WEARABLE SENSORS FOR PRECISION MEDICINE THROUGH PERSONALIZED, HOLISTIC AND CONTEXT-AWARE ANALYTICS

A Dissertation

by

ALI AKBARI

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Roozbeh Jafari
Committee Members,	Jean-Francois Chamberland
	Gerard L. Coté
	Kristen Maitland
	Bobak J. Mortazavi
Head of Department,	Mike McShane

December 2021

Major Subject: Biomedical Engineering

ABSTRACT

Recently, a growing percentage of healthcare takes place outside the traditional walls of clinical care and is tightly coupled with daily experiences. The field of remote and decentralized patient care and precision medicine, catalyzed by the COVID-19 pandemic, benefits from recent developments of wearable sensors and internet of things. Although these devices cannot replace clinical diagnosis, they are capable of complementing clinical care by predicting the onset of disorders that could trigger medical tests and assessing effectiveness of therapeutics. This research builds a suite of algorithms to facilitate deployment of remote health monitoring and precision medicine with the objective of disorder prediction and therapeutic assessment with wearable sensors in day-to-day life. To accomplish these objectives, we need to track minor changes in continuous physiological and behavioral data collected by wearables. However, identifying physiological changes that are not the result of external stimuli, such as daily activities, is challenging due to the imperfection of physiological sensing with wearables in uncontrolled environments. Hence, there is a need for identifying surrounding contexts, e.g., activities, to enable apple-to-apple comparison of physiological parameters. Moreover, inter-subject variabilities and personal baselines should be considered in the process of tracking changes in physiological parameters. Finally, large-scale data collection and processing with wearables is required to assess their effectiveness in real-world applications.

In this research, the problem of context-aware sensing is specifically addressed in the field of activity recognition. The proposed novel extension to existing motion-based methods enables understanding of users' environment through freely available nearables. Experimental results show that leveraging contextual information improves the detection of complex activities that are challenging to be detected by merely motion sensors. A personalization framework is also designed for activity recognition models with a novel uncertainty quantification algorithm to maximize personalization performance while minimizing users' burden. Lastly, to investigate feasibility of using wearables for disease monitoring, a large-scale real-world study with smartwatches and smart rings was conducted. A novel machine learning model is designed to identify pre-symptoms of the disease before diagnosis. The findings validate the potential of wearable sensing for continuous health monitoring in day-to-day life.

DEDICATION

To my loving and supporting family, especially my wife Hananeh Alambeigi and my parents.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Jafari, and my committee members, Dr. Mortazavi, Dr. Cote, Dr. Chamberland, and Dr. Maitland for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, special thanks to my wife for her patience, support, encouragement, and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Jafari [advisor] and Professors Cote and Maitland of the Department of Biomedical Engineering, Professor Chamberland of the Department of Electrical and Computer Engineering, and Professor Mortazavi of the Department of Computer Science and Engineering.

All work for the dissertation was completed independently by the student.

Funding Sources

Graduate study was supported by a research assistantship from Texas A&M University. This work was also made possible in part by the National Science Foundation, under grants CNS-1734039, EEC-1648451, and ECCS-1509063, the National Institute of Health, under grant 1R01EB028106 and the US Defense Threat Reduction Agency (DTRA). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

1. INTRODUCTION*

In recent times, a growing percentage of healthcare takes place outside the traditional walls of clinical care and is tightly coupled with the daily experiences [1]. For example, diabetes that costs the nation over $245 billion each year can be managed through a multitude of daily decisions and stresses that can be supported by continuous physiological monitoring augmented by physical and contextual measurements. Likewise, early diagnosis and management of chronic conditions such as infectious diseases, hypertension, and Parkinson disease as well as mental issues such as dementia and Alzheimer can be improved through continuous monitoring of patients during their normal daily life that also enables targeted and timely personalized interventions. Finally, any health and wellness-related improvements tied to behavior changes such as improved nutrition, exercise and sleep and the cessation of behaviors such as smoking require personalized interventions tailored and responsive to a person's environmental, social and psychological context.

Development of wearable sensing technology, emerging paradigms in mobile sensing, and recent advancements in internet of things (IoT) hold enormous opportunity for continuous measurement of physiological, physical, and contextual parameters.

---

*Reprinted with permission from "Data-driven Context Detection Leveraging Passively Sensed Nearables for Recognizing Complex Activities of Daily Living" by Ali Akbari, Reese Grimsley, and Roozbeh Jafari, 2021. in ACM Transactions on Computing for Healthcare, ACM Transactions on Computing for Healthcare (HEALTH), vol. 2, issue 2, pp. 12:1-12:22, ©2020 by ACM.

Reprinted with permission from "Personalizing Activity Recognition Models through Quantifying Different Types of Uncertainty using Wearable Sensors" by Ali Akbari and Roozbeh Jafari, 2020. in IEEE Transaction on Biomedical Engineering (TBME), vol. 67, issue 9, pp. 2530-2541, ©2020 by IEEE.

Physiological measurements could lead to early detection of infectious diseases through monitoring changes in vital signs during pre-symptomatic phase of the disease. Continuous physiological sensing could also assist physicians in monitoring progression of diseases such as hypertension and heart diseases. Likewise, physical and environmental sensing could assist physicians in monitoring progression of diseases such as dementia and hypertension. Physical and contextual measurements can also enable understanding of human activities of daily living (ADL) that is critical for managing dementia and for providing assistive living services to elderly and people with disabilities. Finally, timely intervention required in chronic situation such as diabetes, asthma, and epilepsy could benefit from continuous monitoring of patient’s health. These are only a few examples of how remote health monitoring and precision medicine could move the needle on healthcare outcomes and costs.

Wearable devices cannot replace the need for clinical diagnosis; however, they are capable of complementing clinical care since they can provide continuous measurement of physiological and behavioral parameters. As Figure 1.1 shows, these continuous measurements can be leveraged to predict onset of disorders that could subsequently trigger certain medical tests for clinical diagnosis, or they can initiate emergency help in critical cases such as stroke or fall. In addition, as depicted in Figure 1.1, wearables can monitor and assess effectiveness of therapeutics after the medical diagnosis. For example, in the case of cardiac rehabilitation, they can monitor the progress of the patient over time by monitoring heartrate and daily activities.

Figure 1.1 - Application of wearable devices in precision medicine

To accomplish the aforementioned objectives, one needs to identify and track minor changes in physiological and behavioral data that are collected continuously and often times over a long period with the wearable sensors. Detecting these small changes, however, could be extremely challenging due to noisy nature of wearables data collected in uncontrolled-environments. Moreover, identifying physiological changes associated with health disorders and distinguishing them from external stimuli, such as daily activities and circadian rhythm could be difficult when the devices are used in day-to-day life. Based on the aforementioned facts, there are several unmet needs that need to be addressed for precision medicine with wearable technology to be effective in day-to-day life. First and most important aspect is multi-modal sensing that provides measurements not only about physiological and physical status of the users but also about their surrounding context such as location, activities, and social interactions. This multimodal-sensing requires a powerful sensor platform augmented with infrastructures for continuous data collection and annotation as well as data fusion techniques. Second, data analytics algorithms are crucial for extracting actionable information from low-level sensors' data in order to understand high-level health-related information. These algorithms need to be personalized and adaptable to consider inter-subject variability of health parameters and to minimize users' burden. Third, putting the user in the loop is vital to 1) ensure high-fidelity data collection and 2) enable just-in-time adaptive interventions (JITAI). Fourth, large-scale data collection and processing is required to assess the effectiveness of precision medicine and remote health monitoring with wearable sensors in a real-world application.

Considering the aforementioned requirements and unmet needs, a successful remote health monitoring platform based on wearable sensors requires a combination of data analytics techniques that can 1) address inter-subject and inter-device variability, 2) analyze heterogenous data from multiple sources, and 3) monitor contextual information with minimum efforts demanded from the end-users in order to enable apple-to-apple comparison of physiological data. Seamless operation of the remote health monitoring platform and minimizing the need for end-user's intervention while maximizing the fidelity of data is crucial to keep users engaged because otherwise the users will lose their interest in using devices that are difficult to interact and/or maintain.

The research presented in this dissertation aims to tackle the aforementioned requirements in order to facilitate the real-world utilization of wearable-based remote health monitoring systems and enable a functional framework for disease monitoring. We tackle those requirements by creating algorithms that can be adapted to new-users and new-devices while minimizing the need for user intervention. We also design a context monitoring framework for understating environmental context without demand for deploying extra infrastructure and hardware. Finally, we build an end-to-end fully functional personalized and context-aware framework based on commercial off-the-shelf (COTS) wearable sensors for monitoring and predicting certain diseases. In the following we will get into the details of the aforementioned requirements for remote health monitoring and the objectives of this research.

