

**Deep Learning for Quantitative Ultrasound and
Multimodal Analysis: Liver Steatosis Diagnosis,
Uncertainty Decomposition, and Diagnosis of Breast
Cancer-Related Lymphedema**

Dorsa Ameri

**A Thesis
in
The Department
of
Electrical and Computer Engineering**

**Presented in Partial Fulfillment of the Requirements
for the Degree of
Master of Applied Science (Electrical and Computer Engineering) at
Concordia University
Montréal, Québec, Canada**

August 2025

© Dorsa Ameri, 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Dorsa Ameri**

Entitled: **Deep Learning for Quantitative Ultrasound and Multimodal Analysis:
Liver Steatosis Diagnosis, Uncertainty Decomposition, and Diagnosis of
Breast Cancer-Related Lymphedema**

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

Master of Applied Science (Electrical and Computer 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:

Dr. Arash Mohammadi Chair

Dr. Arash Mohammadi Examiner

Dr. Paula Lago Examiner

Dr. Hassan Rivaz Supervisor

Approved by _____
Wahab Hamou-Lhadj, Chair
Department of Electrical and Computer Engineering

_____ 2025

Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Deep Learning for Quantitative Ultrasound and Multimodal Analysis: Liver Steatosis Diagnosis, Uncertainty Decomposition, and Diagnosis of Breast Cancer-Related Lymphedema

Dorsa Ameri

Point-of-Care Ultrasound (POCUS) is a portable, cost-effective imaging modality with strong potential to expand access to diagnostic tools in remote and underserved settings. However, its interpretation still depends heavily on expert knowledge, which limits broader clinical adoption. This thesis aims to enhance the interpretability, reliability, and accessibility of POCUS by leveraging deep learning techniques.

The first part of this work presents a Bayesian deep learning framework for the classification of Non-alcoholic fatty liver disease using QUS features extracted from pre-clinical duck experiments. The model not only achieves accurate classification but also provides meaningful uncertainty estimates, helping assess prediction confidence. In the second part, we propose a method to decompose predictive uncertainty into epistemic and aleatoric components in the estimation of Homodyned-K distribution QUS parameters and investigate their relationship with prediction error. The final part introduces a multimodal dataset for the diagnosis of breast cancer-related lymphedema (BCRL) using POCUS. A deep learning pipeline is developed that integrates ultrasound images and clinical features to improve diagnostic performance. Together, these contributions apply deep learning methods to enhance quantitative tissue characterization, uncertainty estimation, and diagnosis, making ultrasound more practical and accessible in everyday healthcare.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Hassan Rivaz, for his exceptional guidance and continuous support throughout the course of my research. I am also sincerely thankful to my mentor, Dr. Ali Kafei Zad Tehrani, for his invaluable guidance and dedicated mentorship. I deeply appreciate their innovative ideas, effective leadership, and thorough reviews, which provided me with an excellent opportunity to learn and thrive in my research. I would like to extend my gratitude to my colleagues at the IMPACT Lab, whose support and collaboration have been incredibly helpful throughout this journey. I have learned a great deal from their valuable insights and experience.

Special thanks to my dear family, partner, and friends, without whose generous love and continuous support, none of this would have been possible. I am deeply grateful for all the love and support that I have received from them.

I am also thankful to my co-authors, Dr. Anna Towers (McGill University Health Centre) and Dr. Ivan M. Rosado-Mendez (University of Wisconsin–Madison), for sharing their valuable expertise and contributions to this thesis through their insightful feedback and ideas.

I would like to thank Dr. Sarkis Meterissian for his support in facilitating access to breast cancer survivors through the McGill University Health Centre. This work would not have been possible without his approval.

I am also grateful to Georgina Cama Lee (Research Associate, Lymphedema Clinic, McGill University Health Centre) for her invaluable efforts in participant recruitment and coordination of data collection sessions.

We acknowledge the support of the Government of Canada's New Frontiers in Research Fund

(NFRF) [NFRFE-2022-00295] and the Natural Sciences and Engineering Research Council of Canada (NSERC).

Contents

List of Figures	viii
List of Tables	x
1 Introduction	1
1.1 Ultrasound Imaging	1
1.1.1 Deep learning in ultrasound imaging	3
1.2 Quantitative Ultrasound	4
1.2.1 Uncertainty Quantification	4
1.3 Breast Cancer-Related Lymphedema	5
1.3.1 Ultrasound and deep learning for BCRL analysis	5
1.4 Research Objective and Contribution	5
1.5 Thesis Outline	7
2 Bayesian Approach with Uncertainty Quantification for Quantitative Ultrasound-Based Diagnosis of Liver Steatosis in a Pre-Clinical Animal Model	8
2.1 Introduction	9
2.2 Materials and Methods	12
2.2.1 Dataset and Data Acquisition	12
2.2.2 Models	13
2.2.3 Training and Hyper-parameter Tuning	15
2.3 Results and Discussion	16

2.4	Conclusion	19
3	Uncertainty Decomposition and Error Margin Detection of Homodyned-K Distribution in Quantitative Ultrasound	20
3.1	Introduction	21
3.2	Materials and Methods	22
3.2.1	Homodyned K-distribution	22
3.2.2	Datasets and data generation	23
3.2.3	Uncertainty Decomposition	24
3.3	Results	25
3.4	Discussion and Conclusion	26
4	Point-of-Care Ultrasound in Diagnosis of Breast Cancer-Related Lymphedema: A Multimodal Database and Deep Learning Approach	29
4.1	Introduction	30
4.2	Related work	33
4.3	Materials and Methods	35
4.3.1	Data collection	35
4.3.2	Dataset preparation and pre-processing	40
4.3.3	Proposed method	40
4.4	Results	44
4.5	Discussions	47
4.6	Conclusion	48
5	Conclusion and Future Work	49
5.1	Conclusion	49
5.2	Future Work	50
	Bibliography	52

List of Figures

Figure 1.1	Verasonics machine and the transducer. This is an open platform that provides access to sequence design and raw data.	2
Figure 1.2	Clarius probe. We used this pocket-sized ultrasound device for the collection of patient data in Chapter 4.	3
Figure 2.1	(A) to (E) Boxplots of different QUS features for the three states. t_0 is before force-feeding, t_2 and t_3 are the states after 7 and 14 days of force-feeding, respectively. The boxes span from the first quartile (Q1) to the third quartile (Q3). (F) Distribution of data points between two classes. The classes 0 and 1 represent t_0 and the aggregation of t_2 and t_3 , respectively.	13
Figure 2.2	The ROC curves for MLP, SVM, Logistic Regression and BNN classifiers. All test samples from the 4 folds are utilized to generate the ROCs.	17
Figure 2.3	(A) Entropy versus probability of prediction of all test samples. (B) Histograms of entropies for BNN’s correct and false predictions.	18
Figure 3.1	The visualization of computing the uncertainties.	25
Figure 3.2	Error versus epistemic and aleatoric uncertainty for simulation data (a) Epistemic uncertainty for $\log_{10}(\alpha)$. (b) Aleatoric uncertainty for $\log_{10}(\alpha)$. (c) Epistemic uncertainty for k . (d) Aleatoric uncertainty for k	27
Figure 4.1	Data collection with POCUS. (a) shows the landmarks where images were acquired. (b) demonstrates how the probe was placed on the arm. For landmarks 2 and 4, we collected additional images with the probe parallel to the muscle fibers (90 degrees rotated compared to the probe direction in the image).	38

Figure 4.2	B-mode images of the arm of a subject diagnosed with mild BCRL (severity level 2) in the right arm. The left and right columns correspond to unaffected and affected arms, respectively. L1 to L5 are landmarks. Images in columns L2 (A) and L4 (A) were collected with the probe placed parallel to the muscle fibers.	39
Figure 4.3	The proposed multimodal architecture that integrates ultrasound images of both arms with tabular data. The POCUS B-mode images are fed to encoders with shared weights, and the absolute difference of the image embeddings is concatenated with tabular embeddings. The final vector was passed through a classifier. The illustration of encoder architecture is reproduced from [1].	42
Figure 4.4	Two-dimensional t-SNE visualization of the concatenated image and tabular features extracted from all validation samples. Each point represents a sample, and colors indicate the ground truth class labels.	46

List of Tables

Table 2.1	Mean $\pm$ standard deviation of test results for different classifiers	17
Table 3.1	Simulation data Pearson’s correlation between error and uncertainties of predicted parameters (SNR = 20, 30, 40)	26
Table 3.2	Aleatoric and epistemic uncertainties for predicted $\log_{10}(\alpha)$ of the investigated experimental phantoms.	27
Table 4.1	Clinical Information of Study Participants	36
Table 4.2	Lymphedema severity scores of the study subjects	37
Table 4.3	Training hyperparameters for experiments	43
Table 4.4	Mean $\pm$ standard deviation of validation results across the three folds, with the best performance highlighted in bold.	46

Chapter 1

Introduction

In this chapter, we provide a brief overview of the fundamentals of the three papers that form this thesis. A more comprehensive and in-depth introduction is provided in each chapter.

1.1 Ultrasound Imaging

Ultrasound waves are sound waves with frequencies higher than 20 kHz, which is higher than the upper limit of human hearing [2]. These waves are mechanical waves that propagate through the material without transverse motion by compressing and expanding longitudinally [3]. Ultrasound imaging is widely used in clinical practice due to its non-invasive nature, affordability, portability, and real-time imaging capabilities [4, 5]. Higher frequencies (7–30 MHz) are used for superficial structures like musculoskeletal and vascular imaging, which offer higher spatial resolution but less penetration depth, while lower frequencies (1–5 MHz) are used for deeper structures like abdominal imaging because of their greater penetration [6]. During an ultrasound examination, sound waves are transmitted into the body, and echoes reflected from tissue structures are received by the transducer and converted into electrical signals. These raw signals, known as radio frequency (RF) data, contain detailed information about the amplitude and phase of the waves that return over time. In order to produce an image that can be clinically interpreted, the envelope of the RF signal is extracted, which represents the signal amplitude, and then converted to grayscale intensity values. This process results in the formation of a B-mode (brightness mode) image, where brighter regions



(a)



(b)

Figure 1.1: Verasonics machine and the transducer. This is an open platform that provides access to sequence design and raw data.

represent stronger echo amplitudes corresponding to acoustic impedance mismatches, which indicate tissue boundaries and structural variations [7]. Despite the advantages of ultrasound, it has a few drawbacks. Due to lower spatial resolution compared to other modalities and the presence of speckle noise, its interpretation can be challenging.

Figure 1.1 shows an image of a Verasonics Vantage 256 ultrasound machine (Verasonics, Kirkland, WA, USA) at the Image Processing and Characterization of Tissue (IMPACT) lab, along with the ultrasound transducer. This state-of-the-art ultrasound machine allows parallel receive and transmit access to 256 transducer elements at the same time. This enables very fast imaging of around 5,000 frames per second for applications such as cardiac imaging and ultrasound elastography [8–10].

Figure 1.2 shows a Clarius Point-of-Care Ultrasound (POCUS) device (Clarius Health, Vancouver, BC, Canada), which we used extensively in Chapter 4 for patient data collection. The device



Figure 1.2: Clarius probe. We used this pocket-sized ultrasound device for the collection of patient data in Chapter 4.

is pocket-sized and battery-operated, and connects wirelessly to a portable device such as a smartphone to display the ultrasound images. Due to its low compute and power ratings, the frame rate is around 25 fps. Nevertheless, we show in Chapter 4 that this probe can be used to conveniently study tissue composition to diagnose lymphedema.

1.1.1 Deep learning in ultrasound imaging

In recent years, deep learning has become a powerful tool in medical imaging and enabled more accurate, efficient, and automated image analysis. While earlier methods relied on predefined rules, deep learning models learn complex patterns directly from the data. This makes them a great tool for handling the variability and noise that are frequently observed in ultrasound images. Deep learning has substantially improved the capabilities of ultrasound imaging, from low-level image reconstruction [11–14] and denoising [15–19] to high-level tasks like classification [20–24], segmentation [25–32], and elastography [8, 33–36], transforming both image quality and clinical decision-making.

1.2 Quantitative Ultrasound

Conventional ultrasound is qualitative and needs visual interpretation of grayscale images, which is often subjective and requires significant expertise to distinguish between subtle variations in tissue appearance. Furthermore, it only displays brightness and discards much of the raw information, such as phase and frequency content, in the original RF signals. Quantitative Ultrasound (QUS) techniques have been developed to address this limitation by extracting quantitative properties of tissue from the RF signals acquired during ultrasound imaging [22, 37–40]. These properties include acoustic parameters such as the attenuation coefficient, backscatter coefficient, and parameters derived from statistical models of the backscattered signals, which can be used as quantitative biomarkers of tissue characteristics [41]. Using this quantitative information, QUS has the potential to simplify and standardize the interpretation of ultrasound images [42]. This enables a more reproducible and operator-independent assessment of tissue properties, which improves the utility of ultrasound in both the diagnosis and monitoring of disease progression [43]. Additionally, unlike B-mode images, which are influenced by system settings and post-processing, QUS features are extracted directly from RF data, which makes them more consistent across different ultrasound machines [44]. Deep learning approaches have been applied to QUS analysis to model complex relationships between quantitative parameters and tissue properties, enhancing tissue characterization and diagnosis.

1.2.1 Uncertainty Quantification

Deep learning has achieved remarkable success in fields such as computer vision and speech recognition [45]. However, in healthcare, where decisions directly impact patient safety, it's important to understand how reliable model predictions are [46–48]. Uncertainty quantification allows clinicians to assess the confidence of a model's output, which helps identify cases where predictions may be unreliable and additional expert review is needed. By providing insights into model confidence, uncertainty estimation makes deep learning models more trustworthy and safer to use in clinical settings [49]

1.3 Breast Cancer-Related Lymphedema

Breast cancer-related lymphedema (BCRL) is a frequent condition following breast cancer treatments, including surgery and radiotherapy, which often damage or remove lymph nodes within the lymphatic system. This damage disrupts lymphatic drainage, resulting in chronic fluid buildup in the affected arm [50–52]. BCRL affects a significant proportion of breast cancer survivors within the first few years post-treatment, with reported symptoms such as swelling, pain, reduced range of motion, and weakness [50]. These symptoms can severely impact patients’ quality of life by causing physical limitations, psychological distress, and an increased risk of infections, ultimately interfering with daily functioning and well-being [53–55].

1.3.1 Ultrasound and deep learning for BCRL analysis

Ultrasound has been demonstrated as an effective tool for detecting tissue changes associated with BCRL [55–59]. Due to the complications of the interpretation of B-mode images, deep learning techniques can be leveraged to enable automated BCRL diagnosis and assessment [25,60]. One of the major challenges that limited ultrasound BCRL analysis is the scarcity of publicly available datasets. The concept is explored in greater detail in Chapter 4.

1.4 Research Objective and Contribution

This thesis presents a set of contributions aimed at improving the interpretability, reliability, and accessibility of ultrasound imaging through advanced deep learning methods. This work addresses these goals from several perspectives, including automated diagnosis, quantitative tissue characterization, and uncertainty estimation. While POCUS can democratize access to ultrasound images by virtue of its low cost and portability, the interpretation of these images still requires expert ultrasound radiologists. Our contributions are a step towards removing these barriers by automating the process so that POCUS can be used more widely in small clinics and remote areas. Specifically, in Chapter 2, we proposed a Bayesian framework to diagnose liver steatosis using quantitative tissue features from pre-clinical duck models. We also quantified predictive uncertainty as a measure of model confidence in classification outcomes. In Chapter 3, we decomposed predictive uncertainty

in QUS parameter estimation into epistemic (uncertainty over the model parameters) and aleatoric (uncertainty inherent in the data) components, using a Bayesian framework to gain insights into the factors contributing to the total uncertainty. By analyzing the relationship of each component with prediction error, we demonstrated that the correlation of both uncertainty components with the error is comparable. Thus, we can use epistemic uncertainty to get a notion of error, as it is more feasible to obtain than aleatoric uncertainty. Finally, in Chapter 4, we proposed a multimodal dataset and a deep learning pipeline for the automated diagnosis of BCRL, combining ultrasound images with clinical tabular data. This work addresses multiple limitations in BCRL research, including the scarcity of datasets, the lack of diagnostic standards, challenges in early detection, and the limited availability of trained specialists. To the best of our knowledge, this is the first automated method for BCRL diagnosis, opening the path to further developments in the field. This thesis has culminated in the following three papers:

- Ameri, D., KZ Tehrani, A., Cloutier, G., Tang, A., Rosado-Mendez, I. M., & Rivaz, H. Bayesian Approach with Uncertainty Quantification for Quantitative Ultrasound-Based Diagnosis of Liver Steatosis in a Pre-Clinical Animal Model. Under review at *SPIE Medical Imaging 2026*.
- Ameri, D., KZ Tehrani, A., Rosado-Mendez, I. M., & Rivaz, H. Uncertainty Decomposition and Error Margin Detection of Homodyned-K Distribution in Quantitative Ultrasound. In *2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS)* (pp. 1-4). IEEE.
- Ameri, D., KZ Tehrani, A., Freiche, B., Meterissian, S. H., Towers, A., & Rivaz, H. Point-of-Care Ultrasound in Diagnosis of Breast Cancer-Related Lymphedema: A Multimodal Database and Deep Learning Approach. Under review at *IEEE Transactions on Biomedical Engineering*.

