Modelling Human Behaviour in Smart Buildings Through Explainable Label-Efficient Learning Methods

Naailah Mahamoodally

A Thesis

in

The Department

of

Concordia Institute for Information Systems Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Quality Systems Engineering) at

Concordia University

Montréal, Québec, Canada

December 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Naailah Mahamoodally**

Entitled: **Modelling Human Behaviour in Smart Buildings Through Explainable Label-Efficient Learning Methods**

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

Master of Applied Science (Quality Systems Engineering)

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

Signed by the Final Examining Committee:

_____ Chair and Examiner
Dr. Abdessamad Ben Hamza

_____ Examiner
Dr. Amin Hammad

_____ Supervisor
Dr. Manar Amayri

Approved by _____
Chun Wang, Chair
Department of Concordia Institute for Information Systems Engineering

_____ 2025

_____
Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Modelling Human Behaviour in Smart Buildings Through Explainable Label-Efficient Learning Methods

Naailah Mahamoodally

Rising energy demands have intensified the pursuit of intelligent building systems capable of adapting to occupant behavior to optimize energy consumption. Occupancy Estimation and Activity Recognition play a central role in enabling data-driven, energy-efficient control strategies that enhance sustainability while maintaining comfort. However, existing approaches often rely on large, labeled datasets collected from a single environment. When deployed in new settings, their performance degrades due to domain shifts caused by differences in sensor layouts, room configurations, and occupant behavior. Acquiring sufficient labeled data for each environment is costly, time-consuming, and raises privacy concerns. This thesis addresses these challenges by developing explainable, label-efficient domain adaptation frameworks that enable models to generalize across buildings. First, we introduce Imbalance–Structure Expansion Reduction (IMB-SER), an interpretable tree-based domain adaptation method that mitigates class imbalance. Second, we propose Uncertainty-aware Transfer via Augmentation for Decision Trees (UTA-DT), a source-free framework that augments target data with uncertainty measures from pretrained source model to achieve cross-domain robustness. Lastly, we develop a novel semi-supervised framework, the Semi-Supervised Mixture of Probabilistic Principal Component Analyzers (Semi-MPPCA), which leverages unlabeled data and is fine-tuned with minimal labeled samples to model human behavior effectively. Building on this foundation, we extend the model by integrating an unsupervised domain adaptation framework, collectively referred to as Probabilistic Hypothesis Anchored Domain Adaptation (PHADA). A comprehensive analysis is conducted under varying label availability scenarios to evaluate performance, robustness, and label efficiency. Together, these contributions advance the development of transparent, label-efficient, and privacy-preserving domain adaptation methods.

Acknowledgments

This thesis marks the culmination of a research endeavour that I could not have completed alone. I am deeply grateful to my supervisor, Dr. Manar Amayri, for her unwavering guidance, patience, and belief in my work. Her faith in my potential and constant encouragement inspired me to think deeply, challenge my own limits, and persevere through every obstacle.

To my co-authors and friends at Concordia University, thank you for the insightful discussions, constructive feedback, shared laughter, and constant encouragement. You made this experience both meaningful and memorable.

Above all, my deepest gratitude goes to my family for their unconditional love, strength, and faith in me. Their support has been my greatest source of inspiration and resilience throughout this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

The 21st century is characterized by rapid urbanization and digitalization, leading to an ever-increasing demand for energy. Within this landscape, the built environment stands as a primary contributor, accounting for approximately 32% of global energy consumption and a significant portion of greenhouse gas emissions [1–7]. Improving building energy efficiency has thus become a central pillar of global climate mitigation strategies [8–10]. The status quo of operating buildings, particularly in terms of heating, ventilation, and air conditioning (HVAC) and lighting, based on fixed schedules and maximum occupancy assumptions, is fundamentally unsustainable, leading to massive energy waste in partially occupied or unoccupied spaces [11–13].

In response to growing demands for energy-efficient building management, the paradigm of the Smart Building has emerged, driven by advances in the Internet of Things (IoT) [14–16]. Through extensive sensor networks, including motion detectors, door contact sensors, temperature and humidity sensors, smart plugs, and microphones, IoT-enabled environments continuously capture information about occupant interactions and environmental conditions. Within this context, two perceptual tasks have gained particular prominence: Occupancy Estimation (OE), which determines the presence and number of occupants, and Human Activity Recognition (HAR), which identifies the specific actions and behaviors taking place [17–21]. Numerous studies have demonstrated that

accurate OE and HAR can significantly optimize energy consumption by enabling intelligent control of HVAC and lighting systems based on real-time occupant behavior, thereby reducing waste while maintaining comfort [22, 23].

The synergy between these two capabilities is what unlocks profound energy-saving potential [19–21]. A building that knows not only that a room is occupied, but also what the occupants are doing, can transition from blunt, reactive control to precise, proactive automation. For instance, the HVAC system, which is often the largest energy consumer, can shift from conditioning entire floors on a 9-to-5 schedule to providing ”follow-me” climate control. It can reduce airflow and setback temperatures in a conference room after a meeting ends, or maintain minimal operation in an empty office zone, only resuming full power when occupancy is detected. Similarly, lighting can be finely tuned to actual needs based on user activity.

This research leverages non-visual IoT sensor data, strategically chosen over computer vision for two compelling reasons. First, it inherently mitigates privacy concerns associated with continuous video monitoring, making it more suitable for deployment in private spaces such as homes and offices. Second, the resulting data streams are typically lower-dimensional and less complex, facilitating the development of computationally lightweight and efficient models suitable for real-time inference on embedded or edge devices. Collectively, these sensors generate a continuous, multi-modal data stream that serves as a digital nervous system for the built environment, enabling intelligent systems to answer not only “Is anyone here, and how many?” but also “What are they doing?”

By leveraging IoT sensor data for real-time activity and occupancy awareness, smart buildings can move beyond limited efficiency to achieve true contextual conservation, eliminating waste without compromising occupant comfort and paving the way for a sustainable future.

1.2 Problem Statement

The practical implementation of data-driven models for activity recognition and occupancy estimation faces significant and interconnected challenges that hinder their widespread adoption in smart buildings. Conventional approaches rely on supervised learning algorithms, which require

large, meticulously labeled datasets for training [24]. However, in real-world settings, acquiring such labeled datasets is a major impediment; the process of annotating sensor readings with corresponding activities or occupant counts is prohibitively costly, time-consuming, and often invasive [19–21, 25–28], creating a fundamental data scarcity problem.

Even when a substantial labeled dataset is assembled for a specific building, models trained on this data frequently exhibit a critical failure to generalize when deployed in new environments [29, 30]. This phenomenon, known as domain shift [31–33], occurs due to the inherent variability between buildings, including differences in sensor types and placements, architectural layouts, and the unique behavioral patterns of different occupant groups. Consequently, a model that performs with high accuracy in one building may suffer a drastic performance decline in another, rendering it unreliable for scalable, real-world applications. Training a new model for each environment would be too costly, time-consuming and not practical [34, 35].

Beyond issues of data scarcity and generalization, the nature of smart building data itself poses substantial analytical challenges. IoT-enabled environments generate continuous, multi-modal streams that are inherently high-dimensional, heterogeneous, and temporally complex. These data capture a broad spectrum of environmental and occupancy-related variables, reflecting dynamic occupant behavior, fluctuating environmental conditions, and sensor noise. While such richness holds great promise for developing adaptive and accurate models, it also introduces significant obstacles in data processing and interpretation. The intricate temporal and non-linear relationships among features complicate the identification of meaningful patterns and demand learning architectures capable of modeling complex dependencies over time [36–38]. Additionally, smart building datasets often exhibit severe class imbalance, where frequent states dominate, while rare yet operationally critical events are underrepresented [19–21, 39]. This imbalance biases models toward the majority classes, limiting their reliability in real-world deployment. Finally, although deep learning models can effectively capture non-linearities, their black-box nature hinders transparency and interpretability, which are key requirements for building operators who must trust, validate, and act upon automated control decisions [40, 41].

Addressing these challenges requires a modeling approach that is both label-efficient and capable of generalizing across diverse environments, while capturing the inherent complexity and

non-linearity of smart building data and maintaining interpretability. Such an approach must effectively leverage information from previously studied environments to perform reliably in new settings with limited labels without extensive retraining. This thesis explores different Domain Adaptation (DA) techniques [42–48] as a promising solution. DA aims to enable models trained on one environment to adapt to new contexts with minimal labeled data, thereby improving robustness and scalability. By facilitating cross-domain generalization, DA provides a practical pathway toward developing data-efficient and transferable Occupancy Estimation and Activity Recognition models for real-world smart-building applications.

1.3 Theoretical background and related works

1.3.1 Theoretical Background - Domain Adaptation

According to previous studies [44, 49], a domain, denoted as $\mathcal{D}$, consists of three main components: an r-dimensional input or feature space $\mathcal{X}$, an output or label space $\mathcal{Y}$, and an associated probability distribution $\mathcal{P}(x, y)$, represented as $\mathcal{D} = \{\mathcal{X}, \mathcal{Y}, \mathcal{P}(x, y)\}$, where $x \in \mathcal{X}$ represents the feature vector, and $y \in \mathcal{Y}$. The marginal distribution, denoted as $P(X)$, captures the distribution of instances within the feature space $\mathcal{X}$, where $X = \{x_i \mid x_i \in \mathcal{X}, i = 1, \ldots, n\}$ [21]. The marginal distribution is denoted with $P(X)$, where X is an instance of the feature space $\mathcal{X}$, $X = \{x_i \mid x_i \in \mathcal{X}, i = 1, \ldots, n\}$ [21]. The source marginal distributions over Y is denoted by:

$$P_Y^S = \int p^S(x, y)\, dx \tag{1}$$

and the target marginal distributions over Y is denoted by

$$P_Y^T = \int p^T(x, y)\, dx \tag{2}$$

In the context of supervised transfer learning, we consider two domains: a source domain $D^S = (X^S, Y^S, p^S)$ and a target domain $D^T = (X^T, Y^T, p^T)$. Various transfer learning models impose different constraints on the relationship between these domains. Our focus lies specifically on Supervised Domain Adaptation, where both source and target tasks share identical feature and label

spaces, i.e., $\mathcal{X}^S, \mathcal{X}^T \subseteq \mathcal{X}$ and $\mathcal{Y} = \mathcal{Y}^S = \mathcal{Y}^T$. In this thesis, we assume $X^S = X^T$ and $Y^S = Y^T$, while acknowledging potential differences in the marginal distributions of the features between the domains. The objective of domain adaptation, as outlined in previous work [50], is to enhance the performance of the decision function f_T of the target task by leveraging knowledge from the source domain D^S.

In domain adaptation, the loss function is pivotal for effective model training across domains. It serves as a measure of the dissimilarity between the distributions of the source and target data, encapsulating the discrepancy that needs to be minimized during training. Given a loss function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}^+$, a source prediction function $f : \mathcal{X}^S \to \mathcal{Y}^S$, and a typically limited target training set S^T drawn i.i.d. from D^T, our goal is to learn a function [49] $f \in \mathcal{F}_T : \mathcal{X}^T \to \mathcal{Y}^T$ with low risk on the target domain given by the following equation:

$$R(f) = \mathbb{E}_{P_T}(x, y)\{\ell(f(x), y)\} \tag{3}$$

Optimizing the loss function enables the model to adapt to the target domain while leveraging knowledge from the source domain, improving generalization across domains. By optimizing the loss function, the model learns to align the representations of the source and target domains, enabling it to generalize well to unseen data from the target domain ensuring robust performance across domains.

Domain shift encompasses three primary categories: prior shift, covariate shift, and concept shift [44, 51]. Prior shift, or class imbalance, pertains to scenarios where posterior distributions are equivalent, $p_s(y|x) = p_t(y|x)$, yet prior distributions of classes differ between domains, $p_s(y) \neq p_t(y)$. Covariate shift occurs when marginal probability distributions differ, $p_s(x) \neq p_t(x)$, while conditional probability distributions remain constant across domains, $p_s(y|x) = p_t(y|x)$. Covariate shift is often attributed to factors such as sample selection bias and missing data. Concept shift, also referred to as data drift, arises when data distributions remain unchanged, $p_s(x) = p_t(x)$, while conditional distributions differ between domains, $p_s(y|x) \neq p_t(y|x)$. In this study, we assume uniform prior shifts across both domains and focus on resolving the Covariate shift problem.

When engaging in knowledge transfer from one domain to another, the two domains need to

exhibit a degree of dependence. This dependency is fundamental for the extraction of meaningful knowledge; otherwise, the insights garnered from the source domain may prove ineffectual in the target domain. Knowledge acquired from the source domain, if improperly applied, has the potential to detrimentally impact the performance of the target model, resulting in what is termed "negative transfer." Negative transfer [21] denotes the phenomenon wherein the transferred knowledge adversely affects the performance of the target model, rather than enhancing it as intended. without knowledge about the difference between source and target domains, negative transfer occurs when knowledge is transferred from highly irrelevant sources [52]. This phenomenon is particularly pronounced in the context of class imbalance [53]. In classification applications, class imbalance may arise from disparities in class conditional distributions, different error costs associated with different classes, or a combination of both factors concurrently. Such imbalance can significantly exacerbate the challenges of knowledge transfer, increasing the risk of the negative transfer and diminishing the effectiveness of transferred knowledge [54]. Therefore, addressing class imbalance becomes imperative in mitigating the risks associated with negative transfer and ensuring the success of knowledge transfer endeavors.

We have explored various methodologies for domain adaptation, each employing Decision Trees owing to their capacity to generate interpretable models through the clear If-else rules they produce. At the core of this methodology lie foundational concepts. Entropy, a measure of uncertainty or disorder in a dataset, is calculated using the formula:

$$H(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \tag{4}$$

where S represents a dataset, c denotes the number of classes, and p_i is the proportion of instances in class i in dataset S. Information Gain (IG) quantifies the reduction in entropy achieved by partitioning a dataset based on a particular attribute, indicating how well a feature separates the data into classes. The formula for IG is given as follows:

$$IG(S, A) = H(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \cdot H(S_v) \tag{5}$$

where $Values(A)$ are the possible values of attribute A, $|S_v|$ is the number of instances in dataset S with value v for attribute A, $|S|$ is the total number of instances in a dataset S and $H(S)$ is the entropy of dataset S_v, the subset of dataset S with value v for attribute A. Node Purity, on the other hand, evaluates the homogeneity of a node by measuring the distribution of classes within it. A node is considered pure if all instances belong to the same class. These concepts guide the Decision Tree algorithm in selecting the most informative features at each node, ultimately facilitating effective knowledge transfer across domains. Decision trees use various criteria to determine the best feature to split on at each node. Common splitting criteria include Gini impurity, Information Gain, and Variance reduction. Similar to entropy, Gini impurity measures the degree of impurity or uncertainty in a dataset. It is calculated differently but serves the same purpose as entropy in determining node purity. It is calculated as follows:

$$\text{Gini}(S) = 1 - \sum_{i=1}^{c} p_i^2 \tag{6}$$

1.3.2 Literature Review

Over the years, enhancing energy efficiency [55] in smart buildings has become a fundamental focus of energy policy initiatives. The primary objective is to optimize energy consumption and distribution, thereby minimizing the environmental footprint of buildings while improving the performance of HVAC systems and ensuring occupant comfort [56, 57]. In this context, Activity Recognition (AR) and Occupancy Estimation (OE) [25, 27, 58–60] have become essential tools for optimizing energy management. While early research focused on supervised methods, the challenges of obtaining labeled data have driven growing interest in semi-supervised learning and domain adaptation approaches.

Occupancy Estimation

Occupancy estimation can be defined as the quantifying of human presence within a designated area which can be done by using different types of smart building data such as ambient sensors [27, 34, 59, 61]. By analyzing data from ambient sensors - such as temperature, humidity, CO_2 levels, and light intensity - alongside other sources like motion detectors, Wi-Fi signals, and HVAC

system logs, occupancy estimation models can provide real-time insights into human presence. These insights enable smarter decision-making, such as adjusting heating, cooling, and lighting systems dynamically to improve energy efficiency and occupant comfort.

[27] presents a methodology for identifying the most informative sensors, such as motion, acoustic, and power consumption sensors, through information gain. Decision tree models are then used to estimate occupancy levels by generating interpretable rules with adjustable thresholds. [34] proposes IMB-SER SMOTETOMEK, a supervised and explainable domain adaptation technique derived from the Structure Expansion Reduction (SER) algorithm, specifically adapted for imbalanced occupancy datasets. This model transfers knowledge from a label-rich source domain to a label-scarce target domain using a wide array of sensors including CO_2, humidity, temperature, and acoustic pressure. [62] introduces a real-time occupancy estimator based on CO_2 measurements, which leverages existing HVAC infrastructure to avoid additional hardware costs. The proposed model is based on a variant of Extreme Learning Machine (ELM) known as Feature-Scaled ELM (FS-ELM), where the input-to-hidden layer weights are initialized using a scaled random matrix. [63] develops a scalable Hidden Markov Model (HMM)-based framework for occupancy detection that incorporates model selection criteria to determine the optimal number of hidden states, improving adaptability to occupancy patterns. [64] applies k-means clustering to air quality data from an air ventilation system to differentiate between occupied and unoccupied states for more efficient energy use. [65] proposes a hybrid statistical learning framework using non-intrusive sensors. The model combines interactive learning with a classifier based on a predictive distribution derived from a Generalized Dirichlet mixture model, optimized through Bayesian variational inference. [66] introduces DA-HOC, a semi-supervised domain adaptation model that estimates occupancy using CO_2 sensor data by transferring knowledge from a source to a target domain with limited labeled samples. [67] presents a Bluetooth-based occupancy detection system using a Deep Q-Network (DQN) reinforcement learning framework. The model interacts with IoT devices and uses experience replay and a target network to stabilize training and support adaptive decision-making. [68] describes a sensor fusion-based approach for estimating occupancy counts using Wi-Fi, CO_2, PIR sensors, and electricity load meters. The estimation is performed using linear regression and a two-layer

feedforward artificial neural network (ANN). [69] proposes a deep learning architecture that combines a convolutional neural network (CNN) for spatial feature extraction with a bidirectional long short-term memory (BiLSTM) network to capture temporal dependencies from advanced metering infrastructure (AMI) data. [70] develops a technique that models complex activity patterns using Wi-Fi signal variations and a modified generative adversarial network (GAN) with self-attention and deep residual convolution, enabling both realistic data synthesis and contextual understanding.

Activity Recognition

Activity Recognition (AR) refers to the automated identification and interpretation of physical actions or tasks, such as walking, sleeping, or cooking, performed by individuals, based on sensor data. This data may be acquired through IoT sensors, wearable sensors, or extracted from video frames or images [71–73]. As a smart building task, it aims to understand occupant behavior to optimize HVAC systems and increase health safety [56].

Computer vision has traditionally been a dominant approach for human activity recognition (HAR) in smart buildings. [74] proposes dynamically adjusting thermostat setpoints to preserve thermal comfort using HAR based on Recurrent Neural Networks (RNNs), which capture sequential patterns in visual data. [75] introduces a Deep Convolutional GAN (DCGAN)-based framework that leverages unsupervised pretraining on unlabeled video data, with the discriminator fine-tuned on labeled samples for activity classification. [76] combines a pre-trained CNN for feature extraction with a deep autoencoder (DAE) to model temporal action dynamics, followed by a quadratic SVM for classification.

However, increasing concerns over privacy have limited the use of camera-based systems, motivating a shift toward non-visual sensing. [77] proposes DeepSense, a framework that utilizes data from Wi-Fi-enabled IoT devices and employs AE-LRCN, a hybrid model integrating autoencoders, CNNs, and LSTMs to denoise inputs, extract spatial features, and capture temporal dependencies. [72] demonstrates that inertial sensor data processed through CNNs can effectively support HAR without relying on video input.

Several domain adaptation approaches have also emerged. [21] explores non-deep PCA-like and PCA-based techniques that construct shared domains for supervised transfer from labeled to

unlabeled settings, validated on smart building datasets [78]. [56] applies unsupervised adversarial domain adaptation (UADA), including Drop-to-Adapt (DTA) and Batch Spectral Penalization with Domain-Adversarial Neural Networks (BSP + DANN), to learn domain-invariant representations using adversarial objectives.

Wearable and ubiquitous device data have also been widely adopted. [79] introduces a hybrid CNN-LSTM model that extracts spatial and temporal features from smartphone sensors to classify activities such as walking, sitting, and stair use. [80] proposes a deep recurrent neural network (DRNN) using smartphone accelerometer, gyroscope, magnetometer, and Google Fit data for HAR. [81] presents a CNN-LSTM model trained on triaxial accelerometer and gyroscope data from wrist-worn sensors to capture both spatial and temporal characteristics of movement. [82] integrates Neural Structured Learning (NSL) with LSTM for body sensor-based activity recognition and applies Kernel-based Discriminant Analysis (KDA) to improve feature clustering and inter-class separation.

Semi-Supervised Learning

Semi-Supervised Learning (SSL) is a machine learning approach that uses both labeled and unlabeled data to improve model performance. By exploring patterns in unlabeled data, SSL reduces the need for large labeled datasets and can be derived from extensions of supervised or unsupervised learning algorithms [83].

Recent clustering-based SSL methods [84] combine supervised objectives with self-supervised clustering signals under a unified cross-entropy loss. However, they lack probabilistic interpretability and offer no mechanisms for dynamic cluster-class alignment. In particular, while the framework imposes class labels on cluster centroids, it lacks mechanisms to resolve incorrect or misaligned clusters that emerge during self-supervised pretraining. Without a means to refine or reinitialize cluster-class associations, performance may degrade under distribution shift or domain noise.

Several recent approaches attempt to address instability and bias in pseudo-labeling. Debiased Self-Training (DST) [85] mitigates confirmation bias by decoupling pseudo-label generation and utilization using two classifier heads and adversarial optimization. However, it remains sensitive to initial model calibration and relies on iterative pseudo-labeling. Similarly, Uncertainty-guided Collaborative Mean-Teacher (UCMT) [86] introduces uncertainty-aware region mixing and co-training,

but its masking mechanisms are complex and tailored to image domains, limiting generalizability and interpretability. Other methods, such as Monte Carlo dropout-based uncertainty estimation [87], also assume reliable uncertainty estimates, an assumption that may not hold early in training or under distribution shift.

For time-series data, TS-TFC [88] combines co-training across temporal and frequency domains with contrastive learning. However, it heavily depends on the quality of pseudo-labels generated via label propagation, which is often compromised by class imbalance and noise. Graph-based methods like FSSLAG [89] offer efficient label inference via anchor graphs, but suffer from sensitivity to hyperparameters, random anchor selection, and computational constraints at scale.

Building upon the limitations of traditional semi-supervised methods, such as confirmation bias in self-training, unrealistic view assumptions in co-training, and scalability issues in graph-based approaches, this research proposes Semi-MPPCA, a novel semi-supervised framework designed for energy-efficient building management. Semi-MPPCA leverages both limited labeled data and abundant unlabeled sensor data to enhance occupancy estimation and activity recognition while reducing labeling costs. Unlike iterative pseudo-labeling methods, it learns a stable latent representation through probabilistic clustering, followed by supervised fine-tuning. It also avoids the independence assumptions of co-training and the parameter sensitivity of graph-based methods. By integrating MPPCA into a differentiable neural architecture with a novel weight-swapping mechanism, Semi-MPPCA achieves interpretable, scalable, and robust performance in real-world smart building scenarios.

Domain Adaptation (DA)

Transfer Learning (TL) emerges as a specialized sub-field of Machine Learning, dedicated to reconciling disparities inherent across related domains [44, 90]. The principal goal is to facilitate a harmonious alignment between domains by transferring knowledge from a source domain, characterized by ample labeled data, to a target domain, where data availability is limited, thereby enabling the trained model to seamlessly generalize its applicability to the domain of interest [59]. Domain Adaption (DA) is a specialized case of TL where the source and target domains differ in terms of marginal data distributions but share identical label sets [91], i.e. the source and target tasks are

the same. The study by [21] presents five Transfer Learning techniques tailored for heterogeneous environments. The algorithm initially employs PCA to transform features, establishing known distributions, followed by distance calculation between features and a subsequent feature mapping algorithm. The models were tested on the Activity Recognition dataset obtained from PIR motion sensor, door/temperature sensor and light switch sensors, and also on the OE dataset collected from power consumption sensors, CO2 concentration sensors, humidity, temperature, door/window contact and acoustic pressure sensors. The study conclusively demonstrates that models utilizing SMOTE exhibit superior accuracy compared to other methods. [59] adapted clustering-based deep unsupervised domain adaptation (UDA) techniques, for estimating occupancy levels. The data has been collected from motion detectors, power consumption, and CO2 concentration sensors at Grenoble Institute of Technology work offices. Three methods, namely Structurally Regularized Deep Clustering (SRDC), Hybrid Structurally Regularized Deep Clustering (H-SRDC), and Distilled Discriminative Clustering for Domain Adaptation (DisClusterDA), have been modified and applied for this purpose reaching an accuracy of 99%. [19] compares unsupervised domain adaptation (UDA) methods for OE, with and without access to labeled source data. Six adapted approaches for UDA with labeled source data include domain separation networks (DSN), cluster alignment with a teacher (CAT), and CAT variants. Additionally, six adapted methods for UDA without labeled source data are examined, such as confidence score weighting adaptation using a joint model data structure (CoWA-JMDS). The models have been tested on AR and OE datasets. [20] investigates occupancy estimation techniques using trained source models for DA, a specialized type of TL, testing three UDA methods: SHOT, HoMM, and SFDA. These methods align data representations between source and target environments through deep learning approaches, achieving accuracy scores of 90%. [92] proposes a novel approach called Adversarial Discriminative Domain Adaptation (ADDA) which combines discriminative modeling, untied weight sharing, and a GAN loss, demonstrating superior effectiveness and simplicity compared to existing methods and achieving state-of-the-art results in unsupervised adaptation and cross-modality object classification tasks. [93] introduces Transfer Component Analysis (TCA) for DA, aiming to learn transfer components across domains in a reproducing kernel Hilbert space, preserving data properties and bringing domain distributions closer. Additionally, the extension of TCA into a semisupervised learning setting

enables encoding label information into transfer components, facilitating dimensionality reduction and improving DA performance. The model has been tested on cross-domain indoor WiFi localization and cross-domain text classification data. [94] proposes a novel deep-learning approach for DA. By incorporating standard layers and a gradient reversal layer, the approach fosters the emergence of discriminative and invariant features across domains, achieving superior performance in image classification tasks, including scenarios with significant domain shifts. [95] applied transfer learning using deep convolutional neural networks to detect cracks on masonry facades with limited data. The study utilized pre-trained CNN models optimized via fine-tuning on a custom dataset of 700 brickwork masonry facade images. The approach demonstrated high effectiveness in detecting cracks, achieving up to 100% accuracy and F1-score, particularly useful in scenarios with limited annotated data. [96] introduces a supervised Active Learning Domain Adapted (ALDA), leveraging the synergy between active learning and DA to facilitate knowledge transfer from a related source domain, demonstrating efficacy through two proposed variants, B-ALDA and O-ALDA.

Imbalanced Data

The imbalanced learning problem [97] addresses the challenges posed by underrepresented data and significant class distribution imbalances in learning algorithms. Tackling imbalanced datasets demand novel insights, algorithms, and tools to effectively extract meaningful information and knowledge representations from vast amounts of raw data. [98] combines predictive modeling using an inverted Dirichlet mixture (IDMM) with a proposed over-sampling approach to address both the small training data and imbalanced domain problems, demonstrated through extensive experiments on synthetic and real smart building sensor data. [99] presents two undersampling strategies utilizing clustering during data preprocessing, with the majority class represented by cluster centers or their nearest neighbors. [100] introduces DataBoost-IM, a novel approach combining boosting with data generation to enhance classifier performance on imbalanced datasets. [100] was evaluated across seventeen synthetic imbalanced datasets, demonstrating high predictive accuracy for both minority and majority classes without sacrificing performance for either. [101] introduces a method to enhance the AdaBoost algorithm by incorporating new weighted vote parameters informed by the classification accuracy of the positive class and the imbalanced index of the data. [102] presents

a novel undersampling technique for addressing data imbalance in real biological datasets. [102] focuses on selecting negative samples based on their proximity to positive samples, leveraging the idea that samples near the decision boundary are more likely to be misclassified. This approach calculates weighted Euclidean distances between negative and positive samples, then selects negative samples closest to positive ones to maintain class balance while minimizing misclassifications, particularly in the context of SVM classifiers.

Explainable AI (XAI)

XAI methodologies serve as indispensable tools in enhancing the transparency of AI models to human users, thereby increasing trust in the algorithms [103]. They achieve this by elucidating the rationales underlying the decision-making processes of AI models as well as identifying potential biases. [41] focuses on applying an interpretable unsupervised machine learning method, Explainable K-means Clustering (ExKMC), to occupancy estimation. By utilizing simple binary decision trees, ExKMC enhances the interpretability of clustering decisions, addressing previous challenges of complex interpretations and difficult labeling. The models were tested on an OE dataset obtained through a sensor network comprising power consumption sensors, CO2 concentration sensors, humidity sensors, temperature sensors, door and window contact sensors, and acoustic pressure sensors. [104] used XGBoost to predict suitable credit selection strategies, and applied SHapley Additive exPlanations (SHAP) to enhance the interpretability of the machine learning decisions. This approach aimed to simplify the credit selection process for architects, reducing the reliance on expert opinions by providing a more data-driven, transparent method. [105] introduces a training strategy that enhances the explainability of AI models for object deep classification (XDC) without sacrificing accuracy, achieved through periodic saliency-based feedback. Explanations are represented as saliency maps, aiding in disentangling domain-specific information and facilitating improved generalization across different domains while maintaining performance on the original domain. [106] introduces XSDA-Net, an explainable supervised domain adaptation framework leveraging case-based reasoning to clarify predictions by highlighting similar regions in source and target train images. [107] introduces DATSA, a Domain Adversarial Tangent Subspace Alignment

network, which aligns local approximations of manifolds across domains without relying on unreliable pseudo labels, addressing limitations of current joint adversarial domain adaptation networks providing explainable adaptations. [108] introduces explainable domain adaptation by offering two explanation mechanisms: one elucidating the alignment of source and target distributions in the model's latent space, and another detailing change in the decision boundary post-adaptation. [109] developed an explainable deep learning model, integrating building characteristics, geometry, and urban morphology with the SHapley Additive exPlanation algorithm into Light Gradient Boosting Machine to predict energy usage and greenhouse gas emissions of residential buildings. [110] used XAI tools like LIME and SHAP on energy-building datasets to assist in removing less important input features from ANN models without compromising the models's accuracy.

1.4 Contributions

The contributions of this thesis are summarized as follows:

1.4.1 Explainable Domain Adaptation for Imbalanced Occupancy Estimation

- Adapt decision-tree–based domain-adaptation methods for real-world occupancy estimation scenarios.
- Develop an imbalance-aware variant (IMB-SER) of the DA framework to improve minority-class performance while preserving overall accuracy.
- Integrate explainability mechanisms using rule comparison and SHAP analyses to ensure transparent and auditable behavior under distribution shift.
- Demonstrate consistent improvements over state-of-the-art tree-based domain-adaptation baselines, validating the effectiveness of imbalance-aware transfer learning for OE.

This research has been published in the Journal of Building Engineering [34].

1.4.2 Explainable Domain Adaptation without Source Data for Activity Recognition

- Develop a source-free domain-adaptation framework that operates without accessing source data, ensuring full privacy preservation.
- Augment target samples using soft and hard predictions from a pretrained source model to support effective cross-domain transfer.
- Integrate entropy and top-class margin as adaptation features to improve robustness under class-distribution and covariate shifts.
- Demonstrate strong cross-domain generalization using interpretable rule-based models suitable for real-world deployment.

This research has been published in the Building Simulation Conference 2025 [111].

1.4.3 Semi-Supervised Mixture of Probabilistic Principal Component Analyzers for Modeling Human Behavior

- Reformulate MPPCA as a neural architecture (NN-MPPCA) optimized through backpropagation rather than Expectation–Maximization.
- Develop a semi-supervised training pipeline enabling pretraining on unlabeled data and efficient few-label fine-tuning.
- Introduce a weight-swapping alignment strategy to correct cluster-label mismatches during supervised adaptation.
- Demonstrate strong label efficiency and improved performance across activity recognition and occupancy estimation tasks.

This research has been published in the Journal of Energy and Buildings [35].

1.4.4 Probabilistic Hypothesis Anchored Domain Adaptation for Modeling Human Behavior

- Extend Semi-MPPCA into a source-free unsupervised domain-adaptation framework for activity recognition and occupancy estimation.
- Transfer the source hypothesis using information maximization and pseudo-label refinement without requiring access to source-domain data.
- Introduce a component-anchor alignment mechanism to preserve mixture geometry and prevent feature drift during adaptation.
- Demonstrate superior performance under severe domain shift and validate the role of anchors through ablation studies.

This research has been submitted to the ACM Buildsys 2026 conference.

1.5 Thesis overview

This section outlines the structure and scope of the thesis and guides the reader to the relevant chapters.

- **Chapter 1: Introduction**

  Motivates the need for behavior-aware, energy-efficient building analytics, formulates the research problem, surveys the state of the art in occupancy estimation (OE), activity recognition (AR), domain adaptation, handling imbalance data and explainable AI, and concludes with a summary of the thesis contributions.

- **Chapter 2: Explainable Domain Adaptation for Imbalanced Occupancy Estimation**

  Introduces IMB-SER, an decision-tree based domain adaptation method adapted to handle imbalanced data. The chapter includes a comparative evaluation between our adapted method and state-of-the-art tree-based domain-adaptation models, and explainability analyses used to evaluate robustness under domain shift.

- **Chapter 3: Explainable Source-Free Domain Adaptation for Activity Recognition**

  Presents UTA-DT, a privacy-preserving domain-adaptation framework that operates without accessing the source data. The chapter describes the feature-augmentation strategy based on predictions and uncertainty cues, followed by cross-domain evaluation and rule-based interpretability analysis.

- **Chapter 4: Semi-Supervised Mixture of Probabilistic Principal Component Analyzers for Modeling Human Behavior**

  Introduces the Semi-MPPCA framework, including the NN-MPPCA architecture finetuned with very limited labeled data in a semi-supervised manner, and a novel cluster–label alignment mechanism. The chapter evaluates the method across OE and AR tasks, demonstrating label efficiency and strong generalization.

- **Chapter 5: Probabilistic Hypothesis-Anchored Domain Adaptation**

  Extends Semi-MPPCA into PHADA, a source-free unsupervised domain-adaptation framework. The chapter presents the adapted source hypothesis transfer strategy with an innovative component-anchor alignment mechanism, extensive experiments under severe domain shift, and evaluates performance against state-of-the-art unsupervised models and ablation studies.

- **Chapter 6: Conclusion**

  Summarizes the main findings, reflects on limitations, and outlines future research directions for explainable, label-efficient, and privacy-preserving domain adaptation in smart-building environments.

Chapter 2

Explainable Domain Adaptation for Imbalanced Occupancy Estimation

The challenges outlined in Chapter 1 highlight the need for domain-adaptation methods that are not only effective but also interpretable and aligned with real deployment constraints in smart buildings. In particular, occupancy estimation systems often suffer from pronounced class imbalance and distribution shift between buildings, which degrade the performance of conventional supervised models. Moreover, most existing domain-adaptation methods for smart-building analytics rely on opaque, black-box models that offer limited interpretability and poor suitability for real deployment. To address these issues, this chapter proposes an imbalance-aware, explainable domain-adaptation framework based on decision trees, validated against state-of-the-art tree-based domain adaptation models.

2.1 Introduction

Today, there is an increasing emphasis on energy management, driven by the rising demand for energy, especially within the building industry. Globally, buildings represent a significant portion of total energy consumption, surpassing the combined energy footprint of all transportation modes [112–114]. They also contribute significantly, accounting for around 30% of greenhouse gas emissions [115–117]. Given the escalating energy expenses [41] and the escalating impacts

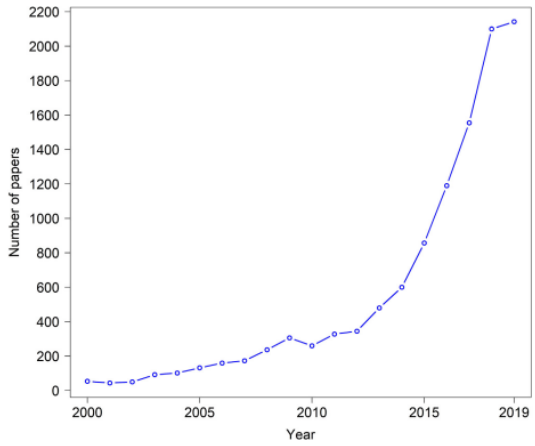


Figure 2.1: Number of studies that used machine learning in the building sector over the past twenty years [1]

of global climate change, enhancing the energy efficiency of buildings has emerged as a critical imperative. The advent of IoT has propelled the concept of smart buildings to the forefront, leveraging sensor technology to monitor and optimize energy usage [118]. This surge in interest has catalyzed a proliferation of research endeavors, notably since 2012, investigating the application of machine learning methodologies within the realm of buildings as shown in Figure 1.1 [1]. Recently researchers have started to focus more on occupancy estimation as it is very helpful in energy efficiency systems [21, 119, 120]. However, challenges persist in accurately estimating occupancy levels, as conventional methods such as sensor-based or camera imaging techniques face limitations concerning communication bandwidth and privacy concerns and counting gates are expensive to deploy at room level inside a whole building[112, 121]. Consequently, the precise and efficient estimation of occupancy emerges as a pivotal challenge in the realization of smart buildings.

Most of the previous research employs conventional supervised machine learning algorithms [21] and has a common assumption: both the training and testing datasets are sourced from identical environments and follow the same distributions [31, 44]. Nonetheless, within the field of occupancy estimation, a model trained within a particular room may exhibit suboptimal performance when applied to another room within the same building, attributable to disparities in environmental conditions. For instance, variations in window count between rooms could influence the light intensity or the spatial configuration of the sensors across distinct room environments may exert unintended

effects on the acquired data. This phenomenon is commonly referred to as domain shift [122]. The ideal course of action entails the acquisition of new data. However, procuring labeled data often proves to be time-consuming, costly, or infeasible [122]. Consequently, we should devise methodologies that leverage existing knowledge from analogous model domains [21]. In such contexts, Domain Adaptation (DA) emerges as a valuable tool for mitigating these challenges. DA ensures that the model can generalize across different environmental conditions and accurately estimate occupancy levels in various building rooms. Another issue is a severe class imbalance in occupancy estimation datasets[54, 123], which detrimentally impacts the model performance. Class imbalance occurs when one class greatly outweighs the other, leading to skewed predictive behavior. In occupancy estimation, where the dominant label represents the absence of occupancy due to its sporadic nature, the model may be biased towards predicting 'absence' in all instances, hindering its accuracy in capturing occupancy variations. Recent research on DA has been predominantly centered on deep learning [120, 124]. Nonetheless, there is a growing recognition of the dangers associated with black box models within the research community [125]. Their opacity makes it challenging to interpret their decisions and understand the underlying mechanisms driving their predictions. This lack of transparency can engender various issues including biased results, untraceable decision-making processes, and obstacles in identifying and rectifying errors or biases.

In this study, we aim to solve the mentioned problems through the adoption of DA methodologies utilizing Decision Trees (DT). This choice is motivated by DTs' capacity to generate interpretable rules, while also demonstrating comparable, if not superior, accuracy compared to deep learning models [121]. Notably, DT models are extensively optimized, evidenced by their sustained performance over the past four decades despite continuous efforts to refine them [126]. Furthermore, DTs exhibit rapid training and testing times, [127], further bolstering their appeal.

We conducted a comparative analysis of DA approaches using decision trees. Initially, we implemented the SER algorithm, which applies expansion and reduction to adapt a decision tree pre-trained on source data to target data, along with its variant SER*, aimed at enhancing minority class representation [54]. We also explored STRUT, which utilizes divergence gain for transfer, and its extension STRUT* for imbalanced data [54]. Additionally, we evaluated GOSDT-DA [126], a framework optimized for class imbalance, and SER-STRUT [49], a random forest method that

leverages divergence measures such as Jensen-Shannon and Kullback-Leibler. To address class imbalances more effectively, we incorporated resampling techniques including SMOTE, ClusterCentroids, and SMOTETOMEK into our models. These methods were benchmarked against a baseline model using a decision tree trained on limited target domain data, providing a foundation for evaluating the effectiveness of our DA strategies.

In this chapter, we made several contributions:

- Adaptation of existing methodologies to effectively handle datasets about Occupancy Estimation.
- Modification of the structure of the DA Decision Tree architecture to accommodate imbalanced datasets.
- Integration of Explainable AI techniques to elucidate the model's behavior on test data and to evaluate similarity.
- Attainment of favorable outcomes, surpassing the accuracy of models without domain adaptation.

The subsequent sections of this chapter are structured into four distinct parts. Section II is dedicated to conducting a comprehensive literature review. In Section III, we delve into a detailed discourse on the algorithms of the different approaches under consideration and we discuss the experimental setups and resultant findings in Section IV. Section V encapsulates the conclusions drawn from our study, as well as an outline of potential future research directions.