1.1. Remote Health Monitoring Platform Requirements

Remote health monitoring platforms as shown in Figure 1.2 consist of sensing devices (i.e., personal and environmental sensors), data analytics, caregivers and the end-users as the most important component. Regarding each of those components, there are several critical factors and unmet gaps that need to be taken care of in order to build a successful remote health platform. Moreover, there are strong mutual interactions between these components in real-life scenarios. For example, the quality of the data analytics algorithms could potentially improve user experience and ultimately user compliance and similarly user intervention could enhance the performance of these algorithms. Likewise, intelligent data analytics techniques could compensate sensing hardware deficits on one hand, and on the flip side high-quality sensing hardware can boost up the performance of data analytics. In the following, we discuss the components shown in Figure 1.2 and introduce technical gaps associated with them.

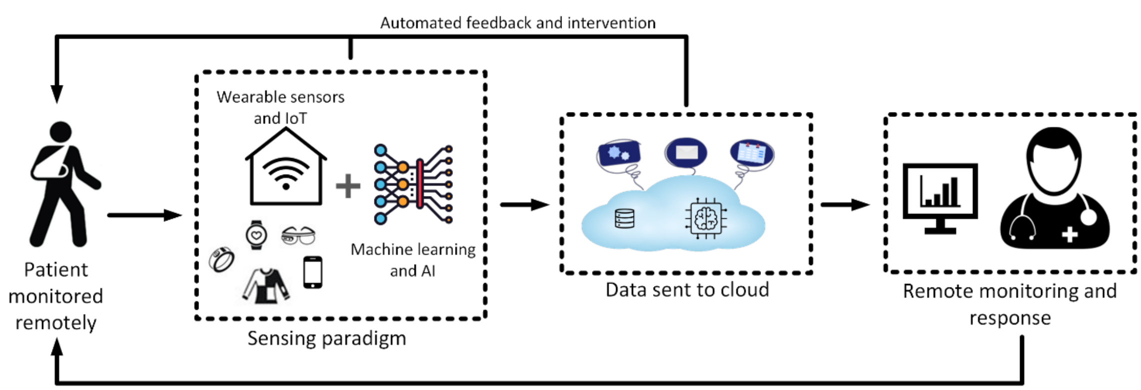


Figure 1.2 - Overview of remote health monitoring systems

1.1.1. Sensing Devices

The complex and nonlinear nature of healthcare challenges, which is continuously increasing, raises the need for holistic sensing of human population. This includes not only physiological status of an individual that requires multimodal measurement of different parameters, but also their physical activities, emotions, daily decisions, food intake, and biological rhythm. Equally important is individual's social and environmental context including but not limited to location, social interactions, and access to caregivers. To gain a holistic view on individual health status multitude of sensing devices is required.

Wearable devices provide a great opportunity for both physical and physiological sensing. Various sensing modalities such as photoplethysmogram (PPG), electrocardiogram (ECG), Bio-impedance (Bio-Z), temperature, and inertial measurement unit (IMU), which is a motion sensor, have been developed in different form factors including wristbands, rings, smart glasses, smart textiles, etc. [2]–[5]. Despite the advancements in wearable sensing technology, these devices to date suffer from two sources of noise including improper contact with the skin and motion artifact due to user's movements. Great amount of research has recently been devoted to design more accurate wearables [2]. In addition, many commercial wearable sensors with different levels of accuracy are available in the market [4], [6]. However, there is still a big demand for signal processing techniques that can remedy the effect of the noise along with robust machine learning algorithms that can address the noise problem [7]–[9].

In addition to the wearable technology, the advancement in IoT has enabled opportunity to capture environmental and social contexts surrounding the users. For example, smart homes equipped with motion cameras can track the activity of residents which is useful for tracking people with dementia or for fall detection for elderly [10], [11]. In light of rapid advancement of IoT and similar technologies, there are ubiquitous systems such as devices equipped with Bluetooth low energy (BLE) or Wi-Fi routers that can be repurposed for context monitoring (e.g., coarse of fine-grained localization) without the need for deploying specific sensors and infrastructure.

Multimodal sensing can be defined in different levels including sensing multitude of physiological parameters, combining physiological and personal contextual information such as physical activities and/or food intake, and finally combining personal measurements with environmental contextual information. Although there have been a long track of research on improving the performance of the individual sensors, there has been less focus on building frameworks that can combine multiple sources of knowledge for holistic health monitoring in real-world scenarios.

1.1.2. Data Analytics for Wearable-based Health Monitoring

Extracting actionable information from sensors' data requires a chain of signal processing and data analysis techniques. These techniques perform a wide range of operations including data cleaning and filtering, noise reduction, pattern recognition, high-level feature extraction and training machine learning models for various objectives including but not limited to detecting anomaly in physiological data, understanding human activities, detecting and predicting chronic diseases, predicting failures, and monitoring disease progress [10], [12]–[14]. Rigorous data analytics techniques have been introduced in the body of literature for physiological monitoring with wearable sensors [15]–[18]; however, there are certain gaps that need further certain gaps associated with the real-world deployment of wearable-based remote health monitoring systems that triggers the need for further research:

- The algorithms need to be context-aware in order to correctly interpret sensors' raw data. For example, an elevated heart rate (HR) during exercise should be considered normal, whereas increased HR during the sleep could indicate a physiological anomaly.
- The algorithms should be tailored to person's physiological baseline. Given high level of inter-subject variability in health data, the data analytics algorithms should be personalized for each user. It is critical though to accomplish personalization with minimum effort required from the end-users since it can significantly affect users' compliance.
- The algorithms should be adaptable to new sensors and devices by leveraging the training algorithms on a known wearable sensor, and expanding them to new sensors, eliminating the need for manual training of data analytics algorithms.
- The algorithms should be expandable and they should be able to work with multimodal sensors data. Fusing information from multiple sources is the key for these algorithms to achieve high accuracy health monitoring during day-to-day life.

The research presented in this dissertation, specifically tackles the above gaps to create a functional remote health monitoring platform using wearable sensors.

1.1.3. Interaction between Remote Health Monitoring Platform and End-users

The end-users lie in the heart of a remote health monitoring system. Building data analytics and machine learning algorithms to extract actionable information requires annotated and labeled data that can be provided by human users. The labels and annotation provided by an end-user is critical for retraining and personalization of these algorithms. In addition, remote health monitoring systems can adapt and improve over time by receiving feedback from the users. However, it is vital to minimize users' interaction required for data annotation since it can directly affect users' compliance. Herein, we leverage intelligent data annotation algorithms that try to minimize the need for users' annotation by choosing the most important part of data that needs users' input.

1.2. Research and Method Overview

In the research presented in this dissertation, we focus on three components of remote health monitoring including context-awareness, personalization for new users, and end-to-end multimodal sensing for disease prediction.

1.2.1. Research approach for Context-aware Activity Monitoring

Human activity recognition can enhance effectiveness of mobile health applications. For instance, in applications such as physical fitness monitoring, diet monitoring, and assisted living it is required to detect complex and high-level ADLs. More complex ADLs, such as cooking and housekeeping that require a combination of physical and cognitive efficiencies are critical for independent living. For instance, caregivers could view this information actionable if they would become aware that patients with dementia cook repeatedly or walk at home too often [19]. In Diabetes or obesity, it is vital to monitor for how long the person is sitting or relaxing, how often they engage in physical activities such as running and exercising, and where and how often they eat [20]. In another example, in the application of stress monitoring, it is critical to understand people's daily routine and its changes over time as the progression of certain mental disorders are associated with less engagement in social and job-related activities [10]. Wearable devices have enabled a convenient way to achieve this goal [21]. However, wearable sensors are usually not aware of the context that they are working in.

Current systems using wearables are not capable of understanding their surroundings, which limits their sensing capabilities. For instance, distinguishing certain activities such as attending a meeting or class, which have similar motion patterns but happen in different contexts, is challenging by merely using wearable motion sensors. Herein, we focus on understanding user's surroundings, i.e., environmental context, to enhance capability of wearables, with focus on detecting complex activities of daily living (ADL).

For an unobtrusive, scalable, and data-driven context detection we use freely available Bluetooth low energy (BLE) data broadcasted by devices in users' surrounding. It is also known as nearables [22]. These nearables could effectively be all devices equipped with Bluetooth such that they are passively detectable from the wireless signals they transmit. It is worth mentioning that, in this study we do not need to deploy any specific BLE device. Instead, we leverage the BLE data broadcasted by any device in the vicinity of the user and rely on the consistency of those devices over time to infer the context, which is further used to improve recognition of ADLs via a single wrist-worn motion sensor. To best of our knowledge, this is the first work that leverages freely available BLE information broadcasted by any device around the users for detection of such environmental context. Yet, free data is not necessarily good data. The reliable BLE devices will be observed consistently within a context, but many others will be inconsistent. Considering only a single outcome within the application space, i.e., a single ADL, we search for consistent patterns in the BLE devices scanned when this ADL was known to occur.