1.5 Thesis Outline

The thesis is organized as follows. Chapter 2 presents a Bayesian deep learning framework for classifying liver steatosis using QUS features extracted from a pre-clinical animal model, demonstrating the diagnostic potential of QUS in liver tissue characterization. Chapter 3 focuses on the decomposition of uncertainty and error margin detection in the estimation of Homodyned-K distribution parameters in QUS, highlighting methods to quantify and interpret uncertainties in tissue parameter estimation. Chapter 4 proposes a multimodal BCRL database integrating imaging and clinical variables, and a deep learning pipeline for classifying POCUS images in the detection of BCRL, using the multimodal database to enhance diagnostic accuracy. Finally, Chapter 5 summarizes the key findings of this work and outlines potential directions for future developments.

Chapter 2

Bayesian Approach with Uncertainty Quantification for Quantitative Ultrasound-Based Diagnosis of Liver Steatosis in a Pre-Clinical Animal Model

This chapter is based on our paper [\[61\]](#).

Metabolic dysfunction-associated steatotic liver disease (MASLD), previously termed non-alcoholic fatty liver disease (NAFLD), is characterized by fat accumulation in liver cells and can progress to severe liver disorders. Traditional diagnostic methods, such as liver biopsy, are costly, invasive, prone to sampling errors, and carry inherent risks. Ultrasound imaging, a noninvasive, accessible, and cost-effective method, is an ideal candidate for this aim. In this study, we proposed a deep learning approach for MASLD diagnosis. We employed Bayesian Neural Networks (BNNs) to classify two MASLD grades (no steatosis versus grades greater than or equal to 1) in force-fed duck livers using quantitative ultrasound (QUS) features. We utilized ultrasound data from force-fed duck livers and employed QUS biomarkers such as shear wave elastography (speed and dispersion), local attenuation coefficient slope, and speckle statistics (parameters of Homodyned K-distribution). To

our knowledge, this is the first application of BNNs for liver steatosis classification using QUS features. Furthermore, prediction uncertainty was quantified using entropy, highlighting the model's robustness and reliability. These findings indicated that combining QUS features with BNNs offers a promising and non-invasive approach for diagnosing MASLD.

2.1 Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD) [62], a prevalent and ever-growing liver issue, is a condition characterized by the accumulation of fat in the hepatocytes. MASLD can progress through different stages, ranging from simple fatty liver (steatosis) to more severe conditions, including metabolic dysfunction-associated steatohepatitis (MASH) (formerly known as non-alcoholic steatohepatitis (NASH)) [62], fibrosis, cirrhosis, and, in some cases, liver cancer. In the initial stage (simple steatosis), the primary feature is the accumulation of fat in hepatocytes. MASH is a more advanced stage characterized by inflammation and liver cell damage in addition to fat accumulation. Persistent inflammation leads to the formation of scar tissue around the liver and nearby blood vessels, which is referred to as fibrosis. Cirrhosis is the advanced scarring of the liver tissue, which disrupts the normal liver structure and function. It increases the risk of liver failure and liver cancer over time [63].

Different MASLD grades reflect the extent of fat accumulation, with grade 1 indicating mild steatosis, grade 2 moderate, and grade 3 severe. While liver biopsy is the established method for MASLD grading, it has drawbacks such as cost, sampling errors, and associated risks. Ultrasound imaging is a great candidate for assessing steatosis grades due to its noninvasive nature, widespread accessibility, and cost-effectiveness. Quantitative ultrasound (QUS) refers to the use of ultrasound technology to obtain quantitative measurements of tissue properties. Physical properties of the tissue can be extracted based on the interactions between ultrasound waves and small variations of density and compressibility known as scatterers. Valuable information about the underlying tissue microstructures, which are not visible through conventional grayscale (B-mode) ultrasound imaging, can be obtained. QUS parameters such as shear wave elastography (dispersion), compression wave local attenuation coefficient slope, and speckle statistics (parameters of Homodyned K-distribution)

can be used as biomarkers for MASLD grading [64].

Vianna *et al.* conducted a study comparing the performance of deep learning models and radiologists in grading hepatic steatosis [65]. This study, which included abdominal B-mode ultrasound images from 199 patients, showed that deep learning models employing B-mode provide comparable performance with human readers for hepatic steatosis detection and grading. Ma *et al.* investigated the classification of NAFLD utilizing 11 machine learning classifiers [66]. They used clinical features obtained in clinical examinations performed by [67] including body mass index (kg/m^2), total cholesterol (mmol/L) and creatinine (mmol/l) to classify NAFLD. Kim *et al.* proposed a learning-based approach to quantify attenuation coefficient (AC) from B-mode ultrasound images [68]. Their results showed that their proposed technique can effectively grade fatty liver disease. In [69], they also improved the robustness of the model to the changes in sensor geometry. Tang *et al.* showed that combining elastography and QUS features using a machine learning model (Random forests) can increase the AUC of classifying not steatohepatitis vs borderline or steatohepatitis in rats from 0.63 (elastography alone) to 0.72 [70]. Sharon *et al.* modified the network architecture to estimate quantitative physical properties of density and speed of sound [71].

In this study, we propose a Bayesian framework to classify two MASLD grades (no steatosis versus grades greater than or equal to 1) using QUS features. To this end, we employ Bayesian Neural Networks (BNNs). The primary objective of the paper is to develop a reliable method that provides an estimate of uncertainty to guide the clinician on the confidence of the classification. The estimated uncertainty can be a notion of the reliability of the outcome. By applying the classification to patches in the image, the uncertainty could be used to guide the tissue collection during biopsy to regions where there is more confidence in the presence of fat infiltration. Another use could be to assess the confidence in detecting changes in steatosis diagnosis as a response to interventions, such as weight loss regimes. Uncertainty estimation has already been implemented in commercial ultrasound machines for shear wave elastography, but this feature has not yet been developed for QUS. Clinicians can use the QUS uncertainty estimates to guide their decision-making process by putting less weight on uncertain estimates or selecting the best frame in the data collection to reduce the uncertainty. The second objective is the development of QUS methods that are grounded in the physical and acoustic properties of the liver tissue, and therefore, are interpretable. The performance

of the BNN is then compared with three other classifiers: a Multilayer Perceptron (MLP), a Support Vector Machine (SVM), and a Logistic Regression model.

We employed QUS features extracted from in vivo ultrasound data collected in a study conducted by the University of Montreal, in which 27 Mulard ducks were force-fed and scanned over a 28-day period [64]. The QUS biomarkers included shear wave speed frequency dispersion (SWS dispersion), compression wave local attenuation coefficient slope (ACS), and parameters of Homodyned K-distribution (HKD), including mean intensity, reciprocal of the scatterer clustering parameter, and coherent-to-diffuse signal ratio. Shear waves are used in ultrasound elastography to measure tissue stiffness. SWS dispersion represents the frequency dependency of the speed of shear wave propagation through the tissue. ACS measures the rate at which ultrasound compression waves attenuate or lose intensity as they travel through tissue. HKD is a statistical model used to describe echo envelope data. The mean intensity (μ_n) represents the average strength or magnitude of the backscattered ultrasound signals within a specific region of the tissue. The reciprocal of the scatterer clustering parameter ($\frac{1}{\alpha}$) is a marker of either heterogeneity or the effective density of random scatterers. The coherent-to-diffuse signal ratio (k) is an indicator of structure in scatterer spatial organization [72], [73]. These features collectively provide information about the structural, mechanical, and statistical properties of the duck liver tissue as captured by ultrasound imaging.

Our contributions can be summarized as follows:

- We employed in vivo data, including QUS features from force-fed duck livers' ultrasound data, to classify two MASLD grades (no steatosis versus stages greater or equal to 1).
- We proposed a deep learning approach for MASLD diagnosis using BNNs. To our knowledge, this is the first application of BNNs for liver steatosis classification using QUS features.
- We compared the performance of BNNs with other classifiers.
- We quantified the uncertainty of predictions in our BNN model using the entropy metric. We also showed that by excluding samples with high entropy, we can achieve improvements and surpass the evaluated classifiers.

2.2 Materials and Methods

2.2.1 Dataset and Data Acquisition

Our dataset originates from a study conducted by the University of Montreal Hospital Research Center [64]. This study has been approved by the Institutional Animal Care Committee of the University of Montreal Hospital Research Center. The data collection was performed on living unanesthetized ducks at the farm in a process including 14 days of pre-force-feeding (ranging from day -13 to day 0), followed by 14 days of force-feeding (ranging from day 1 to day 14). Ultrasound scanning was performed on ducks with a Verasonics V1 programmable system using an ATL L7-4 linear probe driven at 5 MHz [64]. The study conducted by the University of Montreal Hospital Research Center included 5 QUS features, which were studied separately to monitor steatosis grades. In our study, we took a unique approach by combining all those QUS features to classify liver steatosis and to quantify the uncertainty simultaneously. Fig.2.1(A) to (E) represent the distribution of each feature for different states.

We classified data from time point t_0 , before the start of force-feeding, and t_2 and t_3 (days 7 and 14 after the start of the force-feeding), including a dataset of 120 samples. Our classification reflects the clinical importance of differentiating the baseline before force-feeding (t_0) and after at least one week (t_2, t_3), which corresponds to clinically relevant grades of no steatosis versus grades greater than or equal to 1. We assigned t_0 as the first class and both t_2 and t_3 as the second class, which leads to a binary classification problem. There was an imbalance of the data distribution between the two classes that are displayed in Fig.2.1(F).

To enhance the robustness and reliability of the evaluation for this small dataset, we created different folds of training and test data utilizing a customized 4-fold cross-validation algorithm. Our developed algorithm randomly selects a minimum of 30% of samples (equivalent to 36 data points), including at least 6 different duck numbers (representing 20% of the total ducks) during each iteration. We ensured that all samples of a particular duck number were either in the test or training set to avoid data leakage. This methodology guarantees that each test set consists of diverse samples from multiple ducks while maintaining the necessary separation in the dataset based on specific duck numbers. Each classifier was trained and tested 4 times, each time using one of these

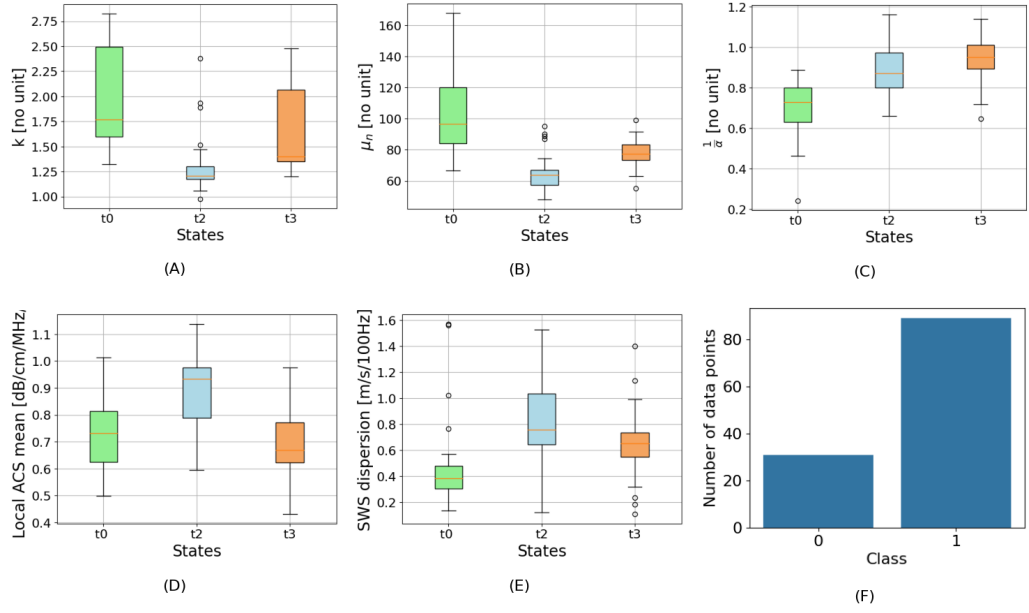


Figure 2.1: (A) to (E) Boxplots of different QUS features for the three states. t_0 is before force-feeding, t_2 and t_3 are the states after 7 and 14 days of force-feeding, respectively. The boxes span from the first quartile (Q1) to the third quartile (Q3). (F) Distribution of data points between two classes. The classes 0 and 1 represent t_0 and the aggregation of t_2 and t_3 , respectively.

sets, and the average and standard deviation of test results were reported.

2.2.2 Models

Bayesian Neural Network (BNN)

The training data, $D = \{X, Y\}$ (X represents data points, Y represents labels), is employed to find the optimum weights (W^*), which can be defined as [74]:

$$W^* = \underset{W}{\operatorname{argmax}} P(D|W) \quad (1)$$

The optimum weights are learned during the training by backpropagation. In BNNs, the weights are sampled from a distribution (usually Gaussian), and the distribution parameters are learned in the training phase. The inference can be formulated as:

$$P(Y|\theta) = \int_W P(Y|W)P(W|\theta)dW \quad (2)$$

where θ is the parameter of the distribution which is learned during the training. Computing the integration over all possible values of W is intractable; therefore, $P(W|\theta)$ can be sampled to obtain the prediction [74].

$$\tilde{Y} = E \left\{ \tilde{Y}_i \right\} \quad (3)$$

where $E(.)$ denotes the expected value, which can be replaced by the sample estimate. $\tilde{Y}_i$ is the sampled prediction value which can be written as:

$$\begin{aligned} \tilde{Y}_i &\sim P(Y|W_i), \\ W_i &\sim P(W|D), \end{aligned} \quad (4)$$

where $\sim$ denotes sampling from the distribution. The prediction of Y can be simply explained as running the trained network with the same input multiple times. Each forward pass of the network has a unique set of weights (W_i), which results in distinct prediction ($\tilde{Y}_i$). The final prediction can be obtained by averaging the values of $\tilde{Y}_i$.

Uncertainty quantification: The total predictive uncertainty in Bayesian modeling can be decomposed into epistemic and aleatoric components [49], [75]. Epistemic uncertainty refers to the uncertainty over the model's parameters, while aleatoric uncertainty captures the inherent noise in input observations. There are several methods to estimate the predictive uncertainty. Gaussian Process is one of the earliest approaches [76]. More recently, [77] employed BNNs for uncertainty quantification in the segmentation of stroke lesions in MRI images, where uncertainty was decomposed into aleatoric and epistemic components. BNNs are gaining popularity due to their solid mathematical foundation for uncertainty estimation [78], [79]. In this study, we opt for BNNs, as they require only a few layers with a limited number of nodes, making them computationally efficient for implementation and training. Calculating aleatoric uncertainty is challenging as it requires multiple observations of the same input data, which may not always be available. In contrast, epistemic uncertainty can be estimated through multiple model inferences within Bayesian frameworks, which is the approach adopted in this study. Notably, Ameri et al. [80] demonstrated that the correlation between error and epistemic uncertainty is comparable to that between error and aleatoric uncertainty in HKD parameter estimation.

Another important topic is the metric to quantify uncertainty. In [81], the standard deviation of the predicted values was employed as the uncertainty of their regression task. Depeweg *et al.* used entropy [82] which measures the average amount of information associated with a probability distribution. Higher entropy indicates higher uncertainty [82]. Entropy is defined as [83]:

$$H(p) = H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log(p_i) \quad (5)$$

where n is the number of samples (forward passes for each test sample which is set to 50 in our study) and p_i is the probability of each sample (forward pass in our study). We opted for using entropy as the metric to quantify the uncertainty of our BNN classifier.

Other Classifiers

We compared the performance of BNN with three other classifiers: a custom Multilayer Perceptron (MLP) with 4 hidden layers, a Support Vector Machine (SVM) using a "sigmoid" kernel from the scikit-learn library, and a Logistic Regression model with an L2 penalty also from the scikit-learn library. The comparison helped us to assess how well BNN performs the classification task.

2.2.3 Training and Hyper-parameter Tuning

To address the data imbalance issue, we assigned different weights to each class during the training process for all classifiers. To monitor the performance of the models, a validation set was separated from the training set during the training of each model.

Bayesian Neural Network(BNN): We implemented a BNN with 3 Bayesian hidden layers with ReLU activation functions and a dropout layer with a dropout rate of 0.1. Adding one more hidden layer did not affect performance. For model training, we opted for weighted Binary Cross Entropy (BCE) loss. The ADAM optimizer is selected due to its relatively higher convergence speed and adaptability to different types of neural network architectures [84]. During training, the model's performance is evaluated on the validation set, and the optimum values of 5×10^{-4} and 200 were selected for learning rate and number of epochs, respectively. We opted for the model with the best

validation AUC and evaluated that on the test set.

Multilayer Perceptron (MLP): An MLP with 4 hidden layers, with ReLU activation functions, and one dropout layer with a dropout rate of 0.1 is trained using a weighted BCE logits loss function and ADAM optimizer. We set the learning rate value to 0.0005 and the number of epochs to 200.

Support Vector Machine (SVM): C-Support Vector Classification (svc) from sklearn. SVM is fitted to the data. Hyperparameter tuning is carried out for the SVM classifier to find optimal values for the model’s hyperparameters. A grid search is performed to identify the optimal regularization parameter (C), kernel type (kernel), and kernel coefficient (gamma). This process results in selecting values of 10 for C, “sigmoid” for the kernel, and 0.01 for gamma.