2.2 Considered methods

Occupancy information, particularly occupancy detection, and estimation, is instrumental in mitigating building energy consumption [21, 128]. By leveraging this insight, smart buildings can strategically manage energy usage by deactivating electrical loads in vacated areas [129], thereby optimizing energy consumption dynamics [59]. Furthermore, smart buildings can reduce the peak energy demand by tailoring the HVAC energy distribution of individual zones based on occupancy

[68]. For example, areas with higher occupancy generating more heat may require more cooling, while areas with fewer occupants or lower thermal loads may need less cooling [128]. Despite the potential benefits of leveraging historical smart building data for occupancy prediction, a notable limitation arises from its sparse availability [129]. Consequently, our objective is to establish a minimally intrusive method for occupancy monitoring, such as leveraging existing building infrastructure sensor data, while concurrently refining techniques for enhancing OE accuracy through knowledge transfer from a comprehensive labeled source model. Subsequently, smart buildings can dynamically adjust heating, lighting, ventilation, and HVAC system settings to optimize energy distribution and consumption [59].

In this section, we provide a comprehensive overview of the selected methodologies, including SER, SER* , STRUT, STRUT*, SER-STRUT [49, 54] and GOSDT [126]. We also present our modified version of SER: IMB-SER variants. Alterations have been made to the algorithms to accommodate imbalanced datasets, and efforts have been made to elucidate the models' behavior through eXplainable Artificial Intelligence (XAI) techniques.

2.2.1 Structure Expansion/Reduction (SER)

The Structure Expansion/Reduction (SER) algorithm entails two iterative procedures aimed at adjusting a decision tree (DT) to better fit target data. Initially, a DT is trained on the source data, denoted as S^S. The expansion phase involves refining the DT rules to accommodate the characteristics of the target data, S^T. Specifically, each leaf node of the DT is expanded into a subtree using the subset of the target data that corresponds to that leaf. This subtree construction utilizes the CART algorithm on the relevant target data. Subsequently, the reduction phase assesses each internal node of the DT for potential simplification and to avoid overfitting. Here, the leaf error and subtree error metrics are computed. The leaf error signifies the classification accuracy if a leaf were to be pruned from the tree, while the subtree error quantifies the classification error of the subtree rooted at a given internal node. If the leaf error is lower than the subtree error, indicating a more accurate classification, the subtree is pruned into a leaf node. This iterative process iterates recursively through the DT, progressively refining its structure to better align with the target data.

The SER algorithm is structured to prioritize expansions before reductions, a design choice

that effectively serves as a form of regularization. This approach maintains the resulting target model closer to the source model than it would if reductions were conducted prior to expansions. A decision tree can be represented as a disjunctive normal form (DNF) formula, where each rule corresponds to a conjunction of literals along a root-to-leaf path. Upon applying SER, if a path u_0 is generated from another path u through expansion, or reduction, we denote the resulting rule as T_T, and the original rule as T_S. If T_T expands T_S, then T_T satisfies T_S, meaning any instance satisfying T_T also satisfies T_S. Conversely, if T_T reduces T_S, then T_S satisfies T_T. This order of operations ensures that for any transformed rule T_T, there exists an original rule T_S, such that either T_T satisfies T_S or vice versa. This property is particularly advantageous for localized refinements or generalizations, fostering consistency with the original model. Conversely, conducting reductions before expansions may lead to divergence from the source model. The framework is suitable for both binary and multi-class classification tasks. The algorithm's pseudo-code is outlined in Algorithm 1.

Algorithm 1 SER

INPUT: Node v, Labeled samples S_v
OUTPUT: Node v
Procedure SER (v, S_v) :

```
% Expansion of Leaves
if v is a leaf then
    v ← BUILDTREE(S_v)
    return v
end if
% Recursive calls over child nodes
SER(v_r, S_{v_r})
SER(v_l, S_{v_l})
% Reduction of current node
if LeafError(v, S_v) ≤ subtreeError(v, S_v) then
    v ← PRUNE(v)
    depth(v) ← 0
    Y(v) ← (arg max_y |{(., Y) ∈ S_v}|)
end if
return v
```

2.2.2 SER*

SER* is an advanced modification of the Structured Expansion Reduction (SER) algorithm, tailored specifically to address challenges associated with leaf loss in scenarios where target data for

a class is sparse, potentially leading to unfavorable pruning outcomes. The SER* methodology enhances the conventional SER framework by focusing on the reduction step, crucial in DA processes. This modified approach strategically avoids pruning any source leaf that represents a minority class, thereby ensuring their retention in the model developed from the DA process. Conversely, it permits the expansion phase to adjust to any changes in the majority class as determined by the target data. This deliberate preservation of minority class leaves counters the inherent bias towards the majority class typically induced by imbalanced datasets. By implementing this strategy, SER* aims to balance the scales, potentially compensating for one bias with another, to enhance the fairness and effectiveness of the model in diverse application scenarios. This nuanced approach to handling class imbalances during the DA process is crucial for developing robust predictive models that maintain high interpretability and reliability in practical applications.

2.2.3 Structure Transfer (STRUT)

The Structure Transfer (STRUT) algorithm operates in a top-down manner, traversing the decision tree from the root to the leaves. At each node, the algorithm assesses whether the node is reachable by examining the target data that reaches it. If a node is unreachable, it is pruned into a leaf; otherwise, the algorithm recalculates the threshold for that node based on the target data. The threshold selection process is based on an optimization problem aiming to maximize the divergence gain (DG), which measures the similarity between previous class distributions while ensuring a local maximum of the information gain (IG), measured by the Gini gain. This procedure involves identifying a new threshold for a node using a set of ordered distinct feature values of instances of the target data. The selected threshold maximizes the DG, subject to constraints that enforce a local maximum of the IG. The pseudo-code for STRUT is presented in Algorithm 2. STRUT accounts for label swaps between the source and target domains during threshold selection, ensuring flexibility in adapting to changes in feature meaning. The algorithm's design is motivated by the observation that decision trees for similar problems should exhibit structural similarity, despite potential differences in feature scales and thresholds. By adapting the decision tree trained on source samples to target samples, STRUT captures possible transformations to maintain consistency with the original model

while accommodating drifts and label swaps in the target domain. To achieve this, STRUT formulates threshold selection as an optimization problem, maximizing DG while ensuring informative thresholds that align with the original distributions calculated during tree induction.

Algorithm 2 STRUT

Procedure STRUT(v, S_v) :

1: **if** $|S_v| = 0$ **then**
2: $v \leftarrow$ PRUNE(v) ▷ Return pruned node
3: **return** v
4: **end if**
5: **if** v is a leaf **then**
6: $v \leftarrow$ CHANGELEAFVALUE(S_v) ▷ Update leaf value
7: **return** v
8: **else**
9: $v \leftarrow$ THRESHOLDSELECTION($\phi_v, S_v, Q^l_{S_v}, Q^r_{S_v}$)
10: STRUT(v_r, S_{v_r}) ▷ Recursive call for right child
11: STRUT(v_l, S_{v_l}) ▷ Recursive call for left child
12: **end if**
13: **return** v

2.2.4 STRUT*

STRUT*, an extension of the STRUT algorithm, addresses the challenges posed by changes in class ratios that are not adequately handled by the divergence gain (DG) used in the original STRUT method. To overcome this limitation, STRUT* incorporates an approach where the optimization problem bypasses the use of DG. Instead, it updates thresholds by recalculating the maximum Gini gain on the target data while retaining the same features. Furthermore, STRUT* is designed to effectively manage homogeneous class imbalance transformations. It modifies the divergence optimization process by replacing terms with a formulation that acknowledges class-specific adjustments. Specifically, for each class k, the modified quantities $Q^l_{S_v}$ and $Q^r_{S_v}$ are recalculated to reflect the weighted contributions of classes based on their presence in the target data relative to the source. This adjustment ensures that if the class proportions are preserved between the source and target datasets, the algorithm simplifies back to the original STRUT configuration, illustrating that STRUT* is a generalization that enhances STRUT's applicability to more complex and varied data distributions. This strategic extension allows STRUT* to maintain the robustness of

the model under varying class imbalances, promoting greater adaptability and effectiveness in DA applications.

2.2.5 SERSTRUT

[49] introduced a mixture model aimed at harnessing the complementary strengths of both the Structure Expansion and Reduction (SER) and Structure Transfer (STRUT) algorithms. Specifically, [49] construct two separate forests using SER and STRUT, respectively, leveraging their distinct strategies for transferring knowledge across domains. Subsequently, [49] defines a mixture model, denoted as SER-STRUT, which serves as a voting ensemble mechanism. The underlying model of SER-STRUT comprises the union of all decision trees generated by both SER and STRUT. SER-STRUT employs a straightforward yet effective strategy of majority voting, aggregating predictions from all transferred decision trees. This ensemble approach capitalizes on the diverse perspectives offered by SER and STRUT, effectively integrating their outputs to enhance the robustness and generalization capability of the final model.

2.2.6 Simple GOSDT-DA (GOSDT)

The optimization of the Generalized Optimal Sparse Decision Tree (GOSDT) model, which aligns closely with Optimal Sparse Decision Trees (OSDT), is proposed by [126]. Conversely, for additive loss functions, the GOSDT model, incorporating dynamic programming with bounds (DPB), demonstrates significant computational efficiency gains. Like conventional dynamic programming approaches, GOSDT decomposes problems into smaller subproblems, which are either recursively solved or processed in parallel. GOSDT employs two central data structures: a priority queue for managing problem scheduling and a dependency graph for tracking inter-problem dependencies. Each problem is classified as either a left or right branch, denoted as p_j^l or p_j^r, respectively, reflecting its association with a specific feature j.

Furthermore, GOSDT introduces the Incremental Similar Support Bound (ISSB), a novel bound that quantifies the discrepancy in support between decision trees with similar root nodes. ISSB enables comparison between child trees grown from analogous branches, bounding the difference in their objective values. Mathematically, ISSB involves computing the maximum support within

the split nodes of two trees and bounding the absolute difference in their objective values based on this support. For trees d and D, differing only by their root nodes, such that the support going to their left and right branches varies by at most ω fraction of observations, $S_{\text{uncertain}}$, defined as $\max(\text{supp}(dsplit), \text{supp}(Dsplit))$, plays a crucial role. The absolute difference between objective values of any child trees d_0 and D_0 grown from d and D, respectively, is bounded by $(\omega+2S_{\text{uncertain}})\cdot \max$. Unlike similar support bounds in other algorithms, ISSB arises from the shared support sets of descendants $\sigma(d)$ and $\sigma(D)$, allowing simultaneous updates of their upper and lower bounds. This bound is particularly advantageous for data with continuous variables, facilitating reuse of prior computations to quickly produce tight upper and lower bounds.

$$|R(d_0, x, y) - R(D_0, x, y)| \leq (\omega + 2S_{\text{uncertain}}) \cdot \max \tag{7}$$

GOSDT leverages a dependency graph framework. At the core of GOSDT lies the concept of asynchronous bound updates, where objective values are computed independently and asynchronously, allowing for efficient propagation of updates through the graph. This approach accelerates decision-making by prioritizing high-priority updates, ensuring that bounds are promptly recalculated and propagated to ancestor nodes. Mathematically, this process involves computing the minimum objective value for each problem in the graph hierarchically, aggregating minimum objectives over child problems to derive the optimal tree structure as follows:

$$R^*(p) = \min_j (R^*(p_{jl}) + R^*(p_{jr})) \tag{8}$$

In addition, GOSDT incorporates fast selective vector sums to expedite computations of subset sums within vectors. By precomputing prefix sum vectors and leveraging the prefix sum trick during calculations, GOSDT significantly accelerates decision tree construction and evaluation. This technique optimizes operations by converting linear time calculations into constant time operations, enhancing the efficiency of the algorithm. The utilization of fast selective vector sums facilitates rapid assessment of support sets and bounds, streamlining the optimization process. The formula for the upper bound is as follows:

$$ub = \lambda + \min\left(\frac{\sum_u s_{uz}^{-u}}{\sum_u s_{uz}^{+u}}\right) \tag{9}$$

2.2.7 IMB-SER: our variant of the SER algorithm to deal with imbalance

While the SER algorithm exhibited commendable performance overall, it remained susceptible to producing suboptimal results, particularly concerning imbalanced classes. SER* did not provide improved results on our real-world OE dataset. In an effort to enhance its efficacy, we embarked on a systematic exploration of various imbalance techniques. These included oversampling, undersampling, and combined over-undersampling methods, namely SMOTE, ClusterCentroids, and SMOTETomek, to better address the disparity in class distribution:

- SMOTE (Synthetic Minority Over-sampling Technique) enhances the minority class representation by generating synthetic samples. It operates by selecting instances that are close within the feature space, drawing lines between these instances, and synthesizing new samples along these lines. This approach helps to bolster the minority class presence effectively without significant information loss.

- ClusterCentroids reduces the size of the majority class by creating representative centroids. It applies a clustering algorithm, such as k-means, to the majority class instances and uses the centroids of these clusters as the new representatives, thereby achieving a more balanced dataset through undersampling.

- SMOTETomek integrates the SMOTE technique with Tomek links to refine class balance further. While SMOTE augments the minority class with synthetic instances, Tomek links are used to remove overlapping samples between classes. This dual strategy enhances the classifier's decision boundaries and overall performance in handling minority classes.

Subsequently, we conducted a comprehensive comparative analysis to evaluate the effectiveness of these techniques in mitigating class imbalances. Furthermore, we introduced a different approach to post-pruning the decision trees, instead of the conventional leaf error pruning method employed in the original SER algorithm. Our proposed pruning strategy leverages cost-complexity pruning, a

technique renowned for its ability to optimize decision tree structures by controlling model complexity. Through these methodological refinements, we aim to bolster the robustness and performance of the SER algorithm, particularly in scenarios characterized by imbalanced class distributions.

Cost complexity pruning, a key component of decision tree optimization, involves adjusting the tree structure based on the cost complexity parameter, denoted as α. This parameter controls the trade-off between model complexity and accuracy by penalizing the addition of nodes to the tree. The path variable obtained from α yields two essential elements: α, representing a range of minimum leaf values for the decision tree, and impurities, indicating the impurity levels at each decision tree node. These values guide the selection of the optimal α parameter, which influences the creation of different decision tree classifiers. By iterating over various α values and training decision tree classifiers with different complexities, we can construct a collection of models. Ultimately, the α parameter helps in determining the appropriate level of tree complexity, thereby enhancing the model's generalization performance and mitigating the risk of overfitting.

In response to the challenges posed by the real-world occupancy estimation dataset, our enhancements to the SER method - including the integration of class imbalance strategies and a revised pruning technique - were designed to improve model performance. Furthermore, to increase the interpretability of our adapted models and elucidate their behavior on unseen data, we incorporated the SHAP (Shapley Additive exPlanations) methodology. This integration allows for a detailed analysis of feature contributions, thereby ensuring that the models are both effective and transparent, which is crucial for their application in smart building environments.

2.3 Experimental setup and results

2.3.1 Research workflow

In the domain of academic research, adhering to a structured workflow is crucial to the successful execution of scholarly endeavors. Such a workflow encompasses a series of methodical steps, beginning with the identification and articulation of the research problem. Through an extensive review of the literature, we embarked on a comprehensive exploration of existing knowledge, identifying pertinent gaps within the domain of imbalanced Occupancy Estimation, based

on domain adaptation. Subsequently, meticulous attention was devoted to method selection, ensuring alignment with the research objectives, scope, and the complexities inherent in the datasets. Rigorous preprocessing techniques were then applied to our Occupancy Estimation (OE) datasets, encompassing data cleansing, re-structuring, and refinement, supplemented by feature engineering methodologies. Following this preparatory phase, we performed data exploration and analysis, aimed at unraveling underlying patterns and insights embedded within the datasets. We have then adapted the selected methodologies to fit our datasets and enhancements were pursued iteratively to optimize model performance. The adapted models were evaluated, and augmented by endeavors to elucidate the underlying decision-making mechanisms through Explainable Artificial Intelligence (XAI). Finally, comprehensive discussions of the findings ensued, accompanied by the identification of avenues for future work endeavors. These methodological stages are summarized in Figure 2 for clarity and reference.

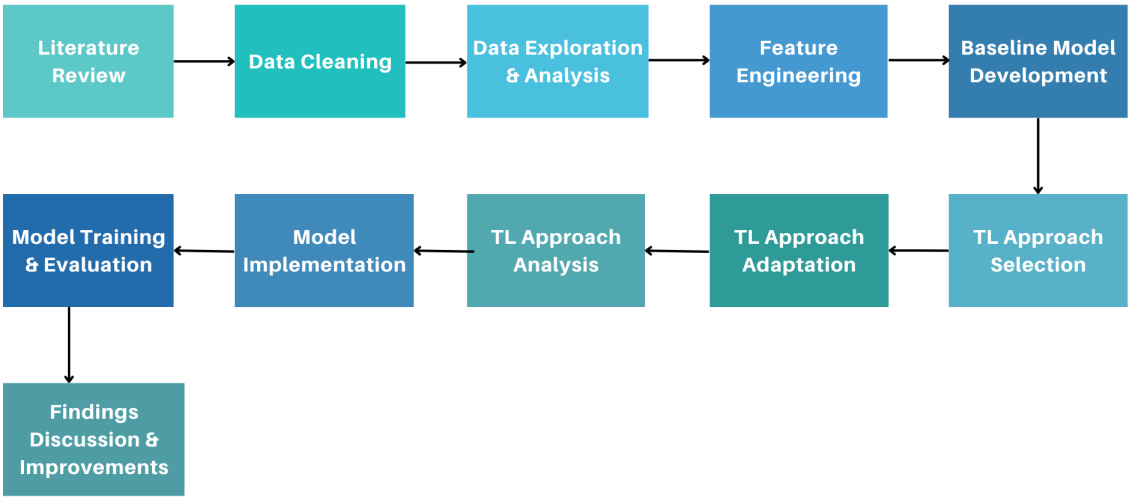


Figure 2.2: Research Workflow

2.3.2 Datasets

Supervised domain adaptation methods facilitate the transfer of knowledge from a source domain, typically rich in labeled data, to a target domain with limited data samples, effectively aligning related but dissimilar domains. Our evaluation of Domain Adaptation of Occupancy Estimation (OE) models utilizes proprietary datasets [19–21, 121] obtained from offices at the Grenoble Institute of Technology. These datasets, collected using ambient sensors, encompass both source and target domains observed within work offices (H355 and H358). The datasets assess levels of occupancy, including no occupant, one occupant, and two occupants, with the H358 dataset comprising 20 days of labeled data leading to 1440 data instances and the H355 dataset containing 6 days of labeled data resulting in 1068 data instances. The spatial distribution of the utilized sensors in the two offices is depicted below. The sensor network setup includes two video cameras for recording real occupancy numbers, an ambiance sensor network measuring temperature, motion, CO2 concentration, power consumption, door and window contact, and acoustic pressure from a microphone. Data are transmitted via the ENOCEAN protocol upon significant value change events to a centralized database with a web application for continuous data retrieval from various sources. In this chapter, office H358 has been used as the source environment while office H355 has been used as the target domain.

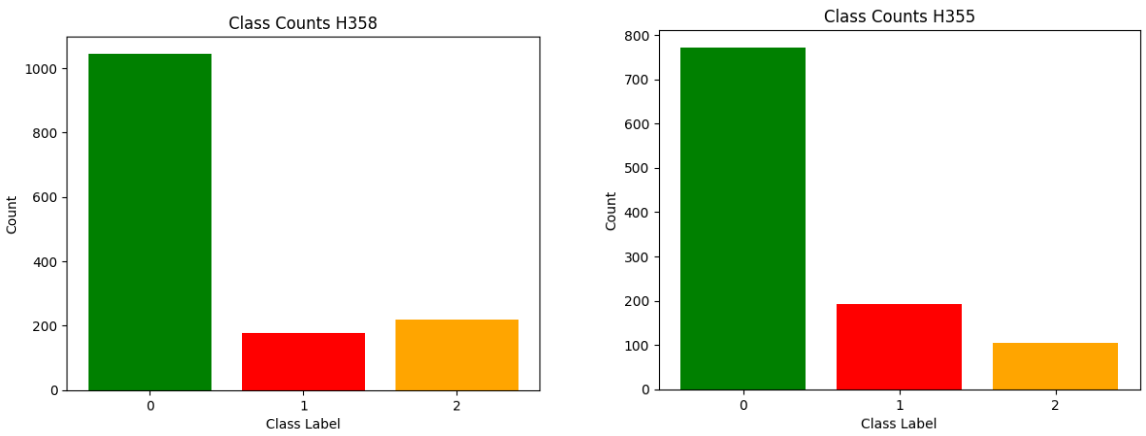
2.3.3 Experimental Setup

In this study, we utilized decision tree-based domain adaptation models with a specific emphasis on extracting interpretable if-else rules. To enhance the interpretability of the models, we constrained the maximum depth of the decision trees to three. For both the SER and STRUT algorithms, we have set the min_impurity_decrease parameter to 0, thereby facilitating aggressive splitting. This approach permits nodes in the decision tree to be partitioned without restraint based on any marginal reduction in impurity, irrespective of its magnitude. Consequently, this may yield a more intricate and deeper tree structure characterized by numerous splits. Although this aggressive splitting strategy might potentially lead to overfitting, it is deemed acceptable in this case, as we subsequently implement pruning techniques to mitigate any resulting overfitting effects. For GOSDT-DA, we

have used the following parameters: regularization: 0.02, depth_budget: 3, time_limit: 60, similar_support: False. Regularization penalizes the tree with more leaves. It is set to a relatively high value to find a sparse tree. Moreover, the dataset exhibits an imbalanced class distribution, where class 0 predominates as the majority class. A summary of label counts for both datasets is provided in Table 2, Figure 5, and Figure 6. To address this class imbalance, we employed imbalance handling libraries available within Sci-kit Imbalanced Learn in our adapted IMB-SER variants.

Table 2.1: Label count

Class	H355	H358
0	772	1045
1	192	177
2	104	218



(a) H358 label count

(b) H355 label count

Figure 2.3: Label distribution for H358 and H355 datasets

2.3.4 Metrics

The most common evaluation metric, accuracy, quantifies the ratio of correctly predicted samples to the total number of samples, expressed as Eq. (10), where TP represents True Positive, TN denotes True Negative, FP signifies False Positive, and FN indicates False Negative. However, one of the focal points of this research is addressing the challenges posed by imbalanced data, where

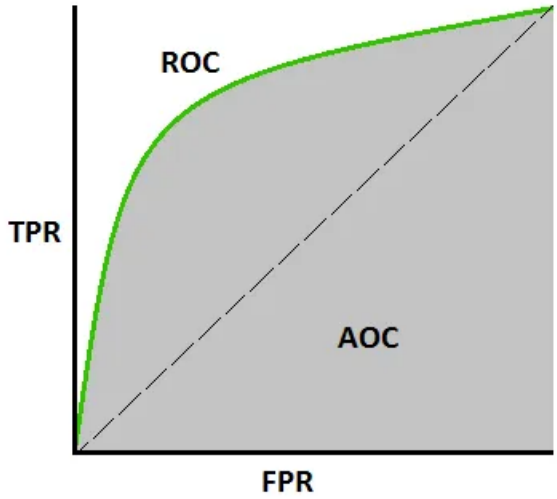


Figure 2.4: AUC-ROC Curve

unequal class distributions can potentially bias model performance and compromise its effectiveness. In fact, accuracy alone may prove unreliable in imbalanced datasets due to its susceptibility to skewing by the majority class. To address this, we employ precision, F1-score, and the ROC-AUC curve as evaluation metrics. Precision, given by Eq. (11), is used to assess the proportion of correctly identified TP instances among all instances predicted as positive by a classification model, providing insight into the model's ability to avoid FP. The F1-score Eq. (14), incorporating both FP and FN, offers a robust solution, particularly in scenarios of significant class imbalance, as it provides a harmonic mean of precision and recall in a single metric, giving equal weight to both measures. Moreover, the AUC-ROC curve provides a visual representation of a model's performance across different thresholds, showing the relationship between true positive rate (TPR) also known as Recall, given by (12), and false positive rate (FPR) given by (13). ROC is a probability curve and AUC represents the degree or measure of separability. The AUC represents the overall discriminatory power of the model, covering the entire area under the ROC curve. The AUC-ROC curve tells how much the model is capable of distinguishing between classes. Figure 5 represents a graphical depiction of the AUC-ROC curve.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

$$\text{Recall / TPR} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{FPR} = \frac{FP}{FP + TN} \tag{13}$$

$$F1\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

2.3.5 Experimental results

To ensure comparability and consistency across all evaluated models, we standardized the allocation of the original dataset for testing purposes, reserving 25% of the entire dataset as the test set. This test set comprises 267 instances, distributed among the classes as follows: 193 instances belong to Class 0, 48 to Class 1, and 26 to Class 2. This structured approach allows for a uniform evaluation framework, facilitating an unbiased comparison of model performance across different classification strategies.

In the course of our experimental analysis, a critical decision involved determining the optimal proportion of the target dataset to allocate for training purposes. An extensive analysis, illustrated in Figure 6, was conducted to identify the minimal percentage of training data that would yield superior performance in terms of AUC score for Structured Expansion Reduction (SER) as it is the base for our variants, relative to the baseline model trained solely on target data which is referred to as Base Model (T) throughout this chapter. Our findings indicate that 10% of the target dataset constitutes this optimal threshold.

Throughout this section, we compare the performance of our models with the Base Model (T) which is trained on 10% of the training set of the target domain, and the Base Model (S) which is trained solely on the source data.

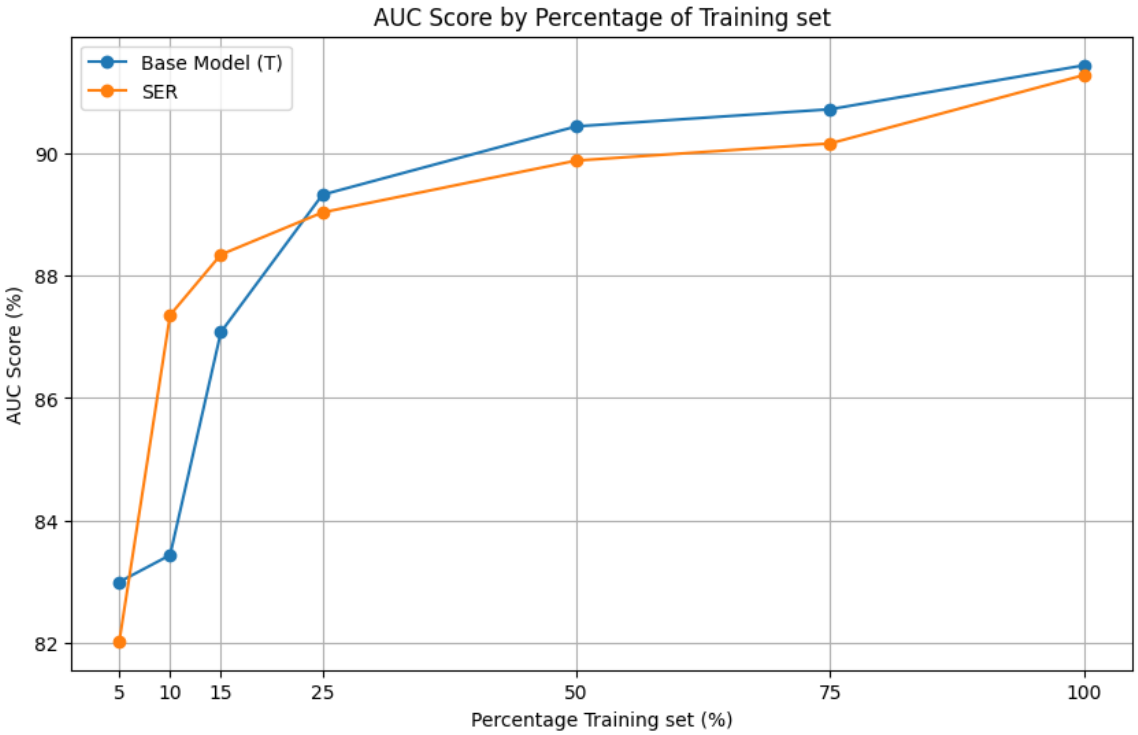


Figure 2.5: Auc score vs percentage training set

2 class OE

Two-level occupancy estimation, akin to occupancy detection, simplifies the representation of occupancy status into binary classes: class 0 indicating absence and class 1 indicating presence. The Base Model (T), trained exclusively on target data, exhibited robust performance metrics, achieving an accuracy of 89.89%, an F1-score of 90.27%, and an AUC of 84.26% as opposed to the Base Model (S), trained solely on source data, registered considerably lower metrics with 78.65% accuracy, 83.32% F1-score, and a significantly diminished AUC of 62.32%. This disparity underscores the challenges posed by the direct application of source-trained models to target datasets, highlighting the impact of domain differences.

Furthermore, to compare our work with commonly employed generalizability methods, we applied K-Fold cross-validation to the base models. To identify the optimal value of K, we trained the model using K values ranging from 2 to 11 and analyzed the resulting metrics. Interestingly, we observed a decline in performance. This was particularly evident in the case of the 6-Fold Base Model (S), where the model exhibited overfitting on the training data and underperformed on the

test set, achieving an accuracy of 72.28%, an F1 score of 83.91%, and an AUC score of 50% but it only predicted values for one class.

Exploring further into DA adaptations, the SER and STRUT models demonstrated variable efficacy. SER maintained commendable performance levels with 89.14% accuracy, 88.59% F1-score, and 82.49% AUC, suggesting effective mitigation of domain shifts. Conversely, STRUT lagged with 83.15% accuracy, 80.90% F1-score, and 70.84% AUC, indicating that certain domain adaptation frameworks may not uniformly benefit all facets of binary classification tasks. Noteworthy are the modified variants, SER* and STRUT*, which displayed mixed results; SER* saw a decrement in performance, whereas STRUT* depicted improvement, underscoring the significance of precise parameter tuning and algorithmic adjustments within DA frameworks.

The hybrid SER-STRUT model mirrored SER's performance metrics, suggesting limited distinct benefits from the integration in this binary setting. The GOSDT-DA model, employing a sophisticated decision tree algorithm, matched the high performance of the target-trained base model, illustrating the potential of advanced tree structures in binary classification scenarios.

Diving deeper into imbalance handling strategies, the IMB-SER variants which incorporated class imbalance techniques including SMOTE, ClusterCentroids, and SMOTETOMEK outshined the other models. Notably, IMB-SER SMOTETOMEK achieved the highest accuracy of 93.63%, an F1-score of 93.59%, and an AUC of 91.43%, attesting to the profound impact of sophisticated sampling strategies in enhancing model robustness against class distribution skews. At the same time, IMB-SER SMOTE has achieved the highest AUC score of 92.06%. These results highlight the critical need for addressing class imbalance in training datasets to optimize the performance of occupancy estimation models.

For comparative analysis, we incorporated several models from [19], which were trained on datasets similar to ours, with comparable features, labels, and data distribution. The CAT (Cluster Alignment with a Teacher) model, a domain adaptation technique that mitigates domain shift by training a teacher model on source data to generate pseudo-labels and align class clusters across domains, achieved an accuracy of 89.80% and an F1-score of 65.22%. The DSN (Domain Separation Networks) method, which separates features into shared and domain-specific spaces to enhance domain generalization while isolating noise, exhibited a more modest performance with an accuracy of

56% and an F1-score of 63.87%. The ATDOC+NC (Auxiliary Target Domain-Oriented Classifier with Nearest Centroid Classifier) model, which leverages source-like instances in the target domain using dynamic class centroids and a memory mechanism to generate high-confidence pseudo-labels, showed strong performance with an accuracy of 88.68% and an F1-score of 90.57%. Notably, AUC scores were not available for these models.

Table 2.2: 2 level Occupancy Estimation Performance Metrics

Method	Accuracy	F1-score	AUC
Base Model (T)	89.89	90.27	84.26
Base Model (S)	78.65	83.32	62.32
5 Fold Base Model (T)	87.64	87.96	82.28
6 Fold Base Model (S)	72.28	83.91	50.00
SER	89.14	88.59	82.49
SER*	88.39	87.73	81.14
STRUT	83.15	80.90	70.84
STRUT*	86.89	88.03	77.60
SER-STRUT	89.14	88.59	82.49
GOSDT-DA	89.89	89.44	83.84
IMB-SER SMOTE	92.13	92.27	92.06
IMB-SER ClusterCentroids	91.39	91.61	86.96
IMB-SER SMOTETOMEK	93.63	93.59	91.43
CAT	89.80	65.22	-
DSN	56.00	63.87	-
ATDOC+NC	88.68	90.57	-

3 class OE

Three-level occupancy estimation assigns numerical values to represent the count of individuals within a given area. In this scenario, class 0 signifies the absence of any individual, while class 1 indicates the presence of a single person. Class 2, on the other hand, denotes the occupancy of more than one individual within the designated space. The experimental results, summarised in Table 4, reveal variations in the performance of the different models evaluated for occupancy estimation.

In the comparative evaluation analysis of various DA models applied to a 3-level occupancy

estimation task, our experimental results reveal notable distinctions in performance across different methodologies and adaptations to class imbalance. The Base Models, trained solely on the source (Base Model (S)) and target data (Base Model (T)), establish a foundational understanding of domain disparities. Specifically, Base Model (T) outperforms Base Model (S) with accuracies of 77.90% vs. 64.79%, F1-scores of 76.39% vs. 65.80%, and AUCs of 83.43% vs. 73.60%, underscoring the challenge of direct domain transfer due to variances in data distribution. Direct domain adaptation yields 78.28% accuracy and an Auc score of 83.71.

In this instance, the 5-Fold Base Model (S) and, notably, the 6-Fold Base Model (T) with an accuracy of 80.52%, an F1-score of 80.71%, and an AUC score of 85.39%, demonstrate improved performance compared to the base models, yielding results comparable to STRUT*. This enhancement in performance may be attributed to the identification of more optimal parameters through K-Fold cross-validation, which also likely mitigated the impact of data imbalance.

Enhanced methodologies such as SER (Structured Expansion Reduction) and STRUT (Structured Transfer Using Trees) demonstrate superior performance over base models, validating the efficacy of DA in mitigating domain shifts. SER achieves an F1 score of 83.17% and an AUC score of 87.36%. Notably, the combined SER-STRUT model which utilizes Random Forests and integrates the strengths of both SER and STRUT approaches outperforms individual configurations, achieving an accuracy of 84.27%, F1-score of 82.63%, and AUC of 88.20%, indicating that hybrid DA frameworks can more effectively leverage cross-domain knowledge. Distinctively, the direct domain adaptation using GOSDT yielded an accuracy of 85.77%, an F1-score of 86.48%, and an AUC of 89.33%.

Further refinement is observed in models addressing class imbalance. IMB-SER variants, which incorporate techniques like SMOTE, ClusterCentroids, and SMOTETOMEK, achieve the highest performance metrics, with IMB-SER SMOTETOMEK standing out as the most effective approach, illustrating the critical impact of sophisticated sampling strategies in enhancing model robustness against skewed class distributions. IMB-SER ClusterCentroids achieves an accuracy of 87.06%, F1-score of 86.78%, and AUC of 90.30%.

The models from [19] demonstrated inferior performance in the 3-class scenario when compared to our proposed models. While the CAT model achieved a comparable accuracy of 89.80% relative

to the domain adaptation methods that informed our approach, its F1-score was notably lower at 65.22%, indicating suboptimal precision and recall balance. Similarly, the DSN model recorded an accuracy of 56% and an F1-score of 63.87%, reflecting limited effectiveness in this context. In contrast, the ATDOC+NC model exhibited great performance, with a robust F1-score of 90.57% and an accuracy of 88.68%, highlighting its superior ability to handle the 3-class classification task.

Collectively, these results underscore the importance of selecting and tuning domain adaptation strategies according to specific domain characteristics and data distribution challenges, especially in imbalanced settings. Moreover, the implementation of domain adaptation strategies to address class imbalance is proven to enhance prediction accuracy and generalization across domains.

Table 2.3: 3 level Occupancy Estimation Performance Metrics

Method	Accuracy	F1-score	AUC
Base Model (T)	77.90	76.39	83.43
Base Model (S)	64.79	65.80	73.60
6 Fold Base Model (T)	80.52	80.71	85.39
5 Fold Base Model (S)	66.66	64.31	75.00
SER	83.15	83.17	87.36
STRUT	82.77	84.88	87.08
SER*	81.65	81.76	86.24
STRUT*	80.52	78.21	85.39
SER-STRUT	84.27	82.63	88.20
GOSDT-DA	85.02	83.34	88.76
IMB-SER SMOTE	84.27	84.30	88.20
IMB-SER ClusterCentroids	85.77	86.48	89.33
IMB-SER SMOTETOMEK	85.02	85.03	88.76
CAT	72.40	66.67	-
DSN	34.80	57.01	-
ATDOC+NC	68.00	64.24	-

We conducted a comprehensive evaluation focusing on the Area Under the Curve (AUC) scores for each class, a critical metric for performance assessment in imbalanced datasets. This analysis, detailed in Figure 7, scrutinizes the performance of base models, the original SER (which serves as the foundation for our enhancements), and the SER variants within a multiclass classification framework. The AUC scores for individual classes provide significant insights into each model's

capability to handle class imbalances effectively. Additionally, the micro AUC scores, which consolidate the performance across all classes, demonstrate that all models exhibit commendable effectiveness, with a progressive improvement observed from the base models to the more sophisticated IMB-SER strategies. Notably, the increase in micro AUC scores as we move from the Base Model (S) with 73.60 to IMB-SER ClusterCentroids with 89.33 underscores the effectiveness of incorporating advanced sampling techniques and DA to enhance overall prediction accuracy in the presence of class imbalance.

Base model (S) exhibited moderate performance for Classes 0 and 2 (AUC of 0.77 for both), but significantly underperformed for Class 1 (AUC of 0.50), suggesting challenges in accurately predicting this class likely due to its underrepresentation or inherent complexity in the source dataset. In contrast, the Base model (T) displays improved performance across all classes, particularly for Class 1 (AUC of 0.82), underscoring the benefits of the direct target domain training in enhancing model accuracy. However, this model showed a decline in performance for Class 2 (AUC of 0.67), potentially indicating overfitting or issues related to class imbalance in the target data. The application of SER demonstrated the utility of combining source and target data, maintaining high AUC scores for Classes 0 and 1 (0.88 and 0.77, respectively) and slightly improving the AUC for Class 2 (0.70) compared to the Base model (T). This highlights SER's capability to leverage cross-domain data to refine predictions.

The enhanced IMB-SER models demonstrate significant performance improvements by employing tailored resampling strategies to address class imbalance. The IMB-SER SMOTE model effectively raises class 1's AUC, maintaining robust scores for classes 0 and 2, illustrating the benefits of synthetic minority oversampling in enriching the dataset. Conversely, the IMB-SER ClusterCentroids model boosts class 0's AUC to 0.90 but sees a reduction for class 1 to 0.79, highlighting potential drawbacks of undersampling which might omit critical data points. Nonetheless, it still achieves the highest micro AUC score. The IMB-SER SMOTETOMEK model, combining SMOTE with Tomek links to remove noisy instances, shows the highest improvements across all individual classes, suggesting its effectiveness in managing complex class distributions and enhancing robustness.