Consistently co-present BLE devices are the basis for extracting these context patterns. With these useful patterns identified, we look back to the entire application domain to find which patterns were important to which ADLs. Using a probabilistic framework, we create an a-posteriori probability for each pattern-ADL pair suggesting the likelihood that an ADL is being performed when a context is observed. In this way, the passively sensed context will directly reduce the set of possible outcomes, i.e., the search space, the model needs to consider. The proposed system can start with a general activity classifier trained only on motion sensors data and learn the context from BLEs in an incremental and personalized fashion. It should be noted that the context training must be personalized since the set of BLE devices that each user visits during their daily living is unique and different from other people. The set of ADLs that we are targeting in here, are unique in a sense that it is very challenging to detect such complex and high-level activities with a single wrist-worn motions sensor in uncontrolled environments. It is important to note that the proposed context identification technique is not limited to ADL recognition application.

1.2.2. Research approach for Personalization

Data gathered by wearable sensors could be analyzed by rigorous machine learning models to build a system to recognize these ADLs [23], [24]. However, a given model trained on a certain user may not generalize well to new users due to variation in how people perform specific activities [25]. Therefore, it is necessary to personalize underlying machine learning models to new users.

We propose an ADL recognition system with a personalization capability. We leverage both active learning and unsupervised learning methods to facilitate the retraining process. The active learning is a supervised process in which the most critical samples identified by the machine learning model and their labels are solicited from the end-user or an oracle in general; these labels are then used to retrain or fine tune the machine learning model [26]. For the active learning component, we leverage the uncertainty of the model on its decision to identify the critical samples where annotations should be requested and used for retraining. We propose a unified Bayesian deep learning framework to model and quantify data and model uncertainties (aleatoric and epistemic) by considering stochasticity on both the parameters of the neural network and the latent variables served as features. Our proposed method extracts the features from the time series automatically through an unsupervised deep learning framework and learns their posterior distribution given the input data through a variational autoencoder (VAE) based model. To account for randomness of the model weights, we utilize the Dropout Bayesian network dropped [27]. To the best of our knowledge, this is the first work that proposes to distinguish different types of uncertainty for active learning via wearable sensors. Moreover, the proposed framework has the ability to learn from unlabeled data through the autoencoder framework, which is leveraged to retrain the parameters of the feature extraction neural network.

1.2.3. Research approach for Wearable-based Health Monitoring

Despite the ability of the wearables to capture high-resolution, continuous and real-time actionable physiological information, a number of unmet gaps and unaddressed challenges remain in identifying diseases such as infection with wearables: 1) inter- and intra-subject variability of physiological responses to the disease, 2) sensor noise and artifacts, and 3) other external stimuli that can lead to physiological changes similar to those caused by the disease. 4) access to commercial wearable devices and collation of real-time data. We conducted a study that covers the largest number of physiological observations integrated across a dual platform of commercial off-the-shelf (COTS) wearables including smart ring and a smartwatch.

The details of this study cannot be revealed per Texas A&M Engineering Experiment Station (TEES) instructions due to confidentiality concerns.

1.3. Research Contributions

This research addresses context monitoring and personalization as critical issues in the field of remote health monitoring and designs an end-to-end wearable sensing system for disease detection. In this section, we introduce contributions of this research in each of those areas.

1.3.1. Context-aware Activity Monitoring

For an unobtrusive, scalable, and data-driven context detection we use freely available Bluetooth low energy (BLE) data broadcasted by devices in users' surrounding. Detecting location and interaction-related context through Bluetooth low energy nearables and use that to narrow down search space. It is worth mentioning that, in this study we do not need to deploy any specific BLE device. Instead, we leverage the BLE data broadcasted by any device in the vicinity of the user and rely on the consistency of those devices over time to infer the context, which is further used to improve recognition of ADLs via a single wrist-worn motion sensor. To best of our knowledge, this is the first work that leverages freely available BLE information broadcasted by any device around the users for detection of such environmental context. Contributions of this research are as follows:

- An unsupervised pattern extraction method for detecting contextual patterns in static and mobile nearables.
- A probabilistic model leveraging the mutual relationship of context and the application domain such that the application helps build context and the resulting context improves performance of the application.
- Search space reduction based on the present context, permitting the context-aware application to use smaller models with higher accuracy rates.

- A case study on recognizing complex ADLs in the wild, built on 100+ days worth of real-world data collected with smartwatches, to demonstrate the effectiveness of our methods without constraining the environment.

1.3.2. Personalized Health Monitoring

In this research, we propose a personalization model applied to the application of ADL recognition with wearable motion sensors. We leverage supervised active learning and unsupervised learning methods to facilitate the personalization process. We designed a new architecture of deep neural networks to quantify data and model-dependent uncertainties for wearable motion sensors. We then designed a technique to choose the most critical samples based on the model uncertainty to be labeled by the user for model retraining.

The contributions of this study are as follows:

- We design a unified framework for automatic feature extraction, classification, and estimation of uncertainty of the classifier to incorporate active learning in human activity recognition.
- We propose a method for quantifying both mode and data uncertainties and demonstrate how differentiating them is essential to achieve a more effective active learning.
- We design a new framework for deep learning for activity recognition that incorporates uncertainty quantification and unsupervised retraining.

- We design an active learning technique leveraging quantified uncertainties with the ability to consider limited capacity of the users to respond to external prompts and solicitation of labels.
- We show how the proposed personalization framework can adapt itself to new users more effectively and quickly compared to the existing paradigms.

1.3.3. Wearable-based Disease Prediction

The details of this study cannot be revealed per Texas A&M Engineering Experiment Station (TEES) instructions due to confidentiality concerns.

2. PRELIMINARIES*

This section expands on some background topics to better contextualize the research described in subsequent sections. We first introduce common wearable sensors for physiological and physical monitoring required for remote health and wellness monitoring. We then define the notion of context in the application of remote health and wellness monitoring. It is followed by discussion about human activity recognition as a critical context for health monitoring with wearables. Afterwards, the concept of inter-subject variability is defined as one of the most important issues in health analytics. Finally, we explain disease prediction through continuous health monitoring via wearables.

2.1. Wearable Sensors for Remote Health Monitoring

Advances in wearable sensing has enabled opportunity to sense different health-related parameters outside of clinics. Wearable sensors are available in different form factors such as smartwatches, smart rings, smart phones, wrist-band sensors, sports shoes, smart glasses, and sensors embedded in clothing. Various physiological and physical parameters can be measured by these wearables from different body locations.

Given the importance of these systems in modern healthcare, many companies have commercialized different types of wearable sensors with diverse sensing capabilities. These COTS wearables provide opportunity for large-scale data collections

*Reprinted with permission from "Data-driven Context Detection Leveraging Passively Sensed Nearables for Recognizing Complex Activities of Daily Living" by Ali Akbari, Reese Grimsley, and Roozbeh Jafari, 2021. in ACM Transactions on Computing for Healthcare, ACM Transactions on Computing for Healthcare (HEALTH), vol. 2, issue 2, pp. 12:1-12:22, ©2020 by ACM.

although they provide researchers with less flexibility compared to custom built research-grade wearable sensors. For example, Garmin smartwatches by Garmin Ltd. (Olathe, Kansas, USA) continuously estimates physiological parameters in a round-the-clock fashion as it is worn. The set of parameters and features collected by the device includes heartrate (HR), respiration rate (RR), pulse oximetry, motion intensity, and sleep reports. Oura rings by Oura Ltd. (Oulu, Finland) provides HR, inter-beat interval (IBI), root mean square of successive differences (RMSSD, *i.e.*, a measure of HRV), RR, temperature, coarse-grained activity, and sleep reports. These are only a few examples of many wearable sensors available in the market.

Physiological sensors worn on different body locations can provide measurements of HR, HRV, RR, pulse oximetry, body temperature, and blood pressure that are required for diagnosis and disease monitoring. These sensors usually work based on photoplethysmography (PPG), Bio-impedance (Bio-Z) or in rare cases electrocardiogram (ECG) signals since ECG is challenging to be acquired from wearable sensors.

Other sensors that can measure physical activity and sleep are also important as they provide important contextual information required for remote health monitoring. Inertial measurement units (IMUs) are motion sensors that can measure acceleration and angular velocity. These sensors are embedded in almost all wearable devices as well as smart phones. Using motion sensors one can monitor activities of daily living or coarse-grained activity level which provides important context not only about the user by also about the quality of the data of the physiological sensors. The quality of the data captured by warble physiological sensors degrades drastically due to motion artifact when the user engages in intense activities.

2.2. Contextual Measurements for Remote Health Monitoring

The context is a broad term that includes any additional information that could help better understand a specific situation. Depending on the main objective the main and the contextual measurements can change. For example, *activity* can serve as context for health monitoring, but in the application of activity recognition the *activity* is the main measurement and location of the user (e.g., home versus at work) and their social interaction can serve as contextual information. Contextual knowledge helps identify the most probable outcomes within an application domain based on the user’s history within that context.

Working with wearable sensors in remote health monitoring applications, the context plays an important role for accurate understanding of the physiological data. In this application, as opposed to clinical data collection, the data is collected while the user can engage in their normal daily activities. In addition to the daily activities, other external stimuli such as stress, daily decisions, food and alcohol consumption can affect the physiological parameters measured by wearables. Therefore, it is critical for wearable-based remote health monitoring systems to be aware of the situation at which the data is collected. Moreover, the knowledge of the contextual information such as location and time is critical for intervention and decision-making systems to make the right decision at the right time. In the specific application of remote health monitoring, the most important contextual information includes activity type and intensity, location, sleep, social interaction, time, and food intake.