Logistic Regression: A logistic regression model is trained with an L2 penalty and a regularization parameter set to C=1.0.

2.3 Results and Discussion

The results obtained by our different classifiers during evaluation are provided in Table 2.1. The mean and standard deviation of the models’ outcomes for 4 data folds are reported. BNN has demonstrated the highest accuracy of 0.956 ± 0.031 and an area under the curve (AUC) of 0.986 ± 0.011 which is comparable to the other classifiers. BNN also outperformed the other classifiers in F1 score and Recall. The receiver operating characteristic (ROC) curves for all classifiers are presented in Fig. 2.2. This figure illustrates the comparative similarity between the AUC values of MLP, Logistic Regression, and BNN classifiers. According to Table 2.1 and Fig. 2.2, in general, all classifiers performed comparably. This can be associated with the availability of a small number of test samples and the high discrimination of QUS features.

To gain further insights into the performance of our BNN classifier, we analyzed the relationship between the model’s prediction and the corresponding entropy values for all test samples (Fig. 2.3)(A). Entropy, in this context, serves as a measure of uncertainty associated with the classifier’s predictions. For correctly classified samples, our analysis showed a notable trend. The scatter plot depicted a major concentration of low entropy values corresponding to the lowest and the highest probabilities of prediction (class 0 and class 1). This observation suggested that the model tends

to have lower entropy when it is more confident about the prediction. Conversely, for misclassified samples, the scatter plot exhibited higher entropy values associated with probabilities around 0.5, which is our threshold for classification. This indicates that providing the uncertainty in the model’s decision-making process can help clinicians assess the confidence of the classification in conjunction with the prediction output.

Table 2.1: Mean $\pm$ standard deviation of test results for different classifiers

Classifier	Accuracy	AUC	F1 Score	precision	Recall
MLP	0.939 ± 0.016	0.980 ± 0.014	0.954 ± 0.013	1.0 ± 0.0	0.913 ± 0.025
SVM	0.948 ± 0.021	0.955 ± 0.026	0.962 ± 0.016	0.987 ± 0.019	0.939 ± 0.017
Logistic Regression	0.940 ± 0.023	0.987 ± 0.025	0.956 ± 0.018	0.987 ± 0.019	0.927 ± 0.027
BNN	0.956 ± 0.031	0.986 ± 0.011	0.967 ± 0.022	0.963 ± 0.037	0.976 ± 0.021

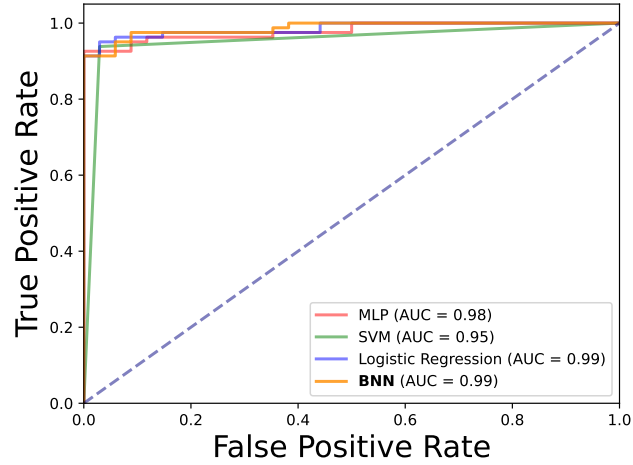


Figure 2.2: The ROC curves for MLP, SVM, Logistic Regression and BNN classifiers. All test samples from the 4 folds are utilized to generate the ROCs.

The histogram of entropy values for the BNN model’s test predictions is provided in Fig. 2.3(B). A clear distinction between the entropy distributions for correct and false predictions can be observed, which highlights the fact that we can consider entropy as an indicator of prediction reliability in our BNN classifier.

In clinical applications, the samples with high uncertainty can be ignored to improve the reliability of the model. To demonstrate that, we excluded samples with entropy higher than 12.5, and

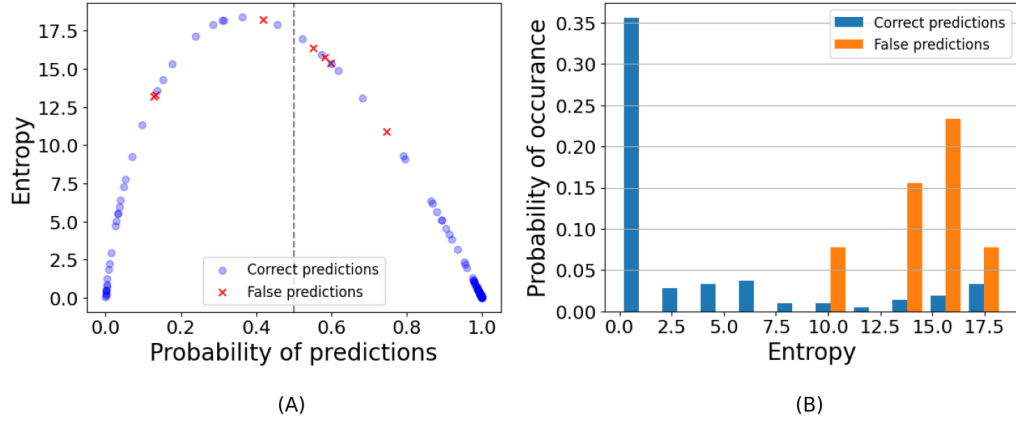


Figure 2.3: (A) Entropy versus probability of prediction of all test samples. (B) Histograms of entropies for BNN's correct and false predictions.

monitored the evaluation metrics. We concluded that this approach enhances model performance during inference. Specifically, we achieved an average test AUC of 0.998, and an average accuracy of 0.969, surpassing all evaluated classifiers. Excluding unreliable samples is feasible in clinical data collection because multiple acquisitions are made during each session. Moreover, when the model expresses high uncertainty for a given sample, alternative strategies can be employed, such as applying test-time augmentation (TTA) to aggregate predictions across multiple augmented views or repeat scans for further analysis.

A limitation of the study is that the results are dependent on the accuracy of the QUS feature extraction. Regarding the translation of the algorithm from animal experiments to humans, it should be noted that the QUS features are physics-based, and as such, have the same signature in humans and ducks because they both follow the same biological mechanism of accumulation of fat in hepatocytes. Nevertheless, their values and distribution might be different. Therefore, if the model is intended to be used for human cases, the current framework can be applied but it should be trained on QUS obtained from human livers. Transfer learning can be used to minimize the requirement for large training data. In addition, the ground truth is extremely hard to obtain in human data (the gold standard for diagnosis is biopsy, which is invasive, or magnetic resonance imaging (MRI), which is costly). Performing experiments on ducks allowed us to bypass this important obstacle.

Employing QUS features rather than B-mode images is indeed the strength of this study, as QUS features provide system-independent information. B-mode images, on the other hand, provide

intensity, which can stem from different tissue properties and imaging settings. B-mode images undergo different post-processing steps, such as beamforming, envelope detection, log compression, and speckle reduction, which highly depend on the system properties. Therefore, by changing the system, the model trained on one system cannot be used for another system. On the other hand, QUS features are physical properties of the tissue, extracted from radio frequency (RF) data. They are independent of the system, which results in achieving a coefficient of variation (CV) as low as 1.5% using different machines [44].

2.4 Conclusion

In this work, we employed BNNs to classify the MASLD state of force-fed ducks using QUS features. We compared the BNN's performance with SVM, Logistic Regression, and MLP. In addition to the modest improvement of the results by BNN, its capability to quantify uncertainty makes it a reasonable choice among other classifiers. We employed the entropy of BNN's predictions as the uncertainty metric to quantify the reliability of the predictions, which is critical in clinical applications.

Chapter 3

Uncertainty Decomposition and Error Margin Detection of Homodyned-K Distribution in Quantitative Ultrasound

This chapter is based on our paper [\[80\]](#).

Homodyned K-distribution (HK-distribution) parameter estimation in quantitative ultrasound (QUS) has been recently addressed using Bayesian Neural Networks (BNNs). BNNs have been shown to significantly reduce computational time in speckle statistics-based QUS without compromising accuracy and precision. Additionally, they provide estimates of feature uncertainty, which can guide the clinician’s trust in the reported feature value. The total predictive uncertainty in Bayesian modeling can be decomposed into epistemic (uncertainty over the model parameters) and aleatoric (uncertainty inherent in the data) components. By decomposing the predictive uncertainty, we can gain insights into the factors contributing to the total uncertainty. In this chapter, we propose a method to compute epistemic and aleatoric uncertainties for HK-distribution parameters (α and k) estimated by a BNN, in both simulation and experimental data. In addition, we investigate the relationship between the prediction error and both uncertainties, shedding light on the interaction between these uncertainties and the error of HK parameters.

3.1 Introduction

Scatterers are microstructures within the tissue that are smaller than the ultrasound wavelength and scatter the ultrasound waves. Quantitative ultrasound (QUS) analyzes the detected backscattered signal to provide insights into the scatterers' structures, which are closely linked to the tissue characteristics. QUS methods have been widely utilized for tissue characterization, including liver fibrosis staging [85–87], breast inclusions classification [88–91], and metastatic lymph nodes detection [92].

QUS methods can be decomposed into two broad categories: spectral-based and envelope-based methods [93]. Spectral-based methods analyze the backscattered signal in the frequency domain to obtain parameters such as backscattering coefficient, attenuation coefficient, and effective scatterer diameter [37, 41]. Envelope-based methods utilize the envelope of the backscattered signal to estimate QUS parameters, such as scatterer number density, and coherency of scatterers [40, 41, 73]. These methods model the envelope of the backscattered RF data by fitting a distribution to the samples [41].

The Homodyned-K distribution can model the envelop data, and its parameters have been found to be correlated with the tissue properties, making them useful in tissue characterization [86, 94]. Envelope statistics, including point-wise signal-to-noise ratio (SNR), skewness, kurtosis, and log-based moments, are commonly used to estimate HK parameters. Estimating HK parameters from known envelope statistics does not have a closed-form solution, and conventional methods of estimating these parameters rely on iterative optimization methods or table search [95–97].

Deep learning methods have recently been employed in QUS [69, 98]. They have also been utilized to estimate the parameters of the HK distribution by employing envelope statistics as input features of the model. Zhou et al. introduced an Artificial Neural Network (ANN) as an estimator of HK parameters [99]. The ANN estimator employed Multilayer Perceptrons (MLPs), which were prone to overfitting. Moreover, ANNs perform as a black box without having a reliability metric, which is crucial to ensure that the model's predictions are trustworthy and can be confidently applied in clinical or research settings. Tehrani et al. addressed these issues by utilizing a Bayesian

Neural Network (BNN) which incorporates probability distributions instead of fixed weights, enabling the estimation of the uncertainty of predictions. Their method improved the estimation of HK-distribution parameters and provided uncertainty quantification which enables the assessment of the reliability of the model's outputs [79, 100].

The total predictive uncertainty in Bayesian modeling can be decomposed into epistemic and aleatoric components [49]. Epistemic uncertainty refers to the uncertainty over the network weights. This uncertainty is often referred to as *model uncertainty*. High epistemic uncertainty indicates that the test input may be an outlier or different compared to the training distribution. As a result, by training the model with more diverse training data, we can reduce this type of uncertainty. Aleatoric uncertainty, on the other hand, captures the noise inherent in the input observations. Addressing this type of uncertainty requires knowledge about unobserved variables, such as additional features in the input, which are often inaccessible. Therefore, it is not always possible to reduce aleatoric uncertainty [49, 75]. By understanding the contributions of the two uncertainty components, we can identify whether increasing the size of the training data or improving the data collection process would be beneficial to reduce the total uncertainty.

In this study, we propose a framework to obtain the total predictive uncertainty of HK-distribution parameters and decompose it into epistemic and aleatoric ones using BNN for simulation and experimental data. We also investigate the relationship between the prediction error and both uncertainties.

3.2 Materials and Methods

3.2.1 Homodyned K-distribution

The Homodyned K-distribution (HK-distribution) is described by the following equation [95]:

$$P_{HK}(A|\epsilon, \sigma^2, \alpha) = A \int_0^\infty u J_0(u\epsilon) J_0(uA) \left(1 + \frac{u^2 \sigma^2}{2}\right)^{-\alpha} du \quad (6)$$

where A represents the envelope of the backscattered echo signal and $J_0(\cdot)$ is the zero-order Bessel function. The coherent signal power is denoted by ϵ^2 , and the diffuse signal power is given

by $2\sigma^2\alpha$ [95]. The scatterer clustering parameter α , and coherent to diffuse scattering ratio $k = \frac{\epsilon}{\sigma\sqrt{\alpha}}$, referred here to as HK parameters, are commonly used in tissue characterization [64, 101]. α and k are correlated with the scatterer number density and the microstructural organization of the scatterers, respectively.

3.2.2 Datasets and data generation

Simulation Data Equation (7) suggested by [102], [99] was employed to produce synthetic samples from HK-distribution:

$$a_i = \sqrt{\left(\sqrt{2k} + X_i\sigma\sqrt{Z_i/\alpha}\right)^2 + \left(Y_i\sigma\sqrt{Z_i/\alpha}\right)^2} \quad (7)$$

where a_i is the generated sample, and X_i and Y_i are the independent and identically distributed samples (i.i.d) from the Normal distribution having zero mean and variance of 1. Z_i is the sample from the Gamma distribution with shape parameter α and scale parameter σ which is set to 1.

We first trained the BNN following the schedule suggested by [79]. The simulation test data were generated for 31 randomly selected $\log_{10}(\alpha)$ values ranging from -0.3 to 1.3 corresponding to α from 0.5 to 20, and 11 values of k ranging from 0 to 1.25 similar to [79] using (7). Moreover, for each value of $\log_{10}(\alpha)$ and k , 10 realizations were generated, leading to 3410 sample sets. Different input observations were used for uncertainty decomposition, which will be elaborated on in Subsection C. In addition, Rayleigh noise was incorporated into the samples at different levels, resulting in data with three different SNRs of 20, 30, and 40 dB, which are computed by $SNR = 10 \log_{10} (E[\text{env}^2]/(2\sigma_N^2))$ where $E[\text{env}^2]$ denotes the power of the envelope samples and σ_N is the scale parameter of Rayleigh noise.

Experimental Data We used data from four experimental phantoms. The first dataset was sourced from a phantom previously described in [103], which is a three-layered phantom with two different scatterer number densities. The phantom was constructed using an emulsion of ultrafiltered milk and water-based gelatin. Glass beads with diameters ranging from 5 to 43 μm (3000E, Potters Industries, Valley Forge, PA, USA) were used as scattering sources. Images of the phantom were

acquired using an 18L6 linear array transducer with a center frequency of 8.9 MHz on a Siemens Acuson S2000 scanner (Siemens Medical Solutions USA, Inc.), previously reported in [103]. The top and bottom layers of the phantom have the same scatterer concentration of 2 g/L, and the middle layer has a higher concentration of 8 g/L. Patches R1 and R2 were selected to extract the statistical features from the top and middle layers, respectively. We obtained 60 patches of envelope data for each of the two specified regions by moving the patches laterally (to avoid any changes in resolution cell size) across several frames.

The other three datasets were acquired from homogeneous phantoms with different numbers of scatterers per resolution cell, previously reported in [22]. The dimensions of the phantoms measured 15 cm $\times$ 5 cm $\times$ 15 cm, and they were made from a homogeneous mixture of agarose gel and glass beads as scattering agents. The diameter range of the glass beads and their concentration in the phantoms are detailed in [22]. For further construction details, including the speed of sound and attenuation coefficient, see [104]. Data was collected with an 18L6 transducer operating at a 10 MHz frequency using an Acuson S2000 scanner (Siemens Medical Solutions, Malvern, PA, USA). The phantoms are referred to as Phantom A (high concentration), Phantom B (medium concentration), and Phantom C (low concentration) having scatterer concentrations of 236, 9, and 3 per mm^3 , respectively. Seventy-two patches were acquired from each phantom.

3.2.3 Uncertainty Decomposition

To estimate the two components of the uncertainty in our model’s predictions, we used the method proposed in [105]:

$$Var(y) \approx \underbrace{E(\hat{y}^2) - E(\hat{y})^2}_{\text{Epistemic Uncertainty}} + \underbrace{E(\hat{\sigma}^2)}_{\text{Aleatoric Uncertainty}} \quad (8)$$

where $E(\hat{y})$ denotes the expected value of the prediction and $\hat{\sigma}^2$ is the predicted variance. Expected values are obtained by Monte Carlo sampling at inference time. According to (8), to estimate each uncertainty component for the simulation data, different observations of the input data and the model weights were required. The trained BNN was used in inference with 50 times sampling of weights for each of the 10 input observations for simulation data, yielding 50×10 different output

estimates. Having the total test data $dt \in \mathbb{R}^{341 \times 10}$, the dimensions of the total prediction were $k, \log_{10}(\alpha) \in \mathbb{R}^{341 \times 10 \times 50}$, where the last two dimensions correspond to different input realizations and different BNN inferences, respectively. As shown in (8), the total predictive uncertainty is the summation of epistemic and aleatoric uncertainties. The epistemic uncertainty was estimated by computing the standard deviation over the 50 model inferences, followed by averaging these deviations over the predictions of 10 different input realizations. On the other hand, the aleatoric uncertainty was estimated by first averaging the predictions across the 50 model inferences, and then computing the standard deviation over the predictions of 10 input realizations. This captures the variations across different realizations and removes the uncertainty of the model in different inferences. The visualization of the calculations is illustrated in Fig. 3.1. A similar approach was followed to compute the epistemic and aleatoric uncertainties of the estimated values of $\log_{10}(\alpha)$ in the experimental data, using different input observations (60 for layered phantom and 72 for homogeneous phantoms) and 50 model inferences.