The comparison of confusion matrices for the Base Model (T), illustrated in Figure 8, and

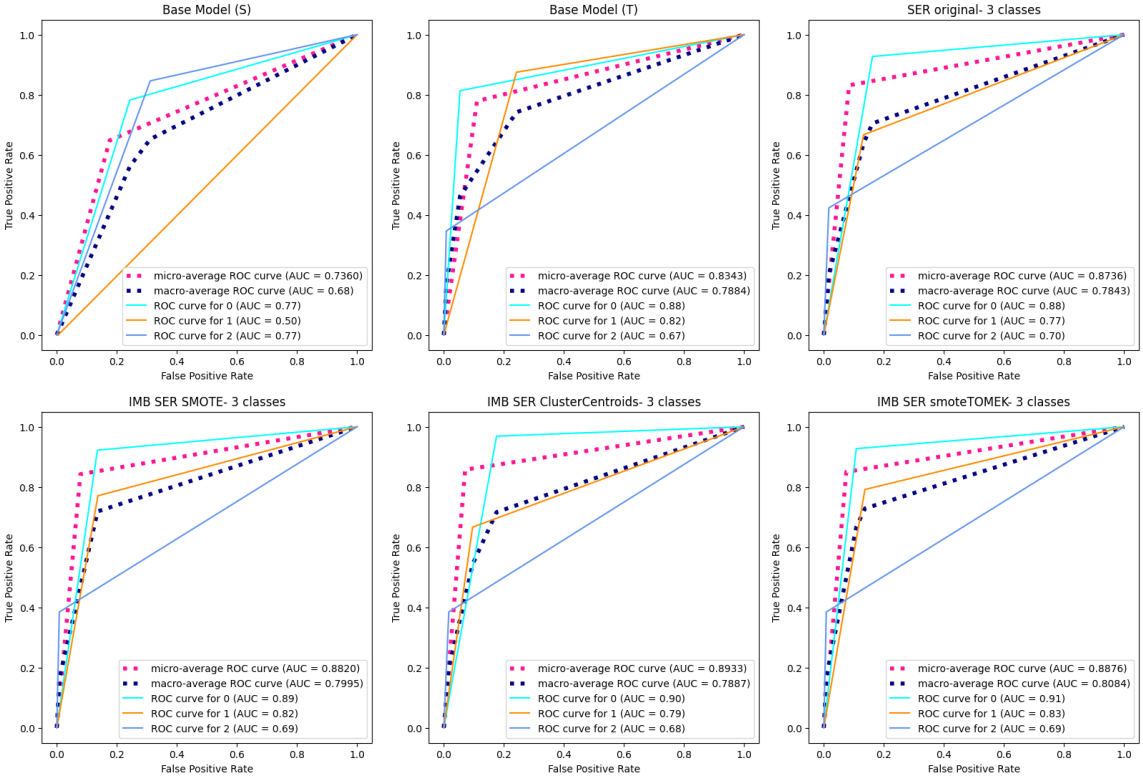
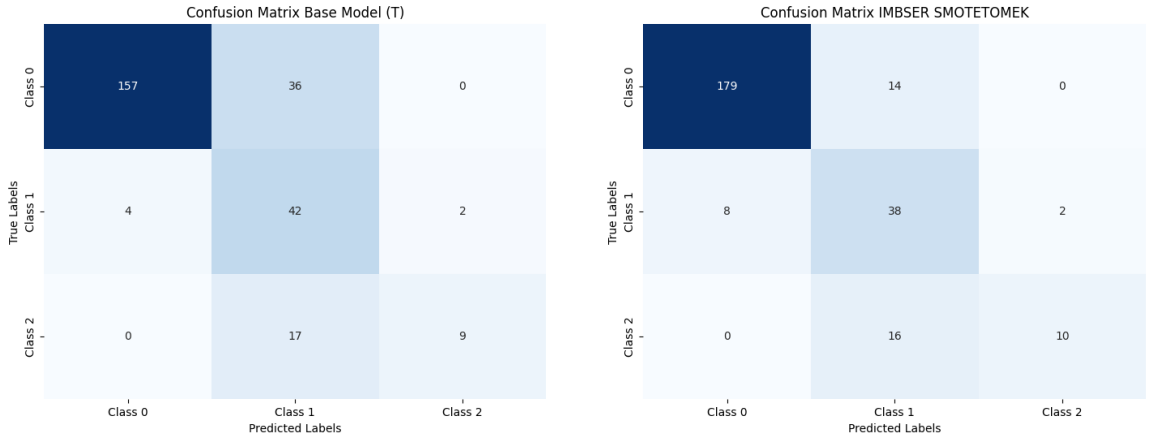


Figure 2.6: AUC scores for base models, SER and IMB-SER

the IMB-SER SMOTETOMEK model, displayed in Figure 9, provides clear evidence of the latter's enhanced performance across all classes. The IMB-SER SMOTETOMEK model significantly improves classification accuracy for Class 0 by reducing misclassifications, indicating better generalization and effective handling of overlapping class characteristics. Although there is a slight decrease in the number of correct predictions for Class 1, the model successfully lowers false positives, suggesting improved precision without a substantial compromise in sensitivity. For Class 2, both models show challenges with misclassifications, but IMB-SER SMOTETOMEK offers a slight increase in correct predictions, reflecting its capability to subtly enhance recognition of this less-represented class through advanced resampling and noise reduction strategies. Overall, the IMB-SER SMOTETOMEK model demonstrates superior performance across all classes, effectively reducing false positives and providing a more balanced approach to class representation. This enhancement makes it especially suitable for applications requiring accurate and reliable predictions across varied class distributions.



(a) Confusion matrix Base Model (T)

(b) Confusion matrix IMB-SER SMOTETOMEK

Figure 2.7: Comparison of confusion matrices for baseline and IMB-SER SMOTETOMEK models

2.3.6 Model Interpretability Discussion

The interpretability of decision rules is a critical factor in occupancy estimation applications within smart buildings, as it reinforces stakeholders' understanding and confidence in the predictive capabilities of the models. Clearly defined and comprehensible decision rules are essential for

straightforward validation and troubleshooting, enhancing the practical deployability of the models. In this research, our objective extends beyond achieving superior performance metrics such as accuracy, F1-score, and AUC; we aim to develop models that generate decision rules that are consistent with those derived from the base target model, thereby ensuring reliability and facilitating a seamless transition to more sophisticated models without compromising stakeholder trust.

In this section, we perform a comparative analysis of the three-class classification scenario. In pursuit of this goal, we retrained the Base Model (T) using the entire training dataset of the target domain to observe the decision rules it would generate with comprehensive labeled data access. We will refer to this model as the Base Model (Target) in the remaining sections of this chapter. This model achieved an accuracy of 87.26%. Interestingly, our best-performing model, the IMB-SER ClusterCentroids, produced decision rules that align closely with those generated by the base target model trained on the full training dataset, particularly in terms of feature contribution. To enhance model interpretability, we have intentionally restricted the depth of all our models to three levels. This limitation inherently introduces a bias by only exposing the top-level feature contributions; however, this constraint is a deliberate trade-off to prioritize explainability within our models. This approach facilitates a clearer understanding of the decision-making processes, at the expense of capturing more granular interactions among features.

The IMB-SER ClusterCentroids rules, illustrated in Figure 10, involve features such as Power, RMS, and CO2. This model uses a mix of feature thresholds to define its classes. For instance, it uses Power at two critical thresholds (27.875 and 40.323) to differentiate among the classes, incorporating both RMS and CO2 as secondary discriminative features. This suggests an attempt to balance the influence of multiple features in determining class membership. The Base Model (Target), on the other hand, shows a more varied use of features across its rules which is illustrated in Figure 11. It relies mainly on Power but also integrates Door_Contact, CO2, and RMS in defining class conditions. The use of Door_Contact alongside Power across a range of conditions suggests a model that is potentially capturing more complex interactions between features than what is seen in the IMB-SER ClusterCentroids rules. In IMB-SER ClusterCentroids, the thresholds for Power are higher compared to the Base Model (Target). This discrepancy may point to a fundamental difference in how each model perceives the influence of Power on occupancy estimation. IMB-SER

ClusterCentroids might be capturing different stages of occupancy scenarios compared to the Base Model (Target) as it trains only on 10% of the training set. Moreover, the Base Model (Target) uses narrower slices of feature space (e.g., Power between 10.625 and 33.875), which might indicate a finer-grained approach to partitioning the data based on these features. In contrast, IMB-SER ClusterCentroids uses broader criteria, potentially making it less sensitive to minor variations in feature values but possibly more robust to noise. Comparing the rules from the Base Model (S), illustrated in Figure 12, reveals a strong reliance on the feature RMS across all its rules, with different thresholds adjusting class decisions based on minor variations in RMS. This influence is somewhat visible in the IMB-SER ClusterCentroids, where RMS also plays a critical role in defining classes, in conjunction with other features. This could suggest that while the IMB-SER ClusterCentroids has adopted a multidimensional approach to feature integration, the underlying impact of RMS observed in the source model persists, affecting how thresholds and feature combinations are configured. We can however notice that the threshold for RMS is higher in IMB-SER ClusterCentroids.

Interestingly, the decision rules generated by the Base Model (Target) and IMB-SER ClusterCentroids exhibit certain similarities when dissected by class. For Class 0, both models formulate rules including Power and RMS. The Base Model (Target) introduces additional complexity by incorporating Door_Contact and CO2, conditions absent in the IMB-SER ClusterCentroids. Simplifying the Base Model (Target) by disregarding Door_Contact and CO2 aligns its rules more closely with those of the IMB-SER ClusterCentroids, both indicating class 0 when Power and RMS are below certain thresholds. For Class 1, the Base Model (Target)ś rules, when excluding Door_Contact, align with those of IMB-SER ClusterCentroids in terms of their reliance on thresholds for Power, RMS, and CO2. Both models suggest class 1 for mid-range values of Power coupled with either high RMS or CO2 is lower than a certain threshold, showcasing their sensitivity to these features for classifying instances into Class 1. Class 2 rules show a direct comparison where both models categorize observations as Class 2 based on high values of Power and CO2 if we disregard RMS for IMB-SER ClusterCentroids. This indicates a consistent understanding across both models that higher power usage combined with high carbon dioxide levels are indicative of Class 2 scenarios. These observations underscore a fundamental alignment in the decision-making criteria between the base and IMB-SER ClusterCentroids models, with variations primarily in feature complexity

if Power <= 27.875 and RMS <= 0.024: **class 0**

if (Power <= 27.875 and RMS > 0.024): **class 1**

if (27.875 < Power <= 40.323 and RMS <= 0.045) or (Power > 40.323 and co2 <= 955.429): **class 1**

if ((Power > 40.323 and co2 > 955.429) or (27.875 < Power <= 40.323 and RMS > 0.045)) : **class 2**

Figure 2.8: IMB-SER ClusterCentroids rules

if (Power <= 10.625) and ((Door_Contact <= 0.5 and RMS <= 0.038) or (Door_Contact > 0.5 and co2 <= 485.0)): **class 0**

if 10.625 < Power <= 33.875 and co2 <= 505.0 : **class 0**

if (10.625 < Power <= 33.875 and co2 > 505.0) or (Power > 33.875 and co2 <= 1265.0)) : **class 1**

if ((Power <= 10.625) and ((Door_Contact > 0.5 and co2 > 485.0) or (Door_Contact <= 0.5 and RMS > 0.038))) : **class 1**

if (Power > 33.875 and co2 > 1265.0) : **class 2**

Figure 2.9: Base model (Target) rules

and the threshold values utilized for classification. The comparative analysis suggests that while both IMB-SER ClusterCentroids and Base Model (Target) effectively utilize a mix of features to estimate occupancy, their strategies differ in terms of the complexity and specificity of the decision boundaries. The influence of the Base Model (S) on IMB-SER ClusterCentroids is evident in the continued importance of RMS, but with adaptations that reflect a more nuanced integration of additional features to address class imbalance effectively. A tree representation of the rules generated by the base model (Target) is shown in Figure 13.

To enhance our analysis of model explainability, we utilized SHAP values, which quantify the contribution of each feature to the predictive outcome. The summary bar plot in Figure 14 delineates the feature impacts across the different models under consideration. For this method, we did not limit the depths of the trees so that we could get a more comprehensive and accurate understanding of the behaviour of the models. This analysis reinforces our previous observations about feature reliance in decision-making. Base Model (Target) utilizes a more balanced feature set, suggesting a

if (RMS <= 0.01) or (RMS <= 0.014 and Power <= 75.5): **class 0**

if (RMS <= 0.014 and Power > 75.5) or (0.067 < RMS <= 0.187): **class 1**

if (RMS > 0.014 and RMS <= 0.067) or (RMS > 0.187): **class 2**

Figure 2.10: Base model (S) rules

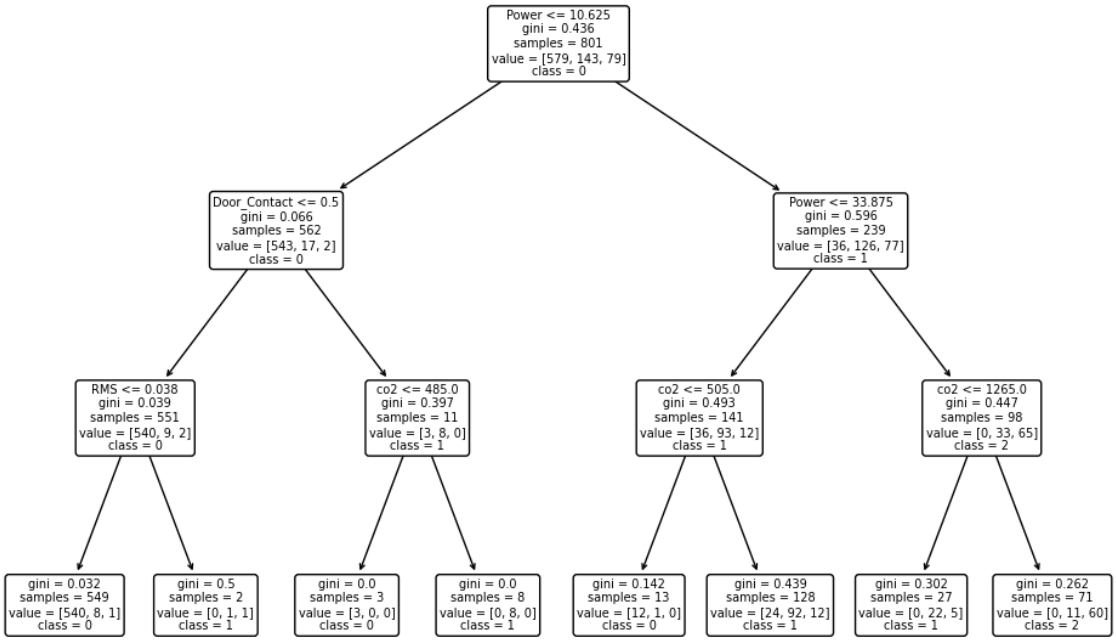


Figure 2.11: Base model (Target) tree representation

comprehensive analysis of the input space. The integration of Power alongside CO2, Temperature, RMS, and Door_Contact showcases the effective use of available features, which could enhance the model's accuracy and applicability in real-world scenarios. In the case of the Base Model (S), even though we notice an over-reliance on RMS, we can observe that when we do not limit the depth, the model is influenced by other features including Power, Temperature, and CO2. Nonetheless, the contribution of RMS still significantly exceed the contribution of the other features. This might indicate that the model is heavily biased towards features that were perhaps more discriminative in the source domain but might not be as effective across different domains.

IMB-SER ClusterCentroids generated rules that focus on Power, RMS, CO2, and Temperature like the Base Model (Target) highlighting these features as crucial determinants in classifying instances. However, we can observe that in IMB-SER ClusterCentroids, there is no contribution of Door_Contact. This might be because no critical data points for the latter were captured in the 10% of the target training data used. We can observe that the Base Model (S) is influencing IMB-SER ClusterCentroids as the contribution of RMS in the latter is much higher compared to Base Model (Target).

We can deduce that the IMB-SER ClusterCentroids and Base Models (Target) are potentially more robust and generalizable compared to the Base Model (S). The SHAP value contribution of multiple and similar features in the Target base model and IMB-SER Cluster Centroids indicate a more sophisticated model that might perform better in practical applications. While for the Base Model (S), a relatively high SHAP value for RMS suggests a high risk of overfitting to the noise or specific patterns in the source data rather than generalizable trends.

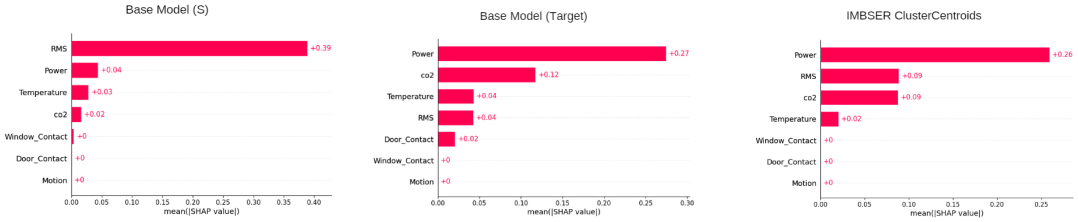


Figure 2.12: Feature contribution comparisons

Moreover, we have used heatmaps to compare the behaviour of the Base Model (Target) and IMB-SER ClusterCentroids on our test dataset, illustrated in Figures 15 and 16 respectively. The

saturation of the colors indicates the magnitude of the impact. A more saturated red means a strong positive impact, and a deeply saturated blue means a strong negative impact. By comparing the saturation across rows, we can identify which features generally have the most substantial impact on the model's predictions. For instance, in both models, power seems to have a greater contribution as it is of a darker color.

By looking at a single column across all features for one sample, we can understand the contribution of each feature to that specific prediction. This can be particularly useful for understanding model behavior in individual cases, which can help in troubleshooting or refining the model. For example, analyzing the 25th test observation reveals that the Power feature significantly impacts predictions for both the Base Model (Target) and IMB-SER ClusterCentroids. We can notice that the Base Model (Target) displays a more subtle coloration across a broader range of contributing features, indicating diverse feature influence. In contrast, the IMB-SER ClusterCentroids model shows more intense colors, suggesting a more focused reliance on fewer key features. This observation highlights differences in how each model weights and integrates feature contributions to determine outcomes.

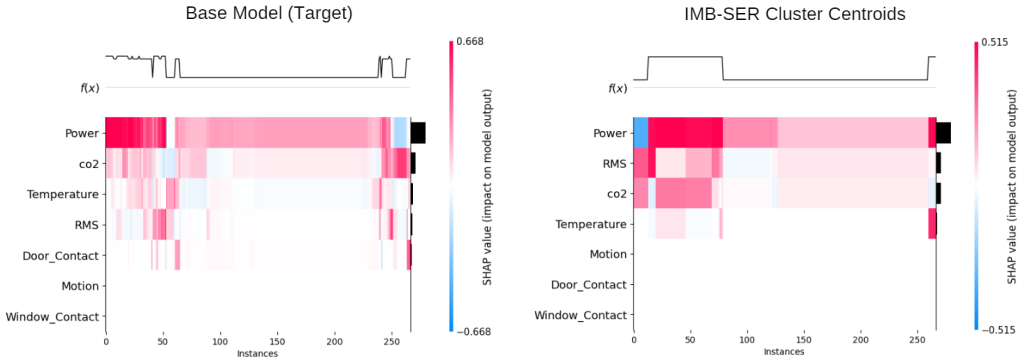


Figure 2.13: Base Model (Target) vs IMB-SER Cluster Centroids Heatmap

The table 5 presents a side-by-side comparative analysis of various DA methods based on three key criteria: Explainability, Performance, and Simplicity. STRUT and its variants (STRUT* and SER-STRUT) show medium levels of explainability and simplicity, with varying levels of performance, none of which reach the excellent mark. In contrast, while SER also scores high in explainability and simplicity, its performance is high but not excellent. The GOSDT-DA method, though it

achieves excellent performance, falls short in both explainability and simplicity, rated as low, which could hinder its practical application despite its effectiveness.

IMB-SER exhibits superior performance, establishing it as the most effective method for achieving high-quality results. From an interpretability standpoint, IMB-SER excels by ensuring that the outcomes are transparent and easily understood, which is essential for fostering trust and encouraging adoption in practical applications. This transparency stems from its foundation in SER, allowing us to trace decision rules from the source tree through their evolution into the target tree. This progression is analogous to the layers in neural network models, where the initial layers capture general patterns and the later layers (comparable to the leaves in decision trees) refine the decision-making process. However, in this case, the structure remains significantly clearer and more interpretable. Moreover, IMB-SER maintains a high degree of simplicity, which not only facilitates a deep understanding of the model's decision-making process but also makes the codebase more accessible for future enhancements.

This analysis clearly positions IMB-SER as the most balanced and effective method, offering superior performance without compromising on explainability or simplicity, making it a more robust choice compared to the other methods evaluated.

Table 2.4: Comparison of DA methods

Method	Explainability	Performance	Simplicity
SER	High	High	High
STRUT	Medium	Low	Medium
SER*	High	Medium	Medium
STRUT*	Medium	Medium	Medium
SER-STRUT	Medium	High	Medium
GOSDT-DA	Low	Excellent	Low
IMB-SER	Excellent	Excelllent	High

2.3.7 Contribution to Science

The contributions of this chapter in the realm of explainable domain adaptation for imbalanced occupancy estimation are significant for advancing scientific understanding and practical applications in Building Management Systems (BMS), energy management optimization, and HVAC systems. Firstly, the chapter addresses the critical challenge of accurately estimating occupancy in smart buildings, a key factor for energy efficiency. By employing domain adaptation methodologies tailored for imbalanced datasets and integrating explainable AI (XAI) techniques, the chapter enhances the interpretability and robustness of occupancy estimation models.

The practical application of these methodologies in BMS and HVAC systems can lead to more precise control of heating, ventilation, and air conditioning based on real-time occupancy data, significantly optimizing energy usage and reducing operational costs. This is especially crucial in larger buildings where energy expenditure is considerable. The use of decision trees in the proposed domain adaptation approach, coupled with explainable AI, allows for the generation of clear, interpretable models that facility managers can easily understand and trust, which is vital for real-world applications.

Furthermore, the explainability aspect brought forth by XAI techniques such as SHAP (Shapley Additive Explanations) enhances the trustworthiness and transparency of the models. This progress in explainable AI not only helps in aligning with regulatory standards for AI systems but also ensures that the stakeholders can understand the decision-making process of the models, leading to broader acceptance and implementation of these technologies in sensitive environments like smart buildings.

Overall, by bridging the gap between advanced AI techniques and practical energy management applications, this research pushes forward the frontiers of smart building technology, promoting sustainability and efficiency in urban development.

2.4 Conclusion

This study systematically investigated and enhanced various explainable domain adaptation methods to address challenges associated with model opacity, imbalanced datasets, and the scarcity

of labeled data, which are prevalent in occupancy estimation datasets. We evaluated multiple established approaches, including SER, STRUT, SER*, STRUT*, SER-STRUT, and GOSDT-DA. Additionally, we developed a novel adaptation of SER, named IMB-SER, specifically designed to manage data imbalance through techniques such as SMOTE, ClusterCentroids, and SMOTE-TOMEK. Our empirical analyses demonstrate that the IMB-SER variants outperform traditional methods across accuracy, F1-score, and AUC metrics, highlighting their effectiveness in providing precise occupancy estimates while ensuring model interpretability. Notably, the IMB-SER SMOTE-TOMEK model exhibited superior performance in binary classification tasks, while the IMB-SER ClusterCentroids model slightly outperformed other variants in three-class classification scenarios. An in-depth analysis of decision rules generated by each model in a three-class classification further emphasized the significance of rule concordance, offering insights into how different domain adaptation approaches formulate their predictions. Our findings suggest that the IMB-SER ClusterCentroids model not only enhances predictive performance but also aligns its decision rules closely with those of a base model trained on target data using the entire training set. The prioritization of model explainability was maintained through the integration of SHAP values, which is essential for the operational deployment of these systems in smart buildings.

This research contributes significantly to explainable domain adaptation within smart building applications, particularly for occupancy estimation by introducing innovative techniques for domain adaptation, addressing imbalanced data, and enhancing model explainability. By adapting existing domain adaptation methodologies to effectively handle datasets specific to occupancy estimation, we addressed a critical gap in managing real-world smart building data and optimizing energy distribution and consumption. The modification of the DA Decision Tree architecture to effectively handle imbalanced datasets represents a novel enhancement, improving the robustness and accuracy of domain adaptation models in these complex environments. Additionally, the integration of Explainable AI (XAI) techniques is a key contribution, as it clarifies model behavior on test data and provides a framework for assessing the similarity between different domain adaptation strategies and original models. This integration ensures that the developed models are both high-performing and interpretable, which is crucial for the deployment of intelligent systems in smart buildings. These contributions pave the way for future research to apply these methodologies in broader settings,

potentially leading to further innovations in domain adaptation and the deployment of intelligent systems across diverse real-world scenarios. Moving forward, we aim to explore the application of these enhanced domain adaptation strategies across heterogeneous domains, thereby broadening their applicability and demonstrating their adaptability. This direction holds promise for substantial advancements in intelligent systems, underscoring the importance of transparent and interpretable decision-making in diverse contexts.

Chapter 3

Explainable Domain Adaptation without Source Data for Activity Recognition

While Chapter 2 focuses on improving supervised domain adaptation under imbalance conditions, many real-world scenarios impose a stricter constraint: the source data cannot be accessed due to privacy or regulatory constraints. This motivates the need for source-free domain-adaptation strategies capable of leveraging knowledge from pretrained models without having to access raw source data. Building on the insights of Chapter 2, this chapter shifts the focus to activity recognition and proposes a privacy-preserving adaptation framework that operates solely on target-domain data enriched with model-based uncertainty cues. This transition marks the next stage in broadening the applicability of domain adaptation to more restrictive deployment contexts.

Introduction

Global energy consumption in buildings has reached unprecedented levels, with demand increasingly outpacing efficiency improvements. According to [130], buildings are responsible for approximately 32% of global energy use, and as reported by [131], HVAC systems alone account for nearly 60% of this consumption. With ongoing population growth, urbanization, and technological development, building energy demands are expected to rise further, underscoring the urgent need for smarter energy management strategies. While constructing energy-efficient smart buildings offers

one path forward, the majority of energy waste occurs in existing structures that dominate the global building stock. For example, [132] reports that buildings constructed after 2009 in the EU consume 30–60% less energy than those built before 1990, highlighting the importance of retrofitting.

Activity Recognition (AR) has emerged as a promising approach to optimizing energy use in existing buildings. [19] argue that AR improves HVAC efficiency by dynamically adjusting system operations based on occupants' real-time behavior, leading to lower energy consumption without compromising comfort. By detecting whether occupants are active, resting, or absent, AR enables intelligent adjustments of heating or cooling of HVAC systems when rooms are unoccupied or during periods of inactivity, such as sleep, where less intense conditioning is required. It can also automate the deactivation of non-essential devices or switch them to low-power mode while directing energy only to occupied zones based on the identified occupant usage of each space, thereby optimizing overall energy distribution. For instance, when AR detects that an occupant is sleeping, the HVAC system can reduce airflow or slightly adjust the temperature to save energy without affecting comfort. In contrast, during active periods like exercising or working, the system can increase ventilation or cooling accordingly. Lighting can also adjust accordingly, increasing brightness for reading and dimming during passive activities like watching TV.

Traditionally, supervised machine learning has been employed for AR in the smart building sector, where the model is trained with data collected in the same environment. [133] proposes to use Random Forest to classify human activities in real time based on simple time-domain statistical features extracted from accelerometer data obtained from low-cost, low-power microcontrollers. [134] proposed a skeleton-based activity recognition method using a fully connected deep LSTM network, where a mixed-norm regularization was introduced to automatically learn co-occurrence patterns among joints. One trending field of AI is computer vision. [135] proposes a human activity recognition approach that leverages Convolutional Neural Networks (CNNs) to automatically extract spatial features from video frames in the Weizmann dataset. By replacing manual feature engineering with deep learning, the method achieves high classification accuracy using models like VGG-16.

As noted by [21], many approaches face the persistent challenge of limited labeled data, due to the high cost, time requirements, and privacy concerns of real-world annotation. Directly reusing

a model trained in a label-rich source domain often yields poor results in new environments due to domain shift—particularly covariate shift, where input feature distributions differ while label relationships remain unchanged. For example, sensor data from two smart homes may vary due to layout differences, despite similar activities. This misalignment can significantly degrade performance. To address this, we adopt a Transfer Learning (TL) framework that leverages patterns learned from a well-labeled source domain to improve performance in a low-label target domain. Rather than training from scratch, TL promotes efficient learning and better generalization across domains. For instance, [51] proposes Domain-Adaptive Decision Trees (DADT), which adjust information gain using estimated target distributions. [136] introduces SHOT, a source-free method that freezes the source classifier and adapts the target encoder through information maximization and pseudo-labeling, enabling alignment without access to source data.

While AR models have seen significant advances, previous works have largely overlooked the critical aspect of model explainability. As a result, many AR systems function as black boxes, offering little to no insight into how specific predictions are made. This lack of transparency hampers our ability to verify, validate and trust model outputs, particularly in real-world applications like smart buildings where safety and reliability are paramount. Furthermore, as production data often differs from training data, model performance can degrade over time, making it crucial to monitor the models to promote responsible AI. The absence of explainability also increases the risk of unintended biases and errors, which can compromise the effectiveness of AR systems in sensitive environments like smart buildings. To address these concerns, [34] introduced IMB-SER SMOTETOMEK, an interpretable domain adaptation method based on decision trees that extends Structure Expansion Reduction (SER) to manage imbalanced occupancy data. Similarly, [41] proposed ExKMC, an explainable K-means clustering algorithm that employs simple binary decision trees to enhance the transparency of unsupervised clustering decisions.

While AR models have seen significant advances, they face two persistent limitations: (1) they struggle to generalize across environments due to domain shift, and (2) they often lack transparency, functioning as black-box models. These issues are especially problematic in real-world smart building deployments, where safety, reliability, and user trust are paramount. Although prior work in domain adaptation addresses covariate shift or source-free adaptation, few approaches explicitly

consider explainability during the transfer process. To address this gap, we aim to develop an interpretable and data-efficient transfer learning framework for activity recognition that performs well under domain shift while remaining transparent and practical for smart building deployment.

We propose a novel methodology named Uncertainty-aware Transfer via Augmentation for Decision Trees (UTA-DT) which leverages the human-readable decision rules, computational efficiency, and robustness to outliers of decision trees. UTA-DT transfers knowledge from a pretrained source model by augmenting target features with soft class probabilities, hard predictions, and uncertainty cues such as entropy and top-class margin. These enriched features enable the decision tree to learn interpretable and domain-adaptive rules without requiring access to source data. To further enhance model transparency, we integrate both local and global explainability techniques. Global explanations, such as tree visualizations and Partial Dependence Plots (PDPs), provide an overarching view of how features influence predictions across the entire dataset, revealing consistent decision patterns and feature importance. In contrast, local explanations, through decision path tracing, allow us to interpret individual predictions by showing the specific rules and thresholds that led to a particular outcome. Together, these techniques offer complementary insights: global tools help assess model behavior at scale, while local tools support debugging, auditing, and understanding edge cases at the instance level. This layered approach allows us to verify the consistency, robustness, and interpretability of UTA-DT's decision-making process. The framework's efficacy is evaluated on benchmark datasets for Activity Recognition (AR), where it consistently achieves high accuracy, significantly outperforming baseline models while enhancing the interpretability of the transfer learning process. These results highlight the value of explainable models for fostering user trust and supporting the deployment of more transparent and accountable Building Management Systems (BMS). Indeed, the future of energy-efficient buildings lies not only in advanced hardware but also in intelligent, behavior-aware systems. By integrating real-time activity recognition with energy control algorithms, our method achieves improved efficiency without the need for human intervention.

The remainder of this chapter is structured as follows: Section 2 outlines the methodology, Section 3 presents the results, and Section 4 discusses the key findings and their implications.

Key Innovations

- Proposes a novel tree-based domain adaptation method that augments target features with soft/hard predictions and uncertainty from a pre-trained source model.
- Introduces entropy and top-class margin as domain adaptation features to improve performance under class distribution and covariate shifts.
- Ensures privacy by relying solely on source model outputs, not raw data.
- Demonstrates strong performance on real-world activity recognition data, outperforming baseline models.

Practical Implications

Our proposed method enables explainable and data-efficient transfer learning for multi-class Activity Recognition problems by leveraging the interpretable nature of decision trees. It improves generalization in low-label target domains, preserves privacy by using only source model outputs, and provides explainability through global and local rule-based insights. This makes our model ideal for high-stakes and resource-constrained environments like smart buildings.

3.1 Methodology

UTA-DT is a novel, interpretable domain adaptation framework based on decision trees. It is designed to enable feature-based knowledge transfer from a label-rich source domain to a label-scarce target domain while preserving data privacy and model interpretability.

3.1.1 Domain Adaptation

The UTA-DT framework begins with a frozen, pretrained decision tree model that has been trained in a supervised manner on source domain data. Importantly, the raw source data is never accessed during the transfer process, thereby preserving privacy which is a critical consideration in contexts such as smart home environments where occupant data must remain protected. The source

model serves as a knowledge-rich hypothesis from which transferable signals can be extracted and leveraged in the target domain.

To enable domain adaptation, the source model is applied to each instance in the target domain to generate soft class probabilities and hard predictions. From these, two additional uncertainty-aware features are computed: Shannon entropy and prediction margin. These outputs are concatenated with the original target features to form an enriched, uncertainty-aware representation that incorporates both target-specific information and latent knowledge from the source.

The augmented features used for transfer in UTA-DT are defined as follows:

(1) **Soft predictions**: The class probability distribution predicted by the source model for a given input instance x:

$$\begin{aligned} \hat{\mathbf{p}}(x) &= [\hat{p}_1(x), \hat{p}_2(x), \ldots, \hat{p}_C(x)]\,, \\ \text{where} \quad \hat{p}_i(x) &= \frac{n_i}{\sum_{j=1}^{C} n_j} \end{aligned} \tag{15}$$

Here, $\hat{p}_i(x)$ denotes the predicted probability for class i, n_i is the number of training samples of class i in the leaf node where x falls, and C is the total number of classes and the denominator $\sum_{j=1}^{C} n_j$: total number of samples in that leaf.

(2) **Hard prediction**: The predicted class label assigned to instance x, defined as the class with the highest probability:

$$\hat{y}(x) = \arg\max_i \hat{p}_i(x) \tag{16}$$

(3) **Entropy**: A scalar value measuring the uncertainty of the prediction, computed as the Shannon entropy of the soft prediction vector:

$$\mathcal{H}(x) = -\sum_{i=1}^{C} \hat{p}_i(x) \log(\hat{p}_i(x) + \varepsilon) \tag{17}$$

where C is the total number of classes, p_i is the predicted probability of class i and ε is a small constant added for numerical stability to avoid $\log(0)$.

(4) **Margin**: The difference between the two highest predicted probabilities, serving as a proxy

for model confidence:

$$\mathcal{M}(x) = \hat{p}_{(1)}(x) - \hat{p}_{(2)}(x) \tag{18}$$

where $p_{(1)}$ is the highest predicted class probability and $p_{(2)}$ represents the the second-highest predicted probability. A larger value indicates a more confident classification.

The resulting augmented feature space is then used to train a decision tree classifier tailored to the target domain. This enriched input enables the transferred model to capture not only local patterns from the target data but also transferable knowledge reflecting the source model's decision structure. As a result, the classifier can construct more accurate and generalizable decision boundaries, even in settings where labeled target data is scarce. The methodology ensures alignment between training and inference pipelines, while preserving interpretability and supporting data-efficient adaptation across domains.

A summary of the new number of features after doing the Augmentation is illustrated in Table 3.1. The augmented feature space used in UTA-DT consists of the original input features d, the soft class probabilities from the source model C, and three additional scalar features: the hard prediction, entropy, and margin. This results in a total dimensionality of $d + C + 3$. For instance, for one case in our experiments, we used $d = 34$ and $C = 10$, yielding a 47-dimensional input for the transferred decision tree.

Table 3.1: Breakdown of the Augmented Feature Space in UTA-DT

Feature Type	**Symbolic Count**	**Example Value**
Original Features	d	34
Soft Predictions	C	10
Hard Prediction	1	1
Entropy	1	1
Margin	1	1
Total Features	$d + C + 3$	47

The pseudocode for the Transfer Learning part is shown in Algorithm 3.

Algorithm 3 UTA-DT

procedure UTA-DT(M_s, X_t, y_t)
  for all $x \in X_\mathrm{t}$ **do**
    $\hat{\mathbf{p}}(x) \leftarrow M_s$.predict_proba$(x)$
    $\hat{y}(x) \leftarrow M_s$.predict$(x)$
    $\mathcal{H}(x) \leftarrow -\sum_i \hat{p}_i(x) \log(\hat{p}_i(x) + \varepsilon)$
    $\mathcal{M}(x) \leftarrow \hat{p}_{(1)}(x) - \hat{p}_{(2)}(x)$
    $x_\mathrm{aug} \leftarrow [x \parallel \hat{\mathbf{p}}(x) \parallel \hat{y}(x) \parallel \mathcal{H}(x) \parallel \mathcal{M}(x)]$
  end for
  $M_t \leftarrow \mathrm{DT}(X_\mathrm{t}^\mathrm{aug},\ y_\mathrm{t})$
end procedure

3.1.2 Explainable AI

To ensure transparency and interpretability in the transfer learning process, UTA-DT incorporates both global and local explainability techniques tailored to decision trees. As a naturally interpretable model class, decision trees enable direct extraction of human-readable rules that govern classification decisions. We leverage this property by applying global explanation methods such as tree rule extraction and visualization, which allow for the inspection of full decision paths from root to leaf. Additionally, Partial Dependence Plots (PDPs) are employed to assess the marginal effect of individual features on the model's predictions, helping to uncover consistent decision patterns across the dataset.

We also implement local explanation methods to better understand how individual predictions are made. Specifically, we trace the decision path taken by each instance through the tree, identifying which feature thresholds led to its classification. Together, these explainability tools provide a comprehensive view of UTA-DT's behavior, facilitating debugging, model auditing, and stakeholder trust in high-stakes domains like smart buildings. An overview of the UTA-DT methodology is depicted in the flowchart presented in Figure 3.1.

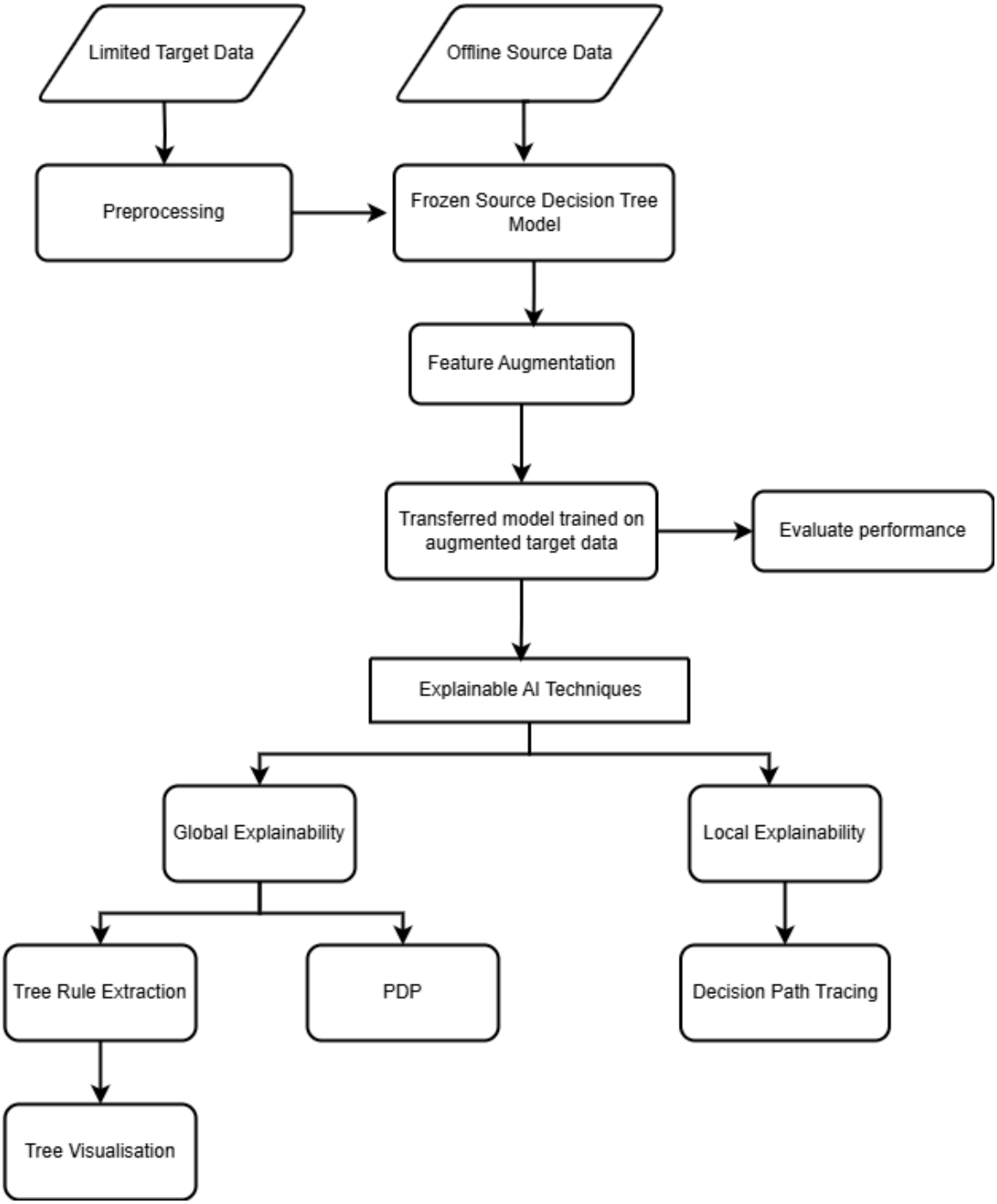


Figure 3.1: Illustrates the UTA-DT pipeline, beginning with limited target and frozen source model, proceeding to model transfer through feature augmentation, and culminating in performance evaluation and explainability via global and local XAI techniques.

3.1.3 Datasets

We use the publicly available CASAS dataset, developed at Washington State University (WSU) by [137] as part of the Center for Advanced Studies in Adaptive Systems. It contains real-world recordings of activities of daily living (ADLs) in smart homes, captured via non-intrusive sensors such as motion detectors, door contacts, temperature sensors, and light switches. Unlike proprietary datasets collected in controlled settings, CASAS offers privacy-preserving, real-world data, making it especially valuable for human activity recognition (HAR), smart home automation, and assistive technologies.

Our experiments follow a domain adaptation setup using two CASAS subsets: HH101 as the source domain and HH105 as the target. Both datasets provide labeled sensor event data from different smart home environments, each with a single resident and unique layout and behavior patterns. This setting reflects real-world deployment challenges where models must generalize across diverse spaces and routines. Each instance is represented by spatial and behavioral features extracted from sensor activations, including activity frequency and duration per space, activity complexity, area transitions, and the number of distinct sensors triggered. Temporal features such as time of day and day of the week were removed during preprocessing to avoid overfitting to environment-specific routines and improve generalization.

