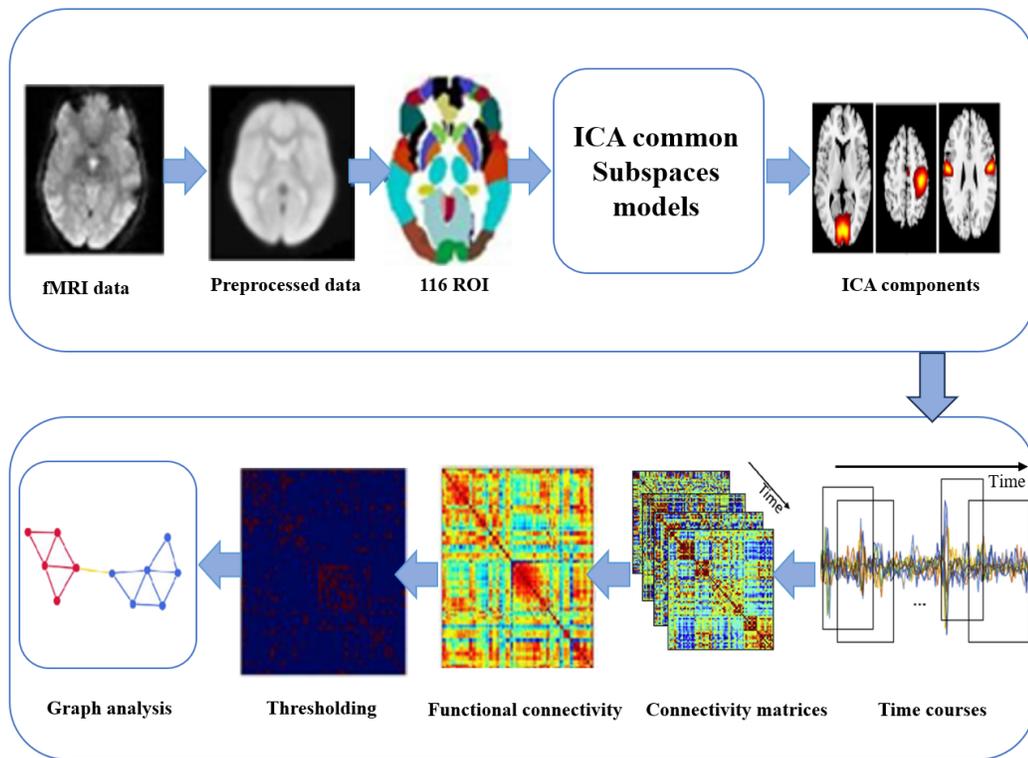


Figure 6.2: The general block diagram for the ICA common subspaces approaches.

brain regions impacted by Schizophrenia compared to healthy controls. Then, to evaluate the generalizability of our model, we further test it on the ADNI dataset, distinguishing between healthy subjects and those with Alzheimer’s disease.

### 6.4.1 Schizophrenia detection using the common subspaces ICABMGGMM

#### Data acquisition and preprocessing

The dataset utilized in this study is the COBRE dataset, which is comprehensively detailed in Subsection 4.5.4. The corresponding data processing steps are thoroughly outlined in Subsection 2.6.4.

#### Evaluation analysis for the common subspace ICABMGGMM

Following the preprocessing stage, the brain was parcellated into 116 regions using the AAL atlas. Then, we applied our models to extract components, time courses, and connectivity matrices

from the data. The average intensity time series for each ROI was then computed and subjected to a band-pass filter with a frequency range of 0.01 Hz to 0.08 Hz. The resulting connectivity matrices were used to identify components influenced by Schizophrenia. These steps for our processing pipeline are illustrated in Fig. 6.2.

After the parcellation step, we apply our proposed ICAMGGMM-CS and ICABMGGMM-CS methods to different subgroups of the large-scale resting-state fMRI dataset. To initialize the model parameters, including the means ( $\mu_j$ ), covariances ( $\Sigma_j$ ), and shape parameters ( $\beta_j$ ), we first employ a K-means clustering approach. Subsequently, the EM algorithm is utilized to compute the posterior probabilities of the common subspace and iteratively update the model parameters. This process is repeated until convergence, at which point the resulting demixing matrix is used to differentiate Schizophrenia patients from healthy controls. The time courses extracted from the ICA common subspace are then used to generate connectivity matrices, allowing us to identify homogeneous subgroups of brain regions affected by Schizophrenia.

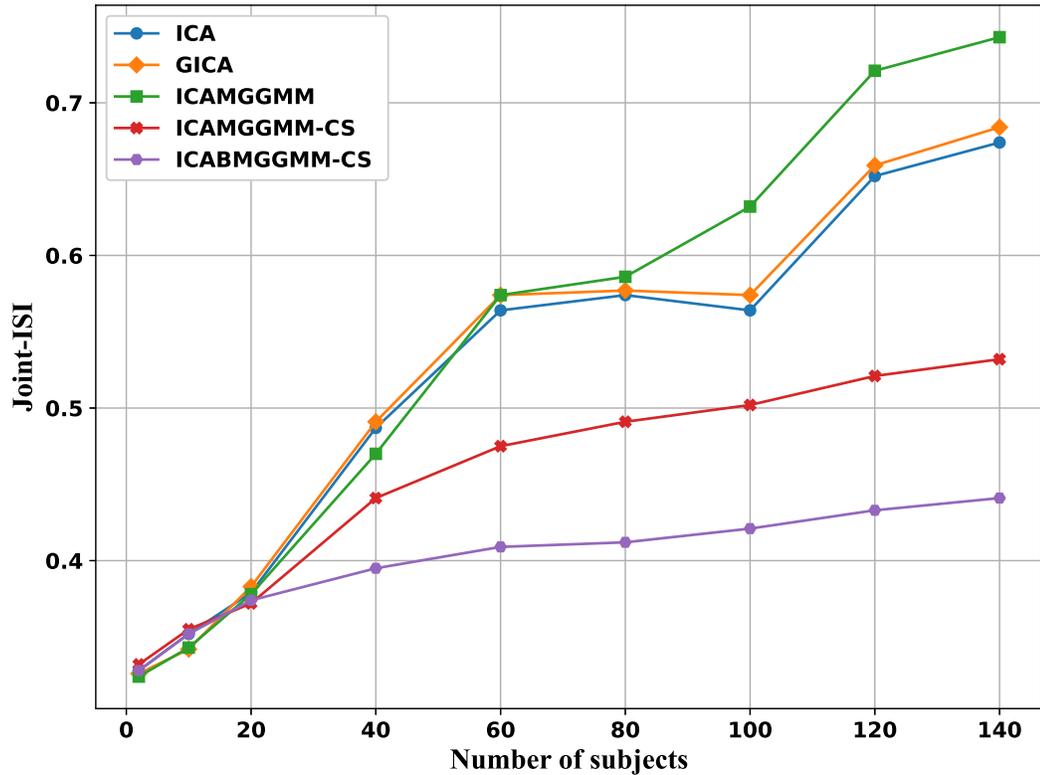


Figure 6.3: The proposed ICABMGGMM-CS and comparing models performance based on the Joint-ISI measure.

To assess the effectiveness of the proposed methods, we utilize BSS performance metrics, including Joint-ISI and spatial correlation, as shown in Figs. 6.3 and 6.4, respectively. In terms of Joint-ISI, our proposed ICAMGGMM-CS and ICABMGGMM-CS frameworks consistently outperform other BSS methods. Joint-ISI evaluates the quality of the separated components for both Schizophrenic and healthy subjects, with lower values indicating reduced interference between the separated components.

The significant improvement of our proposed methods is particularly evident in complex scenarios, such as when separating data from more than 80 subjects, where they exhibit lower Joint-ISI values compared to existing BSS approaches, as illustrated in Fig. 6.3. Similarly, in terms of spa-

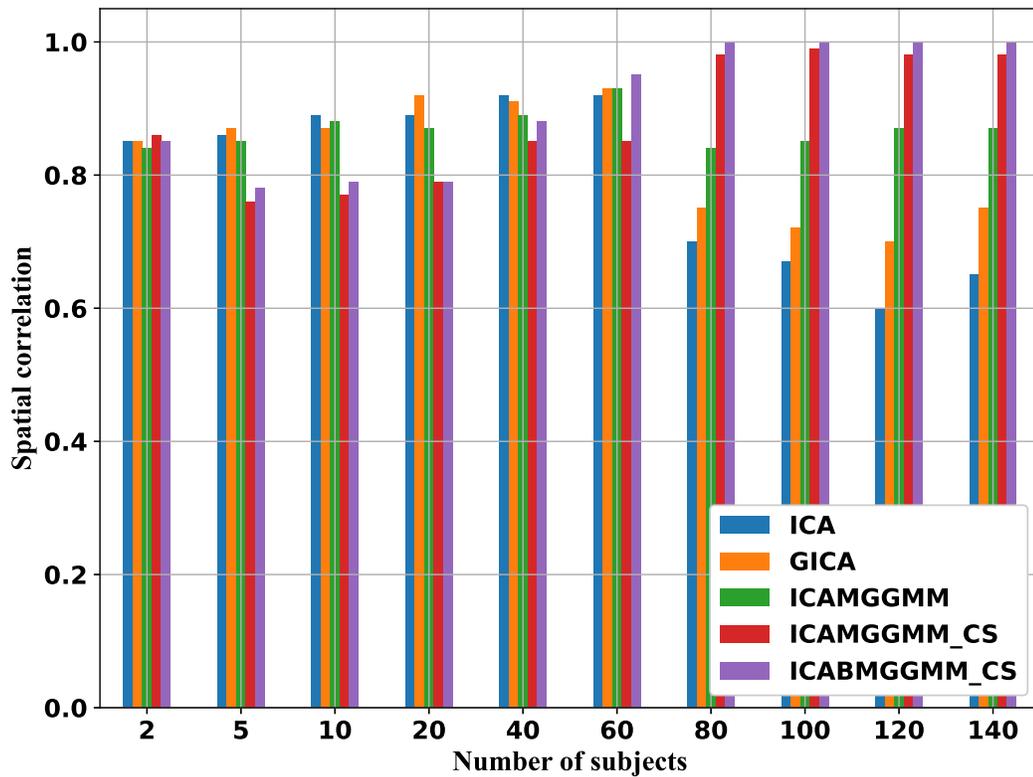


Figure 6.4: The performance of proposed ICABMGGMM-CS and comparing models based on Spatial correlation.

tial correlation, our proposed methods demonstrate superior performance across all evaluated cases, maintaining higher accuracy in capturing spatial patterns. This improvement is especially pronounced in high-dimensional situations, such as when analyzing data from more than 80 subjects, where our models exhibit significant advantages over the compared approaches.

## Functional network identification for Schizophrenia individuals

To evaluate the robustness of our proposed algorithms, we identify brain regions affected by Schizophrenia using the most discriminative ICA components with high correlation coefficients. Specifically, we select the top 100 ICA components, characterized as intrinsic connectivity networks, and divide each class into seven random subgroups of 70 subjects. For each subgroup, there are  $70 \times \frac{(70-1)}{2} = 2415$  unique correlation values. The extracted components are then ranked by their mean correlation values, from highest to lowest. A two-sample t-test is used to produce the discriminative connection maps, which statistically validated the differences between healthy controls and those with Schizophrenia. The FDR technique is used for multiple comparisons, with a corrected significance level of  $p < 0.05$  (minimum  $z > 3.2$ , cluster significance:  $p < 0.05$ ). Table 6.1 presents a comprehensive summary of brain networks, detailing ICA components and their corresponding coordinates. Each network is associated with specific brain regions that align with their recognized functional roles. These networks include the Default Mode network, Basal Ganglia network, Auditory network, Sensorimotor network, and Working Memory network. The DMN spans multiple brain regions, reflecting its broad involvement in various cognitive processes, while networks such as the SMN and WMN demonstrate more localized coordinates, emphasizing their specialized functions. Figs. 6.5 and 6.6 illustrate the results of the ICABMGGMM-CS component group comparison for Schizophrenia patients, highlighting the distinct network activation patterns observed in these Schizophrenic patients. As illustrated in Fig. 6.5, the ICABMGGMM-CS approach reveals significant brain network structures displayed in the right column in comparison to the ICAMGGMM-CS components shown in the left column. This difference highlights the effectiveness of the ICABMGGMM-CS method in separating the components associated with Schizophrenia. Compared to the healthy control group, the Schizophrenia group exhibit increased connectivity across interesting common networks. Notably, regions such as the thalamus, superior frontal gyrus, inferior parietal gyrus, insula, superior temporal gyrus, caudate, and postcentral gyrus show stronger functional connectivity within a common subspace. These regions are particularly relevant to the most discriminative ICA components within networks such as the DMN, BGN, AN,

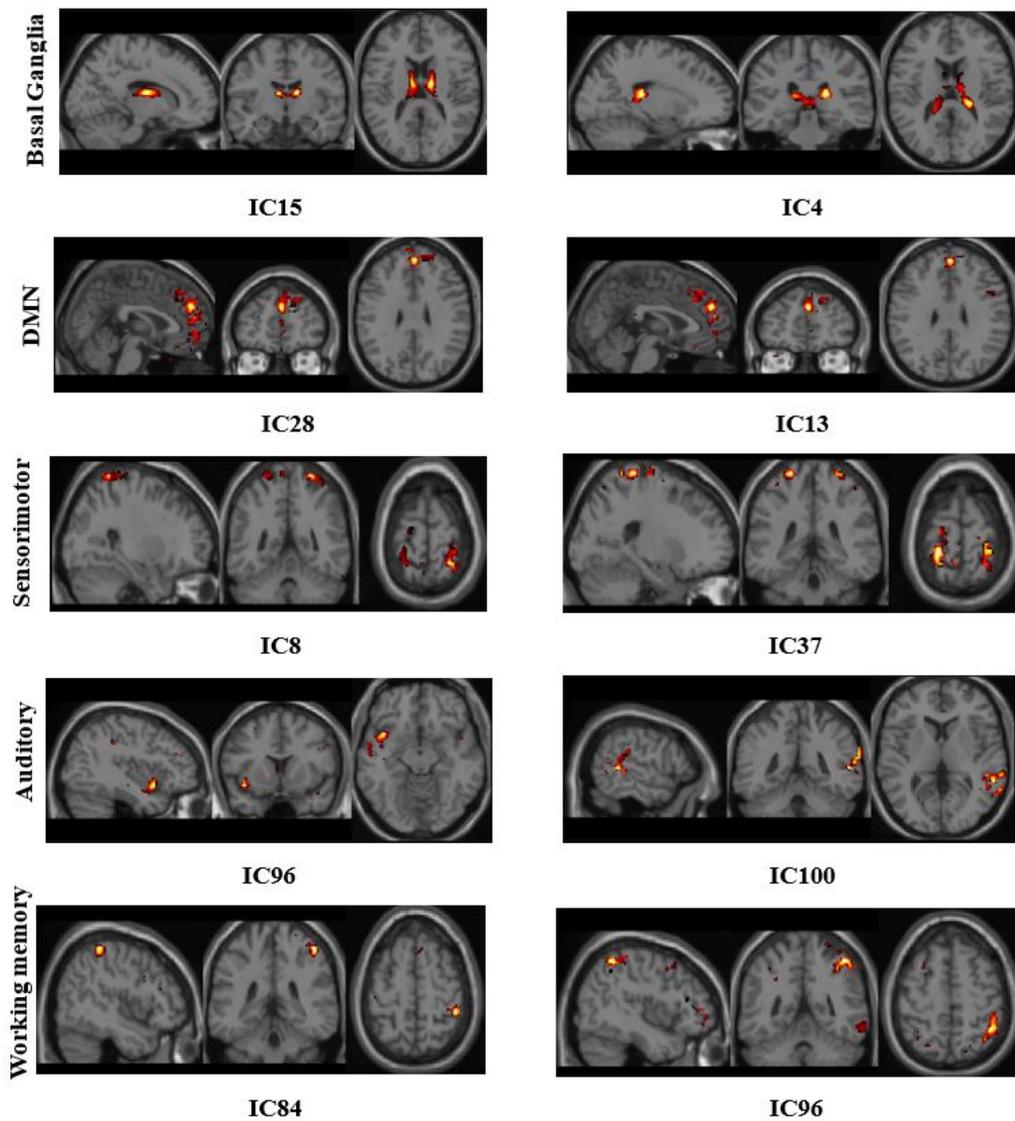


Figure 6.5: The spatial maps of the common subspace ICA components for Schizophrenic patients. The ICAMGGMM-CS components are displayed in the left column, while the ICABMGGMM-CS components are illustrated in the right column.

SMN, and WMN. Previous studies [196, 190] have also implicated abnormalities in these brain regions as contributors to the pathophysiology of Schizophrenia. Our findings align with these studies, reinforcing the role of disrupted connectivity in these areas in the manifestation of Schizophrenia. Our findings also indicate that the precuneus, supplementary motor region, superior parietal gyrus, and middle temporal gyrus may be impacted in the early stages of Schizophrenia. Numerous investigations have found significant abnormalities in these brain regions in subjects with Schizophrenia

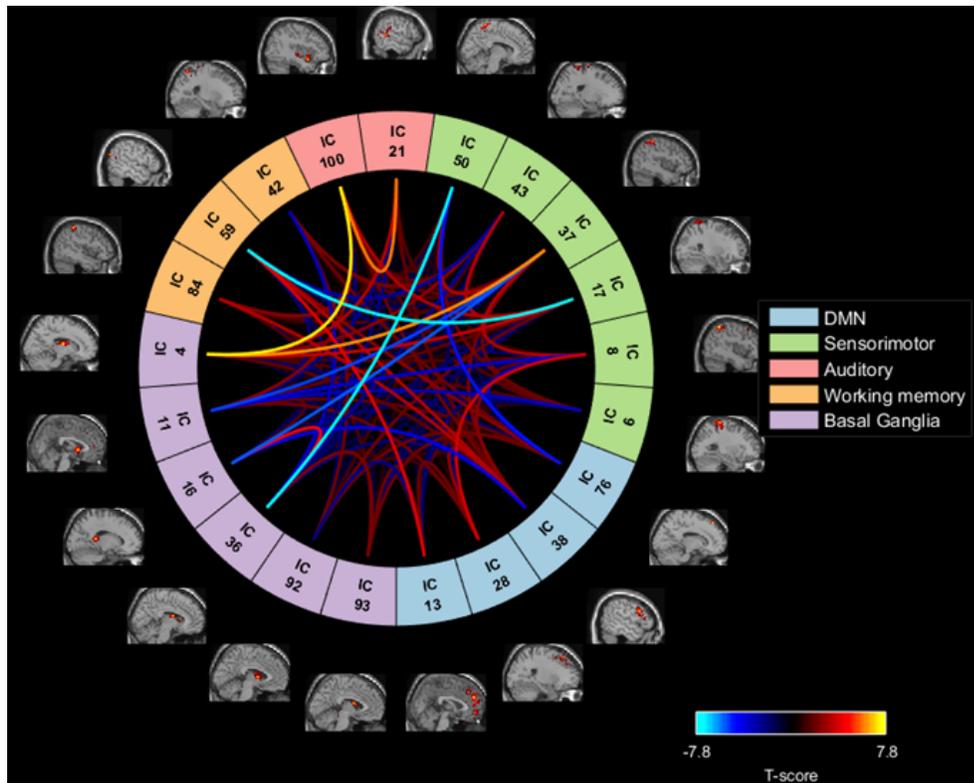


Figure 6.6: The connectogram map of the common subspace ICA components for Schizophrenic patients.

[191, 190, 239, 240].

The two-sample t-test conducted on the functional network connectivity validation reveals that the differences in clustering between healthy controls (HC) and Schizophrenic (SZ) components were statistically significant, confirming the robustness of our proposed BSS framework in improving clustering performance. We conduct a comprehensive graphical analysis using graph-theoretic features such as clustering coefficient, local efficiency, centrality, and small-worldness to assess structural differences. Fig. 6.7 highlights major distinctions in graph elements between SZ and HC subjects. In brain networks, SZ individuals consistently show lower clustering coefficients compared to HC subjects across a threshold range of 0.1 to 0.4. This trend extends to other graph metrics, underscoring significant network alterations associated with SZ. These findings underscore the effectiveness of ICABMGGMM-CS in emphasizing meaningful distinctions, making it a promising method for classification analysis. Its capacity to detect changes in graph-theoretical characteristics further strengthens its potential in distinguishing between HC and SZ components, offering valuable

Table 6.1: ICA components and their coordinates for different brain networks.

Brain Network	ICs	Brain Regions	Abbreviation	Hemisphere	Coordinates	Peak Coordinates
Basal Ganglia network	4	Thalamus	THA	L	-9, -6, 13	12, -10, 16
				R	8, -6, 12	
	11	Caudate	CAU	R	6, 14, 2	0, 8, 0
	15	Thalamus	THA	R	6, -27, 3	-4, -30, 4
	16	Thalamus	THA	L	-5, -9, 13	-18, -36, 12
				R	10, -7, 13	
	36	Caudate	CAU	R	10, 2, 18	6, -4, 18
			L	-8, 2, 16		
92	Caudate	CAU	L	-14, 16, 5	-6, 0, 6	
93	Thalamus	THA	R	7, -7, 16	-6, 0, 16	
Sensorimotor network	6	Postcentral gyrus	PoCG	L	-30, -30, 64	-30, -30, 68
	8	Inferior parietal gyrus excluding supramarginal and angular gyri	IPG	R	42, -50, 56	42, -50, 56
				L	40, -56, 56	
	17	Postcentral gyrus	PoCG	R	26, -35, 74	26, -48, 72
				L	-27, -35, 72	
	37	Postcentral gyrus	PoCG	R	37, -36, 54	42, -36, 60
				L	-46, -36, 50	
	43	Superior parietal gyrus	SPG	L	-22, -48, 72	-22, -48, 72
				R	-31, -36, 72	
	50	Precuneus	PCUN	R	24, -38, 72	6, -44, 68
				L	6, -44, 68	
Default mode network	9	Supplementary motor area	SMA	L	0, -6, 66	0, -6, 66
				R	7, -8, 67	
				L	-9, -46, 68	
		Superior frontal gyrus dorsolateral	SFG	L	-20, -6, 67	
				R	-22, -11, 67	
	13	Superior frontal gyrus medial	SFGmedial	L	0, 48, 30	0, 48, 30
				R	0, 52, 14	
	19	Supplementary motor area	SMA	L	-12, 4, 72	-12, 4, 72
	28	Superior frontal gyrus, dorsolateral	SFG	R	24, 40, 54	22, 32, 54
	76	Superior frontal gyrus, medial	SFGmedial	R	12, 40, 54	12, 40, 54
L				-14, 37, 54		
Auditory network	21	Middle temporal gyrus	MTG	R	56, -44, 6	56, -44, 6
				L	67, -44, 24	
	100	Temporal pole: superior temporal gyrus	TPOsup	L	-42, 12, -18	-42, 12, -18
				R	43, 15, -18	
Working memory network	42	Superior parietal gyrus	SPG	L	-26, -64, 60	-26, -64, 60
	96	Insula	INS	L	-54, -72, 24	-54, -72, 24
				R	-47, 8, 12	
84	Postcentral gyrus	PoCG	R	44, -38, 60	44, -38, 60	

insights into the neurobiological underpinnings of Schizophrenia.

Our findings demonstrate that the ICABMGGMM-CS approach identifies a greater number of significant components in global nodal graph metrics, particularly in characteristic path length (PL) and betweenness centrality. This higher consistency in significant findings indicates that ICABMGGMM-CS offers enhanced sensitivity in capturing network properties that differentiate patient groups from control ones. Fig. 6.8 provides a visual representation of the number of significant components identified across a range of link densities for both PL and betweenness centrality. As seen in the left panel (PL), ICABMGGMM-CS consistently identifies more significant nodes as link density increases, surpassing other models, including ICA and GICA. This pattern is also evident in the right panel (betweenness centrality), where ICABMGGMM-CS maintains a high number of significant nodes even as link density varies. The stability of these findings across multiple link densities underscores the robustness of ICABMGGMM-CS in identifying meaningful differences

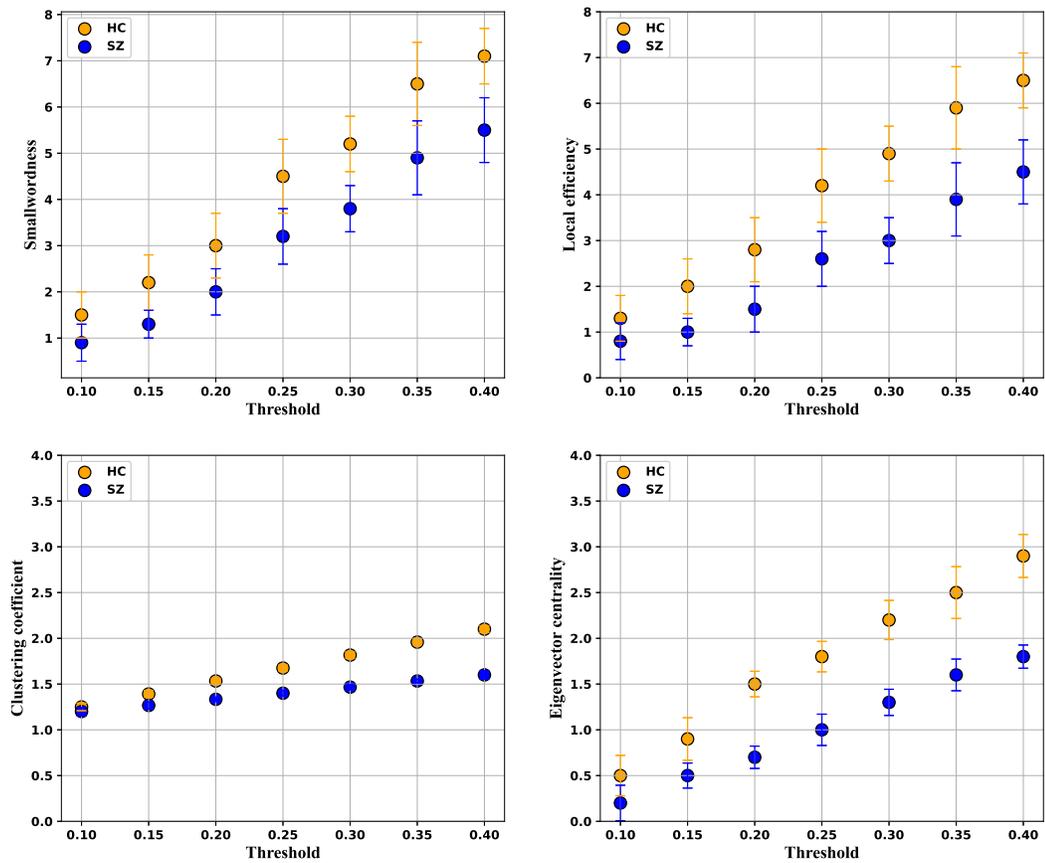


Figure 6.7: The graph analysis for the separated ICABMGMM-CS components.

in network structure between groups.

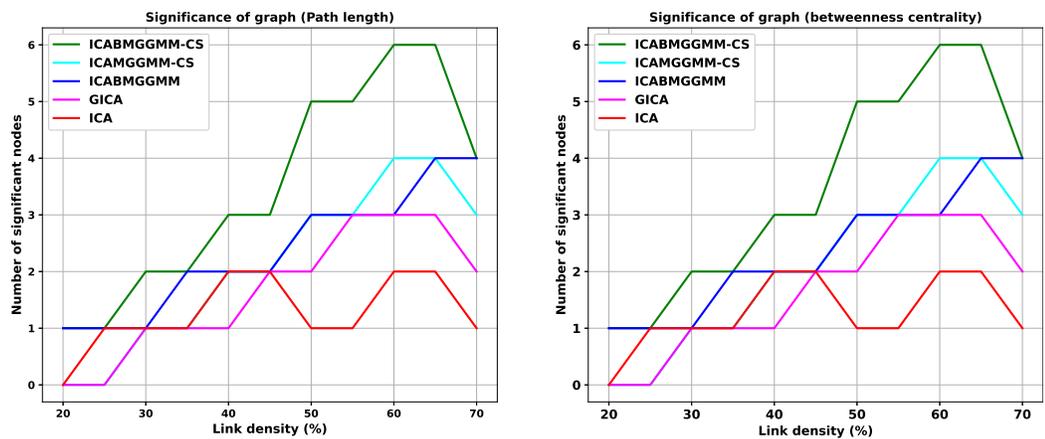


Figure 6.8: Graph analysis for the ICABMGMM approach conducted on COBRE data using path length and betweenness centrality.

