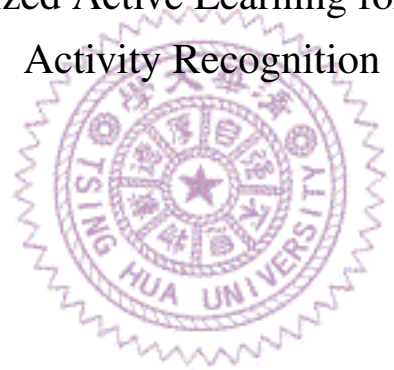
國立清華大學電機資訊學院資訊工程研究所
碩士論文
Department of Computer Science
College of Electrical Engineering and Computer Science
National Tsing Hua University
Master Thesis

將個人化主動學習應用於毫米波雷達的行為辨識
PALM: Personalized Active Learning for mmWave-Based Activity Recognition



109062585
江忻哲
Hsin-Che Chiang

指導教授：徐正炘 博士
Advisor: Cheng-Hsin Hsu, Ph.D.

中華民國 113 年 8 月
August, 2024

中文摘要

人類行為辨識（HAR）在增進人類安全和福祉方面發揮著至關重要的作用，其應用範圍包含飲食管理、駕駛員監控等領域。毫米波（mmWave）雷達由於能夠捕捉細緻的動作，且無穿戴式感測器帶來的不便及視覺感測器可能引發的隱私問題，已成為 HAR 的一種具有前景的技術。本論文提出將個人化主動學習應用於毫米波雷達的行為辨識（PALM: Personalized Active Learning for mmWave）。基於我們先前提出的動態點雲辨識器（DPR: Dynamic Point Cloud Recognizer），PALM 利用不確定性（Uncertainty）選擇性地向使用者詢問最具資訊量的樣本，以訓練個人化模型，解決新用戶的冷啓動問題（Cold Start Problem）。在我們蒐集的進食行為資料集上的實驗表明，PALM 於兩週的主動學習期間達到了 91.08% 的準確率，超越了基線和其他方法。此外，利用我們的駕駛行為資料集進行遷移學習（Transfer Learning），PALM 的準確率比從零開始訓練的基線模型高出 9.87%，曲線下面積（AUC）也提高了 19.48%。這些結果突顯了 PALM 訓練個人化 HAR 模型的有效性，同時將標記工作最小化，使其適合於真實世界應用的廣泛部署。此外，我們顯示 DPR 比最先進的體素化（Voxelization）方法更優，準確率提高了 4.10%，同時減少了 78.29% 的記憶體消耗和 69.64% 的推論時間，實現了資源的有效利用。

Abstract

Human Activity Recognition (HAR) plays a crucial role in enhancing human safety and well-being, with applications ranging from dietary management to driver monitoring. Millimeter-wave (mmWave) radars have emerged as a promising technology for HAR due to their ability to capture fine-grained activities without the inconvenience associated with wearable sensors or the potential privacy issues posed by vision-based sensors. In this thesis, we introduce PALM: Personalized Active Learning for mmWave. Built upon our previously proposed Dynamic Point Cloud Recognizer (DPR), PALM addresses the challenge of cold start for new users by utilizing uncertainty to selectively query the user about the most informative samples while training a personalized model. Experiments on our Food Intake Activity Dataset demonstrate that PALM attains 91.08% accuracy over a two-week active learning period, surpassing the baseline and alternative methods. Furthermore, leveraging transfer learning from our Driver Activity Dataset, PALM achieves a 9.87% higher accuracy and 19.48% improvement in Area Under the Curve (AUC) compared to the baseline model trained from scratch. These results highlight PALM's effectiveness in personalizing HAR models while minimizing labeling effort, making it suitable for widespread deployment in real-world applications. In addition, we show that DPR outperforms state-of-the-art voxelization-based methods, achieving a 4.10% increase in accuracy while reducing memory consumption by 78.29% and inference time by 69.64%, leading to resource efficiency.

致謝

感謝 Cheng-Hsin Hsu 教授和 Shervin Shirmohammadi 教授在本研究中提供的寶貴指導，以及 Yi-Hung Wu 和 Guan-Hua Li 的協助與合作。

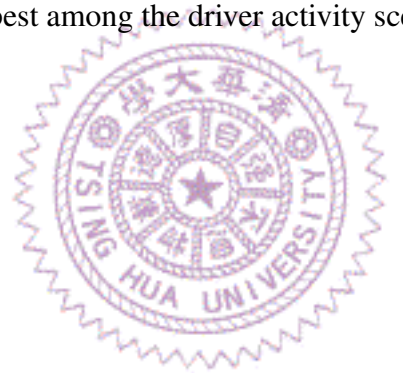
Acknowledgments

I would like to express my sincere gratitude to Prof. Cheng-Hsin Hsu and Shervin Shirmohammadi for their guidance throughout this research, as well as Yi-Hung Wu and Guan-Hua Li for their assistance and collaboration.

Contents

TSING HUA UNIVERSITY

List of Figures



List of Tables

Chapter 1

Introduction

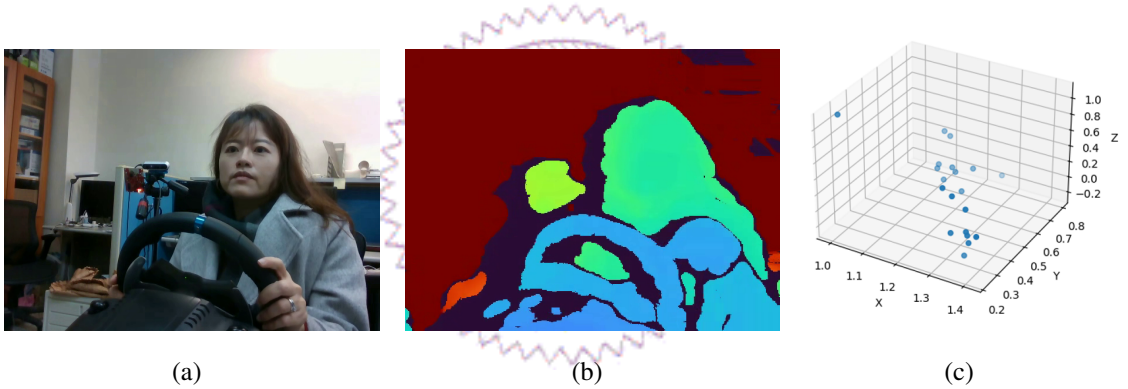
This chapter provides an overview of the research by outlining its scope and limitations, highlighting the key contributions, and presenting the organization of the thesis.

1.1 Scope and Limitation

Human Activity Recognition (HAR) is crucial for enhancing human safety and well-being. By continuously monitoring human activities and behaviors, HAR systems can provide valuable insights and enable a wide range of applications that can significantly improve the quality of life. They can detect and monitor various daily activities, such as walking, food intake, falls, or other emergencies, which can be used to track people's health and mobility, enabling timely interventions, better care management, and potentially life-saving assistance. For instance, chronic diseases have become a leading cause of death worldwide, claiming 41 million lives each year and accounting for 74% of all deaths [1], highlighting the crucial role of dietary management, especially for vulnerable groups such as the elderly, obese, and individuals with dementia. However, traditional methods of dietary tracking rely heavily on subjective self-reporting, leading to inaccurate and unreliable data due to incomplete or inconsistent recording of food portions, ingredients, and meal timings [2]. Another example is Driver Monitoring Systems (DMS), which play a vital role in enhancing road safety by continuously monitoring driver behavior, alertness, and attention, helping to prevent accidents caused by distraction, fatigue, or impairment. Furthermore, driver monitoring can enable advanced driver assistance systems to adapt and respond appropriately, such as initiating automated braking or steering corrections when necessary.

For HAR systems to be widely applicable in real life, they should ideally be: (i) non-intrusive, requiring no attachment of additional devices like smartwatches or throat microphones to minimize discomfort, and (ii) privacy-preserving, eliminating the need for

RGB cameras and other vision-based sensors, which can invade user privacy due to their ability to identify individuals and other sensitive attributes. Based on these requirements, compared to other alternatives, we select in-situ consumer-grade 3D mmWave radars due to several advantages: (i) non-intrusiveness — not requiring any device wear, minimizing user burden; (ii) privacy preservation — capturing body movements as sparse dynamic 3D point clouds, containing only a few dozen points per frame; (iii) robustness — unaffected by lighting conditions, ensuring consistent performance in different real-world scenarios; (iv) power-efficiency — miniaturized and consuming minimal power, allowing for continuous and unobtrusive monitoring. Fig. 1.1 compares different sensor data types, showing that mmWave point clouds preserve user privacy better than RGB-D data. Specifically, the sample data from the RGB and depth cameras provide more detailed and potentially identifiable information, while the mmWave radar data present a sparse point cloud that maintains functionality for activity recognition but significantly reduce privacy risks.



(a) (b) (c)

Figure 1.1: Sample sensor data from: (a) RGB camera, (b) depth camera, and (c) mmWave radar. Compared to RGB-D data, mmWave point clouds present a lower risk of privacy concerns.

Despite these advantages, mmWave-based HAR systems face challenges that hinder their practical deployment and performance. One issue lies in the inefficiency of the common preprocessing technique, *voxelization* [3, 4], in terms of GPU memory usage. The voxelization process converts point cloud data into voxels, stored as a 3D array. Since mmWave radars generate sparse point clouds, often comprising only a few dozen points per frame, most voxels remain empty after voxelization. Furthermore, when point clouds are preprocessed as 3D structures, it becomes more computationally demanding for neural networks to capture spatial information effectively, presenting a challenge for deploying such systems on devices that lack the necessary computational resources.

An arguably more serious problem arises when attempting to recognize activities performed by a new and previously unseen user, as the lack of personalized training data capturing their unique motion patterns and body characteristics can lead to reduced ac-

curacy. In our previous work [5], although the model achieved over 90% accuracy with a conventional 80/20 train-test split, the accuracy dropped to ∼70% when evaluated on new subjects, highlighting the impact of what we term the *cold start problem*. Without sufficient training data encompassing the variability in how different individuals perform the same activities, these systems may struggle to accurately classify the activities of previously unseen subjects, compromising the system's reliability in real-world scenarios.

1.2 Contributions

Our earlier works include the Food Intake Activity Dataset (FIAD) [6], the Driver Activity Dataset (DAD) [7], and the Dynamic Point Cloud Recognizer (DPR) [5], an end-to-end HAR model that operates directly on sparse dynamic 3D point clouds obtained from mmWave radars, reducing memory consumption and computational resources compared to voxelization-based methods [8]. In this thesis, we extend these works and make the following contributions:

- We introduce PALM: Personalized Active Learning for mmWave, an active learning framework built upon DPR. PALM efficiently trains a personalized model for new users by utilizing the quantified uncertainty of its activity label prediction to select the most informative samples from unlabeled activities for labeling.
- We conduct comprehensive experiments to determine optimal system parameters, including labeling budget, number of training epochs per new labeled data acquisition, and efficacy of freezing feature extraction layers.
- We evaluate PALM against state-of-the-art active learning methods, alternative entropy-based methods [9–15], as well as some diversity-based methods using the FIAD dataset, providing a comparative analysis of its performance.
- We investigate PALM's generalizability in a cross-application transfer learning scenario by applying a global model pre-trained with the DAD dataset to a new user in the FIAD dataset.

Through extensive experiments on the FIAD dataset, we demonstrate the effectiveness of our contributions as follows: (i) DPR outperforms previous state-of-the-art voxelization-based methods, achieving a 4.10% increase in classification accuracy while reducing memory consumption by 78.29% and inference time by 69.64%; (ii) PALM outperforms the baseline and four other active learning methods, attaining an accuracy of 91.08% over a two-week active learning period, with an upper bound of 98.25% observed over

an extended period; (iii) among the four entropy-based methods evaluated, Max Entropy achieved the highest accuracy of 92.03% and an AUC improvement of 90.85%, underscoring its effectiveness despite requiring more computational resource; (iv) none of the diversity-based methods outperformed the uncertainty-based ones, highlighting the efficacy of our uncertainty-based approaches in improving model accuracy and efficiency; (v) leveraging a model pre-trained with our Driver Activity Dataset [7], PALM exhibits a 9.87% higher accuracy and a 19.48% improvement in AUC compared to the baseline model trained from scratch, illustrating its ability to benefit from cross-application transfer learning when there is insufficient data for some activities.

1.3 Thesis Organization

The rest of the thesis is organized as follows: Chapter 2 explores related work in HAR and active learning. Chapter 3 overviews the key concepts and technologies underlying mmWave-based HAR systems. Chapter 4 states the problem, including the resource inefficiency in point cloud processing and the cold start problem of new users. Chapter 5 details the methodology of our proposed DPR model and PALM framework. Chapter 6 presents the experiments conducted to assess the performance of our proposed methods. Finally, Chapter 7 provides concluding remarks and discusses future directions.

HUA UNIVER

Chapter 2

Related Work

This chapter provides an overview of existing studies and approaches in two relevant areas: HAR and active learning.

2.1 Human Activity Recognition

We present a brief overview of HAR and refer the readers to surveys such as [16] for more details. The existing literature on HAR can be broadly classified into two categories based on the type of sensors employed: (i) wearable sensors, which move with the user, and (ii) in-situ sensors, which collect data from where they are installed.

Wearable sensors, such as Electromyography (EMG) [17–19] and Electroencephalography (EEG) [20], have been adopted for HAR, as they can be attached to subjects to detect muscle movements. Inertial sensors, like accelerators and gyroscopes [21–26], have also been adopted for HAR, as they are already installed in modern smartphones, watches, and wristbands. Despite decent accuracy [27, 28], attaching these sensors to the human body can be time-consuming, inconvenient, and cause discomfort to the subject, which limits their widespread adoption for day-to-day use.

In-situ sensors, mainly vision-based RGB-D cameras, have been employed to capture RGB-D images [29–31] and videos [32–34] for HAR. However, deploying these sensors in smart home environments could violate individuals' privacy. Researchers have also proposed using Radio Frequency (RF) transceivers [35] for HAR. However, these signals often propagate through noisy Industrial, Scientific, and Medical (ISM) bands. mmWave radars have been adopted for HAR as a more suitable alternative. For instance, Singh et al. [36] proposed voxelizing raw point clouds before inputting the voxels into neural networks, while Wang et al. [37] suggested filtering out environmental noise from the data before voxelization. A more recent work employed a Graph Neural Network (GNN) on dynamic edges generated from sparse point clouds [38]. In light of the advantages of

mmWave radars, we also adopt this sensing modality. However, we address the memory inefficiency of voxelization to enable more practical deployment of mmWave radar-based HAR systems in resource-constrained scenarios.

2.2 Active Learning

Active learning refers to a human-in-the-loop machine learning method in which the learning algorithm can interactively query the user to determine the correct label for specific data points. This method allows the learning algorithm to improve its accuracy over time as it learns the correct labels from the user. Existing active learning approaches can be broadly classified into three scenarios [39]: (i) membership query synthesis, which involves generating synthetic instances for labeling; (ii) pool-based active learning, which selects samples from a pool of unlabeled data; and (iii) stream-based active learning, which makes decisions about labeling samples as they arrive sequentially in a stream.

Membership query synthesis [40] is a seminal approach in the field of active learning. It enables the selection of any samples within the input space for labeling, including synthetically generated ones, such as synthetic images for classification or synthetic sentences for NLP. While this method provides complete freedom to query labels for any sample, it can generate nonsensical samples that human annotators cannot adequately label [41]. In this thesis, we focus on the following two alternative scenarios, utilizing only real mmWave point cloud data streams.