Initial experiments focused on ten energy-relevant activities: Cooking Breakfast, Cooking Lunch, Cooking Dinner, Eating, Toileting, Bathing, Sleeping, Relaxing, Dressing, and Watching TV. These were selected for their routine frequency and distinct energy usage patterns. The task was then expanded to twelve classes with Phone Use and Entertaining Guests, and later to fifteen with Washing Dishes, Personal Hygiene, and Taking Morning Medications. This progressive expansion allowed us to evaluate model robustness in increasingly complex scenarios involving a wider range of overlapping real-world activities.

3.1.4 Metrics

For evaluation, we rely on two primary metrics: accuracy and F1 score. While accuracy (Eq (19)), provides a general measure of overall correctness by comparing total correct predictions to all predictions, it can be misleading in the presence of class imbalance. To complement this, precision (Eq (20)) and recall (Eq (21)) are combined in the F1 score (Eq (22)), making it especially valuable when the cost of false positives and false negatives differs. Together, these metrics provide a more robust and comprehensive assessment of model performance across various activity classes.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{20}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{21}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{22}$$

3.1.5 Experimental setup

To evaluate UTA-DT, we compare it against three baselines: Base T, Base S, and STRUT. Base T is a standard decision tree trained only on limited target domain data and evaluated on held-out target instances. Base S is trained on abundant labeled source data and directly applied to the target test set without adaptation, highlighting the impact of domain shift due to its lack of exposure to target-specific distributions.

We also include STRUT [54], a decision tree-based domain adaptation method that modifies a source-trained tree by adjusting decision thresholds using target data. STRUT traverses the tree top-down, pruning unreachable nodes and recomputing thresholds at reachable ones to align with the target distribution. It selects thresholds by optimizing a criterion that balances divergence gain (alignment of source and target class distributions) and information gain (Gini index). To account

for possible label inversion around thresholds, STRUT evaluates both left-right and right-left splits when computing the optimal split. We selected STRUT as a baseline due to its status as a state-of-the-art tree-based domain adaptation method.

For all models, we set the decision tree depth to max_depth = 5 to ensure interpretability and enable fair comparison. Both source and target datasets were class-balanced. The source dataset included approximately 2000 instances per activity. To reflect real-world constraints, the target dataset was limited to 50 labeled instances per class, split 20% for training and 80% for testing. For the 15-class task, this resulted in 750 instances, with 150 for training and 600 for testing. UTA-DT uses this small training set to generate an augmented target dataset via feature transformation guided by the source model, without accessing raw source data. All models were implemented using Scikit-learn's DecisionTreeClassifier, with default hyperparameters except for max_depth.

3.1.6 Experimental Results

Table 3.2 reports the performance of all models on the 10-class activity recognition task, which includes common daily activities which includes Cooking Breakfast, Cooking Lunch, Cooking Dinner, Eating, Toileting, Bathing, Sleeping, Relaxing, Dressing, and Watching TV. The source-only model (Base S) struggles due to domain shift, yielding an accuracy of 53.55% and an F1 score of 47.79%. The target-only model (Base T), trained solely on a small labeled target subset, performs better with 78.25% accuracy and a 71.58% F1 score. STRUT demonstrates clear gains through adaptation, achieving 91.50% for both metrics. UTA-DT surpasses all baselines, reaching 97.00% accuracy and 97.03% F1 score. This substantial improvement underscores the effectiveness of UTA-DT's uncertainty-aware feature augmentation in capturing transferable patterns from the source model, enabling more robust and generalizable decision-making in the target domain.

Table 3.2: Performance of AR Dataset – 10 activities

Method	Accuracy	F1 score
Base T	78.25	71.58
Base S	53.55	47.79
STRUT	91.50	91.50
UTA-DT	**97.00**	**97.03**

Table 3.3 summarizes the performance of all models on the 12-class activity recognition task. Two new activities, Phone Use and Entertaining Guests, were added to the original ten. This added complexity results in a general decline in performance across models, reflecting the challenge of distinguishing between a broader range of human behaviors. The source-only model (Base S) struggles to adapt, achieving just 48.52% accuracy and 44.58% F1 score, underscoring the limitations of direct model transfer in the presence of domain shift. In contrast, the target-only model (Base T), trained solely on limited labeled target data, performs better with 71.87% on both metrics, though it lacks cross-domain knowledge. STRUT narrows this gap through adaptation, reaching 85.62% accuracy and 85.55% F1 score. UTA-DT once again outperforms all baselines, achieving 88.96% accuracy and 86.30% F1 score. These results highlight UTA-DT's scalability, showing that it can effectively integrate predictive signals and uncertainty to maintain high performance as the classification task becomes more complex.

Table 3.3: Performance of AR Dataset – 12 activities

Method	**Accuracy**	**F1 score**
Base T	71.87	71.87
Base S	48.52	44.58
STRUT	85.62	85.55
UTA-DT	**88.96**	**86.30**

Table 3.4 evaluates performance in the most challenging scenario, where the classification task involves 15 activities. We expand the classification task to 15 activities by adding Washing Dishes, Personal Hygiene, and Taking Morning Medications, introducing greater class diversity and behavioral overlap. As expected, this increased complexity results in a general decline in performance across all models. As a result, Base S drops further to 33.33% accuracy and 31.08% F1. Base T also sees reduced performance, reaching 62.33% on both metrics. STRUT remains competitive with 73.66% accuracy and 70.92% F1. UTA-DT, however, achieves the highest performance once again with 79.00% accuracy and 73.51% F1 score, despite the increased difficulty. Even in the most challenging scenario, UTA-DT continues to deliver competitive performance, illustrating its resilience in learning from sparse labels by effectively incorporating structured knowledge from the source

domain.

Table 3.4: Performance of AR Dataset – 15 activities

Method	Accuracy	F1 score
Base T	62.33	62.33
Base S	33.33	31.08
STRUT	73.66	70.92
UTA-DT	**79.00**	**73.51**

Discussion

To interpret the decision-making process of the UTA-DT model trained on the 10-class activity recognition task, we extract and analyze the full decision tree. As the depth is limited to 5 for interpretability, it is possible that some important features are truncated during tree growth. The extracted rules, illustrated in Figure 3.2, show that sensorElTime-OutsideDoor forms the root node, indicating its strong discriminative power. Downstream splits involve both contextual features like lastMotionLocation, sensorElTime-Bathroom, and prevDominantSensor2, as well as transferred signals such as source_pred, label_distribution_6, and entropy. In the decision rules, source_pred refers to the hard class label predicted by the source model when applied to target instances, while label_distribution_6 denotes the soft prediction probability assigned to class 6 by the same source model.Notably, source_pred governs several key partitions, guiding decisions toward classes like 1, 3, 5, and 7, while uncertainty signals such as entropy, help refine predictions in less confident regions. These patterns confirm that UTA-DT integrates sensor context with source-derived knowledge in a transparent and interpretable way. A visualization of the complete decision tree is provided in Figure 3.3.

To complement the global insights, we also examine the local explainability of UTA-DT by tracing the decision path for an individual instance. As shown in Figure 3.4, the model first checks whether sensorElTime-OutsideDoor is less than or equal to -0.80, indicating that outdoor sensor activity plays a primary role in the initial partitioning. The next condition involves source_pred

```
=== Global Explanation of Transferred Model ===
Global Explanation (Decision Tree Rules):
|--- sensorElTime-OutsideDoor <= -0.80
|   |--- lastMotionLocation <= 0.32
|   |   |--- label_distribution_6 <= -0.14
|   |   |   |--- class: 2
|   |   |--- label_distribution_6 >  -0.14
|   |   |   |--- class: 0
|   |--- lastMotionLocation >  0.32
|   |   |--- sensorElTime-OutsideDoor <= -0.84
|   |   |   |--- class: 8
|   |   |--- sensorElTime-OutsideDoor >  -0.84
|   |   |   |--- class: 4
|--- sensorElTime-OutsideDoor >  -0.80
|   |--- source_pred <= 0.41
|   |   |--- sensorElTime-Bathroom <= -0.28
|   |   |   |--- entropy <= -0.98
|   |   |   |   |--- class: 6
|   |   |   |--- entropy >  -0.98
|   |   |   |   |--- class: 1
|   |   |--- sensorElTime-Bathroom >  -0.28
|   |   |   |--- prevDominantSensor2 <= -0.69
|   |   |   |   |--- class: 3
|   |   |   |--- prevDominantSensor2 >  -0.69
|   |   |   |   |--- class: 5
|   |--- source_pred >  0.41
|   |   |--- lastMotionLocation <= -0.19
|   |   |   |--- class: 9
|   |   |--- lastMotionLocation >  -0.19
|   |   |   |--- class: 7
```

Figure 3.2: Global Explanation rules - 10 classes

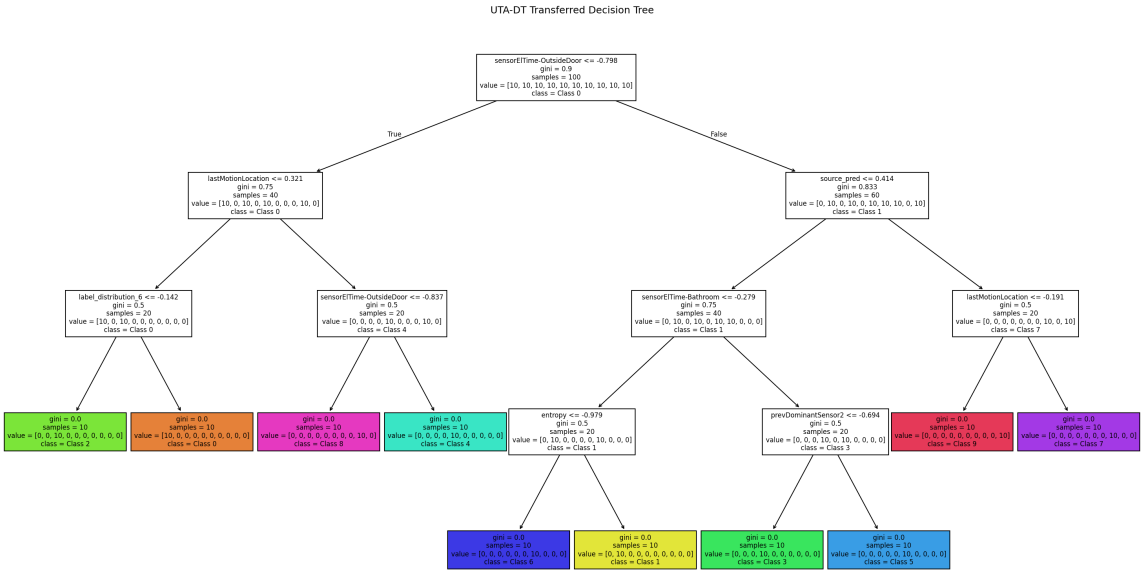


Figure 3.3: UTA-DT Decision Tree Plot

```
=== Local Explanation for the First Instance ===
Decision path for the instance:
Node 0: if sensorElTime-OutsideDoor <= -0.80
Node 8: if source_pred <= 0.41
Node 16: if lastMotionLocation <= -0.19
Node 18: Leaf node
```

Figure 3.4: Local Explanation for first instance

is less or equal 0.41, integrating transferred knowledge from the source model to refine the decision. Finally, the model considers lastMotionLocation is less or equal -0.19 before reaching a leaf node, confirming the class prediction. This interpretable path reveals how both contextual sensor inputs and transferred signals jointly influence the model's decision at an individual level, enabling traceability and trust in high-stakes environments.

Figure 3.5 shows the Partial Dependence Plots (PDPs) of two influential features, *source_pred* and *lastMotionLocation*, with respect to the predicted probability of class 7 corresponding to cooking lunch, in the target domain. The PDP for *source_pred* reveals that as the predicted class from the source model approaches 7, the likelihood of predicting class 7 in the target model increases sharply. This suggests that the transferred model strongly leverages the source model's hard prediction signal

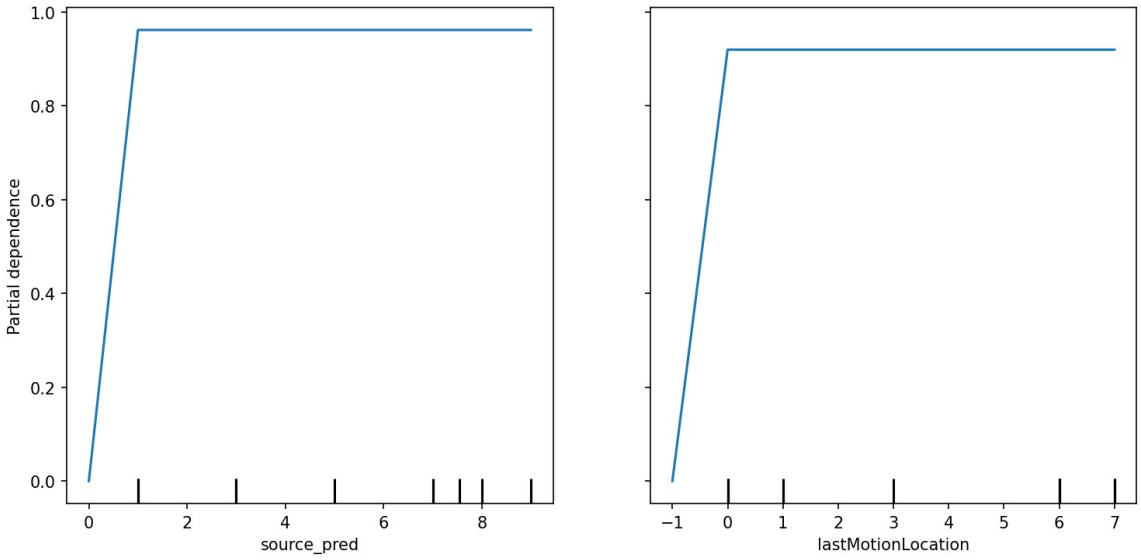


Figure 3.5: PDP plot for class 7

when it aligns with class 7, reinforcing its confidence in this classification.

Similarly, the PDP for *lastMotionLocation* reveals that motion detected in specific areas significantly increases the probability of class 7, highlighting the importance of spatial context. These patterns confirm that the model's predictions are informed by both transferred knowledge and relevant sensor information from the target environment. Such insights enhance transparency and support the broader goal of developing explainable and trustworthy AI for smart building applications.

Conclusion

In this work, we introduced UTA-DT, an interpretable and privacy-preserving domain adaptation framework based on decision trees. UTA-DT enables knowledge transfer from a label-rich source domain to a label-scarce target domain through uncertainty-aware feature augmentation. By incorporating soft predictions, hard labels, entropy, and margin from a pre-trained source model, our method constructs an enriched target feature space that supports more generalizable and robust decision-making. This is achieved without accessing raw source data, making UTA-DT well-suited

for deployment in privacy-sensitive environments such as smart homes. Comprehensive experiments on increasingly complex activity recognition tasks show that UTA-DT consistently outperforms baseline models, validating the effectiveness of its feature-level transfer strategy. Additionally, UTA-DT retains full interpretability, allowing end-users to inspect global decision rules and instance-level predictions through integrated explainability techniques. By supporting interpretable and data-efficient learning, UTA-DT contributes toward enabling intelligent building systems that optimize energy usage based on occupant behavior, without compromising privacy or transparency.

Chapter 4

Semi-Supervised Mixture of Probabilistic Principal Component Analyzers for Modeling Human Behavior

The source-free framework introduced in Chapter 3 addresses privacy constraints but still relies primarily on supervised signals from labeled data, which remain scarce and costly to obtain in smart-building environments. At the same time, deterministic models are often insufficient for capturing the complex, high-dimensional patterns characteristic of IoT sensor data, particularly when labeled samples are limited. This chapter therefore introduces a semi-supervised probabilistic framework based on a neural reformulation of MPPCA that models the latent structure of the data and supports effective representation learning under minimal supervision. By transitioning from purely supervised to semi-supervised learning, this chapter advances the thesis toward more label-efficient and scalable building analytics.

4.1 Introduction

As technology advances, human interaction with their surroundings has fundamentally changed. Today, individuals spend approximately 90% of their daily lives indoors, relying on technology for everyday tasks without needing to step outside [138, 139]. This shift towards an increasingly indoor lifestyle has intensified the demand for comfortable, energy-efficient indoor spaces. Yet, maintaining such environments comes at a significant cost: the building sector is one of the largest consumers of energy, accounting for over 32% of global electricity consumption, with more than half of this energy dedicated to Heating, Ventilation, and Air Conditioning (HVAC) systems [19–21, 59, 138, 140]. Additionally, buildings contribute to nearly one-third of global greenhouse gas emissions, making them a key area for sustainability efforts [138]. Amid growing concerns over energy and environmental crises and the urgency of sustainable development, energy-efficient buildings have become a focal point of research and innovation [41, 128]. Studies indicate that inefficient management strategies are a major contributor to excessive energy use in buildings [141]. In response, Occupancy Estimation (OE) and Activity Recognition (AR) have emerged as key techniques for enhancing energy efficiency and HVAC optimization [19–21, 27, 59, 119]. By detecting and analyzing occupant presence and behavior, these technologies enable real-time adjustments to energy distribution, HVAC scheduling, and indoor environmental quality [70, 110, 142]. Since occupant activities directly influence energy consumption, understanding behavioral patterns is essential for reducing energy waste, improving air quality, and maintaining thermal comfort [117, 143]. Indeed, leveraging intelligent automation in building management can lead to significant energy savings, lower operational costs, and reduced environmental impact [19, 41]. For instance, dynamic HVAC adjustments based on real-time occupancy data can minimize unnecessary heating or cooling, thereby promoting sustainability while ensuring occupant well-being [140, 144].

In recent years, data-driven approaches have advanced significantly, leveraging large datasets to enhance model performance and generalization. Traditional methods for activity recognition (AR) and occupancy estimation (OE) have primarily relied on supervised learning [20, 27]. While effective, these techniques require extensive labeled data to train accurate models. The proliferation of low-cost and high-precision sensors has facilitated the deployment of large-scale sensor networks in

various environments, including smart buildings, industrial monitoring, and intelligent transportation systems. These sensors-ranging from motion detectors and CO_2 sensors to passive infrared (PIR) and environmental monitoring devices-continuously generate vast amounts of real-time data, significantly improving the granularity and accuracy of AR and OE systems. Despite the advantages of sensor-based data collection, a key challenge remains: labeling the collected data. Unlike the automated nature of sensor data acquisition, labeling is manual, time-consuming, and labor-intensive, often requiring expert intervention to annotate events accurately [21]. Additionally, in complex environments where multiple sensors interact, ground truth labels may be ambiguous or prone to inconsistencies.

At the same time, smart building datasets are often large-scale and highly complex [38], generated from the integration of numerous sensors within Internet of Things (IoT) networks. These datasets typically comprise multi-dimensional, heterogeneous data streams that capture a wide range of environmental and occupancy-related variables. While such data richness holds great potential for developing accurate and adaptive models, it also introduces significant challenges in data analysis and interpretation. The intricate structure of complex datasets often results in non-linear relationships among features, making it difficult to identify meaningful patterns or correlations. Non-linearity emerges from the interaction of multiple factors, such as dynamic occupant behavior, varying environmental conditions, and sensor noise, which collectively contribute to the complexity of the data. This non-linearity necessitates the use of advanced machine learning techniques capable of capturing and modeling such relationships. Moreover, the complexity of high dimensional smart building datasets can exacerbate the risk of errors and biases during analysis. For instance, redundant or irrelevant features may obscure important patterns. Ensuring the generalizability of models is particularly critical in smart building research, as it enables robust performance across varying environmental conditions, building layouts, and occupancy patterns.

In light of these challenges, there is a growing need for innovative solutions that can reduce reliance on labeled data while maintaining or even improving model accuracy, particularly when dealing with complex datasets. This has led to the emergence of semi-supervised learning (SSL) methods as a promising alternative. By leveraging both labeled and unlabeled data, SSL aims to bridge the gap between supervised and unsupervised learning, making it possible to train models

with fewer labeled examples while still capturing the underlying data distribution. Traditional SSL methods, such as self-training, co-training, and graph-based approaches, have shown promise in various domains [145]. Self-training iteratively labels unlabeled data and retrains itself using its own predictions as pseudo-labels. However, this approach suffers from several limitations. A major issue is confirmation bias, where the model reinforces its own incorrect predictions, leading to a degradation in performance over iterations [146]. This is particularly problematic when the initial model is poorly calibrated or when the labeled data is insufficient to capture the underlying data distribution [147]. Co-training trains separate models on multiple feature sets that label data for one another, but its effectiveness depends on the often unrealistic assumption of conditional independence [148, 149]. The lack of truly independent views in practice further limits its applicability [150]. Graph-based methods, such as label propagation, are theoretically appealing for semi-supervised learning but face significant practical limitations. Their performance is highly sensitive to graph construction parameters, including similarity metrics and hyperparameters like neighborhood size or kernel bandwidth [151, 152]. Moreover, they scale poorly to large datasets due to the high memory and computational demands of graph construction [153]. These methods also face a trade-off between preserving label accuracy and ensuring smooth label propagation, often resulting in suboptimal classification when the graph poorly reflects the data structure. These challenges underscore the need for more robust and scalable SSL approaches for complex, real-world datasets.

Recently, researchers have increasingly turned to autoencoders for semi-supervised learning tasks, capitalizing on their ability to learn meaningful latent representations of data. Autoencoders operate by compressing input data into a lower-dimensional latent space via an encoder and then reconstructing the original input through a decoder. This process not only reduces dimensionality but also captures essential features, making autoencoders highly effective for tasks with limited labeled data, as they can leverage the structure of unlabeled data to enhance model performance. Their ability to learn robust representations also makes them particularly well-suited for tasks such as pretraining and fine-tuning.

At the same time, Gaussian Mixture Models (GMMs) excel at modeling complex, non-linear, and multi-modal data distributions by assuming that data is generated from a mixture of Gaussian

components. As a statistical model, however, GMMs cannot be directly used for pretraining or fine-tuning in the way that neural networks like autoencoders can. Unlike models that can incrementally update their parameters, GMMs do not inherently retain previously learned knowledge when exposed to new data. Despite these limitations, GMMs remain a powerful tool to learn patterns from unlabeled data and are widely used for clustering and density estimation. While GMMs have been extensively studied, other probabilistic mixture models, such as Mixture of Probabilistic Principal Component Analyzers(MPPCA), have received less attention. MPPCA is closely related to GMMs, as both are probabilistic frameworks designed to model complex data distributions. However, MPPCA extends GMMs by incorporating a dimensionality reduction mechanism. Specifically, MPPCA models each mixture component as a Probabilistic Principal Component Analysis (PPCA) model, capturing data in a lower-dimensional latent space. This allows MPPCA to represent multi-modal distributions like GMMs while also modeling piecewise linear structures in high-dimensional data. In essence, MPPCA can be viewed as a generalization of GMMs, where each Gaussian component is constrained to a lower-dimensional subspace, making it particularly effective for capturing both global and local structures in complex datasets. Despite these advantages, MPPCA, like GMMs, lacks the flexibility for efficient pretraining and fine-tuning. Additionallly, a major limitation of GMMs and MPPCA is the singularity problem, which arises when computing the inverse of covariance matrices during the Expectation-Maximization (EM) training process, particularly in high-dimensional settings. This issue can lead to unstable training and premature termination, hindering the model's ability to generalize effectively.

In the light of these challenges, we propose to model MPPCA with a multilayer Neural Network (MLN), referred to as NN-MPPCA [38], since deep learning models offer greater flexibility and adaptability for fine-tuning tasks. This approach integrates the probabilistic strengths of MPPCA with the flexibility of neural networks, enabling efficient pretraining and fine-tuning while preserving the ability to model complex, non-linear data structures. By replacing the EM algorithm with a backpropagation-based optimization and incorporating a loss function that combines reconstruction error, negative reconstruction probability, and a penalty term to mitigate singularity issues, NN-MPPCA enables stable training and better generalization for non-linear data distributions. Training stops when further minimizing the combined weighted error is impossible, ensuring

a good data fit. This study focuses on enhancing NN-MPPCA, originally an unsupervised clustering technique, by transforming it into a semi-supervised classification model through the incorporation of a fine-tuning step with minimal labeled data. This transformation allows the new model, referred to as Semi-Supervised Mixture of Probabilistic Principal Component Analyzers (Semi-MPPCA), to directly assign clusters to specific activities and occupancy levels, enhancing interpretability and practical applicability. To achieve this, MPPCA is transformed to a neural form, and trained using backpropagation to extract meaningful feature representations from the abundant unlabeled data. The pretrained model then serves as a feature extractor for the classifier. Subsequently, the input and hidden layers are frozen, and only the output layer parameters are refined using few labeled data. This approach leverages the advantages of both unsupervised and supervised learning with only a few labeled data, improving classification accuracy while maintaining the ability to model complex, high-dimensional data distributions. We also observed that incorrect initial cluster assignments cause training delays and local minima issues. To address this, we introduce a novel weight-swapping technique for the last layer, improving performance. Our Semi-MPPCA model outperforms state-of-the-art models when trained with minimal labeled data.

This chapter explores how our proposed Semi-MPPCA framework effectively models nonlinearity in complex datasets while overcoming the limitations of traditional probabilistic methods. In this study, we assume we have no access to labeled data from related tasks or domains and instead focus on a single environment, where we have abundant unlabeled data and only a handful of labeled examples. Our experiments show that incorporating weight-swapping strategies and leveraging semi-supervised learning significantly enhance model accuracy and robustness, particularly outperforming state-of-the-art models when trained with minimal labeled data. This makes Semi-MPPCA a promising alternative for high-dimensional, nonlinear data modeling with very little labeled data.

In this chapter, we present several key contributions:

- We adapt the Mixture of Probabilistic Principal Component Analyzers (MPPCA) framework into a neural network form (NN-MPPCA) by replacing the Expectation-Maximization (EM) algorithm with backpropagation for training. This adaptation allows the framework to function as a pre-trained model, and we further fine-tune the output layer to enable classification

tasks.

- A weight-swapping technique is introduced to address cluster-label misalignment issues, ensuring that learned representations align with their corresponding class labels.

- The proposed model - Semi-MPPCA - is designed to effectively handle both Activity Recognition (AR) and Occupancy Estimation (OE) datasets, marking the first application of this approach in these domains. Extensive experimental evaluations on real-world AR and OE datasets demonstrate its superior performance with minimal labeled data, outperforming state-of-the-art occupancy detection algorithms.

The remainder of this chapter is structured as follows. Section 2 reviews related work, while Section 3 outlines key concepts and details the proposed methodology. Section 4 presents and discusses the results. Section 5 concludes the study and suggests potential directions for future research.

4.2 Methodology

This section outlines the foundational concepts relevant to our study. We begin by formally defining the semi-supervised learning (SSL) problem, followed by an exploration of the neural network-based extension of the Mixture of Probabilistic Principal Component Analyzers (MPPCA). We then detail our enhancements that enable its effective use in a semi-supervised setting. For completeness, an in-depth explanation of Probabilistic Principal Component Analysis (PPCA), MPPCA, and the Expectation-Maximization (EM) training procedure is provided under the Theoretical Background section in the Appendix A.

4.2.1 Problem Formulation

In this work, we focus on the semi-supervised learning (SSL) framework for multi-class classification tasks. Let $\mathbf{X} = \{\mathbf{X}_L, \mathbf{X}_U\}$ denote the entire dataset, where $\mathbf{X}_L = \{(x_i, y_i)\}_{i=1}^{L}$ represents a small labeled subset and $\mathbf{X}_U = \{x_i\}_{i=1}^{U}$ corresponds to a larger unlabeled subset, with $L \ll U$.

The dataset contains K classes, and the labels $y_i \in \{1, 2, \ldots, K\}$ indicate the class to which each labeled sample belongs.

The goal of SSL is to leverage both labeled and unlabeled data to learn a model parameterized by u. This can be formulated as the following optimization problem [83]:

$$\min_{u} \Big\{ \underbrace{\sum_{(x,y)\in \mathbf{X}_L} \mathcal{L}_s(x, y; u)}_{\text{Supervised Loss}} + \underbrace{\sum_{x\in \mathbf{X}_U} \mathcal{L}_u(x; u)}_{\text{Unsupervised Loss}} + \underbrace{\sum_{x\in \mathbf{X}} \mathcal{R}(x; u)}_{\text{Regularization}} \Big\}, \tag{23}$$

where $\mathcal{L}_s$ is the supervised loss, typically cross-entropy loss for multi-class classification, $\mathcal{L}_u$ is the unsupervised loss that encourages consistent predictions on unlabeled data, and $\mathcal{R}$ is a regularization term that imposes additional constraints, such as consistency regularization or entropy minimization.

The semi-supervised approach leverages the underlying structure of the unlabeled data to refine decision boundaries, allowing the model to generalize better even with limited labeled samples. This makes SSL particularly effective in scenarios where labeled data is scarce but a large pool of unlabeled data is available.

4.2.2 Neural network-based mixture of probabilistic principal component analyzers (NN-MPPCA)

Neural Network-based Mixture of Probabilistic Principal Component Analyzers (NN-MPPCA) is a hybrid approach that integrates the MPPCA framework with a neural network estimation process to model responsibilities, enabling better parameter initialization and reducing the risk of convergence to poor local optima. In contrast to the standard MPPCA, NN-MPPCA employ a multilayer perceptron (MLP) as the *estimation network* to model the MPPCA framework. The parameters of the MPPCA are updated through the training process of this estimation network and optimized through back-propagation. The input size of the estimation network is determined by the dimensionality of the observation data, while the output size corresponds to the number of components M in the MPPCA model. For an MPPCA with M components, the outputs of the estimation network are M-dimensional responsibility vectors $\mathbf{R}_n$. Since the responsibility variables R_{ni} must satisfy

$\sum_{i=1}^{M} R_{ni} = 1$, they are computed by applying the softmax function to the output vector $\mathbf{p}_n$. The process of obtaining the responsibility vectors is formulated as follows:

$$\mathbf{p}_n = \text{MLP}(\mathbf{x}_n, \boldsymbol{\theta}) \tag{24}$$

$$\mathbf{R}_n = \text{softmax}(\mathbf{p}_n) \tag{25}$$

where $\mathbf{x}_n$ is the observed input, $\mathbf{p}_n$ is the output of the MLP, and $\boldsymbol{\theta}$ represents the trainable parameters of the estimation network. This formulation ensures that the responsibility vectors $\mathbf{R}_n$ are properly normalized and can be used to update the MPPCA parameters during training.

NN-MPPCA model is trained using mini-batches of N samples. The training process begins with an initialization step to avoid suboptimal local minima, which is a common issue with random initialization in standard MPPCA. Instead, we employ *k-means clustering* to initialize the model parameters. Each sample is temporarily assigned to a cluster, and the initial responsibility values R_{ni} for observation $\mathbf{x}_n$ are determined by the output layer of the estimation network using:

$$R_{ni} = \text{softmax}(\text{estimation\_network}(\mathbf{x}_n)), \tag{26}$$

where R_{ni} represents the responsibility of the i-th mixture component for the n-th sample. Using R_{ni}, the mixture weights π_i and means $\boldsymbol{\mu}_i$ are initialized same as Equations 75, 76 from MPPCA. The covariance matrix, $\mathbf{S}_i$, is calculated as Equation 77.

Let $\{\lambda_{i1} \geq \lambda_{i2} \geq \cdots \geq \lambda_{id}\}$ and $\{\mathbf{u}_{i1}, \mathbf{u}_{i2}, \ldots, \mathbf{u}_{id}\}$ be the eigenvalues and eigenvectors of $\mathbf{S}_i$, respectively. For a given number of principal components q, the noise variance σ_i and loading matrix $\mathbf{W}_i$ are initialized as:

$$\sigma_i = \frac{1}{d-q} \sum_{j=q+1}^{d} \lambda_{ij}, \quad \mathbf{W}_i = \mathbf{U}_{i,q}(\boldsymbol{\Lambda}_{i,q} - \sigma_i^2 \mathbf{I})^{1/2}, \tag{27}$$

$$\text{where } \mathbf{U}_{i,q} = \Big[\mathbf{u}_{i1}, \mathbf{u}_{i2}, \ldots, \mathbf{u}_{iq}\Big] \text{ and}$$

$$\boldsymbol{\Lambda}_{i,q} = \text{diag}(\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{iq})$$

These equations serve as the initial estimates only in the first iteration and are updated in subsequent iterations.

The total loss function for a batch of N samples is defined as:

$$L(\boldsymbol{\theta}) = \frac{\eta_1}{N}\sum_{n=1}^{N} a(\mathbf{x}_n) + \frac{\eta_2}{N}\sum_{n=1}^{N} \mathcal{L}(\mathbf{x}_n, \mathbf{x}'_n) + \eta_3 P(\mathbf{M}), \tag{28}$$

where:

- $a(\mathbf{x}_n)$ is the negative log-likelihood of $\mathbf{x}_n$, given by:

$$a(\mathbf{x}_n) = -\log\left\{\sum_{i=1}^{M} \pi_i (2\pi)^{-d/2} |\mathbf{C}_i|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu}_i)^T \mathbf{C}_i^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_i)\right)\right\} \tag{29}$$

  with

$$\mathbf{C}_i^{-1} = \frac{1}{\sigma_i^2}\left(\mathbf{I} - \mathbf{W}_i \mathbf{M}_i^{-1} \mathbf{W}_i^T\right) \tag{30}$$

- $\mathcal{L}(\mathbf{x}_n, \mathbf{x}'_n)$ is the reconstruction error, computed as:

$$\mathbf{x}'_{ni} = \mathbf{W}_i (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \mathbf{W}_i^T (\mathbf{x}_n - \boldsymbol{\mu}_i) + \boldsymbol{\mu}_i, \tag{31}$$

$$\mathbf{x}'_n = R_{ni} \mathbf{x}'_{ni}, \tag{32}$$

$$\mathcal{L}(\mathbf{x}_n, \mathbf{x}'_n) = \|\mathbf{x}_n - \mathbf{x}'_n\|. \tag{33}$$

- $P(\mathbf{M})$ is a penalty term to avoid singularity issues when computing the inverse of $\mathbf{M}$:

$$P(\mathbf{M}) = \sum_{i=1}^{M}\sum_{j=1}^{d} \frac{1}{M_{ijj}}. \tag{34}$$

The loss function is minimized using backpropagation, with heuristic coefficients η_1, η_2, and η_3 balancing the contributions of the three loss components. After each training epoch, the responsibility values R_{ni} are updated directly by the estimation network, while the other parameters π_i, $\boldsymbol{\mu}_i$, $\mathbf{W}_i$, and σ_i are recomputed using the updated R_{ni} according to Equations 75, 76, 78, and 79 in the Appendix A. These newly updated parameters are then used to compute the loss value for the

next iteration, ensuring an iterative refinement of the model. This iterative process continues until convergence, ensuring that the NN-MPPCA model achieves optimal performance.

4.2.3 Semi Supervised Mixture of Probabilistic Principal Component Analyzers(Semi-MPPCA)

Our approach (Semi-MPPCA) leverages the Neural Network Mixture of Probabilistic Principal Component Analysis (NN-MPPCA) as a pretrained feature extractor network, enabling effective learning from limited labeled data. The NN-MPPCA model is first trained in an unsupervised manner on the available unlabeled data, learning underlying feature representations and capturing local low-dimensional subspaces within each cluster. This pre-training step ensures that the model initializes its parameters based on the intrinsic structure of the data, which is particularly beneficial when labeled data is scarce.

However, since the clustering process is unsupervised, the resulting cluster assignments may not always align with the true class labels. To address this, we introduce a cluster-to-label weight swapping mechanism that realigns the model's output layer to reflect the correct label assignments. After the initial unsupervised training phase, we evaluate the majority class associated with each cluster and use the Hungarian algorithm to determine the optimal mapping between predicted clusters and true labels. The weight matrix of the output layer is then restructured by swapping its columns according to this mapping, ensuring that the neural network correctly associates the learned feature representations with their corresponding class labels. A detailed description of the weight-swapping algorithm is provided as follows. Let:

- $\mathbf{y}^{\text{true}} = (y_1, y_2, ..., y_N)$ be the ground truth labels.
- $\mathbf{y}^{\text{pred}} = (\hat{y}_1, \hat{y}_2, ..., \hat{y}_N)$ be the predicted cluster assignments.
- C be the number of clusters.

We construct a confusion matrix $\mathbf{M} \in \mathbb{R}^{C \times C}$, where each entry M_{ij} represents the count of samples assigned to cluster j while their true label is i:

$$M_{ij} = \sum_{n=1}^{N} \mathbb{1}(y_n = i, \hat{y}_n = j) \tag{35}$$

where $\mathbb{1}(\cdot)$ is the indicator function.

The Hungarian algorithm then finds an optimal mapping $\phi : \{0, ..., C-1\} \to \{0, ..., C-1\}$ by solving:

$$\arg\max_{\phi} \sum_{i=0}^{C-1} M_{i,\phi(i)} \tag{36}$$

This ensures the best possible realignment of clusters to true labels.

Once we obtain the optimal mapping ϕ, we correct the neural network's output layer. The neural network's final layer consists of a weight matrix $W \in \mathbb{R}^{d \times C}$, where d is the number of neurons in the preceding layer, and each column $W_{\cdot,j}$ corresponds to class j.

If the predicted classes $(\hat{y}_1, ..., \hat{y}_N)$ are misaligned, we swap the columns of W according to ϕ:

$$W'_{\cdot,j} = W_{\cdot,\phi(j)} \tag{37}$$

This ensures that the final layer aligns correctly with the corrected pseudo-labels.

Once the correct cluster-label assignments are established through weight swapping, we proceed with the fine-tuning phase. In this step, the input and hidden layers of the pretrained NN-MPPCA model are frozen, while the output layer's parameters remains trainable. The model is then trained using a small set of labeled data, allowing it to refine its predictions while preserving the structural information captured during unsupervised learning. After swapping the final layer weights, we fine-tune the model using labeled samples. The updated model is trained using a cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{C} \mathbb{1}(y_n = c) \log p(c|x_n; \theta) \tag{38}$$

where $p(c|x_n; \theta)$ is the softmax probability output by the neural network for class c.

To ensure stable training, we apply:

- Learning rate scheduling: reducing the learning rate when validation loss plateaus.

- **E**arly stopping: halting training when validation performance stagnates to prevent overfitting, especially given the limited labeled data..

A key innovation of our approach is the extension of MPPCA, which is traditionally an unsupervised clustering model optimized via the Expectation-Maximization (EM) algorithm, into a classification framework through neural network integration and fine-tuning. Unlike conventional statistical models, which have a fixed structure and limited adaptability, our neural network-based approach allows for continuous parameter updates, making it more flexible for various classification tasks.

This methodology ensures that the model leverages unsupervised learning for feature extraction, corrects misaligned cluster-label mappings, and fine-tunes its predictions using limited labeled data, resulting in improved classification accuracy. Semi-MPPCA collectively ensure robust and generalizable performance in the semi-supervised learning setting. The architecture of the neural network is visually depicted in Figure 4.1, offering a clear representation of its layered structure and flow of information. The neural network architecture comprises two hidden layers with sizes 50 and 15, respectively. The output layer, corresponding to the number of target classes, employs a softmax activation function to produce class probabilities. Dropout layers, represented by circles containing a cross, are incorporated to mitigate overfitting by randomly deactivating a subset of neurons during training. The sequential steps involved in the Semi-MPPCA model, from pretraining with unlabeled data to fine-tuning with a limited set of labeled samples, are outlined in Figure 4.2, providing a step-by-step understanding of the model's workflow.

Additionally, Table 4.1 presents a comparative analysis of PCA, PPCA, MPPCA, NN-MPPCA, and semi-supervised NN-MPPCA, emphasizing their fundamental differences and underlying methodologies.