To quantify the effectiveness of each model in identifying relevant brain regions, we use the

two-sample t-test to ROC curves and AUC for each network component. The ROC curves allow us to examine the trade-off between detection power and false alarm probability at various thresholds. For each ROC curve, seven subgroups are randomly chosen from patient components, and the average z-maps are calculated for each group. We then apply a mask to these averaged z-maps at varying z-value thresholds, constructing ROC curves based on the response of active voxels within each threshold range. It is observed that some ICA components, like the major motor component, exhibited higher voxel intensities at lower z-value thresholds, suggesting that even small thresholds yield significant voxel activity. This finding highlights that the location and density of active voxels may play a more crucial role in ROC analysis than merely focusing on high voxel values alone.

Our findings, as shown in Fig. 6.9 reveal that the ICABMGGMM-CS model consistently produced components with superior alignment to the specified anatomical regions described by the AAL atlas, as demonstrated by higher AUC values. Moreover, statistical analysis confirmed that the AUC differences between ICABMGGMM-CS and other models are significant, underscoring the robustness and reliability of ICABMGGMM-CS in capturing meaningful brain network activity. This analysis indicates that ICABMGGMM-CS outperforms other methods in accurately identifying brain regions across multiple network components, making it a promising tool for applications that require precise region-specific analysis, such as in mental health diagnostics and brain function studies.

## **6.4.2 Alzheimer diagnosis using the common subspaces ICABMGGMM**

### **Dataset and preprocessing**

This study employs the ADNI dataset, with a detailed description provided in subsection 4.5.2. The data preprocessing steps are outlined in subsection 4.5.2.

### **Evaluation analysis for the common subspace ICABMGGMM**

To assess the generalizability of our proposed approaches, we apply them to Alzheimer’s diagnosis tasks using data from ADNI, which comprises 120 participants, including individuals with AD and HC. As illustrated in Figs. 6.10 and 6.11, our methods consistently outperform the baseline

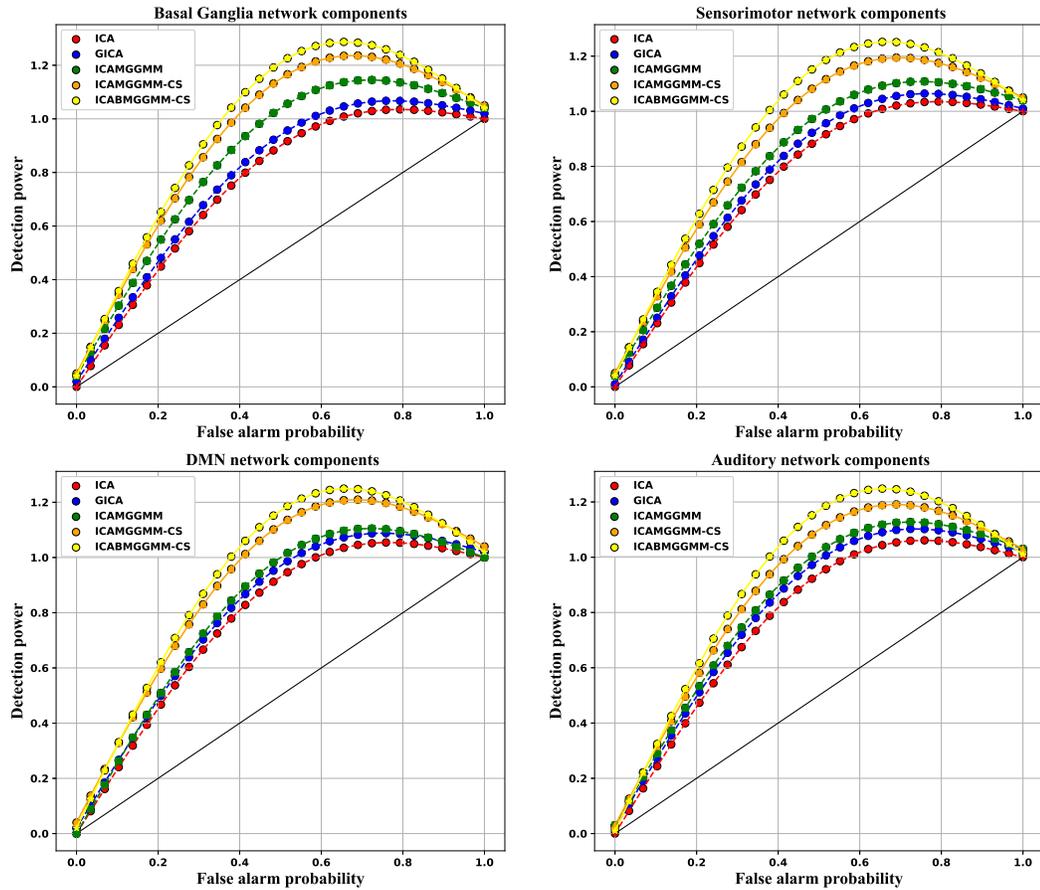


Figure 6.9: ROC curves for ICABMGGMM-CS and compared models based on the resultant from two-sample t-test.

BSS models across various combinations of subjects. The proposed BSS frameworks demonstrate superior performance, particularly in terms of Joint-ISI and spatial correlation metrics. Notably, this improvement is more pronounced when dealing with larger-scale data, involving over 80 subjects. This highlights the limitations of traditional models with simpler distribution assumptions, whereas our approaches offer enhanced flexibility in handling the complex structures inherent in the data.

### Functional network identification for Alzheimer individuals

To evaluate the robustness of our proposed models, we identify brain regions affected by Alzheimer using the most discriminative ICA components with high correlation coefficients. Specifically, we select the top 100 ICA components, and divide each class into six random subgroups of 60 subjects.

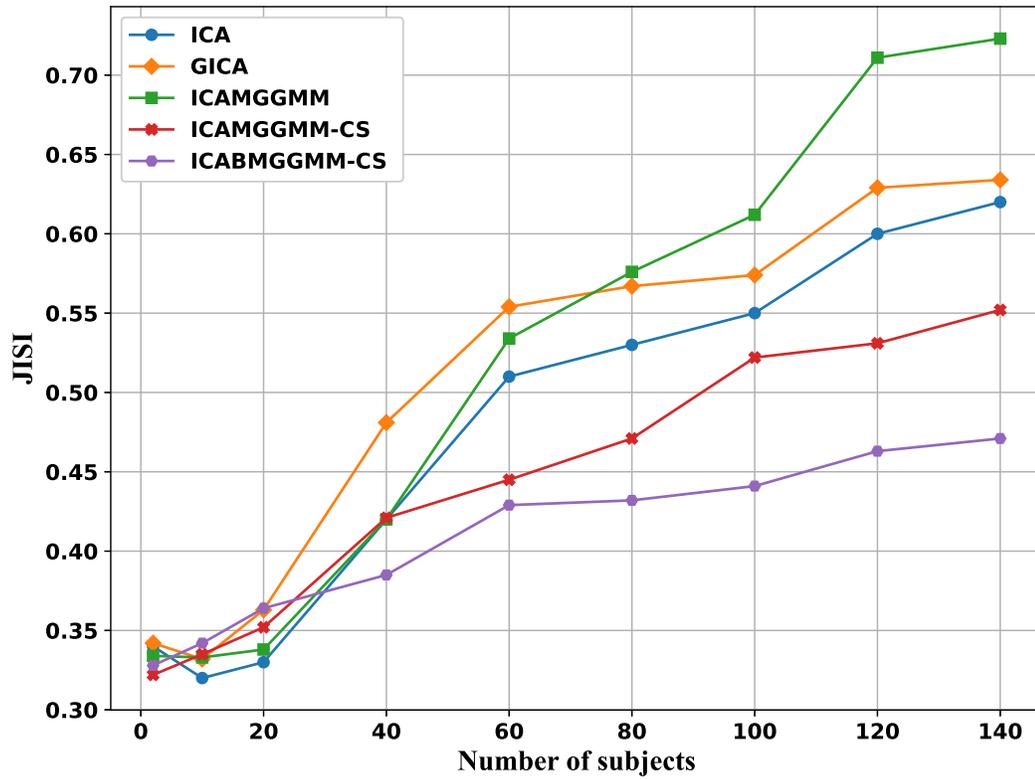


Figure 6.10: Joint ISI metric for proposed and compared models.

For each subgroup, there are  $60 \times \frac{(60-1)}{2} = 1770$  unique correlation values. The extracted components are ranked by their mean correlation values, from highest to lowest. A two-sample t-test conducted on functional network connectivity validation reveals statistically significant differences in clustering between HC and AD components, underscoring the robustness of our proposed BSS framework for enhanced clustering performance. Figs. 6.12 and 6.13 illustrate the most discriminative ICA components along with their corresponding brain networks. These networks include the Visual network, Frontal network, SMN, ADN, BGN, and Cerebellum. The ICABMGGMM-CS components effectively highlight brain regions impacted by Alzheimer’s disease, providing insight into the neural underpinnings of the condition. The most discriminative ICA components identified are IC6, IC27, IC14, IC18, IC58, and IC91. Our findings indicate disruptions in the middle occipital gyrus, superior frontal gyrus (dorsolateral), and superior frontal gyrus (medial). These results align with previous research, which has consistently reported significant reductions in grey and white matter volume within the superior frontal gyrus in Alzheimer’s disease patients [241, 242, 243].

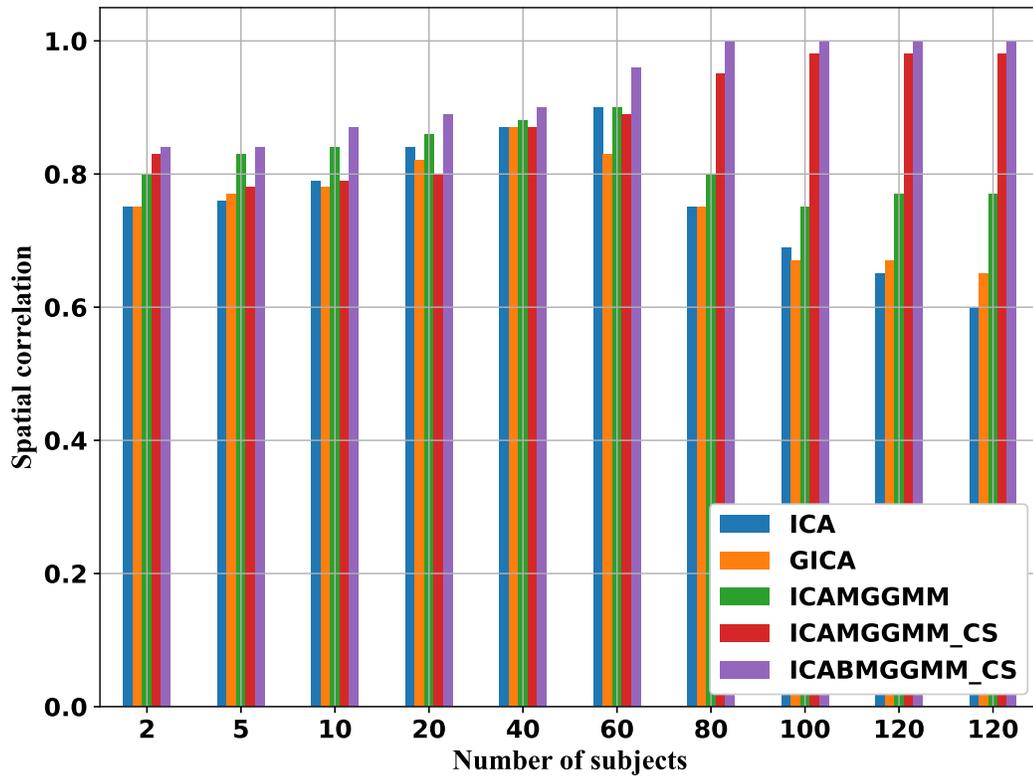


Figure 6.11: Spatial correlation measure for proposed and compared models.

Previous studies have identified abnormalities in the inferior frontal gyrus, superior parietal gyrus, and precuneus in AD [244, 242, 245]; however, the involvement of these regions in early-stage AD remains unclear. Our findings reveal significant differences in these areas between AD patients and healthy controls, supporting their potential relevance in AD pathology. Additionally, we observed significant distinctions in other brain regions between AD and HC subjects, including the middle occipital gyrus [242, 246], postcentral gyrus [243, 242], calcarine cortex [246, 243], and left thalamus [247, 241]. These regions are shown in detail in Table 6.2.

Following the two-sample t-test conducted on the retrieved ICA common components, we perform a comprehensive graphical analysis, including clustering coefficient, local efficiency, centrality, and small-worldness, to assess structural differences. Fig. 6.14 presents the primary distinctions in graph features between AD and HC participants. Notably, the local efficiency for AD participants is consistently lower than that of HC subjects across a threshold range of 0.1 to 0.4 in brain networks. This pattern is similarly reflected across the other graph metrics, indicating pronounced network alterations associated with AD.

Table 6.2: Brain network components and associated anatomical regions.

Brain Network	IC Component	Anatomical Regions	Abbreviation	Hemisphere	Coordinates (X, Y, Z)
Visual network	6	Middle occipital gyrus	MOG	R	30, -86, 6
Frontal network	27	Superior frontal gyrus, dorsolateral	SFG	L	-12, 58, 38
		Superior frontal gyrus, medial	SFGmedial	R	9, 61, 38
		Superior frontal gyrus, medial	SFGmedial	L	-6, 61, 38
Sensorimotor network	14	Superior parietal gyrus	SPG	L	-24, -46, 66
		Postcentral gyrus	PoCG	L	-32, -41, 66
		Precuneus	PCUN	R	9, -55, 66
Auditory network	18	Inferior frontal gyrus, opercular part	IFGoperc	R	50, 18, 30
		Middle occipital gyrus	MOG	R	47, -67, 30
Basal Ganglia network	58	Thalamus	THA	L	-8, -14, 20
Cerebellum network	91	Calcarine	CAL	R	8, -70, 12

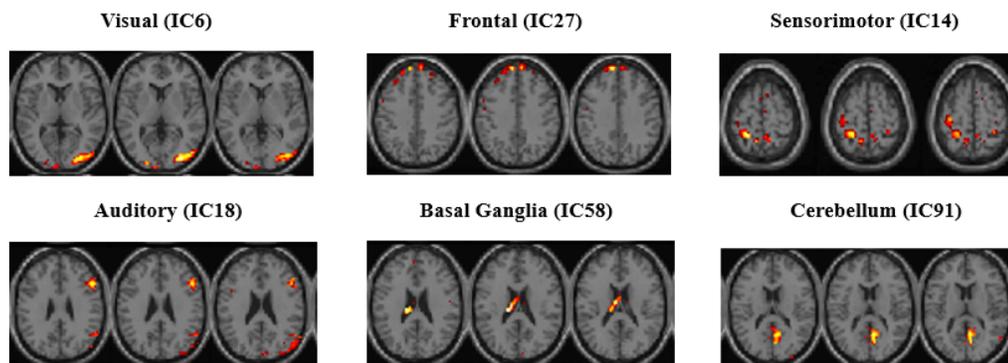


Figure 6.12: The spatial maps of the common subspace ICA components for Alzheimer patients.

Fig. 6.15 demonstrates the performance of the ICABMGGMM-CS model in identifying significant differences between HC and AD patients through the analysis of graph metrics. The results indicate a statistically significant distinction at each tested edge density level, with all FDR-adjusted p-values for the ICABMGGMM-CS model falling below 0.05. The left panel of Fig. 6.15, which shows the significance of nodes based on characteristic path length, highlights that ICABMGGMM-CS consistently identifies a greater number of significant nodes compared to other models, particularly as link density increases. This indicates that ICABMGGMM-CS is more effective at capturing variations in network connectivity that may be associated with cognitive decline in AD. Similarly, the right panel, which examines betweenness centrality, indicates that ICABMGGMM-CS achieves higher node significance across link densities, further supporting its capacity to detect central nodes that could signify important information flow disruptions in AD. These findings position ICABMGGMM-CS as a highly effective tool for classification analysis in clinical research.

To assess the robustness of our proposed models, we conduct a comprehensive ROC curve analysis, as shown in Fig. 6.16, using data from four major brain networks: the BGN, SMN, DMN, and

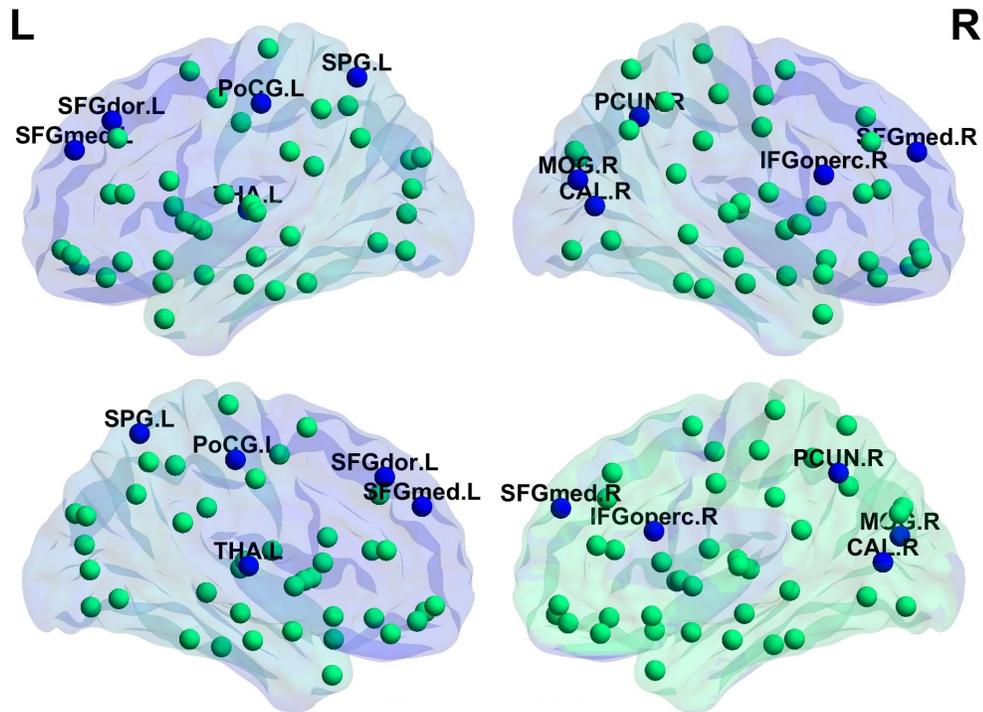


Figure 6.13: Brain regions localization. The blue color represents the affected regions due to Alzheimer whereas the green color indicates the normal regions.

AN. The results highlight that our ICABMGGMM-CS model consistently achieves higher detection power across all examined networks. Specifically, the AUC values for the ICABMGGMM-CS model are significantly greater than those of the other models, indicating an enhanced ability to capture meaningful differences within these brain networks. This increased detection power shows that ICABMGGMM-CS is particularly effective at identifying minor variations in functional connectivity patterns, which are often essential for diagnosing and understanding brain-related disorders. This robustness makes the ICABMGGMM-CS model a valuable tool for large-scale neuroimaging studies.

## 6.5 Conclusion

In this Chapter, we presented the common subspaces ICAMGGMM model, designed to capture shared components within resting-state data. Building on this approach, we further introduced the

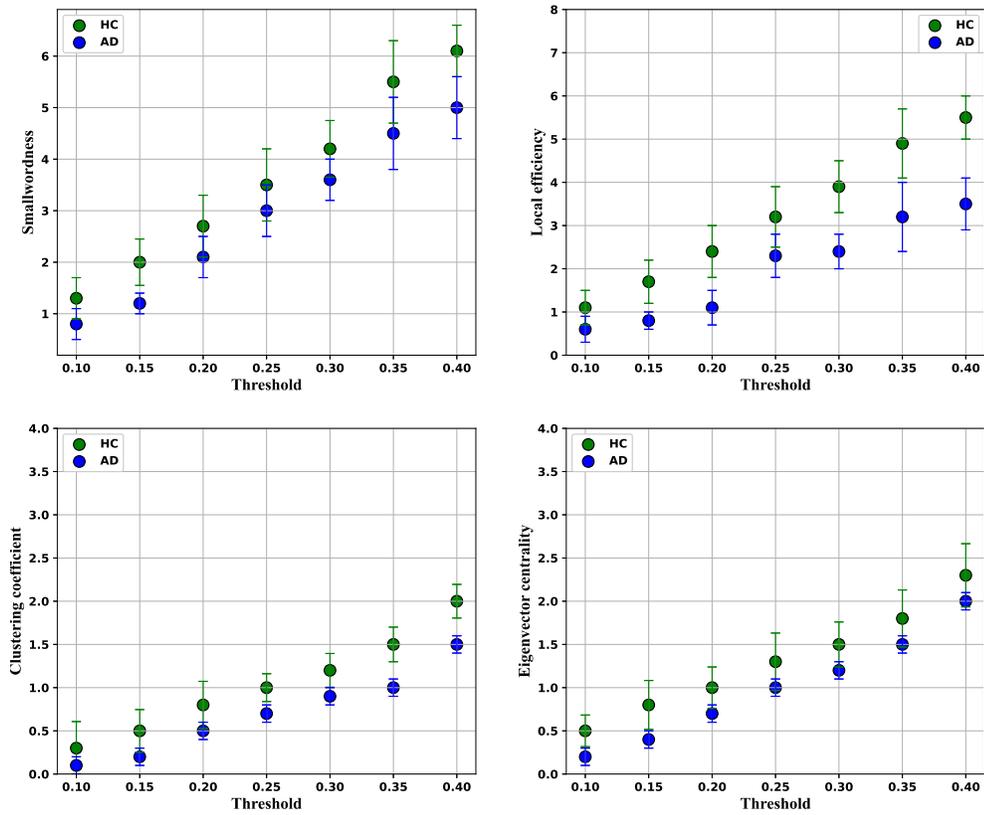


Figure 6.14: Graph analysis for our proposed ICABMGGMM-CS.

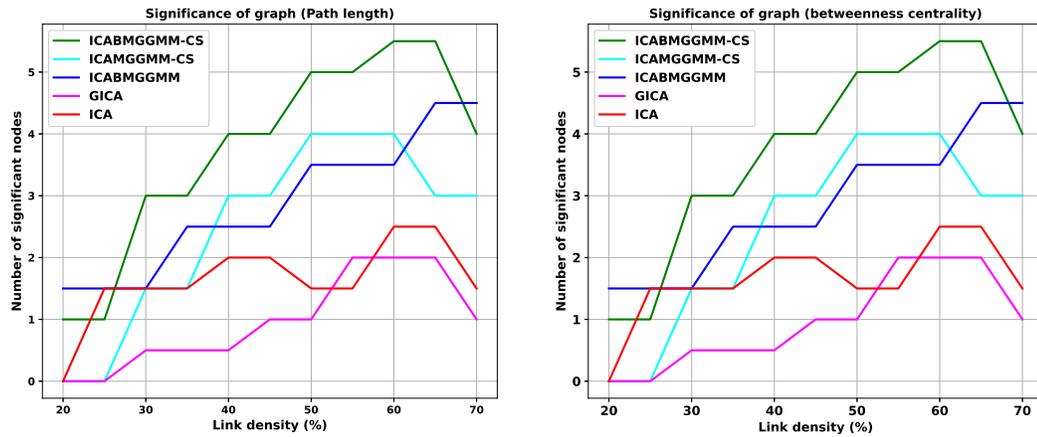


Figure 6.15: Graph analysis for the ICABMGGMM approach conducted on ADNI data using path length and betweenness centrality.

bounded common subspaces ICAMGGMM, optimized for enhanced performance in data confined to bounded support spaces, such as fMRI data. This refined model better adapts to the unique structural properties of such datasets, ensuring improved accuracy and robustness in component analysis.

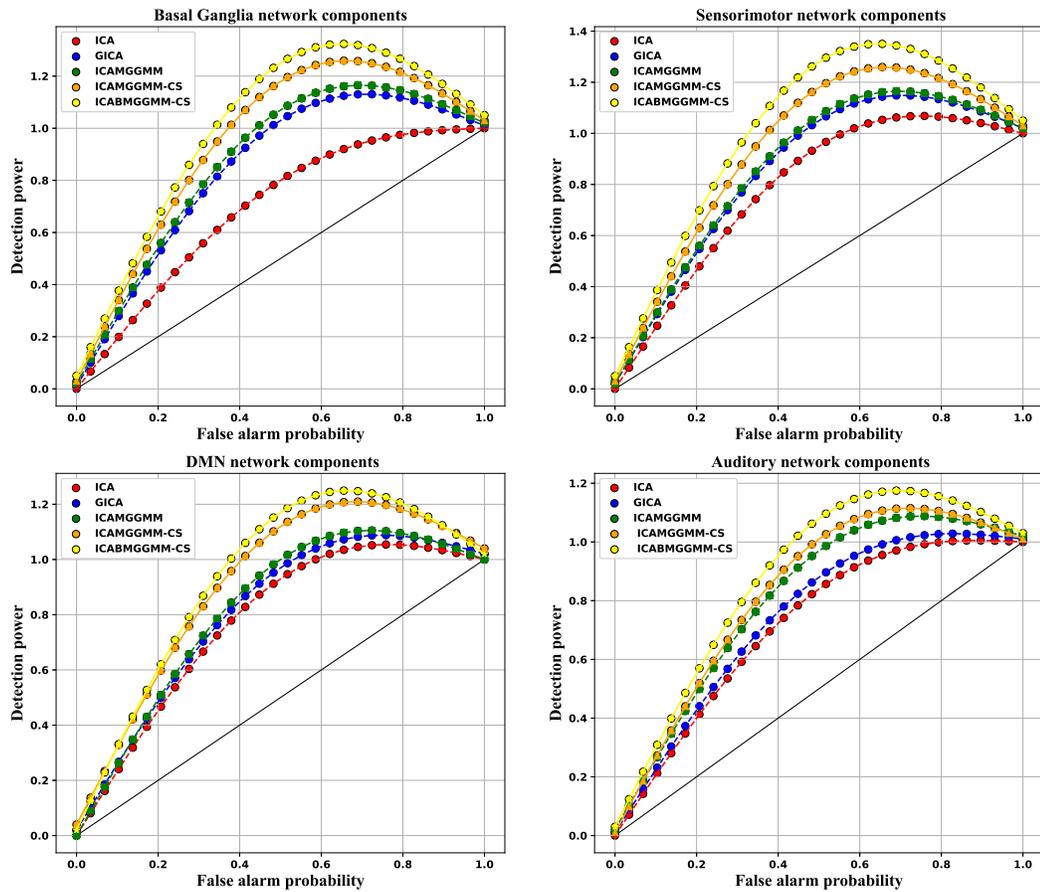


Figure 6.16: ROC curves for ICABMGMM-CS and compared models based on the resultant from two-sample t-test using the ADNI data.

Validation on neuroscience datasets showcased the efficiency of these models in distinguishing various mental disorders. Leveraging the ICABMGMM-CS components, we identified significant differences between mental disorder classes and a healthy control group. To assess the robustness of our models, we conducted two-sample t-tests alongside graphical analyses, demonstrating the model's resilience and its potential for precise classification in mental health diagnostics.