2.3. Activities of Daily Living Recognition

Monitoring activities of daily living (ADL) provides vital contextual information for various applications including health monitoring, assisted living, security and surveillance, and mobile services [28], [29]. For instance, in applications such as physical fitness monitoring, diet monitoring, assisted living, and remote health monitoring it is required to detect complex and high-level activities of daily living (ADL). More complex ADLs, such as cooking and housekeeping that require a combination of physical and cognitive efficiencies are critical for independent living. Therefore, there is a need to monitor those activities, especially among elderly and people with certain diseases and/or disabilities. For instance, caregivers could view this information actionable if they would become aware that patients with dementia cook repeatedly or walk at home too often [19]. In Diabetes or obesity, it is vital to monitor for how long the person is sitting or relaxing, how often they engage in physical activities such as running and exercising, and where and how often they eat [20]. In another example, in the application of stress monitoring, it is critical to understand people's daily routine and its changes over time as the progression of certain mental disorders are associated with less engagement in social and job-related activities [10]. Wearable devices have enabled a convenient way to achieve this goal [21].

Rigorous machine learning and deep learning algorithms have been designed in conjunction with wearable motion sensors for automatic detection of various ADLs [5],

[13], [19], [30], [31].For the goal of ADL recognition with wearables, there is usually a labeled dataset available collected with motion sensors, which is used to train machine learning algorithms, and it is collected under a certain sensing condition including the type and number of the sensors, sensor quality, sensor placement, subjects, and activities to be recognized. All these conditions are possible to be changed in real life.

Researchers have shown that different type of sensors are suitable for recognizing various type of activities. To recognize ambulatory movements such as walking, running, sitting, standing, and climbing stairs using the data collected by accelerometers placed on the body have achieved reasonable accuracy [32], [33]. Recent research has proved the ability of a smartphone to be used for activity recognition goal. Researchers have used phones to recognize gestures and basic activities [34], [35]. For high-level activities that are not as easily distinguishable by body sensors alone, researchers have used environmental sensors to capture interaction between users and different objects [36]–[38].

2.4. Inter-subject Variability

Despite the fact that a wearable device will always aim to measure the same physiological behavior, sensor capabilities will vary between users due to differences in demographics, body composition, health state, and physical attributes. Consequently, performance of the machine learning models trained on these data will vary for a generalized model and will dramatically suffer when it attempts to analyze wearable signals for a new user whose data were not included in the training set. Hence, there is a need for retraining the models for different users.

User-specific variations exist in all physiological and physical measurements. For example, in Figure 2.1 we show the data Bio-Z and ECG signals collected by wearable sensors from 10 participants in steady state. First, we discuss the variations of ECG that are present in this data. Figure 2.1-a shows sample ECG instances collected by wearables that exist for each subject. Here, we show that each subject possesses unique characteristics that impact both the scale and morphology of their typical cardiac cycle characteristics, which are due to both the unique cardiac behaviors that exist within a

(a)

subject and also by the quality of sensing during data collection.

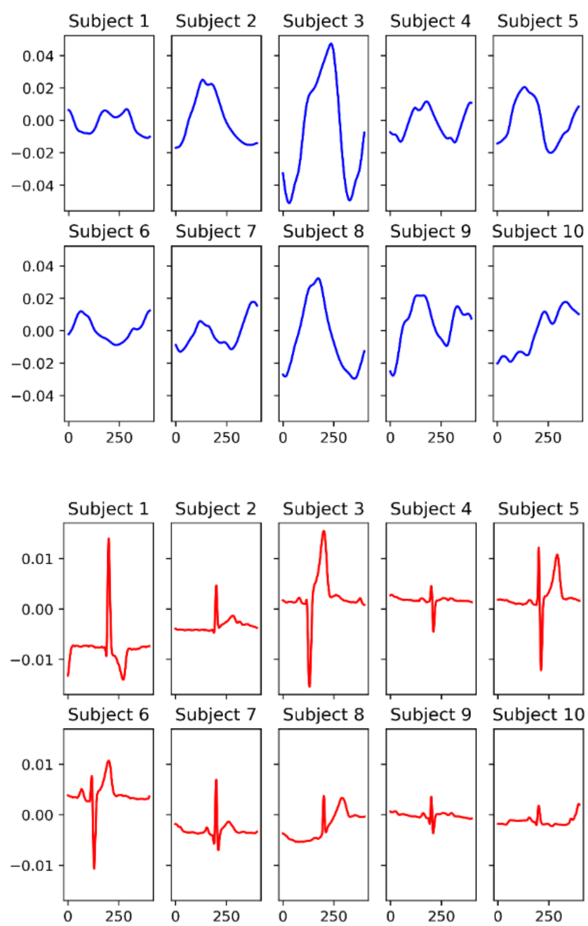


Figure 2.1 - Inter-subject variability of physiological measurements

Now, to consider the variation in Bio-Z signals, we refer to Figure 2.1. We also observe that although general morphology is similar, each signal per subject will vary with respect to amplitude, phase, or delay. Furthermore, similar to the ECG instances, these will also vary from beat-to-beat. This distinctness of Bio-Z signals is best exemplified by the fact that Subjects 5 and 6 have relatively similar ECG types, but their corresponding Bio-Z signals are quite different with respect to scale. As another example, we observe that despite the fact that Subjects 6 and 7 have relatively similar Bio-Z signals, their corresponding ECG types are very different. These types of challenging cases will be best handled with personalized models.

2.5. Disease Monitoring with Wearables

Wearable sensors can be used for continuous monitoring of several diseases including but not limited to diabetes, asthma, Parkinson, dementia, stress, seizure, hypertension, heart failure, and infection [18], [39]–[44]. Here, we focus on explaining the potential of wearable sensing in detecting infection, specifically COVID-19, that its recent pandemic in 2020-2021 has led to millions of deaths worldwide.

Infectious diseases are one of the leading causes of death worldwide [45], [46]. COVID-19 caused by the SARS-CoV-2 virus, affecting the upper and lower respiratory tract [47], [48], led to more than 2,360,000 deaths between January 2020 and February 20211. SARS-CoV-2 is highly contagious and given its long incubation period (2-14 days [49], [50]) it can be easily spread by an infected person prior to observing any symptoms, and in some cases, the infected individual remains asymptomatic. Due to this silent transmission period, the spread of the virus became exponential, and the number of positive cases passed 107 million in one year1. Although it may take a number of days to observe the symptoms, there is an opportunity to identify pre-symptomatic signs, through small physiological changes, e.g., a slight elevation in heart rate (HR) or reduction in certain heart rate variability (HRV) measures could be observed [51]. If an impending infection could be identified pre-symptomatically in a timely manner, potential guidance on self-isolation could reduce the exponential spread of the virus, reduce the burden of contact-tracing on the public health workforce, and provide opportunities for early intervention in high-risk populations.

Infection can cause a chain of symptomatic reactions, negatively impacting vital organs. Although levels of the infection and the viral load might vary for each individual, reports indicate a consensus in the symptoms associated with COVID-19 [52]. The SARS-CoV-2 virus, for example, can damage the cardiovascular and respiratory system in multiple ways [53]. The fever increases evaporative fluid losses which cause the heart to beat faster as blood volume contracts. In addition, the infection inflames the lungs [54], which decreases the oxygenation of the blood. This introduces an extra burden for the heart that is responsible for providing sufficient oxygen to the circulation. These chains of reactions cause elevations in both HR and RR. All these parameters can be continuously measured by wearable devices. Therefore, wearable sensors that can measure these parameters are a good candidate for detecting physiological anomalies associated with infection [55].

Early warning of exposure to infectious diseases, such as COVID-19, is essential to preventing their spread1. This is possible through the continuous analysis of physiological observations captured with wearable devices [56].

3. RELATED WORK*

3.1. Activity monitoring

There is an abundance of prior research in the field of ADL recognition. Different machine learning algorithms have been utilized to perform human activity recognition with varying types of wearable and environmental sensors [57]–[65]. Although promising results have been achieved in these works, two main challenges have not been addressed well: 1) Most of the previous works have focused on hand-crafted features that are useful for simple or low-level human activities such as sitting or walking. While detecting complex ADL needs more complicated features and patterns to be recognized which is not a simple task using the manual features. 2) Most of these systems are designed based on a fixed set of sensors, users, and activities. They have not considered different variations in the distribution of data due to the user or sensor changes, or they have considered just one type of variation. Thus, lacking a robust framework that can address different kinds of variations and can adapt itself to new situations is significant.

Researchers have shown that different type of sensors is suitable for recognizing different type of activities. To recognize ambulatory movements such as walking, running, sitting, standing, and climbing stairs using the data collected by accelerometers

---

*Reprinted with permission from "Data-driven Context Detection Leveraging Passively Sensed Nearables for Recognizing Complex Activities of Daily Living" by Ali Akbari, Reese Grimsley, and Roozbeh Jafari, 2021. in ACM Transactions on Computing for Healthcare, ACM Transactions on Computing for Healthcare (HEALTH), vol. 2, issue 2, pp. 12:1-12:22, ©2020 by ACM.