Table 4.1: Key Differences Between PCA, PPCA, MPPCA, NN-MPPCA, and Semi-MPPCA

Feature	PCA	PPCA	MPPCA	NN-MPPCA	Semi-MPPCA
Model Type	Deterministic	Probabilistic	Mixture of Probabilistic Models	Neural Network	Neural Network
Noise Handling	No	Yes	Yes	Yes	Yes
Handles Missing Data	No	Yes	Yes	Yes	Yes
Clustering Ability	No	No	Yes	Yes	Yes
Singularity Error Handling	No	Yes	No	Yes	Yes
Optimization Method	Eigen-decomposition	EM Algorithm	EM Algorithm	Back-propagation	Back-propagation & Weight swapping
Task Type	Dimensionality Reduction	Dimensionality Reduction	Clustering	Clustering	Classification

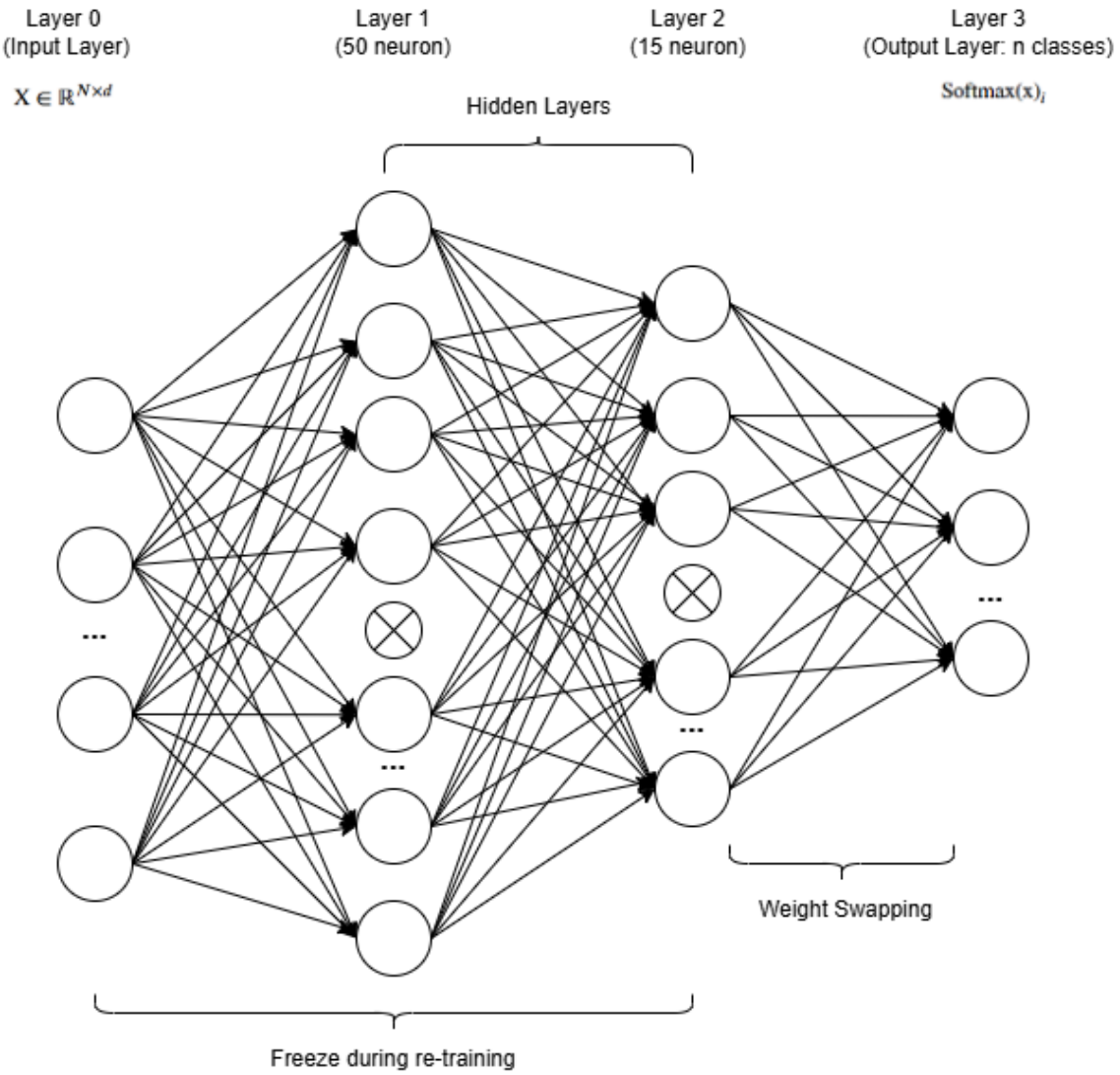


Figure 4.1: Semi-MPPCA Architecture

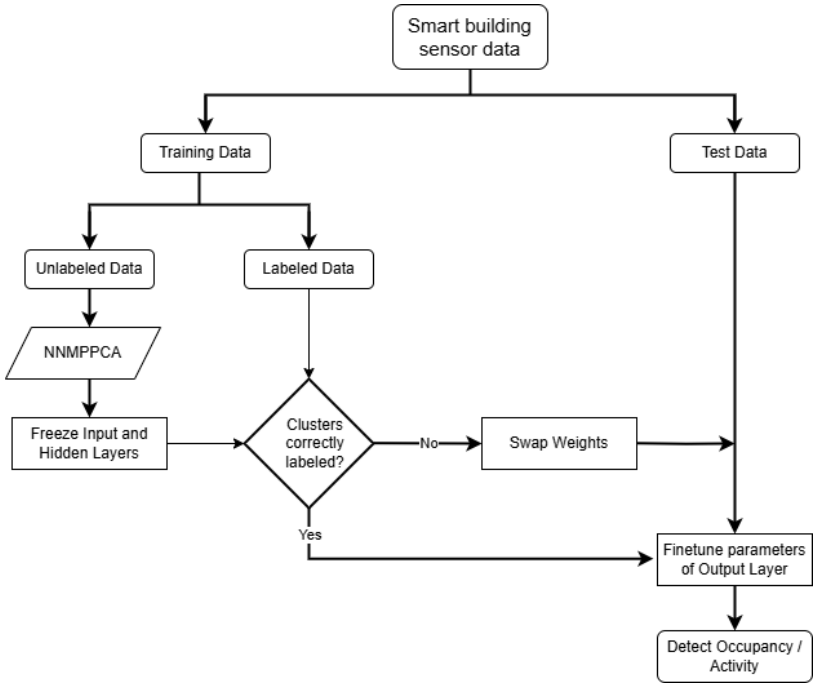


Figure 4.2: Flowchart for logic of Semi-MPPCA

4.3 Experimental setup and results

4.3.1 Datasets

Activity Recognition Dataset

Publicly available activity recognition datasets are relatively scarce, as many studies rely on proprietary datasets collected in controlled environments. This limitation makes real-world datasets particularly valuable for developing and evaluating machine learning models for human activity recognition. This study utilizes the publicly available dataset from the Washington State University (WSU) Center for Advanced Studies in Adaptive Systems (CASAS) [137]. The dataset contains labeled sensor event data collected from multiple smart indoor environments, each occupied by a single resident with distinct spatial layouts and behavioral routines. This reflects real-world deployment scenarios where AR models must generalize across heterogeneous living spaces. This dataset was collected using a variety of sensors, including motion sensors, door contact sensors, temperature sensors, and light switch sensors, capturing a wide range of activities in a smart home environment. Initially, we focused on predicting three key activities that are particularly relevant to

energy management: watching TV, toileting, and cooking breakfast. These activities were chosen due to their distinct energy consumption patterns and their frequent occurrence in daily routines. After establishing a baseline with these three labels, we expanded our analysis to include a more comprehensive set of five activities: watching TV, toileting, cooking breakfast, dressing, and sleeping. This extension allowed us to evaluate the performance of our model across a broader range of activity classes, providing valuable insights into its robustness and generalizability in more diverse and complex scenarios. As a non-intrusive, real-world dataset, it is particularly useful for research on human activity recognition, smart home automation, and assistive technology. Compared to vision-based approaches, this dataset provides a privacy-conscious alternative while supporting machine learning models for behavior analysis. After preprocessing, each instance is represented by a 32-dimensional input vector consisting of spatial and behavioral features derived from sensor activations, including activity frequency and duration across rooms, area transitions, and the number of distinct sensors triggered. Temporal attributes such as the hour of the day or weekday were removed to avoid environment-specific biases and improve generalization. To ensure class balance, we randomly sampled 2,000 instances from each activity class and reset the index.

Occupancy Estimation Dataset

Publicly available occupancy estimation datasets are limited, as many studies rely on data collected from specific buildings or sensor networks. We used a private dataset [21, 27, 119] collected from office spaces at the Grenoble Institute of Technology. The dataset comprises six days of labeled data obtained through a comprehensive sensor network designed for occupancy monitoring. This network includes two video cameras for recording real-time occupancy counts and activities, as well as an ambiance sensing system that captures temperature, relative humidity (RH), motion events, CO2 concentration, power consumption from three laptops, door and window positions, and acoustic pressure using a microphone. Data collection is event-driven, and transmitted via the ENOCEAN protocol upon significant changes. A centralized database with a web interface continuously retrieves and stores data from various sources.

Two classification tasks were defined for the occupancy estimation problem. The first is a binary classification task for occupancy detection, where the model predicts presence versus absence: class

0 corresponds to no occupant, and class 1 represents one or more occupants. The second is a multi-class classification task, where the output is discretized into three categories based on occupant count: (0) absence, (1) a single occupant, and (2) multiple occupants ($\geq$ 2). Given the limitations of environmental sensors in reliably distinguishing fine-grained occupancy levels, we model the occupancy estimation task as a multi-class classification problem. This discretization reflects the practical resolution of sensor signals and enables more robust learning, avoiding ambiguous boundaries between close occupant counts. For example, small fluctuations in CO_2 or noise levels may not reliably differentiate between four and five individuals in a room.

The Occupancy Estimation dataset comprises 1,068 labeled instances which includes 13 input features coming from the mentioned sensors commonly found in low-cost smart building deployments. These features serve as the model inputs, while the output represents occupancy levels. To mitigate class imbalance and ensure fair evaluation across categories, we randomly selected 772 samples per class, resulting in a balanced dataset of 1544 instances for the binary scenario and 2,316 instances for the 3-class scenario. This approach helps prevent model bias toward the majority class and supports a more robust assessment of performance, particularly in the three-class occupancy classification setting.

4.3.2 Metrics

In evaluating our models, we opted to use the F1 score as the primary metric for assessing performance. While accuracy offers a straightforward measure of correctly classified instances, it can sometimes provide an overly simplistic view, especially when subtle trade-offs between precision and recall are critical. The F1 score, defined as the harmonic mean of precision and recall, strikes a balance between these two metrics by considering both false positives and false negatives. This makes it a more comprehensive indicator of model performance, even when the datasets are balanced, as it captures the model's ability to correctly identify positive cases without overlooking misclassifications. To compute the F1 score, we first calculate two key metrics: Precision (P) and Recall (R). Precision measures the ratio of correctly predicted positive instances to the total number of predicted positive instances, while Recall represents the proportion of correctly predicted positive instances out of all actual positive instances. These metrics are defined as follows:

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{39}$$

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{40}$$

The F1 score, which is the harmonic mean of Precision and Recall, provides a balanced assessment of a model's accuracy by accounting for both false positives and false negatives. It is given by:

$$\text{F1 Score} = \frac{2 \times P \times R}{P + R} \tag{41}$$

By integrating both precision and recall into a single value, the F1 score provides a nuanced reflection of our model's predictive capabilities, guiding us in making informed comparisons and improvements.

However, we have also included graphs for accuracy of our models for comparison. Accuracy measures the proportion of correct predictions and is calculated as:

$$\text{Accuracy} = \frac{\text{Correct predictions}}{\text{Total predictions}} \times 100 \tag{42}$$

4.3.3 Baseline models

To evaluate the effectiveness of our model on real-world activity recognition (AR) and occupancy estimation (OE) datasets, we conduct a comparative analysis against several state-of-the-art models. A brief overview of these baseline models is provided below.

- Support Vector Machine (SVM): A supervised learning algorithm that classifies data by constructing an optimal hyperplane to maximize the margin between classes in a high-dimensional space. The model was configured with the default hyperparameters: a radial basis function (RBF) kernel, regularization parameter $C = 1.0$, gamma = 'scale', and a convergence tolerance of $1e^{-3}$.

- k-Nearest Neighbors (KNN): A non-parametric, supervised learning method that classifies a

data point based on the majority class of its k nearest neighbors. Distances between the query point and all other points in the dataset are calculated, and the k closest points determine the class label. The default settings were applied, including $k = 5$ (number of neighbors), metric = 'minkowski', and weights = 'uniform' (equal weight assigned to each neighbor).

- Decision Tree (DT): A non-parametric supervised learning algorithm that uses a hierarchical structure, applying if-else rules at internal nodes and making predictions at the leaf nodes. The model employed default parameters: criterion = 'gini', splitter = 'best', max_depth = None, and min_samples_split = 2.

- AdaBoost: An ensemble method that iteratively combines weak classifiers, adjusting weights to emphasize misclassified instances and progressively improving model accuracy. The default hyperparameters were adopted: base estimator as DecisionTreeClassifier with max_depth = 1 (stump), learning rate set to 1.0, and n_estimators = 50.

- Multi Layer Perceptron (MLP): A deep learning algorithm consisting of an input layer, one or more hidden layers, and an output layer, where each neuron processes a weighted sum of inputs followed by a non-linear activation function. In this study, we compare our semi-supervised NNMPPCA model to a supervised MLP with a similar architecture and initial hyperparameters. The MLP is initialized with an input layer, two hidden layers containing 50 and 15 neurons respectively, and an output layer whose size corresponds to the number of classes. ReLU activation functions are applied to the hidden layers, while the output layer uses sigmoid for binary classification and softmax for multi-class classification. The model is trained using the Adam optimizer with an initial learning rate of 0.002 and an adaptive learning rate strategy, allowing the rate to adjust based on validation loss. This comparison evaluates the performance gains achieved through the semi-supervised approach.

- (Semi-supervised) AutoEncoder: An unsupervised neural network architecture designed to learn a compressed, low-dimensional representation of input data. It consists of two main components: an encoder that maps the input into a latent space and a decoder that reconstructs the input from this compressed representation. The network is trained by minimizing

the mean squared error (MSE) between the original input and its reconstruction, allowing it to capture important features and patterns in the data. In this study, the autoencoder architecture includes an encoder with an input layer matching the feature dimension, followed by two hidden layers with 50 and 15 neurons, using ReLU activation functions. The latent space is represented by a 3-neuron layer with SeLU activation. The decoder mirrors the encoder, consisting of two hidden layers (15 and 50 neurons) and an output layer using SeLU activation to match the input dimensions. The autoencoder is pre-trained for 100 epochs using the Adam optimizer with a learning rate of 0.001 and a batch size of 750. For semi-supervised fine-tuning, the encoder's weights are frozen except for the output layer, and a softmax-activated classification head with 3 output neurons is added. Fine-tuning is performed using 150 epochs, a batch size of 10, and an increased learning rate of 0.003. This approach allows us to compare the autoencoder-based model to both the semi-supervised NN-MPPCA and the supervised MLP models, evaluating the benefits of leveraging unsupervised pre-training followed by targeted fine-tuning.

- Self Training: A semi-supervised meta-estimator that wraps a supervised classifier and iteratively adds high-confidence pseudo-labeled examples from the unlabeled pool to the training set. In our setup, we used criterion='k_best' with k_best=100, meaning the 100 most confident predictions were added per iteration to guide the self-labeling process.

  All baseline models were executed 10 times, with the performance metrics averaged across runs. The data split was performed randomly for each iteration to assess the model's robustness and generalization ability across varying test datasets.

4.3.4 Experimental Setup

The experimental setup of the proposed Semi-MPPCA model involves a two-phase training strategy: pretraining and fine-tuning. The pretraining phase focuses on unsupervised representation learning, while the fine-tuning phase leverages a small set of labeled data for supervised classification.

Pretraining Phase: The Semi-MPPCA model was initialized with an architecture consisting

of an input layer corresponding to the number of features, followed by two hidden layers with 50 and 15 neurons, respectively, and an output layer representing the number of target classes. The activation function used was Scaled Exponential Linear Unit (SELU), coupled with the Glorot Normal initializer for kernel weights. To mitigate overfitting, a dropout rate of 0.2 was applied. The model was trained for 10 epochs with a batch size of 256. The learning rate was set to 0.002, and an adaptive weighting scheme ‘eta=[1, 0.5, 0.005]‘ was employed to balance the loss components during training.

Fine-Tuning Phase: Following pretraining, the output layer’s parameters was fine-tuned using a small subset of labeled data. Specifically, 10 labeled instances were used for validation for each case except for the case when only 10 labeled instance was used for finetuning, and the model was optimized to minimize the validation loss (‘val_loss‘). The Adam optimizer was employed, with an initial learning rate of 0.02. To ensure robustness, early stopping was again applied with a patience of 10 epochs, restoring the best weights according to the minimum validation loss. The learning rate was dynamically adjusted using a ReduceLROnPlateau callback, which monitored the validation loss and reduced the learning rate by a factor of 0.5 if no improvement was observed for 10 consecutive epochs. The minimum learning rate threshold was set at 0.05 to prevent excessively small updates.

This two-phase training approach allowed the model to leverage the abundance of unlabeled data for learning meaningful representations, while the fine-tuning phase enabled the model to adapt effectively to the classification task with limited labeled data. The careful selection of hyperparameters, combined with early stopping and adaptive learning rate adjustments, ensured both stability and performance during training.

For all experiments, we set the random state to `None` and repeated the process 10 times, averaging the performance metrics to ensure robustness. A 30% split was used for testing, while the remaining 70% was reserved for training. To simulate a realistic low-label transfer learning setup, we varied the amount of labeled data used for fine-tuning (e.g., 10, 20, 30, 40, and 50 samples per class for the binary case and 5-class scenario, and 15, 30, 45, 60, and 75 for the three-class case) while the remaining portion was used without any labels for pretraining.

4.3.5 Experimental results

The performance of Semi-MPPCA and the baseline models is assessed on the OE dataset for both binary and three-class classification tasks. For the AR task, the models are evaluated on both three-class and five-class classification tasks to provide a comprehensive and thorough analysis. Each reported value represents the mean performance over 10 iterations, with 30% of the data randomly selected for testing in each run. To evaluate the efficacy of our model, we experimented with varying amounts of labeled data: 10, 20, 30, 40, and 50 labeled data instances for binary and five-class classification. To facilitate the experiments, for the three-class tasks, we used 15, 30, 45, 60, and 75 labeled instances, ensuring balanced splits across classes. This approach provides a robust assessment of model performance across different levels of labeled data availability.

Binary Occupancy Estimation

Table 4.2 presents the F1 scores for binary classification on the OE dataset, comparing Semi-MPPCA with several baseline models across varying amounts of labeled data. The results show that Semi-MPPCA consistently outperforms all baselines, achieving the highest F1 scores across all label sets, from 91.56% with 10 labels to 93.62% with 50. The semi-supervised AutoEncoder, which shares architectural similarities with our model, remains the second-best, reaching 92.63% at 50 labels. The Self-Training classifier also shows strong performance, improving steadily from 85.58% at 10 labels to 92.69% at 50, outperforming traditional models and even slightly surpassing the AutoEncoder at the highest label count. Among supervised methods, SVM, Decision Tree, and AdaBoost plateau around 90 to 91% with 50 labels, while MLP lags slightly behind at 89.33%. These results reaffirm the effectiveness of representation learning in semi-supervised settings and demonstrate that Semi-MPPCA maintains a consistent edge, even when compared with strong semi-supervised alternatives such as Self-Training and AutoEncoder.

Figure 4.3, Figure 4.4 and Figure 4.5 provides a graphical representation of the F1 scores, accuracy and AUC scores respectively across varying amounts of labeled data for all models on the OE binary classification task, offering a clear visual comparison of their performance trends.

Table 4.2: F1 score for Binary Classification on OE Dataset

Method	**10 Labels**	**20 Labels**	**30 Labels**	**40 Labels**	**50 Labels**
SVM	84.51	85.99	88.39	89.44	90.35
KNN	83.23	86.23	86.74	87.12	87.94
DT	81.35	86.58	88.68	89.90	90.59
AdaBoost	83.13	84.78	86.67	89.34	90.93
MLP	80.84	83.12	86.51	86.86	89.33
AutoEncoder	89.05	90.45	91.35	91.96	92.63
Self Training	85.58	89.30	90.79	91.35	92.69
Semi-MPPCA	**91.56**	**91.81**	**92.11**	**92.59**	**93.62**

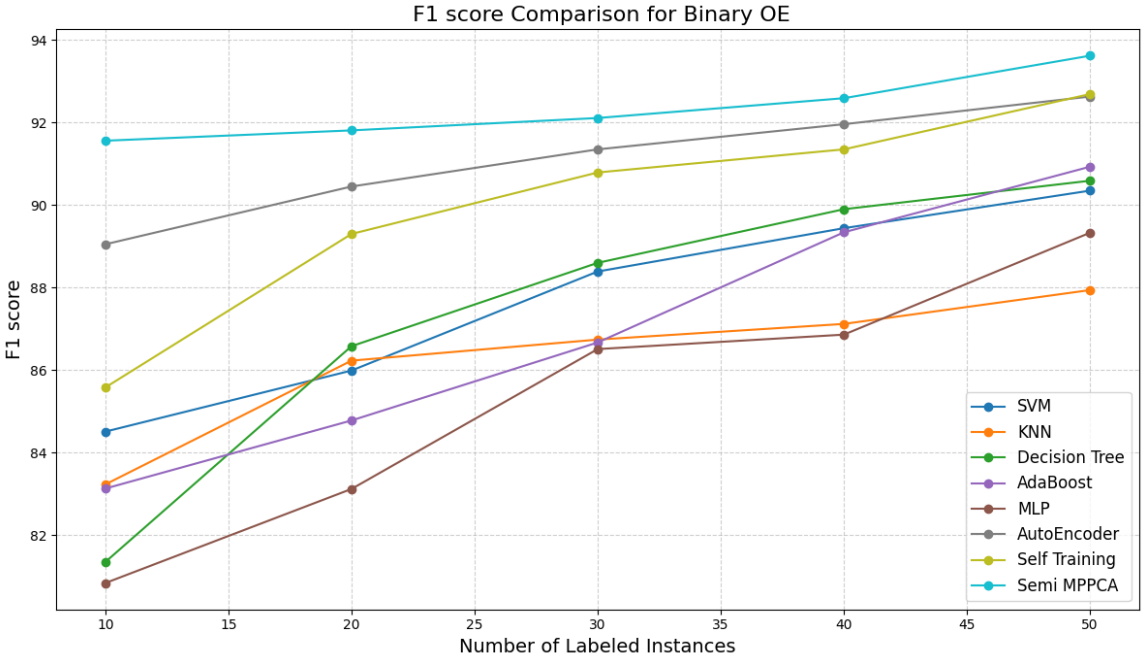


Figure 4.3: F1 score summary for binary OE

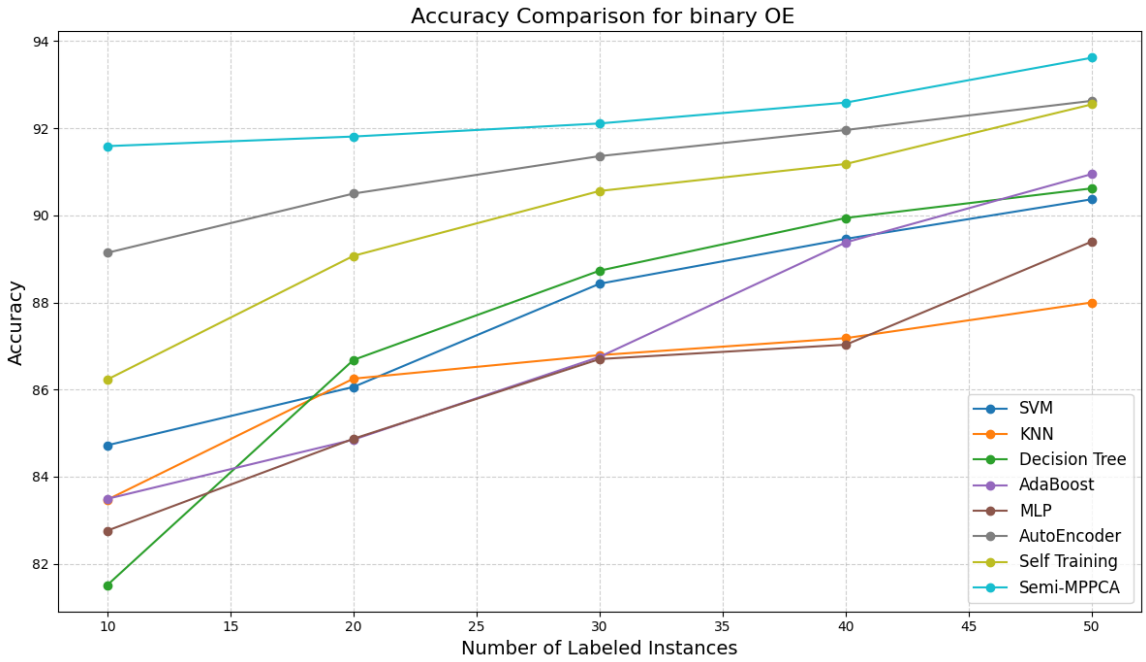


Figure 4.4: Accuracy summary for binary OE

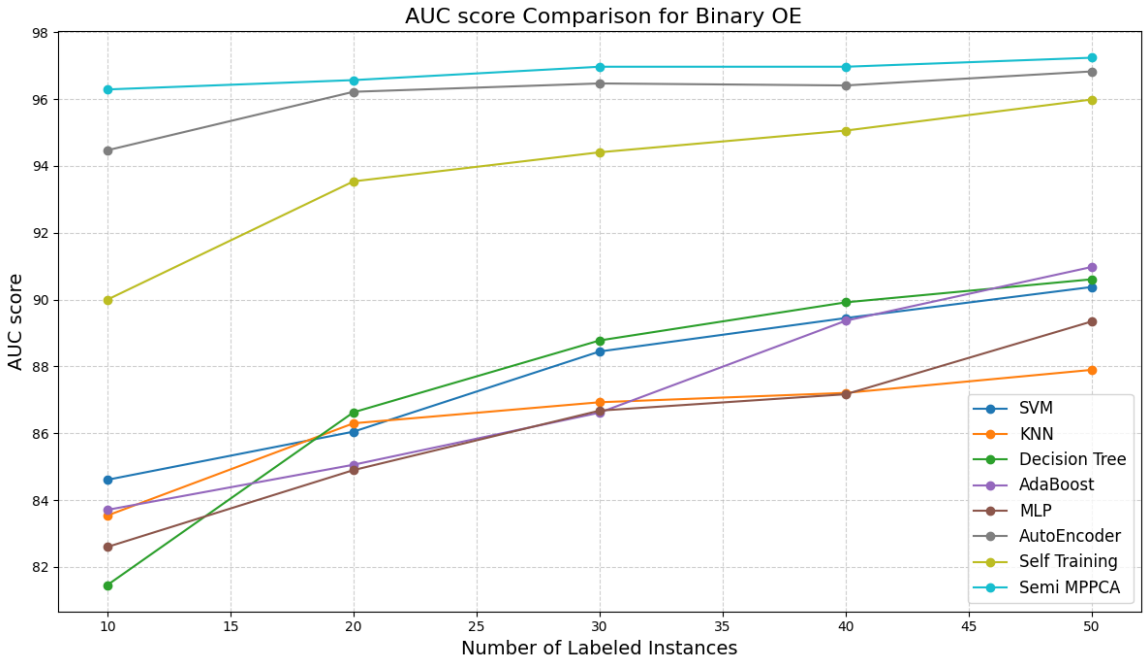


Figure 4.5: AUC score summary for binary OE

3-class Occupancy Estimation

The F1 scores for the OE 3-class classification task, shown in Table 4.3, reveal that our Semi-MPPCA model consistently outperforms all baseline models across all label counts. For instance, with 15 labels, Semi-MPPCA achieves an F1 score of 75.84%, slightly ahead of the next best model, Self-Training, which scores 74.47%. This performance gap widens as the number of labels increases, with Semi-MPPCA reaching an F1 score of 86.20% at 75 labels, outpacing the second-best model, Self-Training (83.82%). This consistent superiority underscores Semi-MPPCA's ability to effectively leverage both labeled and unlabeled data. Similar to our binary classification setup, the AutoEncoder model, which also uses representation learning, performs well, but is ultimately surpassed by Self-Training across all label levels and by Semi-MPPCA across the board.

When comparing with traditional models like SVM, KNN, and MLP, the performance gap remains significant, especially in low-label settings. While DT and AdaBoost show strong performance improvements as labels increase, reaching 80.29% and 81.87% respectively with 75 labels, they are still outperformed by the newer semi-supervised techniques. These results further support the efficacy of incorporating fine-tuning and representation learning strategies, positioning Semi-MPPCA as a competitive approach for multi-class occupancy estimation tasks. This is particularly evident in low-label settings, where traditional models like SVM and KNN struggle due to insufficient labeled data. The results suggest that Semi-MPPCA is a promising approach for tasks with limited labeled data, offering significant improvements over both traditional and advanced models like AutoEncoder. Figure 4.6, Figure 4.7 and Figure 4.8 illustrates the F1 scores, accuracies and AUC scores across different amounts of labeled data for all models on the 3-class OE classification task.

Activity Recognition Results - 3 class

Table 4.4 presents the F1 scores for 3-class classification on the AR dataset, comparing Semi-MPPCA with several baseline models across varying amounts of labeled data. The results demonstrate that Semi-MPPCA consistently outperforms all baseline methods, achieving the highest F1 scores across all label sets. Notably, even with only 15 labeled instances, Semi-MPPCA attains an

Table 4.3: F1 score for 3-class Classification on OE Dataset

Method	**15 Labels**	**30 Labels**	**45 Labels**	**60 Labels**	**75 Labels**
SVM	62.66	69.07	69.27	71.29	72.32
KNN	63.18	69.62	69.87	72.17	72.65
DT	70.33	74.52	77.66	79.33	80.29
AdaBoost	68.88	72.75	76.12	78.17	81.87
MLP	56.65	61.00	63.33	65.01	64.68
AutoEncoder	65.26	73.08	78.99	80.26	81.68
Self Training	74.47	77.60	81.08	82.10	83.82
Semi-MPPCA	**75.84**	**79.48**	**84.30**	**85.04**	**86.20**

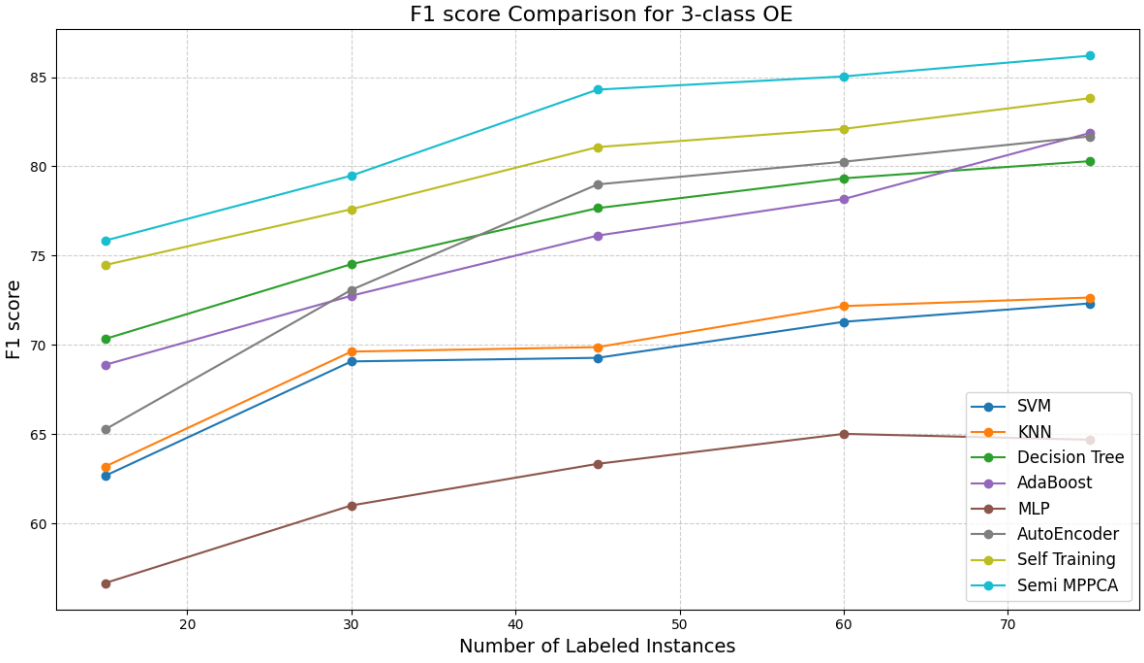


Figure 4.6: F1 score summary for 3-class OE

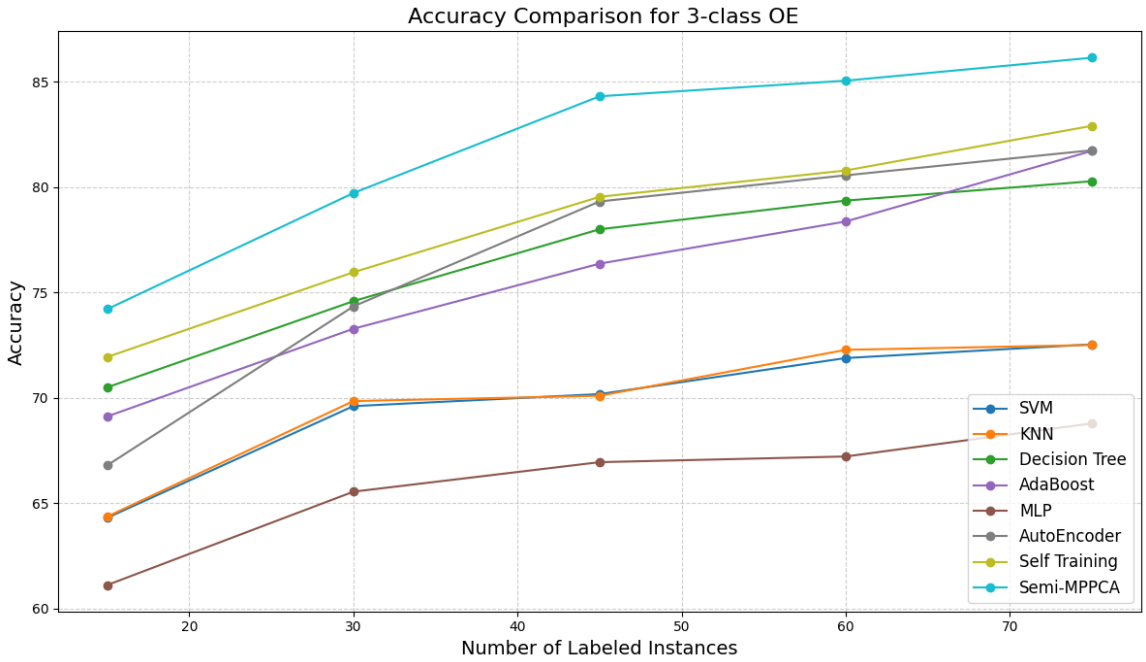


Figure 4.7: Accuracy summary of 3-clas OE

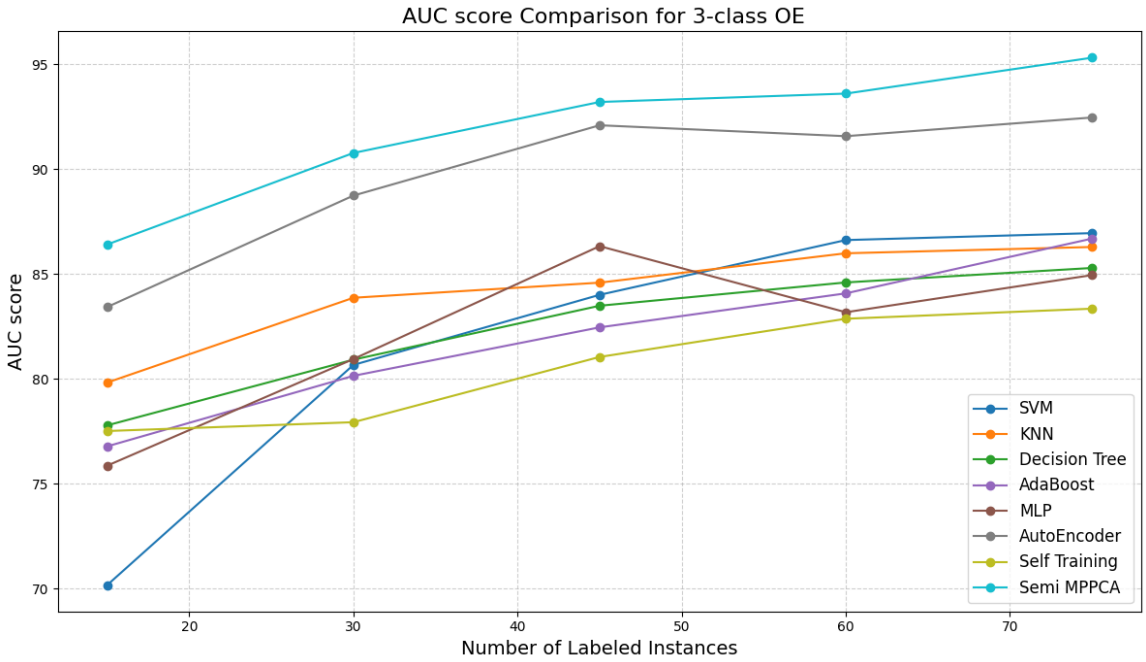


Figure 4.8: AUC score summary for 3-class OE

F1 score of 97.83%, steadily increasing to 99.33% with 75 labels. The semi-supervised AutoEncoder, which shares a similar representation learning approach, emerges as a strong performer, with F1 scores ranging from 91.33% (15 labels) to 98.22% (75 labels), although it is consistently outperformed by Semi-MPPCA. Self-Training also delivers highly competitive results, outperforming both DT and AdaBoost on most label counts, with its F1 scores increasing from 89.21% to 97.68% across the range, just slightly below AutoEncoder.

When compared to traditional supervised models, the advantage of Semi-MPPCA becomes even more pronounced. Decision Trees (DT) and AdaBoost, while performing well, achieve F1 scores of 89.86% and 88.06%, respectively, with 15 labels, and improve to 97.91% and 97.62% with 75 labels. Despite their strong performance, they are consistently surpassed by both AutoEncoder and Self-Training, and ultimately by Semi-MPPCA at every label count. Simpler models like SVM and KNN exhibit more modest improvements, with SVM starting at 43.93% (15 labels) and reaching 60.14% (75 labels), and KNN improving from 63.13% to 84.66% over the same range. These results highlight the limitations of traditional methods in handling the complexities of AR tasks, particularly when labeled data is scarce. Multi-Layer Perceptron (MLP), while showing some improvement, achieves F1 scores of 58.82% (15 labels) to 85.22% (75 labels), still trailing behind the top semi-supervised approaches.

Table 4.4: F1 score for 3-class Classification on AR Dataset

Method	**15 Labels**	**30 Labels**	**45 Labels**	**60 Labels**	**75 Labels**
SVM	43.93	47.90	51.82	55.72	60.14
KNN	63.13	75.95	77.03	82.30	84.66
DT	89.86	93.62	95.89	96.76	97.91
AdaBoost	88.06	92.74	95.67	97.38	97.62
MLP	58.82	64.09	66.67	77.05	85.22
AutoEncoder	91.33	92.19	96.77	97.83	98.22
Self Training	89.21	93.19	96.17	97.05	97.68
Semi-MPPCA	**97.83**	**98.16**	**98.44**	**98.94**	**99.33**

The performance trends for the AR 3-class classification task are visually depicted in Figure 4.9, Figure 4.10, and Figure 4.11, showcasing the F1 scores, accuracy, and AUC scores across different

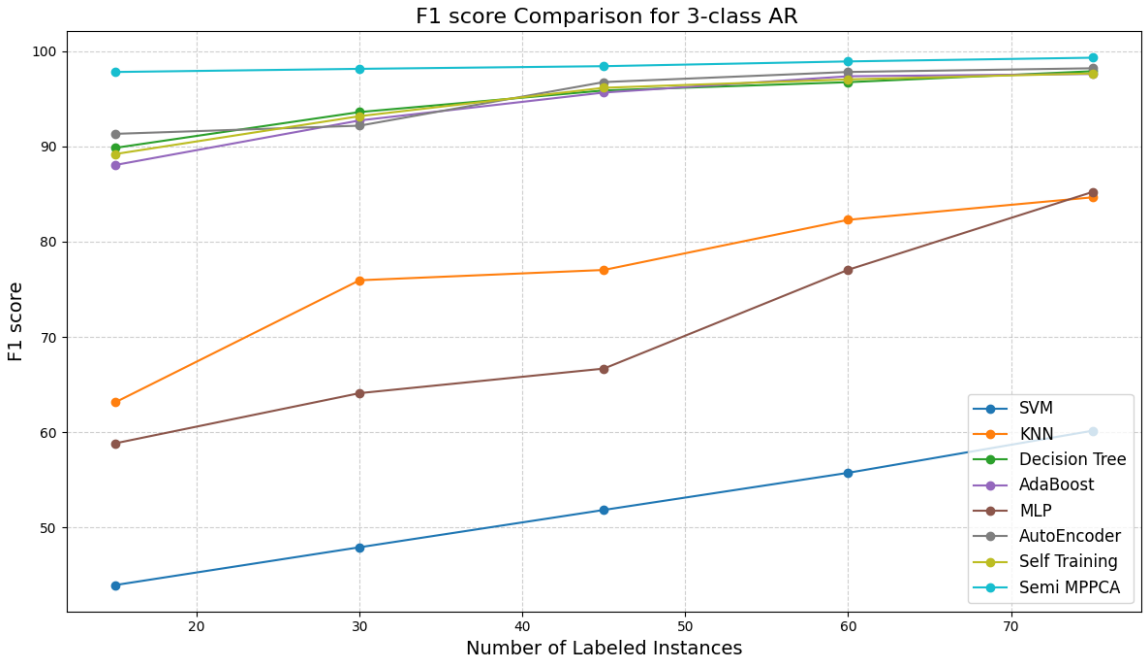


Figure 4.9: F1 score summary for 3-class AR

amounts of labeled data. These figures offer a clear comparison of how each model evolves as more labels are introduced.