## Chapter 7

# Homogeneous Subgroup Identification in Resting State fMRI using Common Subspace Independent Vectors Analysis

In this Chapter, we introduce a novel common subspace independent vector analysis model based on the bounded multivariate generalized Gaussian mixture distribution, termed IVAMGGMM-CSS to solve the JBSS complexity in the high-dimensional data. The common subspace refers to latent sources present in all datasets, forming a low-rank structure. The IVAMGGMM-CSS model is designed to capture subject-specific variability within the data and estimate the common sources. By incorporating correlated features, the IVAMGGMM-CSS can adeptly handle complex, high-dimensional datasets. To enhance robustness and improve the accuracy of source separation, we utilize a bounded indicator function within the model. The IVABMGGMM-CSS framework integrates the flexibility of data-driven methods with the ability to manage noise and artifacts commonly associated with model-based approaches. Moreover, it leverages the interdependence between datasets, significantly enhancing its effectiveness for source separation tasks. To validate the stability and applicability of our proposed model, we conducted extensive simulations and real-world experiments. In practical applications, we applied the IVABMGGMM-CSS model to distinguish subjects with mental diseases from healthy controls, focusing on Schizophrenia and Autism datasets. Our results

highlight the model’s efficacy in identifying these common patterns in rs-fMRI data and pinpointing affected brain regions, thereby showcasing its potential as a powerful tool for mental disease diagnosis.

## 7.1 Introduction

Identifying homogeneous subspaces of patients is a crucial step in precision medicine, which seeks to customize medical treatments for individual patients categorized into specific subgroups. In clinical practice, the heterogeneity of patients presents a significant challenge for mental diseases like bipolar disorder, Schizophrenia, and Autism [248]. Various methods have been employed to cluster behavioural variations, clinical and cognitive scores, or other related metrics to identify subspaces. However, there is limited research that fully considers the multivariate nature of the medical imaging data [249]. Another study demonstrates that individual diversity in brain functional networks identified in fMRI data is consistent and relevant for subject identification via shared and subject-specific dictionary learning (ShSSDL) [250]. This makes the use of fMRI data highly promising for the problem of subgroup identification. However, ShSSDL presumes that time courses are consistent across datasets and cannot detect components shared among subgroups of subjects.

Subjects can be categorized by examining the correlation patterns in their functional network activity, enabling subgroup analysis using fMRI data. The functional connectivity patterns that are obtained from ICA, which is done separately for each subject, are clustered to produce the final subgroups, which is one drawback of the ICA method. This approach does not fully incorporate the multivariate information across all subject datasets. Another recent study introduces the IVA method, which identifies brain areas through the joint analysis of multi-subject data. By considering the dependencies across multiple datasets, IVA extends ICA to enable the joint analysis of multiple datasets. IVA effectively captures inter-subject variability compared to other multi-subject ICA algorithms, such as group ICA. IVA effectively reveals the correlation patterns of latent variables within multi-subject data by leveraging a sophisticated multivariate Gaussian density model.

However, the IVA performance can be inferior in the presence of common source component vectors when there is significant diversity between datasets, such as variability among sources within each source component vectors (SCVs). One reason for this is that many IVA algorithms employ a  $K$ -dimensional multivariate distribution to parameterize each SCV. However, common SCVs often lie in a low-dimensional subspace, which can lead to overparameterization. As a result, the presence of common SCVs can significantly degrade IVA's separation performance and increase running time. Moreover, IVA algorithms experience heightened computational complexity when applied to large-scale fMRI data. Given that IVA is considered one of the most powerful tool for BSS, there is a need for more efficient IVA algorithms that can also perform well with the common SCVs.

In this Chapter, we introduce the IVA multivariate generalized Gaussian mixture for common subspace separation (IVAMGGMM-CSS) model to reliably recover the correlation structure of latent variables across a large number of datasets. By defining the source component vectors, IVAMGGMM-CSS considers the correlation across datasets. Furthermore, strong identification assurances provided by IVAMGGMM-CSS ensure accurate estimation of the correlation structure. IVAMGGMM-CSS utilizes the benefits of IVA as a framework for identifying subgroups. This is achieved by capturing the correlation structure of the underlying datasets through common component vectors.

We also propose a bounded version of IVAMGGMM-CSS to address the limitations of unbounded BSS methods in estimating data shapes within bounded support regions. The IVABMGGMM-CSS algorithm is developed to address major challenges associated with IVA, specifically in terms of computational complexity and performance within common subspace. The IVABMGGMM-CSS method identifies the common subgroups based on each individual SCV, enabling us to assess the contribution of each component that defines the subgroups and enhancing interpretability. After identifying common SCVs using a measure of "correlatedness" among the covariance matrices, the common SCVs are separated from the non-common SCVs. Subsequently, IVABMGGMM-CSS is performed independently on the common and non-common subspaces. This approach enhances the estimation of these common subspaces and significantly reduces the data dimensionality in the IVA stage, thereby lowering the computational complexity. The IVABMGGMM-CSS framework inherits the capability of data-driven methods to manage noise and artifacts typically present in

model-based approaches. It also exploits the interdependence between datasets, making it highly effective for source separation. Additionally, IVABMGGMM-CSS assumes second-order correlation within bounded SCVs, enhancing its flexibility in extracting valuable insights from the multivariate non-Gaussian data. We can summarize the contributions of this Chapter as follows:

- We propose IVAMGGMM for common subspaces (IVAMGGMM-CSS) to effectively capture common subspaces in multivariate data, thereby facilitating the separation of diverse sources in complex scenarios. This model is designed to account for the interdependencies among sources within the multivariate dataset, offering flexibility and robustness in the separation process.
- We introduce bounded IVABMGGMM-CSS to overcome the limitations of unbounded models when applied to data within bounded support regions. This model assumes a second-order correlation within bounded support component vectors, which allows for greater flexibility in extracting knowledge from the data.
- We test the generalization of our proposed models by applying them to various neuroscience applications, including Schizophrenia and Autism.
- We use IVABMGGMM-CSS components to identify brain regions affected by Schizophrenia and Autism.
- To validate the robustness of our models, we employ a two-sample t-test to generate spatial statistical maps for both Schizophrenia and Autism brain volumes.

The observed fMRI signals can be perceived as a linear combination of spatial maps of activation and time courses. Thus, identifying fMRI activation can be considered a BSS problem [251]. We assess the proposed method's performance using both simulated and real-world applications. The simulation results demonstrate that IVABMGGMM-CSS outperforms existing approaches, accurately estimating the number of common SCVs. For real-world applications, we applied the proposed models to two rs-fMRI datasets: one for Schizophrenia and one for Autism. For the Schizophrenia dataset, we used 120 subjects from the CNP data to distinguish between Schizophrenic patients and healthy controls. For the Autism dataset, we used 100 subjects to differentiate between individuals

with Autism and those without. We demonstrate the interpretability and significance of the common subgroups found by IVABMGGMM-CSS using real fMRI data.

The rest of the Chapter is organized as follows: Section 7.2 introduces the literature. In Sections 7.3 and 7.4, we describe the IVA background and proposed algorithms, respectively. The experimental results are discussed in Section 7.5, followed by the conclusion in Section 7.6.

## 7.2 Related Work

Data fusion aims to optimize the exploitation of shared information present in datasets, whether they are multimodal or multiset. This is achieved by enabling full interaction among these datasets, leading to a more informative and comprehensive perspective on the relevant problem [252, 253]. On an intuitive level, these datasets possess shared information while also accommodating individual variations. Therefore, it is preferable to extract both common and distinct information from the datasets and subsequently analyze them separately. Processing this diverse information separately should lead to a more effective data analysis [228, 254, 227].

Due to the significant role of analyzing distinct and common subspaces in the analysis of medical images, several recent studies have specifically concentrated on this aspect. This focus is particularly evident in the fusion of various modalities, such as structural MRI, fMRI, and electroencephalography, or in combining data from different fMRI tasks [255, 256, 232, 257]. Nevertheless, these situations have only been illustrated in the joint decomposition of a limited number of datasets. As previously mentioned, the attractiveness of analyzing distinct and common subspaces extends to multi-subject analyses. The models employed to identify shared and distinct subspaces in fusion studies have not undergone thorough exploration in the context of jointly analyzing more than a few datasets.

Multi-subject data analysis entails the simultaneous examination of at least tens, and more commonly, hundreds, of subjects. The hierarchical ICA algorithm, as proposed by Guo and Tang [258], concurrently determines sources at both the population level and for individual subjects. Nevertheless, as the number of datasets increases, the density model complexity employed in hierarchical

ICA expands, and it neglects consideration of the dependence structure among these sources. A recently introduced ShSSDL technique, presented by Iqbal et al. [250], is designed for multi-subject task fMRI analysis. This method aims to recognize common components across subjects while also identifying subject-specific components. However, ShSSDL operates under the assumption of common time courses across datasets and lacks the capability to pinpoint components that are shared specifically among subgroups of subjects. These investigations highlight the need for a more efficient IVA version that works well with common SCVs.

## **7.3 Background**

### **7.3.1 Independent vector analysis for the bounded multivariate generalized Gaussian mixture model**

In the context of the IVA mixture model, the assumption is made that the observed data arises from a mixture model and is categorized into distinct, non-overlapping classes. The mathematical background for IVA and IVA mixture are introduced in Chapter 2.

## **7.4 Proposed IVA mixture for common subspace analysis**

Analyzing common and individual subspace has proven valuable in detecting distinct biomedical patterns within various populations, aiding in the comprehension of the distinctive characteristics associated with different brain disorders. Most medical imaging data, like fMRI data, involves the collection of information from tens or hundreds of subjects. Analyzing the combined multi-set data enables the exploitation of its rich information, especially when considering diverse datasets. This prompts the exploration of an algorithm capable of identifying both common and distinct subspace within the data.

The dataset observed for the  $m$ th subject is a mixture of three distinct sources:

$$\begin{aligned}
\mathbf{x}^{[m]} &= \mathbf{A}_j^{[m]} \mathbf{s}_j^{[m]} \\
&= \mathbf{A}_{j_C}^{[m]} \mathbf{s}_C^{[m]} + \mathbf{A}_{j_G}^{[m]} \mathbf{s}_G^{[m]} + \mathbf{A}_{j_D}^{[m]} \mathbf{s}_D^{[m]} \\
&= \begin{bmatrix} \mathbf{A}_{j_C}^{[m]} & \mathbf{A}_{j_G}^{[m]} & \mathbf{A}_{j_D}^{[m]} \end{bmatrix} \begin{bmatrix} \mathbf{s}_C^{[m]} \\ \mathbf{s}_G^{[m]} \\ \mathbf{s}_D^{[m]} \end{bmatrix}, \quad m=1, \dots, M
\end{aligned} \tag{7.1}$$

where  $\mathbf{A}_{j_C}^{[m]}$ ,  $\mathbf{A}_{j_G}^{[m]}$ , and  $\mathbf{A}_{j_D}^{[m]}$  are the corresponding mixing matrices of common sources, group-specific sources, and distinct sources, respectively. The first set of SCVs outlines the common sources, where the sources within each SCV display a significant level of correlation. The second set describes subspace with a moderate correlation between sources in each SCV. The third set identifies the distinct sources, where sources within each SCV show a low level of correlation. To construct the IVABMGGMM-CSS, we utilize Equations (2.4) and (7.1). Our IVABMGGMM-CSS method consists of two primary steps: (i) subset analysis and (ii) common subspace identification. To tackle the issue of dimensionality, the subset analysis stage involves performing numerous individual IVABMGGMM decompositions on subsets of randomly chosen subjects from the population. A collection of group-specific and distinct SCVs are established during the common subspace identification stage, in addition to a common subspace that is consistent for the entire population. These steps are elaborated in Fig. 7.1. The specifics of the IVABMGGMM-CSS method will be elaborated on in the following subsections.

#### 7.4.1 IVABMGGMM-CSS: subset analysis

The subset analysis process for IVABMGGMM-CSS, illustrated in Fig. 7.1, involves randomly selecting groups of  $M$  subjects from the entire dataset. For each SCV, an  $M \times M$  correlation matrix is generated, producing  $\frac{M(M-1)}{2}$  distinct correlation values. These correlation values are used to evaluate the similarity of components and to determine whether a source component is consistently present across the  $M$  datasets. SCVs that correspond to common components are expected to yield consistently high correlation values.

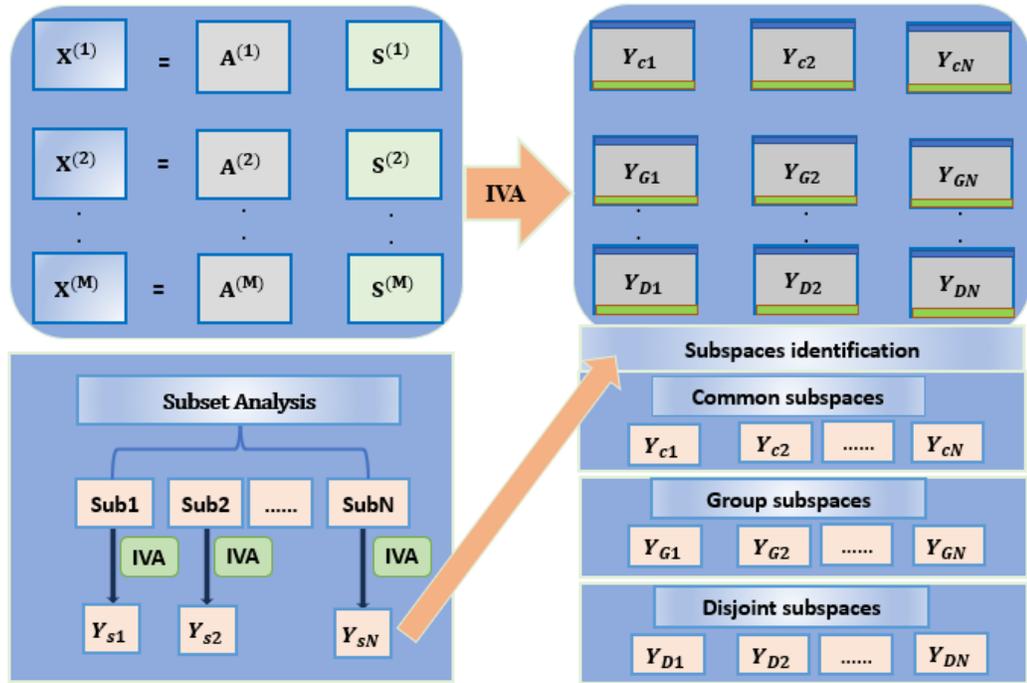


Figure 7.1: IVABMGGMM-CSS framework. The first step is to apply the IVABMGGMM on the multi-subjects data. Then, the identification for the common subgroups is conducted.

To measure the “commonality” of an SCV, the ratio  $Q$  is introduced. This value is computed as the fraction of correlation values exceeding an empirically chosen threshold  $\gamma$ , relative to the total number of correlation values for the SCV:

$$Q = \frac{N_\gamma}{M(M-1)/2} \quad (7.2)$$

In this equation,  $N_\gamma$  represents the correlation values greater than  $\gamma$  within the correlation matrix of the SCV. A higher  $Q$  value, approaching 1, signifies a stronger association with a common component.

Within each subset, SCVs are ordered based on their average correlation values, starting with the highest. This ranking system ensures that SCVs with stronger correlations are listed at the top, while those with weaker correlations appear lower in the order.

## 7.4.2 IVABMGGMM-CSS: common subspace identification

The goal of the second phase is to determine a unified common subspace,  $Y_C$ , that is consistent across all subjects within the group. To ensure uniformity across  $S$  subsets, the number of components in each subspace,  $Y_{sl}$ , is adjusted to a common value,  $C$ . This value is chosen as the maximum of the subset-specific  $C_s$  values, maximizing the pool of potential candidates for constructing the common subspace. The  $C$  SCVs from each subset then serve as candidates for this identification process.

For each subset  $s$ , the mean component,  $\mathbf{y}_{cs}$ , is calculated by averaging the  $M$  components of the  $c$ th SCV. The consistency of each mean component is evaluated using cross-correlation, which measures the average correlation between the mean component of a subset  $s$  and its corresponding components from the remaining  $S - 1$  subsets as:  $p_{cs} = \frac{1}{S-1} \sum_{l \neq s} |p_{csl}|$ . Here,  $p_{csl}$  denotes the Pearson correlation coefficient between the mean components  $\mathbf{y}_{cs}$  and  $\mathbf{y}_{cl}$  for subsets  $s$  and  $l$ , where  $1 \leq l \leq S$  and  $l \neq s$ . To assess the overall consistency of a subset, the average cross-correlation across its  $C$  mean components is computed as:  $p_s = \frac{1}{C} \sum_c p_{cs}$ .

The subset with the highest average cross-correlation  $p_s$  is identified as containing the most consistent components. This subset is subsequently used to define the common subspace shared by all subjects. The complete learning procedure for the IVABMGGMM-CSS model is detailed in Algorithm 7.

## 7.5 Results and discussion

### 7.5.1 Application to simulated data

To test the stability and robustness of our proposed models, we conduct them over multivariate data generated randomly from a multivariate generalized Gaussian distribution. To ensure a comprehensive initialization, we start by employing the K-means algorithm to determine the mixture parameters, which include the mean, covariance, and shape parameters. For robust modelling, the mixing parameter is set to  $1/M$ , thereby normalizing the sum to 1. Subsequently, we apply the EM algorithm, a powerful iterative technique, to meticulously compute the posterior probability in the

---

**Algorithm 7** The common subspace IVABMGGMM algorithm

---

```
1: Input: Dataset  $\mathcal{X}$ 
2: Apply K-means to obtain mean and covariance for each component.
3: Set  $\beta_j = 0.5$ 
4: Randomly initialize  $M$  de-mixing matrices  $\mathbf{W}^{[m]}$ 
5: for  $m = 1, \dots, M$  do
6:   for  $n = 1, \dots, N$  do
7:     Compute  $\mathbf{y}_n^{[m]} = \mathbf{W}^{[m]} \mathbf{x}_n^{[m]}$ ,  $m = 1, \dots, M$ 
8:     Expectation Maximization:
9:     if change in likelihood  $>$  threshold then
10:      E-step
11:      Use Equation (2.8) to compute posterior probabilities
12:    else
13:      M-step
14:      Use Equation (2.10) to update the mixing parameter  $\mathbf{p}_j$ 
15:      Using EM algorithm: Equation (2.14) to update  $\mu_j$ 
16:      Apply Equation (2.17) to update  $\Sigma_j$ 
17:      Apply Equations (2.19) and (2.20) to update  $\beta_j$ 
18:    end if
19:    Compute  $\mathbf{y}_n^{[m]} = \mathbf{W}^{[m]} \mathbf{x}_n^{[m]}$ ,  $m = 1, \dots, M$ 
20:    if  $Q > \gamma$  then
21:       $\mathbf{y}_n^{[m]} \rightarrow$  Common subspace
22:    else
23:       $\mathbf{y}_n^{[m]} \rightarrow$  distinct subspace
24:    end if
25:  end for
26: end for
27: Repeat steps 5 to 26 until convergence.
28: Return the model's parameters  $\hat{\mu}_j, \hat{\beta}_j, \hat{\Sigma}_j$ , and common components  $Y_C$ 
```

---

E-step and refine the mixture parameters in the M-step.

To ensure the integrity of our model, we define a small tolerance value, enabling precise convergence assessment. This iterative process persists until our model achieves convergence, guided by the pursuit of optimal parameters for effective source separation. We set a maximum limit of 1000 iterations for each model with a stringent convergence threshold of 0.001, ensuring computational efficiency without compromising accuracy. As a result, we derive a de-mixing matrix, represented as  $\mathbf{W}^{[m]}$ , facilitating the accurate estimation of sources through the transformation  $\mathbf{y}_n^{[m]} = \mathbf{W}^{[m]} \times \mathbf{x}_n^{[m]}$ . Fig. 7.2 shows the effectiveness of our models IVAMGGMM-CSS and IVABMGGMM-CSS in separating the common mixing sources in terms of the joint-ISI. Lower joint-ISI values depicted in the figure correspond to reduced interference within the separated signals, exemplifying the superior performance of our models in mitigating interference. However, the base models overparameterize

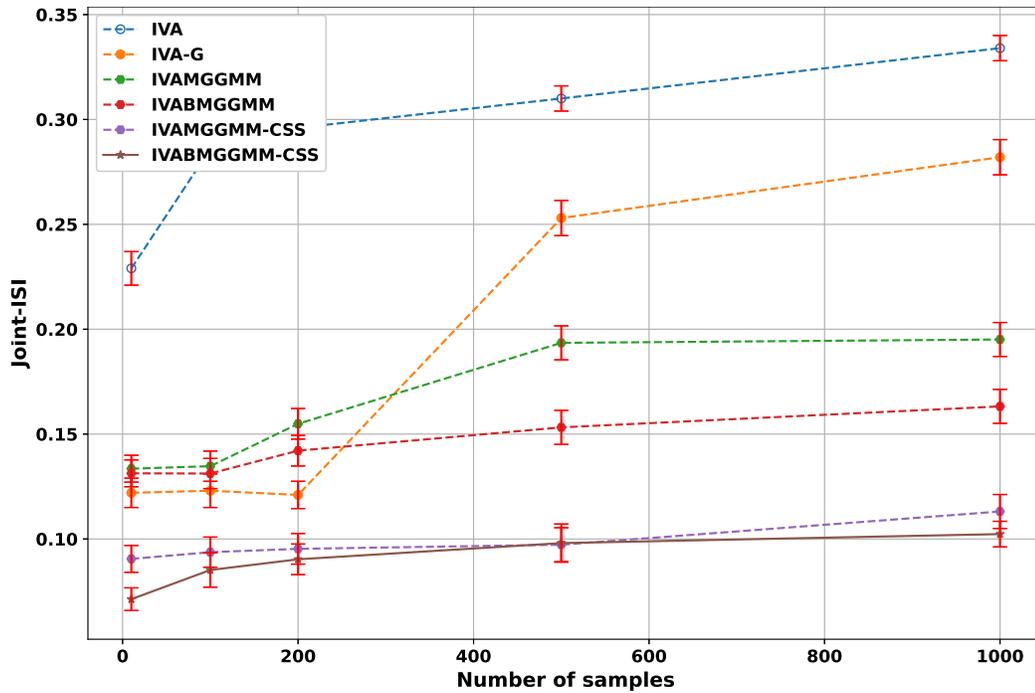


Figure 7.2: The joint-ISI metric for proposed IVAMGGMM-CSS and IVABMGGMM-CSS algorithms in terms of sample number.

the common SCVs, leading to poor performance.

To rigorously evaluate the performance of our algorithms, we employ the normalized interference-to-signal ratio (ISR). This metric is derived by averaging results from 100 independent runs, ensuring a robust and comprehensive assessment. The flexible underlying density models of IVAMGGMM-CSS and IVABMGGMM-CSS demonstrate superior performance as the sample size increases, as illustrated in Fig. 7.3. In contrast, other models, constrained by their basic density structures, do not perform well. Among the six algorithms evaluated, IVABMGGMM-CSS notably achieves the best overall performance.

To validate the robustness of our models, we assess their performance in terms of normalized ISR as a function of shape parameter  $\beta_j$ . In Fig. 7.4, we observe that for the small values of  $\beta_j$ , i.e, high sparsity, the proposed IVAMGGMM-CSS and IVABMGGMM-CSS exhibit better performance. In the case of low sparsity, i.e, high values for  $\beta$ , we notice the poor performance for base models due to their simple distribution function. In total, among the different values of  $\beta_j$ , all conducted models perform well in the case where  $\beta_j = 0.5$ . The IVABMGGMM-CSS exhibits

the best performance among all models due to the flexibility in its underlying density function.

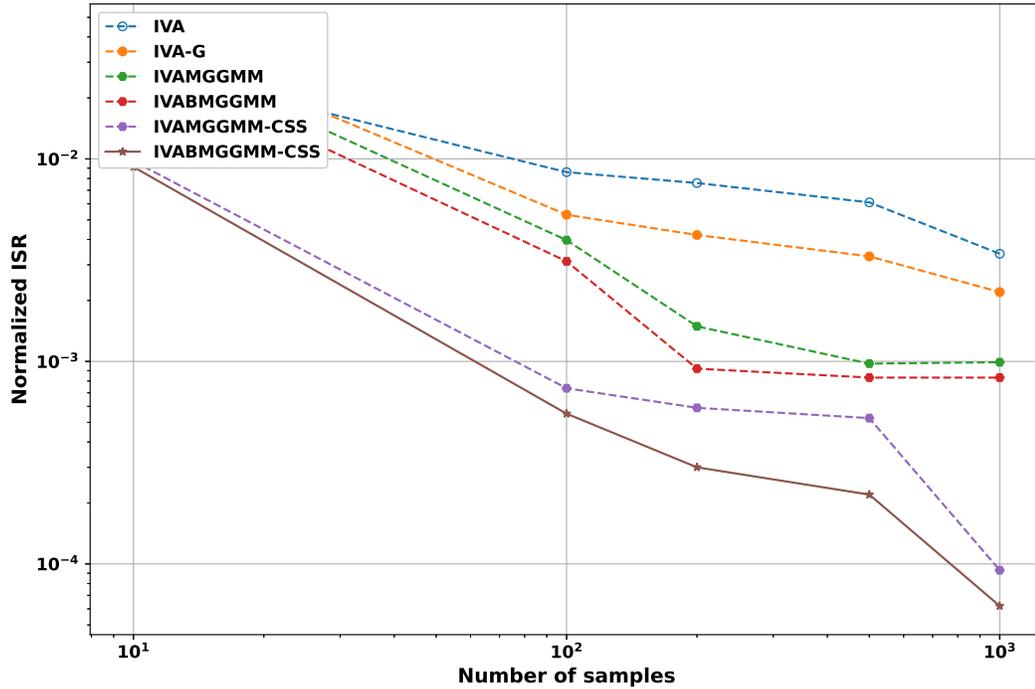


Figure 7.3: Normalized ISR for our proposed models and base models as a function of number of samples.

We further assess our algorithms' efficacy by evaluating the absolute correlation between true and estimated sources. In Fig. 7.5, we observe that IVAMGGMM-CSS perform significantly better than other base models in terms of spatial correlation. However, IVABMGGMM-CSS provide the best performance among all conducted models due to its flexibility in estimating the sources. In this experiment, results are averaged over 200 independent runs.

## 7.5.2 Schizophrenic diagnosis using the common subspace IVAMGGMM and IVABMGGMM

### Dataset

The UCLA consortium provides the consortium for neuropsychiatric phenomics (CNP) dataset, which includes data from 50 Schizophrenia patients and 121 healthy individuals. Resting-state fMRI data for the CNP were acquired using a T2\*-weighted EPI sequence with the following parameters: TE = 30 ms, TR = 2 s, matrix size = 64, and 34 slices. For this investigation, we used a subset of

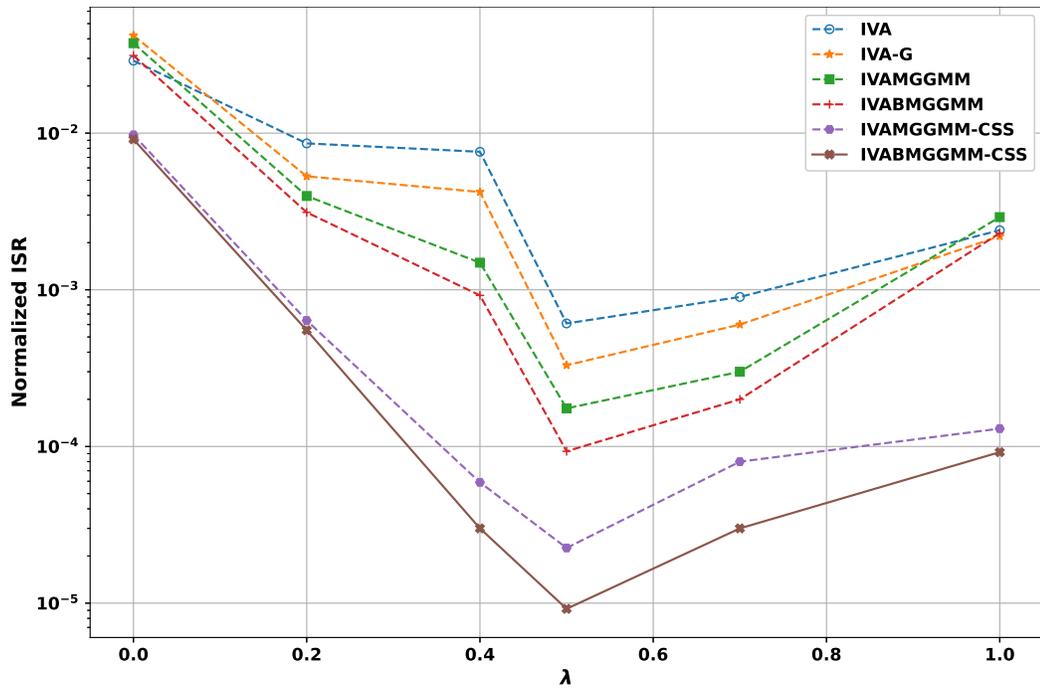


Figure 7.4: The performance of IVA methods based on the normalized average ISR as a function of shape parameter  $\beta_j$ . Each point represents the result of 100 separate runs.