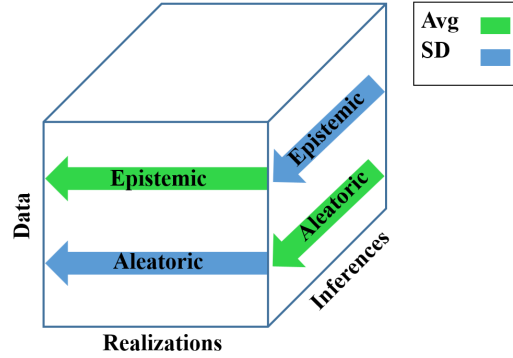


Figure 3.1: The visualization of computing the uncertainties.

3.3 Results

The Root Mean Square Error ($RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$) of the model's predictions is calculated where n represents the number of data points, y_i and $\hat{y}_i$ denote the ground truth labels and model's predictions, respectively. Fig. 3.2 illustrates the scatter plots representing the error values versus each uncertainty (epistemic and aleatoric) for estimated parameters $\log_{10}(\alpha)$ and k for the simulation data with different SNRs. The correlation coefficients between the error and each

uncertainty component for the simulation data at various SNR levels are provided in Table 3.1. All p -values were below 0.01, indicating the statistical significance of the reported correlation values. According to the plots in Fig. 3.2 and the correlation values in Table 3.1, a relatively high correlation between the error and each uncertainty was observed. For data with SNRs of 30 and 40, aleatoric uncertainty shows a stronger correlation with error than epistemic uncertainty for both parameters $\log_{10}(\alpha)$ and k . On the other hand, for data with an SNR of 20, epistemic uncertainty demonstrates a higher correlation with error compared to aleatoric one, again for both parameters. Based on Fig. 3.2 (b) and (d), there is an apparent lower bound for the error as a function of the aleatoric uncertainty. Determining the lower bound for the error can be crucially important in tissue characterization since it allows researchers to acquire the minimum prediction error values without knowing the ground truth.

Table 3.2 shows the uncertainties for different values of estimated $\log_{10}(\alpha)$ in the experimental phantoms. The analysis of uncertainty values in the plots in Fig. 3.2 and Table 3.2 (for simulation and experimental data, respectively), indicates that aleatoric uncertainty constitutes a larger proportion of the total uncertainty than the epistemic one. Therefore, adding more training data would not significantly reduce the overall predictive uncertainty here. If data acquisition is feasible, strategies such as using larger patches and angular compounding may help reduce aleatoric uncertainty, resulting in total uncertainty reduction.

Table 3.1: Simulation data Pearson’s correlation between error and uncertainties of predicted parameters (SNR = 20, 30, 40)

Parameters	Epistemic	Aleatoric	Total
$\log_{10}(\alpha)$	0.66, 0.75, 0.78	0.63, 0.80, 0.84	0.64, 0.81, 0.85
k	0.44, 0.52, 0.38	0.24, 0.71, 0.75	0.36, 0.72, 0.75

3.4 Discussion and Conclusion

Previous studies that quantified the uncertainty of HK parameter estimations [24, 25] mainly focused on epistemic uncertainty, as they only accounted for uncertainty over the model parameters. Moreover, Tehrani et al. in [98] estimated an uncertainty that is similar to aleatoric uncertainty. In

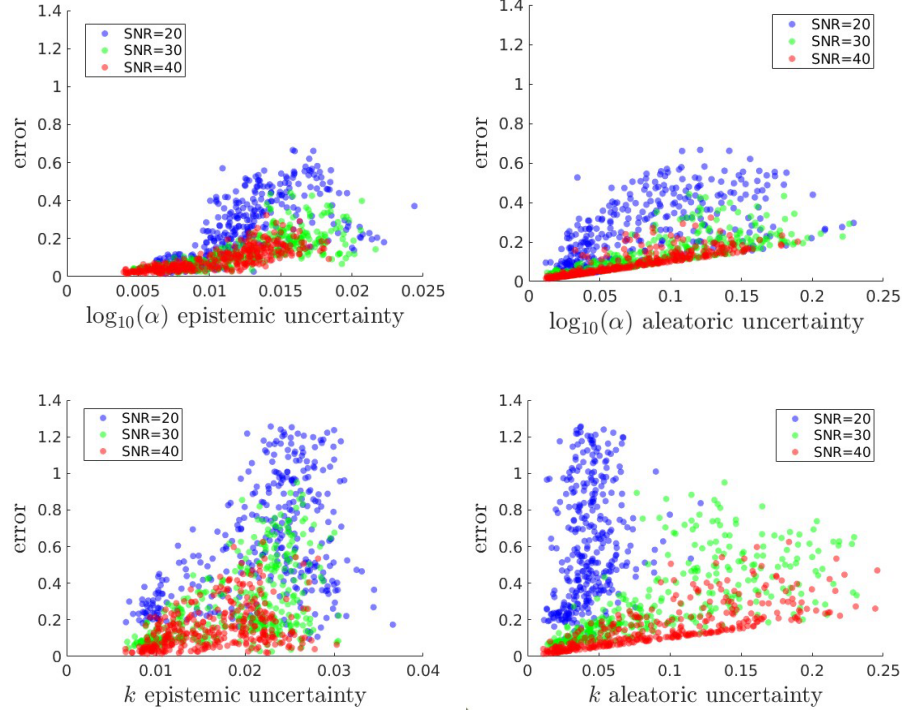


Figure 3.2: Error versus epistemic and aleatoric uncertainty for simulation data **(a)** Epistemic uncertainty for $\log_{10}(\alpha)$. **(b)** Aleatoric uncertainty for $\log_{10}(\alpha)$. **(c)** Epistemic uncertainty for k . **(d)** Aleatoric uncertainty for k .

Table 3.2: Aleatoric and epistemic uncertainties for predicted $\log_{10}(\alpha)$ of the investigated experimental phantoms.

Phantoms	Epistemic	Aleatoric	Interquartile range [25%, 75%] of predicted $\log_{10}(\alpha)$
Layered-R1	0.012	0.125	[0.854, 1.034]
Layered-R2	0.013	0.063	[1.081, 1.126]
A (High)	0.013	0.210	[0.723, 1.074]
B (Medium)	0.006	0.137	[0.387, 0.574]
C (Low)	0.010	0.173	[-0.311, -0.137]

this study, we acquired both uncertainties and investigated their correlation with error.

Calculating aleatoric uncertainty is challenging, as it requires multiple observations of the input data, which may not always be available. In contrast, epistemic uncertainty is easier to obtain through multiple inferences of the model in Bayesian frameworks. Our results in Section III show that the correlation between error and epistemic uncertainty is comparable to that between error and aleatoric uncertainty. This is valuable since epistemic uncertainty is always accessible, whereas aleatoric uncertainty may not be, especially in *in vivo* data where obtaining multiple observations from the same tissue is difficult.

In this study, we introduced a framework to quantify and decompose the predictive uncertainty into epistemic and aleatoric components for Homodyned K-distribution (HK-distribution) parameters using Bayesian Neural Networks (BNNs). Our results showed that the main contributor to the total uncertainty of the BNN model’s predictions is the aleatoric uncertainty. Moreover, investigating the relationship between prediction errors and each uncertainty component leads to identifying a lower bound for the error values.

Chapter 4

Point-of-Care Ultrasound in Diagnosis of Breast Cancer-Related Lymphedema: A Multimodal Database and Deep Learning Approach

This chapter is based on our paper [\[106\]](#).

Breast Cancer-Related Lymphedema (BCRL) is a common complication after breast cancer treatment, leading to chronic arm swelling and reduced quality of life. Early detection remains challenging due to limitations in current assessment methods and a shortage of expert clinicians. Point-of-care ultrasound (POCUS) imaging offers a cost-effective, non-invasive tool for BCRL evaluation; however, the visual disparity between affected and unaffected arms can be subtle. To address this, we introduce a multimodal dataset containing B-mode images, radio-frequency (RF) data, perometry-based arm volume measurements, the number of dissected lymph nodes, the time elapsed since surgery, and body mass index (BMI) from 57 subjects. Additionally, the dataset includes an assessment of the severity of BCRL by an expert clinician (A.T.). We also propose a deep learning-based framework for automated BCRL diagnosis using multimodal data. Our method first employs a Siamese-based architecture to extract and compare features from bilateral arm images,

addressing subject-specific variability. We further improve the method by incorporating clinical tabular data and investigating the effect of multimodal fusion and supervised contrastive learning on the model’s performance. To our knowledge, this is the first automated deep learning approach for BCRL diagnosis. Using three-fold cross-validation, the proposed Multimodal model achieved an AUC of 86.8%, an accuracy of 83.3%, and an F1 score of 76.6%. It outperformed the two other variants, Unimodal-A and Unimodal-B, in BCRL classification across most metrics. The dataset will be made publicly available to facilitate further research in lymphedema detection and other related medical imaging applications at <http://data.sonography.ai>. The source code and trained model will also be released at <http://code.sonography.ai>.

4.1 Introduction

Breast Cancer-Related Lymphedema (BCRL) is a common complication arising from breast cancer treatments such as surgery and radiotherapy, which often involve the removal or damage of lymph nodes within the lymphatic drainage system. This disruption leads to chronic accumulation of lymphatic fluid in the affected arm [50–52]. At least 10%, and up to 60%, of women report experiencing one or more upper-body symptoms between 6 months and 3 years following breast cancer surgery [50]. BCRL symptoms commonly include pain, swelling, increased arm volume, poor range of motion, tingling, stiffness, weakness, and numbness [50]. The condition negatively impacts patients’ quality of life by causing physical limitations, psychological distress, an increased risk of infections, and disruption of daily activities and overall well-being [53–55].

According to the International Society of Lymphology (ISL) staging system [107], BCRL progression includes several clinical stages (0-III). Stage 0 refers to a subclinical phase with impaired lymph transport but without visible swelling. Stage I involves early fluid accumulation that reduces with limb elevation. Stage II presents more persistent swelling with changes in tissue structure, and Stage III is characterized by severe swelling and fibrosis [107]. Studies have shown that in the later stages (II-III), the skin and subcutaneous fat undergo significant changes in both layer thickness and composition [55], [108–110], as well as changes in the mechanical properties [111]. In [112], it was shown that there are significant differences in echogenicity and tissue thickness between the

affected and unaffected arms.

Early diagnosis of BCRL and a standardized assessment protocol are essential for initiating timely treatment and minimizing the long-term and irreversible complications. The current approach for evaluating BCRL relies on various techniques, including using perometry to calculate volume difference, circumferential tape measurements, and subjective tissue evaluations [113–115]. While these methods are widely used, they have notable limitations, particularly in identifying lymphedema during its early stages. In addition, the shortage of expert clinicians in the field poses an additional challenge to the proper diagnosis and assessment of BCRL.

Medical imaging has great potential in BCRL diagnosis and assessment. Among available modalities, ultrasound stands out due to its cost-effectiveness, non-invasiveness, portability, and ability to provide real-time imaging. It has been confirmed to be a valuable tool for BCRL evaluation [55–59]. Ultrasound enables the visual examination of various tissue compartments, including the skin, subcutis, and muscle, to detect fluid accumulation and changes in texture or thickness [112]. B-mode ultrasound, in particular, produces two-dimensional images where tissue echogenicity is depicted as varying levels of brightness. However, the affected and unaffected arms by BCRL often look similar in B-mode ultrasound images [25]; therefore, distinguishing between them using these images remains challenging, even for experienced radiologists. This highlights the need for an automated approach to assist clinicians in accurately diagnosing BCRL from ultrasound images.

Deep learning has emerged as the state-of-the-art (SOTA) approach for a wide range of medical image analysis tasks and is increasingly being adopted for analyzing medical ultrasound images [116]. In the context of lymphedema, deep learning-based methods show promise for improving diagnosis and assessment using ultrasound imaging [25]. In this study, we explore the potential of such methods to diagnose BCRL by analyzing differences in the composition, texture, and thickness of tissue layers between the affected and unaffected arms. The use of deep learning for automated BCRL assessment can potentially overcome the challenges in lymphedema diagnosis, including the lack of standardized diagnosis thresholds, difficulty in early detection, and the shortage of trained specialists, by making advanced diagnostic tools more accessible. These methods can also potentially lead to the reduction of inter-observer differences between clinicians' diagnoses.

Medical images are not the sole source of information for tissue characterization and disease

diagnosis. Clinical assessments and physiological measurements, such as body weight and various blood parameters, can provide complementary information that imaging alone cannot capture. These additional inputs are referred to as “tabular data”. Integrating tabular data with medical images has the potential to enhance diagnostic accuracy. Several approaches have been proposed to effectively fuse this multimodal data [117–120].

The scarcity of publicly available datasets has limited image analysis research for BCRL. To the best of our knowledge, no automatic method currently exists for the diagnosis of BCRL. In this study, we addressed this gap by introducing a publicly available multimodal BCRL dataset comprising B-mode images, radio-frequency (RF) data, perometry-based arm volume measurements, the number of dissected lymph nodes, the time elapsed since surgery, and body mass index (BMI) from 57 subjects. Regarding ultrasound data, only B-mode images are utilized in our current analysis; the RF data are provided to support future research. The primary contributions of this work are as follows:

- (1) We introduce the first multimodal BCRL dataset from 57 breast cancer survivors. We conducted an extensive data collection process from October 2023 to May 2025, focusing on subjects who had undergone breast surgery and lymph node dissection. Our dataset comprises 798 RF frames and 1254 B-mode images acquired from 5 landmarks, addressing the limited availability of standardized datasets for investigating BCRL. The provided dataset also includes an expert clinician’s assessments of BCRL severity, perometer arm volume measurements, and photographs of each subject’s arms. Additional subject-specific information, including the number of dissected lymph nodes, time passed since surgery, and body mass index (BMI), is provided to support comprehensive analyses. The dataset is utilized for BCRL detection in our study. It can also facilitate applications and studies in other fields, such as tissue characterization. We will make the dataset publicly available after the acceptance of the manuscript.
- (2) We propose a deep learning method to diagnose BCRL using our collected ultrasound dataset. To the best of our knowledge, this is the first deep learning approach for BCRL classification. The proposed method does not require any segmentation and is subject-independent, meaning

it is designed to account for the inherent variability in tissue texture and body size, which differ significantly between individuals. This variability presents unique challenges in the classification, as it requires the model to generalize across diverse patterns and morphologies. The proposed solution employs a Siamese-based architecture with ResNet-50 as the encoders that compares images of the left and right arms of each subject to capture subject-specific tissue characteristics, reducing the need for a large training dataset. Due to the complication of dataset collection and the high variability of arm tissue compositions among the subjects, there was no prior study to automatically classify BCRL.

- (3) We enhance the method’s performance by fusing imaging and tabular data. Perometry-derived arm volume difference, number of dissected lymph nodes, time elapsed since surgery, and BMI are incorporated alongside the ultrasound images. Each of these features can be correlated with the risk of developing lymphedema. By combining structural insights from ultrasound images with subject-specific information from tabular data, our model leverages complementary information to improve classification accuracy.
- (4) We explore the use of a contrastive learning approach on the image data features. Specifically, we employ the Supervised Contrastive (SupCon) loss proposed by [121]. This approach uses the labeled information to pull together the clusters of samples belonging to the same class, while simultaneously pushing apart clusters of samples from different classes within the batch.

4.2 Related work

Among studies on BCRL assessment using ultrasound, skin and subcutis thickness in women with BCRL has been examined using dual-frequency ultrasound [122]. Their findings indicated that the degree and uniformity of swelling in these layers are correlated with BCRL, a conclusion supported by later studies [55, 123]. In addition, texture analysis of ultrasound images from affected and unaffected limbs was performed in [108], and structural changes in the subcutaneous tissue caused by lymphedema were explored in [124].

Deep learning methods have been widely utilized for ultrasound data analysis for different tasks such as disease detection and tissue characterization [34, 125, 126]. Due to the limited availability of labeled medical image datasets, transfer learning has become a widely adopted and effective approach for leveraging knowledge from models pre-trained on large-scale datasets [30, 127]. In [23], an Inception-ResNet-v2 deep CNN pre-trained on the ImageNet dataset [128] was fine-tuned on liver B-mode ultrasound image sequences to detect steatosis level. Lazo *et al.* compared the performance of different CNNs in breast cancer tumor classification of breast ultrasound images using transfer learning [129]. In [21], pre-trained SOTA CNN networks and different pre-trained vision transformers (ViTs) were fine-tuned to classify breast ultrasound images to diagnose cancerous masses, with ResNet-50 achieving the best performance among CNN models and ViT-B/32 yielding the best results among ViT models. Yang *et al.* fine-tuned a ResNet-18 pre-trained on ImageNet for ultrasound image classification of thyroid nodules [130]. In [131], a feature fusion framework was proposed for thyroid ultrasound image classification, combining texture features derived from pathology data with deep image features extracted by a ResNet-based model using transfer learning.