Activity Recognition Results - 5 class

The updated results for the 5-class activity recognition scenario are presented in Table 4.5. Across all label budgets, the Semi-MPPCA once again achieves the strongest performance, demonstrating clear and consistent superiority over all baselines. With only 10 labeled samples, Semi-MPPCA attains an F1 score of **77.87%**, substantially outperforming the next-best methods, AutoEncoder (70.25%) and AdaBoost (67.90%). As the number of labels increases, all models show performance gains; however, the Semi-MPPCA maintains a widening advantage at every scale.

With 20 labels, Semi-MPPCA reaches **84.51%**, significantly above Self-Training (73.63%), AdaBoost (77.05%), and DT (75.15%). At 30 labels, the model achieves **86.58%**, outperforming strong baselines such as AdaBoost (82.48%) and AutoEncoder (80.61%). The margin remains pronounced at 40 labels, where Semi-MPPCA rises to **88.27%**, surpassing DT (82.79%), Self-Training (82.73%), and AutoEncoder (82.05%). Finally, with 50 labels, Semi-MPPCA reaches

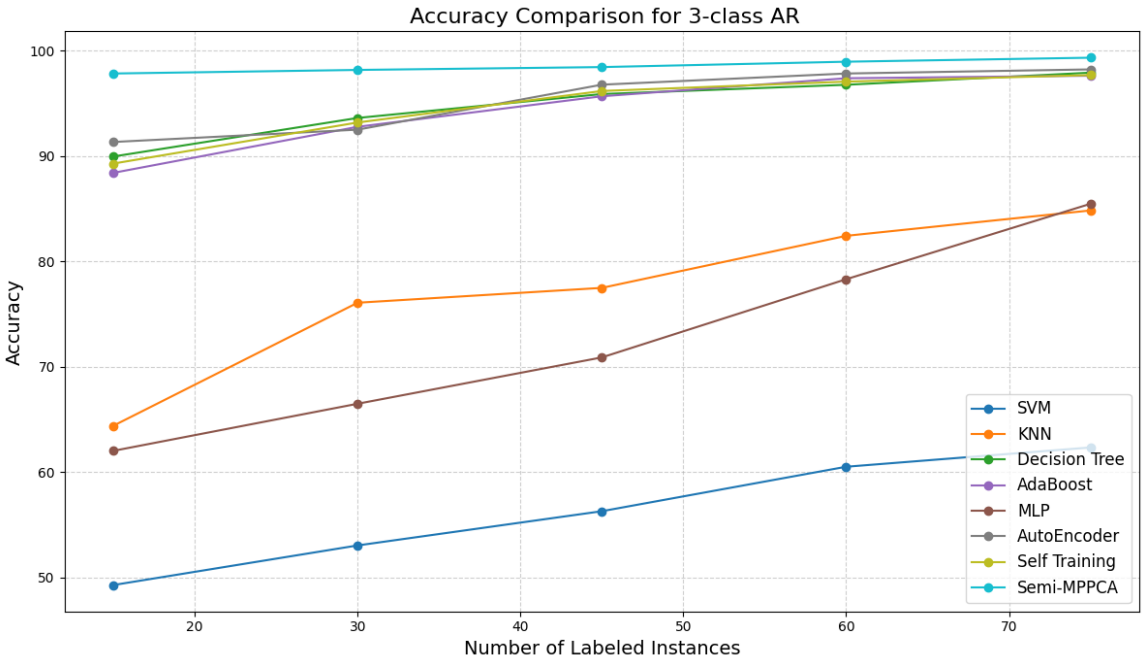


Figure 4.10: Accuracy summary for 3-class AR

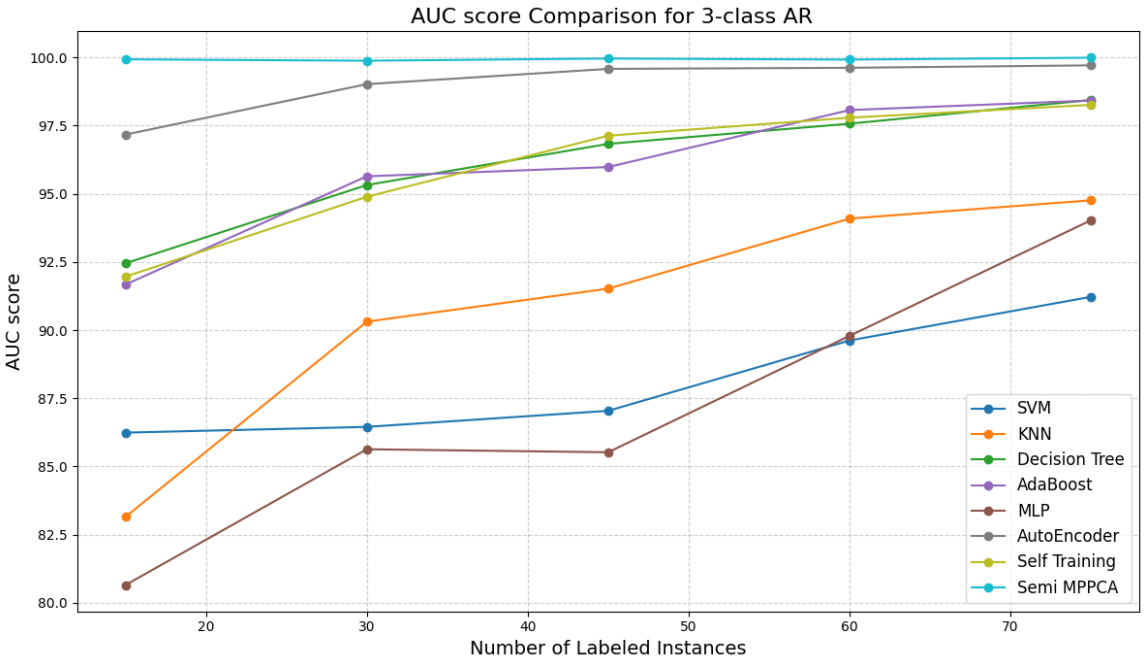


Figure 4.11: AUC score summary for 3-class AR

its peak performance of **89.00%**, retaining a notable lead over AdaBoost (85.96%), Self-Training (85.58%), DT (85.28%), and AutoEncoder (85.12%).

These results confirm the strong label efficiency and robustness of Semi-MPPCA, which consistently provides the best performance across all label regimes.

Table 4.5: F1 score for 5-class Classification on AR Dataset

Method	**10 Labels**	**20 Labels**	**30 Labels**	**40 Labels**	**50 Labels**
SVM	53.78	57.37	67.17	72.07	78.47
KNN	26.07	43.95	54.22	59.35	61.24
DT	62.26	75.15	80.39	82.79	85.28
AdaBoost	67.90	77.05	82.48	84.15	85.96
MLP	24.85	31.38	36.44	38.12	44.25
AutoEncoder	70.25	71.19	80.61	82.05	85.12
Self Training	67.02	73.63	79.43	82.73	85.58
Semi-MPPCA	**77.87**	**84.51**	**86.58**	**88.27**	**89.00**

Figure 4.12, Figure 4.13 and Figure 4.14 provides a graphical representation of the F1 scores, accuracy and AUC scores respectively across varying amounts of labeled data for all models on the AR 5-class classification task, offering a clear visual comparison of their performance trends.

4.3.6 Discussion

The results across the four classification tasks reveal consistent trends that highlight the effectiveness of Semi-MPPCA compared to both traditional machine learning models and deep learning methods. A clear pattern emerges, with Semi-MPPCA achieving the highest performance across all tasks, especially as the number of labeled samples increases. While simpler tasks like binary classification show relatively strong results for all models, more complex multi-class problems expose the limitations of traditional approaches. AutoEncoder, leveraging deep feature extraction, often performs competitively, but still falls short of Semi-MPPCA's capabilities. We can also observe a striking trend of the steadily increasing F1 scores across all models as the number of labeled samples grows.

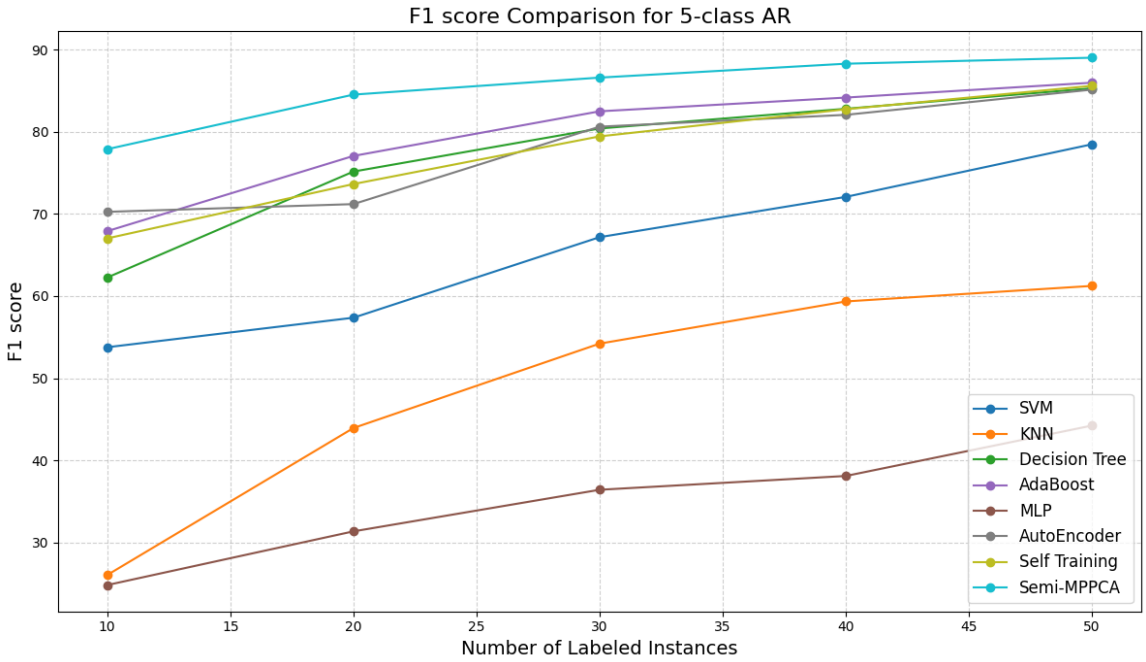


Figure 4.12: F1 score summary for 5-class AR

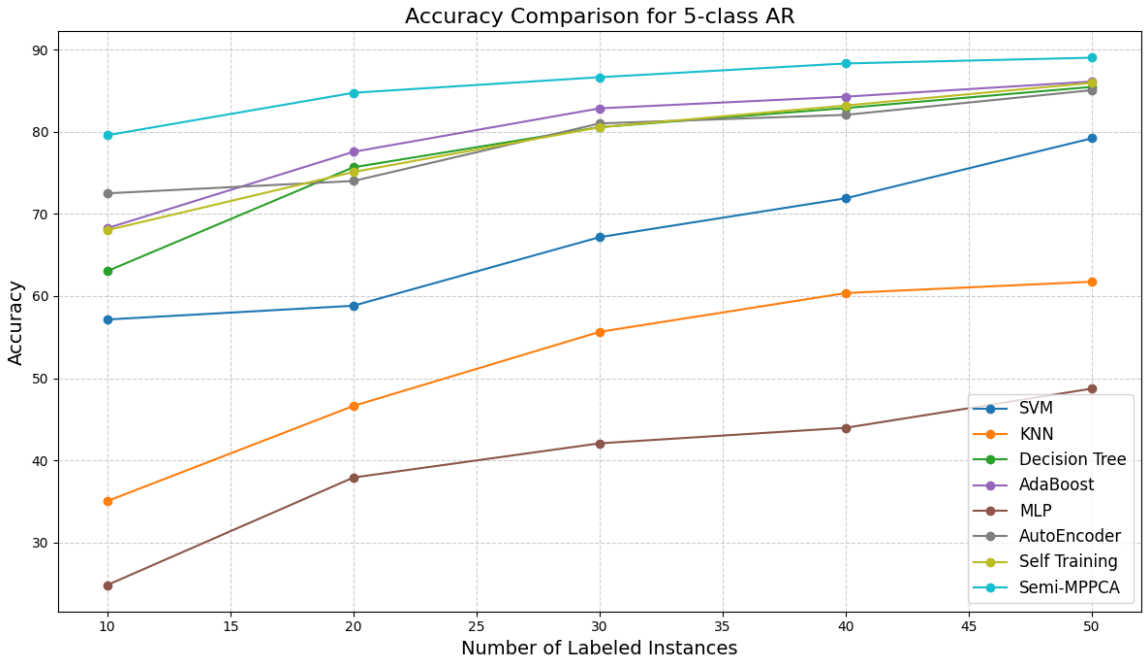


Figure 4.13: Accuracy summary for 5-class AR

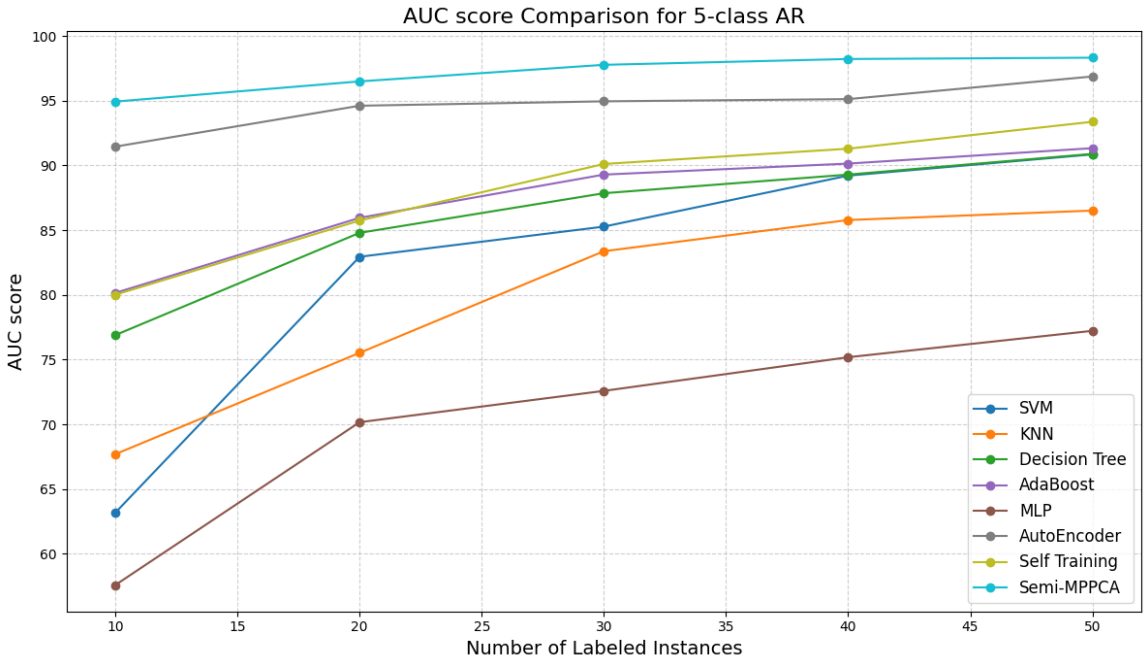


Figure 4.14: AUC score summary for 5-class AR

Interestingly, classical methods such as Decision Tree and AdaBoost show strong baseline performance, particularly in binary and 3-class tasks. However, as the tasks become more intricate, like the 5-class activity recognition problem, their improvement plateaus, revealing a struggle to capture complex class relationships. Meanwhile, MLP presents inconsistent results, performing adequately in simpler settings but failing to generalize effectively in multi-class scenarios. This suggests that MLP may require more extensive hyperparameter tuning or larger labeled datasets to achieve comparable outcomes. It is also important to highlight that in all scenarios, the relatively weaker performance of MLP emphasizes the benefits of incorporating MPPCA within a neural framework, as opposed to relying solely on traditional neural network architectures.

Overall, these insights underscore the superiority of Semi-MPPCA, not only in binary and 3-class problems but also in more challenging multi-class settings. The results validate the method's capacity to balance labeled and unlabeled data, offering a promising solution for semi-supervised classification tasks, while shedding light on the relative strengths and weaknesses of both classical and deep learning models. Its hybrid approach, which combines probabilistic modeling with deep learning, provides a scalable and adaptable solution for tasks requiring both accuracy and

interpretability.

Table 4.6: Class-wise Performance of Semi-MPPCA on OE (30 Labels)

Method	**class 0**	**class 1**	**class 2**
Accuracy	92.67	64.94	81.47
F1-score	87.58	69.93	80.43
Precision	83.01	75.76	79.41
AUC	96.83	85.57	90.24

To assess Semi-MPPCA's ability to generalize across occupancy levels, we report per-class performance metrics in Table 4.6, including accuracy, F1-score, precision, and AUC on Occupancy Estimation data for the 30-label scenario. The model demonstrates strong performance on class 0 (low occupancy), with an accuracy of 92.67% and an AUC of 96.83%. However, performance on class 1 is notably lower across all metrics, suggesting some difficulty in distinguishing medium occupancy patterns. In contrast, class 2 shows balanced results, indicating that Semi-MPPCA can generalize reasonably well to higher occupancy levels despite limited labeled data. These findings highlight both the robustness of the model on dominant classes and the need for further improvement on more ambiguous or transitional occupancy states.

4.3.7 Contribution, limitation, and future direction

The proposed Semi-MPPCA model introduces several key innovations to the field of representation learning and semi-supervised classification. First, we adapt the Mixture of Probabilistic Principal Component Analysis (MPPCA) framework into a neural network form by replacing the traditional Expectation-Maximization (EM) algorithm with backpropagation. This transformation enables the model to leverage gradient-based optimization, enhancing scalability and flexibility. The model is then used as a pre-trained architecture that utilizes unsupervised learning to capture complex feature representations from unlabeled data, followed by fine-tuning the output layer by using a cluster-then-label approach for classification tasks. To address the challenge of cluster-label misalignment which is a common issue in unsupervised and semi-supervised settings, we introduce a novel weight-swapping technique. This mechanism dynamically aligns learned representations with their corresponding class labels, ensuring more accurate mapping between clusters

and target outputs. Our extensive experimental evaluations demonstrate the superior performance of Semi-MPPCA with minimal labeled data, consistently outperforming traditional occupancy detection algorithms and several state-of-the-art machine learning models. Semi-MPPCA showcases how representation learning, when combined with finetuning, can enhance performance in data-scarce environments. This work paves the way for further exploration of hybrid models in machine learning.

Despite its strong performance, the Semi-MPPCA model has certain limitations. The model demonstrates reduced effectiveness when applied to datasets with binary-encoded features, where the input data primarily consists of simple, linear patterns (e.g., features represented by 0s and 1s). In such cases, the model give modest performance. This indicates that Semi-MPPCA is most effective in scenarios where complex, non-linear relationships exist, allowing the model to fully exploit its representation learning capacity. Additionally, as a deep learning-based model, Semi-MPPCA benefits from a substantial amount of unlabeled data during the pretraining phase. In cases where the quantity of unlabeled data is limited, the model's ability to learn robust feature representations may be compromised. The computational overhead associated with training a neural network-based framework also poses a challenge, particularly for real-time or resource-constrained environments.

To address these limitations and further enhance the capabilities of Semi-MPPCA, several promising avenues for future research can be explored. One potential direction is the integration of domain adaptation techniques, allowing the model to generalize more effectively across different but related datasets. By aligning feature distributions between source and target domains, this approach could improve the model's robustness when applied to new environments with varying data distributions. Another valuable extension would involve incorporating Explainable AI (XAI) methods to provide greater interpretability of the learned representations and classification decisions. Enhancing model transparency could foster trust in critical applications, such as energy-efficient building management, by offering insights into how and why specific predictions are made.

4.4 Conclusion

Sensor data plays a critical role in modern machine learning applications, providing rich, real-time information about environmental conditions, user behaviors, and system states. In this chapter, we utilize data collected from a variety of sensors, including motion sensors, door contact sensors, temperature sensors, and light switch sensors, to capture detailed activity patterns and to detect occupancy. We proposed the Semi-Supervised Mixture of Probabilistic Principal Component Analyzers (Semi-MPPCA) model, an innovative extension of the MMPCA framework adapted into a neural network architecture for Activity Recognition (AR) and Occupancy Estimation (OE). By replacing the traditional Expectation-Maximization (EM) algorithm with backpropagation with a custom defined loss function to ensure that the model converge to the optimum, Semi-MPPCA leverages representation learning followed by a cluster-then-label approach of semi-supervised learning, pretraining on unlabeled data, and fine-tuning on few labeled data, effectively addressing the challenge of data scarcity in classification tasks. Additionally, we introduced a weight-swapping technique to mitigate cluster-label misalignment, enhancing the model's alignment between learned feature representations and class labels. Our extensive experimental evaluations on both OE and AR datasets demonstrate the effectiveness of the proposed model. The results consistently show that Semi-MPPCA outperforms traditional machine learning models such as SVM, KNN, DT, as well as advanced methods like AdaBoost and semi-supervised AutoEncoder, particularly with minimal labeled data. Notably, Semi-MPPCA achieves superior F1 scores across various label proportions, underscoring its robustness and adaptability. For the OE dataset, Semi-MPPCA consistently outperformed most baseline models. In the binary classification task, it achieved an F1 score of 91.56% with 10 labels, steadily increasing to 93.62% with 50 labels, surpassing semi-supervised AutoEncoder and other base models. In the 3-class classification task, Semi-MPPCA obtained an F1 score of 73.84% with 15 labels and peaked at 86.20% with 75 labels, demonstrating its robustness with minimal labeled data. Similarly for 3-class AR, Semi-MPPCA achieved an impressive 99.33% accuracy at 75 labels, starting with 97.83% at 15 labels. Semi-MPPCA also outperforms the other base models for 5-class AR reaching an F1 score of 67.28% at 50 labels, demonstrating that Semi-MPPCA remains effective even in complex classification scenarios. Furthermore, by maintaining

strong performance with limited supervision, Semi-MPPCA presents a significant contribution to the scientific community by advancing semi-supervised learning methodologies and their applications in real-world domains. In the context of smart buildings, where sensor networks continuously collect vast amounts of environmental and occupancy data, the ability to leverage unlabeled data is crucial for optimizing energy consumption and automation systems. While Semi-MPPCA proves effective in handling complex, nonlinear datasets, like many deep learning models, its reliance on a substantial amount of unlabeled data for effective pretraining poses a limitation, as smaller datasets may not fully leverage its capabilities. Future research can focus on incorporating domain adaptation techniques to enhance the model's transferability across diverse datasets, ensuring its applicability to new environments with minimal retraining. Additionally, integrating Explainable AI (XAI) methods could provide deeper insights into the decision-making process of Semi-MPPCA, fostering trust and transparency. By advancing these directions, we aim to further enhance practical relevance in the development of smarter, more responsive building systems to optimize energy consumption.

Chapter 5

Probabilistic Hypothesis Anchored Domain Adaptation for Modeling Human Behavior

Although the semi-supervised Semi-MPPCA framework developed in Chapter 4 effectively leverages unlabeled data, it does not inherently resolve the challenges posed by domain shift between different buildings. To address this gap, this chapter extends Semi-MPPCA into a fully unsupervised, source-free domain-adaptation framework that can transfer probabilistic structure across environments without accessing source-domain data. By incorporating a source hypothesis-transfer mechanism and introducing a novel anchor-based alignment strategy, our proposed model tackles cross-domain adaptation while preserving the mixture geometry learned from the source. This chapter therefore completes the methodological trajectory by tackling domain shift under the most restrictive setting: absence of target labels, very limited unlabeled target data, and no access to raw source-domain data.

5.1 Introduction

The building sector accounts for a substantial portion of global energy consumption [35, 58, 109, 154]. In fact, the ongoing rise in population and rapid urbanization are further amplifying energy demand, leading to a significant increase in the energy requirements of buildings [155]. Given such a large share, even modest improvements in building energy efficiency can yield major savings and reduce carbon emissions. Researchers have found out that one of the most promising strategies involves modeling human behavior such as occupancy and activity patterns to enable intelligent, data-driven control of building systems [21]. For instance, dynamically adjusting heating, ventilation and air-conditioning based on real-time occupancy level of that space optimize the balance between occupants comfort and energy consumption [156]. Moreover, integrating activity recognition with occupancy estimation enables a deeper understanding of occupants' contextual behavior, allowing energy management systems to proactively adapt to different usage patterns. Such intelligent control can, for example, pre-cool active areas, dim lighting in unoccupied zones, or regulate appliance usage according to activity type, thereby enhancing overall energy efficiency without compromising comfort [58, 157, 158].

Despite the rapid progress of data-driven modeling, traditional approaches for activity recognition (AR) and occupancy estimation (OE) remain heavily reliant on supervised learning paradigms [58]. While these methods achieve strong performance in controlled settings, their effectiveness significantly deteriorates in real-world environments due to the fundamental challenge of label scarcity. Acquiring high-quality annotations is inherently difficult, as it typically requires manual observation, synchronized video recordings, or specialized equipment. These processes are not only costly and time-consuming but also intrusive, raising substantial privacy concerns and hindering large-scale deployment [34, 56, 159, 160].

A further complication arises from the heterogeneous nature of sensor systems across different buildings. Traditional machine learning algorithms assume that the training set and the test set follow the same data distribution. However, very often this is not the case. Models trained on data from one specific environment often struggle when applied to another environment due to domain shifts [32, 44, 161–163]. These shifts, which include variations in sensor configurations,

spatial layouts, and occupant behaviors, alter the underlying data distribution, causing significant performance degradation when models are applied to new, unseen contexts. The convergence of these challenges, including the impracticality of supervised learning at scale, data privacy issues, and the poor generalization across environments, forms the critical motivation for this work.

This sensitivity to distributional changes and the lack of labeled datasets underscores the need for generalizable, label-efficient learning techniques. Recently, domain adaptation (DA) [44, 51] has emerged as a powerful approach that leverages labeled data from a source domain together with unlabeled data from a target domain. By aligning the underlying feature distributions between domains, DA enables models to generalize across different environments without extensive relabeling, thereby enhancing scalability while maintaining data privacy. Over the past decade, researchers have increasingly explored deep domain adaptation as a means to bridge distribution gaps between datasets [42, 164–166], with particular emphasis on the unsupervised closed-set scenario, where domains share identical label spaces but lack labeled target samples. One prominent line of research in domain adaptation focuses on minimizing distributional divergence between source and target domains by aligning their statistical moments, a principle that underlies methods based on Maximum Mean Discrepancy (MMD). For example, [167] proposes a discriminative Maximum Mean Discrepancy (MMD) approach for domain adaptation that explicitly balances domain alignment and class separability. Their method introduces controlled intra-class constraints and reweighted inter-class distances to preserve discriminative structure while aligning domains, resulting in improved cross-domain generalization.

Lately, source-free domain adaptation (SFDA) [45, 168] has gained significant traction as it enables knowledge transfer without accessing source data, addressing growing concerns over data privacy, security, and storage constraints. [169] proposes G-SFDA (Generalized Source-Free Domain Adaptation) which adapts a source-trained model using only unlabeled target data while preserving source performance. It employs Local Structure Clustering (LSC) to group target features by semantic similarity and Sparse Domain Attention (SDA) to activate domain-specific channels and regularize gradients for stable adaptation. [170] introduces SF(DA)², a source-free domain adaptation framework that reinterprets adaptation through feature-space data augmentation. By building an augmentation graph and applying spectral neighborhood clustering along with implicit feature

augmentation and feature disentanglement losses, the method achieves efficient, class-consistent adaptation without prior transformation knowledge or high computational cost.

Recent works have introduced prototype- or anchor-based adaptation frameworks that align target features with class-level representations rather than raw source samples, effectively reducing sampling variability, class imbalance, and privacy concerns. [171] proposes Prototype-oriented Conditional Transport (PCT) framework learns class prototypes and aligns target features through a probabilistic bi-directional transport process, while estimating target class proportions. The method learns class prototypes, which are latent representations that characterize each class, rather than aligning raw source and target samples. Each prototype acts as a learnable parameter in the latent space, serving as a stable class-level reference for alignment. This approach ensures every class is represented during training, reduces sensitivity to outliers, and allows source-free adaptation by aligning target features to these prototypes instead of individual source instances.

While recent advances in domain adaptation have achieved impressive progress in addressing data scarcity, privacy, and cross-domain generalization, these methods often assume relatively simple or well-structured data distributions. However, modeling human behavior in smart buildings poses significant challenges due to the highly non-linear, dynamic, and high-dimensional nature of sensor data [35, 38, 172]. Traditional statistical and linear models often fail to capture complex temporal dependencies and latent relationships between heterogeneous signals such as motion, temperature, light, and CO_2 levels. These interdependencies are crucial for accurately inferring occupancy and activity patterns, which drive energy optimization and comfort control. Therefore, a model capable of learning rich, non-linear representations while maintaining interpretability and robustness is essential for advancing intelligent building systems.

To address these challenges, we first developed the Semi-Supervised Mixture of Probabilistic Principal Component Analyzers (Semi-MPPCA) [35]. This model transforms the traditional Mixture of Probabilistic PCA into a neural network framework [38] that supports both pretraining and fine-tuning. During pretraining, unlabeled data are used to capture complex, high-dimensional, nonlinear latent structures and probabilistic relationships within the feature space. A small labeled subset is then used to fine-tune the model, aligning mixture components with their corresponding

classes. To further enhance cluster-class consistency, a weight-swapping alignment strategy is introduced, improving the interpretability and discriminative power of the latent representation. This formulation effectively combines the uncertainty modeling of probabilistic methods with the flexibility of neural architectures, enabling robust semi-supervised learning even under extreme label scarcity.

Building on this foundation, we extend the Semi-MPPCA framework into a source-free domain adaptation setting through the proposed Probabilistic Hypothesis-Anchored Domain Adaptation (PHADA). In this case, the pretrained Semi-MPPCA serves as the source hypothesis, with the classifier frozen and only the feature extractor adapted to an unlabeled target domain. PHADA follows the SHOT-IM paradigm, optimizing an information-maximization objective and a pseudo-labeling mechanism to refine target features without source data access. To further stabilize adaptation, we introduce a novel component-anchor alignment strategy, which constrains evolving target representations to remain consistent with the probabilistic geometry learned in the source domain. This integration yields a label-efficient, interpretable, and privacy-preserving framework for cross-building behavioral modeling. By operating without labeled target data or source samples, PHADA minimizes the risk of negative transfer while maintaining generalization across diverse building environments. Its design combines label efficiency, domain robustness, and data privacy within a unified probabilistic framework.

Our work makes several contributions to advancing label-efficient, source-free domain adaptation.

- We propose a probabilistic semi-supervised source model that employs a Semi-MPPCA backbone, a neural extension of the Mixture of Probabilistic Principal Component Analyzers (MPPCA) that extracts the underlying latent structure of nonlinear, high-dimensional multi-sensor smart-building data through mixtures of low-dimensional probabilistic subspaces. This formulation enables robust semi-supervised learning by leveraging unlabeled data for representation learning and minimal labeled samples for cluster-to-class alignment.

- We extend this model into a source-free domain adaptation framework and introduce a novel component-anchor alignment mechanism that preserves the probabilistic geometry of the

source mixture while adapting to unlabeled target data. Combined with an information-maximization objective, this strategy stabilizes adaptation and mitigates class collapse without requiring access to source data or labels.

- We demonstrate that PHADA achieves dual label efficiency, as it not only adapts successfully to the target domain in a fully unsupervised manner, but also learns effectively from limited labeled data in the source domain through semi-supervised training (Semi MPPCA), maintaining strong performance under domain shift.

The remainder of this chapter is organized as follows. Section 1 introduces the motivation and key contributions of PHADA. Section 2 details the proposed methodology and section 3 presents the experimental setup and results. Lastly, section 4 concludes the chapter with final remarks and directions for future research.

5.2 Methodology

Building upon the Semi-MPPCA framework, this section introduces the proposed Probabilistic Hypothesis-Anchored Domain Adaptation (PHADA) model, which extends Semi-MPPCA to the unsupervised domain adaptation (UDA) setting. While Semi-MPPCA demonstrated the ability to model nonlinear and high-dimensional sensor data with limited labeled samples, it assumes the availability of at least a small proportion of labeled data. In contrast, PHADA addresses the practical scenario where labeled data from the target domain are entirely unavailable, and direct access to the source data is restricted due to privacy or storage constraints. Instead, adaptation relies solely on the pretrained source model, making the approach *source-free* and label-efficient.

5.2.1 Source HypOthesis Transfer (SHOT)

In real-world smart-building environments, models trained on one building often fail to generalize to another due to variations in sensor placement, architecture, and occupant behavior. These variations induce a *domain shift*, where the target data distribution $p_t(x)$ differs from the source distribution $p_s(x)$, resulting in degraded performance if models are transferred directly. PHADA mitigates this problem by transferring the knowledge embedded in the source hypothesis (the learned

parameters of the Semi-MPPCA model) to the target domain without relying on labeled target data.

Formally, the pretrained source model is denoted as:

$$f_s(x) = h_s(g_s(x)), \tag{43}$$

where $g_s : \mathcal{X}_s \to \mathbb{R}^d$ represents the feature extractor that maps sensor inputs to a d-dimensional latent space, and $h_s : \mathbb{R}^d \to \mathbb{R}^K$ is the classifier that outputs the posterior probabilities over K classes. The corresponding target model is $f_t(x) = h_t(g_t(x))$. Following the principle of Source Hypothesis Transfer (SHOT), the classifier weights are transferred and frozen such that $h_t = h_s$, while the target feature extractor g_t is updated to align with the fixed decision boundaries of the source hypothesis.

5.2.2 SHOT - Information Maximization (SHOT-IM)

The adaptation process is guided by an information-theoretic objective that simultaneously promotes prediction confidence and class-level diversity among unlabeled target samples. The first component is the conditional entropy term, which encourages the model to produce confident, low-entropy predictions for each target instance:

$$\mathcal{L}_{ent}(f_t; \mathcal{X}_t) = -\mathbb{E}_{x_t \sim \mathcal{X}_t} \sum_{k=1}^{K} \delta_k(f_t(x_t)) \log \delta_k(f_t(x_t)), \tag{44}$$

where $\delta_k(f_t(x_t))$ denotes the predicted probability for class k given a target input x_t. Minimizing $\mathcal{L}_{ent}$ reduces predictive uncertainty and encourages outputs close to one-hot vectors.

However, conditional entropy minimization alone can lead to degenerate solutions, such as mapping all target samples to a single class. To prevent this, SHOT-IM incorporates a second term that promotes global prediction diversity across the entire target dataset:

$$\mathcal{L}_{div}(f_t; \mathcal{X}_t) = \sum_{k=1}^{K} \hat{p}_k \log \hat{p}_k, \tag{45}$$

where

$$\hat{p}_k = \mathbb{E}_{x_t \sim \mathcal{X}_t}[\delta_k(f_t(x_t))]$$

is the mean predicted probability of class k over the target domain. This term is minimized when the marginal distribution of predictions approaches the uniform distribution, thereby enforcing balanced class usage and preventing collapse.

Together, $\mathcal{L}_{ent}$ and $\mathcal{L}_{div}$ constitute the Information Maximization (IM) objective, ensuring that target outputs are individually confident yet globally diverse, even in the absence of source data.

5.2.3 Source Hypothesis Transfer Augmented with Self-supervised Pseudo-labeling

Building on the probabilistic latent representation learned by Semi-MPPCA, the pseudo-label refinement mechanism in PHADA operates directly within the responsibility-weighted embedding space. Unlike deterministic feature embeddings, each target representation $g_t(x_t)$ reflects a soft mixture activation, allowing pseudo-labels to form around flexible component boundaries. This probabilistic formulation yields a more stable refinement process that remains consistent with the Semi-MPPCA backbone, as both mixture responsibilities and pseudo-labels co-evolve during adaptation.

Once the information-maximization objective has partially aligned the target encoder g_t with the frozen classifier h_t, PHADA sharpens decision boundaries through a self-supervised prototype-based refinement. The goal is to construct class prototypes directly from target features and to iteratively improve pseudo-label quality without relying on source data.

Let $f_t = h_t \circ g_t$ denote the current target hypothesis, and let $p_k(x_t) = \delta_k(f_t(x_t))$ be the predicted probability for class k. We begin by estimating initial class prototypes via probability-weighted averaging:

$$c_k^{(0)} = \frac{\sum_{x_t \in \mathcal{X}_t} p_k(x_t)\, g_t(x_t)}{\sum_{x_t \in \mathcal{X}_t} p_k(x_t)} \quad \text{for } k = 1, \ldots, K. \tag{46}$$

These soft prototypes emphasize high-confidence samples while attenuating noisy predictions. Using these initial prototypes, each target instance is assigned a pseudo-label via nearest-prototype matching using cosine distance D_f:

$$\hat{y}_t = \arg\min_{k \in \{1,\ldots,K\}} D_f\Big(g_t(x_t),\, c_k^{(0)}\Big). \tag{47}$$

To reduce initialization bias and better reflect the underlying structure of the target domain, we update the prototypes using hard assignments:

$$c_k^{(1)} \;=\; \frac{\sum_{x_t \in \mathcal{X}_t} \mathbf{1}[\hat{y}_t = k]\, g_t(x_t)}{\sum_{x_t \in \mathcal{X}_t} \mathbf{1}[\hat{y}_t = k]}, \tag{48}$$

and optionally reassign labels using the updated prototypes (one refinement round is typically sufficient):

$$\hat{y}_t \;=\; \arg\min_k D_f\Big(g_t(x_t),\, c_k^{(1)}\Big)\,.$$

The final pseudo-labels provide a self-supervised training signal through a cross-entropy objective with the frozen classifier:

$$\mathcal{L}_{\mathrm{PL}} \;=\; \mathbb{E}_{(x_t,\hat{y}_t)\in \mathcal{X}_t \times \hat{\mathcal{Y}}_t} \left[\sum_{k=1}^{K} \mathbf{1}[k = \hat{y}_t]\, \log \delta_k(h_t(g_t(x_t))) \right]. \tag{49}$$

This pseudo-label term complements the information-maximization objective: while $\mathcal{L}_{\mathrm{IM}}$ promotes globally confident and diverse predictions, $\mathcal{L}_{\mathrm{PL}}$ enforces local class consistency by pulling features toward class prototypes, thereby improving separability in the target domain.

5.2.4 Component-Anchor Alignment

In PHADA, the SHOT-IM framework is embedded within the probabilistic latent space learned by Semi-MPPCA. Unlike deterministic feature extractors, Semi-MPPCA represents each input as a mixture of probabilistic subspaces, expressed through soft responsibilities $\mathbf{R}(x) = [r_1(x), \ldots, r_K(x)]$, where $r_i(x)$ denotes the posterior responsibility of component i. The target encoder g_t therefore operates in a responsibility-aware latent space, and adaptation must preserve the mixture structure learned from the source domain.

The adaptation procedure begins by transferring the pretrained Semi-MPPCA parameters $(\mu_i, W_i, \sigma_i, \pi_i)$ from the source domain. The classifier h_t remains frozen, while g_t is optimized using the information-maximization objective over unlabeled target samples. This aligns the target responsibilities with the source hypothesis but does not guarantee that the adapted components remain geometrically consistent with their source counterparts.

A key limitation of SHOT-IM is that entropy-based alignment alone provides no constraint on how target features relate to the mixture structure learned by Semi-MPPCA. As a result, component assignments may drift during adaptation, weakening the correspondence between source and target latent spaces. To prevent such drift, PHADA introduces a *component-anchor alignment* mechanism that explicitly preserves the geometric organization of the source mixture components.

For each mixture component i, Semi-MPPCA provides responsibility-weighted statistics computed over the source domain:

$$\mu_i^S = \frac{\sum_{x_s \in \mathcal{X}_s} r_i(x_s)\, g_s(x_s)}{\sum_{x_s \in \mathcal{X}_s} r_i(x_s)}, \tag{50}$$

$$\Sigma_i^S = \frac{\sum_{x_s \in \mathcal{X}_s} r_i(x_s)\, (g_s(x_s) - \mu_i^S)(g_s(x_s) - \mu_i^S)^\top}{\sum_{x_s \in \mathcal{X}_s} r_i(x_s)}, \tag{51}$$

$$\pi_i^S = \frac{1}{|\mathcal{X}_s|} \sum_{x_s \in \mathcal{X}_s} r_i(x_s). \tag{52}$$

Analogous target statistics $(\mu_i^T, \Sigma_i^T, \pi_i^T)$ are computed from the responsibilities $r_i(x_t)$ and features $g_t(x_t)$ over the target set $\mathcal{X}_t$:

$$\mu_i^T = \frac{\sum_{x_t \in \mathcal{X}_t} r_i(x_t)\, g_t(x_t)}{\sum_{x_t \in \mathcal{X}_t} r_i(x_t)}. \tag{53}$$

$$\Sigma_i^T = \frac{\sum_{x_t \in \mathcal{X}_t} r_i(x_t)\, (g_t(x_t) - \mu_i^T)(g_t(x_t) - \mu_i^T)^\top}{\sum_{x_t \in \mathcal{X}_t} r_i(x_t)}. \tag{54}$$

$$\pi_i^T = \frac{1}{|\mathcal{X}_t|} \sum_{x_t \in \mathcal{X}_t} r_i(x_t). \tag{55}$$

PHADA then penalizes deviations between source and target component statistics using the following anchor loss:

$$\mathcal{L}_{\text{anchor}} = w_\mu \sum_{i=1}^{K} \|\mu_i^T - \mu_i^S\|_2^2 \;+\; w_\Sigma \sum_{i=1}^{K} \|\Sigma_i^T - \Sigma_i^S\|_F^2 \;+\; w_\pi \sum_{i=1}^{K} \pi_i^T \big(\log \pi_i^T - \log \pi_i^S\big), \tag{56}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and w_μ, w_Σ, and w_π control the relative contributions of mean, covariance, and mixture-weight alignment. The last term corresponds to the Kullback–Leibler divergence $\mathrm{KL}(\pi^T \,\|\, \pi^S)$ between source and target mixture weights.