Pool-based active learning [42] involves selecting the most informative sample from a static pool of unlabeled data. This method is well-suited for tasks such as image classification [43] and text classification [44], where large volumes of unlabeled data can be gathered simultaneously. Several studies have attempted to apply pool-based active learning to HAR [45–47]. However, HAR involves capturing real sensor data of human activities, unlike image or text data, where large datasets can be collected simultaneously. Therefore, relying solely on pool-based scenarios is insufficient; we should also consider scenarios where data is acquired incrementally.

Stream-based (online) active learning [48, 49] involves processing data that arrives continuously, requiring the learner to make real-time decisions on whether to query each sample for labeling without knowledge of future instances. One of the earliest applications of this approach is spam filtering [50], where a filter is constantly updated by selecting the most informative emails for labeling. Stream-based active learning can be broadly divided into two classes: (i) batch-based (window-based) methods, where a buffer accumulates a batch of samples, and best-out-of-window sampling is used to select samples for querying, and (ii) single-pass methods, where each sample requires an immediate

decision on whether to query. Existing studies on applying stream-based active learning to HAR [51–53] focus on datasets with wearable sensors. In contrast, we compare both pool-based and stream-based scenarios using datasets featuring in-situ mmWave radars. To the best of our knowledge, we are among the first to employ stream-based active learning for mmWave radar-based HAR.

Chapter 3

Background

This chapter overviews the key concepts and technologies underlying mmWave-based HAR systems. We examine various sensor modalities employed in HAR and then specifically investigate mmWave radar technology. Subsequently, we explore deep learning, uncertainty quantification, and active learning, all of which play crucial roles in enhancing the performance and adaptability of mmWave-based HAR systems.

3.1 Human Activity Recognition

Human Activity Recognition (HAR) has emerged as a pivotal field in ubiquitous computing, with applications spanning healthcare, smart homes, and human-computer interaction. HAR's primary objective is automatically identifying and classifying human activities based on sensor data. Over the years, researchers have explored a diverse array of sensors to capture human activities, each with its own set of advantages and limitations.

Wearable sensors have been extensively utilized in HAR to capture personalized physiological and kinematic data. These sensors, such as accelerometers, gyroscopes, electromyography (EMG), and electroencephalography (EEG), offer high precision and individual-specific data. Accelerometers and gyroscopes collectively capture motion data, with accelerometers detecting linear acceleration and gravitational forces, while gyroscopes measure angular velocity and orientation. EMG sensors measure the electrical activity of skeletal muscles, providing insights into muscle engagement during various activities. On the other hand, EEG sensors record the brain's electrical activity, offering a window into cognitive states that may correlate with specific activities. Despite their capabilities, wearable sensors present significant usability challenges. Users often find them intrusive, uncomfortable, or burdensome to wear consistently. This can lead to non-compliance and data gaps in long-term studies, potentially compromising the reliability and continuity of activity recognition systems. The need to regularly charge or replace

batteries in wearable devices further adds to the user burden, making them less than ideal for continuous, long-term monitoring scenarios.

In response to these limitations, researchers have turned to in-situ sensors embedded within the environment. These sensors offer an alternative approach that mitigates some usability issues associated with wearable devices. Vision-based sensors, such as RGB-D cameras, have been widely explored due to their ability to capture color and depth information, enabling detailed analysis of human posture and movement. However, these sensors raise significant privacy concerns, continuously recording potentially sensitive visual data within private spaces. Cameras in personal environments like homes or healthcare facilities can be perceived as intrusive, potentially altering natural behavior patterns and raising ethical questions about data storage and access. Wi-Fi sensing represents another category of in-situ sensing technology for HAR. These systems leverage existing Wi-Fi infrastructure to detect human activities through signal perturbations. While non-intrusive and utilizing ubiquitous infrastructure, Wi-Fi sensing systems are susceptible to interference from other devices operating in the crowded ISM (Industrial, Scientific, and Medical) bands. This interference can compromise the reliability and accuracy of activity recognition, particularly in densely populated urban environments or areas with multiple overlapping Wi-Fi networks.

Given the limitations of wearable sensors, vision-based sensors, and Wi-Fi sensing, mmWave radar has emerged as a desirable option for HAR applications. mmWave radar offers a promising balance between non-intrusiveness and robust sensing capabilities. These systems can capture detailed motion information without recording identifiable visual data, addressing both privacy and usability concerns. The following section explores the fundamental principles and variants of mmWave radar systems, focusing on their application in HAR.

3.2 mmWave Radar

Millimeter-wave (mmWave) radar technology, operating in the 30-300 GHz frequency range, has gained significant traction in HAR due to its unique advantages. At its core, radar systems operate by emitting electromagnetic waves and analyzing the reflected signals to detect and localize objects. mmWave radar, in particular, leverages high-frequency signals to achieve fine spatial resolution and sensitivity to subtle movements, making it ideal for HAR applications. The short wavelengths of mmWave signals allow for compact antenna designs, facilitating the integration of radar systems into consumer devices and smart home infrastructure.

Radar systems can be broadly categorized into pulse-based radar and continuous-wave

radar. Pulse-based radar systems emit short, high-power pulses of electromagnetic energy and measure the time delay of the reflected signals. This approach offers excellent range resolution, allowing for precise distance measurements to multiple targets. However, pulse-based systems may have limitations in detecting slow-moving targets due to the pulsed nature of the transmitted signal. Continuous-wave (CW) radar, on the other hand, continuously transmits a signal, enabling the detection of Doppler shifts to measure target velocity. Basic CW radar excels at measuring velocity but cannot accurately determine range. To overcome this limitation, advanced variants like Frequency-Modulated Continuous-Wave (FMCW) radar have been developed, combining the benefits of continuous transmission with range measurement capabilities.

FMCW radar represents a sophisticated form of continuous-wave radar that addresses the range measurement limitations of basic CW systems. FMCW radar modulates the transmitted signal's frequency over time, typically using a linear chirp. By analyzing the frequency difference between transmitted and received signals, FMCW radar can simultaneously measure targets' range and radial velocity. This capability is valuable in HAR applications, where body parts' position and movement contribute to activity classification. The high resolution achievable with FMCW systems, often at the millimeter level, enables the detection of subtle human movements. This fine-grained motion capture is crucial for distinguishing between similar activities or detecting micro-gestures, which may be challenging with other sensing modalities. Moreover, FMCW radar's ability to measure velocity through Doppler shifts allows for capturing speed-related features of human motion, enriching the feature set available for activity classification algorithms.

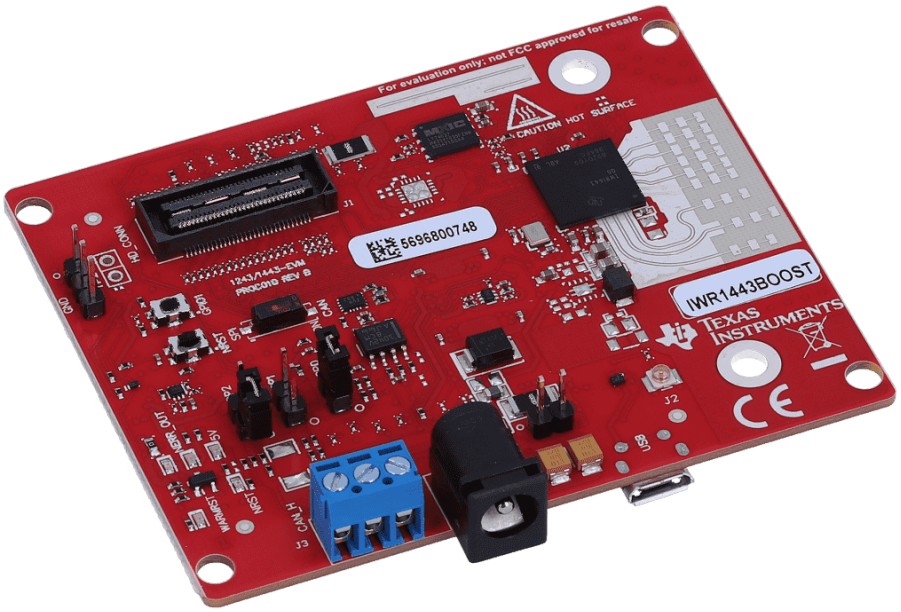


Figure 3.1: The Texas Instrument IWR1443BOOST module.

Given these advantages, we employ the IWR1443BOOST mmWave radar by Texas Instruments, which utilizes FMCW technology. Fig. 3.1 shows the IWR1443BOOST module [54]. The rich, multidimensional data provided by FMCW radar - including range, velocity, and angle information - serves as an excellent input for advanced machine learning algorithms, particularly deep learning models that can automatically extract relevant features from complex input spaces. The following section explores the specifics of deep learning techniques related to HAR, exploring how these advanced algorithms process and interpret the complex data generated by sensors such as mmWave radar.

3.3 Deep Learning

The field of Artificial Intelligence (AI) encompasses systems that can perform tasks requiring human-like intelligence, ranging from problem-solving to perception and language understanding. Within this broad domain, Machine Learning (ML) focuses on algorithms that can learn from and make predictions or decisions based on data without being explicitly programmed for each specific task. Deep Learning (DL), in turn, represents a specialized subset of ML that utilizes artificial neural networks with multiple layers to learn hierarchical representations of data.

At the heart of deep learning are neural networks, computational models inspired by the structure and function of biological neural networks. These networks are composed of interconnected nodes, or neurons, organized in layers. The input layer receives raw data features, such as the multidimensional data captured by mmWave radars. This input then passes through one or more hidden layers, where each neuron applies a weighted sum of its inputs followed by a non-linear activation function. These hidden layers progressively transform and abstract the input data, learning to represent increasingly complex features. Finally, the output layer produces the final prediction or classification, such as the identified human activity. The power of deep learning lies in its ability to automatically learn relevant features directly from raw data, reducing the need for manual feature engineering. This is particularly advantageous in HAR applications, where the optimal features for distinguishing between activities may not be immediately apparent or vary across different sensor modalities and activity types.

Several deep learning architectures have shown promise in HAR applications, each with strengths and characteristics. Convolutional Neural Networks (CNNs) have proven particularly effective in processing grid-like data, such as spectrograms or range-Doppler maps generated from radar signals. CNNs use convolutional layers to automatically learn spatial hierarchies of features, making them well-suited to capturing the spatial and temporal patterns inherent in human activities. Recurrent Neural Networks (RNNs) and their

advanced variants, such as Long Short-Term Memory (LSTM) networks, are designed to process sequential data, making them effective for capturing temporal dependencies in activity patterns. These architectures maintain an internal state that can remember information over time, allowing them to model the evolving nature of human activities. This temporal modeling is crucial in HAR, as many activities are defined not just by instantaneous postures or movements but by sequences of actions over time.

The application of deep learning to HAR offers several key advantages. First, learning hierarchical representations allows these models to capture both low-level motion primitives and high-level activity concepts within a single framework. This can lead to more robust and generalizable activity recognition systems. Second, the scalability of deep learning models means they can potentially improve as more data becomes available, making them well-suited to the ever-growing volumes of sensor data in ubiquitous computing environments. Furthermore, transfer learning techniques allow knowledge gained from one HAR task to be applied to related tasks, enabling more efficient learning in new domains or for new activities. This is particularly valuable in real-world deployments, where the ability to quickly adapt to new users, environments, or activity types is crucial.

However, the power of deep learning in HAR also comes with challenges. The "black box" nature of deep neural networks can make it difficult to interpret their decision-making processes, which may be problematic in applications where explainability is essential, such as healthcare monitoring. Additionally, the large number of parameters in deep models can lead to overfitting, especially when labeled training data is limited. This necessitates careful selection of the model size, hyperparameters, and regularization strategies such as dropout, which we will evaluate in Section 6.2.

3.4 Active Learning

Active learning is a human-in-the-loop machine learning paradigm in which the learner interactively queries the user to determine the correct label for specific data points. This approach is particularly relevant to HAR systems, where obtaining labeled data can be time-consuming, expensive, and sometimes intrusive. By strategically selecting which data points to label, active learning aims to maximize model performance while minimizing the labeling effort, a crucial consideration in real-world HAR deployments.

Membership Query Synthesis represents one of the earliest and most theoretically intriguing approaches to active learning. In this framework, the learning algorithm generates synthetic examples for labeling, potentially exploring the entire feature space to find the most informative samples. While powerful in theory, applying membership query synthesis to HAR presents significant challenges. The primary difficulty lies in ensuring

the physical feasibility of synthesized sensor data. Unlike domains such as image classification, where synthetic images can be readily generated and interpreted, creating realistic sensor data that accurately represent human activities is far more complex. For instance, synthetic radar signals or accelerometer data must adhere to actual human movements' physical constraints and temporal coherence. Moreover, even if physically plausible data can be generated, it may confuse human annotators accustomed to labeling real sensor data. An expert might struggle to confidently label a synthesized radar signature that doesn't match the patterns they're familiar with from real-world activities.

Pool-based active learning, which assumes a pool of unlabeled data from which the model selects the most informative samples for labeling, aligns better with many HAR scenarios. Collecting vast amounts of unlabeled sensor data from users' daily activities is often feasible in a typical HAR deployment. The challenge then becomes identifying which subset of this data should be labeled to improve the model's performance most effectively. The pool-based approach typically follows an iterative process. A small subset of the data is initially labeled to train a preliminary model. This model is then used to assess the entire pool of unlabeled data, employing various strategies to identify the most informative samples. For example, the system might locate sequences of mmWave radar data where the model's predictions fluctuate rapidly between similar activities (e.g., eating and drinking). These sequences likely represent transitional movements or edge cases that, once labeled, could significantly improve the model's ability to distinguish between related activities. Once the most informative samples are identified, they are presented to human annotators for labeling. This newly labeled data is then added to the training set to update the model. This iterative approach allows the HAR system to progressively refine its understanding of activities, focusing human labeling effort on the most challenging or ambiguous cases.

Stream-based (or online) active learning presents a dynamic alternative well-suited to real-time HAR systems. In this paradigm, the model decides whether to request a label for each incoming sample in real time. This approach is especially relevant for HAR scenarios with continuous data streams, such as wearable devices or smart home systems constantly monitoring user activities. The critical challenge in stream-based active learning for HAR is making rapid, effective decisions about which samples to label while keeping pace with the incoming data stream. Various strategies have been developed to address this challenge. One common approach uses an uncertainty threshold, where labels are requested for samples exceeding a certain level of prediction uncertainty. This could be particularly useful for detecting and learning new or uncommon activities as they occur. Another strategy employs a budget-based approach, limiting the number of label requests within a given time window or data volume. This helps manage the burden

on human annotators and prevents the system from becoming overly reliant on frequent human input. In a HAR context, this might translate to requesting labels for no more than a certain number of activity segments per day, focusing on those the model finds most ambiguous or informative.

The integration of active learning with HAR systems offers several significant benefits. Firstly, it allows for more efficient use of human annotation resources, focusing labeling efforts on the most informative or challenging samples. This is particularly valuable in HAR, where the cost and intrusiveness of obtaining labeled activity data can be substantial. Secondly, active learning enables HAR systems to adapt more effectively to new users or environments. Since individual users may have unique ways of performing activities, or the system may encounter new contexts (e.g., different home layouts for a smart home system), active learning can help identify and learn these variations with minimal additional labeling effort. Furthermore, active learning can assist in addressing the challenge of concept drift in HAR, where the characteristics of activities may change over time due to factors like changes in user behavior or sensor degradation. By continuously identifying uncertain or misclassified samples, the system can adapt to these changes, maintaining high performance over extended periods.