Deep learning has been leveraged in BCRL studies to identify and segment the structural layers within arm tissue. In [25], a Gated Shape Convolutional Neural Network was modified to segment different layers of arm tissue in BCRL patients. This contributed to the staging of lymphedema since BCRL affects the thickness of tissue layers, particularly the subcutaneous fat layer [55]. A recent study proposed a hybrid structure-oriented Transformer (HSformer) designed to segment subcutaneous fat, skeletal muscle, and bone in arm musculoskeletal ultrasound images to aid BCRL diagnosis [60]. The model incorporates hierarchical-consistency position encoding and structure-biased attention mechanisms to handle the layered morphology of ultrasound images. Their experiments, conducted on the dataset introduced by Goudarzi *et al.* [25], demonstrated SOTA segmentation performance.

Siamese networks, originally introduced for signature verification tasks, have gained popularity in vision applications that require pairwise image comparisons [132]. In [133], the Siamese network has been effectively used for learning image similarity by training on matching and non-matching image pairs using contrastive loss. The effectiveness of Siamese networks for unsupervised representation learning was further demonstrated in SimSiam, which showed that meaningful features

can be learned without negative pairs, large batches, or momentum encoders [134]. In [135], a Siamese network architecture based on a VGG encoder was utilized, employing cross-entropy loss to train a binary classifier distinguishing matching and non-matching pairs. In [136], a Siamese network was utilized to grade disease severity on a continuous spectrum using medical images. An attention-based Siamese network was proposed in [137] to address few-shot and fine-grained medical image recognition challenges, demonstrating superior performance on COVID-19 lung image classification with limited training samples. This network was shown to outperform traditional models like SVM and Discriminant Analysis for classifying high-dimensional radiomic features extracted from T2-weighted breast MRI, particularly in a low-sample setting [138]. Chen et al. [139] proposed a Siamese network augmented with anatomical symmetry analysis for detecting pelvic fractures in X-ray images, achieving state-of-the-art performance by leveraging contrastive feature learning to detect clinically relevant asymmetries. Additionally, a Siamese CNN was proposed for learning medical image representations with less supervision, relying only on binary image pair information, and achieved performance comparable to single supervised CNNs in content-based image retrieval [140].

Recently, multimodal vision language models have shown promise in integrating imaging and nonimaging data for improved classification and segmentation [119,120,141]. Fusion of tabular data and vision can be considered to leverage information from both sources [142]. In [143], a contrastive learning approach was proposed to jointly learn features from medical images and clinical tabular data. Similarly, Tehrani *et al.* introduced a method that leverages statistical features in conjunction with ultrasound images through supervised fusion and deep supervision [22].

4.3 Materials and Methods

4.3.1 Data collection

Ethics approval for human data collection was granted by the McGill University Health Centre (MUHC). *In vivo* data collection took place at the MUHC lymphedema Clinic, and all participants provided written consent before taking part in the study. The study included 57 female breast cancer survivors who had undergone breast surgery and lymph node dissection. The selection criteria

ensured at least 3 dissected nodes and no evidence of metastasis or cancer recurrence. 40.4% of our subjects were not affected by BCRL, and the rest were affected by unilateral BCRL in stage II. Table 4.1 presents relevant clinical information of the study subjects.

Table 4.1: Clinical Information of Study Participants

Parameter	Value (mean $\pm$ SD)
Total number of subjects	57
Age	55.04 $\pm$ 10.52
BMI (kg/m ²)	27.03 $\pm$ 4.50
Calibrated arm volume difference (ml)	8.70 $\pm$ 13.68
Number of dissected nodes	9.82 $\pm$ 6.25
Time passed since surgery (years)	5.34 $\pm$ 6.63

Ultrasound data were acquired using an L15 HD high-frequency linear point-of-care ultrasound (POCUS) probe (Clarius Mobile Health, Vancouver, BC, Canada), with a center frequency of 10 MHz. The imaging depth and gain were set to 3.8 cm and 75% of maximum gain, respectively. Both left and right arms of each subject were scanned. Ultrasound images and RF data were collected from the five landmarks along the ventral side of the arm, as shown in Fig. 4.1(a). The landmarks were defined based on anatomical distances to ensure consistency across subjects. They were located as follows:

- The first landmark is at 25% of the distance between the wrist crease and the elbow crease.
- The second landmark is at 75% of the same interval.
- The third landmark is on the elbow crease itself.
- The fourth landmark is at 25% of the interval between the elbow crease and the glenohumeral joint.
- The fifth landmark is at 75% of the same interval.

This structured approach ensured comprehensive imaging coverage of the arms. Fig. 4.1(b) illustrates the probe placement on the arm, positioned perpendicular to the muscle fibers. Samples of the ultrasound images are illustrated in Fig. 4.2. For landmarks 2 and 4, additional images were acquired with the probe aligned parallel to the muscle fibers for potential use in future studies (Fig. 4.2,

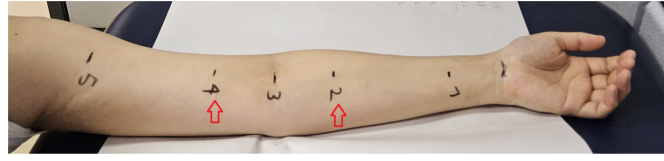
Landmark 2 (A) and Landmark 4 (A)). In this study, only ultrasound images from landmarks 2 and 4 (highlighted in Fig. 4.1), acquired with the probe perpendicular to the muscle fibers, were selected for analysis, as lymphedema typically manifests more frequently in these regions. Overall, we provided $57 \times 14 = 798$ RF acquisitions, where each subject contributed data from 10 landmarks on both arms, specifically, one RF frame from each landmark 1, 3, and 5, and two frames from each landmark 2 and 4. Additionally, two extra B-mode frames were included for landmarks 2 and 4, as detailed in Section 4.3.3, resulting in a total of $57 \times 22 = 1254$ image frames.

Before data acquisition, arm volume measurements were obtained for each participant using a perometer. After data collection, an expert physician with more than 30 years of experience in the evaluation of lymphedema (A.T.) examined each participant to assess the severity of BCRL. The evaluation was done based on the segmental volume differences below or above the elbow, the consistency of the tissue, and the amount of inflammation. The physician assigned a custom severity score on a scale from 0 to 5, where 0 indicates no lymphedema; 1, very mild; 2, mild; 3, moderate; 4, severe; and 5, very severe. Intermediate values between these levels were also used, as shown in Table 4.2.

Table 4.2: Lymphedema severity scores of the study subjects

Sevrity score	Number of subjects
No lymphedema (0)	24
Very mild lymphedema (1)	11
Very mild to mild lymphedema (1-2]	13
Mild to moderate lymphedema (2-3]	5
Moderate to severe lymphedema (3-4]	3
Severe to very severe lymphedema (4-5]	1

The currently available ultrasound BCRL dataset proposed in [25] includes B-mode images and RF data of 39 subjects, including 20 healthy women and 19 breast cancer survivors with stage II unilateral BCRL, and severity levels of moderate to severe (3-4). Our dataset covers a 46% higher number of subjects. All of our subjects are breast cancer survivors, 40.4% of them were not affected by BCRL, and the rest were affected by unilateral BCRL of stage II with specified severities from a wide range of very mild to severe (1-4). Additionally, our landmarks were located on the ventral side of the arm, while in the previous dataset, the dorsal side of the arm was scanned. We modified



(a)



(b)

Figure 4.1: Data collection with POCUS. (a) shows the landmarks where images were acquired. (b) demonstrates how the probe was placed on the arm. For landmarks 2 and 4, we collected additional images with the probe parallel to the muscle fibers (90 degrees rotated compared to the probe direction in the image).

the location of data collection as the ventral side is where lymphedema manifests earliest and most often. A critical final distinction is that our dataset also provides multimodal information, including arm perometry volume measurements, the number of dissected nodes, time elapsed since surgery, and body mass index (BMI) as a complementary resource.

We should also note that 57.9% of our subjects are breast cancer survivors shortly after surgery (within one year) who may develop lymphedema. Some of the patients were already affected by lymphedema, and for the majority of affected ones, the BCRL was not easily identifiable without the expertise of our experienced clinician due to the small amount of swelling. Also, the arm volume difference derived by perometry might not be significant. Therefore, automating the diagnosis helps due to the scarcity of expert physicians and equipment for lymphedema diagnosis.

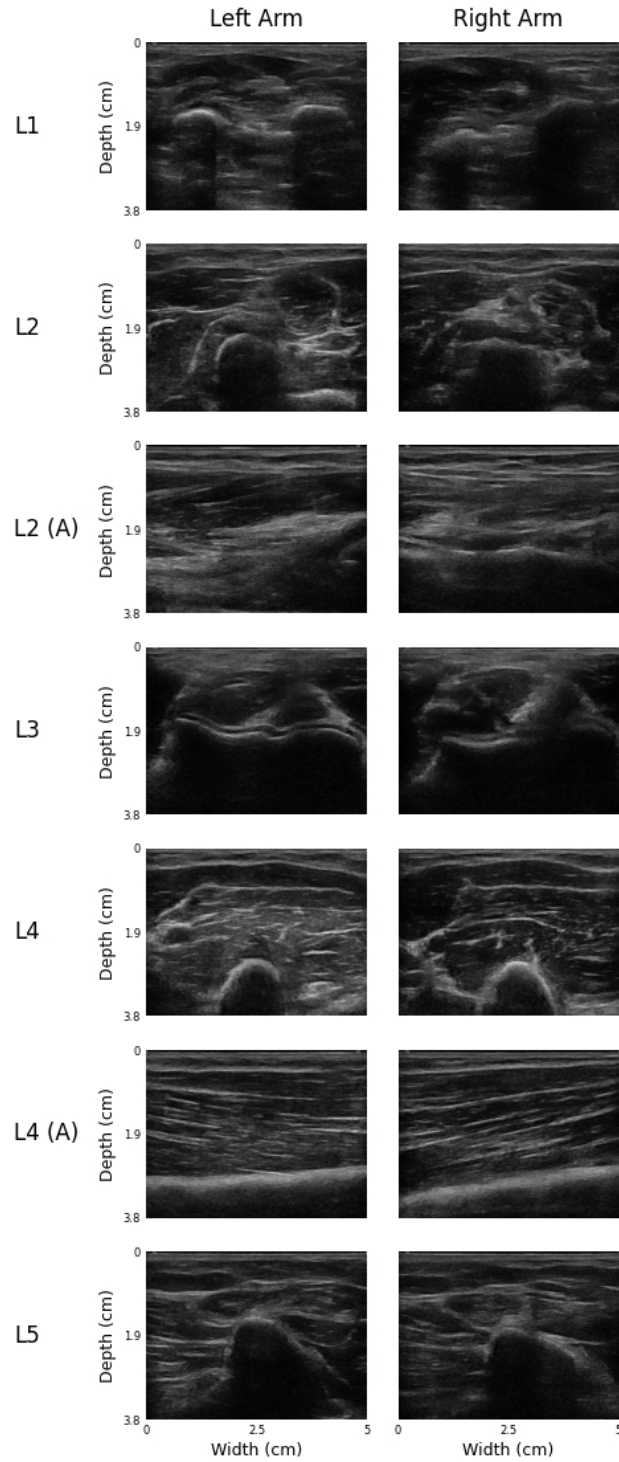


Figure 4.2: B-mode images of the arm of a subject diagnosed with mild BCRL (severity level 2) in the right arm. The left and right columns correspond to unaffected and affected arms, respectively. L1 to L5 are landmarks. Images in columns L2 (A) and L4 (A) were collected with the probe placed parallel to the muscle fibers.

4.3.2 Dataset preparation and pre-processing

Ultrasound Image Data

The input consisted of grayscale ultrasound B-mode images with an original size of 581×446 pixels. All images were resized to 448×448 pixels and normalized to the range 0 to 1. To enhance model generalization, we applied data augmentation techniques. Because of the special physical properties of ultrasound images, the augmentation techniques that respect ultrasound physics and can be applied are limited. Specifically, we used random cropping up to 5% and random rotations up to ± 5 degrees. We also added Gaussian noise with zero mean and a standard deviation scaled by 0.001.

Tabular Data

We incorporated relevant clinical variables as additional features in our analysis. The tabular data included the calibrated arm volume difference, the number of dissected lymph nodes, the time elapsed since surgery (in years), and BMI. The average values are presented in Table 4.1. The volume difference between the arms is often associated with the presence and severity of lymphedema. This difference was derived using a perometer and calibrated by dividing the value by the volume of the reference arm, which is the unaffected arm in BCRL patients or the contralateral arm to the surgery side in healthy subjects. A higher number of dissected lymph nodes during surgery can increase the risk of developing lymphedema. The time passed since surgery has been shown to elevate the risk of lymphedema [50]. Obesity or higher BMI can also be correlated with an increased risk of lymphedema [51]. Gaussian noise with a standard deviation of 0.01 was added to tabular features during training prior to integration with the imaging data.

4.3.3 Proposed method

Since the tissue composition and the ratio of fat and muscle layer thickness may vary between individuals, analyzing each image in isolation can be too sensitive to subject-specific tissue characteristics. To address this challenge, we designed a subject-independent method that accounts for the inherent variability in tissue texture and size between individuals. Specifically, we employed

a Siamese-based architecture in which two identical encoders with shared weights were utilized to extract features from both left and right arm images of each subject. By incorporating bilateral comparisons, the model becomes more robust to subject-specific differences and focuses on the disparity between the two arms. The overview of the proposed method is illustrated in Fig. 4.3.

We selected three frames with a 2.4-second difference from each landmark to increase the number of training samples. Three frames from the left landmark were paired with all three frames of the right landmark, resulting in $3 \times 3 = 9$ image pairs. Each pair was considered as an individual sample. A binary classification task was defined, where image pairs from subjects either without lymphedema or with very mild lymphedema (severity score = 1) were assigned to the negative class, and those from subjects diagnosed with mild to severe lymphedema (severity score > 1) were labeled as the positive class. The rationale behind assigning very mild cases to the negative class is that tissue changes may not have developed or be very subtle, such that they are closer to negative cases.

We employed a ResNet-50 [144] encoder pre-trained on ImageNet [128] using the DINO (Distillation with No Labels) learning approach [145]. DINO is a self-supervised learning method that enables models to learn rich visual representations without requiring labeled data. It works by training a student network to match the output of a momentum teacher network, encouraging the emergence of semantically meaningful features. The final classification head of the ResNet-50 was removed, resulting in a 2048-dimensional (2048-D) feature representation for each image. The absolute difference between two feature vectors produced by each Siamese branch was computed to capture the disparity between the two images. It was then fed to a multi-layer perceptron (MLP), including two fully connected layers with Leaky ReLU activations and a dropout rate of 0.2, producing a 64-D image embedding. It was then passed through a classifier sub-network to generate the final prediction.

To incorporate clinical information, tabular data were encoded as a 4-D feature vector, which was processed through two fully connected layers with a dropout rate of 0.5, resulting in a 32-D tabular embedding. This embedding was concatenated with the 64-D image embedding and passed through a classifier sub-network for prediction.

Additionally, we evaluated the discriminative capacity of the tabular features in isolation. The

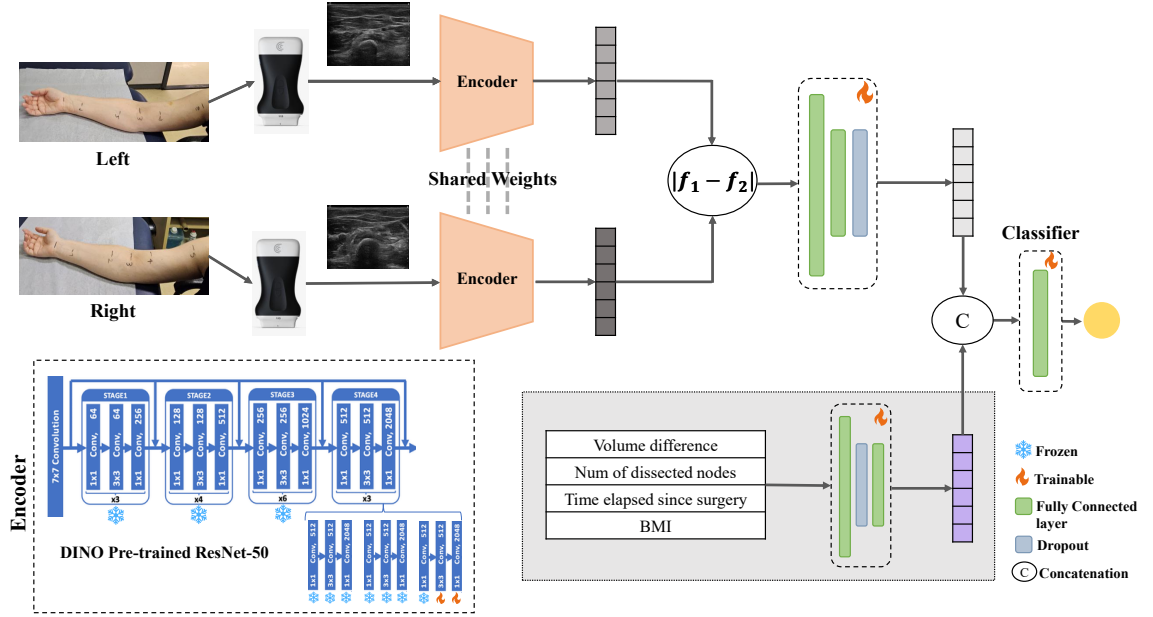


Figure 4.3: The proposed multimodal architecture that integrates ultrasound images of both arms with tabular data. The POCUS B-mode images are fed to encoders with shared weights, and the absolute difference of the image embeddings is concatenated with tabular embeddings. The final vector was passed through a classifier. The illustration of encoder architecture is reproduced from [1].