Reprinted with permission from "Personalizing Activity Recognition Models through Quantifying Different Types of Uncertainty using Wearable Sensors" by Ali Akbari and Roozbeh Jafari, 2020. in IEEE Transaction on Biomedical Engineering (TBME), vol. 67, issue 9, pp. 2530-2541, ©2020 by IEEE.

placed on the body have achieved reasonable accuracy [32], [33]. Recent research has proved the ability of a smartphone to be used for activity recognition goal. Researchers have used phones to recognize gestures and basic activities [34], [35]. For high-level activities that are not as easily distinguishable by body sensors alone, researchers have used environmental sensors to capture interaction between users and different objects [36]–[38]. Accelerometers and RFID tags are used in these works. Other researchers utilized other environmental sensors including motion detectors and door contact sensors to recognize ADL [66]–[68].

To address the challenge of extracting informative features for complex activities, deep learning has been recently used. Several prior investigations used data from inertial measurement units (IMU) or motion sensors to automatically extract features and detect human gestures and activities using convolutional neural networks (CNN). In one study, axes of the IMUs were treated like different channels of an image provided to a three-layer CNN and outperformed traditional machine learning algorithms with respect to the accuracy [23]. A CNN based approach that captures temporal dependency of a signal was proposed to perform activity recognition with a cell phone's accelerometer [69]. These investigations used 1D convolutional units which can only consider temporal patterns of the signal but not the relationship between different axes of a sensor. To overcome this problem, data from an accelerometer and a gyroscope were converted into images to be used in a CNN for detecting basic human activities, which outperformed state-of-the-art in terms of both recognition accuracy and computational efficiency [70]. Experimental results showed that CNN achieved significant speed-up in computing and detecting the final class while also improving overall classification accuracy compared to the state-of-the-art models such as support vector machine (SVM) [71].

Deep learning methods that use more than just CNN for ADL recognition have also been developed. A combination of CNN and recurrent neural network (RNN) was used to detect modes of locomotion as well as hand gestures [72]. In this article, four convolutional layers were used to extract features of the signal and two recurrent layers were used to model temporal behavior. The article showed that using RNN in addition to CNN could slightly increase accuracy. However, the authors did not evaluate how adding RNN increases the computational complexity of the model, and they did not explain the tradeoff between improving accuracy vs. increasing computational time. Another similar deep neural network with the ability to detect concurrent activities was introduced that offered a new type of layer called a coding layer to allow the deep learning model to simultaneously make predictions for all activities in progress [73]. A multimodal multi-stream deep learning framework to tackle the egocentric activity recognition problem was proposed [74]. This model used the data from both the video and wearable sensors including an accelerometer and a gyroscope. CNN was used to extract features from video and RNN were used to model wearable sensors' signal, and finally, a two-layer fusion approach was proposed to combine different modalities and detect human activities. Although all these studies showed the ability of deep neural network to use different sensors and extract the features automatically as well as its superiority over traditional methods, none focused on the robustness aspects of the systems with respect to the heterogeneity of sensors and human activities. In fact, the investigators proposed a fixed model trained for specific sensor and users and did not investigate the effect of changes on their algorithms.

Multiple prior investigations focused on transfer learning for activity recognition in order to address the challenge of sensor or user changes. A CNN-based model for ADL recognition was proposed that was able to adapt itself to each individual with a small amount of additional data. This model inserted a special layer with a small number of free parameters between each of two CNN layers and estimated the free parameters using a small amount of data from each new subject [75]. This technique needs to collect at least a few labeled data of the new setup to retrain the deep network that could be burdensome for the users and is not truly seamless. There are many other studies that enabled the concept of transfer learning in this area using traditional machine learning approaches, but not deep learning.

A transfer learning technique was proposed based on unsupervised clustering of new data in order to personalize an ADL model for new users [76]. The activity recognition knowledge of one IMU sensor was transferred to a newly worn IMU in another study. In this work, a label propagation technique was utilized to refine the labels using both old (supervised) and new sensor (unsupervised) data and then the modified labels were used for training a new model for the new sensor from scratch [77]. The authors showed that their technique improved the performance compared to using the labels that are generated by the old sensor for training the new one. An unsupervised similarity-based method for integrating new sensors was proposed by a group of researchers. They established a similarity metric between data points in the space of new sensor by building a similarity graph using the features from the old and new sensor. This was used for moving the labels assigned by the original classifier if they are not consistent with similarity considerations, and the adjusted labeled data were used to train a new classifier [78]. However, with these techniques, a new model needs to be trained from scratch using the generated labels. The classification step was formulated under covariate shift, which assumes that the distributions of input points change between the training and testing phases, to adapt classification model to the data of new users by estimating the class-posterior probability [79]. However, this technique relies on the assumption that the class-posterior probability does not change in between training and test phases that might not be correct in the case of changing the sensors. Challenges of an opportunistic ADL recognition system to train a newly added sensor with the available sensing devices to recognize activities at runtime were investigated [80]. All these attempts used handcrafted features that affect both performance and the computational efficiency of the techniques as well as re-deploying the algorithms for new ADL recognition scenarios.

While most of the aforementioned studies evaluated the performance of their algorithms under controlled experiments in which the participants were asked to perform only the desired activities, a couple of researchers tried to apply ADL recognition algorithms on real-life scenarios. A smart home project running in Washington State University collected the data of several residents living in different smart homes equipped with motion, temperature, and water/gas usage sensors. Using this dataset the researchers recognized and tracked functional activities that people perform in their own homes and everyday settings [81]. In this paper, they looked at approaches to perform real-time recognition of ADL with focus on recognizing independent and joint activities among multiple residents. However, they did not take into account the challenge of sensor or user variations. Moreover, they could not achieve high accuracy (66% on average) because of the limited sensors that were used. Simple environmental sensors were the only type of sensor used in this study while adding wearables could significantly increase the performance. Using the same smart home data, researchers tried to recognize the activities that do not belong to a predefined class [82]. They used hidden Markov models (HMM) to recognize the activities. However, this challenge is addressed with binary sensors, while it is more difficult to solve such a problem using continuous sensors. There are other works that tried to design ADL recognition for smart homes and tested the system in real life scenarios [83]–[87]. However, all of them assumed that all sensors are always available. Thus, lack of a robust framework that can leverage a considerable set of heterogeneous sensors for ADL recognition and can adapt itself to different sensor and user variations hinders developing context-aware applications in smart healthcare.

3.2. Contextual Information

Context-awareness in mobile computing is considered an integral component of the ubiquitous computing paradigm, i.e., the third-wave of computing [88], [89]. The idea that an understanding of the current situation will help computational models is by no means a new one. There has been much work into the structure of context and its usage such that it can be applied to various applications through a centralized framework [90], [91]. Often, these studies build context or suggest building context by merging the output of heterogeneous sensors, e.g., GPS, proximity sensors, and microphone into some structured yet flexible format, such as an ontology, for other applications to use [19], [92]. Since we fixate on the user-centric context, we consider what the user's mobile devices are capable of sensing.

A common approach to context detection is to use technology embedded in the infrastructure of the interesting contexts, such as specific rooms in the home or the user's desk and common meeting rooms at work [93], [94]; these may be considered "logical" or "semantic" locations. Prior investigations have studied using Bluetooth beacons/metadata to detect location [95], social interaction [96], and activity recognition [19], in both supervised and unsupervised manners. Bluetooth-based localization was used to push mobile advertisements to user's cellphones [95]. This system performs localization based on the presence of a known Bluetooth device in the vicinity of a fixed Bluetooth sensor. Therefore, this system needs to know user's Bluetooth address and phone number as well as specific location of the Bluetooth sensor. Another study proposed a probabilistic matrix factorization method for identifying both time characteristic and people involved in a person's social circle using Bluetooth scans [96]. Although this is an unsupervised method that reveals the social circles in which a person is involved, this system assumes that all the Bluetooth packets belong to the cellphones of the study participants. In other words, it ignores the BLE packets coming from stationary devices such as a smart TV. Another study on ADL recognition placed one known BLE beacon into each room, and used the RSSI from each beacon as an input feature for classification; the classifier thus indirectly uses context [19], [93]. Yet if the location and identity of each beacon is a priori knowledge, it becomes trivial to determine the user's logical location [19]. This paper used BLE beacons placed in known locations to identify context, and using that it could extensively improve the current activity recognition systems by detecting complex and fine-grained ADLs, which are challenging to be detected via only motion sensors. They explored activities such as sitting and eating, sitting on sofa vs. bed, standing and using sink vs. talking, walking outdoor to indoor vs. indoor to outdoor, and lying on bed vs. sofa. They showed that by adding location information to the motion data, the accuracy improved from ~78% to ~85%. However, the primary issue here is that the user must deploy infrastructure and provide corresponding information to the context detection system. In all these studies, some prior knowledge about the type of devices and their owners is assumed to be accessible. In contrast our proposed system does not leverage any prior knowledge about the type of BLE devices and their specific location. Moreover, it does not need one to deploy any device in particular locations.