However, implementing active learning in HAR systems also presents challenges. Balancing the desire for optimal sample selection with the need for rapid activity recognition is still an ongoing area of research. Furthermore, the real-time nature of many HAR applications requires careful consideration of the computational overhead introduced by active learning algorithms. Therefore, in Section 6.3, we comprehensively evaluate various active learning methods in the context of mmWave-based HAR. We explore the impact of different system parameters, such as labeling budget constraints, which directly affect the trade-off between annotation effort and model performance. Additionally, we investigate model update strategies to optimize the balance between computational efficiency and recognition accuracy. Through these investigations, we aim to provide insights into the most effective active learning approaches for real-world HAR applications, considering both performance and practical implementation constraints.

Chapter 4

Problem Statement

This chapter details the two critical challenges of mmWave-based HAR systems: (i) resource inefficiency in point cloud processing and (ii) the cold start problem of new users.

4.1 Resource Inefficiency in Point Cloud Processing

Directly processing mmWave point clouds with neural networks poses a challenge due to the varying number of points across temporal frames. For instance, fast-moving objects generate numerous radar reflection points, while stationary objects produce few or none. Voxelization offers a solution by converting the point cloud data into a structured 3D grid format. However, this approach presents a trade-off: Smaller voxels capture finer details, leading to higher accuracy, but also increase memory consumption. Fig. 4.1 illustrates the result of an experiment where we implemented FIA [8], a state-of-the-art voxelization-based food intake activity recognition method, and varied the voxel size to observe its impact on accuracy and memory consumption. It shows that there exists a practical limit beyond which further reduction in voxel size becomes infeasible due to GPU memory constraints. Voxelization often requires at least thousands of voxels to represent the data with sufficient detail, but mmWave radars collect sparse point clouds, typically yielding only a few dozen points per frame. Consequently, after voxelization, most voxels become zero, representing empty space. This approach is highly inefficient in terms of GPU memory utilization.

Additionally, point clouds processed as a 3D structure necessitates using three-dimensional neural networks (e.g., 3D-CNNs [55]) instead of their two-dimensional counterparts. This incurs a significant computational burden and demands substantial processing resources, hindering the deployment of such systems on resource-constrained single-board computers, such as the Raspberry Pi or Arduino. We address this issue by exploring more memory-efficient representations, such as *feature maps*, which preserve essential

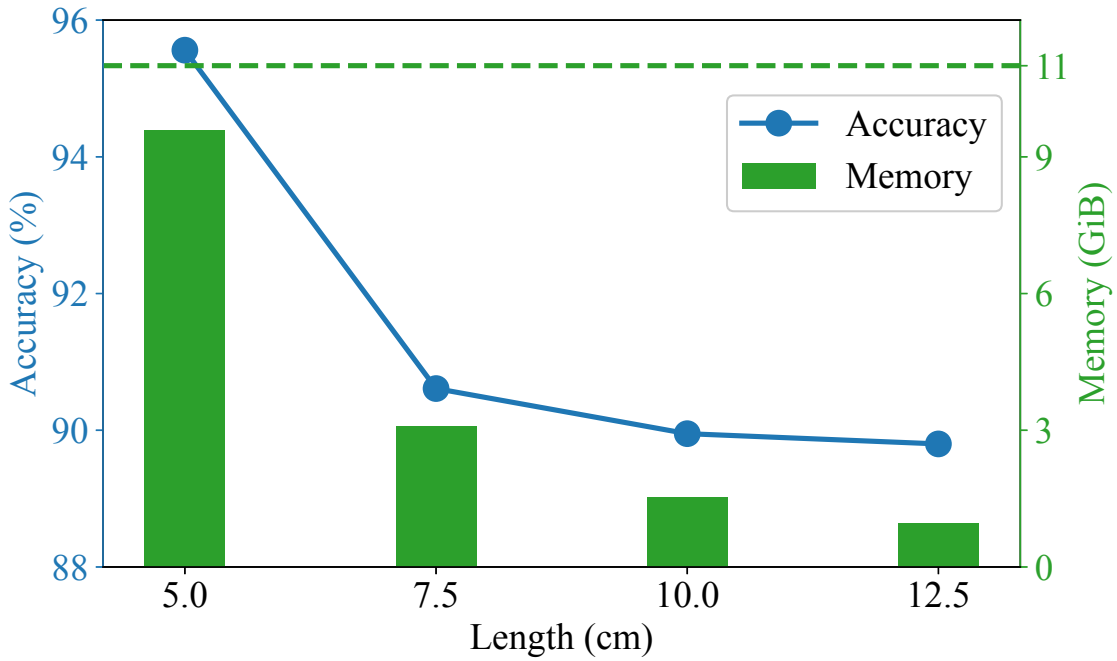


Figure 4.1: Impact of voxel size on accuracy and memory. Voxels with side lengths below ∼4.77 cm exceed our GPU memory limit of 11 GiB (dashed line).

features for accurate activity recognition while consuming fewer resources. Detailed explanations of this approach can be found in Section 5.1.

4.2 Cold Start Problem

In the context of HAR, the cold start problem emerges when the system encounters new users whose data were not included in the training process, resulting in a noticeable decline in accuracy. In our previous work [5], we randomly split our dataset, comprising 24 subjects (each performing 12 activities), into 80% for training and 20% for testing. This splitting method allowed samples from the same activity and subject to appear in both the training and test sets, likely contributing to the high test accuracy exceeding 90%. The model benefited from encountering similar samples during training, thereby performing well during testing. However, this scenario does not accurately reflect real-world situations where new users whose data the model has never encountered are introduced to the system. We investigate this by conducting a leave-one-out test, where the model was trained on the data from 23 subjects and tested on the remaining one. The results revealed a drop of almost 20% in recognition accuracy. Details of the experiment can be found in Section 6.3.

The cold start problem is particularly pronounced with mmWave point clouds because this data type is relatively scarce, unlike more common ones such as text and images, which benefit from extensive public datasets and pre-trained models. This is especially detrimental for today's deep learning models, which thrive on large datasets. Instead of attempting to improve generalization, which is difficult under these constraints, we focus on training a personalized model suitable for an individual user. To achieve this, we need to collect data from the new user for training purposes. However, requesting extensive labeling of samples is not feasible. In this context, active learning plays a crucial role by identifying the most informative samples for label querying, maximizing the model's accuracy while minimizing the number of queried labels. Detailed explanations of this approach can be found in Section 5.2.

Chapter 5

Methodology

This chapter introduces our proposed Dynamic Point Cloud Recognizer (DPR) model [5] and Personalized Active Learning for mmWave (PALM) framework, which extends DPR. Table 5.1 presents a list of symbols utilized throughout the thesis.

Table 5.1: List of Symbols

	Symbol	Description
Feature Map	x	x coordinate of the radar point
	y	y coordinate of the radar point
	z	z coordinate of the radar point
	v	Velocity of the radar point
	i	Intensity of the radar point
	r	Range (distance) between the radar and the radar point
	b	Bearing (angle) of the radar point relative to the radar
DPR	L	Input length of the LSTM layer
	N	Number of LSTM layers
	H	Number of hidden LSTM states
	D	Dropout rate during training
	B	Indicator for bidirectional LSTM employment
PALM	B	Labeling budget
	D	Dropout rate during inference
	K	Number of samples queried per active learning cycle
	E	Number of epochs updated per active learning cycle
	F	Indicator for freezing feature extraction layers

5.1 Dynamic Point Cloud Recognizer (DPR)

Preprocessing. DPR is designed to recognize activities from mmWave point clouds directly. It addresses the resource constraints outlined in Section 4.1 by converting each point cloud frame into a feature map instead of voxels during the preprocessing stage. This method draws inspiration from MARS [56], originally designed for rehabilitation applications. Each point cloud frame contains a maximum of 64 points[1], with each point comprising seven features: coordinates x, y, z, velocity v, intensity i, range r, and bearing b. Since r and b represent 2D polar coordinates, we can choose to utilize either $\{x, y, z\}$ or $\{r, b\}$. In our earlier work [5], we compared the two approaches and found that employing $\{x, y, z\}$ provides a slight advantage in accuracy, possibly due to the inclusion of the third dimension. Consequently, we discard $\{r, b\}$ and only adopt five features for each point. We spatially arrange these 64 points into 8×8 feature maps (sorted based on x, y, and then z, and zero-padded as necessary) with the five feature channels stacked. Utilizing these feature maps instead of voxelization effectively treats each point cloud frame as a 2D image representation (albeit with five channels instead of the typical three in RGB images). The reduction in dimensionality saves memory by avoiding the creation of largely empty voxel grids. Consequently, the computational load imposed on the neural network is significantly reduced, allowing faster and more efficient processing.

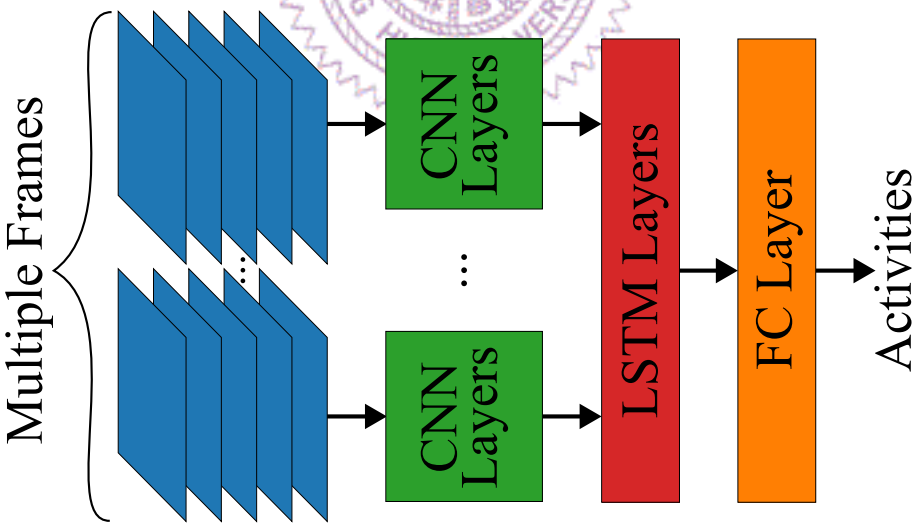


Figure 5.1: The neural network structure of DPR.

[1] According to the specification of our mmWave radar model (Texas Instrument IWR1443BOOST).

Neural network structure. Fig. 5.1 illustrates the neural network structure of DPR. Given that mmWave point clouds encompass spatial and temporal information, it is crucial to employ neural network structures that capture both of them effectively. The combination of Convolutional Neural Network (CNN) [57] and Long Short-Term Memory (LSTM) [58] offers a balanced, efficient, and robust solution. CNNs effectively extract spatial features by leveraging convolutional operations that detect local patterns, while LSTMs model temporal dependencies to maintain long-range temporal relationships. In DPR, multiple preprocessed frames are first passed through CNN layers for spatial features, then through LSTM layers for temporal features, and finally through fully connected layers for classification. In the context of mmWave-based HAR, this approach is superior to alternatives such as 3D-CNNs [55], which are computationally intensive and less efficient at handling long-term dependencies. ConvGRUs [59] and TCNs [60], though simpler and faster, may not capture the complexity of spatial features as effectively as CNNs or the long-term temporal dependencies as robustly as LSTMs. Transformers [61], despite their prowess in handling sequence data, often require vast amounts of data and computational power, making them less practical in our application.

Choice of CNN model. Since each point cloud frame is treated as a 2D image, widely-adopted CNN models for images, such as AlexNet [57], GoogLeNet [62], and ResNet [63], can be employed with minimal adjustments. Experiments in our earlier work [5] suggested that among these models (along with the original one employed by MARS [56]), ResNet produces the best results. Specifically, ResNet-34 outperformed both its 18- and 50-layer counterparts, suggesting that unquestioningly increasing the number of layers does not necessarily yield better results, likely due to overfitting. Therefore, while DPR is compatible with various CNN models, we adopt ResNet-34 as the CNN model for evaluation.

System parameters. DPR is characterized by several key parameters, including L, the output length of each CNN layer and the input length of the LSTM layer; N, the number of LSTM layers; H, the number of hidden LSTM states; D, the dropout rate for regularization; and B, a boolean indicating the use of Bidirectional LSTM. Optimal values for these parameters are determined through experiments specified in Section 6.2.

5.2 Personalized Active Learning for mmWave (PALM)

The objective of PALM is to mitigate the cold start problem in mmWave-based HAR systems. However, the limited availability of training data for mmWave point clouds makes it infeasible to develop a model that accurately works for any user. We address this challenge by creating a personalized model, ensuring that the model learns and adapts to the unique characteristics of the new user. Since we aim to classify samples into predefined sets of human activities rather than cluster similar samples without predefined categories, obtaining labeled data from new users is essential for training. Nonetheless, requesting a large amount of labeled data from users is impractical due to the associated burden and time consumption. Therefore, the system must identify and query labels for only the most critical samples - specifically, those where the model is most uncertain. We achieve this through active learning [39], a machine-learning strategy designed to enhance model performance by selectively querying labels for the most informative samples, thereby reducing the labeling burden on the user while maximizing the accuracy of the personalized model.

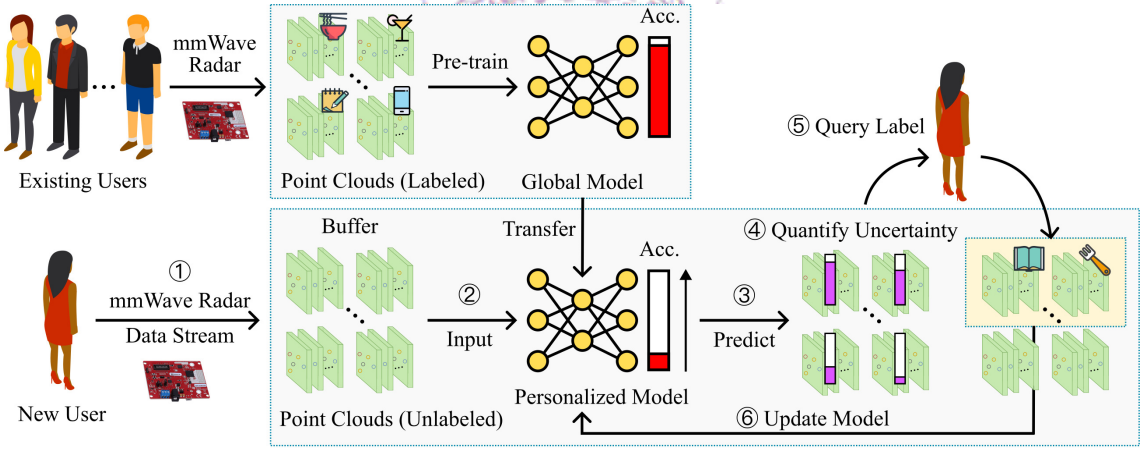


Figure 5.2: An overview of the PALM framework. Acc. stands for accuracy, and the height of the red bar illustrates the level of accuracy, while the height of the purple bar illustrates the level of uncertainty.

System overview. Fig. 5.2 provides an overview of the PALM framework. Before deployment with new users, a global model is pre-trained using labeled data in a dataset of existing users. Upon the arrival of a new user, the model weights are transferred to the personalized model for initialization. Subsequently, the daily active learning cycle begins with the following steps:

1. Capture the activity data stream from the new user and store it in a buffer.
2. Input the buffered samples into the personalized model for prediction.
3. Predict probabilities for classification from the model.
4. Quantify the uncertainty associated with each prediction.
5. Select samples with high uncertainty for label querying.
6. Update the personalized model using the newly acquired labeled data.
7. Repeat steps 3–6 until the labeling cost exceeds the labeling budget.
8. Clear the buffer and conclude the cycle for the day.