4-D tabular feature vector was processed through the MLP to obtain a 32-D embedding, which was then passed into the classifier to predict BCRL status.

These three configurations are referred to as Unimodal-A (image-only), Multimodal (image + tabular), and Unimodal-B (tabular-only), respectively. A comprehensive performance comparison of these models is provided in Section 4.4.

Training details

Given the limited size of the dataset, we adopted a three-fold cross-validation strategy for performance evaluation. To prevent data leakage, we ensured that images from the same participant did not appear in both the training and validation sets. Each validation fold contained 33.3% of the data and maintained a balanced class distribution.

To mitigate overfitting arising from the limited dataset size, we adopted a partial fine-tuning strategy. Specifically, all layers of the pre-trained ResNet-50 except for the last two convolutional

layers were frozen. In addition to two convolutional layers of the ResNet-50, all embedding layers responsible for projecting the image and tabular features into a latent space, as well as the final classification layer, were fine-tuned in training.

The training was conducted for 200 epochs using a weighted Binary Cross-Entropy (BCE) loss (9), where a weight p_c was assigned to the positive class to address the class imbalance due to the smaller number of samples in class 1. Optimization was performed with the Adam optimizer, and a weight decay of 10^{-5} was applied for regularization. A step decay learning rate scheduler was employed, reducing the learning rate (base lr) by a factor of γ every s epoch. The specific values of p_c , base learning rate, γ , and s used in each experiment are listed in Table 4.3.

$$\mathcal{L}_{\text{BCE}} = (1 - y) \cdot x + \log(1 + e^{-|x|}) + p_c \cdot y \cdot (-x + \log(1 + e^{-|x|})) \quad (9)$$

Table 4.3: Training hyperparameters for experiments

Experiment	p_c	base lr	γ	s
Unimodal-A	1.4	1e-5	0.9	20
Multimodal	1.1	1e-5	0.9	20
Unimodal-B	1.1	1e-4	0.6	50
Unimodal-A+CL	1.4	1e-5	0.9	20

Contrastive Learning

Contrastive learning is a self-supervised framework that trains models to pull semantically similar pairs closer and push dissimilar ones apart in the embedding space. Positive pairs are typically created through data augmentations of the same sample, while negatives are drawn from other samples of the batch. This approach has become popular in recent years and has helped improve self-supervised learning methods [146–148]. Contrastive learning has also been extended to supervised settings by utilizing label information to pull together embeddings of the same-class samples and push apart those from different classes, resulting in improved performance over traditional supervised losses [121].

We investigated a supervised contrastive learning strategy and applied it to the image feature embeddings ($|f_1 - f_2|$ in Fig. 4.3). The supervised contrastive (SupCon) loss [121] can be defined

as:

$$\mathcal{L}_{\text{out}}^{\text{sup}} = \sum_{i \in I} \mathcal{L}_{\text{out},i}^{\text{sup}} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a / \tau)} \quad (10)$$

Let $z = \text{Proj}(\text{Enc}(\tilde{x})) \in \mathbb{R}^{D_P}$ denote the projected feature representation obtained by applying a projection head to the encoder output. Here, the $\cdot$ symbol represents the inner (dot) product, and $\tau \in \mathbb{R}^+$ is a positive scalar temperature parameter that scales the logits. We set $\tau = 0.1$ in our experiments. For each anchor index i , $P(i) \equiv \{p \in A(i) \mid \tilde{y}_p = \tilde{y}_i\}$ denotes the set of positive indices (i.e., samples belonging to the same class as i), and $A(i)$ is a member of the set of all indices excluding i . In this setup, each anchor has multiple positive pairs and $2N - 2$ negative pairs, where N is the batch size. We adopted the $\mathcal{L}_{\text{out}}^{\text{sup}}$ loss variant (10), where the averaging over positive pairs is performed outside of the logarithm. We investigated the impact of incorporating supervised contrastive (SupCon) loss in training the model with ultrasound image data. Initially, a ResNet-50 encoder was trained using the SupCon loss, where the loss was computed on image embeddings obtained after passing through a projection head. Subsequently, transfer learning was employed by initializing the encoder of the Unimodal-A architecture with these pre-trained weights, which were kept frozen during further training. The embedding layers (with the dropout rate increased to 0.5 and an additional dropout layer with a rate of 0.5 added to mitigate overfitting) and the classification head of the Unimodal-A model were fine-tuned on the POCUS image data to adapt the network for the BCRL classification task. The hyperparameters for training this model are presented in Table 4.3. The method is referred to as Unimodal-A+CL, and its results are presented in Section 4.4.

4.4 Results

The classification results for the different model variants are presented in Table 4.4. Each model was evaluated using standard classification metrics, including accuracy, area under the receiver operating characteristic curve (AUC), F1 score = $2(\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$, precision, and recall. Results are reported as the mean $\pm$ standard deviation across three validation folds.

To strengthen the validity and reproducibility of the results, five independent instances of each model were trained with different seeds. Two evaluation strategies were employed. In the first strategy (model averaging), classification metrics were computed separately for each trained instance, and the final reported values represent the average of these metrics for the best validation loss across the five runs. In the second strategy, an ensembling approach was applied to improve generalizability [149]. Specifically, prediction probabilities of the best model from the five trained instances were averaged to generate the final predictions, yielding more robust and accurate performance estimates.

Among the model variants, the ensemble variant of the Multimodal model achieved the highest performance across all evaluation metrics except AUC and Precision with a marginal decrease compared to Unimodal-B (Table 4.4). Its overall performance outperformed both Unimodal-A and Unimodal-B, highlighting the effectiveness of combining POCUS imaging and tabular clinical data in enhancing classification efficacy. These findings confirm that leveraging complementary information from POCUS images and tabular clinical data enables the network to capture discriminative features that single-modality approaches cannot fully represent.

We also examined the impact of incorporating supervised contrastive learning (denoted Unimodal-A+CL) for training with image data. This approach did not outperform the weighted BCE-trained Unimodal-A model, suggesting that in this setting, supervised training with weighted BCE loss remains more effective for optimizing overall performance.

We also visualized the learned feature embeddings using two-dimensional t-SNE [150], as shown in Fig. 4.4, where the concatenated image and tabular features from all validation folds are combined, and points are colored according to their class labels. It can be observed that feature vectors corresponding to the same class tend to cluster together in the t-SNE space, indicating a good class separability in the learned embedding space.

Overall, these results demonstrate that multimodal learning with the Multimodal architecture provides an effective framework for automated BCRL diagnosis, establishing a strong baseline for future advancements in deep learning-based lymphedema assessment.

Table 4.4: Mean $\pm$ standard deviation of validation results across the three folds, with the best performance highlighted in bold.

Methods	AUC (%)	Accuracy (%)	F1 score (%)	Precision (%)	Recall (%)
Averaged					
Unimodal-A	68.4 ± 2.0	64.1 ± 1.4	57.0 ± 2.2	54.4 ± 4.3	57.0 ± 7.9
Unimodal-A+CL	60.8 ± 3.4	56.2 ± 3.8	51.6 ± 4.8	48.7 ± 7.0	60.0 ± 6.8
Unimodal-B	78.3 ± 3.9	76.2 ± 4.3	58.8 ± 7.3	75.5 ± 12.7	53.9 ± 5.5
Multimodal	89.0 ± 5.7	83.2 ± 5.5	78.3 ± 8.9	77.2 ± 2.2	81.4 ± 16.5
Ensemble					
Unimodal-A	67.6 ± 3.7	64.3 ± 3.4	55.2 ± 1.68	55.1 ± 7.4	56.3 ± 5.5
Unimodal-A+CL	64.1 ± 6.3	60.4 ± 3.4	50.4 ± 6.5	51.5 ± 9.7	52.4 ± 16.4
Unimodal-B	91.3 ± 5.8	78.9 ± 5.3	67.6 ± 11.7	80.5 ± 4.8	58.9 ± 14.6
Multimodal	91.0 ± 6.9	87.0 ± 3.9	83.5 ± 7.3	79.3 ± 1.7	89.1 ± 15.4

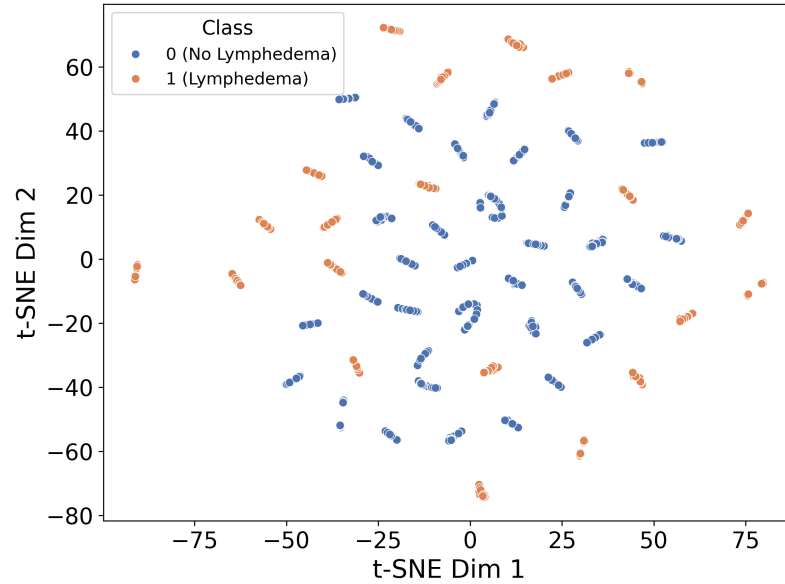


Figure 4.4: Two-dimensional t-SNE visualization of the concatenated image and tabular features extracted from all validation samples. Each point represents a sample, and colors indicate the ground truth class labels.

4.5 Discussions

This study demonstrates the feasibility of automated deep learning-based BCRL detection using POCUS imaging, establishing a baseline for future advancements in deep learning methods for lymphedema detection. We also investigated the integration of tabular clinical data with imaging. These findings highlight the potential of deep learning to support early and accurate BCRL diagnosis, aiding clinical decision-making and facilitating the development of standardized assessment protocols.

To support these investigations, we systematically organized and prepared the acquired multi-modal data and made it publicly available. A comprehensive set of experiments was conducted to identify optimal network architectures, fine-tuning strategies, training hyperparameters, and effective strategies for addressing challenges associated with small datasets, such as data augmentation and partial network fine-tuning.

The study has a few limitations. First, due to the limited dataset size, we employed three-fold cross-validation without a separate test set, which may impact the generalizability of the reported results. Despite data collection for over one year, budget restrictions limited our cohort to 57 subjects, which constrained data diversity and sample size.

Second, the current study focused only on ultrasound images acquired from landmarks 2 and 4, where BCRL is most likely to manifest. However, for some individuals, lymphedema may develop in other regions of the arm, introducing potential label noise into the dataset. Future work could incorporate data from other landmarks provided in this study and further expand the dataset by collecting data from the entire arm to improve the classification performance.

Additionally, for cases of very mild lymphedema (severity score = 1), tissue changes may not have developed yet or may be undetectable, which complicates the accurate labeling of these cases, which can be another source of label noise.

Lastly, the absence of existing automated methods for POCUS-based BCRL classification prevents direct benchmarking against prior approaches. We anticipate that releasing our dataset will encourage further research in this area and inspire new methods that can be compared to our baseline.

In addition to expanding the dataset size and diversity, future directions could explore incorporating RF data (included in this dataset), shear wave elastography [151], and quantitative ultrasound [92] for tissue characterization. Exploring self-supervised learning [148] and co-teaching [152] strategies may help mitigate label noise. A larger dataset would also enable the model to stratify BCRL cases by severity, which supports multi-class classification. Lastly, integrating additional clinical data, such as patient questionnaires, could further improve diagnostic accuracy and reflect real-world clinical workflows more effectively.

4.6 Conclusion

In this study, we proposed a deep learning framework for automated BCRL detection using ultrasound imaging, exploring subject-independent modeling and multimodal data fusion. To our knowledge, this is the first automated method developed for BCRL classification. The superior performance of our Multimodal network across most evaluation metrics demonstrates the effectiveness of combining ultrasound imaging with tabular clinical data to improve classification performance. The proposed method has the potential to assist clinicians in accurately diagnosing BCRL. Furthermore, such tools could support the standardization of BCRL assessment and help address the shortage of specialists by making advanced diagnostic capabilities more widely accessible.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In Chapter 2, we presented a Bayesian deep learning model trained on QUS features extracted from *in vivo* ultrasound data of force-fed duck livers to classify liver steatosis. This approach not only demonstrated the feasibility of applying deep learning to liver tissue characterization but also introduced a probabilistic framework capable of estimating prediction uncertainty, which is essential in clinical decision-making.

Chapter 3 presented the uncertainty decomposition of QUS parameter estimation. We developed a Bayesian framework for quantifying both epistemic and aleatoric uncertainties in the estimation of Homodyned-K distribution parameters and demonstrated how these uncertainties relate to prediction error.

In Chapter 4, we proposed the first automated deep learning method for the detection of BCRL using ultrasound. By developing a multimodal dataset that combines ultrasound images with clinical information, we showed that fusing image and tabular features improves classification performance. This work supports the development of AI-assisted tools to enhance the diagnostic accuracy and expand access to BCRL screening in settings without dedicated ultrasound experts.

5.2 Future Work

- Chapter 2: Future work should focus on improving the accuracy of QUS feature extraction, as the model’s performance is inherently dependent on the quality of these features. Incorporating additional QUS parameters may further enhance classification performance. With access to larger datasets, multi-class classification could be explored to distinguish between different grades of liver steatosis. While the use of a pre-clinical duck model offers a practical solution to the challenge of limited ground truth in human studies, validating the proposed framework on human liver data remains necessary. Applying the same deep learning pipeline to human cases will require retraining the model on human-specific QUS features. To address the limited availability of annotated human data, transfer learning techniques should be considered.
- Chapter 3: This chapter introduced a Bayesian framework to quantify and decompose predictive uncertainty in estimating Homodyned K distribution parameters. Currently, we use simulated and experimental phantom data. Other QUS biomarkers and also *in vivo* data can also be incorporated to improve the method’s clinical relevance. The proposed uncertainty estimates can also be leveraged for frame selection, where frames with high predictive uncertainty are excluded or deprioritized during analysis or training to improve overall model reliability.
- Chapter 4: Future work could address limitations related to dataset size and diversity. Expanding data collection to include more subjects and other landmarks provided in this study, and further expanding the dataset by collecting data from the entire arm, could enhance generalizability and reduce label noise. As automated methods for ultrasound-based BCRL classification are currently lacking, we anticipate that releasing our dataset will facilitate future benchmarking and stimulate further research. Additionally, incorporating other modalities, such as RF data (included in this dataset), shear wave elastography, and QUS, can provide tissue characterization. Exploring self-supervised or semi-supervised learning strategies may help address the issue of label noise. Finally, incorporating more clinical information, such

as patient questionnaires, physical examinations, and full medical history, could enhance diagnostic accuracy and better align the model with real-world clinical practices. Caution, however, must be taken when publicly releasing comprehensive patient data to make sure human subjects remain anonymous.

- Chapters 2 to 4: Future work could extend the multimodal fusion approach from Chapter 4 by using probabilistic uncertainty estimation methods presented in Chapter 2. This will provide uncertainty estimates for BCRL classification, especially for borderline cases. It can also validate the proposed Bayesian frameworks from Chapters 2 and 3 using clinical datasets similar to the multimodal dataset presented in Chapter 4, using transfer learning and domain adaptation to account for differences between phantom, animal and clinical ultrasound data. Finally, future work could explore self-supervised and semi-supervised learning techniques, as suggested in Chapter 4, within a Bayesian learning framework established in Chapters 2 and 3. This could reduce the challenges associated with limited labeled data and enhance model robustness in liver steatosis and BCRL applications.