Sensing absolute location via a GPS sensor is another option [97]–[99]. The location captured by GPS has been used to narrow down the search space for classifying the type of food from images [91]. However, several issues preclude GPS from being a good context sensor: 1) it only detects location and is not capable of detecting user interactions as well as other environmental information, 2) it requires line of sight and thus, cannot be used indoors, 3) absolute location has a privacy stigma, and 4) it is not power efficient [100].

Another approach to coarse-grained localization for context detection is through unsupervised fingerprinting of Wi-Fi Access Points [101], [102]; this does not require the user to place infrastructure in their environment, but is only applicable to locational context. These studies facilitate automatic detection of context on passively sensed data. Yet the problem is that they cannot identify social aspects of context well, if at all. While it might be possible to determine social context by passively scanning all Wi-Fi packets, this would require unattainable permissions on commercial devices or specialized hardware [103]. In addition, these techniques need recalibration for each new location that puts extra burden on users. Prior works have attempted to use Bluetooth devices on their own, but for constrained scenarios. These studies often perform statistical feature extraction on Bluetooth scans for use in a classifier [94], [104]. However, just like studies which used known devices as features, whether by presence alone or RSSI value, this will at best lead to an indirect learning of context within the model. These techniques may place high importance on specific features whose values depend on context. Over time, the context may evolve, meaning the model may lose its effectiveness over time. To our knowledge, few studies in general exist which facilitate purely data-driven detection of context. Further, no such studies exist for automatically detecting both locational and social contexts using passively observed nearables. We note that our algorithm does not explicitly derive social interactions from BLE beacons. In fact, since our proposed algorithm considers all regularly observed BLEs from any devices around the user for context modeling, the identified contexts could potentially be related to interaction between users (e.g., when the BLE comes from a friend's device), actual location (e.g., when BLE belongs to a stationary device such as a smart TV in the living room), or objects and devices (e.g., when BLE belongs to user's vehicle). However, our algorithm is blind to the exact type of social interaction. In fact, the main advantage of the proposed method over existing BLE-based methods is learning the context from all BLE packets possible instead of limiting the system to a set of pre-known BLE devices or placing BLE beacons in particular locations.

3.3. Personalization

ADL recognition has become an important component of context-aware systems in the last decade [105]. This contextual information provides valuable knowledge to better interpret biomedical signals, diagnose with more confidence, intervene and treat more efficiently, detect emergency situations, make proper decisions based on the context, assist people with chronic disabilities to perform their daily activities independently, and monitor patients more precisely [106], [107]. Accurate and robust recognition of human activities requires gathering massive amount of labeled data to train powerful machine learning models [58], [59]. Different machine learning algorithms have been utilized to perform human activity recognition with various types of wearable and environmental sensors [64], [108], [109]. All these investigations have focused on hand-crafted features that are useful for simple or low-level human activities such as sitting or walking. However, detecting complex ADL needs more complicated features and patterns which cannot be executed easily with manually handcrafted features.

To address the challenge of extracting informative features for complex activities, deep learning has been recently used [13], [21], [73]. Several prior investigations have used data from inertial measurement units (IMU) or motion sensors to automatically extract features by leveraging convolutional neural networks (CNN) for detecting human gestures and activities [23], [69], [71], [72]. In all these efforts, the principal aim has been to create generalized models for all users. The performance of these systems, however, degrade significantly when they are used for a new user who performs the activities differently form the set of training subjects [110].

To address the problem of inter-subject variability in human activities, the researchers have designed personalization techniques [111], [112]. An unsupervised retraining technique based on an ensemble model was designed for personalizing activity recognition models [113]. In this study, the model trained on the data of old subjects was used to assign labels to the data of a new subject, and then these data were used to update the ensemble model. However, it has been shown that with using merely unsupervised data it is hard to achieve high accuracy in personalization [114]. Active learning is a widely used technique in this area that seeks to identify important samples and interacts with the user to acquire labels for those samples. Those labeled data is then used to fully or partially retrain the classification models [115], [116]. Active learning techniques based on the entropy and the random forest committee of classifiers were investigated in prior studies. The entropy-based methods typically select data samples to solicit a label on the basis of the highest information gain. In random forest-based methods, a forest of 100 trees was trained and the disagreement between the output of those trees for each sample was used to select the samples to obtain label for [117].

Uncertainty of the classifier is a widely used criterion to identify vital samples that require label [118], [119]. An active learning technique based on the uncertainty of the SVM model was developed for detecting exercise activities [118]. A logistic regressor was cascaded to the SVM's output to estimate the SVM classifier's uncertainty. This work, however, does not take into account user's burden associated with soliciting labels, does not consider different types of uncertainty, and it uses handcrafted features. A prior study provided an active learning method under constrained query budget [120]. All these works only rely on the labeled data solicited from the user and they do not take advantage of abundance of unlabeled data available in the ADL recognition platforms. A personalization method integrating both unsupervised and supervised retraining was developed using model uncertainty for active learning [111]. This study quantifies the uncertainty using the similarity between the new data and the training data in dense regions. However, neither this work nor other uncertainty-based active learning methods consider the source of uncertainty when designing the active learning technique.

There are mainly two types of uncertainties present in sensor data. Aleatoric or data-dependent uncertainty is related to noisy sensor measurements and cannot be mitigated by increasing the training data, whereas the epistemic or model-dependent uncertainty is due to lack of enough training data and it can be alleviated by enhancing the training set. Therefore, considering the source of uncertainty is critical to identify the vital samples that can enhance the performance of personalization. In deep learning, often Softmax functions is used at the end of the pipeline to measure model uncertainty; however, it has been shown that it does not always capture model uncertainty [27], [121]. A framework has been designed to learn mappings from input data to aleatoric uncertainty in image recognition [27]. The authors compose these together with epistemic uncertainty approximations. However, this model learns aleatoric uncertainty based on the assumption that in the training phase they have access to examples of the disturbed data (e.g., highly textured input images or far objects) to train the system, which is not always the case in ADL recognition. In ADL recognition, the disturbance in sensor measurements could be the result of electrical noises or sensor movements with respect to the body, where examples of such a noisy data may not be available during the training phase. Therefore, the system should be able to estimate the aleatoric uncertainty with no need to observe the examples in the training phase. To address the aforementioned issues and to maximize the effectiveness of personalization while minimizing the user interaction, we propose a deep learning assisted method for measuring different types of uncertainties.

3.4. Infectious Disease Detection with Wearables

Various wearable sensors have been used for predicting and/or detecting infection through monitoring changes in vital signs associated with various infections [122]. Several wearable-based systems tried to address detection and prediction of the novel coronavirus (SARS-CoV-2) that led to a pandemic (2020-2021) due to its highly contagious nature and resulted major disruption to public health and a huge loss of human life [123].

A wearable-based framework was designed to facilitate daily and pervasive detection of SARS-CoV-2/COVID-19 by combining off-the-shelf wearables with deep neural networks [123]. This study used synthetic data generation to alleviate the need for large datasets. They leveraged a fully dense network with no convolutional layers for feature extraction purposes. The notion of context and noise in wearable data was not addressed in this paper. Another study designed a smartphone app to collect smartwatch data as well as self-reported symptoms for infection detection [124]. A novel score based on RHR normalized by step count measured by smartwatches was proposed for detecting infection [51]. A CNN-based model was designed based on the data collected with Fitbit devices for detecting infection using RHR, HRV, and RR [125]. Based on self-reported symptoms alone, they could also predict the need for hospitalization.

Although prior studies have conducted compelling analysis on the physiological changes associated with COVID-19, only some pursued the prognosis of infection in the pre-symptomatic phase [51]. Furthermore, most of these investigations considered a limited set of physiological parameters and a single wearable device or provided limited evaluation of their proposed framework. We conducted a study that covers the largest number of physiological observations integrated across a dual platform of smartring and a smartwatch.

A few other studies reported the effect of the infection on physiological measurements provided by wearables without designing any machine learning model for infection detection/prediction. It was shown that the resting heartrate (RHR) rises significantly in 2 days surrounding the symptoms onset, and the step count remains lower than normal for days after the symptom onset [55]. Another study determined whether a watch measuring wrist temperature could accurately identify patients who are infected [126]. They showed that there was a significantly higher average maximum temperature in infected people compared with healthy people. Changes in respiratory rate was also analyzed to predict the risk of COVID-19 infection. RR, RHR, and HRV were measured using a strap and a gradient boosted classifier was trained for infection detection and prediction. This study leveraged the measurements during the sleep time to reject motion noise but it did not consider physical activities that could be an important contextual information indicating if an individual is sick.

4. DATA-DRIVEN CONTEXT DETECTION LEVERAGING PASSIVELY-SENSED NEARABLES FOR RECOGNIZING COMPLEX ADL*

In this section we study usage of passively-sensed nearables for identifying environmental context and leveraging that for detecting complex and high-level activities of daily living (ADL). Wearable sensors are usually not aware of the context that they are working in. The context is a broad term that includes any additional information that could help better understand a specific situation. For activity recognition with motion sensors' data, this extra information could be the location of the user (e.g., home versus at work), their social interaction, and the time of day. Since the set of possible activities in each of these contexts could be different, understanding this contextual information improves the ability of the system to detect activities with complex motion patterns.