Algorithm 1 gives a detailed step-by-step procedure. The following paragraphs provide an in-depth explanation of the system's components and operation, focusing on: (i) the labeling budget, (ii) uncertainty quantification, (iii) output variability, (iv) query strategy, and (v) model update strategy.

Algorithm 1 Personalized Active Learning for mmWave (PALM)

Require: labeling budget B, number of queries K, number of epochs E

Pre-train a global model and transfer the weights to the personalized model
for each active learning cycle **do**
    Store the incoming data stream to the buffer
    $c \leftarrow 0$ ▷ labeling cost
    while $c < B$ **do**
        Make prediction(s) for each sample
        Quantify uncertainty for each sample based on the prediction(s)
        Query the labels for the samples with the top K highest uncertainties
        $c \leftarrow c + K$
        Train the personalized model for E epochs with the queried samples
    end while
end for

Labeling budget. A critical factor in an active learning system is the labeling budget, denoted as B, which, in our case, represents the number of label queries that can be made each day. A higher B allows for rapid model improvement by providing more labeled data in a shorter time, but it also risks overwhelming or annoying the user with frequent labeling requests. Conversely, a lower B reduces the user's burden but may slow the model's learning progress. Therefore, determining an optimal labeling budget is crucial for maintaining user engagement while ensuring efficient model enhancement. We will evaluate different values of B in Section 6.3.

Uncertainty quantification. Uncertainty of a machine learning model's prediction is a good metric for determining the chances of misclassification for that prediction [64]; in other words, uncertainty can be used to identify doubtful label predictions. In the context of HAR, uncertainty can be crucial in identifying which samples (those with the highest uncertainty) the model should prioritize for acquiring labels [53]. Although the ISO GUM [65] defines Type A standard uncertainty as the experimental standard deviation of the mean of multiple observations, it presents no methods for quantifying the uncertainty of ordinal quantities or nominal properties inherent in classification tasks; however, due to the need of knowing the doubt in a classification result, researchers and practitioners employ various approaches to estimate uncertainty in classification [66]. These methods include Least Confidence, Margin Sampling [9, 10], Deviation of Predictions [11, 12], and Information Entropy [14, 15]. Least Confidence (LC) selects the sample for which the model is least confident in its prediction, i.e., the sample with the lowest predicted probability for the most likely class. Given a sample x,

$$LC(x) = 1 - p_{best} \tag{5.1}$$

where p_{best} is the probability of the most likely class of sample x. Margin Sampling calculates the difference between the top two predicted class probabilities, with a smaller margin indicating higher uncertainty:

$$M(x) = p_{best} - p_{second} \tag{5.2}$$

where p_{best} and p_{second} are the probabilities of the most and second most likely classes, respectively. This is also called Best-vs.-Second-Best Margin. It is worth noting that other variations exist, such as Best-vs.-Worst Margin and Multi-class Margin. Deviation of Predictions, which is similar to Type A uncertainty quantification and uses multiple observations (or predictions, in the case of classification), quantifies the uncertainty as the deviation of the predicted probabilities across multiple predictions:

$$D(x) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(p_{best}^{m} - \overline{p}_{best})^2} \tag{5.3}$$

where M is the number of predictions, p^m_{best} is the probability of the most likely class in prediction m, and $\overline{p}_{best}$ is the average probability of the most likely class across all M predictions. Finally, information entropy is quantified as:

$$H(x) = -\sum_{i=1}^{C} p_i \log(p_i) \tag{5.4}$$

where C indicates the total number of classes, and p_i indicates the average predictive probability of class i. PALM employs Information Entropy, which has been shown to be one of the best indicators of uncertainty for classification [64]. We will evaluate PALM's performance against other uncertainty quantification methods in Section 6.3.

Multiple predictions. The Deviation of Prediction method described above necessitates multiple observations, i.e., multiple predictions. This can be achieved through several approaches [67] including Monte Carlo (MC) dropout, MC batch normalization, Deep Ensembles, and others. We employ MC dropout due to its simplicity and efficiency, which requires minimal modifications to the existing model architecture and can be easily integrated without significantly increasing computational overhead. MC dropout involves randomly dropping neurons in the neural network during inference, with a dropout rate of D. We employ the same D value used for regularization during training, the optimal value of which will be determined through the experiment specified in Section 6.2.

We can also incorporate multiple predictions to calculate Information Entropy. One such method is Entropy Mean, which computes the entropy of the mean prediction across numerous runs, potentially capturing a more robust representation of the model's uncertainty. Another approach is Max Entropy, which identifies the maximum entropy among a set of predictions, focusing on the most uncertain instances. Additionally, the Bayesian Active Learning by Disagreement (BALD) method [13] offers a nuanced perspective by calculating the difference between Entropy Mean (the entropy of the mean prediction) and Mean Entropy (the mean of individual predicted entropies). A higher BALD score indicates higher uncertainty on average, as different instances of the model disagree. We will compare these entropy-based methods in Section 6.3.

Query strategy. After quantifying the uncertainty of each prediction, we select the samples with the highest uncertainty for label querying. This ensures that labeling efforts focus on the most doubtful cases, which are likely to enhance the model's performance the most. Rather than selecting a single sample at a time, choosing the top K samples with the highest uncertainties is possible. However, we opt for $K = 1$ to minimize user burden. Asking users to label multiple samples at once can be intrusive and time-consuming, potentially leading to decreased user engagement and cooperation. Limiting the queries to one sample at a time ensures that the labeling process remains manageable and less disruptive for users, promoting long-term participation and data quality.

Diversity. In addition to uncertainty, we can utilize *diversity* [68] to ensure the model is exposed to a broad range of scenarios. Unlike uncertainty, which focuses on the ambiguity of the model predictions, diversity emphasizes the variability of the selected samples. By combining uncertainty and diversity using weights, we can balance the need to resolve ambiguous predictions with the need to cover a wide range of data distributions. Diversity can be quantified using the L1- or L2-norm, which measures the differences between feature vectors. These feature vectors can be derived from the raw input data or the outputs of feature extraction layers (CNN and LSTM) just before they are fed into the classification layers (fully connected layers). Additionally, *reinforcement learning* [69] can dynamically decide whether to prioritize uncertainty or diversity in each active learning cycle based on the model's current state. This approach allows for a more adaptive and efficient learning process, tailoring the selection strategy to the evolving needs of the model. However, incorporating diversity does not necessarily guarantee improved performance. For instance, prioritizing diversity might introduce samples that, although different, do not contribute significantly to reducing uncertainty, thereby diluting the effectiveness of the active learning cycle. Therefore, we will empirically evaluate the impact of these configurations in Section 6.3.

Model update strategy. We utilize the newly obtained labels to update the personalized model. There are two main approaches for that [48]: complete re-training and incremental training. Complete re-training is typically employed when a substantial amount of labels is obtained simultaneously, making the previous model irrelevant. Instead, we use incremental training, which fine-tunes the model to preserve existing knowledge without starting from scratch. A critical parameter in this process is the number of epochs, E, used for updating the model upon receiving new labeled data. This parameter directly influences the trade-off between computational efficiency and the degree of model refinement. Additionally, we introduce a binary parameter $F \in \{\mathit{True}, \mathit{False}\}$ to control the freezing of feature extraction layers (CNN and LSTM) during the update process. When $F = \mathit{True}$, only the classification layers (fully connected layers) are fine-tuned, while $F = \mathit{False}$ allows for updating the entire neural network architecture. The decision to freeze layers can conserve computational resources but may result in lower accuracy. In Section 6.3, we will evaluate different values of E and both states of F to identify the optimal configuration.

Chapter 6

Evaluations

In this chapter, we evaluate our proposed DPR model and PALM framework for mmWave-based HAR, focusing on both global and personalized model performances. Our primary objectives are to assess the accuracy and efficiency of the global model, investigate the necessity and benefits of personalized models, and explore the impact of various active learning strategies and transfer learning techniques on model performance.

6.1 Experimental Setup

Given the scarcity of publicly available mmWave-based human activity datasets, we have curated our own two datasets: (i) Food Intake Activity Dataset (FIAD) [6] and (ii) Driver Activity Dataset (DAD) [7]. Both datasets consist of sparse point clouds captured using the IWR1443BOOST model from Texas Instruments. The FIAD dataset comprises 24 subjects performing 12 activities, including three eating-related, three drinking-related, and six other daily activities, each performed for two minutes. The DAD dataset includes 15 subjects performing 11 driving activities, with four classified as safe and 11 as unsafe, each performed for one minute. Detailed descriptions of the specific activities included in FIAD and DAD are provided in Tables 6.1 and 6.2, respectively.

Table 6.1: Activities Included in the FIAD Dataset

Activity	Description	Time/Rep
a01	Drinking tea with a cup	4 sec
a02	Drinking tea with a bottle	4 sec
a03	Drinking tea with a straw	4 sec
a04	Eating a burger with hands	4 sec
a05	Eating fruit with a fork	4 sec
a06	Eating noodles with chopsticks	4 sec
a07	Sitting still	Continuous
a08	Picking up a call	4 sec
a09	Wiping mouth with a tissue	4 sec
a10	Writing on a piece of paper	4 sec
a11	Reading a book	4 sec
a12	Scrolling a smartphone	Continuous

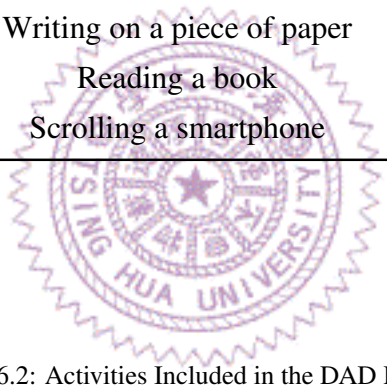


Table 6.2: Activities Included in the DAD Dataset

Activity	Description	Time/Rep
a01	Waiting	Continuous
a02	Driving safely	Continuous
a03	Changing gears	4 sec
a04	Checking mirrors	4 sec
a05	Drinking water	4 sec
a06	Touching hair	4 sec
a07	Talking to passengers	4 sec
a08	Checking the phone	4 sec
a09	Picking up a call	4 sec
a10	Nodding	Continuous
a11	Reaching sideways	4 sec

We primarily evaluate DPR and PALM with the FIAD dataset, which provides a diverse range of activities and ample data to test our methods thoroughly. Additionally, we conduct experiments involving transfer learning, where we pre-train the global model using the DAD dataset and then apply it to a new user in the FIAD dataset. This approach allows us to assess PALM's generalization capability in cross-application scenarios, testing its adaptability and robustness when transferred to a different activity recognition task.

We divide the mmWave point cloud data stream into individual samples using a four-second window (with a one-second stride), as most activities were recorded at a tempo of four seconds per repetition (with the exceptions noted in the tables). This division resulted in approximately 34,560 samples for the FIAD dataset (24 subjects $\times$ 12 activities $\times$ 120 seconds for each activity) and 9,900 samples for the DAD dataset (15 subjects $\times$ 11 activities $\times$ 60 seconds for each activity).

We implement our proposed methods using PyTorch 1.10 and conduct experiments on a Linux server with 2.1 GHz Intel Xeon CPUs and NVIDIA 1080Ti GPUs with 11 GiB of memory. For model training, we minimize the cross-entropy loss using the Adam optimizer with a learning rate of 0.001. These configurations are consistently utilized throughout this thesis unless otherwise specified.

6.2 Global Model

We evaluate the global model using the commonly adopted 80/20 train-test split [70]. This setting represents the upper bound performance, where all data are labeled. The resulting model serves as the pre-trained model for active learning, which we evaluate in the next section.

DPR parameters. We conduct pilot tests to determine the default system parameters for DPR. A suitable range of parameters was selected for each: $L \in \{\underline{39}, 256, 576\}$, $N \in \{\underline{1}, 2, 3\}$, $H \in \{\underline{64}, 128, 256\}$, $D \in \{0.1, \underline{0.3}, 0.5\}$, and $B \in \{\underline{\textit{True}}, \textit{False}\}$, with the default parameters underlined. Given the numerous combinations, we varied parameters individually while keeping others fixed at their default values. Fig. 6.1 presents the accuracy results under various parameter settings, enabling us to identify the optimal parameters: $L^* = 39$, $N^* = 1$, $H^* = 128$, $D^* = 0.3$, and $B^* = \textit{True}$. The dropout rate $D^* = 0.3$ will also be used during inference to generate multiple predictions with MC dropout, as described earlier.

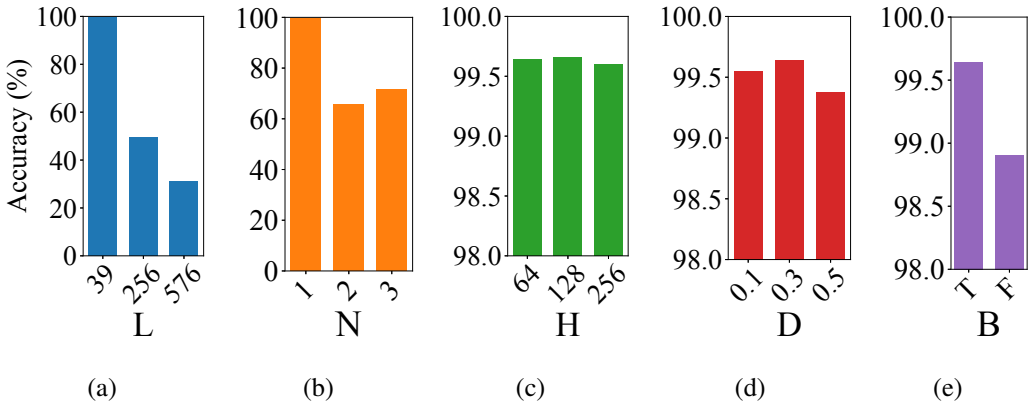


Figure 6.1: Accuracy comparison of DPR's system parameters: (a) L, (b) N, (c) H, (d) D, and (e) B.

Results. We evaluate DPR against FIA [8], a state-of-the-art voxelization-based food intake activity recognition method. While both approaches utilize a combination of CNN and LSTM, FIA requires a more resource-heavy 3D-CNN to process voxelized data. Using the optimal parameters, DPR achieves an accuracy of 99.66%, a modest improvement of +4.10% over FIA's accuracy of 95.56%. Fig. 6.2 shows the confusion matrix of DPR, with the strong diagonal pattern indicating high classification accuracy across all activities. Fig. 6.3 illustrates the accuracy improvements of DPR over FIA for each activity, ranging from 1.12% to 8.14%, demonstrating that DPR consistently outperforms FIA across all activities. More significantly, DPR demonstrates substantial efficiency gains in both memory consumption and computational resources: DPR consumes only 2131 MiB, a 78.29% reduction compared to FIA's 9817 MiB, while having an average inference time of 31.75 ms per sample, a 69.64% reduction compared to FIA's 104.58 ms. This dual advantage in memory and computational efficiency highlights the efficacy of using feature maps over voxelization, positioning DPR as a more resource-efficient and scalable solution for deployment on constrained devices. Fig. 6.4 demonstrates the superiority of DPR over FIA in terms of classification accuracy, memory consumption, and inference time.