Bibliography

- [1] L.-R. Yeh, Y. Zhang, J.-H. Chen, Y.-L. Liu, A.-C. Wang, J.-Y. Yang, W.-C. Yeh, C.-S. Cheng, L.-K. Chen, and M.-Y. Su, “A deep learning-based method for the diagnosis of vertebral fractures on spine mri: retrospective training and validation of resnet,” *European Spine Journal*, vol. 31, no. 8, pp. 2022–2030, 2022.
- [2] T. J. Hall, “Aapm/rsna physics tutorial for residents: topics in us: beyond the basics: elasticity imaging with us,” *Radiographics*, vol. 23, no. 6, pp. 1657–1671, 2003.
- [3] J. A. Jensen, “A model for the propagation and scattering of ultrasound in tissue,” *The Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 182–190, 1991.
- [4] W. D. O’Brien Jr, “Ultrasound–biophysics mechanisms,” *Progress in biophysics and molecular biology*, vol. 93, no. 1-3, pp. 212–255, 2007.
- [5] J. A. Zagzebski, “Essentials of ultrasound physics,” (*No Title*), 1996.
- [6] B. B. Goldberg, J.-B. Liu, and F. Forsberg, “Ultrasound contrast agents: a review,” *Ultrasound in medicine & biology*, vol. 20, no. 4, pp. 319–333, 1994.
- [7] R. S. Cobbold, *Foundations of biomedical ultrasound*. Oxford university press, 2006.
- [8] M. Ashikuzzaman, A. Héroux, A. Tang, G. Cloutier, and H. Rivaz, “Displacement tracking techniques in ultrasound elastography: From cross-correlation to deep learning,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2024.

- [9] A. K. Tehrani, M. Ashikuzzaman, and H. Rivaz, “Lateral strain imaging using self-supervised and physically inspired constraints in unsupervised regularized elastography,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1462–1471, 2022.
- [10] M. Mirzaei, A. Asif, M. Fortin, and H. Rivaz, “3d normalized cross-correlation for estimation of the displacement field in ultrasound elastography,” *Ultrasonics*, vol. 102, p. 106053, 2020.
- [11] S. Goudarzi, A. Asif, and H. Rivaz, “Ultrasound beamforming using mobilenetv2,” in *2020 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2020, pp. 1–4.
- [12] S. Goudarzi, A. Basarab, and H. Rivaz, “A unifying approach to inverse problems of ultrasound beamforming and deconvolution,” *IEEE Transactions on Computational Imaging*, vol. 9, pp. 197–209, 2023.
- [13] S. Goudarzi and H. Rivaz, “Deep reconstruction of high-quality ultrasound images from raw plane-wave data: A simulation and in vivo study,” *Ultrasonics*, vol. 125, p. 106778, 2022.
- [14] S. Goudarzi, A. Basarab, and H. Rivaz, “Inverse problem of ultrasound beamforming with denoising-based regularized solutions,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 10, pp. 2906–2916, 2022.
- [15] H. Asgariandehkordi, S. Goudarzi, M. Sharifzadeh, A. Basarab, and H. Rivaz, “Denoising plane wave ultrasound images using diffusion probabilistic models,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2024.
- [16] H. Asgariandehkordi, S. Goudarzi, A. Basarab, and H. Rivaz, “Deep ultrasound denoising using diffusion probabilistic models,” in *2023 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2023, pp. 1–4.
- [17] S. Goudarzi and H. Rivaz, “Deep ultrasound denoising without clean data,” in *Medical Imaging 2023: Ultrasonic Imaging and Tomography*, vol. 12470. SPIE, 2023, pp. 131–136.
- [18] S. Goudarzi, A. Asif, and H. Rivaz, “Fast multi-focus ultrasound image recovery using generative adversarial networks,” *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1272–1284, 2020.

- [19] —, “Multi-focus ultrasound imaging using generative adversarial networks,” in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE, 2019, pp. 1118–1121.
- [20] A. Nasiri-Sarvi, M. S. Hosseini, and H. Rivaz, “Vision mamba for classification of breast ultrasound images,” in *Deep Breast Workshop on AI and Imaging for Diagnostic and Treatment Challenges in Breast Care*. Springer, 2024, pp. 148–158.
- [21] B. Gheflati and H. Rivaz, “Vision transformers for classification of breast ultrasound images,” in *2022 44th annual international conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 480–483.
- [22] A. K. Tehrani, M. Amiri, I. M. Rosado-Mendez, T. J. Hall, and H. Rivaz, “Ultrasound scatterer density classification using convolutional neural networks and patch statistics,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 8, pp. 2697–2706, 2021.
- [23] M. Byra, G. Styczynski, C. Szmigielski, P. Kalinowski, Ł. Michałowski, R. Paluszkiwicz, B. Ziarkiewicz-Wróblewska, K. Zieniewicz, P. Sobieraj, and A. Nowicki, “Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images,” *International journal of computer assisted radiology and surgery*, vol. 13, pp. 1895–1903, 2018.
- [24] B. Behboodi, H. Rasaee, A. K. Tehrani, and H. Rivaz, “Deep classification of breast cancer in ultrasound images: more classes, better results with multi-task learning,” in *Medical Imaging 2021: Ultrasonic Imaging and Tomography*, vol. 11602. SPIE, 2021, pp. 170–175.
- [25] S. Goudarzi, J. Whyte, M. Boily, A. Towers, R. D. Kilgour, and H. Rivaz, “Segmentation of arm ultrasound images in breast cancer-related lymphedema: a database and deep learning algorithm,” *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 9, pp. 2552–2563, 2023.

- [26] M. Y. Ansari, Y. Yang, P. K. Meher, and S. P. Dakua, "Dense-ppsp-unet: A neural network for fast inference liver ultrasound segmentation," *Computers in biology and medicine*, vol. 153, p. 106478, 2023.
- [27] J. Chowdary, P. Yogarajah, P. Chaurasia, and V. Guruviah, "A multi-task learning framework for automated segmentation and classification of breast tumors from ultrasound images," *Ultrasonic imaging*, vol. 44, no. 1, pp. 3–12, 2022.
- [28] F. Marzola, N. Van Alfen, J. Doorduyn, and K. M. Meiburger, "Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment," *Computers in Biology and Medicine*, vol. 135, p. 104623, 2021.
- [29] M. Amiri, R. Brooks, B. Behboodi, and H. Rivaz, "Two-stage ultrasound image segmentation using u-net and test time augmentation," *International journal of computer assisted radiology and surgery*, vol. 15, no. 6, pp. 981–988, 2020.
- [30] M. Amiri, R. Brooks, and H. Rivaz, "Fine-tuning u-net for ultrasound image segmentation: different layers, different outcomes," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2510–2518, 2020.
- [31] —, "Fine tuning u-net for ultrasound image segmentation: which layers?" in *MICCAI Workshop on Domain Adaptation and Representation Transfer*. Springer, 2019, pp. 235–242.
- [32] B. Behboodi and H. Rivaz, "Ultrasound segmentation using u-net: learning from simulated data and testing on real data," in *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2019, pp. 6628–6631.
- [33] A. K. Tehrani and H. Rivaz, "Mpwc-net++: evolution of optical flow pyramidal convolutional neural network for ultrasound elastography," in *Medical Imaging 2021: Ultrasonic Imaging and Tomography*, vol. 11602. SPIE, 2021, pp. 14–23.

- [34] —, “Displacement estimation in ultrasound elastography using pyramidal convolutional neural network,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 67, no. 12, pp. 2629–2639, 2020.
- [35] H. Rivaz, E. M. Boctor, M. A. Choti, and G. D. Hager, “Real-time regularized ultrasound elastography,” *IEEE transactions on medical imaging*, vol. 30, no. 4, pp. 928–945, 2010.
- [36] H. Rivaz, E. Boctor, P. Foroughi, R. Zellars, G. Fichtinger, and G. Hager, “Ultrasound elastography: a dynamic programming approach,” *IEEE transactions on medical imaging*, vol. 27, no. 10, pp. 1373–1377, 2008.
- [37] N. Jafarpisheh, L. Castaneda-Martinez, H. Whitson, I. M. Rosado-Mendez, and H. Rivaz, “Physics-inspired regularized pulse-echo quantitative ultrasound: Efficient optimization with admm,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 70, no. 11, pp. 1428–1441, 2023.
- [38] N. Jafarpisheh, T. J. Hall, H. Rivaz, and I. M. Rosado-Mendez, “Analytic global regularized backscatter quantitative ultrasound,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 68, no. 5, pp. 1605–1617, 2020.
- [39] Z. Vajihi, I. M. Rosado-Mendez, T. J. Hall, and H. Rivaz, “Low variance estimation of backscatter quantitative ultrasound parameters using dynamic programming,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 65, no. 11, pp. 2042–2053, 2018.
- [40] H. Rivaz, R. Zellars, G. Hager, G. Fichtinger, and E. Boctor, “9c-1 beam steering approach for speckle characterization and out-of-plane motion estimation in real tissue,” in *2007 IEEE Ultrasonics Symposium Proceedings*. IEEE, 2007, pp. 781–784.
- [41] M. L. Oelze and J. Mamou, “Review of quantitative ultrasound: Envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 63, no. 2, pp. 336–351, 2016.

- [42] M. L. Oelze, J. F. Zachary, and W. D. O'Brien, "Differentiation of tumor types in vivo by scatterer property estimates and parametric images using ultrasound backscatter," in *IEEE Symposium on Ultrasonics, 2003*, vol. 1. IEEE, 2003, pp. 1014–1017.
- [43] J. Mamou and M. L. Oelze, *Quantitative ultrasound in soft tissues*. Springer, 2013.
- [44] A. Stewart and D. Reid, "Precision of quantitative ultrasound: comparison of three commercial scanners," *Bone*, vol. 27, no. 1, pp. 139–143, 2000.
- [45] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [46] B. Ghoshal and A. Tucker, "Deep learning for detecting covid-19 lung infections from chest ct images using uncertainty quantification and interpretable visualization," *Scientific Reports*, vol. 11, no. 1, pp. 1–16, 2021.
- [47] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, 2021.
- [48] H. Rasaee and H. Rivaz, "Explainable ai and susceptibility to adversarial attacks: A case study in classification of breast ultrasound images," in *2021 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2021, pp. 1–4.
- [49] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" *Advances in neural information processing systems*, vol. 30, 2017.
- [50] S. C. Hayes, K. Johansson, N. L. Stout, R. Prosnitz, J. M. Armer, S. Gabram, and K. H. Schmitz, "Upper-body morbidity after breast cancer: incidence and evidence for evaluation, prevention, and management within a prospective surveillance model of care," *Cancer*, vol. 118, no. S8, pp. 2237–2249, 2012.
- [51] M. R. Fu, "Breast cancer-related lymphedema: Symptoms, diagnosis, risk reduction, and management," *World journal of clinical oncology*, vol. 5, no. 3, p. 241, 2014.

- [52] S. Hayes, “Review of research evidence on secondary lymphoedema: incidence, prevention, risk factors and treatment. national breast and ovarian cancer centre,” *NSW*, 2008.
- [53] C. J. Moffatt, P. J. Franks, D. C. Doherty, A. Williams, C. Badger, E. Jeffs, N. Bosanquet, and P. Mortimer, “Lymphoedema: an underestimated health problem,” *Qjm*, vol. 96, no. 10, pp. 731–738, 2003.
- [54] D. Bojinović-Rodić, S. Popović-Petrović, S. Tomić, S. Markež, and D. Živanić, “Upper extremity function and quality of life in patients with breast cancer related lymphedema,” *Vojnosanitetski pregled*, vol. 73, no. 9, pp. 825–830, 2016.
- [55] J. Y. Park, J. Y. Jeon, and S. Cha, “Ultrasonographic features of the skin and subcutis: correlations with the severity of breast cancer–related lymphedema,” *Ultrasonography*, vol. 43, no. 4, p. 284, 2024.
- [56] C. Pirri, C. Ferraretto, N. Pirri, L. Bonaldo, R. De Caro, S. Masiero, and C. Stecco, “Ultrasound examination of skin, fasciae and subcutaneous tissue: Optimizing rehabilitation for secondary upper limb lymphedema,” *Diagnostics*, vol. 14, no. 24, p. 2824, 2024.
- [57] Y. Jeon, J. Beom, S. Ahn, and S.-K. Bok, “Ultrasonographic evaluation of breast cancer-related lymphedema,” *Journal of Visualized Experiments: JoVE*, no. 119, p. 54996, 2017.
- [58] K. Suehiro, N. Morikage, M. Murakami, O. Yamashita, M. Samura, and K. Hamano, “Significance of ultrasound examination of skin and subcutaneous tissue in secondary lower extremity lymphedema,” *Annals of vascular diseases*, vol. 6, no. 2, pp. 180–188, 2013.
- [59] E. Giray and İ. Yağcı, “Interrater and intrarater reliability of subcutaneous echogenicity grade and subcutaneous echo-free space grade in breast cancer-related lymphedema,” *Lymphatic research and biology*, vol. 17, no. 5, pp. 518–524, 2019.
- [60] L. Chen, Y. Wang, Z. Zhao, H. Liao, D. Zhang, H. Han, and F. Chen, “Hybrid-structure-oriented transformer for arm musculoskeletal ultrasound segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 621–631.

- [61] D. Ameri, A. K. Z. Tehrani, G. Cloutier, A. Tang, I. M. Rosado-Mendez, and H. Rivaz, “Bayesian approach with uncertainty quantification for quantitative ultrasound-based diagnosis of liver steatosis in a pre-clinical animal model,” 2026, under review in *SPIE Medical Imaging 2026*.
- [62] M. E. Rinella and S. Sookoian, “From nafld to masld: updated naming and diagnosis criteria for fatty liver disease,” *Journal of Lipid Research*, vol. 65, no. 1, 2024.
- [63] G. Vernon, A. Baranova, and Z. Younossi, “Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults,” *Alimentary pharmacology & therapeutics*, vol. 34, no. 3, pp. 274–285, 2011.
- [64] M. Gesnik, M. Bhatt, M.-H. R. Cardinal, F. Destrempe, L. Allard, B. N. Nguyen, T. Alquier, J.-F. Giroux, A. Tang, and G. Cloutier, “In vivo ultrafast quantitative ultrasound and shear wave elastography imaging on farm-raised duck livers during force feeding,” *Ultrasound in medicine & biology*, vol. 46, no. 7, pp. 1715–1726, 2020.
- [65] P. Vianna, S.-I. Calce, P. Boustros, C. Larocque-Rigney, L. Patry-Beaudoin, Y. H. Luo, E. Aslan, J. Marinos, T. M. Alamri, K.-N. Vu *et al.*, “Comparison of radiologists and deep learning for us grading of hepatic steatosis,” *Radiology*, vol. 309, no. 1, p. e230659, 2023.
- [66] H. Ma, C.-f. Xu, Z. Shen, C.-h. Yu, Y.-m. Li *et al.*, “Application of machine learning techniques for clinical predictive modeling: a cross-sectional study on nonalcoholic fatty liver disease in china,” *BioMed research international*, vol. 2018, 2018.
- [67] C. Xu, C. Yu, H. Ma, L. Xu, M. Miao, and Y. Li, “Prevalence and risk factors for the development of nonalcoholic fatty liver disease in a nonobese chinese population: the zhejiang zhenhai study,” *Official journal of the American College of Gastroenterology— ACG*, vol. 108, no. 8, pp. 1299–1304, 2013.
- [68] M.-G. Kim, S. Oh, Y. Kim, H. Kwon, and H.-M. Bae, “Learning-based attenuation quantification in abdominal ultrasound,” in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII* 24. Springer, 2021, pp. 14–23.

- [69] S. Oh, M.-G. Kim, Y. Kim, G. Jung, H. Kwon, and H.-M. Bae, “Sensor geometry generalization to untrained conditions in quantitative ultrasound imaging,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 780–789.
- [70] A. Tang, F. Destrempes, S. Kazemirad, J. Garcia-Duitama, B. N. Nguyen, and G. Cloutier, “Quantitative ultrasound and machine learning for assessment of steatohepatitis in a rat model,” *European radiology*, vol. 29, pp. 2175–2184, 2019.
- [71] T. Sharon, H. Naaman, Y. Eder, and Y. C. Eldar, “Real-time quantitative ultrasound and radar knowledge-based medical imaging,” in *Medical Imaging with Deep Learning, short paper track*, 2023.
- [72] A. Ozturk, V. Kumar, T. T. Pierce, Q. Li, M. Baikpour, I. Rosado-Mendez, M. Wang, P. Guo, S. Schoen Jr, Y. Gu *et al.*, “The future is beyond bright: The evolving role of quantitative us for fatty liver disease,” *Radiology*, vol. 309, no. 2, p. e223146, 2023.
- [73] G. Cloutier, F. Destrempes, F. Yu, and A. Tang, “Quantitative ultrasound imaging of soft biological tissues: a primer for radiologists and medical physicists,” *Insights into Imaging*, vol. 12, pp. 1–20, 2021.
- [74] L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun, “Hands-on bayesian neural networks—a tutorial for deep learning users,” *IEEE Computational Intelligence Magazine*, vol. 17, no. 2, pp. 29–48, 2022.
- [75] L. R. Chai, “Uncertainty estimation in bayesian neural networks and links to interpretability,” *Master’s Thesis, Massachusetts Institute of Technology*, 2018.
- [76] I. Bilonis and N. Zabarar, “Multi-output local gaussian process regression: Applications to uncertainty quantification,” *Journal of Computational Physics*, vol. 231, no. 17, pp. 5718–5746, 2012.