4.1. Context-aware Activity Recognition

Current state-of-the-art systems that use wearables are not capable of distinguishing a large number of complex activities, which may appear similar regarding movement of the hand but with vital differences in context, such as bicep curl exercise versus eating or running in the gym versus running in home. With the development of IoT technology, a wide range of information could be accessible from the devices in a

*Reprinted with permission from "Data-driven Context Detection Leveraging Passively Sensed Nearables for Recognizing Complex Activities of Daily Living" by Ali Akbari, Reese Grimsley, and Roozbeh Jafari, 2021. in ACM Transactions on Computing for Healthcare, ACM Transactions on Computing for Healthcare (HEALTH), vol. 2, issue 2, pp. 12:1-12:22, ©2020 by ACM.

particular environment [93]. Our goal is to leverage freely available data from "nearables" to identify context that could be associated to the user's surroundings, i.e., location, social groups, and nearby objects, in order to reduce the search space for the application of recognizing complex ADLs.

Detecting more complex ADLs with merely motion sensors is extremely challenging, if even possible, due to similarity in motion signature of various activities [19]. There is a lack in studies on detecting complex ADLs in the body of literature. To distinguish between those complex activities, understanding the users' surrounding is the key. In this case, knowing the context helps to narrow down the set of probable activities, which in turn facilitates the activity classification. Understanding the context can similarly assist wearables in any other sensing and recognition tasks.

For an unobtrusive, scalable, and data-driven context detection we use freely available Bluetooth low energy (BLE) data broadcasted by devices in users' surrounding. It is also known as nearables [22]. These nearables could effectively be all devices equipped with Bluetooth such that they are passively detectable from the wireless signals they transmit. It is worth mentioning that, in this study we do not need to deploy any specific BLE device. Instead, we leverage the BLE data broadcasted by any device in the vicinity of the user and rely on the consistency of those devices over time to infer the context, which is further used to improve recognition of ADLs via a single wrist-worn motion sensor. To best of our knowledge, this is the first work that leverages freely available BLE information broadcasted by any device around the users for detection of such environmental context.

The creation of context in a completely unsupervised way is not a trivial task; it becomes significantly more difficult if many elements used to build the context are unreliable. This is the case when considering passively sensed devices in the user's vicinity because many nearables could be mobile. For the given application of activity recognition, a particular activity may occur in several different contexts; further, there may be multiple distinct activities possible within the same context. In this way, the relation between the user's action and context may not be 1:1. Here we focus on complex ADLs such that motion signals alone are not enough to distinguish all activities. Examples include attending a meeting or class, cooking, eating, working at one's desk, etc. Knowledge of the environment can help alleviate this issue. Figure 4.1 shows the effect of context w.r.t. ADL recognition. No context is used in Figure 4.1-a where any of n activities may be detected by the model. However, context helps estimate which activities are actually feasible, shown in Figure 4.1-b. In this study, contextual knowledge is derived from passively sensed BLE devices.

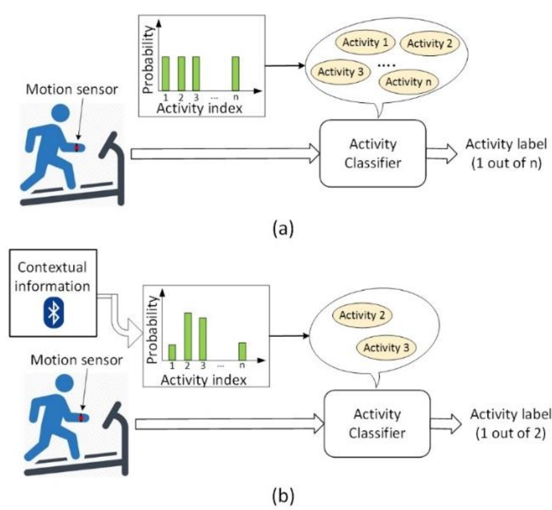


Figure 4.1 - Classification search space (a) without and (b) with context

4.2. Methodology

We assume that there is a general activity classifier trained on the data of the motion sensor embedded in a smartwatch. This classifier has a limited accuracy due to similarity in the motion patterns of complex activities. As the users start using the smartwatch during their normal daily life, the system starts to learn the context by looking at the patterns of BLE devices around the user. The overview of the proposed system is shown in Figure 4.2. When a new user starts using the device, the general activity classifier is used to assign labels to each sample using motion data, and the most certain labels are chosen for context learning.

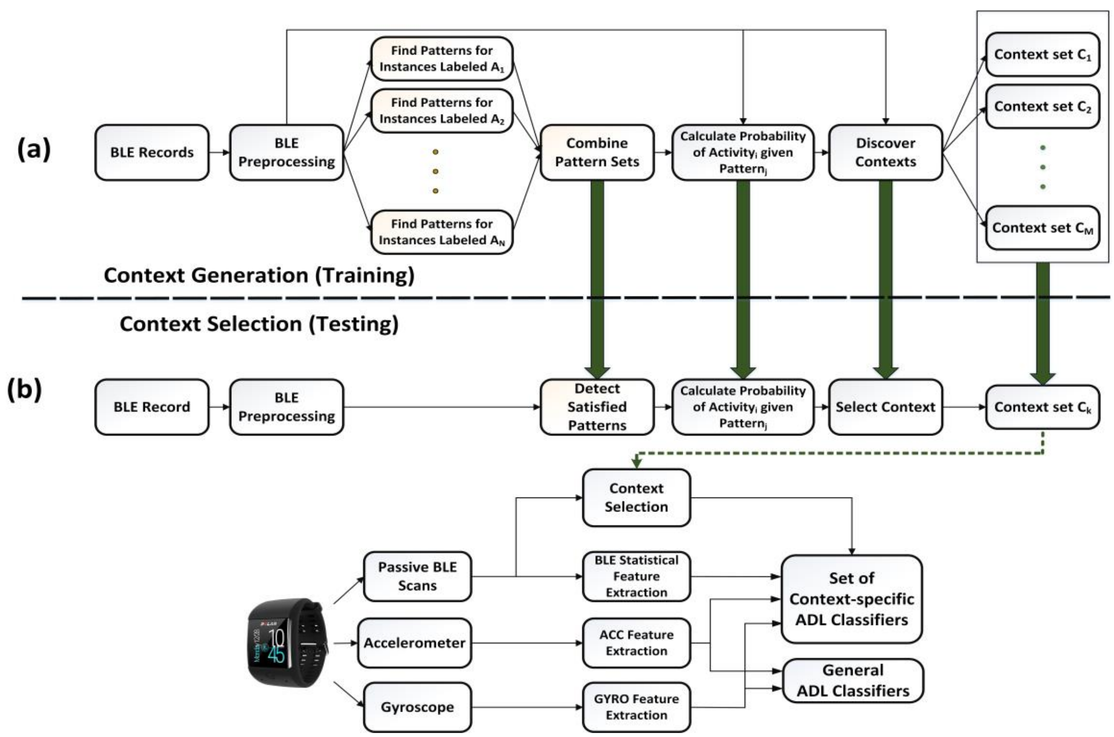


Figure 4.2 - Overview of the proposed system for context learning

In addition to motion data, we collect BLE scans broadcasted by devices in the vicinity of the user. After a period of data collection, as shown in Figure 4.2-a, the context is created separately for each possible action the user may take based off of patterns extracted from BLE devices scanned while that action was being performed. A probabilistic model then relates these possible outcomes to the observed contexts which is used to narrow the search space to probable actions only. A separate classifier is trained for each set of actions that share context that is called context-specific classifier. For a new sample then, as shown in Figure 4.2-b, the set of satisfied patterns is identified based on observed BLE devices. Using the probabilistic model, the patterns are then mapped to the set of probable actions; these are used to select the appropriate context-specific classifier, which is trained only on that reduced set of possible outcomes. When a new unknown context is observed, again the most reliable labels generated by the general classifier are gathered to retrain the context patterns. In fact, the context learning is an ongoing process that never stops. The ongoing training is required because the set of BLE devices around a user can dynamically change over time. In the following, we first explain our context training approach including BLE preprocessing, extracting context patterns and associating them to the activities in a probabilistic fashion. Lastly, in Section 4.2.3, we explain the process of labeling the data on-the-fly by leveraging general activity classifier and context-specific activity classifier. Before introducing the details of our approach, notation definitions are summarized in Table 4.1 to enhance the readability of the equations.