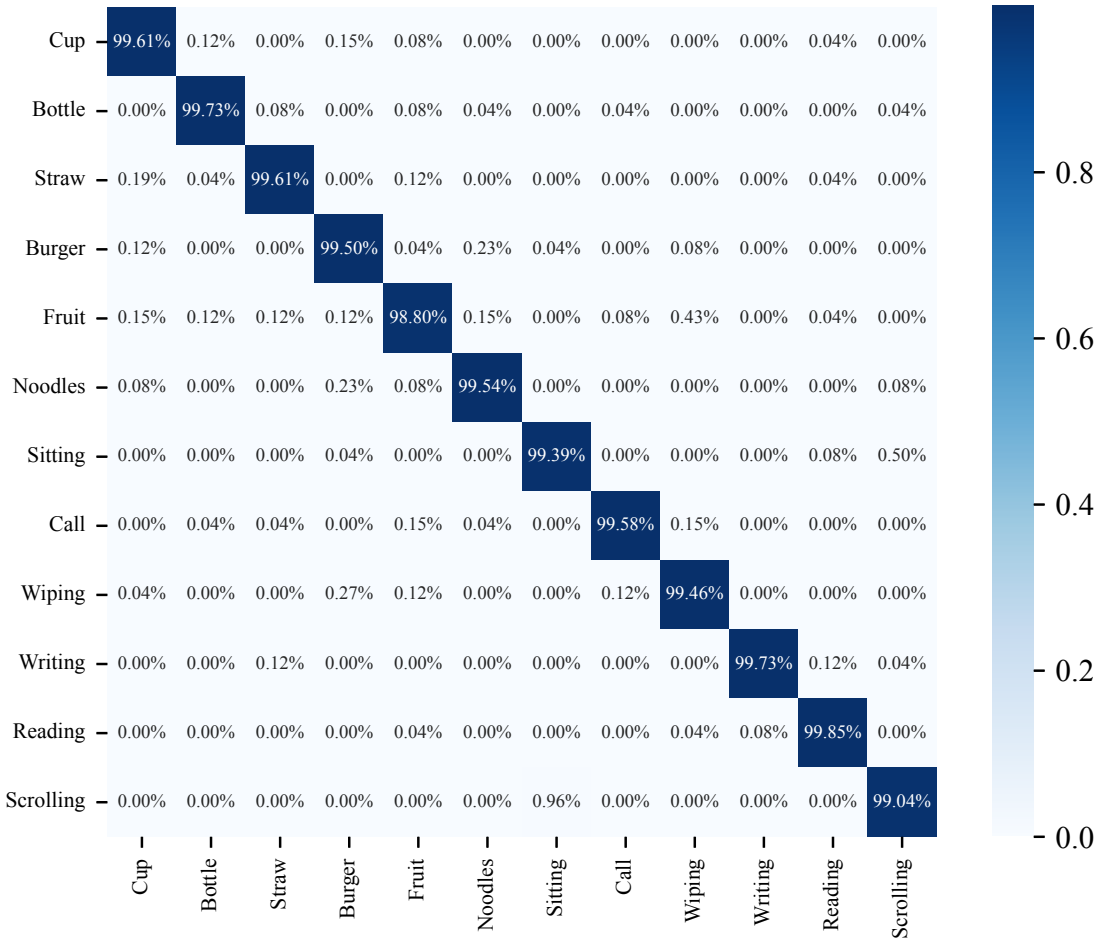


Figure 6.2: Confusion matrix the global model (80/20 train-test split). The prominent diagonal indicates high classification accuracy across all activities.

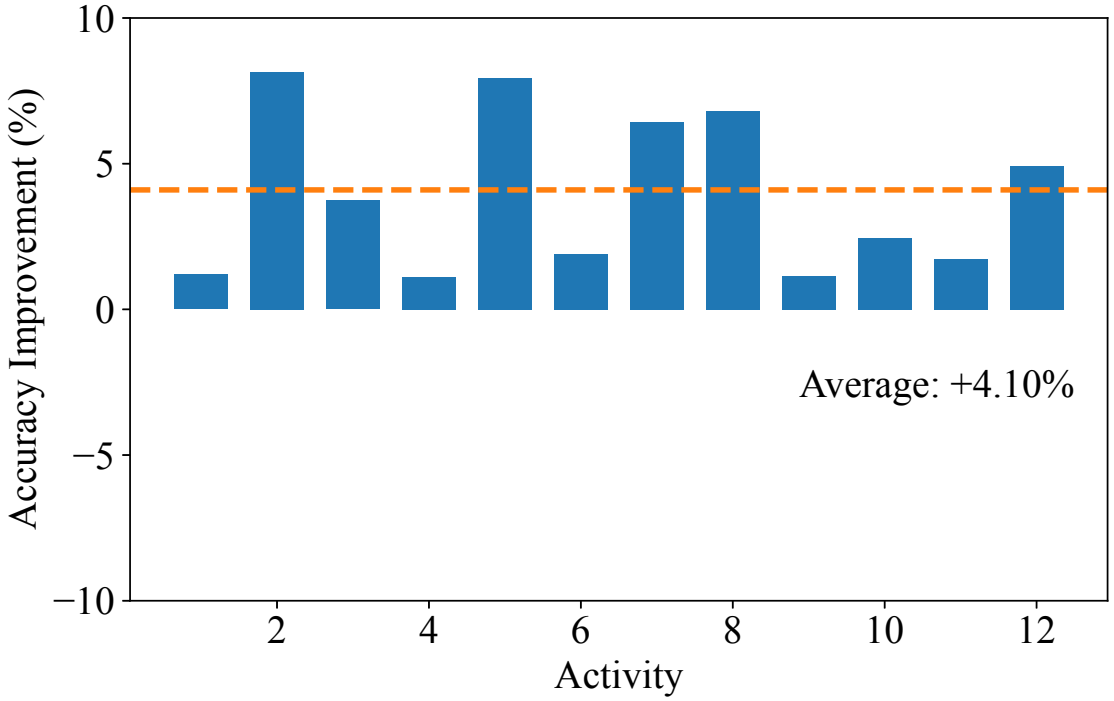
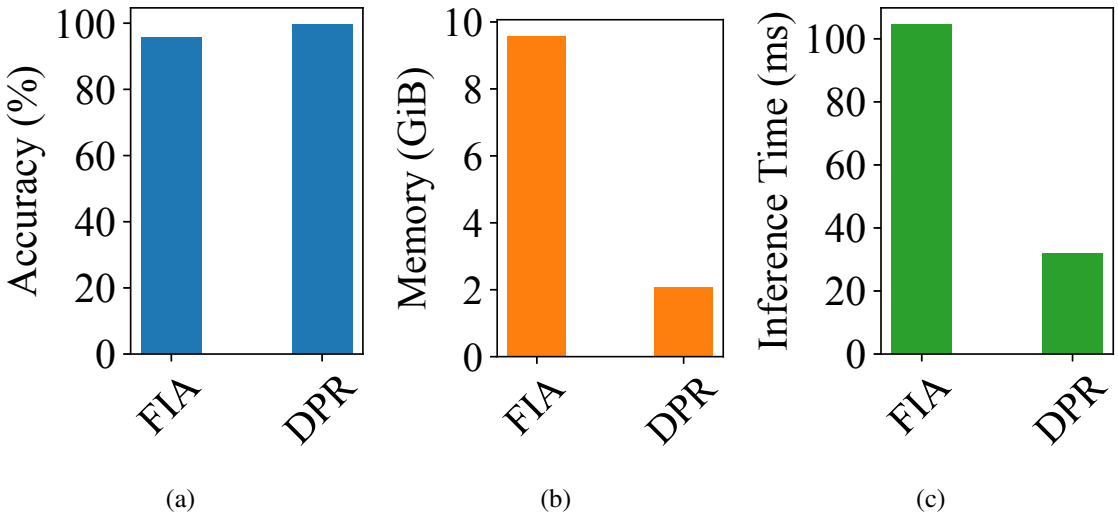


Figure 6.3: Accuracy improvements of DPR compared to FIA for each activity. Positive values indicate superior performance of DPR across all activities.



(a) (b) (c)

Figure 6.4: Comparison between FIA and DPR on: (a) classification accuracy, (b) memory consumption, (c) inference time.

6.3 Personalized Model

As a baseline, we apply the global model directly to a new and unseen subject, representing the lower bound performance. Fig. 6.5 presents the results of a 24-fold leave-one-out test, where each subject is treated as the new user while the remaining 23 subjects serve as existing users. The results demonstrate considerable inter-subject variability in accuracy, ranging from 38.84% to 85.42%, with an average of 73.24% (indicated by the dashed line). Notably, subject 23 exhibits particularly low accuracy. Upon further inspection, we discover that subject 23 exhibits movement patterns that differ significantly from other subjects, highlighting the necessity for personalized models. Fig. 6.6 shows the confusion matrix of the global model when applied to subject 23 as the new user before any personalization. This serves as a baseline for evaluating the performance of PALM, which we employ to train a personalized model for subject 23 in the next section.

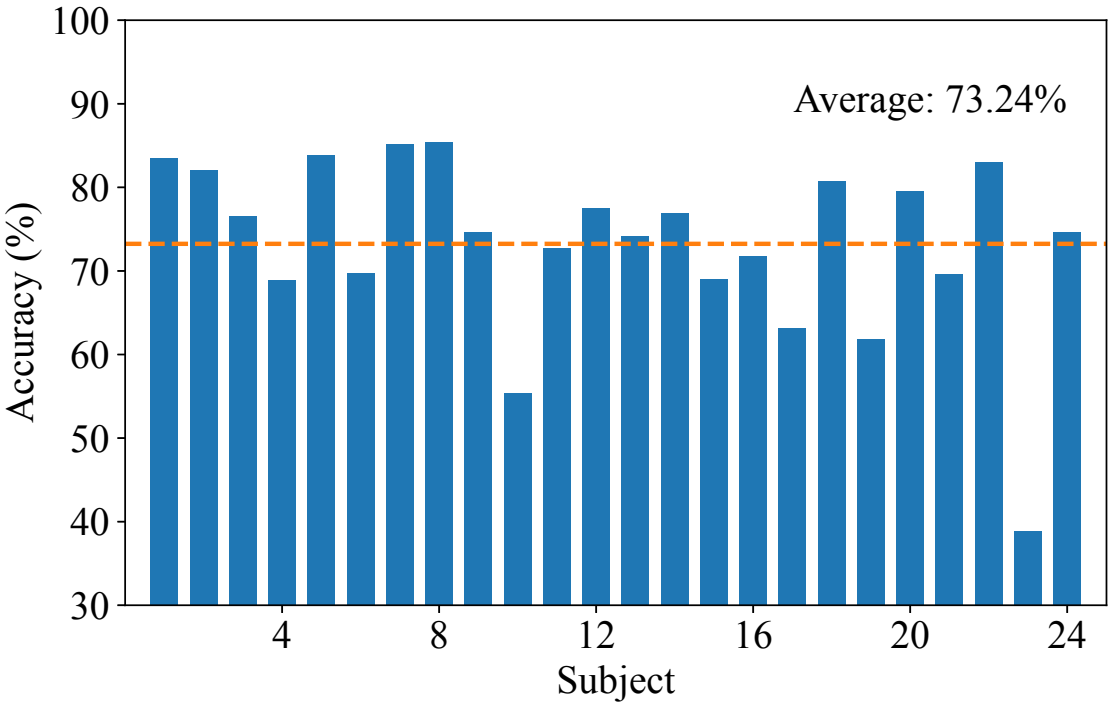


Figure 6.5: Baseline accuracy of the global model applied directly to new users without personalization. Each bar represents a subject serving as the new user in a leave-one-out test.

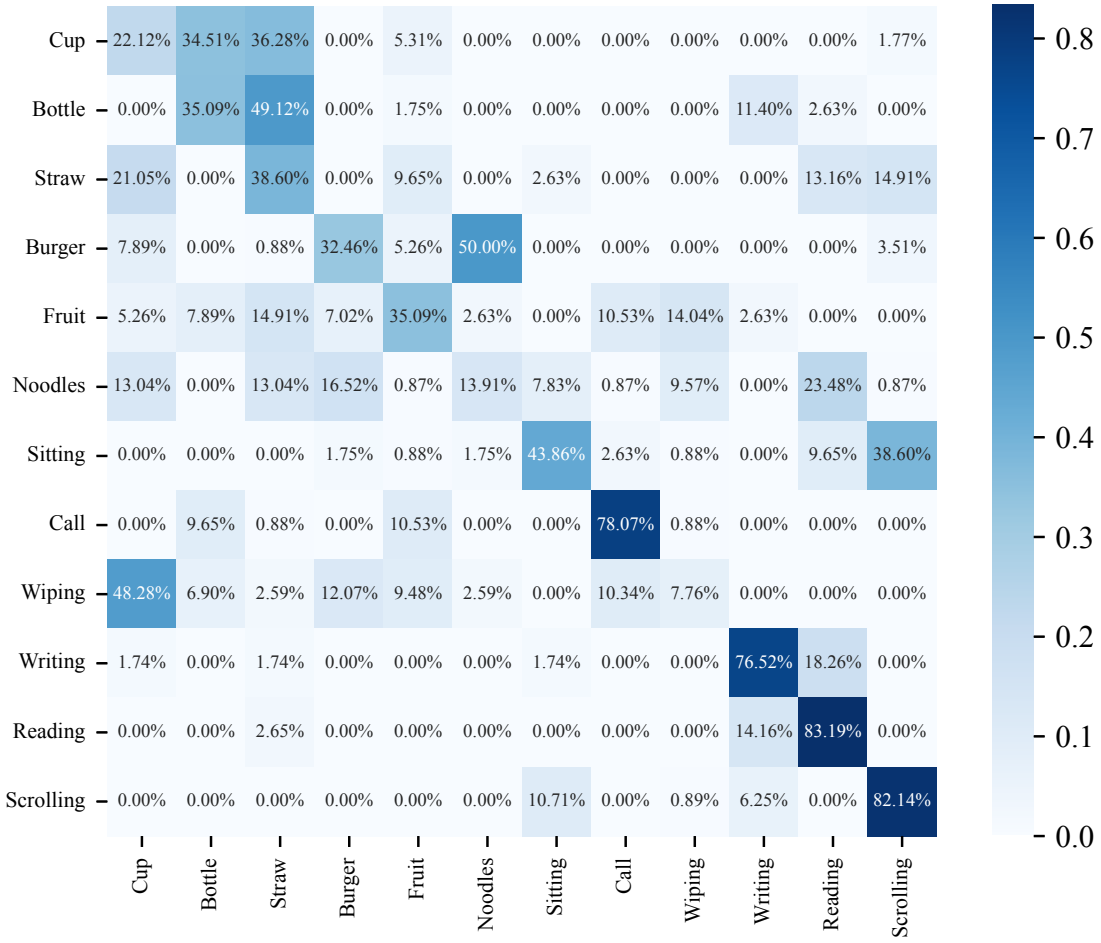


Figure 6.6: Confusion matrix of the global model's performance on subject 23 as the new user, representing the worst-case scenario among all subjects. This serves as the baseline for evaluating the effectiveness of PALM.