Combining the information-maximization objective, pseudo-label refinement, and the component-anchor loss, the target encoder g_t is optimized by minimizing

$$\mathcal{L}_{\mathrm{PHADA}} \;=\; \mathcal{L}_{\mathcal{L}_{div}} \;+\; \mathcal{L}_{\mathcal{L}_{ent}} \;-\; \beta\,\mathcal{L}_{\mathrm{PL}} \;+\; \lambda_a\,\mathcal{L}_{\mathrm{anchor}}, \tag{57}$$

where $\beta > 0$ is a balancing hyper-parameter and $\lambda_a > 0$ balances probabilistic structure preservation through the anchor loss. Throughout adaptation, the classifier h_t remains frozen while only g_t is updated. In practice, prototypes and pseudo-labels are refreshed once per epoch, and the joint use of $\mathcal{L}_{\mathrm{IM}}$, $\mathcal{L}_{\mathrm{anchor}}$, and $\mathcal{L}_{\mathrm{PL}}$ consistently improves performance over information maximization alone.

PHADA unifies information-theoretic adaptation and probabilistic alignment within a source-free learning paradigm. The SHOT-IM pseudo-label objective enables confident and balanced predictions in the target domain, while the component-anchor alignment ensures that the adaptation remains consistent with the probabilistic geometry of the source model. This dual mechanism provides PHADA with strong label efficiency, interpretability, and stability under domain shift, making it highly suitable for adaptive modeling in intelligent building environments.

5.3 Experimental Setup & Results

5.3.1 Datasets

This study builds on the datasets employed in our previous work, Semi-MPPCA, which are now restructured to support domain adaptation experiments. We use four datasets representing complementary aspects of human-centered sensing: two publicly available CASAS datasets for Activity Recognition (AR) [137] and two private datasets collected at the Grenoble Institute of Technology for Occupancy Estimation (OE) [21, 27, 34, 35, 119]. In the prior study, one dataset from each category was used for in-domain training and evaluation. In this extension, we incorporate a second dataset for each task to construct source–target domain pairs, thereby enabling the evaluation of

cross-environment generalization under the proposed domain adaptation framework.

Activity Recognition (AR)

The CASAS dataset comprises labeled sensor event data from multiple smart homes equipped with motion, door, temperature, and light sensors. For this study, we select data from two distinct environments (hh101 as the source domain and hh102 as the target domain) to simulate domain shifts arising from differences in spatial configuration and activity distribution. The same three (Watching TV, Toileting and Cooking breakfast) and five activity classes (Watching TV, Toileting, Cooking breakfast, Dressig and Sleeping) from our prior setup are retained to ensure comparability, and the preprocessing protocol, including feature extraction, normalization, and class balancing, remains unchanged.

Occupancy Estimation (OE)

The private Grenoble dataset provides labeled occupancy information alongside environmental and motion features acquired through an ambient sensing network. In this study, we extend this setting by incorporating a second dataset, using office H358 as the source domain and office H355 as the target domain, both located within the Grenoble Institute of Technology and instrumented with comparable sensor infrastructures. This inclusion enables an evaluation of cross-building generalization, reflecting deployment scenarios in which a model trained in one space must adapt to another with similar sensing modalities but different spatial layouts and occupancy behaviors. Following prior work, both datasets support binary and three-class occupancy prediction tasks. Although the original data include richer occupancy annotations, our experiments focus on three levels:0 , 1, and 2. All datasets were collected from heterogeneous Internet of Things (IoT) sensors, including power consumption, CO_2 concentration, humidity, temperature, door and window contact sensors, and acoustic pressure measurements.

By leveraging the same datasets in a domain-adaptive setting, this study maintains methodological continuity with our prior work while extending its scope to cross-domain generalization which

is a critical step toward scalable, environment-agnostic smart building analytics.

5.3.2 Base Models

K-means

K-means is an unsupervised learning algorithm that partitions data into clusters by assigning each point to the nearest cluster centroid. It iteratively updates centroids to minimize the distance within clusters, grouping similar data points together.

K-means AutoEncoder

K-means AutoEncoder combines an autoencoder neural network with K-means clustering. The autoencoder learns a low-dimensional latent representation of the input data by compressing and reconstructing it, capturing essential features while reducing dimensionality. K-means is then applied to cluster these latent features, enabling improved grouping compared to clustering raw data. This approach leverages deep feature learning with traditional clustering for better unsupervised classification.

5.3.3 Cluster Alignment with a Teacher

Cluster Alignment with a Teacher (CAT) [173] is a deep unsupervised domain adaptation method that enhances alignment by incorporating discriminative class-conditional structures from both source and target domains. It utilizes a teacher-student paradigm where an implicit ensemble teacher model provides pseudo labels for the unlabeled target data, enabling reliable discovery of class structures in the target feature space. CAT optimizes a combination of supervised loss on the source data and

two unsupervised objectives: a discriminative clustering loss that encourages intra-class compactness and inter-class separation,

$$L_c(X) = \frac{1}{|X|^2} \sum_{i=1}^{|X|} \sum_{j=1}^{|X|} [\delta_{ij} d(f(x_i), f(x_j)) + (1 - \delta_{ij}) \max(0, m - d(f(x_i), f(x_j)))] \tag{58}$$

where d measures feature distance, δ_{ij} indicates whether instances share the same (pseudo) label, and m is a margin.

and a cluster alignment loss that matches source and target class centroids,

$$L_a = \frac{1}{K} \sum_{k=1}^{K} \|\lambda_{s,k} - \lambda_{t,k}\|_2^2 . \tag{59}$$

where $\lambda_{s,k}$ and $\lambda_{t,k}$ are computed as

$$\lambda_{s,k} = \frac{1}{|X_{s,k}|} \sum_{x_i^s \in X_{s,k}} f(x_i^s), \qquad \lambda_{t,k} = \frac{1}{|X_{t,k}|} \sum_{x_i^t \in X_{t,k}} f(x_i^t). \tag{60}$$

where $X_{s,k}$ denotes the subset of X_s containing all source samples with ground-truth label k, and $X_{t,k}$ denotes the subset of X_t consisting of all target samples assigned to class k by the teacher classifier h.

5.3.4 SHOT-IM

The Source Hypothesis Transfer Info-Max (SHOT-IM) [136] model is a source-free domain adaptation method that optimizes an information maximization objective to adapt a well-trained source classifier to the target domain. Since the classifier head remains fixed, adaptation is achieved solely by updating the feature extractor to produce confident and diverse target predictions. SHOT-IM minimizes conditional entropy while maximizing marginal entropy, encouraging well-separated target clusters.

5.3.5 PHADA without Anchors

PHADA (without anchors) variant corresponds to the proposed model but excludes its main innovation: the anchor-guided alignment mechanism. We first train a Semi-MPPCA model on the source domain and we then apply SHOT-IM pseudo-labeling on minimal target data for unsupervised domain adaptation. We include this variant to illustrate the contribution of our anchor-based mechanism, allowing us to directly compare performance and show how the introduction of anchors leads to more stable and accurate domain alignment.

5.3.6 Metrics

To evaluate model performance, we primarily rely on the F1 score. While accuracy provides a simple measure of overall correctness, it often overlooks the trade-off between precision and recall, especially when both false positives and false negatives are of concern. The F1 score, defined as the harmonic mean of precision and recall, captures this balance by combining both aspects of classification performance into a single value. This makes it a more reliable indicator of a model's discriminative ability, even under balanced data conditions.

To compute the F1 score, two key quantities are first determined: Precision (P) and Recall (R), given by Equation 61 and Equation 62 respectively. Precision measures the proportion of correctly predicted positive instances among all predicted positives, whereas Recall measures the proportion of actual positive instances that are correctly identified. These metrics are expressed as:

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{61}$$

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{62}$$

The F1 score, given by Equation 63, integrates both Precision and Recall into a single metric that rewards models achieving a balanced trade-off between the two:

$$\text{F1 Score} = \frac{2 \times P \times R}{P + R} \tag{63}$$

By encapsulating both types of classification errors, the F1 score provides a nuanced and informative measure of predictive performance, enabling more reliable model comparisons and refinements.

For completeness, we also report accuracy as a supplementary measure. Accuracy, given by Equation 64, reflects the overall proportion of correctly classified instances relative to the total number of predictions and is defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100 \tag{64}$$

While accuracy offers a broad overview of model correctness, the F1 score remains the preferred evaluation criterion in this study due to its sensitivity to both precision and recall, ensuring a balanced and equitable assessment of model performance.

5.3.7 Experimental setup

The experimental setup for PHADA closely follows the protocol used in our previous Semi-MPPCA study, ensuring consistency and enabling fair comparisons. All models, including PHADA and the baseline approaches, are evaluated on the OE dataset for binary and three-class classification tasks, and on the AR dataset for three-class and five-class settings to provide a comprehensive performance assessment. Each reported result represents the mean over 10 independent runs, with 30% of the data randomly selected as the test set in each iteration. To analyze the impact of target-domain data availability in the unsupervised adaptation setting, we vary the number of unlabeled target samples used during training. Specifically, we consider 10, 20, 30, 40, and 50 unlabeled instances for the binary and five-class tasks, and 15, 30, 45, 60, and 75 unlabeled instances for the three-class tasks, using balanced sampling across classes. This design enables a thorough evaluation of PHADA across various target-domain settings without using labeled data. For activity recognition (AR), the source domain uses the same 32-dimensional representation described in Chapter 5. As in earlier experiments, 2,000 samples per class are initially available; however, for PHADA we

restrict source supervision to only 50 labeled samples per class, reflecting the highly constrained conditions assumed in this chapter. On the target side, we construct a balanced dataset by sampling 2,000 instances per class, reserving 30% for testing. From the remaining target samples, only a very small subset of unlabeled data is retained for adaptation, consistent with the strict PHADA evaluation setting described above. For occupancy estimation (OE), we adopt the same preprocessing steps used in previous chapters. In this case, however, we operate directly on the original imbalanced OE dataset (with distributional details provided in Chapter 2). As in AR, PHADA limits source supervision to 50 labeled samples per class. After creating the target set and holding out 30% for testing, only a restricted portion of the remaining unlabeled target samples is used for unsupervised adaptation. This configuration allows PHADA to be evaluated under an imbalanced, low-supervision scenario, closely reflecting the constraints encountered in real smart-building environments.

K-means is implemented using scikit-learn with its default initialization strategy and hyperparameter configuration. For the K-means Autoencoder Variant, we use an encoder–decoder architecture consisting of a single hidden layer of 64 ReLU units and an 8-dimensional latent space. The autoencoder is trained for 10 epochs using MSE loss, the Adam optimizer (lr = 1e-3), and a batch size of 64. After training, latent representations from both source and target domains are extracted and clustered with scikit-learn's default KMeans implementation, using a number of clusters equal to the number of classes. Cluster assignments are then mapped to class labels via majority voting on the training portion of the data.

For CAT, source and target datasets are loaded into PyTorch DataLoaders with a batch size of 128, shuffling the source data while iterating over the target data in parallel. The backbone consists of an input layer matching the feature dimension, a hidden layer of 256 ReLU units with batch normalization, and a 128-dimensional feature output. A single linear layer serves as the classifier head. Training proceeds for 50 epochs using the Adam optimizer (lr = 1e-3) with gradient clipping at a norm of 5.0. The EMA teacher model uses momentum 0.98, and all computations are performed on GPU when available. Target-domain evaluation is performed using the EMA teacher network.

For SHOT-IM, the source model is a supervised feedforward neural classifier with an input dimension matching the feature space, a 128-unit hidden layer, and an output layer with `num_classes` units. It is trained for 10 epochs with batch size 128 using the Adam optimizer (lr = 0.001) and

sparse categorical cross-entropy on scaled source data. SHOT-IM adaptation is then performed by cloning the source model, freezing the classifier head, and updating only the encoder using unlabeled target data (batch size 512) for 18 epochs. Optimization uses SGD with momentum 0.9, Nesterov acceleration, and a polynomial learning-rate decay schedule (initial lr = 0.006, $\gamma = 7.2$, power = 0.6). The model minimizes the InfoMax loss, and gradients are clipped to a norm of 5.0.

PHADA and its ablation without anchors follow the same source-model training protocol as Semi-MPPCA. To enhance label efficiency, the source model is trained using only 50 labeled instances per class. The domain adaptation stage mirrors the SHOT-IM configuration. PHADA without anchors adds pseudo-labeling to improve target supervision, while the full PHADA model further incorporates anchor-based alignment.

5.3.8 Results

Binary Occupancy Estimation

The results in Table 5.1 provide a detailed comparison of binary occupancy estimation performance across several base models using the OE Dataset. PHADA, the proposed framework, consistently outperforms the various baseline models as the number of unlabeled target instances increases. With 10 unlabeled instances, PHADA achieves an F1 score of 91.19, surpassing its ablation without Anchors (89.28%) as well as the baseline models SHOT-IM (80.95%) and K-means AutoEncoder (78.30%). As the number of unlabeled instances rises to 50, PHADA's score continues to climb, reaching 94.45, while SHOT-IM reaches 86.41 and K-means AutoEncoder 85.05%. This indicates that PHADA's strong performance is primarily due to its ability to effectively leverage knowledge from the labeled source dataset, facilitating robust extraction of domain-invariant features even when faced with domain shift. Examining K-means and its AutoEncoder variant shows scores increasing from 78.03% and 78.30% at 10 instances, to 83.93% and 85.05% at 50 instances, respectively. CAT lags considerably in lower unlabeled samples (67.11% at 10), though its relative improvement by 50 instances climbs to 80.30%; still consistently beaten by PHADA at each data increment. Notably, SHOT-IM is the strongest of the competitive baselines, starting at 80.95% and

ending at 86.41%, but its growth rate does not match PHADA's. PHADA without Anchors also shows high scores, maintaining above 89% at all settings, but the full PHADA framework improves upon this by approximately 2-3 F1 points at each step, confirming the benefit of anchoring in unsupervised adaptation. Overall, results across all models highlight that increasing unlabeled target data enhances unsupervised domain adaptation performance, but PHADA's design more efficiently leverages this data, gaining a performance advantage as the problem scales.

Table 5.1: F1 score for binary Classification on OE Dataset

Method	**10 samples**	**20 samples**	**30 samples**	**40 samples**	**50 samples**
K-means	78.03	80.01	81.13	83.41	83.93
K-means AutoEncoder	78.30	79.27	80.42	83.86	85.05
CAT	67.11	72.43	77.57	78.04	80.30
SHOT-IM	80.95	82.25	84.75	85.49	86.41
PHADA without Anchors	89.28	90.60	90.70	90.75	91.23
PHADA	**91.19**	**91.86**	**93.21**	**93.34**	**94.45**

3-class Occupancy Estimation

The 3-class occupancy estimation results in Table 5.2 further reinforce PHADA's superior adaptation capability. PHADA consistently achieves the highest F1 scores across all unlabeled target instances, ranging from 87.86% at 15 labels to 92.27% at 75 labels. Its performance improvement over the next best variant, PHADA without anchors, somewhat similar the binary case, still highlights the pivotal role of anchoring in enhancing domain alignment and class discrimination. Baseline methods show mixed trends: K-means steadily improves but plateaus near the higher label counts, while its AutoEncoder variant continues to improve steadily, while CAT performs poorly at low unlabeled data amounts (44.51% at 15 labels), though improving to 71.21% at 75 labels. This again illustrates CAT's difficulty with limited target data, in contrast to PHADA which maintains strong classification even in low-resource scenarios. SHOT-IM performs competitively among baselines, but PHADA outperforms it while also using fewer source labels, illustrating PHADA's greater label efficiency and robustness. Overall, these results demonstrate PHADA's strength in leveraging

source knowledge through anchor-guided adaptation to achieve consistent high accuracy across increasing target data, further validating the framework's domain adaptation design in multi-class occupancy estimation.

Table 5.2: F1 score for 3-class Classification on OE Dataset

Method	**15 samples**	**30 samples**	**45 samples**	**60 samples**	**75 samples**
K-means	72.62	76.39	76.41	76.65	76.76
K-means AutoEncoder	72.15	76.06	80.00	80.44	83.18
CAT	44.51	58.53	59.63	68.23	71.21
SHOT-IM	70.56	72.57	77.75	78.60	80.60
PHADA without Anchors	85.98	86.63	86.92	88.19	89.47
PHADA	**87.86**	**88.86**	**89.89**	**90.21**	**92.27**

3-class Activity Recognition

The 3-class activity recognition results in Table 5.3 demonstrate PHADA's superior performance and robustness compared to baseline methods. PHADA achieves the highest F1 scores across all unlabeled target data sizes, starting at 93.17% with 15 labels and improving to 95.16% at 75 labels, outperforming PHADA without anchors and the strong SHOT-IM baseline. Baselines such as K-means and K-means AutoEncoder show gradual improvement with more data but remain far behind. CAT exhibits a peculiar trend with low performance initially (29.62% at 15 labels) but a spike at 75 labels, suggesting instability or sensitivity to data size. SHOT-IM performs strongly, especially at larger label sizes, but PHADA's consistent advantage highlights its superior capability in leveraging source knowledge efficiently and adapting to the target domain with fewer labeled source samples. The anchoring mechanism further contributes to stable improvements by providing effective domain alignment. These results confirm PHADA's effectiveness in human activity recognition tasks, combining label efficiency and robust feature adaptation to handle varying amounts of unlabeled target data while maintaining high classification accuracy.

Table 5.3: F1 score for 3-class Classification on AR Dataset

Method	15 samples	30 samples	45 samples	60 samples	75 samples
K-means	32.13	38.82	41.15	42.75	44.72
K-means AutoEncoder	37.03	52.37	65.45	72.74	73.88
CAT	29.62	29.71	42.35	46.42	57.94
SHOT-IM	84.08	87.20	86.68	91.07	92.23
PHADA without Anchors	91.96	92.99	93.40	93.43	94.15
PHADA	**93.17**	**93.41**	**93.94**	**94.42**	**95.16**

5-class Activity Recognition

The 5-class activity recognition results in Table 5.4 highlight PHADA's ability to sustain high and steadily improving classification performance as task complexity increases. Unlike the previous sub-sections, here the gains of PHADA over baselines are especially meaningful given the increased number of classes, which typically exacerbates the domain shift challenge. PHADA achieves F1 scores starting from 80.60% with just 10 labels, progressing steadily to 86.75% at 50 labels, outperforming its variant without anchors (from 78.66% to 84.33%) and all baseline models, including SHOT-IM, which ranges from 75.69% to 85.59%. Traditional clustering approaches like K-means variants show limited improvement, with K-means AutoEncoder rising to 56.58% at 50 labels and CAT performing the weakest overall (below 34% throughout). This contrast highlights PHADA's effectiveness in complex multi-class activity classification scenarios where class overlap and domain shift pose heightened challenges.

Table 5.4: F1 score for 5-class Classification on AR Dataset

Method	10 samples	20 samples	30 samples	40 samples	50 samples
K-means	31.86	33.28	34.93	36.85	37.30
K-means AutoEncoder	33.36	39.07	41.56	45.25	56.58
CAT	19.75	23.55	28.60	32.85	33.45
SHOT-IM	75.69	75.92	77.45	80.98	85.59
PHADA without Anchors	78.66	79.80	80.28	83.28	84.33
PHADA	**80.60**	**82.72**	**83.51**	**85.97**	**86.75**

Overall, K-means and its AutoEncoder extension provide a useful baseline and show progressive

improvements with greater unlabeled data. As seen in the results, K-means AutoEncoder consistently yields higher F1 scores than K-means across all tasks, demonstrating superior representation learning capacity. However, like K-means, its gains taper off in more challenging multi-class or low-label scenarios where domain shifts are more pronounced. Their relatively limited ability to leverage source domain knowledge and align complex target domain distributions restricts their adaptability compared to PHADA.

Although CAT is recognized as a strong unsupervised domain adaptation method and achieves substantial improvements when more data is available, its performance drops when trained with very limited target data. Unlike PHADA, CAT struggles to maintain strong adaptation capability in low-data regimes, whereas PHADA effectively leverages source knowledge and its anchoring mechanism to deliver robust performance even with very sparse target data.

It is important to note that PHADA achieves superior results not only because of its adaptation mechanism, but also due to its label efficiency. Specifically, PHADA is designed to use fewer labeled samples from the source domain (semi-supervised setting), whereas the SHOT-IM baseline typically relies on a well-performing source model that is trained using fully supervised data. Although SHOT-IM delivers strong results among the baselines, its effectiveness is highly dependent on the availability of substantial source label information. In contrast, PHADA consistently outperforms SHOT-IM across all unlabeled target data settings, demonstrating that it not only performs better but also does so with access to significantly fewer labeled samples, highlighting its superior label efficiency and adaptability.

The inclusion of anchors in PHADA clearly demonstrates its importance and effectiveness. PHADA with anchors consistently outperforms the version without anchors by a noticeable margin across all unlabeled data settings. The anchoring mechanism serves as stable reference points that facilitate better alignment of feature distributions between the source and target domains and helps mitigate negative transfer. This results in more precise and semantically consistent feature adaptation, which translates into higher F1 scores. Therefore, the improvement seen with anchors underscores their crucial role in enhancing PHADA's ability to reduce domain shift and improve classification accuracy under unsupervised domain adaptation scenarios, highlighting its suitability for real-world deployment.

5.3.9 Discussion

The baseline methods compared include Source, Fine-tune Target, and Target. The Source model is trained exclusively on source domain data and evaluated directly on the target data without any form of domain adaptation, serving as a natural lower bound for cross-domain performance. The Fine-tune Target variant builds upon a Semi MPPCA model pre-trained on source data, which is then fine-tuned on the limited labeled target data and tested on the target domain, without any explicit domain adaptaion strategies. Lastly, the Target model is a semi-MPPCA trained directly on the target dataset under a semi-supervised setting, serving as an upper bound for performance since it has access to target domain labels during training.

It is important to clarify the choice of baseline models used for comparison and their significance in evaluating the effectiveness of PHADA. The Source model serves as a fundamental benchmark illustrating the limitations of direct transfer without adaptation, exposing domain shift challenges. The Fine-tune Target model reflects practical scenarios where limited labeled target data is used to adapt a pre-trained source model, though without employing a dedicated domain adaptation framework. The Target model represents an idealized upper bound, trained with access to sufficient labeled target data, setting the standard for achievable performance in each classification task.

Analyzing the results for 30 labeled target instances (30 unlabeled target samples in the case of PHADA) in Table 5.5, the Source models consistently yield poor accuracy and F1 scores, such as 27.57% accuracy and 11.92% F1 on binary occupancy estimation, highlighting the significant domain shift between source and target. The Fine-tune Target models markedly improve over Source, for example demonstrating accuracy gains from 13.08% to 72.43% in the 3-class occupancy case; however, their gains vary and are limited in more complex settings such as 5-class activity recognition, where accuracy remains low at 20.00%. The Target models, with full access to target labels, achieve solid performance across all cases, including 92.52% accuracy in binary occupancy and 80.95% in 5-class activity recognition, setting a meaningful performance ceiling.

PHADA surpasses all baselines across all classification tasks with accuracy and F1 scores close to or exceeding those of the Target models, e.g., obtaining 92.99% accuracy and 93.21% F1 in binary occupancy, and 84.10% accuracy and 83.51% F1 in 5-class activity recognition. Notably, PHADA

achieves these top results using only 30 unlabeled target instances while leveraging knowledge from source data and employing anchor-guided domain adaptation. This demonstrates PHADA's superior capability in bridging domain gaps effectively and efficiently, improving performance beyond naive source-only training or limited fine-tuning, and even outperforming the upper bound represented by semi supervised target training.

Table 5.5: Performance comparison score for 3-class Classification on OE and AR Dataset (30 labels)

Method	**Variant**	**Accuracy**	**F1**
Source	Binary OE	27.57	11.92
	3-class OE	13.08	8.25
	3-class AR	33.67	17.37
	5-class AR	23.45	14.03
Fine-tune Target	Binary OE	86.92	86.10
	3-class OE	72.43	52.46
	3-class AR	42.17	31.93
	5-class AR	20.00	6.67
Target	Binary OE	92.52	92.39
	3-class OE	81.31	77.41
	3-class AR	92.17	92.16
	5-class AR	80.95	80.98
PHADA	Binary OE	**92.99**	**93.21**
	3-class OE	**88.32**	**88.86**
	3-class AR	**93.42**	**93.41**
	5-class AR	**84.10**	**83.51**

The one-shot PHADA results, summarized in Table 5.6, demonstrate the model's remarkable ability to adapt with extremely limited target data, using only a single unlabeled target instance per class for training. Despite this extreme data scarcity, PHADA achieves strong accuracy and F1 scores across tasks, such as 87.85% in binary occupancy estimation and 68.85% in the more complex 5-class activity recognition. This highlights its robust generalization and domain alignment capabilities, enabling meaningful adaptation even in highly constrained scenarios where traditional methods would struggle.

PHADA's one-shot adaptation capability demonstrates its ability to capture domain-invariant features while effectively leveraging source knowledge, minimizing reliance on labeled target data.

This scenario reflects real-world conditions where acquiring large annotated or even unlabeled target datasets is often impractical or costly. By achieving competitive performance using only a single unlabeled instance per class, PHADA reduces annotation burdens, accelerates deployment, and maintains robust domain alignment across diverse environments. Moreover, one-shot adaptation challenges conventional domain adaptation methods, which typically require substantial target data to learn domain-invariant features and often suffer from overfitting or poor generalization in low-data regimes. PHADA's anchor-guided approach, however, enables effective knowledge transfer and domain alignment despite the acute scarcity of target samples, highlighting its robustness and flexibility. Studying this extreme low-data setting highlights the framework's scalability, flexibility, and potential for real-world applications, while also motivating further research in ultra-low-data adaptive learning for efficient and deployable AI systems.

Table 5.6: One-shot PHADA

Method	**Accuracy**	**F1**
Binary OE	87.85	87.85
3-class OE	72.90	73.66
3-class AR	79.75	77.94
5-class AR	68.85	65.96

5.4 Conclusion

Modern smart buildings increasingly rely on data-driven models to optimize energy consumption, enhance occupant comfort, and support sustainability objectives. However, achieving accurate occupancy estimation and activity recognition remains challenging due to the scarcity of labeled data, privacy constraints, and significant domain shifts between buildings with different layouts, sensors, and behavioral patterns. Traditional supervised methods are further limited by their reliance on extensive labeled data and their difficulty modeling the complex, nonlinear, and high-dimensional structure of smart-building sensor streams. These obstacles highlight the need for unsupervised and source-free domain adaptation methods that can generalize across environments while minimizing reliance on annotated data. This work presented PHADA (Probabilistic Hypothesis-Anchored Domain Adaptation), a novel framework designed to achieve robust and label-efficient adaptation in

smart-building environments. Building upon the Semi-MPPCA model, employed to address the challenge of modeling complex, high-dimensional, heterogeneous smart-building data, PHADA integrates probabilistic representation learning with the Source Hypothesis Transfer (SHOT) pseoudo-label framework to enable source-free domain adaptation. By freezing the source classifier and optimizing the target feature extractor through an information maximization objective, PHADA ensures confident and diverse predictions on unlabeled target data. Furthermore, the component-anchor alignment mechanism preserves the probabilistic geometry of the source mixture during adaptation, helping stabilize the target feature space and preventing class collapse even without access to source data. Extensive experiments across occupancy estimation and activity recognition datasets and varying amounts of unlabeled target data demonstrate that PHADA consistently outperforms competitive baselines, including SHOT-IM, CAT, and K-means variants and its ablation PHADA without Anchors, while maintaining strong performance in low-data and one-shot adaptation scenarios. The results demonstrate that PHADA achieves robust unsupervised domain adaptation with minimal unlabeled target data, maintaining high accuracy across diverse environments and adapting effectively even in ultra-low-data scenarios. Its anchor component enforces stable alignment by anchoring target components to the source mixture structure., while the semi-supervised source model enables efficient knowledge transfer, highlighting PHADA's potential for real-world deployment in settings where labeled data are scarce. Overall, PHADA contributes a flexible and privacy-preserving approach to domain adaptation, capable of generalizing across buildings without requiring extensive labeling efforts. Future work can extend PHADA toward open-set domain adaptation (OSDA), multi-source adaptation, and improved interpretability, further strengthening its value for intelligent building management and energy-efficient infrastructure.

Chapter 6

Conclusion

The growing need for energy-efficient and context-aware building management systems has intensified efforts to model occupant behavior accurately and robustly. Yet, the development of reliable Occupancy Estimation (OE) and Activity Recognition (AR) systems remains hindered by three main persistent challenges: the scarcity of labeled data, the prevalence of domain shifts across buildings, and the inherently complex, high-dimensional nature of multimodal IoT sensor streams. This thesis addressed these challenges by proposing a suite of explainable, label-efficient, and privacy-preserving domain adaptation frameworks that enable behavioral models to generalize across diverse smart-building environments.

Across the four research components, the thesis introduced methods that progressively relax labeled data requirements, increase interpretability, and enhance transferability. The first contribution, IMB-SER, extended structure-based transfer learning for decision trees to address class imbalance which is an issue often overlooked in occupancy estimation but crucial for capturing rare, operationally significant states. By integrating imbalance-aware restructuring and rule-level interpretability analyses, IMB-SER demonstrated that accuracy, fairness toward minority classes, and transparency can be jointly optimized.

The second contribution, UTA-DT, advanced source-free domain adaptation for AR by leveraging uncertainty cues derived from a pretrained source model. By augmenting target data with hard and soft predictions, entropy and margin-based uncertainty signals, UTA-DT enabled robust cross-domain adaptation without accessing source samples, directly addressing privacy constraints

commonly encountered in building deployments. Its rule-based nature ensured interpretability while maintaining strong generalization performance.

The third contribution, Semi-MPPCA, moved beyond decision trees toward probabilistic deep learning by reformulating the Mixture of Probabilistic Principal Component Analyzers as a neural architecture capable of pretraining on unlabeled data and fine-tuning with minimal labels. Through a combination of probabilistic modeling and a weight-swapping alignment strategy, Semi-MPPCA effectively leveraged the abundance of unlabeled smart-building data with only minimal labeled supervision while maintaining interpretability through mixture structures.

The final and central contribution, PHADA, extended Semi-MPPCA into a fully source-free unsupervised domain adaptation framework. By combining information maximization, probabilistic pseudo-labeling, and a novel component-anchor alignment mechanism, PHADA preserved the geometric and probabilistic structure learned in the source domain while adapting to shifting target distributions. Extensive experiments, including ablation studies, showed that the anchor mechanism is critical for preventing drift and class collapse, and established PHADA as a highly label-efficient and privacy-preserving UDA solution.

Together, these contributions form a coherent progression: from interpretable, imbalance-aware tree-based methods (IMB-SER), to privacy-preserving adaptation without source data (UTA-DT), to probabilistic semi-supervised learning (Semi-MPPCA), and ultimately to robust source-free domain adaptation with probabilistic anchors (PHADA). The collective findings demonstrate that effective behavior modeling in smart buildings does not need to rely on large labeled datasets or direct access to sensitive data. Instead, through principled domain adaptation, uncertainty modeling, and probabilistic latent structures, models can be both transparent and resilient to domain shifts.

6.1 Limitations and Future Work

While the proposed methods substantially improve cross-domain generalization, several avenues remain for future exploration. First, current models focus on closed-set adaptation; extending them to open-set or evolving activity spaces would increase real-world applicability. Second, temporal modeling remains an open challenge; future work may integrate recurrent or attention-based

architectures with probabilistic mixtures to better capture long-range dependencies. Third, incorporating multi-source adaptation and continual learning would allow models to aggregate knowledge across many domains while retaining privacy. Finally, advancing interpretable probabilistic explainability techniques could strengthen operator trust and facilitate human–machine collaboration in building operations.

In summary, this thesis contributes a structured path toward scalable, explainable, and label-efficient behavioral modeling in smart buildings. By simultaneously addressing domain shift, label scarcity, and privacy constraints, it provides the foundation for next-generation adaptive building management systems that optimize energy consumption while supporting more sustainable, intelligent, and occupant-responsive built environments.

Appendix A

Appendix

.1 Probabilistic Principal Component Analysis (PPCA)

Probabilistic Principal Component Analysis (PPCA) extends traditional PCA by formulating a probabilistic model for dimensionality reduction and latent variable inference. Given an observed d-dimensional data vector $\mathbf{x} \in \mathbb{R}^d$, PPCA assumes that $\mathbf{x}$ is generated from a lower-dimensional latent variable $\mathbf{z} \in \mathbb{R}^q$ ($q < d$) according to the linear Gaussian model:

$$\mathbf{x} = \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon} \tag{65}$$

where $\mathbf{W} \in \mathbb{R}^{d \times q}$ is the loading matrix, $\boldsymbol{\mu} \in \mathbb{R}^d$ is the mean vector, and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ represents isotropic Gaussian noise. The latent variable $\mathbf{z}$ is assumed to follow a standard normal distribution:

$$\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}) \tag{66}$$

Integrating out $\mathbf{z}$ yields the marginal distribution of the observed data:

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T + \sigma^2 \mathbf{I}\right) \tag{67}$$

The covariance structure $\mathbf{W}\mathbf{W}^T + \sigma^2 \mathbf{I}$ captures both the principal components of the data and the residual noise, making PPCA equivalent to standard PCA in the limit as $\sigma^2 \to 0$.

The maximum likelihood estimates of $\mathbf{W}$ and σ^2 can be obtained via the Expectation-Maximization (EM) algorithm. During the E-step, the posterior distribution over the latent variables is computed:

$$p(\mathbf{z}|\mathbf{x}) = \mathcal{N}\left(\mathbf{M}^{-1}\mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu}),\, \sigma^2\mathbf{M}^{-1}\right)$$

where $\mathbf{M} = \mathbf{W}^T\mathbf{W} + \sigma^2\mathbf{I}$. In the M-step, the parameters $\mathbf{W}$ and σ^2 are updated to maximize the likelihood.

Unlike standard PCA, which finds deterministic projections, PPCA estimates latent variables probabilistically, allowing for Bayesian inference and handling missing data effectively. The maximum likelihood solution for $\mathbf{W}$ can be obtained through Expectation-Maximization (EM) or Singular Value Decomposition (SVD) techniques. PPCA provides a robust probabilistic framework for dimensionality reduction, improving interpretability and flexibility in machine learning applications.

.2 Mixture of Probabilistic Principal Component Analyzers (MPPCA)

The mixture of probabilistic principal component analyzers (MPPCA) extends PPCA by modeling complex, high-dimensional, and heterogeneous datasets using multiple local PPCA models. While PPCA assumes that the entire dataset follows a single global probabilistic structure, MPPCA partitions the data into K clusters, each represented by its own PPCA component. Each cluster is modeled as a local Gaussian distribution with a low-dimensional subspace, and the overall data distribution is expressed as a weighted sum of these Gaussian components. This approach effectively combines dimensionality reduction with clustering, making it particularly well-suited for analyzing heterogeneous data.

MPPCA is closely related to Gaussian Mixture Models (GMMs), but it offers the additional advantage of reducing dimensionality within each cluster. Moreover, GMM models full covariance structures (or diagonal, spherical), which can become computationally expensive and unstable in high dimensions. MPPCA imposes a low-dimensional structure on the covariance matrix within each cluster, making it more efficient for high-dimensional data by combining clustering with dimensionality reduction. Thus, MPPCA can be seen as a special case or extension of GMM that

incorporates a low-dimensional subspace (via PPCA) within each Gaussian component, making it particularly suitable for clustering high-dimensional data with underlying low-rank structure.

Assuming the data consists of N observations $\mathbf{x}_n \in \mathbb{R}^D$, MPPCA models the data as being generated from a mixture of M PPCA components. Each component represents a cluster with its own local low-dimensional subspace.

Each observation $\mathbf{x}_n$ is modeled with a latent variable $\mathbf{z}_{ni} \in \mathbb{R}^q$, with $q < D$, corresponding to each component i. The completed data log-likelihood is expressed as:

$$L_C = \sum_{n=1}^{N} \sum_{i=1}^{M} r_{ni} \ln \{\pi_i p(\mathbf{x}_n, \mathbf{z}_{ni})\} \tag{68}$$

where $\pi_i \geq 0$ are the mixing proportions such that $\sum_{i=1}^{M} \pi_i = 1$, and r_{ni} indicates the responsibility of component i for generating observation $\mathbf{x}_n$.

Each component is characterized by its own mean vector $\mu_i \in \mathbb{R}^D$, factor loading matrix $\mathbf{W}_i \in \mathbb{R}^{D \times q}$, and noise variance σ_i^2. The model parameters are optimized using the Expectation-Maximization (EM) algorithm.

For the E-step, the posterior responsibility for component given observation is computed as:

$$R_{ni} = p(i|\mathbf{x}_n) = \frac{p(\mathbf{x}_n|i)\pi_i}{p(\mathbf{x}_n)} \tag{69}$$

where the likelihood of the component is:

$$p(\mathbf{x}_n|i) = (2\pi)^{-D/2} |\mathbf{C}_i|^{-1/2} \exp\{-\frac{1}{2}(\mathbf{x}_n - \mu_i)^T \mathbf{C}_i^{-1} (\mathbf{x}_n - \mu_i)\} \tag{70}$$

and the covariance of the component is:

$$\mathbf{C}_i = \sigma_i^2 \mathbf{I} + \mathbf{W}_i \mathbf{W}_i^T \tag{71}$$

The expectations of the latent variables are computed as:

$$\langle \mathbf{z}_{ni} \rangle = \mathbf{M}_i^{-1} \mathbf{W} i^T (\mathbf{x}_n - \mu_i) \tag{72}$$

$$\langle \mathbf{z}_{ni}\mathbf{z}_{ni}^T \rangle = \sigma_i^2 \mathbf{M} i^{-1} + \langle \mathbf{z}_{ni} \rangle \langle \mathbf{z}_{ni} \rangle^T \tag{73}$$

where:

$$\mathbf{M}_i = \sigma_i^2 \mathbf{I} + \mathbf{W}_i^T \mathbf{W}_i \tag{74}$$

The M-Step consists of 2 stages.

Stage 1: Ignoring latent variables Maximizing the complete-data log-likelihood for the mixing coefficients and means:

$$\tilde{\pi}_i = \frac{1}{N} \sum_{n=1}^{N} R_{ni} \tag{75}$$

$$\tilde{\boldsymbol{\mu}}_i = \frac{\sum_{n=1}^{N} R_{ni} \mathbf{x}_n}{\sum_{n=1}^{N} R_{ni}} \tag{76}$$

Stage 2: Incorporating latent variables Define the scatter matrix:

$$\mathbf{S}_i = \frac{1}{\tilde{\pi}_i N} \sum_{n=1}^{N} R_{ni} (\mathbf{x}_n - \tilde{\boldsymbol{\mu}}_i)(\mathbf{x}_n - \tilde{\boldsymbol{\mu}}_i)^T \tag{77}$$

The loading matrix and noise variance are updated as follows:

$$\tilde{\mathbf{W}}_i = \mathbf{S}_i \mathbf{W}_i \left(\sigma_i \mathbf{I} + \mathbf{M}_i^{-1} \mathbf{W}_i^T \mathbf{S}_i \mathbf{W}_i \right)^{-1} \tag{78}$$

$$\tilde{\sigma}_i^2 = \frac{1}{D} \mathrm{tr} \left(\mathbf{S}_i - \mathbf{S}_i \mathbf{W}_i \mathbf{M}_i^{-1} \tilde{\mathbf{W}}_i^T \right) \tag{79}$$

The EM algorithm alternates between the E-step and M-step:

(1) Compute responsibilities using (10).

(2) Update the mixing coefficients and means using (16) and (17).

(3) Update loading matrices and noise variances using (19) and (20).

This iterative procedure progressively refines the parameter estimates by alternating between assigning data points to mixture components and updating the component parameters. Each iteration is guaranteed to increase the likelihood until convergence to a locally optimal solution is achieved.

Bibliography

[1] Xilei Dai, Junjie Liu, and Xin Zhang. A review of studies applying machine learning models to predict occupancy and window-opening behaviours in smart buildings. *Energy and Buildings*, 223:110159, 2020.

[2] Kyle A Barber and Moncef Krarti. A review of optimization based tools for design and control of building energy systems. *Renewable and Sustainable Energy Reviews*, 160:112359, 2022.