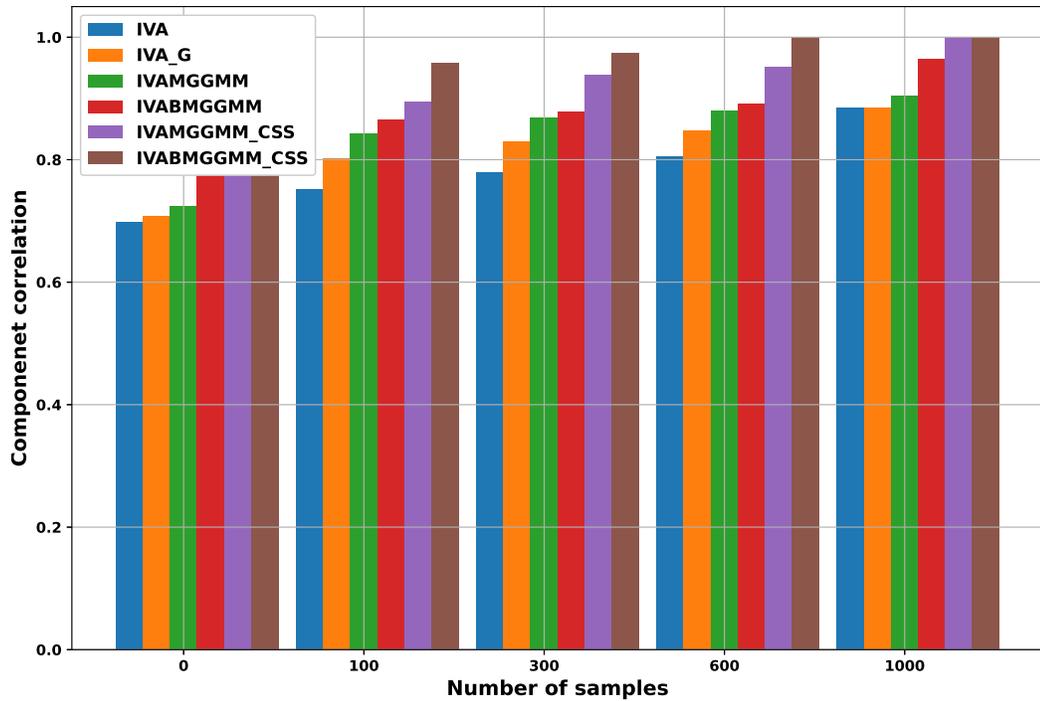


Figure 7.5: Common discriminative components correlation.

120 participants, comprising 50 individuals with Schizophrenia and 70 healthy controls. Fig. 7.6 presents a statistical summary of these data.

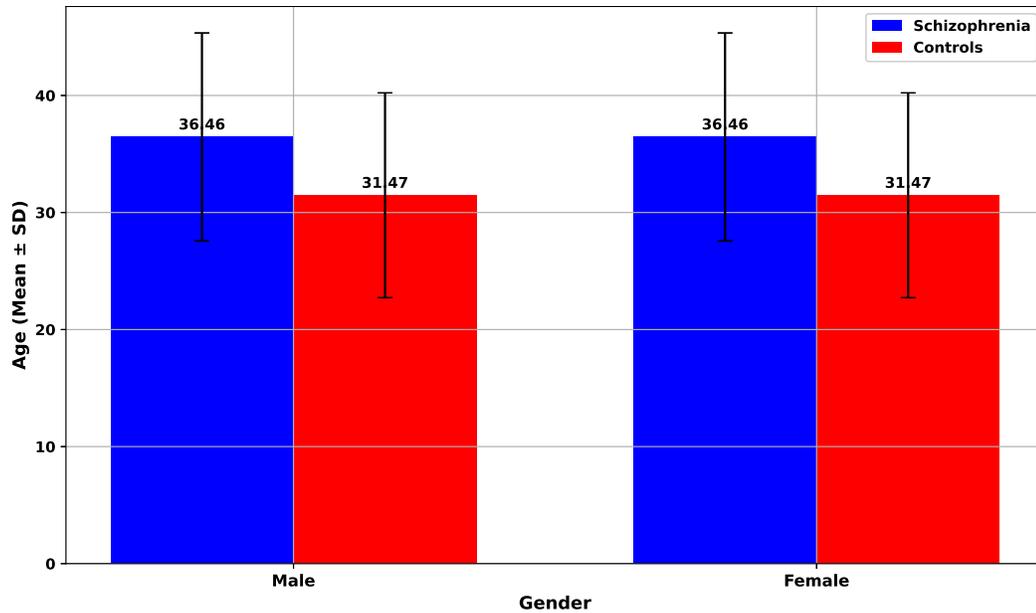


Figure 7.6: Participant demographics in CNP data.

### CNP data preprocessing

The data preprocessing was conducted using the DPABI software tool. The primary steps included removing initial time points, performing slice timing correction, correcting for head movements, co-registering images, applying spatial normalization, and smoothing the data. Initially, we removed the first ten time points for each subject. For the slice time correction, we chose the 34th slice as a reference slice. We aligned the high-resolution structural images with the BOLD space to synchronize the structural and functional images. Then, the fMRI data were aligned with the MNI template by mapping the brains with varying sizes and shapes into a standardized space, thereby minimizing individual brain differences. Finally, spatial smoothing was applied to reduce subtle variations in brain structures among different subjects and enhance the signal-to-noise ratio. This was achieved using a Gaussian kernel with a full width at FWHM of 4 mm.

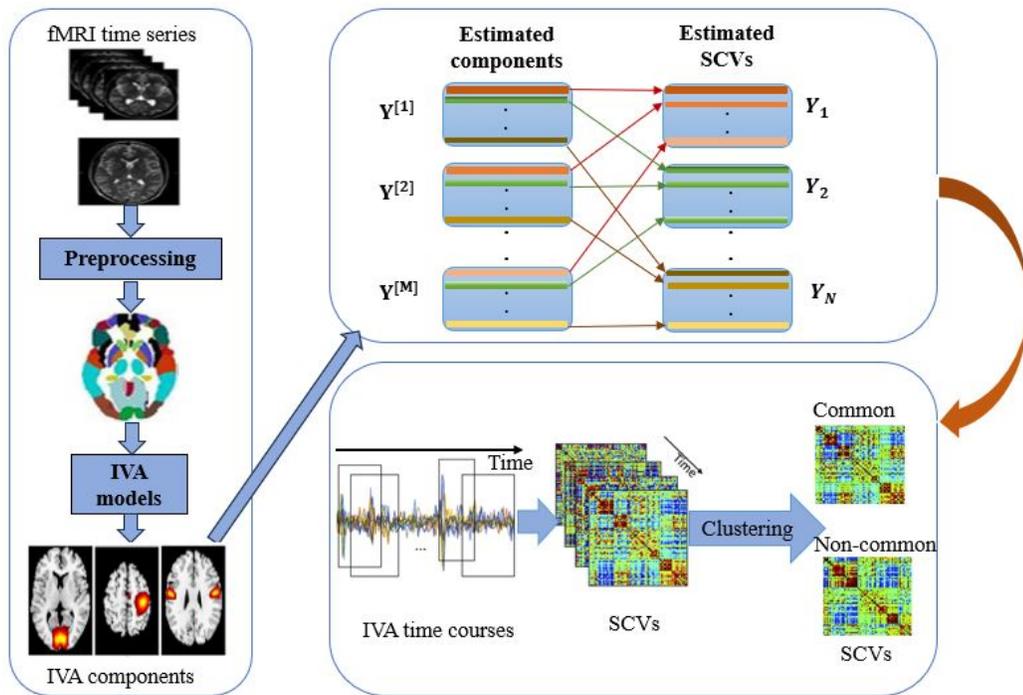


Figure 7.7: The general framework illustrates the comprehensive proposed pipeline for fMRI data analysis. Our approach is conducted over the SCVs to identify the common and non-common subgroups.

### Experimental setup

After the preprocessing stage, the brain was parcelled into 116 regions using the AAL atlas. Then, we applied our IVA models to the data to extract the IVA components, time courses, and connectivity matrices. The average intensity time series for each ROI was then extracted, and it was subsequently passed through a band-pass filter with a frequency range of 0.01 Hz to 0.08 Hz. The connectivity matrices were used to identify the components affected by Schizophrenia. We masked each volume to exclude non-brain voxels and flattened the result, yielding an observation vector of  $V = 56,230$  voxels (samples). These steps of our pipeline are illustrated in Fig. 7.7.

### Experimental results for the common subspace IVABMGMM

After the parcellation of brain into 116 brain regions, we conducted our proposed IVAMGMM and IVABMGMM for the common subspace to this parcelled large-scale rs-fMRI data. Initially, we used the k-means algorithm to initialize the parameters of our model, including the mean  $\mu_j$ ,

covariance  $\Sigma_j$ , and shape parameters  $\beta_j$ . Subsequently, we employed the EM algorithm to compute the posterior probability of the common subspace and iteratively update the model's parameters. This iterative process continued until convergence, at which point the resulting demixing matrix was utilized to differentiate Schizophrenic patients from healthy controls. The time courses of the IVA common subspace were then used to extract connectivity matrices, ultimately identifying the homogeneous subgroups of the brain regions affected by Schizophrenia. We employed various BSS metrics to evaluate the performance of the proposed models including SDR, SAR, SIR, and Joint-ISI.

Table 7.1 displays the performance comparison of IVAMGGMM, IVABMGGMM for the common subspace, and other models, such as IVA and IVA-G, across various subject combinations. The findings clearly indicate that our proposed models outperform the basic models, especially in cases with lower subject combinations. This superior performance is even more evident as the number of subject combinations increases. The IVAMGGMM-CSS model achieves SDR values ranging from 8.73 for 2 subjects to 5.36 for 120 subjects, illustrating enhanced performance stability in identifying homogeneous subgroup structures. The IVABMGGMM-CSS model achieves the highest SDR value of 9.63 at a combination of 2 subjects and maintains relatively high SDR values at larger combinations, such as 6.32 at 120 subjects. All models exhibit a decrease in SDR as the combination increases. This trend indicates that increased combination complexity makes it harder for the models to obtain high SDR values. Models that incorporate common subspace (IVAMGGMM-CSS and IVABMGGMM-CSS) demonstrate the greatest resilience to higher combinations, maintaining superior SDR values compared to traditional models. The improved effectiveness of our models in handling high-dimensional data highlights their flexibility and stability, making them particularly effective at distinguishing between Schizophrenic patients and control subjects. In contrast, the base models exhibit poor performance due to their simplistic probability density function distribution and susceptibility to overparameterization of the common SCVs.

The heat map in Fig. 7.8 displays the SIR values for various models across different combinations. The models evaluated include IVA, IVA-G, IVAMGGMM, IVABMGGMM, IVAMGGMM-CSS, and IVABMGGMM-CS, with combinations ranging from 2 to 120 individuals. The IVA model exhibits the lowest overall SIR values, starting at 6.83 for a combination of 2 and decreasing

Table 7.1: Performance of IVAMGGMM-CSS and IVABMGGMM-CSS using SDR separation measure.

Combinations/Models	SDR					
	IVA	IVA-G	IVAMGGMM	IVABMGGMM	IVAMGGMM-CSS	IVABMGGMM-CSS
2	5.39	6.89	7.64	7.94	<b>8.73</b>	<b>9.63</b>
10	4.72	6.05	7.05	7.14	<b>8.53</b>	<b>8.93</b>
20	3.59	5.66	6.56	6.95	<b>7.63</b>	<b>8.63</b>
40	2.88	5.71	6.07	6.25	<b>7.33</b>	<b>7.43</b>
60	2.85	4.73	5.87	5.89	<b>6.76</b>	<b>7.13</b>
80	2.81	4.07	5.17	5.29	<b>6.28</b>	<b>6.86</b>
100	1.85	3.92	4.34	4.65	<b>5.34</b>	<b>6.93</b>
120	1.87	2.39	3.38	4.37	<b>5.36</b>	<b>6.32</b>

to 3.78 at a combination of 120 subjects. In contrast, the IVAMGGMM-CSS model outperforms other models, maintaining high SIR values between 8.94 and 5.96. The IVABMGGMM-CSS model achieves the highest SIR values, ranging from 10.63 at a combination of 2 to 7.25 at a combination of 120 subjects. The IVABMGGMM-CSS model demonstrates the highest robustness and best performance across all combinations, indicating that integrating the full covariance matrix and common subspace methods significantly enhances the model’s ability to identify the homogeneous common patterns. Similarly, the IVAMGGMM-CSS model also performs well, although slightly below the IVABMGGMM-CSS, highlighting the effectiveness of the common subspace technique.

In terms of SAR, we evaluated the proposed models to differentiate between healthy and Schizophrenic subjects. Fig. 7.9 illustrates the performance of these models on the CNP dataset. We tested various subject combinations to validate the robustness of our models in high-dimensional scenarios and analyzed the common SCVs. As shown in the figure, our models are highly effective in distinguishing between individuals with Schizophrenia and those without. This improvement demonstrates the stability and capability of our models in analyzing rs-fMRI common structures.

Fig. 7.10 illustrates the comparison of Joint-ISI performance across different numbers of subjects for six algorithms: IVA, IVA-G, IVAMGGMM, IVABMGGMM, IVAMGGMM-CSS, and IVABMGGMM-CSS. As the number of subjects increases from 2 to 120, the Joint-ISI values generally increase, indicating a degradation in the performance of the algorithms with complex data. IVAMGGMM-CSS and IVABMGGMM-CSS show the smallest Joint-ISI values, suggesting they are more effective for large-scale fMRI datasets and extracting the common patterns. In contrast, IVA and IVA-G exhibit a high increase in Joint-ISI values, indicating they might be less efficient

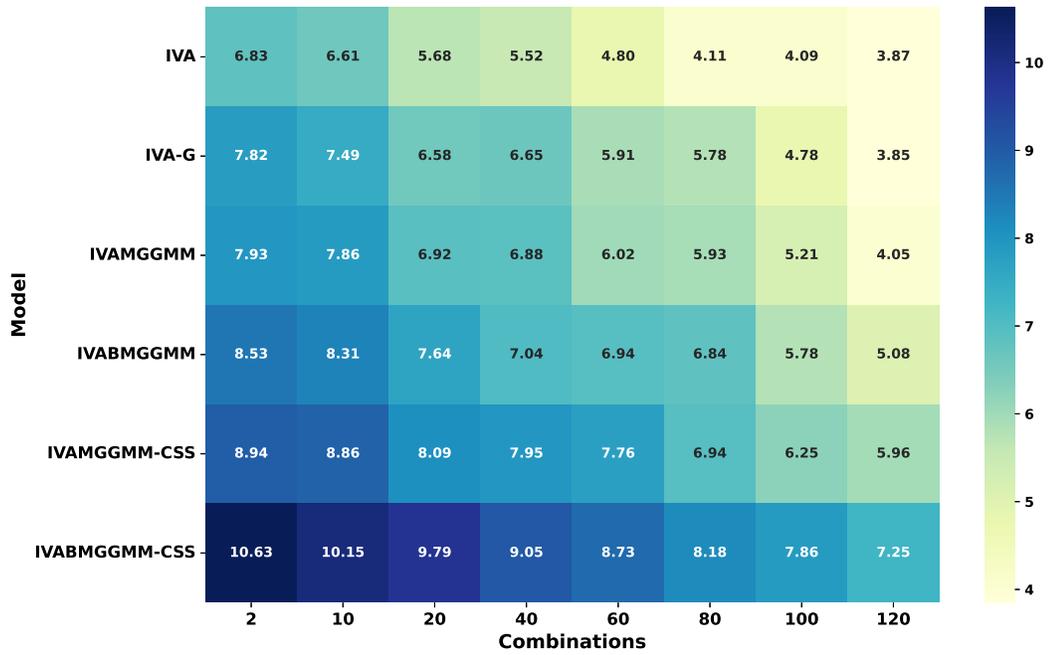


Figure 7.8: Heatmap for the IVA conducted models across a range of combination settings. Each cell represents the performance score, with low performance (cool colors) and high performance (warm colors).

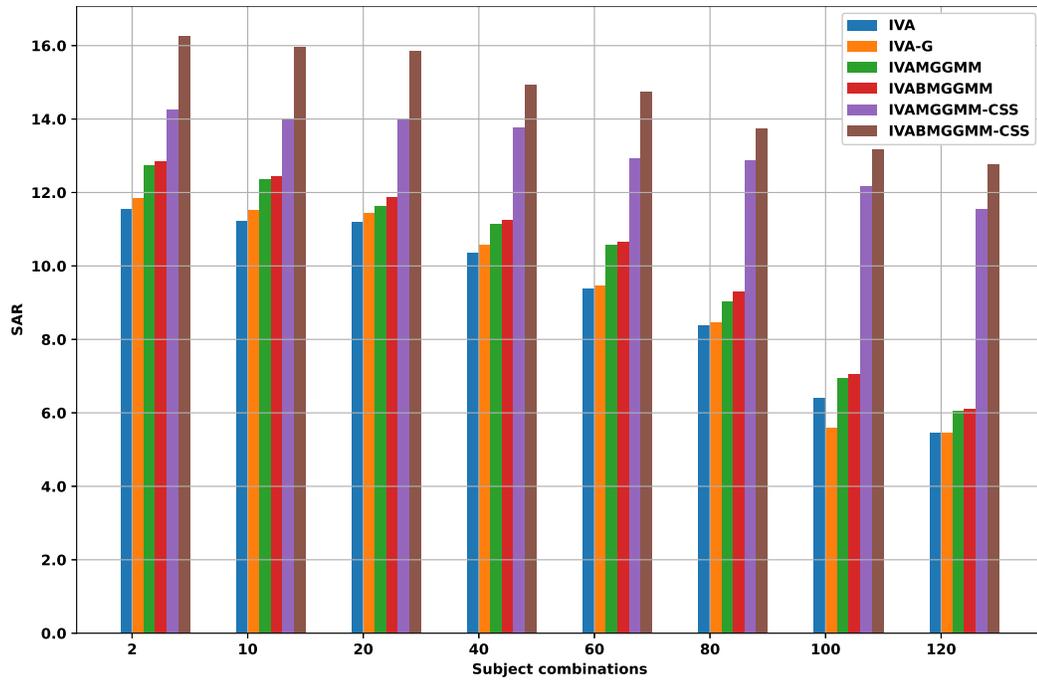


Figure 7.9: The performance of conducted models using SAR measure.

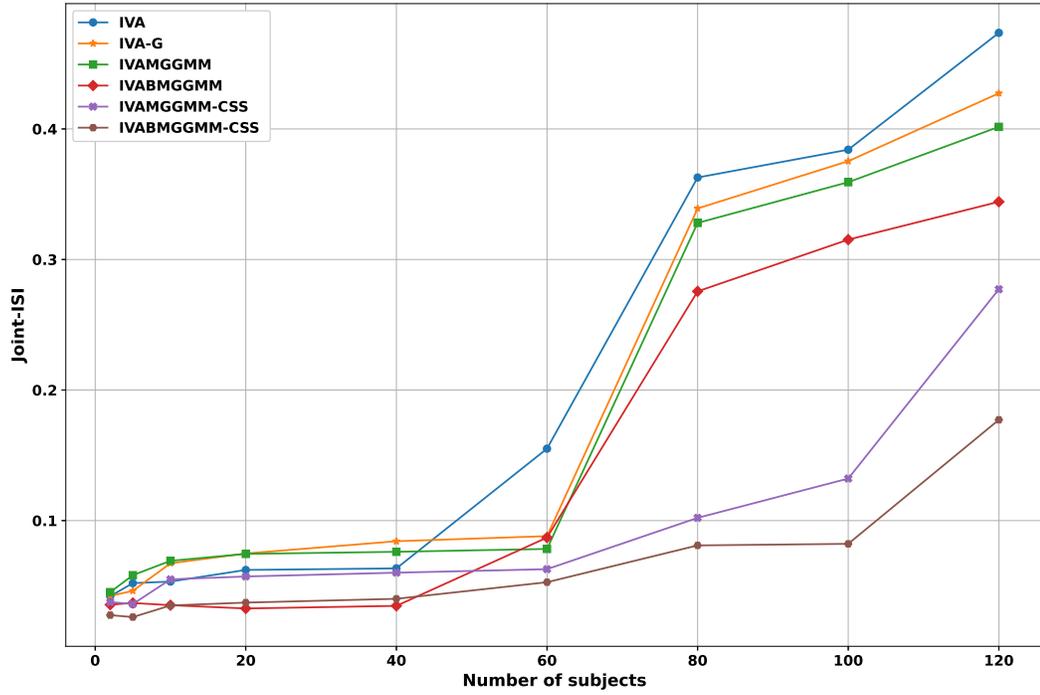


Figure 7.10: The IVABMGGMM-CSS performance based on the Joint-ISI metric.

in handling larger numbers of subjects and analyzing the common SCVs. The figure highlights the importance of selecting appropriate algorithms based on the dataset size to achieve optimal performance in terms of Joint-ISI.

### Brain network identification and statistical analysis for Schizophrenic patients

To assess the robustness of our proposed algorithms, we extracted the affected brain regions due to Schizophrenia using the most discriminant IVA components with high correlation coefficients. Here, we selected the top 100 IVA components, which characterized as intrinsic connectivity networks. Each class is divided into six random subgroups of 60 subjects. For each SCV, there are  $60 \times \frac{(60-1)}{2} = 1770$  unique correlation values. The SCVs are ranked by their mean correlation values, ranging from high to low. In the Schizophrenic subsets, over 85% of the first 35 SCVs have  $Q \geq 0.96$ , indicating that these SCVs are prevalent across individuals in these subgroups. The number of common SCVs ( $C$ ) among the six Schizophrenic group subsets are 33, 26, 20, 24, 25, and 38, respectively. For the healthy subjects class, the numbers are 21, 22, 27, 29, 25, and 30, respectively. Among the twelve values, the number of candidates,  $C$ , in the subspace  $Y_{sl}$  is determined to be 38,

the highest observed. The 25th percentile of the cross-correlation values of these  $C$  candidates is used as the threshold  $\gamma^{[m]}$  for identifying associated components. We use the MNI template for grey matter (GM) and cerebrospinal fluid (CSF) to distinguish these components. Based on the median value of  $C_{GM} - C_{CSF}$ , the components are divided into two distinct groups. Components with  $C_{GM} - C_{CSF}$  values above the median are more likely to be associated with higher probabilities of resting-state networks, while components with values below the median are more likely to represent ventricle effects. Table 7.2 introduces a comprehensive summary of the brain networks along with their IVA components and coordinates. Each network is associated with specific brain areas that align with their known functional roles. These networks include the DMN, Executive function network, Visual network, Salience network, Auditory network, Sensorimotor network, Frontal network, and Working memory network. The DMN network is broadly engaged across various brain regions, whereas networks like the executive function and visual networks have more localized coordinates indicative of their specialized functions. Results of the IVABMGMM-CSS component group comparison for Schizophrenic patients are shown in Fig. 7.11.

The discriminative connectivity maps were created using a two-sample t-test to statistically validate the differences between healthy controls and individuals with Schizophrenia. Using the FDR method for multiple comparisons, the significance level was set at a corrected  $p < 0.05$  (minimum  $z > 1.9$ , cluster significance:  $p < 0.05$ ). Table 7.2 presents the most discriminative IVA common components. Compared to the healthy control group, the Schizophrenic group exhibited increased connectivity across interesting common networks. Specifically, the common subspace with higher functional connectivity is observed in brain networks such as the DMN, dorsal DMN, Executive function network, Visual network, Salience network, Auditory network, and sensorimotor network. These regions are localized to the superior frontal gyrus, lingual gyrus (LING), FFG, supplementary motor area, superior temporal gyrus (STG), and postcentral gyrus (PoCG). Additionally, regions with increased FC in the frontal network are found in the superior frontal gyrus, dorsolateral (SFG). STG, FFG and SFG show disruptions in brain functions in Schizophrenic subjects. These findings align with the defective regions identified by several studies [196, 190], which highlight the involvement of these brain areas in Schizophrenia. Our results further suggest that the LING gyrus may be affected in the early stages of Schizophrenia. Numerous studies have reported

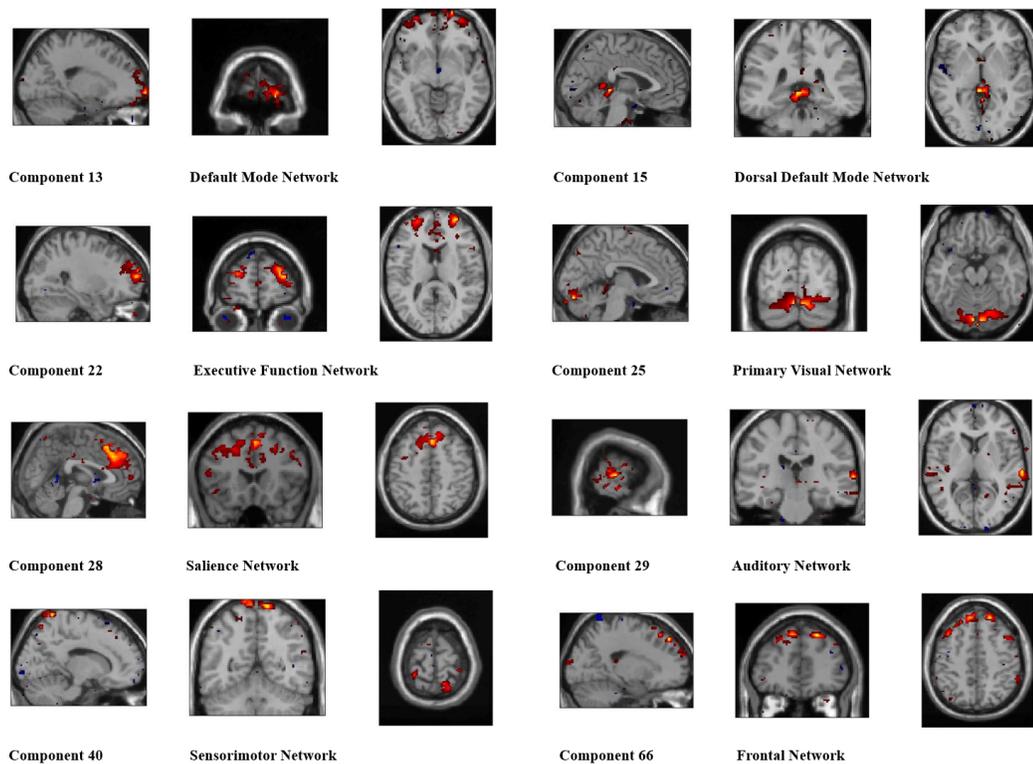


Figure 7.11: The spatial maps of the IVABMGGMM-CSS components for Schizophrenia.

significant abnormalities in the LING and PoCG in individuals with Schizophrenia [191, 190]. In our study, the LING and PoCG were significantly smaller in Schizophrenic patients compared to healthy controls.