PALM parameters. To determine optimal parameters, we evaluate over a 10-day active learning period, focusing on: (a) the labeling budget B, (b) the number of training epochs E each time we obtain new labeled data, and (c) the feature extraction layer freezing parameter F. Given the context of food intake activities, where individuals typically consume three meals per day, we initially set $B = 3$ and evaluated $E = \{1, 10, 20, 40\}$ with $F =$ *False* (no freezing). The resulting accuracies, as illustrated in Fig. 6.7(a), suggest that increasing the number of training epochs can be beneficial, although excessive updates do not necessarily yield improvements in accuracy. Subsequently, we set $E = 20$ and evaluated $B = \{3, 6, 9, 12\}$. The accuracies presented in Fig. 6.7(b) indicate that while querying more samples generally enhances performance, excessive queries can lead to diminishing returns, potentially imposing an undue burden on users. Finally, with $B = 9$, we compare model performance for both values of F. The results, illustrated in Fig. 6.7(c), demonstrate that $F =$ *True* (freezing feature extraction layers) indeed leads to lower accuracy compared to $F =$ *False* (full model training). This finding suggests that effective personalization necessitates updating both the feature extraction and classification layers to better capture new users' unique motion patterns and characteristics. Based on these empirical findings, we establish the optimal configuration as $E = 20$, $B = 9$, and $F =$ *False* for subsequent PALM evaluations.

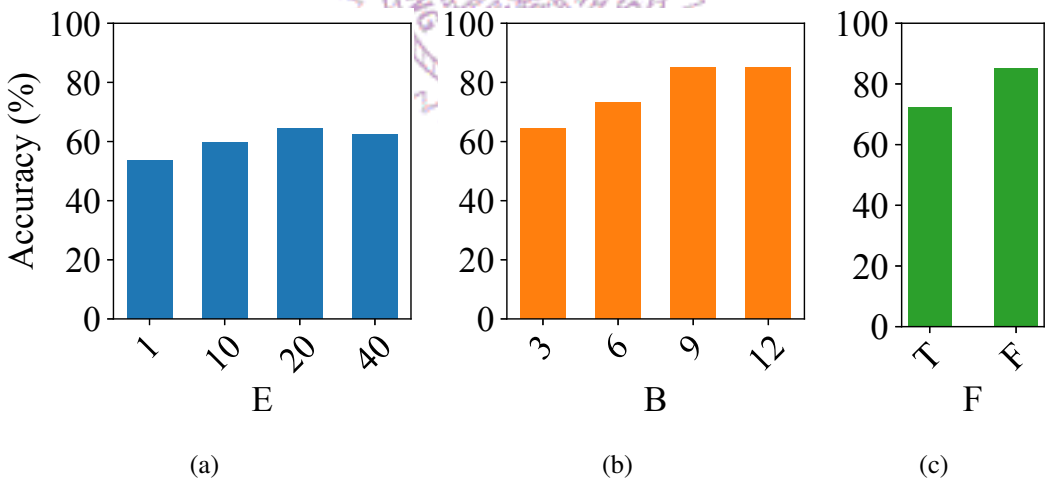
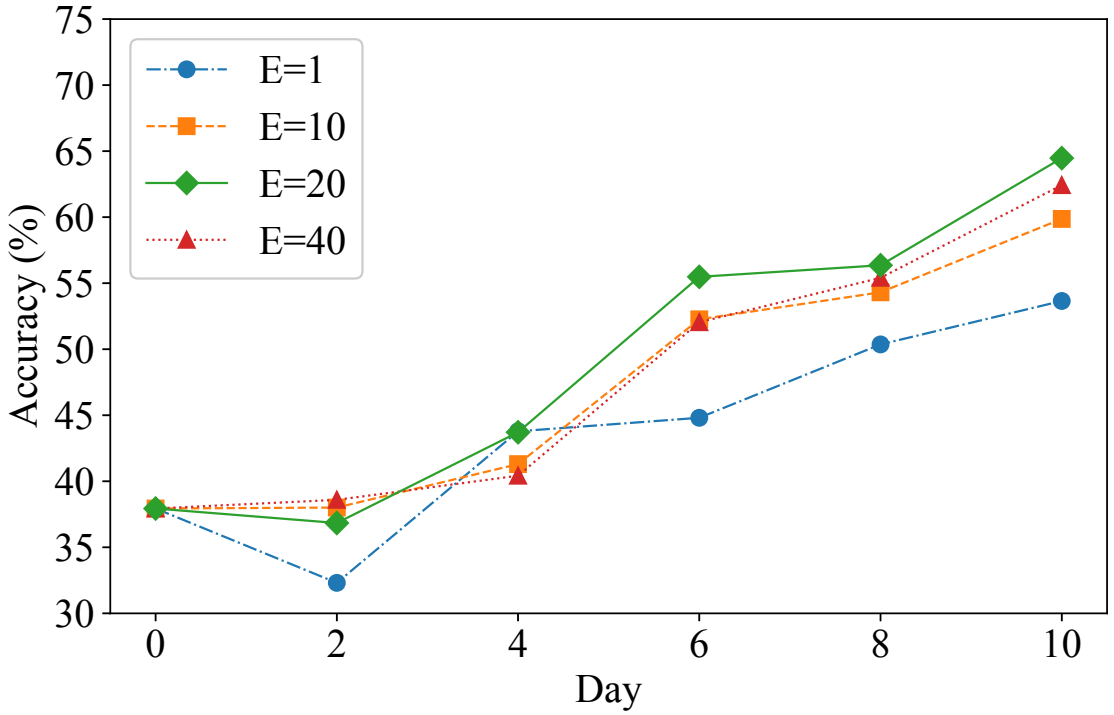
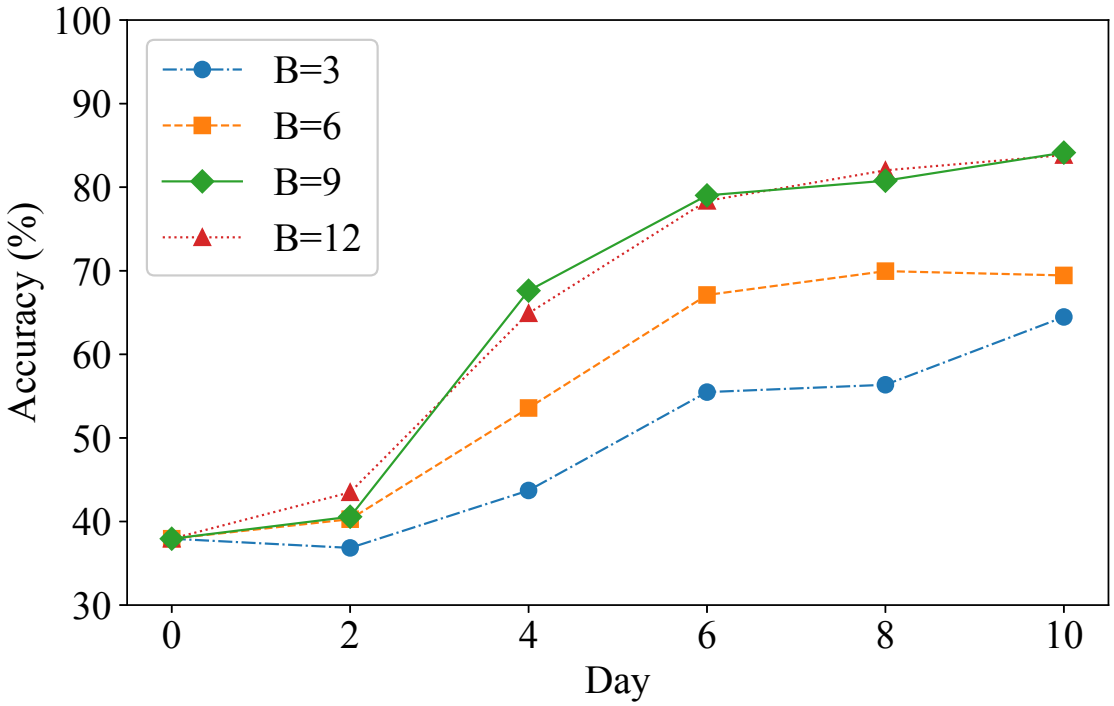


Figure 6.7: Accuracy comparison of PALM's system parameters: (a) E, (b) B, and (c) F.



(a)



(b)

Figure 6.8: Learning curves for parameters: (a) E, and (b) B. Increasing the E from 20 to 40 does not improve accuracy, and increasing B from 9 to 12 yields diminishing returns.

Results. Table 6.3 presents the evaluation of PALM compared to other methods over a two-week active learning period, using two key metrics: (a) classification accuracy, which measures the model's overall performance, and (b) Area Under the Curve (AUC), which captures the cumulative performance throughout the active learning process. The baseline model, which does not employ active learning, achieved an accuracy of 37.94%. In contrast, all active learning methods demonstrated substantial improvements. Random Query and Least Confidence methods achieved accuracies slightly below 90%. Margin Sampling performed better, with an accuracy of 90.86%. Deviation of Predictions only achieved an accuracy of 83.04%, suggesting that it may not effectively capture the true model uncertainty in mmWave-based HAR, possibly due to overemphasis on activities with inherently higher variability. Nevertheless, utilizing Information Entropy from a single prediction, our proposed PALM outperformed all the other methods, achieving the highest accuracy of 91.08%, with an upper bound of 98.25% observed over a more extended period (2 months). Furthermore, PALM exhibited the most substantial improvement in AUC compared to the baseline, with an increase of 87.66% and an upper bound of 137.86%. This superior performance is also evident in the learning curves shown in Fig. 6.9, where PALM's curve consistently remains above those of other methods. Fig. 6.10 shows the confusion matrix of PALM, where most of the activities are correctly recognized again. These results highlight PALM's effectiveness in training personalized models for mmWave-based HAR.

HUA UNIVERSITY

Table 6.3: Active Learning Evaluation

Method	**Acc.**	**AUC vs. Baseline**
Baseline (Without Active Learning)	37.94%	–
Random Query	89.55%	+66.93%
Least Confidence	87.21%	+79.02%
Margin Sampling	90.86%	+71.46%
Deviation of Predictions	83.04%	+68.27%
PALM (Entropy) (Single Prediction)	**91.08%**	**+87.66%**
PALM (Upper Bound)	**98.25%**	**+137.86%**

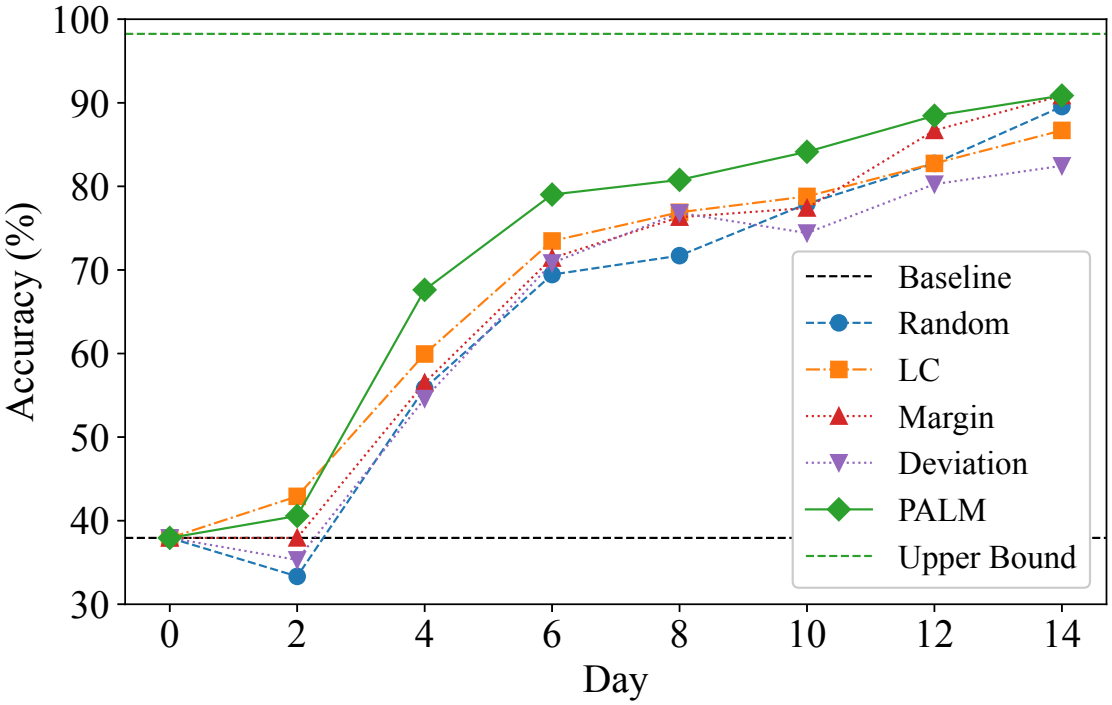


Figure 6.9: Learning curves comparing different methods over a two-week active learning period, where PALM's curve consistently remains above those of other methods.

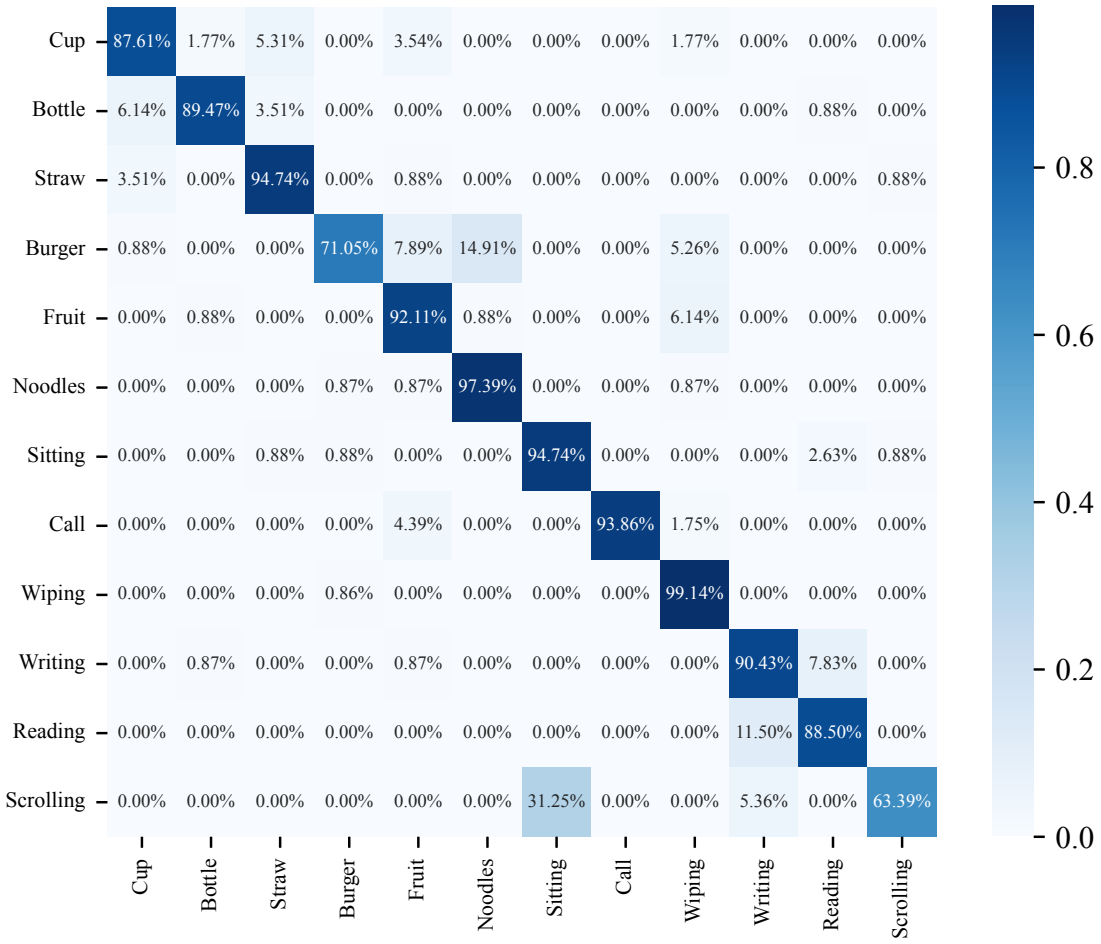


Figure 6.10: Confusion matrix of the personalized model's performance on subject 23 (PALM, after a 14-day active learning period).