[3] Yiqun Pan, Mingya Zhu, Yan Lv, Yikun Yang, Yumin Liang, Ruxin Yin, Yiting Yang, Xiaoyu Jia, Xi Wang, Fei Zeng, et al. Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. *Advances in Applied Energy*, 10:100135, 2023.

[4] Qing Li, Lianying Zhang, Limao Zhang, and Xianguo Wu. Optimizing energy efficiency and thermal comfort in building green retrofit. *Energy*, 237:121509, 2021.

[5] J Aguilar, Alberto Garces-Jimenez, Maria Dolores R-Moreno, and Rodrigo García. A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings. *Renewable and Sustainable Energy Reviews*, 151:111530, 2021.

[6] Guofeng Qiang, Shu Tang, Jianli Hao, Luigi Di Sarno, Guangdong Wu, and Shaoxing Ren. Building automation systems for energy and comfort management in green buildings: A critical review and future directions. *Renewable and Sustainable Energy Reviews*, 179:113301, 2023.

[7] Maedeh Motalebi, Ali Rashidi, and Mohammad Mahdi Nasiri. Optimization and bim-based lifecycle assessment integration for energy efficiency retrofit of buildings. *Journal of Building Engineering*, 49:104022, 2022.

[8] Natural Resources Canada. The canada green buildings strategy: Transforming canada's buildings sector for a net-zero and resilient future. https://natural-resources.canada.ca/energy-efficiency/building-energy-efficiency/canada-green-buildings-strategy-transforming-canada-s-buildings-sector-n 2025. Accessed: 2025-11-18.

[9] Schneider Electric. Four proven strategies for unlocking building decarbonization. https://blog.se.com/sustainability/2025/02/04/four-strategies-unlocking-building-decarbonization/, February 2025. Accessed: 2025-11-18.

[10] Canada Green Building Council (CAGBC). Budget 2025: Progress and gaps. https://www.cagbc.org/news-resources/cagbc-news/budget-2025-progress-and-gaps/, November 2025. Accessed: 2025-11-18.

[11] Xing Lu, Saptarshi Bhattacharya, Himanshu Sharma, Veronica Adetola, and Zheng O'Neill. Sensor impact evaluation in commercial buildings: The case of occupancy-centric controls. *Energy and Buildings*, 267:112134, 2022.

[12] Prashant Anand, David Cheong, and Chandra Sekhar. A review of occupancy-based building energy and ieq controls and its future post-covid. *Science of the Total Environment*, 804: 150249, 2022.

[13] Brodie W Hobson, Tareq Abuimara, H Burak Gunay, and Guy R Newsham. How do buildings adapt to changing occupancy? a natural experiment. In *Building Simulation 2021*, volume 17, pages 3465–3472. IBPSA, 2021.

[14] Daniel Minoli, Kazem Sohraby, and Benedict Occhiogrosso. Iot considerations, requirements, and architectures for smart buildings—energy optimization and next-generation building management systems. *IEEE Internet of Things Journal*, 4(1):269–283, 2017.

[15] partnership21. Smart buildings and iot, May 2024. URL https://www.greendesignconsulting.com/single-post/smart-buildings-and-iot.

[16] Hari Prasanna Das, Yu-Wen Lin, Utkarsha Agwan, Lucas Spangher, Alex Devonport, Yu Yang, Ján Drgoňa, Adrian Chong, Stefano Schiavon, and Costas J Spanos. Machine learning for smart and energy-efficient buildings. *Environmental Data Science*, 3:e1, 2024.

[17] Manar Amayri, Stephane Ploix, Hussain Kazmi, Quoc-Dung Ngo, and EL Abed EL Safadi. Estimating occupancy from measurements and knowledge using the bayesian network for energy management. *Journal of Sensors*, 2019(1):7129872, 2019.

[18] Giovanni Diraco, Gabriele Rescio, Andrea Caroppo, Andrea Manni, and Alessandro Leone. Human action recognition in smart living services and applications: context awareness, data availability, personalization, and privacy. *Sensors*, 23(13):6040, 2023.

[19] Jawher Dridi, Manar Amayri, and Nizar Bouguila. Unsupervised domain adaptation with and without access to source data for estimating occupancy and recognizing activities in smart buildings. *Building and Environment*, 243:110651, 2023.

[20] Jawher Dridi, Manar Amayri, and Nizar Bouguila. Unsupervised domain adaptation without source data for estimating occupancy and recognizing activities in smart buildings. *Energy and Buildings*, 303:113808, 2024.

[21] Jawher Dridi, Manar Amayri, and Nizar Bouguila. Transfer learning for estimating occupancy and recognizing activities in smart buildings. *Building and Environment*, 217:109057, 2022.

[22] Omar Mata, Juana Isabel Mendez, Pedro Ponce, Therese Peffer, Alan Meier, and Arturo Molina. Energy savings in buildings based on image depth sensors for human activity recognition. *Energies*, 16(3):1078, 2023.

[23] Mi-Lim Kim, Keon-Jun Park, and Sung-Yong Son. Occupancy-based energy consumption estimation improvement through deep learning. *Sensors*, 23(4):2127, 2023.

[24] Omar Bouhamed, Manar Amayri, and Nizar Bouguila. Weakly supervised occupancy prediction using training data collected via interactive learning. *Sensors*, 22(9):3186, 2022.

[25] Manar Amayri, Stephane Ploix, Nizar Bouguila, and Frederic Wurtz. Estimating occupancy using interactive learning with a sensor environment: Real-time experiments. *IEEE Access*, 7:53932–53944, 2019. doi: 10.1109/ACCESS.2019.2911921.

[26] Zhiwen Luo, Manar Amayri, Wentao Fan, and Nizar Bouguila. Cross-collection latent beta-liouville allocation model training with privacy protection and applications. *Applied Intelligence*, 53(14):17824–17848, 2023.

[27] Manar Amayri, Abhay Arora, Stephane Ploix, Sanghamitra Bandhyopadyay, Quoc-Dung Ngo, and Venkata Ramana Badarla. Estimating occupancy in heterogeneous sensor environment. *Energy and Buildings*, 129:46–58, 2016. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2016.07.026. URL https://www.sciencedirect.com/science/article/pii/S0378778816306223.

[28] Muhammad Waseem Ahmad, Monjur Mourshed, David Mundow, Mario Sisinni, and Yacine Rezgui. Building energy metering and environmental monitoring–a state-of-the-art review and directions for future research. *Energy and Buildings*, 120:85–102, 2016.

[29] Amos Storkey et al. When training and test sets are different: characterizing learning transfer. *Dataset shift in machine learning*, 30(3-28):6, 2009.

[30] Mayee Chen, Karan Goel, Nimit S Sohoni, Fait Poms, Kayvon Fatahalian, and Christopher Re. Mandoline: Model evaluation under distribution shift. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 1617–1629. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/chen21i.html.

[31] Marvin Zhang, Henrik Marklund, Nikita Dhawan, Abhishek Gupta, Sergey Levine, and Chelsea Finn. Adaptive risk minimization: Learning to adapt to domain shift. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in*

Neural Information Processing Systems, volume 34, pages 23664–23678. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/c705112d1ec18b97acac7e2d63973424-Paper.pdf.

[32] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2507–2516, 2019.

[33] Karin Stacke, Gabriel Eilertsen, Jonas Unger, and Claes Lundström. Measuring domain shift for deep learning in histopathology. *IEEE journal of biomedical and health informatics*, 25 (2):325–336, 2020.

[34] Naailah Mahamoodally, Jawher Dridi, and Manar Amayri. Explainable domain adaptation for imbalanced occupancy estimation. *Journal of Building Engineering*, 97:110613, 2024.

[35] Naailah Mahamoodally, Viet Tra, and Manar Amayri. Semi-supervised mixture of probabilistic principal component analyzers for modeling human behavior. *Energy and Buildings*, page 116145, 2025.

[36] IoT for Buildings. Iot integration challenges in commercial buildings. https://iotforbuildings.com/iot-integration-challenges/, 2025. Accessed: YYYY-MM-DD.

[37] Xiachong Lin, Arian Prabowo, Imran Razzak, Hao Xue, Matthew Amos, Sam Behrens, and Flora D Salim. A gap in time: The challenge of processing heterogeneous iot data in digitalized buildings. *IEEE Pervasive Computing*, 2025.

[38] Viet Tra, Manar Amayri, and Nizar Bouguila. Unsupervised outlier detection using neural network-based mixtures of probabilistic principal component analyzers for building chiller fault diagnosis. *Building and Environment*, 225:109620, 2022. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2022.109620. URL https://www.sciencedirect.com/science/article/pii/S0360132322008502.

[39] Tom Narock, J Michael Johnson, Justin Singh-Mohudpur, and Arash Modaresi Rad. Building occupancy type classification and uncertainty estimation using machine learning and open data. *Environmental Data Science*, 4:e10, 2025.

[40] Ahmad Almadhor, Nejib Ghazouani, Belgacem Bouallegue, Natalia Kryvinska, Shtwai Alsubai, Moez Krichen, Abdullah Al Hejaili, and Gabriel Avelino Sampedro. Digital twin based deep learning framework for personalized thermal comfort prediction and energy efficient operation in smart buildings. *Scientific Reports*, 15(1):24654, 2025.

[41] Kumar Prabhakaran, Jawher Dridi, Manar Amayri, and Nizar Bouguila. Explainable k-means clustering for occupancy estimation. *Procedia Computer Science*, 203:326–333, 2022.

[42] Peeyush Singhal, Rahee Walambe, Sheela Ramanna, and Ketan Kotecha. Domain adaptation: challenges, methods, datasets, and applications. *IEEE access*, 11:6973–7020, 2023.

[43] Jiangtao Peng, Yi Huang, Weiwei Sun, Na Chen, Yujie Ning, and Qian Du. Domain adaptation in remote sensing image classification: A survey. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:9842–9859, 2022.

[44] Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R. Arabnia. A brief review of domain adaptation. In Robert Stahlbock, Gary M. Weiss, Mahmoud Abou-Nasr, Cheng-Ying Yang, Hamid R. Arabnia, and Leonidas Deligiannidis, editors, *Advances in Data Science and Information Engineering*, pages 877–894, Cham, 2021. Springer International Publishing. ISBN 978-3-030-71704-9.

[45] Jingjing Li, Zhiqi Yu, Zhekai Du, Lei Zhu, and Heng Tao Shen. A comprehensive survey on source-free domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5743–5762, 2024.

[46] David Berthelot, Rebecca Roelofs, Kihyuk Sohn, Nicholas Carlini, and Alex Kurakin. Adamatch: A unified approach to semi-supervised learning and domain adaptation. *arXiv preprint arXiv:2106.04732*, 2021.

[47] Shiqi Yang, Shangling Jui, Joost Van De Weijer, et al. Attracting and dispersing: A simple approach for source-free domain adaptation. *Advances in Neural Information Processing Systems*, 35:5802–5815, 2022.

[48] Mengqiu Xu, Ming Wu, Kaixin Chen, Chuang Zhang, and Jun Guo. The eyes of the gods: A survey of unsupervised domain adaptation methods based on remote sensing data. *Remote Sensing*, 14(17):4380, 2022.

[49] Noam Segev, Maayan Harel, Shie Mannor, Koby Crammer, and Ran El-Yaniv. Learn on source, refine on target: A model transfer learning framework with random forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1811–1824, 2017. doi: 10.1109/TPAMI.2016.2618118.

[50] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning, 2020.

[51] Jose M. Alvarez, Kristen M. Scott, Bettina Berendt, and Salvatore Ruggieri. Domain adaptive decision trees: Implications for accuracy and fairness. In *2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23. ACM, June 2023. doi: 10.1145/3593013.3594008. URL http://dx.doi.org/10.1145/3593013.3594008.

[52] Wen Zhang, Lingfei Deng, Lei Zhang, and Dongrui Wu. A survey on negative transfer. *IEEE/CAA Journal of Automatica Sinica*, 10(2):305–329, February 2023. ISSN 2329-9274. doi: 10.1109/jas.2022.106004. URL http://dx.doi.org/10.1109/JAS.2022.106004.

[53] Liang Ge, Jing Gao, Hung Ngo, Kang Li, and Aidong Zhang. On handling negative transfer and imbalanced distributions in multiple source transfer learning. *Statistical Analysis and Data Mining*, 7, 08 2014. doi: 10.1002/sam.11217.

[54] Ludovic Minvielle, Mounir Atiq, Sergio Peignier, and Mathilde Mougeot. Transfer learning on decision tree with class imbalance. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 1003–1010, 2019. doi: 10.1109/ICTAI.2019.00141.

[55] Lorenzo Belussi, Benedetta Barozzi, Alice Bellazzi, Ludovico Danza, Anna Devitofrancesco, Carlo Fanciulli, Matteo Ghellere, Giulia Guazzi, Italo Meroni, Francesco Salamone, et al. A review of performance of zero energy buildings and energy efficiency solutions. *Journal of building engineering*, 25:100772, 2019.

[56] Jawher Dridi, Manar Amayri, and Nizar Bouguila. Unsupervised adversarial domain adaptation for estimating occupancy and recognizing activities in smart buildings. In *Proceedings of the 2024 9th International Conference on Intelligent Information Technology*, pages 114–123, 2024.

[57] Carlos A Santos Silva, Manar Amayri, and Kaustav Basu. Characterization of energy demand and energy services using model-based and data-driven approaches. In *Towards Energy Smart Homes: Algorithms, Technologies, and Applications*, pages 229–248. Springer, 2021.

[58] Jawher Dridi, Manar Amayri, and Nizar Bouguila. Unsupervised domain adaptation with and without access to source data for estimating occupancy and recognizing activities in smart buildings. *Building and Environment*, 243:110651, 2023. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2023.110651. URL https://www.sciencedirect.com/science/article/pii/S0360132323006789.

[59] Jawher Dridi, Manar Amayri, and Nizar Bouguila. Unsupervised clustering-based domain adaptation for estimating occupancy and recognizing activities in smart buildings. *Journal of Building Engineering*, 85:108741, 2024. ISSN 2352-7102. doi: https://doi.org/10.1016/j.jobe.2024.108741. URL https://www.sciencedirect.com/science/article/pii/S2352710224003097.

[60] Soroush Samareh Abolhassani, Azar Zandifar, Negar Ghourchian, Manar Amayri, Nizar Bouguila, and Ursula Eicker. Improving residential building energy simulations through occupancy data derived from commercial off-the-shelf wi-fi sensing technology. *Energy and Buildings*, 272:112354, 2022. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2022.112354. URL https://www.sciencedirect.com/science/article/pii/S0378778822005254.

[61] Chujie Lu. Enhancing real-time nonintrusive occupancy estimation in buildings via knowledge fusion network. *Energy and Buildings*, 303:113812, 2024. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2023.113812. URL https://www.sciencedirect.com/science/article/pii/S0378778823010423.

[62] Chaoyang Jiang, Mustafa K. Masood, Yeng Chai Soh, and Hua Li. Indoor occupancy estimation from carbon dioxide concentration. *Energy and Buildings*, 131:132–141, 2016. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2016.09.002. URL https://www.sciencedirect.com/science/article/pii/S0378778816308027.

[63] Samr Ali and Nizar Bouguila. Towards scalable deployment of hidden markov models in occupancy estimation: A novel methodology applied to the study case of occupancy detection. *Energy and Buildings*, 254:111594, 2022.

[64] Usman Habib and Gerhard Zucker. Automatic occupancy prediction using unsupervised learning in buildings data. In *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, pages 1471–1476, 2017. doi: 10.1109/ISIE.2017.8001463.

[65] Jiaxun Guo, Manar Amayri, Nizar Bouguila, and Wentao Fan. A hybrid of interactive learning and predictive modeling for occupancy estimation in smart buildings. *IEEE Transactions on Consumer Electronics*, 67(4):285–293, 2021. doi: 10.1109/TCE.2021.3131943.

[66] Irvan B Arief-Ang, Flora D Salim, and Margaret Hamilton. Da-hoc: semi-supervised domain adaptation for room occupancy prediction using co2 sensor data. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, pages 1–10, 2017.

[67] Md Fazlay Rabbi Masum Billah, Nurani Saoda, Jiechao Gao, and Bradford Campbell. Ble can see: A reinforcement learning approach for rf-based indoor occupancy detection. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)*, IPSN '21, page 132–147, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380980. doi: 10.1145/3412382.3458262. URL https://doi.org/10.1145/3412382.3458262.

[68] Brodie W. Hobson, Daniel Lowcay, H. Burak Gunay, Araz Ashouri, and Guy R. Newsham. Opportunistic occupancy-count estimation using sensor fusion: A case study. *Building and Environment*, 159:106154, 2019. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2019.05.032. URL https://www.sciencedirect.com/science/article/pii/S0360132319303506.

[69] Cong Feng, Ali Mehmani, and Jie Zhang. Deep learning-based real-time building occupancy detection using ami data. *IEEE Transactions on Smart Grid*, 11(5):4490–4501, 2020. doi: 10.1109/TSG.2020.2982351.

[70] Qizhen Zhou, Jianchun Xing, Qiliang Yang, Xu Wang, Wenjie Chen, Yixin Mo, and Bowei Feng. Enabling non-intrusive occupant activity modeling using wifi signals and a generative adversarial network. *Energy and Buildings*, 249:111228, 2021. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2021.111228. URL https://www.sciencedirect.com/science/article/pii/S0378778821005120.

[71] Charmi Jobanputra, Jatna Bavishi, and Nishant Doshi. Human activity recognition: A survey. *Procedia Computer Science*, 155:698–703, 2019.

[72] Antonio Bevilacqua, Kyle MacDonald, Aamina Rangarej, Venessa Widjaya, Brian Caulfield, and Tahar Kechadi. Human activity recognition with convolutional neural networks. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part III 18*, pages 541–552. Springer, 2019.

[73] Djamila Romaissa Beddiar, Brahim Nini, Mohammad Sabokrou, and Abdenour Hadid. Vision-based human activity recognition: a survey. *Multimedia Tools and Applications*, 79 (41):30509–30555, 2020.

[74] Omar Mata, Juana Isabel Méndez, Pedro Ponce, Therese Peffer, Alan Meier, and Arturo Molina. Energy savings in buildings based on image depth sensors for human activity recognition. *Energies*, 16(3), 2023. ISSN 1996-1073. doi: 10.3390/en16031078. URL https://www.mdpi.com/1996-1073/16/3/1078.

[75] Unaiza Ahsan, Chen Sun, and Irfan A. Essa. Discrimnet: Semi-supervised action recognition from videos using generative adversarial networks. *CoRR*, abs/1801.07230, 2018. URL http://arxiv.org/abs/1801.07230.

[76] Amin Ullah, Khan Muhammad, Ijaz Ul Haq, and Sung Wook Baik. Action recognition using optimized deep autoencoder and cnn for surveillance data streams of non-stationary environments. *Future Generation Computer Systems*, 96:386–397, 2019. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2019.01.029. URL https://www.sciencedirect.com/science/article/pii/S0167739X18318533.

[77] Han Zou, Yuxun Zhou, Jianfei Yang, Hao Jiang, Lihua Xie, and Costas J. Spanos. Deepsense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6, 2018. doi: 10.1109/ICC.2018.8422895.

[78] Emmanuel Munguia Tapia, Stephen S Intille, and Kent Larson. Activity recognition in the home using simple and ubiquitous sensors. In *International conference on pervasive computing*, pages 158–175. Springer, 2004.

[79] Sakorn Mekruksavanich and Anuchit Jitpattanakul. Lstm networks using smartphone data for sensor-based human activity recognition in smart homes. *Sensors*, 21(5), 2021. ISSN 1424-8220. doi: 10.3390/s21051636. URL https://www.mdpi.com/1424-8220/21/5/1636.

[80] Abdul Rehman Javed, Raza Faheem, Muhammad Asim, Thar Baker, and Mirza Omer Beg. A smartphone sensors-based personalized human activity recognition system for sustainable smart cities. *Sustainable Cities and Society*, 71:102970, 2021. ISSN 2210-6707. doi: https://doi.org/10.1016/j.scs.2021.102970. URL https://www.sciencedirect.com/science/article/pii/S2210670721002560.

[81] Ronald Mutegeki and Dong Seog Han. A cnn-lstm approach to human activity recognition. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 362–366, 2020. doi: 10.1109/ICAIIC48513.2020.9065078.

[82] Md Zia Uddin and Ahmet Soylu. Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning. *Scientific Reports*, 11(1):16455, 2021.

[83] Xiangli Yang, Zixing Song, Irwin King, and Zenglin Xu. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(9):8934–8954, 2023. doi: 10.1109/TKDE.2022.3220219.

[84] Enrico Fini, Pietro Astolfi, Karteek Alahari, Xavier Alameda-Pineda, Julien Mairal, Moin Nabi, and Elisa Ricci. Semi-supervised learning made simple with self-supervised clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3187–3197, June 2023.

[85] Baixu Chen, Junguang Jiang, Ximei Wang, Pengfei Wan, Jianmin Wang, and Mingsheng Long. Debiased self-training for semi-supervised learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 32424–32437. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/d10d6b28d74c4f0fcab588feeb6fe7d6-Paper-Conference.pdf.

[86] Zhiqiang Shen, Peng Cao, Hua Yang, Xiaoli Liu, Jinzhu Yang, and Osmar R. Zaiane. Co-training with high-confidence pseudo labels for semi-supervised medical image segmentation, 2023. URL https://arxiv.org/abs/2301.04465.

[87] Xu Zheng, Chong Fu, Haoyu Xie, Jialei Chen, Xingwei Wang, and Chiu-Wing Sham. Uncertainty-aware deep co-training for semi-supervised medical image segmentation. *Computers in Biology and Medicine*, 149:106051, 2022.

[88] Zhen Liu, Qianli Ma, Peitian Ma, and Linghao Wang. Temporal-frequency co-training for time series semi-supervised learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 8923–8931, 2023.

[89] Fang He, Rong Wang, and Weimin Jia. Fast semi-supervised learning with anchor graph for large hyperspectral images. *Pattern Recognition Letters*, 130:319–326, 2020. ISSN

0167-8655. doi: https://doi.org/10.1016/j.patrec.2018.08.008. URL https://www.sciencedirect.com/science/article/pii/S0167865518304021. Image/Video Understanding and Analysis (IUVA).

[90] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3:1–40, 2016.

[91] Novi Patricia and Barbara Caputo. Learning to learn, from transfer learning to domain adaptation: A unifying perspective. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[92] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017.

[93] Sinno Jialin Pan, Ivor W. Tsang, James T. Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2011. doi: 10.1109/TNN.2010.2091281.

[94] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1180–1189, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/ganin15.html.

[95] Stamos Katsigiannis, Saleh Seyedzadeh, Andrew Agapiou, and Naeem Ramzan. Deep learning for crack detection on masonry façades using limited data and transfer learning. *Journal of Building Engineering*, 76:107105, 2023. ISSN 2352-7102. doi: https://doi.org/10.1016/j.jobe.2023.107105. URL https://www.sciencedirect.com/science/article/pii/S2352710223012846.

[96] Avishek Saha, Piyush Rai, Hal Daumé, Suresh Venkatasubramanian, and Scott L DuVall. Active supervised domain adaptation. In *Machine Learning and Knowledge Discovery in*

Databases: European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III 22, pages 97–112. Springer, 2011.

[97] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009. doi: 10.1109/TKDE.2008.239.

[98] Jiaxun Guo, Manar Amayri, Fatma Najar, Wentao Fan, and Nizar Bouguila. Occupancy estimation in smart buildings using predictive modeling in imbalanced domains. *Journal of Ambient Intelligence and Humanized Computing*, 14(8):10917–10929, 2023.

[99] Wei-Chao Lin, Chih-Fong Tsai, Ya-Han Hu, and Jing-Shang Jhang. Clustering-based undersampling in class-imbalanced data. *Information Sciences*, 409-410:17–26, 2017. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2017.05.008. URL https://www.sciencedirect.com/science/article/pii/S0020025517307235.

[100] Hongyu Guo and Herna L. Viktor. Learning from imbalanced data sets with boosting and data generation: the databoost-im approach. *SIGKDD Explor. Newsl.*, 6(1):30–39, jun 2004. ISSN 1931-0145. doi: 10.1145/1007730.1007736. URL https://doi.org/10.1145/1007730.1007736.

[101] Wenyang Wang and Dongchu Sun. The improved adaboost algorithms for imbalanced data classification. *Information Sciences*, 563:358–374, 2021.

[102] Ashish Anand, Ganesan Pugalenthi, Gary B Fogel, and PN Suganthan. An approach for classification of highly imbalanced data using weighting and undersampling. *Amino acids*, 39:1385–1391, 2010.

[103] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion*, 99:101805, 2023. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2023.101805. URL https://www.sciencedirect.com/science/article/pii/S1566253523001148.

[104] Yixin Li, Xiaodong Li, Dingyuan Ma, and Wei Gong. Exploring green building certification credit selection: A model based on explainable machine learning. *Journal of Building Engineering*, 95:110279, 2024. ISSN 2352-7102. doi: https://doi.org/10.1016/j.jobe.2024.110279. URL https://www.sciencedirect.com/science/article/pii/S2352710224018473.

[105] Andrea Zunino, Sarah Adel Bargal, Riccardo Volpi, Mehrnoosh Sameki, Jianming Zhang, Stan Sclaroff, Vittorio Murino, and Kate Saenko. Explainable deep classification models for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3233–3242, 2021.

[106] Vidhya Kamakshi and Narayanan C Krishnan. Explainable supervised domain adaptation. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2022. doi: 10.1109/IJCNN55064.2022.9892273.

[107] Christoph Raab, Manuel Röder, and Frank-Michael Schleif. Domain adversarial tangent subspace alignment for explainable domain adaptation. *Neurocomputing*, 506:418–429, 2022. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2022.07.074. URL https://www.sciencedirect.com/science/article/pii/S0925231222009377.

[108] Szymon Bobek, Sławomir Nowaczyk, Sepideh Pashami, Zahra Taghiyarrenani, and Grzegorz J Nalepa. Towards explainable deep domain adaptation. In *European Conference on Artificial Intelligence*, pages 101–113. Springer, 2023.

[109] Yan Zhang, Bak Koon Teoh, Maozhi Wu, Jiayu Chen, and Limao Zhang. Data-driven estimation of building energy consumption and ghg emissions using explainable artificial intelligence. *Energy*, 262:125468, 2023. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2022.125468. URL https://www.sciencedirect.com/science/article/pii/S0360544222023507.

[110] Thamsanqa Tsoka, Xianming Ye, YangQuan Chen, Dunwei Gong, and Xiaohua Xia. Explainable artificial intelligence for building energy performance certificate labelling classification. *Journal of Cleaner Production*, 355:131626, 2022. ISSN 0959-6526. doi:

https://doi.org/10.1016/j.jclepro.2022.131626. URL https://www.sciencedirect.com/science/article/pii/S0959652622012422.

[111] Naailah Mahamoodally, Jawher Dridi, and Manar Amayri. Explainable domain adaptation without source data for activity recognition. In *Proceedings of the Building Simulation 2025 Conference*, page To be published, Brisbane, Australia, Aug 2025. Accepted, to appear.

[112] Richard Cantin, Andrea Kindinis, and Pierre Michel. New approaches for overcoming the complexity of future buildings impacted by new energy constraints. *Futures*, 44(8):735–745, 2012. ISSN 0016-3287. doi: https://doi.org/10.1016/j.futures.2012.05.001. URL https://www.sciencedirect.com/science/article/pii/S0016328712001139.

[113] Saman Taheri, Paniz Hosseini, and Ali Razban. Model predictive control of heating, ventilation, and air conditioning (hvac) systems: A state-of-the-art review. *Journal of Building Engineering*, 60:105067, 2022. ISSN 2352-7102. doi: https://doi.org/10.1016/j.jobe.2022.105067. URL https://www.sciencedirect.com/science/article/pii/S2352710222010750.

[114] Abed Al Waheed Hawila and Abdelatif Merabtine. A statistical-based optimization method to integrate thermal comfort in the design of low energy consumption building. *Journal of Building Engineering*, 33:101661, 2021. ISSN 2352-7102. doi: https://doi.org/10.1016/j.jobe.2020.101661. URL https://www.sciencedirect.com/science/article/pii/S2352710220312420.

[115] Yang Geng, Wenjie Ji, Zhe Wang, Borong Lin, and Yingxin Zhu. A review of operating performance in green buildings: Energy use, indoor environmental quality and occupant satisfaction. *Energy and Buildings*, 183:500–514, 2019. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2018.11.017. URL https://www.sciencedirect.com/science/article/pii/S0378778818315378.

[116] United Nations Environment Programme. Buildings and climate change: Summary for decision makers, 2009. URL https://wedocs.unep.org/20.500.11822/32152.

[117] Alessandro Franco and Francesco Leccese. Measurement of co2 concentration for occupancy estimation in educational buildings with energy efficiency purposes. *Journal of Building Engineering*, 32:101714, 2020. ISSN 2352-7102. doi: https://doi.org/10.1016/j.jobe.2020.101714. URL https://www.sciencedirect.com/science/article/pii/S2352710219316912.

[118] Syed Faisal Abbas Shah, Muhammad Iqbal, Zeeshan Aziz, Toqir A. Rana, Adnan Khalid, Yu-N Cheah, and Muhammad Arif. The role of machine learning and the internet of things in smart buildings for energy efficiency. *Applied Sciences*, 12(15), 2022. ISSN 2076-3417. doi: 10.3390/app12157882. URL https://www.mdpi.com/2076-3417/12/15/7882.

[119] Manar Amayri and Stéphane Ploix. Decision tree and parametrized classifier for estimating occupancy in energy management. In *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 397–402. IEEE, 2018.

[120] Haneul Choi, Chai Yoon Um, Kyungmo Kang, Hyungkeun Kim, and Taeyeon Kim. Application of vision-based occupancy counting method using deep learning and performance analysis. *Energy and Buildings*, 252:111389, 2021. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2021.111389. URL https://www.sciencedirect.com/science/article/pii/S0378778821006733.

[121] Manar Amayri, Abhay Arora, Stéphane Ploix, Sanghamitra Bandhyopadyay, Dung Ngo, and Venkata Badarla. Estimating occupancy in heterogeneous sensor environment. *Energy and Buildings*, 129, 07 2016. doi: 10.1016/j.enbuild.2016.07.026.

[122] Ievgen Redko, Emilie Morvant, Amaury Habrard, Marc Sebban, and Younès Bennani. A survey on domain adaptation theory: learning bounds and theoretical guarantees, 2022.

[123] Xinhao Li, Jingjing Li, Lei Zhu, Guoqing Wang, and Zi Huang. Imbalanced source-free domain adaptation. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, page 3330–3339, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450386517. doi: 10.1145/3474085.3475487. URL https://doi.org/10.1145/3474085.3475487.

[124] Seonghyeon Kim, Seokwoo Kang, Kwang Ryel Ryu, and Giltae Song. Real-time occupancy prediction in a large exhibition hall using deep learning approach. *Energy and Buildings*, 199:216–222, 2019. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2019.06.043. URL https://www.sciencedirect.com/science/article/pii/S0378778819304815.

[125] Xiyang Hu, Cynthia Rudin, and Margo Seltzer. Optimal sparse decision trees, 2023.

[126] Jimmy Lin, Chudi Zhong, Diane Hu, Cynthia Rudin, and Margo Seltzer. Generalized and scalable optimal sparse decision trees, 2022.

[127] Norberto Goussies, Sebastián Ubalde, and Marta Mejail. Transfer learning decision forests for gesture recognition. *Journal of Machine Learning Research*, 15:3667–3690, 11 2014.

[128] Zhenghua Chen, Chaoyang Jiang, and Lihua Xie. Building occupancy estimation and detection: A review. *Energy and Buildings*, 169:260–270, 2018.

[129] Vishal Garg and N.K. Bansal. Smart occupancy sensors to reduce energy consumption. *Energy and Buildings*, 32(1):81–87, 2000. ISSN 0378-7788. doi: https://doi.org/10.1016/S0378-7788(99)00040-7. URL https://www.sciencedirect.com/science/article/pii/S0378778899000407.

[130] Jerson A Pinzon, Pedro P Vergara, Luiz CP Da Silva, and Marcos J Rider. Optimal management of energy consumption and comfort for smart buildings operating in a microgrid. *IEEE transactions on smart grid*, 10(3):3236–3247, 2018.

[131] Mahmoud Kassas. Modeling and simulation of residential hvac systems energy consumption. *Procedia computer science*, 52:754–763, 2015.

[132] Xiu'e Yang, Shuli Liu, Yuliang Zou, Wenjie Ji, Qunli Zhang, Abdullahi Ahmed, Xiaojing Han, Yongliang Shen, and Shaoliang Zhang. Energy-saving potential prediction models for large-scale building: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 156:111992, 2022.

[133] Nhung Tran Thi Hong, Giang L Nguyen, Nguyen Quang Huy, Do Viet Manh, Duc-Nghia Tran, and Duc-Tan Tran. A low-cost real-time iot human activity recognition system based on wearable sensor and the supervised learning algorithms. *Measurement*, 218:113231, 2023.

[134] Wentao Zhu, Cuiling Lan, Junliang Xing, Wenjun Zeng, Yanghao Li, Li Shen, and Xiaohui Xie. Co-occurrence feature learning for skeleton based action recognition using regularized deep lstm networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

[135] Samundra Deep and Xi Zheng. Leveraging cnn and transfer learning for vision-based human activity recognition. In *2019 29th international telecommunication networks and applications conference (ITNAC)*, pages 1–4. IEEE, 2019.

[136] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International conference on machine learning*, pages 6028–6039. PMLR, 2020.

[137] Diane J Cook. Learning setting-generalized activity models for smart spaces. *IEEE intelligent systems*, 2010(99):1, 2010.

[138] Clyde Zhengdao Li, Limei Zhang, Xin Liang, Bing Xiao, Vivian W.Y. Tam, Xulu Lai, and Zhe Chen. Advances in the research of building energy saving. *Energy and Buildings*, 254:111556, 2022. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2021.111556. URL https://www.sciencedirect.com/science/article/pii/S0378778821008409.

[139] Xiaodong Cao, Xilei Dai, and Junjie Liu. Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. *Energy and buildings*, 128:198–213, 2016.

[140] T Hong and HW Lin. Occupant behavior: Impact on energy use of private offices, asim 2012-1st asia conf. *Int. Build. Perform. Simul. Assoc*, 12, 2013.

[141] Liu Yang, Haiyan Yan, and Joseph C Lam. Thermal comfort and building energy consumption implications–a review. *Applied energy*, 115:164–173, 2014.

[142] Ke Zhang, Jun Zhang, Pei-Dong Xu, Tianlu Gao, and David Wenzhong Gao. Explainable ai in deep reinforcement learning models for power system emergency control. *IEEE Transactions on Computational Social Systems*, 9(2):419–427, 2022. doi: 10.1109/TCSS.2021.3096824.

[143] Yan Ding, Shuxue Han, Zhe Tian, Jian Yao, Wanyue Chen, and Qiang Zhang. Review on occupancy detection and prediction in building simulation. In *Building Simulation*, pages 1–24. Springer, 2022.

[144] Frauke Oldewurtel, David Sturzenegger, and Manfred Morari. Importance of occupancy information for building climate control. *Applied energy*, 101:521–532, 2013.

[145] Jesper E Van Engelen and Holger H Hoos. A survey on semi-supervised learning. *Machine learning*, 109(2):373–440, 2020.

[146] Zehui Zhao, Laith Alzubaidi, Jinglan Zhang, Ye Duan, and Yuantong Gu. A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. *Expert Systems with Applications*, 242:122807, 2024.

[147] Barret Zoph, Golnaz Ghiasi, Tsung-Yi Lin, Yin Cui, Hanxiao Liu, Ekin Dogus Cubuk, and Quoc Le. Rethinking pre-training and self-training. *Advances in neural information processing systems*, 33:3833–3845, 2020.

[148] Kamal Nigam and Rayid Ghani. Analyzing the effectiveness and applicability of co-training. In *Proceedings of the Ninth International Conference on Information and Knowledge Management*, CIKM '00, page 86–93, New York, NY, USA, 2000. Association for Computing Machinery. ISBN 1581133200. doi: 10.1145/354756.354805. URL https://doi.org/10.1145/354756.354805.

[149] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*,

COLT' 98, page 92–100, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 1581130570. doi: 10.1145/279943.279962. URL https://doi.org/10.1145/279943.279962.

[150] Jizong Peng, Guillermo Estrada, Marco Pedersoli, and Christian Desrosiers. Deep co-training for semi-supervised image segmentation. *Pattern Recognition*, 107:107269, 2020.

[151] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*, pages 912–919, 2003.

[152] Zixing Song, Xiangli Yang, Zenglin Xu, and Irwin King. Graph-based semi-supervised learning: A comprehensive review. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):8174–8194, 2022.

[153] Partha Pratim Talukdar and Koby Crammer. New regularized algorithms for transductive learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 442–457. Springer, 2009.

[154] Md Washim Akram, Muhammad Firdaus Mohd Zublie, Md Hasanuzzaman, and Nasrudin Abd Rahim. Global prospects, advance technologies and policies of energy-saving and sustainable building systems: A review. *Sustainability*, 14(3):1316, 2022.

[155] Razak Olu-Ajayi, Hafiz Alaka, Ismail Sulaimon, Funlade Sunmola, and Saheed Ajayi. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *Journal of Building Engineering*, 45:103406, 2022.

[156] Veena Chidurala and Xinrong Li. Occupancy estimation using thermal imaging sensors and machine learning algorithms. *IEEE Sensors Journal*, 21(6):8627–8638, 2021.

[157] Manar Amayri, Samer Ali, Nizar Bouguila, and Stephane Ploix. Machine learning for activity recognition in smart buildings: A survey. *Towards Energy Smart Homes: Algorithms, Technologies, and Applications*, pages 199–228, 2021.

[158] Mohammad Esrafilian-Najafabadi and Fariborz Haghighat. Occupancy-based hvac control using deep learning algorithms for estimating online preconditioning time in residential buildings. *Energy and Buildings*, 252:111377, 2021.

[159] Mattia Litrico, Alessio Del Bue, and Pietro Morerio. Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7640–7650, 2023.

[160] Jinyu Yang, Jingjing Liu, Ning Xu, and Junzhou Huang. Tvt: Transferable vision transformer for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 520–530, 2023.

[161] Kun Zhang, Bernhard Schölkopf, Krikamol Muandet, and Zhikun Wang. Domain adaptation under target and conditional shift. In *International conference on machine learning*, pages 819–827. Pmlr, 2013.

[162] Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2021.

[163] Yifei Ding, Minping Jia, Jichao Zhuang, Yudong Cao, Xiaoli Zhao, and Chi-Guhn Lee. Deep imbalanced domain adaptation for transfer learning fault diagnosis of bearings under multiple working conditions. *Reliability Engineering & System Safety*, 230:108890, 2023.

[164] Siyu Zhang, SU Lei, GU Jiefei, LI Ke, ZHOU Lang, and Michael Pecht. Rotating machinery fault detection and diagnosis based on deep domain adaptation: A survey. *Chinese Journal of Aeronautics*, 36(1):45–74, 2023.

[165] Xiaofeng Liu, Chaehwa Yoo, Fangxu Xing, Hyejin Oh, Georges El Fakhri, Je-Won Kang, Jonghye Woo, et al. Deep unsupervised domain adaptation: A review of recent advances and perspectives. *APSIPA Transactions on Signal and Information Processing*, 11(1), 2022.

[166] Zhipeng He, Yongshi Zhong, and Jiahui Pan. An adversarial discriminative temporal convolutional network for eeg-based cross-domain emotion recognition. *Computers in biology and medicine*, 141:105048, 2022.

[167] Wei Wang, Haojie Li, Zhengming Ding, Feiping Nie, Junyang Chen, Xiao Dong, and Zhihui Wang. Rethinking maximum mean discrepancy for visual domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 34(1):264–277, 2021.

[168] Yuqi Fang, Pew-Thian Yap, Weili Lin, Hongtu Zhu, and Mingxia Liu. Source-free unsupervised domain adaptation: A survey. *Neural Networks*, 174:106230, 2024.

[169] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Generalized source-free domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8978–8987, October 2021.

[170] Uiwon Hwang, Jonghyun Lee, Juhyeon Shin, and Sungroh Yoon. Sf $(da)^2$: Source-free domain adaptation through the lens of data augmentation. *arXiv preprint arXiv:2403.10834*, 2024.

[171] Korawat Tanwisuth, Xinjie Fan, Huangjie Zheng, Shujian Zhang, Hao Zhang, Bo Chen, and Mingyuan Zhou. A prototype-oriented framework for unsupervised domain adaptation. *Advances in Neural Information Processing Systems*, 34:17194–17208, 2021.

[172] Alessandro Aliberti, Lorenzo Bottaccioli, Enrico Macii, Santa Di Cataldo, Andrea Acquaviva, and Edoardo Patti. A non-linear autoregressive model for indoor air-temperature predictions in smart buildings. *Electronics*, 8(9):979, 2019.

[173] Zhijie Deng, Yucen Luo, and Jun Zhu. Cluster alignment with a teacher for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9944–9953, 2019.