These findings provide a comprehensive mapping of the brain networks affected in Schizophrenia, underscoring the heterogeneity and complexity of the Schizophrenia disorder. The identified components and their associated regions offer valuable insights for the proposed IVABMGGMM-CSS to preserve more subject variability and providing more detailed identification of these common brain networks.

### 7.5.3 Autism detection using the common subspace IVAMGGMM and IVABMGGMM

Autism spectrum disorder (ASD) is a developmental syndrome characterized by impairments in social communication and interaction, leading to restricted and repetitive behaviours and interests. It manifests in a wide range of symptoms and severity levels, with its underlying causes not yet

Table 7.2: The most discriminate IVA components and peak activation in the t-maps of the resting-state networks.

Anatomical description	IVA Components	T value	Coordinates (X, Y, Z)
Default mode network Superior frontal gyrus, medial	13	2.3	13, 65, 6
Dorsal Default mode network Lingual gyrus	15	1.96	-8, -36, -5
Executive function network Superior frontal gyrus	22	2.2	26, 60, 14
Visual network Fusiform gyrus	25	2.5	23, -78, -18
Saliience network Supplementary motor area	28	2.3	0, 20, 50
Auditory network Superior temporal gyrus	29	1.98	66, -24, 6
Sensorimotor network Postcentral gyrus	40	2.4	12, -54, 74
Frontal network Superior frontal gyrus, dorsolateral	66	2.3	-24, 38, 48

fully understood. Advances in neuroimaging techniques, including EEG and fMRI, have provided researchers with valuable insights into the neural mechanisms underlying ASD. Studies using fMRI have identified various brain networks with typical functioning in individuals with ASD, such as the social brain network and default mode network [259, 260]. Researchers also explored the functional connectivity between various brain regions and how it might be altered in individuals with ASD. Focusing on tasks and resting states, Rafiee et al.'s recent review [261] summarized the most recent developments in ASD research.

## Dataset

In this work, the publicly available fMRI Autism brain imaging data exchange (ABIDE) database was used [262]. For analysis and prediction, we selected 70 individuals with ASD and 50 normal controls, aged 9 – 13 years, who were scanned at Georgetown University and Children’s National Medical Centre Institutional Review Boards. Children with ASD were recruited through the Centre for ASD at Children’s National Medical Centre, while control children were collected from the Washington, DC, area community via advertisements in public venues and pediatrician offices. The 3T Siemens Allegra scanner was used to collect the data. The primary parameters for the resting-state functional images were: TR/TE = 2000/30 ms, flip angle =  $90^\circ$ , number of slices = 43, slice thickness = 2.5 mm, FOV =  $192 \times 192$  mm, and voxel size =  $3.0 \times 3.0 \times 2.5$  mm<sup>3</sup>. T1-weighted images were collected using the following parameters: TR/TE = 2530/3.25 ms, flip angle =  $7^\circ$ , slice thickness = 1.33 mm, and voxel size =  $1.3 \times 1.0 \times 1.3$  mm<sup>3</sup>.

## ASD data preprocessing

The rs-fMRI data were processed using SPM12 [180] and DPABI [72] software tools. The preprocessing involved several crucial steps: the initial 10 time-point volumes were discarded to establish magnetization equilibrium, slice timing correction was applied for interleaved acquisition, and subject motion was corrected through realignment. Functional and structural images were then co-registered. After normalizing the images to the EPI template, they were smoothed using a Gaussian kernel with a FWHM of 4 mm. Using the AAL atlas, each volume was divided into 116 ROIs. The average intensity time series was extracted for each ROI, and to minimize non-neuronal influences on the BOLD signal fluctuations, a band-pass filter ranging from 0.01 Hz to 0.08 Hz was applied.

## Experimental results

After the preprocessing step, we apply PCA to retrieve the most informative components ( $V = 76,950$ ). The reduced data is then used as input for our BSS framework. We tested our proposed models on the extracted time-course fMRI data to distinguish between Autism patients and healthy

subjects. Starting with the K-means algorithm to initialize the IVABMGGMM-CSS parameters, we evaluated the convergence of our model using these parameters with a small threshold value. This process was repeated for up to 1000 iterations to ensure the stability of our models. We used demixing metrics to separate the Autism time courses from the healthy ones for the best run. The complexity of the experiment was gradually increased, starting with two subjects and progressing to the full dataset of 100 subjects, validating the performance of our models against the complexity of the fMRI data.

Table 7.3: Performance of IVAMGGMM-CSS and IVABMGGMM-CSS using SDR separation measure.

Combinations/Models	SDR					
	IVA	IVA-G	IVAMGGMM	IVABMGGMM	IVAMGGMM-CSS	IVABMGGMM-CSS
2	15.39	14.87	14.53	15.02	<b>18.78</b>	<b>25.72</b>
5	14.72	15.15	14.54	14.72	<b>16.70</b>	<b>23.47</b>
10	13.59	13.70	13.92	14.30	<b>14.89</b>	<b>20.63</b>
20	12.88	12.85	12.73	12.95	<b>13.12</b>	<b>17.53</b>
40	10.85	10.61	11.03	11.58	<b>12.44</b>	<b>16.03</b>
60	6.81	7.79	8.14	8.63	<b>11.46</b>	<b>14.61</b>
80	4.85	6.09	7.63	7.84	<b>9.65</b>	<b>12.75</b>
100	3.87	2.08	3.71	4.62	<b>7.84</b>	<b>10.66</b>

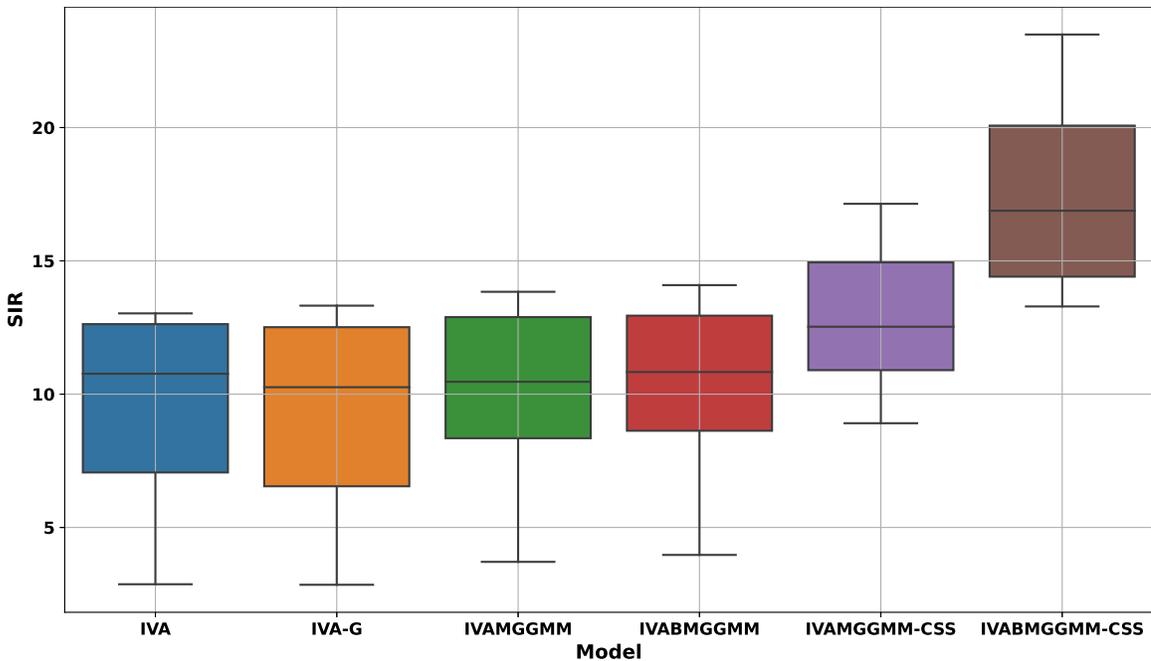


Figure 7.12: The IVABMGGMM-CSS performance in separating the Autism subjects using SIR measurement.

Table 7.3 details the performance of various models, including IVA, IVA-G, IVAMGGMM, IV-ABMGGMM, IVAMGGMM-CSS, and IVABMGGMM-CSS. As illustrated, the proposed models consistently outperform the baseline models across all combinations. This is particularly evident in high-dimensional combinations, where our models demonstrate superior performance in distinguishing Autism patients from healthy controls. Specifically, for the combination of 100 subjects, the IVABMGGMM-CSS model achieves the highest SDR value (SDR = 10.66), indicating it as the best-performing model among those evaluated models. In contrast, the baseline models show a decline in performance as the number of subjects increases. This reflects that the base models over-parameterize the common SCVs, leading to the poor performance to separate the common patterns in the ASD patients. The results clearly demonstrate that the IVABMGGMM-CSS model achieves the highest SDR values across all combinations, reflecting its superior performance in minimizing distortion and considering the common ASD structures. This model’s ability to maintain higher SDR values, especially in more complex scenarios, highlights its robustness and effectiveness in managing intricate data challenges.

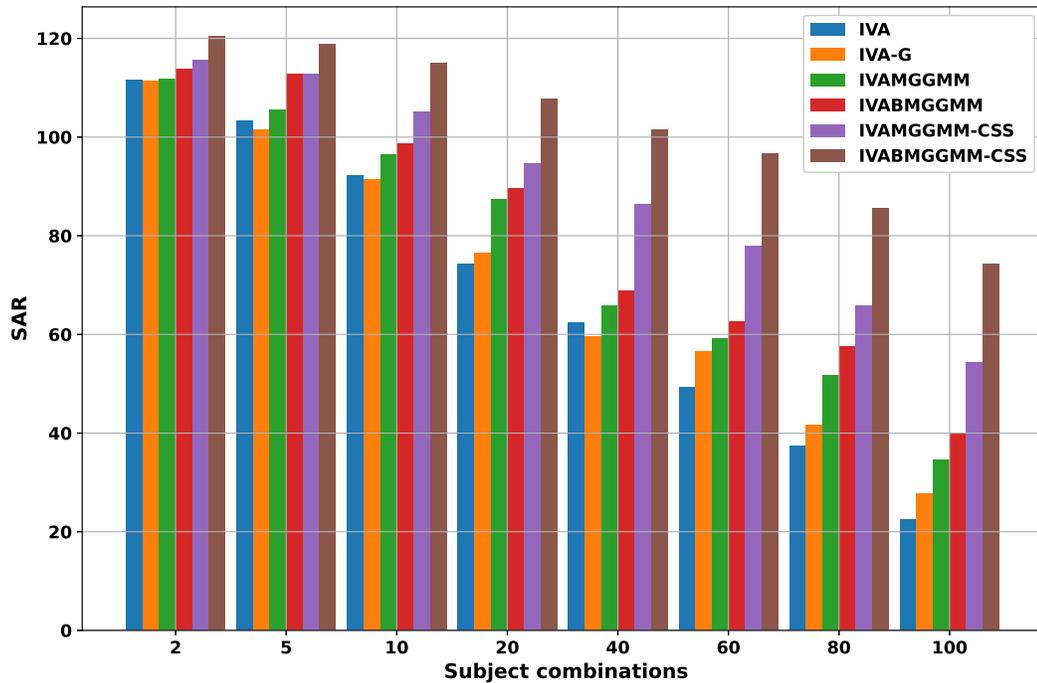


Figure 7.13: SAR measurements for our proposed models compared to base models.

Fig. 7.12 exhibits the boxplot comparing SIR for our proposed models and the baseline models.

This comparison is made across both low and high levels of subject combinations. The boxplot clearly illustrates the superior performance of the IVABMGGMM-CSS model, which achieves the highest median SIR values among all the models evaluated. This model's ability to maintain higher SIR values, despite exhibiting some variability, highlights its robustness and effectiveness in handling complex data scenarios. The IVAMGGMM-CSS model also shows significant performance improvements, reinforcing the benefits of incorporating additional constraints and methodologies in enhancing model efficacy to consider the common SCVs in ASD. In contrast, the baseline models (IVA and IVA-G) exhibit the lowest median SIR values, reflecting their limitations in minimizing interference. The modest improvements observed in IVAMGGMM and IVABMGGMM suggest incremental gains over the baseline models. The superiority of our models becomes particularly evident in high-dimensional scenarios. Notably, for combinations of 80 and 100 subjects, our proposed IVABMGGMM-CSS model achieves SIR values of 14.04 and 13.29, respectively, demonstrating significant improvement over the traditional methods.

Fig. 7.13 presents the SAR metrics comparison across various models IVA, IVA-G, IVAMGGMM, IVABMGGMM, IVAMGGMM-CSS, and IVABMGGMM-CSS for different subject combinations. The chart provides a clear visualization of the SAR performance as the number of subject combinations increases from 2 to 100. Notably, the IVABMGGMM-CSS model consistently achieves the highest SAR values across all combinations, demonstrating its superior performance in separating signals from interference and analyzing the common ASD patterns. This is particularly evident in high-dimensional scenarios, such as the 80 and 100 subject combinations, where IVABMGGMM-CSS reaches SAR values of approximately 85 and 74, respectively. The baseline models, IVA and IVA-G, exhibit lower SAR values, indicating less effectiveness in managing interference as the complexity of the data increases. The intermediate performance of IVAMGGMM and IVABMGGMM models, compared to their common subgroup versions, highlights the significant improvements gained through the advanced configurations of IVAMGGMM-CSS and IVABMGGMM-CSS. Overall, this figure underscores the robustness and efficacy of the proposed models to separate Autism patients from normal controls, particularly the IVABMGGMM-CSS, in handling high-dimensional data, the common SCVs, and achieving superior signal reconstruction quality.

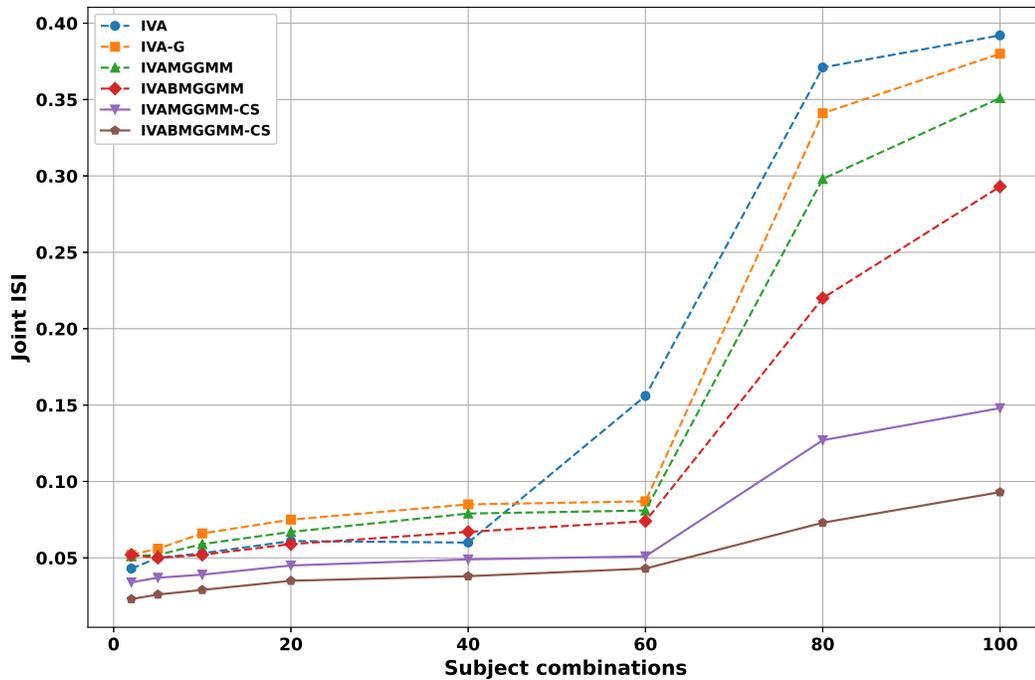


Figure 7.14: The performance of IVABMGGMM-CSS in terms of Joint-ISI.

To validate the separation quality, we use the Joint-ISI to measure the interference in the estimated time-courses. Fig. 7.14 illustrates the performance of various models IVA, IVA-G, IVAMGGMM, IVABMGGMM, IVAMGGMM-CSS, and IVABMGGMM-CSS in terms of Joint-ISI across different combinations of subjects. The graph demonstrates a clear distinction in performance as the number of combinations increases from 2 to 100. Notably, the IVABMGGMM-CSS model consistently achieves the lowest Joint-ISI values, indicating its superior ability to minimize interference and enhance the time-courses separation. Specifically, at combinations of 80 and 100 subjects, the IVABMGGMM-CSS model maintains a low Joint-ISI, demonstrating its robustness and effectiveness in handling high-dimensional data scenarios and extracting the common ASD structures. In contrast, baseline models such as IVA and IVA-G show a significant increase in Joint-ISI with more combinations, reflecting their diminished performance in more complex scenarios. Also, this indicates that the PDF of these base models overparameterizes the common SCVs in the ASD data. Intermediate models like IVAMGGMM and IVABMGGMM perform better than the baseline models but are outperformed by their common subspace versions. Overall, the figure underscores

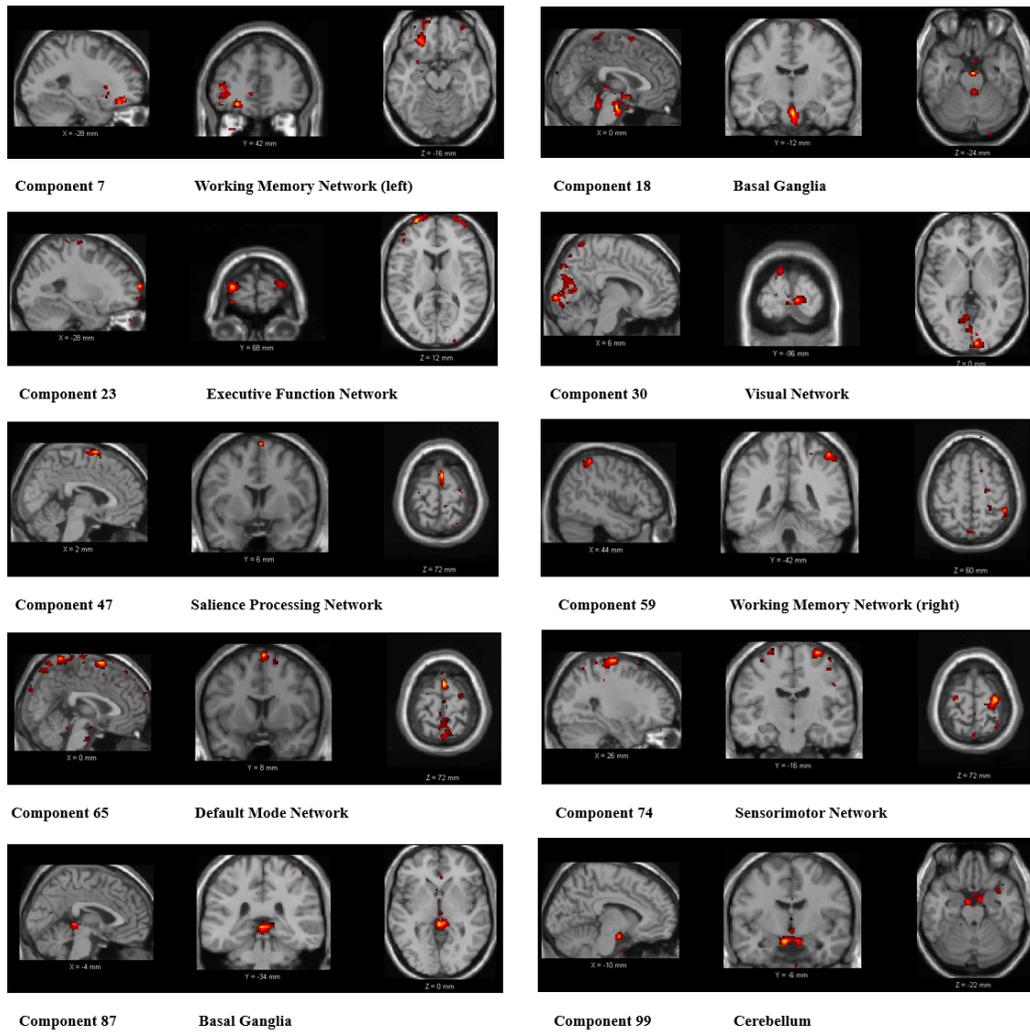


Figure 7.15: The most discriminating IVA components for Autism patients.

the enhanced performance of IVABMGGMM-CSS, showcasing its capability to maintain low interference levels and high signal integrity across varying combinations of subjects. This superior performance highlights the potential of IVABMGGMM-CSS in applications requiring high fidelity and precision in signal separation. It also underscores the advantage of incorporating the common subgroup analysis into the cost function of the proposed IVABMGGMM-CSS method.

### Brain network identification and statistical analysis for Autism patients

We validated our proposed model using a two-sample t-test to identify the most significant brain regions affected by Autism. For this analysis, we applied the FDR correction method with

Table 7.4: Peak activation in the t-maps of the resting-state networks that show group differences.

Brain network	IVA Components	Anatomical description	Abbreviation	Hemisphere	T value	Coordinates (X, Y, Z)
Working memory network (left)	7	Superior frontal gyrus, medial orbital	PFCventmed	L	1.96	-28, 42, -16
		Inferior frontal gyrus, triangular part	IFGtriang	L		
Basal Ganglia	18	Lobule I, II of vermis	VER1_2	L	2.3	2, -35, -17
	87	Lingual gyrus	LING	R		5, -34, 0
Executive function network	23	Superior frontal gyrus, dorsolateral	SFG	L	2.4	-28, 63, 13
		Superior frontal gyrus, dorsolateral	SFG	R		
Visual network	30	Calcarine fissure and surrounding cortex	CAL	R	1.98	13, -94, 0
		Calcarine fissure and surrounding cortex	CAL	L		0, -82, 0
		Lingual gyrus	LING	L		-8, -72, 0
		Middle occipital gyrus	MOG	L		-8, -102, 2
Salience network	47	Supplementary motor area	SMA	R	2.0	2, 6, 72
Working memory network (right)	59	Superior parietal gyrus	SPG	R	2.3	43, -42, 61
Default mode network	65	Supplementary motor area	SMA	L	2.5	0, 8, 71
		Precuneus	PCUN	R		0, -51, 70
Sensorimotor network	74	Precentral gyrus	PreCG	R	2.7	26, -16, 72
		Precentral gyrus	PreCG	L		-29, -16, 71
Cerebellum	99	Parahippocampal gyrus	PHG	R	3.2	21, 8, -31

$p < 0.05$  and a threshold greater than 1.96. Fig. 7.15 illustrates the IVABMGGMM-CSS components and their corresponding brain networks, demonstrating our model’s capability to extract the brain regions impacted by Autism. The most affected common brain networks include the Working memory network left and right, Basal ganglia (BG), Executive function network, Visual network, Salience network, Default mode network, Sensorimotor network, and Cerebellum. As shown in Fig. 7.15 and Table 7.4 these brain networks are linked to specific brain regions. IV7 and IV59 include the superior frontal gyrus (medial orbital) and inferior frontal gyrus (triangular portion) for the left WMN, and the superior parietal gyrus for the right WMN. These regions are critical for preserving and manipulating information over short periods of time, reflecting the working memory problems commonly reported in Autism, implying that the IVABMGGMM-CSS maintained more subject variability. Recent neuroimaging research [263] found similar abnormalities in the superior and inferior frontal gyruases, highlighting their crucial role in ASD-related cognitive difficulties. IV18 and IV87 highlight the vermis’ lobules I and II, and the lingual gyrus. The basal ganglia have a role in both motor control and cognitive functions such as learning and memory. These findings are consistent with a recent study [264], which found a reduction in the lingual gyrus of ASD patients relative to healthy people. Alterations in these regions may lead to the repetitive behaviours

and motor impairments that are commonly observed in Autism, underscoring the Basal ganglia's significance in Autism's neurobiology.

IVA23 is associated with the dorsolateral superior frontal gyrus in both hemispheres. Executive processes such as planning, decision-making, and cognitive flexibility are frequently compromised in people with Autism. The involvement of the dorsolateral prefrontal cortex indicates potential therapies to improve executive functioning. This aligns with the study [263], which reports unusual activity patterns in the superior frontal gyrus in ASD patients. The calcarine fissure, surrounding cortex, lingual gyrus, and middle occipital gyrus are all part of the IV30, which is crucial for visual perception and processing. Variations in these domains underscore the need for measures to improve visual processing and raise the possibility that people with Autism may interpret visual information differently. The IV47 is concerned with the supplemental motor area. The salience network is responsible for identifying and responding to important stimuli. Alterations here may contribute to social and sensory processing impairments in Autism, indicating a target for therapy. Our findings indicate that IV65, which is activated in the supplementary motor region and precuneus, has a role in self-referential thinking and mind wandering. Disruptions in the DMN are associated with social and communication impairments in Autism, underscoring their importance in understanding the illness. A recent study [265] also found the precuneus as a major region impacted in ASD, relating its dysfunction to the introspective and self-related cognitive difficulties reported in the disorder. The IV74 highlights the precentral gyrus, which is responsible for processing sensory input and motor output. Variations in this network may be related to sensory sensitivities and motor deficits observed in Autism, highlighting the importance of focused sensorimotor integration treatments. The IV99 involves the parahippocampal gyrus, highlighting the cerebellum's role in motor control, coordination, and cognitive functioning. Our findings show that the parahippocampal gyrus, which is known for its function in memory encoding and retrieval, has abnormal activity patterns in ASD patients. This corresponds with current literature demonstrating problems in memory and spatial navigation in ASD patients [266, 264]. The cerebellum's role in Autism suggests that it is an important target for future research and therapies.

## 7.6 Conclusion

In this Chapter, we introduced the IVAMGGMM-CSS and IVABMGGMM-CSS as powerful tools to identify mental disorders from the healthy controls. In the IVAMGGMM-CSS method, we used the full covariance matrix to extract the correlated features in the multivariate data, like resting state fMRI datasets. For the complex scenarios, we proposed the IVABMGGMM-CSS algorithm to consider the high-dimensional data fall in the bounded support regions and analyze the common components in these mental diseases. We validated our proposed methods using simulated and real-world applications. For that, we conducted our models to distinguish between the Schizophrenic patients and the normal controls. Then, we used the IVABMGGMM-CSS components to identify the affected brain regions due to the Schizophrenia disease. Our proposed models demonstrated a strong ability to distinguish Schizophrenic subjects from healthy controls. Furthermore, we employed our methods on the Autism data to differentiate between the Autism time-courses and the normal ones. We also applied the two sample t-tests on the IVABMGGMM-CSS components to extract the abnormal brain regions affected by Autism disease. Using these experiments, we highlight the effectiveness of our model in separating the abnormal time-courses from the normal ones. Additionally, we emphasized that incorporating the analysis of common subspaces in our models significantly enhances the detection of these structures in the brain. Through various experiments, our models demonstrated their ability to generalize across different data types and scales.

## Chapter 8

# Conclusion and future work

In this chapter, we summarize the main contributions of this dissertation and highlight future research directions.

### 8.1 Conclusions

In this dissertation, we introduced the bounded generalized Gaussian mixture model to address the limitations of unbounded models in representing data confined within bounded support regions. Building upon this foundation, we proposed the ICABMGGMM and IVABMGGMM frameworks to overcome the respective shortcomings of standard ICA and IVA. The proposed models demonstrated significant performance improvements across a diverse range of applications, as validated through comprehensive experimental evaluations. To enhance their clustering capabilities, we integrated these models into a hidden Markov model framework. This integration led to the development of the ICABMGGMM-HMM and IVABMGGMM-HMM, which serve as robust BSS methods. These advanced methods have proven to be highly effective in applications such as human action recognition and speech recognition, showcasing their versatility and superior performance.