Entropy-based methods. Table 6.4 further evaluates various entropy-based methods compared to the baseline without active learning. Interestingly, Entropy Mean, which calculates the entropy from the mean of 50 predictions, shows a slight decrease in performance compared to the single prediction approach. This suggests that simply averaging multiple predictions may not necessarily lead to enhanced performance. The most effective method among those evaluated is Max Entropy, which selects the maximum entropy value from 50 predictions. This method achieves the highest accuracy of 92.03% and the largest AUC improvement of 90.85% relative to the baseline. The superior performance of this approach indicates that considering the most uncertain predictions may be more informative for active learning than averaging across multiple predictions. Lastly, BALD yields an accuracy of 89.25% and an 80.49% improvement in AUC. While this approach demonstrated enhancement over the baseline, it did not outperform the other entropy-based methods in this evaluation. Notably, while Max Entropy achieves the highest accuracy by a small margin, all methods except Entropy (Single Prediction) require more computational resources, as they involve making multiple predictions for each sample. This increase in computational cost should be carefully considered when selecting an approach for practical applications.

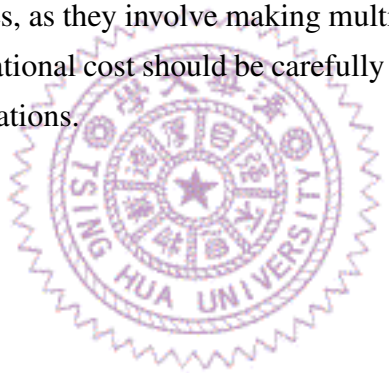


Table 6.4: Entropy-based Methods Evaluation

Method	**Acc.**	**AUC vs. Baseline**
Baseline (Without Active Learning)	37.94%	–
Entropy (Single Prediction)	91.08%	+87.66%
Entropy Mean (50 Predictions)	89.99%	+86.83%
Max Entropy (50 Predictions)	**92.03%**	**+90.85%**
BALD (50 Predictions)	89.25%	+80.49%

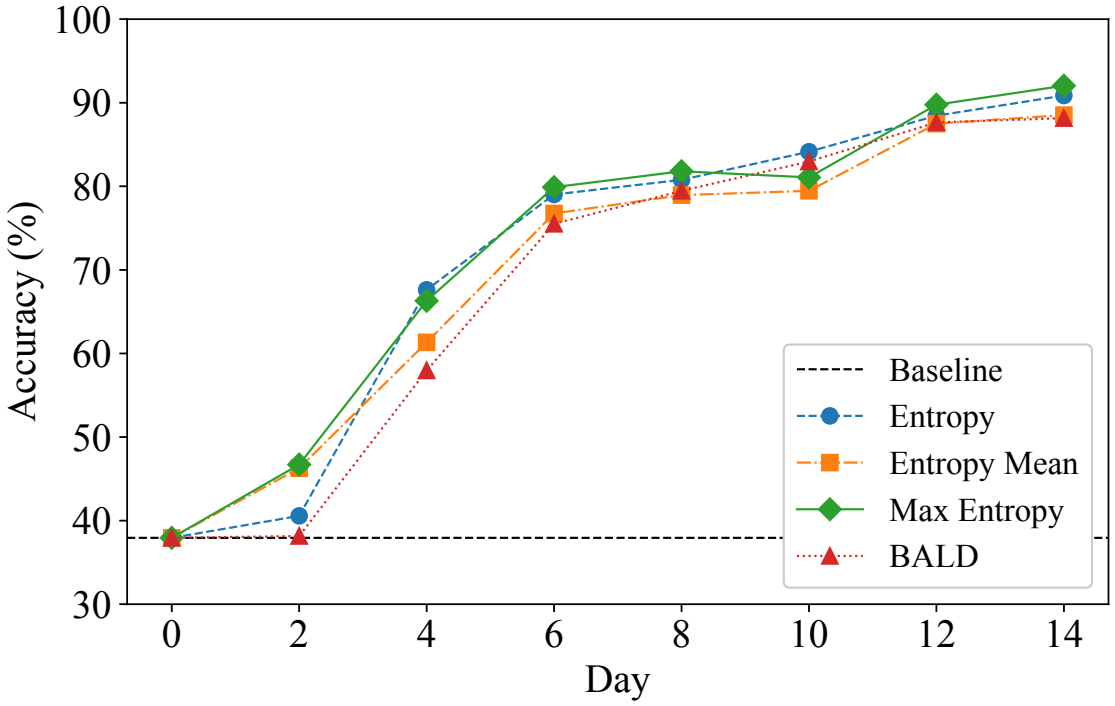


Figure 6.11: Learning curves comparing different entropy-based methods over a two-week active learning period.

Diversity-based methods. Table 6.5 and Fig. 6.12 show the results of assessing whether combining diversity with uncertainty could enhance PALM's performance. Initially, a weighted combination of uncertainty and diversity was employed using equal weights (0.5, 0.5), and two diversity calculation methods are compared: L1-norm and L2-norm. L2-norm achieves an accuracy of 80.04% compared to 79.09% for L1-norm. Subsequently, the weighting scheme is adjusted to (0.7, 0.3), favoring uncertainty. This modification yields improved results, with an accuracy of 88.23%, suggesting that uncertainty is indeed a more potent indicator of labeling importance. An alternative approach using the output of the feature extraction layer (LSTM) instead of raw input features for diversity calculation is also evaluated. However, this method achieves a slightly lower accuracy of 85.23%. Finally, a reinforcement learning approach dynamically chooses between labeling the highest uncertainty sample or the highest diversity sample based on the day-to-day accuracy gain as the reward. This method achieves an accuracy of only 79.61%. Notably, while it eventually learns to favor uncertainty over diversity after many days, its initial exploration phase results in suboptimal performance, making it less effective than using a fixed weighting scheme. Remarkably, none of the methods incorporating diversity outperforms the uncertainty-only approach, which maintains the highest accuracy at 91.08%. Several factors may contribute to this phenomenon:

- Dataset characteristics: The dataset may possess inherent properties that make uncertainty a more reliable indicator of informative samples than diversity.
- Diversity measurement limitations: The chosen diversity metrics (L1-norm and L2-norm) may not adequately capture the true diversity of samples in the feature space for this particular task.
- Interaction effects: There could be complex interactions between uncertainty and diversity that are not effectively captured by simple linear combinations or reinforcement learning approaches.
- Saturation effect: As the model improves, the remaining uncertain samples might inherently represent diverse regions of the feature space, making explicit diversity considerations redundant.

These findings underscore the complexity of integrating diversity into active learning strategies and highlight the need for further research to develop more sophisticated methods that can consistently leverage both uncertainty and diversity to enhance model performance across various tasks and datasets.

Table 6.5: Diversity-based Methods Evaluation

Method	**Acc.**	**AUC**
Uncertainty Only (Entropy, Single Prediction)	**91.08%**	–
0.5 Uncertainty + 0.5 Diversity (L1-norm) (Raw Input)	79.09%	-18.85%
0.5 Uncertainty + 0.5 Diversity (L2-norm) (Raw Input)	80.04%	-21.32%
0.7 Uncertainty + 0.3 Diversity (L2-norm) (Raw Input)	88.23%	-7.90%
0.7 Uncertainty + 0.3 Diversity (L2-norm) (LSTM Output)	85.23%	-8.14%
Reinforcement Learning (Highest Uncertainty or Diversity)	79.61%	-14.92%

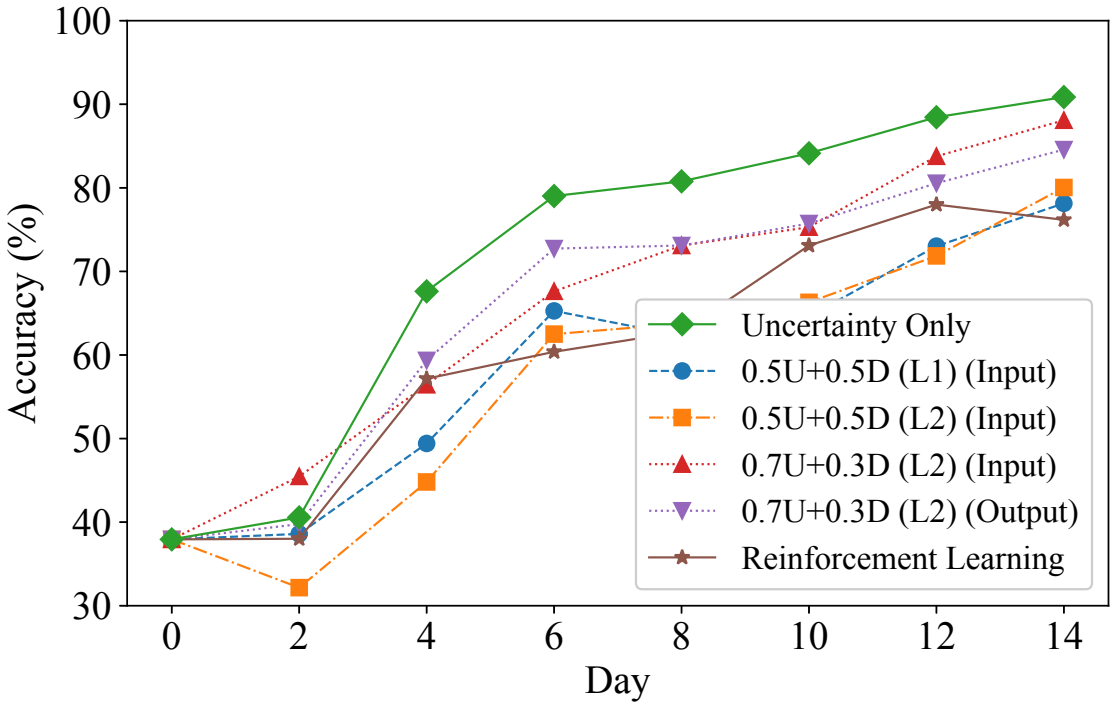


Figure 6.12: Learning curves comparing different ways to incorporate diversity with uncertainty over a two-week active learning period.

Transfer learning. Transfer learning enables the adaptation of knowledge from one domain to another by initializing a model with pre-trained weights. We apply this technique to enhance food intake activity recognition using models pre-trained on driver activity data. Specifically, we utilize a direct one-to-one mapping of activity labels between datasets, e.g., a01 in DAD (Table 6.1) to a01 in FIAD (Table 6.2), rather than manually aligning similar activities, since fine-tuning the classification layers is necessary for both approaches. Table 6.6 presents the evaluation results. The baseline model, trained from scratch without transfer learning, achieved an accuracy of 80.99%. As expected, the upper bound model pre-trained specifically on food intake activities demonstrated superior performance with an accuracy of 98.25% and an AUC improvement of 50.75%. Interestingly, leveraging data from other activity domains proved beneficial even without pre-trained data specific to food intake activities. For instance, models pre-trained on the DAD dataset exhibited notable improvements over the baseline. Among these, the model pre-trained using front-facing camera data achieved the highest accuracy of 90.86%, with an AUC increase of 19.48%. The learning curves in Fig. 6.13 indicate that the front-facing camera pre-trained model consistently performs above the baseline and other camera angles. This aligns with our intuition, as our FIAD dataset also employs a front-facing camera angle. These findings highlight the importance of transfer learning with pre-trained data, demonstrating that our proposed PALM can benefit even from models pre-trained with mmWave data of other types of activities.

Table 6.6: Transfer Learning Evaluation

Method	**Acc.**	**AUC**
Baseline (Train from Scratch)	80.99%	–
Transfer (Driver Activities, Body Cam)	90.06%	+10.29%
Transfer (Driver Activities, Face Cam)	**90.86%**	**+19.48%**
Transfer (Driver Activities, Hands Cam)	86.04%	+8.06%
Upper Bound (Food Intake Activities)	98.25%	+50.75%

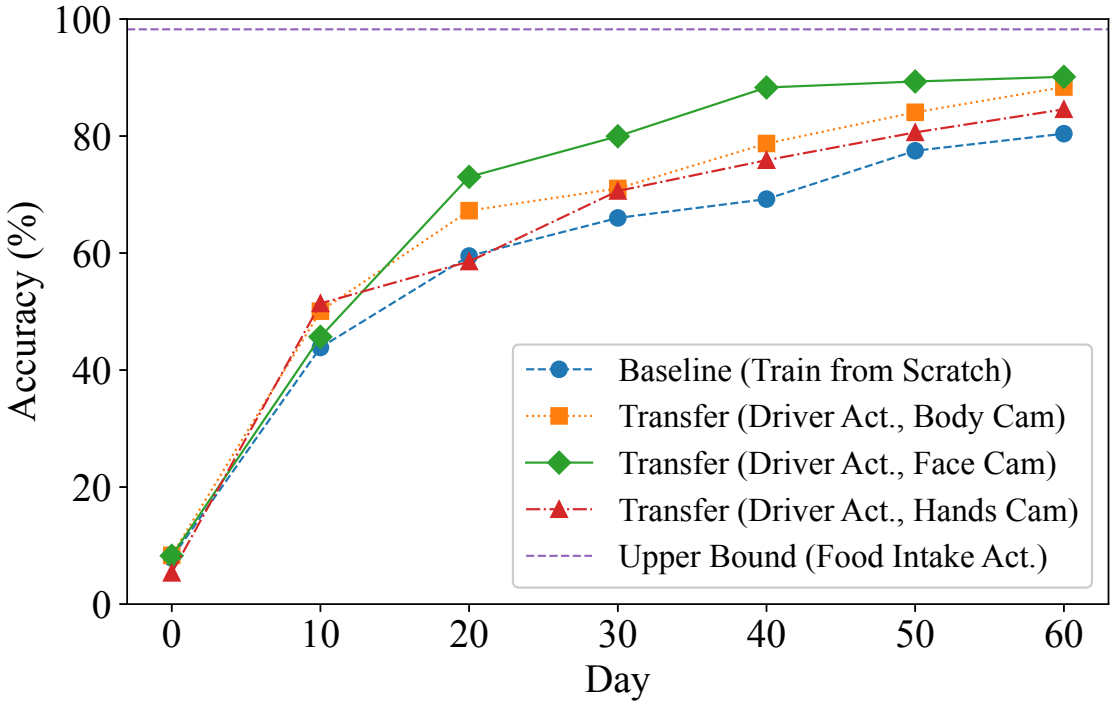


Figure 6.13: Learning curves comparing different transfer learning scenarios for food intake activity recognition, where the model pre-trained with face camera data performs the best among the driver activity scenarios.

6.4 Summary of Findings

Evaluating the global and personalized models for food intake activity recognition reveals several key insights. First, the global model performs the best compared to the state-of-the-art voxelization-based method:

- An accuracy improvement of 4.10%, achieving 99.66%.
- A memory consumption reduction of 78.29%.
- An inference time reduction of 69.64%.

Second, the personalized model trained using our PALM framework over a 14-day active learning period achieves the best performance compared to the state-of-the-art active learning methods:

- Accuracy of 91.08%, with an upper bound of 98.25%.
- Largest AUC improvement of 87.66%, with an upper bound of 137.86%.