- [77] Y. Kwon, J.-H. Won, B. J. Kim, and M. C. Paik, “Uncertainty quantification using bayesian neural networks in classification: Application to biomedical image segmentation,” *Computational Statistics & Data Analysis*, vol. 142, p. 106816, 2020.
- [78] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher *et al.*, “A survey of uncertainty in deep neural networks,” *Artificial Intelligence Review*, vol. 56, no. Suppl 1, pp. 1513–1589, 2023.
- [79] A. K. Tehrani, G. Cloutier, A. Tang, I. M. Rosado-Mendez, and H. Rivaz, “Homodyned k-distribution parameter estimation in quantitative ultrasound: Autoencoder and bayesian neural network approaches,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2024.
- [80] D. Ameri, A. K. Tehrani, I. M. Rosado-Mendez, and H. Rivaz, “Uncertainty decomposition and error margin detection of homodyned-k distribution in quantitative ultrasound,” in *2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS)*. IEEE, 2024, pp. 1–4.
- [81] A. K. Tehrani, I. M. Rosado-Mendez, and H. Rivaz, “Homodyned k-distribution: parameter estimation and uncertainty quantification using bayesian neural networks,” in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023, pp. 1–4.
- [82] S. Depeweg, J.-M. Hernandez-Lobato, F. Doshi-Velez, and S. Udluft, “Decomposition of uncertainty in bayesian deep learning for efficient and risk-sensitive learning,” in *International Conference on Machine Learning*. PMLR, 2018, pp. 1184–1193.
- [83] D. W. Robinson, “Entropy and uncertainty,” *Entropy*, vol. 10, no. 4, pp. 493–506, 2008.
- [84] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [85] Z. Zhou, J. Fang, A. Cristea, Y.-H. Lin, Y.-W. Tsai, Y.-L. Wan, K.-M. Yeow, M.-C. Ho, and P.-H. Tsui, “Value of homodyned k distribution in ultrasound parametric imaging of hepatic steatosis: an animal study,” *Ultrasonics*, vol. 101, p. 106001, 2020.

- [86] M. Bhatt, L. Yazdani, F. Destrempes, L. Allard, B. N. Nguyen, A. Tang, and G. Cloutier, "Multiparametric in vivo ultrasound shear wave viscoelastography on farm-raised fatty duck livers: Human radiology imaging applied to food sciences," *Poultry Science*, vol. 100, no. 4, 2021.
- [87] J. Timaná, H. Chahuara, L. Basavarajappa, A. Basarab, K. Hoyt, and R. Lavarello, "Simultaneous imaging of ultrasonic relative backscatter and attenuation coefficients for quantitative liver steatosis assessment," *Scientific Reports*, vol. 13, no. 1, p. 8898, 2023.
- [88] S. Muhtadi, "Breast tumor classification using intratumoral quantitative ultrasound descriptors," *Computational and Mathematical Methods in Medicine*, vol. 2022, 2022.
- [89] A. Chowdhury, R. R. Razzaque, S. Muhtadi, A. Shafiullah, E. U. I. Abir, B. S. Garra, and S. K. Alam, "Ultrasound classification of breast masses using a comprehensive nakagami imaging and machine learning framework," *Ultrasonics*, vol. 124, p. 106744, 2022.
- [90] F. Destrempes, I. Trop, L. Allard, B. Chayer, J. Garcia-Duitama, M. El Khoury, L. Lalonde, and G. Cloutier, "Added value of quantitative ultrasound and machine learning in bi-rads 4–5 assessment of solid breast lesions," *Ultrasound in Medicine & Biology*, vol. 46, no. 2, pp. 436–444, 2020.
- [91] M. Khairalseed and K. Hoyt, "High-resolution ultrasound characterization of local scattering in cancer tissue," *Ultrasound in medicine & biology*, vol. 49, no. 4, pp. 951–960, 2023.
- [92] C. Hoerig, K. Wallace, M. Wu, and J. Mamou, "Classification of metastatic lymph nodes in vivo using quantitative ultrasound at clinical frequencies," *Ultrasound in Medicine & Biology*, vol. 49, no. 3, pp. 787–801, 2023.
- [93] H. Kim and T. Varghese, "Hybrid spectral domain method for attenuation slope estimation," *Ultrasound in medicine & biology*, vol. 34, no. 11, pp. 1808–1819, 2008.
- [94] M. Byra, A. Nowicki, H. Wróblewska-Piotrkowska, and K. Dobruch-Sobczak, "Classification of breast lesions using segmented quantitative ultrasound maps of homodyned k distribution parameters," *Medical physics*, vol. 43, no. 10, pp. 5561–5569, 2016.

- [95] F. Destrempes, J. Porée, and G. Cloutier, “Estimation method of the homodyned k-distribution based on the mean intensity and two log-moments,” *SIAM journal on imaging sciences*, vol. 6, no. 3, pp. 1499–1530, 2013.
- [96] D. P. Hruska and M. L. Oelze, “Improved parameter estimates based on the homodyned k distribution,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 56, no. 11, 2009.
- [97] Y. Liu, B. He, Y. Zhang, X. Lang, R. Yao, and L. Pan, “A study on a parameter estimator for the homodyned k distribution based on table search for ultrasound tissue characterization,” *Ultrasound in Medicine & Biology*, 2023.
- [98] A. K. Tehrani, I. M. Rosado-Mendez, H. Whitson, and H. Rivaz, “A deep learning approach for patchless estimation of ultrasound quantitative parametric image with uncertainty measurement,” in *Medical Imaging 2023: Ultrasonic Imaging and Tomography*, vol. 12470. SPIE, 2023.
- [99] Z. Zhou, A. Gao, W. Wu, D.-I. Tai, J.-H. Tseng, S. Wu, and P.-H. Tsui, “Parameter estimation of the homodyned k distribution based on an artificial neural network for ultrasound tissue characterization,” *Ultrasonics*, vol. 111, p. 106308, 2021.
- [100] A. K. Tehrani, I. M. Rosado-Mendez, and H. Rivaz, “Homodyned k-distribution: parameter estimation and uncertainty quantification using bayesian neural networks,” *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, <https://arxiv.org/abs/2211.00175>, 2023.
- [101] G. Ghoshal, R. J. Lavarello, J. P. Kemmerer, R. J. Miller, and M. L. Oelze, “Ex vivo study of quantitative ultrasound parameters in fatty rabbit livers,” *Ultrasound in medicine & biology*, vol. 38, no. 12, pp. 2238–2248, 2012.
- [102] D. P. Hruska and M. L. Oelze, “Improved parameter estimates based on the homodyned k distribution,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 56, no. 11, pp. 2471–2481, 2009.

- [103] K. Nam, I. M. Rosado-Mendez, L. A. Wirtzfeld, G. Ghoshal, A. D. Pawlicki, E. L. Madsen, R. J. Lavarello, M. L. Oelze, J. A. Zagzebski, W. D. O'Brien Jr *et al.*, "Comparison of ultrasound attenuation and backscatter estimates in layered tissue-mimicking phantoms among three clinical scanners," *Ultrasonic imaging*, vol. 34, no. 4, pp. 209–221, 2012.
- [104] I. M. Rosado-Mendez, "Advanced spectral analysis methods for quantification of coherent ultrasound scattering: Applications in the breast," Ph.D. dissertation, The University of Wisconsin-Madison, 2014.
- [105] L. H. Lee, E. Bradburn, A. T. Papageorghiou, and J. A. Noble, "Calibrated bayesian neural networks to estimate gestational age and its uncertainty on fetal brain ultrasound images," in *International Workshop on Advances in Simplifying Medical Ultrasound*. Springer, 2020, pp. 13–22.
- [106] D. Ameri, A. K. Z. Tehrani, B. Freiche, S. H. Meterissian, A. Towers, and H. Rivaz, "Point-of-care ultrasound in diagnosis of breast cancer-related lymphedema: A multimodal database and deep learning approach," 2026, under review in *IEEE Transactions on Biomedical Engineering*.
- [107] I. S. of Lymphology Executive Committee *et al.*, "The diagnosis and treatment of peripheral lymphedema: 2020 consensus document of the international society of lymphology. 2020."
- [108] T. Ashikaga, D. Burns, P. O'Brien, K. B. Schaberg, and D. Huston, "Texture analysis of post breast cancer lymphedema ultrasound images obtained using a portable device—a pilot study," *Lymphatic research and biology*, vol. 3, no. 3, pp. 147–155, 2005.
- [109] S. G. Rockson, V. Keeley, S. Kilbreath, A. Szuba, and A. Towers, "Cancer-associated secondary lymphoedema," *Nature reviews Disease primers*, vol. 5, no. 1, p. 22, 2019.
- [110] A. Szuba and S. G. Rockson, "Lymphedema: anatomy, physiology and pathogenesis," *Vascular medicine*, vol. 2, no. 4, pp. 321–326, 1997.

- [111] H. S. Hashemi, S. Fallone, M. Boily, A. Towers, R. D. Kilgour, and H. Rivaz, "Assessment of mechanical properties of tissue in breast cancer-related lymphedema using ultrasound elastography," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 66, no. 3, pp. 541–550, 2018.
- [112] C. S. Perez, C. Mestriner, L. T. Ribeiro, F. W. Grillo, T. W. Lemos, A. A. Carneiro, R. R. d. J. Guirro, and E. C. Guirro, "Relationship between lymphedema after breast cancer treatment and biophysical characteristics of the affected tissue," *PLoS One*, vol. 17, no. 4, p. e0264160, 2022.
- [113] J. T. Hidding, P. B. Viehoff, C. H. Beurskens, H. W. van Laarhoven, M. W. Nijhuis-van der Sanden, and P. J. van der Wees, "Measurement properties of instruments for measuring of lymphedema: systematic review," *Physical therapy*, vol. 96, no. 12, pp. 1965–1981, 2016.
- [114] C. Seward, M. Skolny, C. Brunelle, M. Asdourian, L. Salama, and A. G. Taghian, "A comprehensive review of bioimpedance spectroscopy as a diagnostic tool for the detection and measurement of breast cancer-related lymphedema," *Journal of surgical oncology*, vol. 114, no. 5, pp. 537–542, 2016.
- [115] R. Taylor, U. W. Jayasinghe, L. Koelmeyer, O. Ung, and J. Boyages, "Reliability and validity of arm volume measurements for assessment of lymphedema," *Physical therapy*, vol. 86, no. 2, pp. 205–214, 2006.
- [116] Y. Wang, X. Ge, H. Ma, S. Qi, G. Zhang, and Y. Yao, "Deep learning in medical ultrasound image analysis: a review," *Ieee Access*, vol. 9, pp. 54 310–54 324, 2021.
- [117] J. Schilcher, A. Nilsson, O. Andlid, and A. Eklund, "Fusion of electronic health records and radiographic images for a multimodal deep learning prediction model of atypical femur fractures," *Computers in Biology and Medicine*, vol. 168, p. 107704, 2024.
- [118] C. Cui, H. Yang, Y. Wang, S. Zhao, Z. Asad, L. A. Coburn, K. T. Wilson, B. A. Landman, and Y. Huo, "Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: a review," *Progress in Biomedical Engineering*, vol. 5, no. 2, p. 022001, 2023.

- [119] H. Rasaee, T. Koleilat, and H. Rivaz, “Groundingdino-us-sam: Text-prompted multi-organ segmentation in ultrasound with lora-tuned vision-language models,” *arXiv preprint arXiv:2506.23903*, 2025.
- [120] T. Koleilat, H. Asgariandehkordi, H. Rivaz, and Y. Xiao, “Medclip-sam: Bridging text and image towards universal medical image segmentation,” in *International conference on medical image computing and computer-assisted intervention*. Springer, 2024, pp. 643–653.
- [121] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, “Supervised contrastive learning,” *Advances in neural information processing systems*, vol. 33, pp. 18 661–18 673, 2020.
- [122] R. H. Mellor, N. L. Bush, A. W. Stanton, J. C. Bamber, J. R. Levick, and P. S. Mortimer, “Dual-frequency ultrasound examination of skin and subcutis thickness in breast cancer-related lymphedema,” *The breast journal*, vol. 10, no. 6, pp. 496–503, 2004.
- [123] E. Iker, C. K. Mayfield, D. J. Gould, and K. M. Patel, “Characterizing lower extremity lymphedema and lipedema with cutaneous ultrasonography and an objective computer-assisted measurement of dermal echogenicity,” *Lymphatic research and biology*, vol. 17, no. 5, pp. 525–530, 2019.
- [124] K. Niimi, M. Hirai, H. Iwata, and K. Miyazaki, “Ultrasonographic findings and the clinical results of treatment for lymphedema,” *Annals of vascular diseases*, vol. 7, no. 4, pp. 369–375, 2014.
- [125] P. Huang, S. Hu, B. Peng, J. Zhang, X. Wu, and X. Wang, “Robustly optimized deep feature decoupling network for fatty liver diseases detection,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 68–78.
- [126] Y.-M. Kim, M.-G. Kim, S.-H. Oh, G. Jung, H.-J. Lee, S.-Y. Kim, H.-S. Kwon, S.-I. Choi, and H.-M. Bae, “Quantitative assessment of thyroid nodules through ultrasound imaging analysis,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 711–720.

- [127] A. Menegola, M. Fornaciali, R. Pires, S. Avila, and E. Valle, “Towards automated melanoma screening: Exploring transfer learning schemes,” *arXiv preprint arXiv:1609.01228*, 2016.
- [128] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [129] J. F. Lazo, S. Moccia, E. Frontoni, and E. De Momi, “Comparison of different cnns for breast tumor classification from ultrasound images,” *arXiv preprint arXiv:2012.14517*, 2020.
- [130] J. Yang, X. Shi, B. Wang, W. Qiu, G. Tian, X. Wang, P. Wang, and J. Yang, “Ultrasound image classification of thyroid nodules based on deep learning,” *Frontiers in Oncology*, vol. 12, p. 905955, 2022.
- [131] X. Zhao, X. Shen, W. Wan, Y. Lu, S. Hu, R. Xiao, X. Du, and J. Li, “Automatic thyroid ultrasound image classification using feature fusion network,” *IEEE Access*, vol. 10, pp. 27 917–27 924, 2022.
- [132] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, “Signature verification using a” siamese” time delay neural network,” *Advances in neural information processing systems*, vol. 6, 1993.
- [133] I. Melekhov, J. Kannala, and E. Rahtu, “Siamese network features for image matching,” in *2016 23rd international conference on pattern recognition (ICPR)*. IEEE, 2016, pp. 378–383.
- [134] X. Chen and K. He, “Exploring simple siamese representation learning,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 15 750–15 758.
- [135] E. Solomon, A. Woubie, and E. S. Emiru, “Deep learning based face recognition method using siamese network,” *arXiv preprint arXiv:2312.14001*, 2023.
- [136] M. D. Li, K. Chang, B. Bearce, C. Y. Chang, A. J. Huang, J. P. Campbell, J. M. Brown, P. Singh, K. V. Hoebel, D. Erdoğmuş *et al.*, “Siamese neural networks for continuous disease

- severity evaluation and change detection in medical imaging,” *NPJ digital medicine*, vol. 3, no. 1, p. 48, 2020.
- [137] Z. Huang, Y. Wang, W. Xin, X. Lin, H. Li, H. Chen, Y. Lao, and X. Chen, “Application of attention-based siamese composite neural network in medical image recognition,” *arXiv preprint arXiv:2304.09783*, 2023.
- [138] A. Mahajan, J. Dormer, Q. Li, D. Chen, Z. Zhang, and B. Fei, “Siamese neural networks for the classification of high-dimensional radiomic features,” in *Proceedings of Spie—the International Society for Optical Engineering*, vol. 11314, 2020, p. 113143Q.
- [139] H. Chen, Y. Wang, K. Zheng, W. Li, C.-T. Chang, A. P. Harrison, J. Xiao, G. D. Hager, L. Lu, C.-H. Liao *et al.*, “Anatomy-aware siamese network: Exploiting semantic asymmetry for accurate pelvic fracture detection in x-ray images,” in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*. Springer, 2020, pp. 239–255.
- [140] Y.-A. Chung and W.-H. Weng, “Learning deep representations of medical images using siamese cnns with application to content-based image retrieval,” *arXiv preprint arXiv:1711.08490*, 2017.
- [141] T. Koleilat, H. Asgariandehkordi, H. Rivaz, and Y. Xiao, “Biomedcoop: Learning to prompt for biomedical vision-language models,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 14 766–14 776.
- [142] S.-C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren, “Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines,” *NPJ digital medicine*, vol. 3, no. 1, p. 136, 2020.
- [143] P. Hager, M. J. Menten, and D. Rueckert, “Best of both worlds: Multimodal contrastive learning with tabular and imaging data,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 23 924–23 935.

- [144] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [145] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, “Emerging properties in self-supervised vision transformers,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 9650–9660.
- [146] A. v. d. Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” *arXiv preprint arXiv:1807.03748*, 2018.
- [147] O. Henaff, “Data-efficient image recognition with contrastive predictive coding,” in *International conference on machine learning*. PMLR, 2020, pp. 4182–4192.
- [148] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *International conference on machine learning*. PmLR, 2020, pp. 1597–1607.
- [149] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, “Ensemble deep learning: A review,” *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, 2022.
- [150] L. v. d. Maaten and G. Hinton, “Visualizing data using t-sne,” *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [151] B. Wang, Q. Guo, J.-Y. Wang, Y. Yu, A.-J. Yi, X.-W. Cui, and C. F. Dietrich, “Ultrasound elastography for the evaluation of lymph nodes,” *Frontiers in Oncology*, vol. 11, p. 714660, 2021.
- [152] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, and M. Sugiyama, “Co-teaching: Robust training of deep neural networks with extremely noisy labels,” *Advances in neural information processing systems*, vol. 31, 2018.