Our experiments revealed a notable limitation of the proposed ICABMGGMM and IVABMGGMM models when applied to large-scale datasets. To address this issue, we developed the adaptive constrained ICAMGGMM and IVAMGGMM, which have shown promising performance in identifying mentally challenged subjects. Experimental results demonstrate a significant improvement in

distinguishing brain volumes associated with mental diseases from those of healthy individuals. To evaluate the generalization capability of our models, we further introduced bounded versions of the adaptive constrained ICAMGGMM and IVAMGGMM. These enhanced models were successfully applied to ECG separation tasks, further demonstrating their adaptability and effectiveness across diverse applications.

Finally, to enhance the scalability of our methods, we incorporated the ICABMGGMM and IVABMGGMM into a common subspaces analysis framework. This integration allows the models to effectively account for common and non-common components in large-scale datasets, which are often overparameterized by other BSS methods. These models are particularly well-suited for applications that require the identification of common components, such as rs-fMRI data. This capability is crucial for capturing similar patterns, which play a significant role in the analysis and diagnosis of mental diseases. By leveraging this framework, our refined models are better equipped to adapt to the unique structural properties of such datasets, ensuring improved accuracy and robustness in component analysis. To evaluate the robustness of our models, we conducted two-sample t-tests and performed graphical analyses. These assessments confirmed the resilience of our models and their potential for precise classification in mental health diagnostics. Specifically, our experiments demonstrated the models' effectiveness in distinguishing abnormal time courses from normal ones. Furthermore, the models exhibited a remarkable ability to generalize across different data types and scales, solidifying their versatility and applicability in diverse scenarios.

## **8.2 Future Directions**

The success achieved by the proposed methods opens several promising avenues for future research. One research area of interest involves the study of dynamic functional network connectivity (dFNC) for the rs-fMRI datasets. The dFNC analysis has become increasingly important, revealing that FNC evolves throughout scanning sessions. Traditional stationary FNC analysis assumes constant connectivity patterns, offering only an average measure and limiting insights into the brain's dynamic behavior. In contrast, studies have shown multiple FNC states in both task-related and resting-state fMRI data, helping identify biomarkers for disorders such as Schizophrenia, Autism,

Alzheimer, and mild cognitive impairment. However, most dFNC methods assume the spatial structure of networks remains stationary, overlooking spatial variability. To overcome the aforementioned limitations in identifying dynamic brain activity, a model that captures both spatial and temporal changes would provide a more detailed understanding of the brain's dynamic processes.

Another promising direction is to explore the explainability and interpretability of our models. By integrating explainable artificial intelligence (XAI) methodologies, we aim to enhance the usability and transparency of our proposed models, particularly for mental health disease detection. XAI tools provide valuable insights for both researchers and clinicians, enabling a deeper understanding of the underlying processes and improving the accuracy and interpretability of the models used in mental health diagnostics. By translating complex algorithmic operations into accessible and actionable insights, these tools bridge the gap between computational models and human comprehension. This integration not only empowers clinicians with more informed decision-making capabilities but also drives significant advancements in precision diagnostics, ultimately contributing to better outcomes in mental health care.

## Appendix A

# Proof of BMGGMM parameters estimation (Chapter 2)

The complete log-likelihood function for MGGMM is given by:

$$\begin{aligned} \mathcal{L}(\Theta, Z, \mathcal{X}) &= \sum_{n=1}^N Z_{nj} \left\{ \log(\beta_j) - \log \Gamma\left(\frac{D}{2\beta_j}\right) - \frac{D}{2\beta_j} \log(2) - \frac{D}{2} \log(\mathbf{c}_j) \right. \\ &\quad \left. - \frac{1}{2} \Sigma_j - \frac{1}{2(\mathbf{c}_j)\beta_j} \left( (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j) \right)^{\beta_j} \right\} \end{aligned} \quad (\text{A.1})$$

### A.1 Mean estimation

To estimate the mean, the process of maximizing (A.1) with respect to  $\boldsymbol{\mu}_j$  is detailed in Equations (A.2) to (A.3).

$$\begin{aligned} \frac{\partial \mathcal{L}(\Theta, Z, \mathcal{X})}{\partial \boldsymbol{\mu}_j} &= \sum_{n=1}^N \hat{Z}_{nj} \left\{ \frac{\beta_j}{2(\mathbf{c}_j)\beta_j} \left[ \left( (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j) \right)^{\beta_j - 1} \left( (\mathbf{x}_n - \boldsymbol{\mu}_j) \Sigma_j^{-1} + (\mathbf{x}_n - \boldsymbol{\mu}_j) \Sigma_j^{-1} \right) \right] \right. \\ &\quad \left. - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left\{ \frac{\beta_j}{2(\mathbf{c}_j)\beta_j} \left[ \mathbf{y}^{\beta_j - 1} \Sigma_j^{-1} \left( (\mathbf{u} - \boldsymbol{\mu}_j) + (\mathbf{u} - \boldsymbol{\mu}_j) \right) \right] \right\} du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \right\} \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\Theta, Z, \mathcal{X})}{\partial^2 \boldsymbol{\mu}_j} &= \sum_{n=1}^N \hat{Z}_{nj} \left( \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left\{ (\beta_j - 1) \left[ (\mathbf{x}_n - \boldsymbol{\mu}_j) \Sigma_j^{-1} \right]^2 \mathbf{y}^{(\beta_j-2)} - \left( \Sigma_j^{-1} + \Sigma_j^{-1} \right) \mathbf{y}^{(\beta_j-1)} \right\} \right. \\
&+ \left. \frac{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left[ \mathbf{y}_u^{(\beta_j-1)} \left( \Sigma_j^{-1} \left\{ (\mathbf{u} - \boldsymbol{\mu}_j) + (\mathbf{u} - \boldsymbol{\mu}_j) \right\} \right) \right] du \right)^2}{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) du \right)^2} \right. \\
&- \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left( \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left[ \mathbf{y}_u^{(\beta_j-1)} \left( \Sigma_j^{-1} \left\{ (\mathbf{u} - \boldsymbol{\mu}_j) + (\mathbf{u} - \boldsymbol{\mu}_j) \right\} \right) \right] \right)^2 du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \\
&- \frac{1}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \int_{\Omega_j} f(\mathbf{u}|\xi_j) \left\{ \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} (\beta_j - 1) \left[ \left( \Sigma_j^{-1} \left\{ (\mathbf{u} - \boldsymbol{\mu}_j) \right\} \right)^2 \mathbf{y}_u^{(\beta_j-2)} \right. \right. \\
&- \left. \left. \left( \Sigma_j^{-1} + \Sigma_j^{-1} \right) \mathbf{y}_u^{(\beta_j-1)} \right] du \right\}
\end{aligned} \tag{A.3}$$

where  $\mathbf{y} = (\mathbf{x}_n - \boldsymbol{\mu}_j) \Sigma_j^{-1} + (\mathbf{x}_n - \boldsymbol{\mu}_j)$  and  $\mathbf{y}_u = (\mathbf{u} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{u} - \boldsymbol{\mu}_j)$ .

$$\begin{aligned}
|F(\boldsymbol{\mu}_j)| &= - \prod_{d=1}^D \sum_{n=1}^N \hat{Z}_{nj} \left( \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left\{ (\beta_j - 1) \left( (\mathbf{x}_n - \boldsymbol{\mu}_j) \Sigma_j^{-1} + (\mathbf{x}_n - \boldsymbol{\mu}_j) \Sigma_j^{-1} \right)^2 \right. \right. \\
&\times \left. \left. \mathbf{y}^{(\beta_j-2)} - \left( \Sigma_j^{-1} + \Sigma_j^{-1} \right) \mathbf{y}^{(\beta_j-1)} \right\} \right. \\
&+ \left. \frac{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left[ \mathbf{y}_u^{(\beta_j-1)} \Sigma_j^{-1} \left\{ (\mathbf{u} - \boldsymbol{\mu}_j) + (\mathbf{u} - \boldsymbol{\mu}_j) \right\} \right] du \right)^2}{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) du \right)^2} \right. \\
&- \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left( \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left[ \mathbf{y}_u^{(\beta_j-1)} \Sigma_j^{-1} \left\{ (\mathbf{u} - \boldsymbol{\mu}_j) + (\mathbf{u} - \boldsymbol{\mu}_j) \right\} \right] \right)^2 du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \\
&- \frac{1}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \int_{\Omega_j} f(\mathbf{u}|\xi_j) \left\{ \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} (\beta_j - 1) \left( \Sigma_j^{-1} \left\{ (\mathbf{u} - \boldsymbol{\mu}_j) \right. \right. \right. \\
&+ \left. \left. \left. (\mathbf{u} - \boldsymbol{\mu}_j) \right\} \right)^2 \mathbf{y}_u^{(\beta_j-2)} - \left( \Sigma_j^{-1} + \Sigma_j^{-1} \right) \mathbf{y}_u^{(\beta_j-1)} \right\} du
\end{aligned} \tag{A.4}$$

where  $\mathbf{y} = (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j)$  &  $\mathbf{y}_u = (\mathbf{u} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{u} - \boldsymbol{\mu}_j)$ . Taking into account the approximations for estimating the mean parameter as discussed in Section 2.2,  $|F(\boldsymbol{\mu}_j)|$  can be written as:

$$\begin{aligned}
|F(\boldsymbol{\mu}_j)| &= - \prod_{d=1}^D \sum_{n=1}^N \hat{Z}_{nj} \left( \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left\{ (\beta_j - 1) \left( (\mathbf{x}_n - \boldsymbol{\mu}_j)^{\Sigma_j^{-1}} + (\mathbf{x}_n - \boldsymbol{\mu}_j)^{\Sigma_j^{-1}} \right)^2 \right. \right. \\
&\times \left. \left. \mathbf{y}^{(\beta_j-2)} - \left( \Sigma_j^{-1} + \Sigma_j^{-1} \right) \mathbf{y}^{(\beta_j-1)} \right\} \right. \\
&+ \left. \frac{\left( \sum_{v=1}^V \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left[ \mathbf{y}_s^{(\beta_j-1) \Sigma_j^{-1}} \left\{ (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j) + (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j) \right\} \right] H(\mathbf{s}_{j_v} | j) \right)^2}{\left( \sum_{v=1}^V H(\mathbf{s}_{j_v} | j) \right)^2} \right. \\
&- \left. \frac{\sum_{v=1}^V \left( \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left[ \mathbf{y}_s^{(\beta_j-1) \Sigma_j^{-1}} \left\{ (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j) + (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j) \right\} \right] \right)^2 H(\mathbf{s}_{j_v} | j)}{\sum_{v=1}^V H(\mathbf{s}_{j_v} | j)} \right. \\
&- \left. \frac{1}{\sum_{v=1}^V H(\mathbf{s}_{j_v} | j)} \sum_{v=1}^V \left\{ \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} (\beta_j - 1) \left( \Sigma_j^{-1} \left\{ (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j) + (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j) \right\} \right)^2 \right. \right. \\
&\times \left. \left. \mathbf{y}_s^{(\beta_j-2)} - \left( \Sigma_j^{-1} + \Sigma_j^{-1} \right) \mathbf{y}_s^{(\beta_j-1)} \right\} H(\mathbf{s}_{j_v} | j) \right) \tag{A.5}
\end{aligned}$$

where the term  $\mathbf{y}_s$  is represented as:  $\mathbf{y}_s = (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j)$ .

## A.2 Covariance estimation

To estimate the covariance matrix, we derived Equation (A.1) with respect to  $\Sigma_j$  as in Equations (A.7) and (A.7).

$$\begin{aligned}
\frac{\partial \mathcal{L}(\Theta, Z, \mathcal{X})}{\partial \Sigma_j} &= \sum_{n=1}^N \hat{Z}_{nj} \left\{ \frac{\beta_j}{(\mathbf{c}_j)^{\beta_j}} \sum_{n=1}^N \mathbf{y}^{\beta_j-1} (\mathbf{x}_n - \boldsymbol{\mu}_j)^{\Sigma_j^{-2}} (\mathbf{x}_n - \boldsymbol{\mu}_j)^T - \frac{1}{2} \Sigma_j^{-1} \right. \\
&- \left. \frac{\int_{\Omega_j} f(\mathbf{u} | \xi_j) \left\{ \frac{\beta_j}{(\mathbf{c}_j)^{\beta_j}} \sum_{n=1}^N \mathbf{y}_u^{\beta_j-1} (\mathbf{u} - \boldsymbol{\mu}_j)^{\Sigma_j^{-2}} (\mathbf{u} - \boldsymbol{\mu}_j)^T - \frac{1}{2} \Sigma_j^{-1} \right\} du}{\int_{\Omega_j} f(\mathbf{u} | \xi_j) du} \right\} \tag{A.6}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\Theta, Z, \mathcal{X})}{\partial^2 \Sigma_j} &= \sum_{n=1}^N \hat{Z}_{nj} \left[ \frac{1}{2} \Sigma_j^{-2} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( (1 - \beta_j) \mathbf{y}^{(\beta_j-2)} U^2 \right. \right. \\
&\quad \left. \left. - 2(\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-3} (\mathbf{x}_n - \boldsymbol{\mu}_j) \mathbf{y}^{(\beta_j-1)} \right) \right. \\
&\quad \left. + \frac{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) \left( -\frac{1}{2} \Sigma_j^{-1} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( \mathbf{y}_u^{(\beta_j-1)} U_u \right) \right) du \right)^2}{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) du \right)^2} \right. \\
&\quad \left. - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left( -\frac{1}{2} \Sigma_j^{-1} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( \mathbf{y}_u^{(\beta_j-1)} U_u \right) \right)^2 du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \right. \\
&\quad \left. - \frac{1}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \int_{\Omega_j} f(\mathbf{u}|\xi_j) \left\{ \frac{1}{2} \Sigma_j^{-2} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( (1 - \beta_j) \right. \right. \right. \\
&\quad \left. \left. \times \mathbf{y}_u^{(\beta_j-2)} U_u^2 - 2(\mathbf{u} - \boldsymbol{\mu}_j)^T \Sigma_j^{-3} (\mathbf{u} - \boldsymbol{\mu}_j) \mathbf{y}_u^{(\beta_j-1)} \right) \right\} du \right] \quad (\text{A.7})
\end{aligned}$$

where  $U = (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-2} (\mathbf{x}_n - \boldsymbol{\mu}_j)$  and  $U_u = (\mathbf{u} - \boldsymbol{\mu}_j)^T \Sigma_j^{-2} (\mathbf{u} - \boldsymbol{\mu}_j)$ .

$$\begin{aligned}
|F(\Sigma_j)| &= - \sum_{n=1}^N \hat{Z}_{nj} \left( \frac{1}{2} \Sigma_j^{-2} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( (1 - \beta_j) \mathbf{y}^{(\beta_j-2)} U^2 \right. \right. \\
&\quad \left. \left. - 2(\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-3} (\mathbf{x}_n - \boldsymbol{\mu}_j) \mathbf{y}^{(\beta_j-1)} \right) \right. \\
&\quad \left. + \frac{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) \left( -\frac{1}{2} \Sigma_j^{-1} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( \mathbf{y}_u^{(\beta_j-1)} U_u \right) \right) du \right)^2}{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) du \right)^2} \right. \\
&\quad \left. - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left( -\frac{1}{2} \Sigma_j^{-1} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( \mathbf{y}_u^{(\beta_j-1)} U_u \right) \right)^2 du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \right. \\
&\quad \left. - \frac{1}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \int_{\Omega_j} f(\mathbf{u}|\xi_j) \left\{ \frac{1}{2} \Sigma_j^{-2} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( (1 - \beta_j) \right. \right. \right. \\
&\quad \left. \left. \times \mathbf{y}_u^{(\beta_j-2)} U_u^2 - 2(\mathbf{u} - \boldsymbol{\mu}_j)^T \Sigma_j^{-3} (\mathbf{u} - \boldsymbol{\mu}_j) \mathbf{y}_u^{(\beta_j-1)} \right) \right\} du \right) \quad (\text{A.8})
\end{aligned}$$

where the terms  $U$  and  $U_u$  are represented as:  $U = (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-2} (\mathbf{x}_n - \boldsymbol{\mu}_j)$  and  $U_u = (\mathbf{u} - \boldsymbol{\mu}_j)^T \Sigma_j^{-2} (\mathbf{u} - \boldsymbol{\mu}_j)$  respectively. Considering approximations for the covariance matrix estimation as described in Section 2.2,  $|F(\Sigma_j)|$  is given as:

$$\begin{aligned}
|F(\Sigma_j)| &= -\sum_{n=1}^N \hat{Z}_{nj} \left( \frac{1}{2} \Sigma_j^{-2} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( (1 - \beta_j) \mathbf{y}^{(\beta_j-2)} U^2 \right. \right. \\
&\quad \left. \left. - 2(\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-3} (\mathbf{x}_n - \boldsymbol{\mu}_j) \mathbf{y}^{(\beta_j-1)} \right) \right) \\
&\quad + \frac{\left( \sum_{v=1}^V \left( \frac{-1}{2} \Sigma_j^{-1} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} (\mathbf{y}_s^{(\beta_j-1)} U_s) \right) H(\mathbf{s}_{j_v} | j) \right)^2}{\left( \sum_{v=1}^V H(\mathbf{s}_{j_v} | j) \right)^2} \\
&\quad - \frac{\sum_{v=1}^V \left( \frac{-1}{2} \Sigma_j^{-1} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} (\mathbf{y}_s^{(\beta_j-1)} U_s) \right)^2 H(\mathbf{s}_{j_v} | j)}{\sum_{v=1}^V H(\mathbf{s}_{j_v} | j)} \\
&\quad - \frac{1}{\sum_{v=1}^V H(\mathbf{s}_{j_v} | j)} \sum_{v=1}^V \left\{ \frac{1}{2} \Sigma_j^{-2} + \frac{\beta_j}{2(\mathbf{c}_j)^{\beta_j}} \left( (1 - \beta_j) \mathbf{y}_s^{(\beta_j-2)} U_s^2 \right. \right. \\
&\quad \left. \left. - 2(\mathbf{s}_{j_v} - \boldsymbol{\mu}_j)^T \Sigma_j^{-3} (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j) \mathbf{y}_s^{(\beta_j-1)} \right) \right\} H(\mathbf{s}_{j_v} | j) \tag{A.9}
\end{aligned}$$

where  $U_s = (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j)^T \Sigma_j^{-2} (\mathbf{s}_{j_v} - \boldsymbol{\mu}_j)$ .

### A.3 Shape parameter estimation

Similarly, we derived Equation (A.1) with respect to  $\beta_j$  to estimate the shape parameter: Equations (A.11) and (A.11) explain steps to get the shape parameter.

$$\begin{aligned}
\frac{\partial \mathcal{L}(\Theta, Z, \mathcal{X})}{\partial \beta_j} &= \sum_{n=1}^N \hat{Z}_{nj} \left\{ Q - \frac{\mathbf{y}^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left[ \log(\mathbf{c}_j) - \log \left( (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j) \right) \right] \right. \\
&\quad \left. - \frac{\int_{\Omega_j} f(\mathbf{u} | \xi_j) \left\{ Q - \frac{\mathbf{y}_u^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left[ \log(\mathbf{c}_j) - \log \left( (\mathbf{u} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{u} - \boldsymbol{\mu}_j) \right) \right] \right\} du}{\int_{\Omega_j} f(\mathbf{u} | \xi_j) du} \right\} \tag{A.10}
\end{aligned}$$

where  $Q = \frac{1}{\beta_j} + \frac{d}{2\beta_j} \psi\left(\frac{d}{2\beta_j}\right) + \frac{d}{2\beta_j^2} \log(2)$ ,  $\mathbf{y} = (\mathbf{x}_n - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_j)$  and  $\mathbf{y}_u = (\mathbf{u} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{u} - \boldsymbol{\mu}_j)$ .

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\Theta, Z, \mathcal{X})}{\partial^2 \beta_j} &= \sum_{n=1}^N \hat{Z}_{nj} \left\{ Q + \frac{\mathbf{y}^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}) \right)^2 \right. \\
&\quad + \frac{\left[ \int_{\Omega_j} f(\mathbf{u}|\xi_j) \left( Q + \frac{\mathbf{y}_u^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}_u) \right) \right) du \right]^2}{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) du \right)^2} \\
&\quad - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left[ \left( Q + \frac{\mathbf{y}_u^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}_u) \right) \right) \right]^2 du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \\
&\quad \left. - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left[ Q + \frac{\mathbf{y}_u^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}_u) \right) \right]^2 du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \right\} \quad (\text{A.11})
\end{aligned}$$

$$\begin{aligned}
|F(\beta_j)| &= - \sum_{n=1}^N \hat{Z}_{nj} \left( Q + \frac{\mathbf{y}^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}) \right)^2 \right. \\
&\quad + \frac{\left[ \int_{\Omega_j} f(\mathbf{u}|\xi_j) \left( Q + \frac{\mathbf{y}_u^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}_u) \right) \right) du \right]^2}{\left( \int_{\Omega_j} f(\mathbf{u}|\xi_j) du \right)^2} \\
&\quad - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left[ \left( Q + \frac{\mathbf{y}_u^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}_u) \right) \right) \right]^2 du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \\
&\quad \left. - \frac{\int_{\Omega_j} f(\mathbf{u}|\xi_j) \left[ Q + \frac{\mathbf{y}_u^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}_u) \right) \right]^2 du}{\int_{\Omega_j} f(\mathbf{u}|\xi_j) du} \right) \quad (\text{A.12})
\end{aligned}$$

where  $Q = \frac{1}{\beta_j} + \frac{d}{\beta_j^3} \psi\left(\frac{d}{2\beta_j}\right) + \left(\frac{d}{2\beta_j}\right)^2 \psi\left(\frac{d}{2\beta_j}\right) + \frac{d}{\beta_j^3} \log(2)$ . Using approximations for the shape parameter estimation as explained in Section 2.2,  $|F(\beta_j)|$  can be written as:

$$\begin{aligned}
|F(\beta_j)| &= -\sum_{n=1}^N \hat{Z}_{nj} \left( Q + \frac{\mathbf{y}^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}) \right) \right)^2 \\
&+ \frac{\left[ \sum_{v=1}^V \left( Q + \frac{\mathbf{y}^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}_s) \right) \right) H(\mathbf{s}_{j_v}|j) \right]^2}{\left( \sum_{v=1}^V H(\mathbf{s}_{j_v}|j) \right)^2} \\
&- \frac{\sum_{v=1}^V \left[ \left( Q + \frac{\mathbf{y}^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}_s) \right) \right) \right]^2 H(\mathbf{s}_{j_v}|j)}{\sum_{v=1}^V H(\mathbf{s}_{j_v}|j)} \\
&- \frac{\sum_{v=1}^V \left[ Q + \frac{\mathbf{y}^{\beta_j}}{2(\mathbf{c}_j)^{\beta_j}} \left( \log(\mathbf{c}_j) - \log(\mathbf{y}_s) \right) \right]^2 H(\mathbf{s}_{j_v}|j)}{\sum_{v=1}^V H(\mathbf{s}_{j_v}|j)} \quad (A.13)
\end{aligned}$$

## Appendix B

# ICABMGGMM-HMM and IVABMGGMM-HMM parameter estimation procedures (Chapter 3)

### B.1 Estimation of $\Lambda$

To update the parameters of our proposed models, we maximize the final term in Equation (3.7) with respect to  $\Lambda_k$ . Here,  $\Lambda_k$  is a set of parameters of the  $k$ th state emission probability distribution, where  $\Lambda_k = [p, \mu, \Sigma, \beta, A, b]$ . Starting from Gaussian mixture model, the complete log-likelihood function at state  $k$  is given by:

$$\mathcal{L}(\mathcal{X}, Z, \Theta) = \sum_{n=1}^N \sum_{j=1}^M Z_{nkj} \log \left( p(\mathbf{x}_n | \xi_{kj}) \mathbf{P}_{kj} \right) \quad (\text{B.1})$$

By substituting each  $Z_{nkj}$  with its expectation, defined as the posterior probability that the  $n$ th observation corresponds to the  $m$ th component of the mixture model at state  $k$ , we obtain:

$$\hat{Z}_{nkj} = \frac{p(\mathbf{x}_n | \xi_{kj}) \mathbf{P}_{kj}}{\sum_{j=1}^M p(\mathbf{x}_n | \xi_{kj}) \mathbf{P}_{kj}} \quad (\text{B.2})$$

By considering the gradient of the log-likelihood with respect to  $\mathbf{p}_{kj}$ ,  $\boldsymbol{\mu}_{kj}$ ,  $\Sigma_{kj}$  and  $\boldsymbol{\beta}_{kj}$  at state  $k$  for the  $j$ th component of the bounded multivariate generalized Gaussian mixture (BMGGM), Equation (3.7) can be maximized and provides a closed form solution for all mixture model parameters as follows:

$$\hat{\mathbf{p}}_{kj} = \frac{1}{N} \sum_{n=1}^N p(kj|\mathbf{x}_n) \quad (\text{B.3})$$

$$\hat{\boldsymbol{\mu}}_{kj} = \frac{\sum_{n=1}^N \hat{Z}_{nkj} \left\{ \mathbf{x}_n - \frac{\sum_{v=1}^V (\mathbf{s}_{kjv} - \boldsymbol{\mu}_{kj}) H(\mathbf{s}_{kjv}|kj)}{\sum_{v=1}^V H(\mathbf{s}_{kjv}|kj)} \right\}}{\sum_{n=1}^N \hat{Z}_{nkj}} \quad (\text{B.4})$$

$$\hat{\Sigma}_{kj} = \frac{1}{\sum_{n=1}^N \hat{Z}_{nkj}} \sum_{n=1}^N \hat{Z}_{nkj} \left( \frac{\boldsymbol{\beta}_{kj}}{\mathbf{c}_j \boldsymbol{\beta}_{kj}} \sum_{n=1}^N \mathbf{y}^{\boldsymbol{\beta}_{kj}-1} (\mathbf{x}_n - \boldsymbol{\mu}_{kj}) (\mathbf{x}_n - \boldsymbol{\mu}_{kj})^T - \frac{\sum_{v=1}^V \left[ \frac{\boldsymbol{\beta}_{kj}}{\mathbf{c}_j \boldsymbol{\beta}_{kj}} \sum_{n=1}^N \mathbf{y}^{\boldsymbol{\beta}_{kj}-1} (\mathbf{s}_{kjv} - \boldsymbol{\mu}_{kj}) (\mathbf{s}_{kjv} - \boldsymbol{\mu}_{kj})^T - \Sigma_{kj} \right] H(\mathbf{s}_{kjv}|kj)}{\sum_{v=1}^V H(\mathbf{s}_{kjv}|kj)} \right) \right) \quad (\text{B.5})$$

$$\hat{\beta}_{k+1} = \hat{\beta}_{kj} - \frac{\alpha(\hat{\beta}_{kj})}{\hat{\alpha}(\hat{\beta}_{kj})} \quad (\text{B.6})$$

$$\alpha(\hat{\beta}_{kj}) = \frac{\sum_{n=1}^N \hat{Z}_{nkj} \left\{ Q - \frac{\mathbf{y}^{\boldsymbol{\beta}_{kj}}}{2\mathbf{c}_j \boldsymbol{\beta}_{kj}} \left[ \log(\mathbf{y}) - \log(\mathbf{c}_j) \right] - \frac{\sum_{v=1}^V \left\{ Q - \frac{\mathbf{y}_s^{\boldsymbol{\beta}_{kj}}}{2\mathbf{c}_j \boldsymbol{\beta}_{kj}} \left[ \log(\mathbf{y}_s) - \log(\mathbf{c}_j) \right] \right\} H(\mathbf{s}_{kjv}|kj)}{\sum_{v=1}^V H(\mathbf{s}_{kjv}|kj)} \right\}}{\sum_{n=1}^N \hat{Z}_{nkj}} \quad (\text{B.8})$$

where  $\mathbf{s}_{kj_v}$  represents the random variable that is derived from the generalized Gaussian distribution for the  $m$ th component of the mixture model.  $V$  is the number of random variables  $\mathbf{s}_{kj_v}$ ,  $c$  is the scale parameter,  $\mathbf{y} = (\mathbf{x}_n - \boldsymbol{\mu}_{kj})^T \boldsymbol{\Sigma}_{kj}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_{kj})$ ,  $\mathbf{y}_s = (\mathbf{s}_{kj_v} - \boldsymbol{\mu}_{kj})^T \boldsymbol{\Sigma}_{kj}^{-1} (\mathbf{s}_{kj_v} - \boldsymbol{\mu}_{kj})$ ,  $Q = \frac{1}{\beta_{kj}} + \frac{d}{2\beta_{kj}} \psi\left(\frac{d}{2\beta_{kj}}\right) + \frac{d}{2\beta_{kj}^2} \log(2)$ ,  $\mathbf{y}_u = (\mathbf{u} - \boldsymbol{\mu}_{kj})^T \boldsymbol{\Sigma}_{kj}^{-1} (\mathbf{u} - \boldsymbol{\mu}_{kj})$  and  $\psi$  is the digamma function. The computation of derivative of log-likelihood with respect to  $\mathbf{p}_{kj}$ ,  $\boldsymbol{\mu}_{kj}$ ,  $\boldsymbol{\Sigma}_{kj}$  and  $\boldsymbol{\beta}_{kj}$  is given in A. For the ICA parameter estimation, the complete computation to estimate basis functions  $A_{kj}$  and bias vector  $b_{kj}$  is given in Chapter 2.