Third, among four entropy-based methods, Max Entropy achieved the highest accuracy of 92.03% and AUC improvement of 90.85%. Fourth, none of the diversity-based methods outperforms the uncertainty-based ones, showing the efficacy of our uncertainty-based approaches. Finally, transfer learning evaluations demonstrated that pre-training on related activities, particularly using data captured with similar camera angles, substantially benefits model performance, achieving accuracy improvements up to 90.86%.

Chapter 7

Conclusion

In this chapter, we summarize the key findings of our thesis and highlight the advancements made in addressing the challenges of mmWave-based HAR, particularly focusing on resource efficiency and personalized model accuracy.

7.1 Concluding Remark

In this thesis, we explored the potential of mmWave radars for HAR and addressed the challenges associated with resource inefficiency in point cloud processing and accuracy degradation for new users. By introducing DPR and PALM, we proposed solutions to train memory-efficient personalized models effectively. Experimental results on our Food Intake Activity Dataset [6] demonstrate the effectiveness of our approach: (i) DPR outperforms previous state-of-the-art voxelization-based methods, achieving a 4.10% increase in classification accuracy while reducing memory consumption by 78.29% and inference time by 69.64%; (ii) PALM outperforms the baseline and four other active learning methods, attaining an accuracy of 91.08% over a two-week active learning period, with an upper bound of 98.25% observed over an extended period; (iii) among the four entropy-based methods evaluated, Max Entropy achieved the highest accuracy of 92.03% and an AUC improvement of 90.85%, underscoring its effectiveness despite requiring more computational resource; (iv) none of the diversity-based methods outperformed the uncertainty-based ones, highlighting the efficacy of our uncertainty-based approaches in improving model accuracy and efficiency; (v) leveraging a model pre-trained with our Driver Activity Dataset [7], PALM exhibits a 9.87% higher accuracy and a 19.48% improvement in AUC compared to the baseline model trained from scratch, illustrating its ability to benefit from cross-application transfer learning when there is insufficient data for some activities.

7.2 Future Directions

While our current work shows significant improvements in mmWave-based HAR, there are several promising directions for future research.

- First, we can enhance the sample selection strategy in active learning by exploring alternative methods of quantifying uncertainty and diversity.
- Second, we can formulate the querying decision as a reinforcement learning problem, expanding the selection beyond merely choosing between the highest uncertainty and highest diversity samples to include a broader range of options.
- Third, we can investigate the application of PALM to other domains [71] beyond food intake and driver activity recognition. This could help assess its generalizability and identify any necessary domain-specific adaptations.
- Fourth, integrating PALM with other sensing modalities or data sources [72, 73] could create more comprehensive and accurate HAR systems, addressing some limitations inherent to mmWave radar-based approaches.
- Lastly, exploring *federated learning* [74] for PALM could allow for privacy-preserving personalized models by enabling decentralized learning from distributed data sources without sharing raw data.

Bibliography

[1] WHO. (2023) Noncommunicable diseases. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

[2] L. Harnack, L. Steffen, D. Arnett, S. Gao, and R. Luepker, "Accuracy of estimation of large food portions," *Journal of the American Dietetic Association*, vol. 104, no. 5, pp. 804–806, 2004.

[3] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 652–660.

[4] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4490–4499.

[5] H.-C. Chiang, Y.-H. Wu, S. Shirmohammadi, and C.-H. Hsu, "Memory-efficient high-accuracy food intake activity recognition with 3D mmWave radars," in *Proc. of the ACM International Workshop on Multimedia Assisted Dietary Management (MADiMa)*, 2023, pp. 33–41.

[6] Y.-H. Wu, H.-C. Chiang, S. Shirmohammadi, and C.-H. Hsu, "A dataset of food intake activities using sensors with heterogeneous privacy sensitivity levels," in *Proc. of the ACM Multimedia Systems Conference (MMSys)*, 2023, pp. 416–422.

[7] G.-H. Li, H.-C. Chiang, Y.-C. Li, S. Shirmohammadi, and C.-H. Hsu, "A driver activity dataset with multiple RGB-D cameras and mmWave radars," in *Proc. of the ACM Multimedia Systems Conference (MMSys)*, 2024, pp. 360–366.

[8] Y.-H. Wu, Y. Chen, S. Shirmohammadi, and C.-H. Hsu, "AI-assisted food intake activity recognition using 3D mmWave radars," in *Proc. of the ACM International Workshop on Multimedia Assisted Dietary Management (MADiMa)*, 2022, pp. 81–89.

[9] D. Roth and K. Small, "Margin-based active learning for structured output spaces," in *Proc. of the European Conference on Machine Learning (ECML)*. Springer, 2006, pp. 413–424.

[10] M.-F. Balcan, A. Broder, and T. Zhang, "Margin based active learning," in *Proc. of the International Conference on Computational Learning Theory (COLT)*. Springer, 2007, pp. 35–50.

[11] J. M. Dolezal, A. Srisuwananukorn, D. Karpeyev, S. Ramesh, S. Kochanny, B. Cody, A. S. Mansfield, S. Rakshit, R. Bansal, M. C. Bois *et al.*, "Uncertainty-informed deep learning models enable high-confidence predictions for digital histopathology," *Nature communications*, vol. 13, no. 1, p. 6572, 2022.

[12] B. Ghoshal, A. Tucker, B. Sanghera, and W. Lup Wong, "Estimating uncertainty in deep learning for reporting confidence to clinicians in medical image segmentation and diseases detection," *Computational Intelligence*, vol. 37, no. 2, pp. 701–734, 2021.

[13] N. Houlsby, F. Huszár, Z. Ghahramani, and M. Lengyel, "Bayesian active learning for classification and preference learning," *arXiv preprint arXiv:1112.5745*, 2011.

[14] L. Mari, C. Narduzzi, G. Nordin, and S. Trapmann, "Foundations of uncertainty in evaluation of nominal properties," *Measurement*, vol. 152, p. 107397, 2020.

[15] A. Possolo, "Statistical models and computation to evaluate measurement uncertainty," *Metrologia*, vol. 51, no. 4, p. S228, 2014.

[16] H. Raeis, M. Kazemi, and S. Shirmohammadi, "Human activity recognition with device-free sensors for well-being assessment in smart homes," *IEEE Instrumentation & Measurement Magazine*, vol. 24, no. 6, pp. 46–57, 2021.

[17] H. Liu and T. Schultz, "A wearable real-time human activity recognition system using biosensors integrated into a knee bandage," in *Proc. of the International Conference on Biomedical Electronics and Devices (BioDevices)*, 2019, pp. 47–55.

[18] A. Moin, A. Zhou, A. Rahimi, A. Menon, S. Benatti, G. Alexandrov, S. Tamakloe, J. Ting, N. Yamamoto, Y. Khan *et al.*, "A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition," *Nature Electronics*, vol. 4, no. 1, pp. 54–63, 2021.

[19] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Intelligent human-computer interaction based on surface EMG gesture recognition," *IEEE Access*, vol. 7, pp. 61 378–61 387, 2019.

[20] A. Salehzadeh, A. Calitz, and J. Greyling, "Human activity recognition using deep electroencephalography learning," *Biomedical Signal Processing and Control*, vol. 62, p. 102094, 2020.

[21] A. Stisen, H. Blunck, S. Bhattacharya, T. Prentow, M. Kjaergaard, A. Dey, T. Sonne, and M. Jensen, "Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition," in *Proc. of the ACM Conference on Embedded Networked Sensor Systems (SenSys)*, 2015, pp. 127–140.

[22] N. Ahmed, J. Rafiq, and M. Islam, "Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model," *Sensors*, vol. 20, no. 1, pp. 317:1–317:19, 2020.

[23] M. Hassan, M. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, vol. 81, pp. 307–313, 2018.

[24] G. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, vol. 7, pp. 133 190–133 202, 2019.

[25] S. Balli, E. Saugbacs, and M. Peker, "Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm," *Measurement and Control*, vol. 52, no. 1-2, pp. 37–45, 2019.

[26] S. Mekruksavanich and A. Jitpattanakul, "Smartwatch-based human activity recognition using hybrid LSTM network," in *Proc. of the IEEE Sensors Conference*, 2020, pp. 1–4.

[27] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.

[28] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proc. of the ACM International Conference on Multimedia (MM)*, 2015, pp. 1307–1310.

[29] A. Franco, A. Magnani, and D. Maio, "A multimodal approach for human activity recognition based on skeleton and RGB data," *Pattern Recognition Letters*, vol. 131, pp. 293–299, 2020.

[30] T. Singh and D. Vishwakarma, "A deeply coupled ConvNet for human activity recognition using dynamic and RGB images," *Neural Computing and Applications*, vol. 33, no. 1, pp. 469–485, 2021.

[31] F. Baradel, C. Wolf, and J. Mille, "Human activity recognition with pose-driven attention to RGB," in *Proc. of the British Machine Vision Conference (BMVC)*, 2018, pp. 1–14.

[32] Z. Wharton, A. Behera, Y. Liu, and N. Bessis, "Coarse temporal attention network (CTA-Net) for driver's activity recognition," in *Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 1279–1289.

[33] A. Iosifidis, E. Marami, A. Tefas, and I. Pitas, "Eating and drinking activity recognition based on discriminant analysis of fuzzy distances and activity volumes," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 2201–2204.

[34] K. Verma and B. Singh, "Deep multi-model fusion for human activity recognition using evolutionary algorithms," *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 7, no. 2, 2021.

[35] M. G. Moghaddam, A. A. N. Shirehjini, and S. Shirmohammadi, "A WiFi-based method for recognizing fine-grained multiple-subject human activities," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[36] A. Singh, S. Sandha, L. Garcia, and M. Srivastava, "RadHAR: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proc. of the ACM Workshop on Millimeter-wave Networks and Sensing Systems (mmNets)*, 2019, pp. 51–56.

[37] Y. Wang, H. Liu, K. Cui, A. Zhou, W. Li, and H. Ma, "m-Activity: Accurate and real-time human activity recognition via millimeter wave radar," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8298–8302.

[38] P. Gong, C. Wang, and L. Zhang, "MMPoint-GNN: Graph neural network with dynamic edges for human activity recognition through a millimeter-wave radar," in *Proc. of the International Joint Conference on Neural Networks (IJCNN)*, 2021, pp. 1–7.

[39] B. Settles, "Active learning literature survey," University of Wisconsin-Madison Department of Computer Sciences, Tech. Rep., 2009.

[40] D. Angluin, "Queries and concept learning," *Machine learning*, vol. 2, pp. 319–342, 1988.

[41] E. B. Baum and K. Lang, "Query learning can work poorly when a human oracle is used," in *Proc. of the International Joint Conference on Neural Networks (IJCNN)*, vol. 8. Beijing China, 1992, p. 8.

[42] P. Kumar and A. Gupta, "Active learning query strategies for classification, regression, and clustering: a survey," *Journal of Computer Science and Technology*, vol. 35, pp. 913–945, 2020.

[43] X. Li and Y. Guo, "Adaptive active learning for image classification," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 859–866.

[44] S. C. Hoi, R. Jin, and M. R. Lyu, "Large-scale text categorization by batch mode active learning," in *Proc. of the International Conference on World Wide Web (WWW)*, 2006, pp. 633–642.

[45] B. Longstaff, S. Reddy, and D. Estrin, "Improving activity classification for health applications on mobile devices using active and semi-supervised learning," in *Proc. of the International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE, 2010, pp. 1–7.

[46] R. Liu, T. Chen, and L. Huang, "Research on human activity recognition based on active learning," in *Proc. of the International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 1. IEEE, 2010, pp. 285–290.

[47] S. Bagaveyev and D. J. Cook, "Designing and evaluating active learning methods for activity recognition," in *Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2014, pp. 469–478.

[48] D. Cacciarelli and M. Kulahci, "Active learning for data streams: a survey," *Machine Learning*, vol. 113, no. 1, pp. 185–239, 2024.

[49] E. Lughofer, "On-line active learning: A new paradigm to improve practical useability of data stream modeling methods," *Information Sciences*, vol. 415, pp. 356–376, 2017.

[50] D. Sculley, "Online active learning methods for fast label-efficient spam filtering." in *Proc. of the International Conference on Email and Anti-Spam (CEAS)*, vol. 7, 2007, p. 143.

[51] T. Miu, P. Missier, and T. Plötz, "Bootstrapping personalised human activity recognition models using online active learning," in *Proc. of the IEEE International Conference on Computer and Information Technology (CIT)*. IEEE, 2015, pp. 1138–1147.

[52] G. K. Gudur, P. Sundaramoorthy, and V. Umaashankar, "ActiveHARNet: Towards on-device deep bayesian active learning for human activity recognition," in *Proc. of the International Workshop on Deep Learning for Mobile Systems and Applications (EMDL)*, 2019, pp. 7–12.

[53] A. Akbari and R. Jafari, "Personalizing activity recognition models through quantifying different types of uncertainty using wearable sensors," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 9, pp. 2530–2541, 2020.

[54] Texas Instrument. (2024) IWR1443BOOST evaluation board — TI.com. [Online]. Available: https://www.ti.com/tool/IWR1443BOOST

[55] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 4489–4497.

[56] S. An and U. Ogras, "MARS: mmWave-based assistive rehabilitation system for smart healthcare," *ACM Transactions on Embedded Computing Systems*, vol. 20, no. 5s, pp. 1–22, 2021.

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[58] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[59] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic mr image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 280–290, 2018.

[60] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[61] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[62] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[63] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[64] S. Shirmohammadi, M. H. Amiri, and H. Al Osman, "Uncertainty as a predictor of classification accuracy in machine leaning-assisted measurements," *IEEE Instrumentation & Measurement Magazine*, vol. 27, no. 7, October 2024.

[65] Joint Committee for Guides in Metrology (JCGM), *Guide to the Expression of Uncertainty in Measurement*. Aenor, 1993.

[66] S. Shirmohammadi, "Visualizing uncertainty in machine learning-assisted measurements," *IEEE Instrumentation & Measurement Magazine*, vol. 26, no. 7, pp. 20–27, October 2023.

[67] H. Al Osman and S. Shirmohammadi, "Machine learning in measurement part 2: Uncertainty quantification," *IEEE Instrumentation & Measurement Magazine*, vol. 24, no. 3, pp. 23–27, 2021.

[68] Y. Yang, Z. Ma, F. Nie, X. Chang, and A. G. Hauptmann, "Multi-class active learning by uncertainty sampling with diversity maximization," *International Journal of Computer Vision*, vol. 113, pp. 113–127, 2015.

[69] M. Fang, Y. Li, and T. Cohn, "Learning how to active learn: A deep reinforcement learning approach," *arXiv preprint arXiv:1708.02383*, 2017.

[70] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation," University of Texas at El Paso, Tech. Rep. UTEP-CS-18-09, 2018.

[71] S. Ranasinghe, F. Al Machot, and H. C. Mayr, "A review on applications of activity recognition systems with regard to performance and evaluation," *International Journal of Distributed Sensor Networks*, vol. 12, no. 8, p. 1550147716665520, 2016.

[72] E. De-La-Hoz-Franco, P. Ariza-Colpas, J. M. Quero, and M. Espinilla, "Sensor-based datasets for human activity recognition–a systematic review of literature," *IEEE Access*, vol. 6, pp. 59 192–59 210, 2018.

[73] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *Ieee Access*, vol. 8, pp. 83 791–83 820, 2020.

[74] L. Ahmed, K. Ahmad, N. Said, B. Qolomany, J. Qadir, and A. Al-Fuqaha, "Active learning based federated learning for waste and natural disaster image classification," *IEEE Access*, vol. 8, pp. 208 518–208 531, 2020.