# Bibliography

- [1] P. Scanlon, D. P. Ellis, and R. B. Reilly, "Using broad phonetic group experts for improved speech recognition," *IEEE transactions on audio, speech, and language processing*, vol. 15, no. 3, pp. 803–812, 2007.
- [2] A. Algumaei, M. Azam, F. Najjar, and N. Bouguila, "Bounded multivariate generalized gaussian mixture model using ica and iva," *Pattern Analysis and Applications*, pp. 1–30, 2023.
- [3] A. Algumaei, M. Azam, M. Amayri, and N. Bouguila, "Ica and iva bounded multivariate generalized gaussian mixture based hidden markov models," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106345, 2023.
- [4] A. Algumaei, M. Azam, and N. Bouguila, "Adaptive constrained ivamggmm: Application to mental disorders detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024.
- [5] A. Algumaei, M. Azam, and N. Bouguila, "Novel approach for ecg separation using adaptive constrained ivabmgmm," *Digital Signal Processing*, vol. 149, p. 104476, 2024.
- [6] A. Algumaei, M. Azam, and N. Bouguila, "Mental disorder detection using non-orthogonal adaptive constrained ivabmgmm," in *2024 IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP)*, vol. 1, pp. 65–70, IEEE, 2024.
- [7] A. Algumaei, A. M. Azam, Muhammad, and N. Bouguila, "Adaptive constrained icabmgmm: application to ecg blind source separation," in *The 31st International Conference on Neural Information Processing (ICONIP 2024)*, Springer, 2025.

- [8] T. Elguebaly and N. Bouguila, "Bayesian learning of finite generalized Gaussian mixture models on images," *Signal Processing*, vol. 91, no. 4, pp. 801–820, 2011.
- [9] A. Sefidpour and N. Bouguila, "Spatial color image segmentation based on finite non-Gaussian mixture models," *Expert Systems with Applications*, vol. 39, no. 10, pp. 8993–9001, 2012.
- [10] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annual review of statistics and its application*, vol. 6, pp. 355–378, 2019.
- [11] M. S. Allili, N. Bouguila, and D. Ziou, "Finite general Gaussian mixture modeling and application to image and video foreground segmentation," *Journal of Electronic Imaging*, vol. 17, no. 1, p. 013005, 2008.
- [12] W. Mauersberger, "Experimental results on the performance of mismatched quantizers," *IEEE Transactions on Information Theory*, vol. 25, no. 4, pp. 381–386, 1979.
- [13] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and kullback-leibler distance," *IEEE transactions on image processing*, vol. 11, no. 2, pp. 146–158, 2002.
- [14] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," in *Fundamental Papers in Wavelet Theory*, pp. 494–513, Princeton University Press, 2009.
- [15] J. Miller and J. Thomas, "Detectors for discrete-time signals in non-Gaussian noise," *IEEE Transactions on Information Theory*, vol. 18, no. 2, pp. 241–250, 1972.
- [16] S. Meignen and H. Meignen, "On the modeling of small sample distributions with generalized Gaussian density in a maximum likelihood framework," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1647–1652, 2006.
- [17] N. Farvardin and J. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless sources," *IEEE Transactions on Information Theory*, vol. 30, no. 3, pp. 485–497, 1984.

- [18] F. Najar, S. Bourouis, N. Bouguila, and S. Belghith, “Unsupervised learning of finite full covariance multivariate generalized Gaussian mixture models for human activity recognition,” *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 18669–18691, 2019.
- [19] F. Najar, S. Bourouis, A. Zaguia, N. Bouguila, and S. Belghith, “Unsupervised human action categorization using a riemannian averaged fixed-point learning of multivariate ggmm,” in *International Conference Image Analysis and Recognition*, pp. 408–415, Springer, 2018.
- [20] J. Lindblom and J. Samuelsson, “Bounded support Gaussian mixture modeling of speech spectra,” *IEEE transactions on speech and audio processing*, vol. 11, no. 1, pp. 88–99, 2003.
- [21] T. M. Nguyen, Q. J. Wu, and H. Zhang, “Bounded generalized Gaussian mixture model,” *Pattern Recognition*, vol. 47, no. 9, pp. 3132–3142, 2014.
- [22] M. Azam and N. Bouguila, “Speaker verification using adapted bounded Gaussian mixture model,” in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, pp. 300–307, IEEE, 2018.
- [23] P. Hedelin and J. Skoglund, “Vector quantization based on Gaussian mixture models,” *IEEE transactions on speech and audio processing*, vol. 8, no. 4, pp. 385–401, 2000.
- [24] G. J. McLachlan and T. Krishnan, *The EM algorithm and extensions*, vol. 382. John Wiley Sons, 2007.
- [25] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.
- [26] N. Bouguila and D. Ziou, “High-dimensional unsupervised selection and estimation of a finite generalized dirichlet mixture model based on minimum message length,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 10, pp. 1716–1731, 2007.
- [27] H. Akaike, “A new look at the statistical model identification,” *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

- [28] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
- [29] J. J. Oliver, R. A. Baxter, and C. S. Wallace, "Unsupervised learning using mml," in *ICML*, pp. 364–372, Citeseer, 1996.
- [30] C. S. Wallace and D. L. Dowe, "Minimum message length and kolmogorov complexity," *The Computer Journal*, vol. 42, no. 4, pp. 270–283, 1999.
- [31] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, 1998.
- [32] S. Bhinge, Y. Levin-Schwartz, G.-S. Fu, B. Pesquet-Popescu, and T. Adali, "A data-driven solution for abandoned object detection: Advantages of multiple types of diversity," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1347–1351, IEEE, 2015.
- [33] S. Makeig, A. J. Bell, T.-P. Jung, T. J. Sejnowski, *et al.*, "Independent component analysis of electroencephalographic data," *Advances in neural information processing systems*, pp. 145–151, 1996.
- [34] T.-W. Lee, M. S. Lewicki, and T. J. Sejnowski, "Unsupervised classification with non-Gaussian mixture models using ica," *Advances in neural information processing systems*, pp. 508–514, 1999.
- [35] G. Safont, A. Salazar, L. Vergara, and A. Rodríguez, "New applications of sequential ica mixtures models compared with dynamic bayesian networks for eeg signal processing," in *2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks*, pp. 397–402, IEEE, 2013.
- [36] T.-W. Lee and M. S. Lewicki, "Unsupervised image classification, segmentation, and enhancement using ica mixture models," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 270–279, 2002.

- [37] M. Azam and N. Bouguila, “Unsupervised keyword spotting using bounded generalized Gaussian mixture model with ica,” in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1150–1154, IEEE, 2015.
- [38] M. Azam and N. Bouguila, “Bounded generalized Gaussian mixture model with ica,” *Neural Processing Letters*, vol. 49, no. 3, pp. 1299–1320, 2019.
- [39] T.-W. Lee, *Independent component analysis: theory and applications*. PhD thesis, 1997.
- [40] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE transactions on audio, speech, and language processing*, vol. 15, no. 1, pp. 70–79, 2006.
- [41] J. Vía, M. Anderson, X.-L. Li, and T. Adali, “A maximum likelihood approach for independent vector analysis of Gaussian data sets,” in *2011 IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1–6, IEEE, 2011.
- [42] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adali, “Independent vector analysis, the kotz distribution, and performance bounds,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3243–3247, IEEE, 2013.
- [43] A. Dasgupta and A. E. Raftery, “Detecting features in spatial point processes with clutter via model-based clustering,” *Journal of the American statistical Association*, vol. 93, no. 441, pp. 294–302, 1998.
- [44] P. Kasarapu and L. Allison, “Minimum message length estimation of mixtures of multivariate Gaussian and von mises-fisher distributions,” *Machine Learning*, vol. 100, no. 2, pp. 333–378, 2015.
- [45] T. Elguebaly and N. Bouguila, “Background subtraction using finite mixtures of asymmetric gaussian distributions and shadow detection,” *Machine vision and applications*, vol. 25, no. 5, pp. 1145–1162, 2014.
- [46] A. Salazar, *On Statistical Pattern Recognition in Independent Component Analysis Mixture Modelling*, vol. 4. Springer Science & Business Media, 2012.

- [47] T.-W. Lee, M. S. Lewicki, and T. J. Sejnowski, "Ica mixture models for unsupervised classification of non-Gaussian classes and automatic context switching in blind signal separation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1078–1089, 2000.
- [48] T.-W. Lee and M. S. Lewicki, "The generalized Gaussian mixture model using ica," in *International Workshop on Independent Component Analysis (ICA'00)*, pp. 239–244, 2000.
- [49] M. Anderson, T. Adali, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2011.
- [50] M. Anderson, X.-L. Li, and T. Adali, "Nonorthogonal independent vector analysis using multivariate Gaussian model," in *International Conference on Latent Variable Analysis and Signal Separation*, pp. 354–361, Springer, 2010.
- [51] L. Cuadros-Rodríguez, E. Pérez-Castaño, and C. Ruiz-Samblás, "Quality performance metrics in multivariate classification methods for qualitative analysis," *TrAC Trends in Analytical Chemistry*, vol. 80, pp. 612–624, 2016.
- [52] C. Ferri, J. Hernández-Orallo, and R. Modroi, "An experimental comparison of performance measures for classification," *Pattern recognition letters*, vol. 30, no. 1, pp. 27–38, 2009.
- [53] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.
- [54] M. Selim, "Perioperative stroke," *New England journal of medicine*, vol. 356, no. 7, pp. 706–713, 2007.
- [55] P. Bentley, J. Ganesalingam, A. L. C. Jones, K. Mahady, S. Epton, P. Rinne, P. Sharma, O. Halse, A. Mehta, and D. Rueckert, "Prediction of stroke thrombolysis outcome using ct brain machine learning," *NeuroImage: Clinical*, vol. 4, pp. 635–640, 2014.
- [56] R. O. Duda, P. E. Hart, and D. G. Stork, "Unsupervised learning and clustering," *Pattern classification*, pp. 517–601, 2001.

- [57] A. C. Geller, S. M. Swetter, K. Brooks, M.-F. Demierre, and A. L. Yaroch, "Screening, early detection, and trends for melanoma: current status (2000-2006) and future directions," *Journal of the American Academy of Dermatology*, vol. 57, no. 4, pp. 555–572, 2007.
- [58] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pp. 197–206, 2007.
- [59] C. Patterson, "The state of the art of dementia research: New frontiers," *World Alzheimer Report*, vol. 2018, 2018.
- [60] N. Theera-Umpon, I. Poonkasem, S. Auephanwiriyaikul, and D. Patikulsila, "Hard exudate detection in retinal fundus images using supervised learning," *Neural Computing and Applications*, vol. 32, no. 17, pp. 13079–13096, 2020.
- [61] J. van der Togt and T. H. Rasmussen, "Toward evidence-based hr," *Journal of Organizational Effectiveness: People and Performance*, 2017.
- [62] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, pp. 461–464, 1978.
- [63] H. Bozdogan, "Model selection and akaike's information criterion (aic): The general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.
- [64] D. Peel and G. J. McLachlan, "Robust mixture modelling using the t distribution," *Statistics and computing*, vol. 10, no. 4, pp. 339–348, 2000.
- [65] J. Rissanen, *Stochastic complexity in statistical inquiry*, vol. 15. World scientific, 1998.
- [66] P. Varady, L. Wildt, Z. Benyó, and A. Hein, "An advanced method in fetal phonocardiography," *Computer Methods and programs in Biomedicine*, vol. 71, no. 3, pp. 283–296, 2003.
- [67] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

- [68] H. Almstrom, G. Ekman, O. Axelsson, U. Ulmsten, S. Cnattingius, K. Maesel, and K. AArstrom, "Comparison of umbilical-artery velocimetry and cardiotocography for surveillance of small-for-gestational-age fetuses," *The Lancet*, vol. 340, no. 8825, pp. 936–940, 1992.
- [69] R. Gribonval, C. Févotte, and E. Vincent, "Bss eval toolbox user guide," *IRISA Technical Report*, 2005.
- [70] D. Bhugra, "The global prevalence of schizophrenia," *PLoS medicine*, vol. 2, no. 5, p. e151, 2005.
- [71] J. Deslauriers, W. Racine, P. Sarret, and S. Grignon, "Preventive effect of  $\alpha$ -lipoic acid on pre-pulse inhibition deficits in a juvenile two-hit model of schizophrenia," *Neuroscience*, vol. 272, pp. 261–270, 2014.
- [72] C.-G. Yan, X.-D. Wang, X.-N. Zuo, and Y.-F. Zang, "Dpabi: data processing & analysis for (resting-state) brain imaging," *Neuroinformatics*, vol. 14, pp. 339–351, 2016.
- [73] K. J. Friston, S. Williams, R. Howard, R. S. Frackowiak, and R. Turner, "Movement-related effects in fmri time-series," *Magnetic resonance in medicine*, vol. 35, no. 3, pp. 346–355, 1996.
- [74] J. Ashburner, "A fast diffeomorphic image registration algorithm," *Neuroimage*, vol. 38, no. 1, pp. 95–113, 2007.
- [75] J. A. Spencer, "Clinical overview of cardiotocography," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 100, pp. 4–7, 1993.
- [76] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [77] T. Adali, M. Anderson, and G.-S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, 2014.

- [78] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.
- [79] M. Azam and N. Bouguila, “Speaker classification via supervised hierarchical clustering using ica mixture model,” in *Image and Signal Processing: 7th International Conference, ICISP 2016, Trois-Rivières, QC, Canada, May 30-June 1, 2016, Proceedings 7*, pp. 193–202, Springer, 2016.
- [80] M. Azam and N. Bouguila, “Blind source separation as pre-processing to unsupervised keyword spotting via an ica mixture model,” in *2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 833–836, IEEE, 2018.
- [81] T.-W. Lee, “Independent component analysis,” in *Independent component analysis*, pp. 27–66, Springer, 1998.
- [82] D. A. Bridwell, S. Rachakonda, R. F. Silva, G. D. Pearlson, and V. D. Calhoun, “Spatiospectral decomposition of multi-subject eeg: evaluating blind source separation algorithms on real and realistic simulated data,” *Brain topography*, vol. 31, no. 1, pp. 47–61, 2018.
- [83] H. Lv, Z. Wang, E. Tong, L. M. Williams, G. Zaharchuk, M. Zeineh, A. N. Goldstein-Piekarski, T. M. Ball, C. Liao, and M. Wintermark, “Resting-state functional mri: everything that nonexperts have always wanted to know,” *American Journal of Neuroradiology*, vol. 39, no. 8, pp. 1390–1399, 2018.
- [84] K. Itakura, Y. Bando, E. Nakamura, K. Itoyama, K. Yoshii, and T. Kawahara, “Bayesian multichannel audio source separation based on integrated source and spatial models,” *IEEE/ACM transactions on audio, speech, and language processing*, vol. 26, no. 4, pp. 831–846, 2018.
- [85] T. Kim, I. Lee, and T.-W. Lee, “Independent vector analysis: definition and algorithms,” in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, pp. 1393–1396, IEEE, 2006.
- [86] L. Rabiner and B. Juang, “An introduction to hidden markov models,” *iee assp magazine*, vol. 3, no. 1, pp. 4–16, 1986.

- [87] M. Bicego, U. Castellani, and V. Murino, "A hidden markov model approach for appearance-based 3d object recognition," *Pattern Recognition Letters*, vol. 26, no. 16, pp. 2588–2599, 2005.
- [88] M. Cholewa and P. Głomb, "Estimation of the number of states for gesture recognition with hidden markov models based on the number of critical points in time sequence," *Pattern Recognition Letters*, vol. 34, no. 5, pp. 574–579, 2013.
- [89] E. L. Andrade, S. Blunsden, and R. B. Fisher, "Hidden markov models for optical flow analysis in crowds," in *18th international conference on pattern recognition (ICPR'06)*, vol. 1, pp. 460–463, IEEE, 2006.
- [90] F. Jiang, Y. Wu, and A. K. Katsaggelos, "Abnormal event detection from surveillance video by dynamic hierarchical clustering," in *2007 IEEE international conference on image processing*, vol. 5, pp. V–145, IEEE, 2007.
- [91] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [92] N. Bouguila, W. Fan, and M. Amayri, *Hidden Markov Models and Applications*. Springer, 2022.
- [93] O.-W. Kwon and T.-W. Lee, "Phoneme recognition using ica-based feature extraction and transformation," *Signal Processing*, vol. 84, no. 6, pp. 1005–1019, 2004.
- [94] H. Zhao, K. Zhao, H. Liu, and F. Yu, "Improved mfcc feature extraction combining symmetric ica algorithm for robust speech recognition.," *Journal of multimedia*, vol. 7, no. 1, 2012.
- [95] H. Lee, "Simultaneous blind separation and recognition of speech mixtures using two microphones to control a robot cleaner," *International Journal of Advanced Robotic Systems*, vol. 10, no. 2, p. 103, 2013.

- [96] T. Yamada, A. Tawari, and M. M. Trivedi, "In-vehicle speaker recognition using independent vector analysis," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pp. 1753–1758, IEEE, 2012.
- [97] M. Bressan, D. Guillet, and J. Vitria, "Using an ica representation of high dimensional data for object recognition and classification," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–I, IEEE, 2001.
- [98] S. Bhinge, Z. Boukouvalas, Y. Levin-Schwartz, and T. Adali, "Iva for abandoned object detection: Exploiting dependence across color channels," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2494–2498, IEEE, 2016.
- [99] I. Winkler, S. Haufe, and M. Tangermann, "Automatic classification of artifactual ica-components for artifact removal in eeg signals," *Behavioral and brain functions*, vol. 7, no. 1, pp. 1–15, 2011.
- [100] P. P. Acharjee, R. Phlypo, L. Wu, V. D. Calhoun, and T. Adali, "Gradient artifact removal in concurrently acquired eeg data using independent vector analysis," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5859–5863, IEEE, 2014.
- [101] S. Bhinge, R. Mowakeaa, V. D. Calhoun, and T. Adali, "Extraction of time-varying spatiotemporal networks using parameter-tuned constrained iva," *IEEE transactions on medical imaging*, vol. 38, no. 7, pp. 1715–1725, 2019.
- [102] X. Zhu and D. R. Hunter, "Clustering via finite nonparametric ica mixture models," *Advances in Data Analysis and Classification*, vol. 13, no. 1, pp. 65–87, 2019.
- [103] Q. Yang, H. Hu, and W. Gui, "Fatigue state recognition based on improved ica-hmm," *International Journal of Modelling, Identification and Control*, vol. 22, no. 4, pp. 334–343, 2014.

- [104] Z. Li, J. Han, J. Sun, Y. He, and F. Chu, "Fault recognition method based on independent component analysis and hidden markov model," *Journal of Vibration and Control*, vol. 13, no. 2, pp. 125–137, 2007.
- [105] M. Vrigkas, C. Nikou, and I. A. Kakadiaris, "A review of human activity recognition methods," *Frontiers in Robotics and AI*, vol. 2, p. 28, 2015.
- [106] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local svm approach," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3, pp. 32–36, IEEE, 2004.
- [107] J.-P. Hosom, "Speaker-independent phoneme alignment using transition-dependent states," *Speech Communication*, vol. 51, no. 4, pp. 352–368, 2009.
- [108] P. Karsmakers, K. Pelckmans, J. A. Suykens, and H. V. hamme, "Fixed-size kernel logistic regression for phoneme classification.," in *INTERSPEECH*, pp. 78–81, 2007.
- [109] S. M. Siniscalchi, P. Schwarz, and C.-H. Lee, "High-accuracy phone recognition by combining high-performance lattice generation and knowledge based rescoring," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 4, pp. IV–869, IEEE, 2007.
- [110] A.-r. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE transactions on audio, speech, and language processing*, vol. 20, no. 1, pp. 14–22, 2011.
- [111] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, 2012.
- [112] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE transactions on Consumer Electronics*, vol. 57, no. 1, pp. 76–84, 2011.

- [113] H. Gonçalves, A. Ocneanu, M. Bergés, and R. Fan, “Unsupervised disaggregation of appliances using aggregated consumption data,” in *The 1st KDD workshop on data mining applications in sustainability (SustKDD)*, 2011.
- [114] O. Parson, S. Ghosh, M. Weal, and A. Rogers, “An unsupervised training method for non-intrusive appliance load monitoring,” *Artificial Intelligence*, vol. 217, pp. 1–19, 2014.
- [115] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, “Non-intrusive appliance load monitoring using low-resolution smart meter data,” in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 535–540, IEEE, 2014.
- [116] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, “A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring,” *Applied energy*, vol. 185, pp. 331–344, 2017.
- [117] J. Z. Kolter and M. J. Johnson, “Redd: A public data set for energy disaggregation research,” in *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA*, vol. 25, pp. 59–62, 2011.
- [118] K. He, L. Stankovic, J. Liao, and V. Stankovic, “Non-intrusive load disaggregation using graph signal processing,” *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1739–1747, 2016.
- [119] B. Zhao, L. Stankovic, and V. Stankovic, “On a training-less solution for non-intrusive appliance load monitoring using graph signal processing,” *IEEE Access*, vol. 4, pp. 1784–1799, 2016.
- [120] H. Altrabalsi, V. Stankovic, J. Liao, and L. Stankovic, “Low-complexity energy disaggregation using appliance load modelling,” *Aims Energy*, vol. 4, no. 1, pp. 884–905, 2016.
- [121] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, “An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition,” *Bell System Technical Journal*, vol. 62, no. 4, pp. 1035–1074, 1983.

- [122] L. E. Baum and J. A. Eagon, “An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology,” *Bulletin of the American Mathematical Society*, vol. 73, no. 3, pp. 360–363, 1967.
- [123] L. E. Baum and G. Sell, “Growth transformations for functions on manifolds,” *Pacific Journal of Mathematics*, vol. 27, no. 2, pp. 211–227, 1968.
- [124] G. D. Forney, “The viterbi algorithm,” *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [125] A. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.
- [126] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [127] S. B. Wang, A. Quattoni, L.-P. Morency, D. Demirdjian, and T. Darrell, “Hidden conditional random fields for gesture recognition,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, vol. 2, pp. 1521–1527, IEEE, 2006.
- [128] S.-i. Amari, “Backpropagation and stochastic gradient descent method,” *Neurocomputing*, vol. 5, no. 4-5, pp. 185–196, 1993.
- [129] T. Fawcett, “An introduction to roc analysis,” *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [130] M. J. Black and A. D. Jepson, “Estimating optical flow in segmented images using variable-order parametric models with local deformations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 972–986, 1996.
- [131] J. Morris and E. Fosler-Lussier, “Conditional random fields for integrating local discriminative classifiers,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 617–628, 2008.

- [132] P. Matejka, *Phonotactic and acoustic language recognition*. PhD thesis, PhD thesis, 2009.
- [133] P. Schwarz, “Phoneme recognition based on long temporal context. ph. d. thesis. brno university of technology, 2008,” 2008.
- [134] H. Fujihara and M. Goto, “Three techniques for improving automatic synchronization between music and lyrics: Fricative detection, filler model, and novel feature vectors for vocal activity detection,” in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 69–72, IEEE, 2008.
- [135] M. Gruhne, C. Dittmar, and K. Schmidt, “Phoneme recognition in popular music.,” in *ISMIR*, pp. 369–370, 2007.
- [136] S. Furui, “50 years of progress in speech and speaker recognition research,” *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. 1, no. 2, pp. 64–74, 2005.
- [137] V. Zue, S. Seneff, and J. Glass, “Speech database development at mit: Timit and beyond,” *Speech communication*, vol. 9, no. 4, pp. 351–356, 1990.
- [138] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, “Epileptic seizure detection in eegs using time–frequency analysis,” *IEEE transactions on information technology in biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.
- [139] L. Muda, M. Begam, and I. Elamvazuthi, “Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques,” *arXiv preprint arXiv:1003.4083*, 2010.
- [140] A. Hyvärinen and E. Oja, “Independent component analysis: algorithms and applications,” *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
- [141] S. Makino, T.-W. Lee, and H. Sawada, *Blind speech separation*, vol. 615. Springer, 2007.
- [142] C. J. James and C. W. Hesse, “Independent component analysis for biomedical signals,” *Physiological measurement*, vol. 26, no. 1, p. R15, 2004.

- [143] C. A. Shah, M. K. Arora, and P. K. Varshney, "Ica mixture model for unsupervised classification of non-gaussian classes in multi/hyperspectral imagery," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, vol. 5093, pp. 569–580, SPIE, 2003.
- [144] W. Lu and J. C. Rajapakse, "ICA with reference," *Neurocomputing*, vol. 69, no. 16-18, pp. 2244–2257, 2006.
- [145] Z.-L. Zhang, "Morphologically constrained ICA for extracting weak temporally correlated signals," *Neurocomputing*, vol. 71, no. 7-9, pp. 1669–1679, 2008.
- [146] Z.-L. Sun and L. Shang, "An improved constrained ICA with reference based unmixing matrix initialization," *Neurocomputing*, vol. 73, no. 4-6, pp. 1013–1017, 2010.
- [147] M. De Vos, L. De Lathauwer, and S. Van Huffel, "Spatially constrained ICA algorithm with an application in EEG processing," *Signal Processing*, vol. 91, no. 8, pp. 1963–1972, 2011.
- [148] Z. Wang, "Fixed-point algorithms for constrained ICA and their applications in fmri data analysis," *Magnetic resonance imaging*, vol. 29, no. 9, pp. 1288–1303, 2011.
- [149] W. Lu and J. C. Rajapakse, "Approach and applications of constrained ica," *IEEE transactions on neural networks*, vol. 16, no. 1, pp. 203–212, 2005.
- [150] W. Lu and J. Rajapakse, "Constrained independent component analysis," *Advances in neural information processing systems*, vol. 13, 2000.
- [151] P. A. Rodriguez, M. Anderson, X.-L. Li, and T. Adalı, "General non-orthogonal constrained ica," *IEEE Transactions on Signal Processing*, vol. 62, no. 11, pp. 2778–2786, 2014.
- [152] Y. Du and Y. Fan, "Group information guided ica for fmri data analysis," *Neuroimage*, vol. 69, pp. 157–197, 2013.
- [153] H. Yang, T. Vu, Q. Long, V. Calhoun, and T. Adalı, "Identification of homogeneous subgroups from resting-state fmri data," *Sensors*, vol. 23, no. 6, p. 3264, 2023.

- [154] S. Jun, M. Kim, M. Oh, and H.-M. Park, "Robust speech recognition based on independent vector analysis using harmonic frequency dependency," *Neural Computing and Applications*, vol. 22, pp. 1321–1327, 2013.
- [155] M. Pakravan and M. B. Shamsollahi, "Joint, partially-joint, and individual independent component analysis in multi-subject fmri data," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 7, pp. 1969–1981, 2019.
- [156] T. Kim, "Real-time independent vector analysis for convolutive blind source separation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1431–1438, 2010.
- [157] Z. Boukouvalas, G.-S. Fu, and T. Adalı, "An efficient multivariate generalized Gaussian distribution estimator: Application to iva," in *2015 49th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–4, IEEE, 2015.
- [158] Z. Boukouvalas, S. Said, L. Bombrun, Y. Berthoumieu, and T. Adalı, "A new riemannian averaged fixed-point algorithm for mggd parameter estimation," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2314–2318, 2015.
- [159] S. Bhinge, Q. Long, Y. Levin-Schwartz, Z. Boukouvalas, V. D. Calhoun, and T. Adalı, "Non-orthogonal constrained independent vector analysis: Application to data fusion," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2666–2670, IEEE, 2017.
- [160] Y. Shi, W. Zeng, N. Wang, and L. Zhao, "A new constrained spatiotemporal ica method based on multi-objective optimization for fmri data analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 9, pp. 1690–1699, 2018.
- [161] P. Cunningham, M. Cord, and S. J. Delany, "Supervised learning," *Machine learning techniques for multimedia: case studies on organization and retrieval*, pp. 21–49, 2008.
- [162] Z. Wang, M. Xia, Z. Jin, L. Yao, and Z. Long, "Temporally and spatially constrained ica of fmri data analysis," *PLoS One*, vol. 9, no. 4, p. e94211, 2014.

- [163] N. Wang, W. Zeng, and D. Chen, “A novel sparse dictionary learning separation (sdls) model with adaptive dictionary mutual incoherence constraint for fmri data analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2376–2389, 2016.
- [164] Z. Long, Z. Wang, J. Zhang, X. Zhao, and L. Yao, “Temporally constrained ica with threshold and its application to fmri data,” *BMC Medical Imaging*, vol. 19, no. 1, pp. 1–14, 2019.
- [165] Q.-H. Lin, J. Liu, Y.-R. Zheng, H. Liang, and V. D. Calhoun, “Semiblind spatial ica of fmri using spatial constraints,” *Human brain mapping*, vol. 31, no. 7, pp. 1076–1088, 2010.
- [166] T. Rasheed, Y.-K. Lee, and T.-S. Kim, “Constrained spatiotemporal ica and its application for fmri data analysis,” in *13th International Conference on Biomedical Engineering: ICBME 2008 3–6 December 2008 Singapore*, pp. 555–558, Springer, 2009.
- [167] J. Laney, K. Westlake, S. Ma, E. Woytowicz, and T. Adali, “Capturing subject variability in data driven fmri analysis: A graph theoretical comparison,” in *2014 48th Annual conference on information sciences and systems (CISS)*, pp. 1–6, IEEE, 2014.
- [168] A. M. Michael, M. Anderson, R. L. Miller, T. Adali, and V. D. Calhoun, “Preserving subject variability in group fmri analysis: performance evaluation of gica vs. iva,” *Frontiers in systems neuroscience*, vol. 8, p. 106, 2014.
- [169] S. Bhinge, Q. Long, V. D. Calhoun, and T. Adali, “Spatial dynamic functional connectivity analysis identifies distinctive biomarkers in schizophrenia,” *Frontiers in neuroscience*, vol. 13, p. 1006, 2019.
- [170] C.-Y. Zhang, W.-X. Li, L.-D. Kuang, and Q.-H. Lin, “Incorporating spatial sparsity constraint into complex iva of multi-subject complex-valued fmri data,” in *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, IEEE, 2023.
- [171] Q. Long, S. Bhinge, V. D. Calhoun, and T. Adali, “Independent vector analysis for common subspace analysis: Application to multi-subject fmri data yields meaningful subgroups of schizophrenia,” *NeuroImage*, vol. 216, p. 116872, 2020.

- [172] A. H. Khan, M. Taseska, and E. A. Habets, "A geometrically constrained independent vector analysis algorithm for online source extraction," in *Latent Variable Analysis and Signal Separation: 12th International Conference, LVA/ICA 2015, Liberec, Czech Republic, August 25-28, 2015, Proceedings 12*, pp. 396–403, Springer, 2015.
- [173] D. M. Holtzman, J. C. Morris, and A. M. Goate, "Alzheimer's disease: the challenge of the second century," *Science translational medicine*, vol. 3, no. 77, pp. 77sr1–77sr1, 2011.
- [174] K. Blennow, M. J. de Leon, and H. Zetterberg, "Alzheimer's disease," *The Lancet*, vol. 368, no. 9533, pp. 387–403, 2006.
- [175] J. Ye, T. Wu, J. Li, and K. Chen, "Machine learning approaches for the neuroimaging study of alzheimer's disease," *Computer*, vol. 44, no. 4, pp. 99–101, 2011.
- [176] J. Han, T. Kim, S. Lee, J. Park, J. Lee, Y. Huh, J. Park, and K. Kim, "327 diagnostic stability of mild cognitive impairment subtype," *Asian Journal of Psychiatry*, no. 4, pp. S65–S66, 2011.
- [177] H. K. Yap, N. Kamaldin, J. H. Lim, F. A. Nasrallah, J. C. H. Goh, and C.-H. Yeow, "A magnetic resonance compatible soft wearable robotic glove for hand rehabilitation and brain imaging," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 25, no. 6, pp. 782–793, 2016.
- [178] S. Goh, Z. Dong, Y. Zhang, S. DiMauro, and B. S. Peterson, "Brain imaging evidence that mitochondrial dysfunction is a neurobiological subtype of autism spectrum disorder," *JAMA psychiatry*, vol. 71, no. 6, p. 665, 2014.
- [179] Y. Zhang, S. Wang, G. Ji, and Z. Dong, "Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 10, no. 1, pp. 116–117, 2015.
- [180] K. J. Friston, "Statistical parametric mapping," *Neuroscience databases: a practical guide*, pp. 237–250, 2003.

- [181] J. A. Maldjian, P. J. Laurienti, R. A. Kraft, and J. H. Burdette, “An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fmri data sets,” *Neuroimage*, vol. 19, no. 3, pp. 1233–1239, 2003.
- [182] X. Guo, Z. Wang, K. Li, Z. Li, Z. Qi, Z. Jin, L. Yao, and K. Chen, “Voxel-based assessment of gray and white matter volumes in alzheimer’s disease,” *Neuroscience letters*, vol. 468, no. 2, pp. 146–150, 2010.
- [183] K. Wang, M. Liang, L. Wang, L. Tian, X. Zhang, K. Li, and T. Jiang, “Altered functional connectivity in early alzheimer’s disease: A resting-state fmri study,” *Human brain mapping*, vol. 28, no. 10, pp. 967–978, 2007.
- [184] S. H. Hojjati, A. Ebrahimzadeh, A. Khazaei, A. Babajani-Feremi, A. D. N. Initiative, *et al.*, “Predicting conversion from mci to ad using resting-state fmri, graph theoretical approach and svm,” *Journal of neuroscience methods*, vol. 282, pp. 69–80, 2017.
- [185] F. Agosta, M. Pievani, C. Geroldi, M. Copetti, G. B. Frisoni, and M. Filippi, “Resting state fmri in alzheimer’s disease: beyond the default mode network,” *Neurobiology of aging*, vol. 33, no. 8, pp. 1564–1578, 2012.
- [186] G. Hickok and D. Poeppel, “The cortical organization of speech processing,” *Nature reviews neuroscience*, vol. 8, no. 5, pp. 393–402, 2007.
- [187] A. Khazaei, A. Ebrahimzadeh, and A. Babajani-Feremi, “Identifying patients with alzheimer’s disease using resting-state fmri and graph theory,” *Clinical Neurophysiology*, vol. 126, no. 11, pp. 2132–2141, 2015.
- [188] T. R. Insel, “Rethinking schizophrenia,” *Nature*, vol. 468, no. 7321, pp. 187–193, 2010.
- [189] D. Shi, H. Zhang, G. Wang, X. Yao, Y. Li, S. Wang, and K. Ren, “Neuroimaging biomarkers for detecting schizophrenia: A resting-state functional mri-based radiomics analysis,” *Heliyon*, vol. 8, no. 12, p. e12276, 2022.

- [190] A. H. Algumaei, R. F. Algunaïd, M. A. Rushdi, and I. A. Yassine, “Feature and decision-level fusion for schizophrenia detection based on resting-state fmri data,” *Plos one*, vol. 17, no. 5, p. e0265300, 2022.
- [191] R. F. Algunaïd, A. H. Algumaei, M. A. Rushdi, and I. A. Yassine, “Schizophrenic patient identification using graph-theoretic features of resting-state fmri data,” *Biomedical Signal Processing and Control*, vol. 43, pp. 289–299, 2018.
- [192] O. R. Phillips, K. H. Nuechterlein, K. A. Clark, L. S. Hamilton, R. F. Asarnow, N. S. Hageman, A. W. Toga, and K. L. Narr, “Fiber tractography reveals disruption of temporal lobe white matter tracts in schizophrenia,” *Schizophrenia research*, vol. 107, no. 1, pp. 30–38, 2009.
- [193] S. K. Goldsmith, R. M. Shapiro, and J. N. Joyce, “Disrupted pattern of d2 dopamine receptors in the temporal lobe in schizophrenia: a postmortem study,” *Archives of General Psychiatry*, vol. 54, no. 7, pp. 649–658, 1997.
- [194] X. Li, Q. Liu, Z. Chen, Y. Li, Y. Yang, X. Wang, X. Guo, B. Luo, Y. Zhang, H. Shi, *et al.*, “Abnormalities of regional brain activity in patients with schizophrenia: A longitudinal resting-state fmri study,” *Schizophrenia Bulletin*, p. sbad054, 2023.
- [195] M. Kong, T. Chen, S. Gao, S. Ni, Y. Ming, X. Chai, C. Ling, and X. Xu, “Abnormal network homogeneity of default-mode network and its relationships with clinical symptoms in antipsychotic-naïve first-diagnosis schizophrenia,” *Frontiers in Neuroscience*, vol. 16, 2022.
- [196] D. Shi, Y. Li, H. Zhang, X. Yao, S. Wang, G. Wang, and K. Ren, “Machine learning of schizophrenia detection with structural and functional neuroimaging,” *Disease markers*, vol. 2021, pp. 1–12, 2021.
- [197] G. Polanczyk, M. S. De Lima, B. L. Horta, J. Biederman, and L. A. Rohde, “The worldwide prevalence of adhd: a systematic review and metaregression analysis,” *American journal of psychiatry*, vol. 164, no. 6, pp. 942–948, 2007.

- [198] E. Hoekzema, S. Carmona, J. A. Ramos-Quiroga, V. Richarte Fernandez, R. Bosch, J. C. Soliva, M. Rovira, A. Bulbena, A. Tobena, M. Casas, *et al.*, “An independent components and functional connectivity analysis of resting state fmri data points to neural network dysregulation in adult adhd,” *Human brain mapping*, vol. 35, no. 4, pp. 1261–1272, 2014.
- [199] D. Dai, J. Wang, J. Hua, and H. He, “Classification of adhd children through multimodal magnetic resonance imaging,” *Frontiers in systems neuroscience*, vol. 6, p. 63, 2012.
- [200] C. Tang, Y. Wei, J. Zhao, and J. Nie, “Different developmental pattern of brain activities in adhd: a study of resting-state fmri,” *Developmental neuroscience*, vol. 40, no. 3, pp. 246–257, 2018.
- [201] J. H. Yoo, D. Kim, J. Choi, and B. Jeong, “Treatment effect of methylphenidate on intrinsic functional brain network in medication-naïve adhd children: a multivariate analysis,” *Brain imaging and behavior*, vol. 12, no. 2, pp. 518–531, 2018.
- [202] P. Sörös, E. Hoxhaj, P. Borel, C. Sadohara, B. Feige, S. Matthies, H. H. Müller, K. Bachmann, M. Schulze, and A. Philipsen, “Hyperactivity/restlessness is associated with increased functional connectivity in adults with adhd: a dimensional analysis of resting state fmri,” *BMC psychiatry*, vol. 19, pp. 1–11, 2019.
- [203] U. of Freiburg, *Freiburg-seizure-prediction-project*. University of Freiburg, 2007.
- [204] Q.-H. Zou, C.-Z. Zhu, Y. Yang, X.-N. Zuo, X.-Y. Long, Q.-J. Cao, Y.-F. Wang, and Y.-F. Zang, “An improved approach to detection of amplitude of low-frequency fluctuation (alf) for resting-state fmri: fractional alf,” *Journal of neuroscience methods*, vol. 172, no. 1, pp. 137–141, 2008.
- [205] Y. Zang, T. Jiang, Y. Lu, Y. He, and L. Tian, “Regional homogeneity approach to fmri data analysis,” *Neuroimage*, vol. 22, no. 1, pp. 394–400, 2004.
- [206] A. Algumaei, M. Azam, M. Amayri, and N. Bouguila, “Adaptive constrained icamgmm: An improvement over ica,” in *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 1–7, IEEE, 2023.

- [207] C. Wang, Y. Ren, R. Zhang, C. Wang, X. Ran, J. Shen, Z. Zhao, W. Tao, Y. Yang, W. Ren, *et al.*, “Schizophrenia classification and abnormalities reveal of brain region functional connection by deep-learning multiple sparsely connected network,” *Biomedical Signal Processing and Control*, vol. 96, p. 106580, 2024.
- [208] L.-H. Lee, C.-H. Chen, W.-C. Chang, P.-L. Lee, K.-K. Shyu, M.-H. Chen, J.-W. Hsu, Y.-M. Bai, T.-P. Su, and P.-C. Tu, “Evaluating the performance of machine learning models for automatic diagnosis of patients with schizophrenia based on a single site dataset of 440 participants,” *European Psychiatry*, vol. 65, no. 1, p. e1, 2022.
- [209] S. Ramkiran, A. Sharma, and N. P. Rao, “Resting-state anticorrelated networks in schizophrenia,” *Psychiatry Research: Neuroimaging*, vol. 284, pp. 1–8, 2019.
- [210] L. Liu, L.-B. Cui, X.-S. Wu, N.-B. Fei, Z.-L. Xu, D. Wu, Y.-B. Xi, P. Huang, K. M. von Deneen, S. Qi, *et al.*, “Cortical abnormalities and identification for first-episode schizophrenia via high-resolution magnetic resonance imaging,” *Biomarkers in Neuropsychiatry*, vol. 3, p. 100022, 2020.
- [211] Z. Zhang and C. Sun, “A numerical study on multi-site damage identification: A data-driven method via constrained independent component analysis,” *Structural Control and Health Monitoring*, vol. 27, no. 10, p. e2583, 2020.
- [212] N. Lassance, V. DeMiguel, and F. Vrins, “Optimal portfolio diversification via independent component analysis,” *Operations Research*, vol. 70, no. 1, pp. 55–72, 2022.
- [213] S. Wein, A. M. Tomé, M. Goldhacker, M. W. Greenlee, and E. W. Lang, “A constrained ica-emd model for group level fmri analysis,” *Frontiers in Neuroscience*, vol. 14, p. 221, 2020.
- [214] M. Azam and N. Bouguila, “Multivariate bounded support asymmetric generalized gaussian mixture model with model selection using minimum message length,” *Expert Systems with Applications*, vol. 204, p. 117516, 2022.

- [215] R. Song, G. Wang, J. Cheng, A. Liu, C. Li, and X. Chen, “Constrained independent vector extraction of quasi-periodic signals from multiple data sets,” *Signal Processing*, vol. 189, p. 108296, 2021.
- [216] K. Goto, T. Ueda, L. Li, T. Yamada, and S. Makino, “Geometrically constrained independent vector analysis with auxiliary function approach and iterative source steering,” in *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 757–761, IEEE, 2022.
- [217] S. Bhinge, *Adaptive Constrained Independent Vector Analysis: Application to Large-scale fMRI Analysis*. PhD thesis, University of Maryland, Baltimore County, 2020.
- [218] T. Nakatani, R. Ikeshita, K. Kinoshita, H. Sawada, N. Kamo, and S. Araki, “Switching independent vector analysis and its extension to blind and spatially guided convolutional beamforming algorithms,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1032–1047, 2022.
- [219] L. Li and K. Koishida, “Geometrically constrained independent vector analysis for directional speech enhancement,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 846–850, IEEE, 2020.
- [220] J. Pan and W. J. Tompkins, “A real-time qrs detection algorithm,” *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.
- [221] “Database for fetal heartbeat separation.,” *Available online: <https://homes.esat.kuleuven.be/~smc/daisy/daisydata.html>*.
- [222] G. B. Moody, R. G. Mark, and A. L. Goldberger, “Physionet: a web-based resource for the study of physiologic signals,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 70–75, 2001.
- [223] N. G. C. M. M. S. Bentley, P., “The pascal classifying heart sounds challenge 2011,” *Available online: [www.peterjbentley.com/heartchallenge/index.html](http://www.peterjbentley.com/heartchallenge/index.html)*, 2011.

- [224] O. R. A. Almanifi, A. F. Ab Nasir, M. A. M. Razman, R. M. Musa, and A. P. A. Majeed, "Heartbeat murmurs detection in phonocardiogram recordings via transfer learning," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 10995–11002, 2022.
- [225] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B.-W. On, "Heartbeat sound signal classification using deep learning," *Sensors*, vol. 19, no. 21, p. 4819, 2019.
- [226] A. Chao, S. Ng, and L. Wang, "Listen to your heart: Feature extraction and classification methods for heart sounds," 2018.
- [227] M. Schouteden, K. Van Deun, T. F. Wilderjans, and I. Van Mechelen, "Performing disco-sca to search for distinctive and common information in linked data," *Behavior research methods*, vol. 46, pp. 576–587, 2014.
- [228] A. K. Smilde, I. Måge, T. Naes, T. Hankemeier, M. A. Lips, H. A. Kiers, E. Acar, and R. Bro, "Common and distinct components in data fusion," *Journal of Chemometrics*, vol. 31, no. 7, p. e2900, 2017.
- [229] V. D. Calhoun, T. Adali, G. D. Pearlson, and K. A. Kiehl, "Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data," *Neuroimage*, vol. 30, no. 2, pp. 544–553, 2006.
- [230] M. Jacob, J. Ford, B. Roach, V. Calhoun, and D. Mathalon, "145. semantic priming abnormalities in schizophrenia: An erp-fmri fusion study," *Schizophrenia bulletin*, vol. 43, no. Suppl 1, p. S76, 2017.
- [231] J. Mangalathu-Arumana, E. Liebenthal, and S. A. Beardsley, "Optimizing within-subject experimental designs for jica of multi-channel erp and fmri," *Frontiers in neuroscience*, vol. 12, p. 13, 2018.
- [232] T. Adali, Y. Levin-Schwartz, and V. D. Calhoun, "Multimodal data fusion using source separation: Two effective models based on ica and iva and their properties," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1478–1493, 2015.

- [233] V. D. Calhoun, T. Adali, K. A. Kiehl, R. Astur, J. J. Pekar, and G. D. Pearlson, "A method for multitask fmri data fusion applied to schizophrenia," *Human brain mapping*, vol. 27, no. 7, pp. 598–610, 2006.
- [234] J. Sui, H. He, G. D. Pearlson, T. Adali, K. A. Kiehl, Q. Yu, V. P. Clark, E. Castro, T. White, B. A. Mueller, *et al.*, "Three-way (n-way) fusion of brain imaging data based on mcca+ jica and its application to discriminating schizophrenia," *NeuroImage*, vol. 66, pp. 119–132, 2013.
- [235] J. Sui, H. He, Q. Yu, J. Chen, J. Rogers, G. D. Pearlson, A. Mayer, J. Bustillo, J. Canive, and V. D. Calhoun, "Combination of resting state fmri, dti, and smri data to discriminate schizophrenia by n-way mcca+ jica," *Frontiers in human neuroscience*, vol. 7, p. 235, 2013.
- [236] D. Hirjak, M. Rashidi, K. M. Kubera, G. Northoff, S. Fritze, M. M. Schmitgen, F. Sambataro, V. D. Calhoun, and R. C. Wolf, "Multimodal magnetic resonance imaging data fusion reveals distinct patterns of abnormal brain structure and function in catatonia," *Schizophrenia bulletin*, vol. 46, no. 1, pp. 202–210, 2020.
- [237] J. Sui, E. Castro, H. He, D. Bridwell, Y. Du, G. D. Pearlson, T. Jiang, and V. D. Calhoun, "Combination of fmri-smri-eeeg data improves discrimination of schizophrenia patients by ensemble feature selection," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3889–3892, IEEE, 2014.
- [238] S. Liu, H. Wang, M. Song, L. Lv, Y. Cui, Y. Liu, L. Fan, N. Zuo, K. Xu, Y. Du, *et al.*, "Linked 4-way multimodal brain differences in schizophrenia in a large chinese han population," *Schizophrenia bulletin*, vol. 45, no. 2, pp. 436–449, 2019.
- [239] Y. Xiao, Z. Yan, Y. Zhao, B. Tao, H. Sun, F. Li, L. Yao, W. Zhang, S. Chandan, J. Liu, *et al.*, "Support vector machine-based classification of first episode drug-naïve schizophrenia patients and healthy controls using structural mri," *Schizophrenia research*, vol. 214, pp. 11–17, 2019.
- [240] C. Zhao, J. Zhu, X. Liu, C. Pu, Y. Lai, L. Chen, X. Yu, and N. Hong, "Structural and functional brain abnormalities in schizophrenia: a cross-sectional study at different stages of the

- disease,” *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 83, pp. 27–32, 2018.
- [241] S. Liu, C. Jie, W. Zheng, J. Cui, and Z. Wang, “Investigation of underlying association between whole brain regions and alzheimer’s disease: A research based on an artificial intelligence model,” *Frontiers in aging neuroscience*, vol. 14, p. 872530, 2022.
- [242] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, and T.-F. Yuan, “Detection of subjects and brain regions related to alzheimer’s disease using 3d mri scans based on eigenbrain and machine learning,” *Frontiers in computational neuroscience*, vol. 9, p. 66, 2015.
- [243] H. Chen, W. Li, X. Sheng, Q. Ye, H. Zhao, Y. Xu, F. Bai, and A. D. N. Initiative, “Machine learning based on the multimodal connectome can predict the preclinical stage of alzheimer’s disease: a preliminary study,” *European Radiology*, vol. 32, pp. 448–459, 2022.
- [244] I. Eliasova, L. Anderkova, R. Marecek, and I. Rektorova, “Non-invasive brain stimulation of the right inferior frontal gyrus may improve attention in early alzheimer’s disease: a pilot study,” *Journal of the neurological sciences*, vol. 346, no. 1-2, pp. 318–322, 2014.
- [245] J. Pan, Q. Zuo, B. Wang, C. P. Chen, B. Lei, and S. Wang, “Decgan: Decoupling generative adversarial network for detecting abnormal neural circuits in alzheimer’s disease,” *IEEE Transactions on Artificial Intelligence*, 2024.
- [246] H. Lin, J. Jiang, Z. Li, C. Sheng, W. Du, X. Li, and Y. Han, “Identification of subjective cognitive decline due to alzheimer’s disease using multimodal mri combining with machine learning,” *Cerebral Cortex*, vol. 33, no. 3, pp. 557–566, 2023.
- [247] M. De Oliveira, M. Balthazar, A. D’abreu, C. Yasuda, B. Damasceno, F. Cendes, and G. Castellano, “Mr imaging texture analysis of the corpus callosum and thalamus in amnesic mild cognitive impairment and mild alzheimer disease,” *American Journal of Neuroradiology*, vol. 32, no. 1, pp. 60–66, 2011.
- [248] D. B. Dwyer, C. Cabral, L. Kambeitz-Ilankovic, R. Sanfelici, J. Kambeitz, V. Calhoun, P. Falkai, C. Pantelis, E. Meisenzahl, and N. Koutsouleris, “Brain subtyping enhances the

- neuroanatomical discrimination of schizophrenia,” *Schizophrenia bulletin*, vol. 44, no. 5, pp. 1060–1069, 2018.
- [249] O. Veatch, J. Veenstra-VanderWeele, M. Potter, M. Pericak-Vance, and J. Haines, “Genetically meaningful phenotypic subgroups in autism spectrum disorders,” *Genes, Brain and Behavior*, vol. 13, no. 3, pp. 276–285, 2014.
- [250] A. Iqbal, A.-K. Seghouane, and T. Adali, “Shared and subject-specific dictionary learning (shssdl) algorithm for multisubject fmri data analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 11, pp. 2519–2528, 2018.
- [251] X. Hu, H. Huang, B. Peng, J. Han, N. Liu, J. Lv, L. Guo, C. Guo, and T. Liu, “Latent source mining in fmri via restricted boltzmann machine,” *Human brain mapping*, vol. 39, no. 6, pp. 2368–2380, 2018.
- [252] A. P. James and B. V. Dasarathy, “Medical image fusion: A survey of the state of the art,” *Information fusion*, vol. 19, pp. 4–19, 2014.
- [253] V. D. Calhoun and T. Adali, “Feature-based fusion of medical imaging data,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 711–720, 2008.
- [254] V. D. Calhoun and J. Sui, “Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness,” *Biological psychiatry: cognitive neuroscience and neuroimaging*, vol. 1, no. 3, pp. 230–244, 2016.
- [255] K. Dontaraju, S.-J. Kim, M. Akhonda, and T. Adali, “Capturing common and individual components in fmri data by discriminative dictionary learning,” in *2018 52nd asilomar conference on signals, systems, and computers*, pp. 1351–1356, IEEE, 2018.
- [256] Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, “Quantifying the interaction and contribution of multiple datasets in fusion: Application to the detection of schizophrenia,” *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1385–1395, 2017.
- [257] A. Klami, S. Virtanen, and S. Kaski, “Bayesian canonical correlation analysis,” *Journal of Machine Learning Research*, vol. 14, no. 4, 2013.

- [258] Y. Guo and L. Tang, “A hierarchical model for probabilistic independent component analysis of multi-subject fmri studies,” *Biometrics*, vol. 69, no. 4, pp. 970–981, 2013.
- [259] A. Puli and A. Kushki, “Toward automatic anxiety detection in autism: A real-time algorithm for detecting physiological arousal in the presence of motion,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 646–657, 2019.
- [260] J. Liu, Z. Wang, H. Qin, Y. Wang, J. Deng, H. Li, Q. Xu, X. Xu, and H. Liu, “Social recognition of joint attention cycles in children with autism spectrum disorders,” *IEEE Transactions on Biomedical Engineering*, 2023.
- [261] F. Rafiee, R. Rezvani Habibabadi, M. Motaghi, D. M. Yousem, and I. J. Yousem, “Brain mri in autism spectrum disorder: narrative review and recent advances,” *Journal of Magnetic Resonance Imaging*, vol. 55, no. 6, pp. 1613–1624, 2022.
- [262] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, *et al.*, “The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism,” *Molecular psychiatry*, vol. 19, no. 6, pp. 659–667, 2014.
- [263] W. Cao, H. Zhu, Y. Li, Y. Wang, W. Bai, U. Lao, Y. Zhang, Y. Ji, S. He, and X. Zou, “The development of brain network in males with autism spectrum disorders from childhood to adolescence: evidence from fnirs study,” *Brain sciences*, vol. 11, no. 1, p. 120, 2021.
- [264] B. Qin, L. Wang, J. Cai, T. Li, and Y. Zhang, “Functional brain networks in preschool children with autism spectrum disorders,” *Frontiers in Psychiatry*, vol. 13, p. 896388, 2022.
- [265] E. T. Rolls, Y. Zhou, W. Cheng, M. Gilson, G. Deco, and J. Feng, “Effective connectivity in autism,” *Autism Research*, vol. 13, no. 1, pp. 32–44, 2020.
- [266] N. Chaitra, P. Vijaya, and G. Deshpande, “Diagnostic prediction of autism spectrum disorder using complex network measures in a machine learning framework,” *Biomedical Signal Processing and Control*, vol. 62, p. 102099, 2020.