Table 4.1 - Term definition for context detection

Term	Definition
M	Set of all observed BLE beacons during data collection
m_j	The MAC address of the j^{th} element in M
R	The set of BLE MAC addresses (i.e., record) scanned during some one-minute segment of time
B_i	A binary vector of size $\|M\|$ representing the BLE record at the i^{th} timestamp
b_j	A binary element corresponding to j^{th} MAC in M
x_i	Sensor readings at i^{th} timestamp
y	The output label at i^{th} timestamp
L	Total number of possible output labels
A	Set of output labels $A = \{1, \ldots, L\}$
a	An output label where $a \in A$
n_a	Total number of samples with label a
p_k	k^{th} context pattern which is composed of one or more BLE beacons
K	Total number of context patterns
Co_p^a	The coverage of pattern p for label a; i.e., the percentage of samples labeled a for which context pattern p is observed
Φ	An L by K matrix that includes the probability of action a_i given context pattern p_j in i^{th} row and j^{th} column
P^R	Set of patterns satisfied by record R
ϕ^R	A subset of columns in Φ that corresponds to the patterns P^R satisfied by record R
Θ	A small lower threshold to filter out patterns that are rarely present during an action
$IoU(p_1, p_2)$	Intersect over union calculated between two patterns
N_{p_1}	Number of samples/instances in which pattern p_1 is observed
S_u	Motion signal, i.e., either acceleration or angular velocity, along axis u

4.2.1. BLE Representation and Preprocessing

Many BLE devices periodically transmit ‘advertisement’ packets containing basic information about a ‘beacon’, allowing us to collect a small amount of data about each advertising device in the user’s vicinity without any handshaking between the transmitter and receiver. Traditionally, beacons are BLE devices whose primary purpose is to advertise information about the environment; for this study, we refer to all BLE devices as beacons because we are only interested in their advertisement packet which indicates their presence in the environment. BLE is an appropriate modality for detecting both the social and locational components of environmental context because it is a protocol common to both static and mobile devices. The former may be telemetry beacons, infrastructural sensors or other stationary electronics while the latter may be smartphones, wireless headphones, wearable electronics, etc. Clearly, static devices would be effective for determining location similar to studies using Wi-Fi Access Point fingerprints [100], [101], [104], and mobile devices, when exhibiting consistent patterns with respect to each other, could help determine social groups [96]. However, distinguishing between static and mobile devices is difficult as both the user and the BLE device could be mobile. Please note that the main assumption here is that there is no information available about the type of the device when its BLE packet is received. This is a realistic assumption because: 1) many BLE packets include the MAC id of the device but there is no information about the type of the device in their packet, and 2) relying on the type of the devices will significantly limit scalability of the system. Thus, we choose to model all beacons as equivalently purposed, and we refer to all BLE devices as ‘beacons’ regardless of their physical purpose.

We log the 6-byte MAC address of each device that we scan. Collecting all MAC addresses of beacons over many days will nonetheless result in noisy data. For instance, many people encountered in a public place will not be seen again. Clearly, most of these beacons will be noise. We use a simple rule to filter out beacons which are obvious noise: if a device is present for a short period of time (less than 15 minutes a day in our study) then it is not useful. This simple filtering rule removes a large portion of beacons; for example, for one of our subjects, we filter from around 20,000 beacons detected across seven days down to around 8,200. With this reduced set of beacons, we aim to find patterns in the BLE records to help understand the user's context.

Figure 4.3 - Mobility of user and nearby BLE beacons

As depicted in Figure 4.3, the user's context may be represented by the devices near them at a point in time. This provides a fingerprint of the user's context. However, a single scan may miss some nearby beacons due to differences in the scanning period versus advertisement intervals. For this reason, we represent the user's current environment using the set of BLE beacons that are detected over multiple scans within a relatively short interval of time (one minute for this study based on the limitations of the smartwatch device to complete a scan of all nearby BLE devices). Let us consider this one-minute scan to be a BLE 'record'. Intuitively, we can represent a record simply by making a binary vector B of length $|M|$ where M is the set of all observed beacons during data collection period after filtering. Each element of a binary vector corresponds to a MAC address such that an element corresponding to beacon b_i is 1 if MAC address m_i was present and zero otherwise. Equation 4.1 formalizes this, where R is the set of all MAC addresses scanned during a one-minute segment.

$$B = (b_1, b_2, \dots, b_{|M|}) \quad , \quad B[i] = b_i = \begin{cases} 1 & , \quad m_i \in R \\ 0 & , \quad o.w. \end{cases} \tag{4.1}$$

4.2.2. Data-Driven Context Detection

When the user consistently visits a particular context, certain patterns should exist in the BLE scans corresponding to that particular context. The context patterns should be generalizable in the sense that they should not be too sensitive to the requirement of many observed beacons being co-present. For instance, in the training data, we might see three beacons that are always present together when the user is at home. If one of those beacons is no longer present during the testing time, the system should still be able to recognize the home context. At the same time, the patterns should also be selective. Beacons that are present in too many different contexts, such as the cellphone of the user, are not selective enough since they cannot distinguish different contexts. Therefore, we are interested in extracting consistent, selective, and generalized patterns of beacons that help understand the users' surrounding and facilitate ADL recognition.

4.2.2.1. Context Pattern Formulation

Recall that in this study, "context" refers to information about the user's surroundings; we assume that there is no provided information available about the true context of the user. In other words, there is no supervision and consequently, no labels available for the context which is a realistic assumption in real-world scenarios. In general, we use no a priori knowledge about the context. Although we have knowledge on the user's actions, which herein is provided by a pretrained motion-based activity classifier (or it can be provided by the user, see Section 4.2.3), there is not a 1:1

correlation between a possible action outcome and the context. Thus, we build the context for each outcome separately from the others. That is, for each action, we describe the context in which that action might be performed by extracting patterns from BLE records.

A BLE context pattern is a set of different BLE beacons that are consistently present in a specific context. To be formal, assume that (X,Y) is all of the training data where $x_i \in X$, $i=1,\ldots,n$ is the sensor readings, *e.g.*, accelerometer and gyroscope, for i^{th} data sample and $y_i \in Y$ is the label for i^{th} sample. $Y=\{1,\ldots,L\}$ where L is the total number of possible labels. Please note that the label y_i could either be generated by a pretrained motion sensor-based activity classifier or provided by any external source such as the user. In Section 4.2.3, we will explain how we leverage the labels generated by the motion-based classifier to avoid burdening the users. We assume that there is a context associated with each sample, which is unknown and should be learnt for each user. In addition to the input data x_i, we have a set of observed beacons B^i where $B^i=(b_1^i,b_2^i,\ldots,b_{|M|}^i)$ is a vector of beacons in which b_j^i is 1 if j^{th} beacon is present in i^{th} sample and 0 otherwise.

Definition 3.1 (Context Pattern). p_k is called a “context pattern” consisting of a set of specific beacons. Here $k=1,\ldots,K$ is the index of the pattern where K is the total number of context patterns extracted from the data. With this definition, one might consider a context pattern as an atomic location or environment; for instance, each pattern may indicate a different room, device or person around the user.

There is always a trade-off between the selectivity and generalizability of context patterns as mentioned before. To deal with this tradeoff, we design a modified version of hierarchical agglomerative clustering (HAC) to generate the patterns, termed accumulative hierarchical agglomerative clustering (AHAC), along with a probabilistic framework for mapping the context patterns to the actions w.r.t. the application domain using an action-context probability (ACP) estimation algorithm. AHAC allows us to have general and selective patterns at the same time. By estimating the probability of user actions given contexts (Section 4.2.2.3), we prioritize the probable actions then to narrow down the search space.

4.2.2.2. Context Pattern Learning

This section discusses how to extract meaningful patterns that are both generalized and selective from an abundance of observed BLE beacons. Recall that we create the context patterns for each possible outcome separately. For an action a, assume that $\mathcal{B}^a$ is the set of all training records that have label a where $\mathcal{B}^a=\{B_1^a, \ldots, B_{na}^a,\}$ and n_a is the total number of samples with label a. Recall that B is a BLE record represented as a one-hot-encoded vector that contains $|M|$ beacons, where $|M|$ is the total number of different beacons in the dataset. We first remove the beacons that are rarely seen during action a using Equation 4.2:

$$P^a = \left\{ j: \left(\frac{1}{n_a} \sum_{k=1}^{n_a} B_k^a[j] \right) > \Theta \} \quad, j = 1, \ldots, |M| \right\} \tag{4.2}$$

where j is the j^{th} beacon in the record, P^a is the set of all beacons that we keep to generate patterns for action a, and Θ is a constant threshold.

After this stage, one can just keep the single beacons as representation of the context for action a. However, such a basic method has a poor selectivity as it fails to deal with beacons such as the user's cellphone that is present in many different contexts. To address this issue, we can merge beacons that are often present together in order to create more specific and representative patterns. For this goal, we can utilize a clustering algorithm to put beacons into clusters whose contents represent more specific context patterns. We modify the hierarchical agglomerative clustering (HAC) to fit into our problem and call it Accumulative HAC (AHAC). Algorithm 4.1 shows the steps of AHAC algorithm that is done only during the offline training phase. Intuitively, AHAC elects to keep the unmerged components of a cluster in the entire set rather than removing them once they are merged. In fact, we grow the set of patterns to contain both generalized and selective patterns altogether. Later, by looking at the distribution of patterns in various actions we can effectively select the most probable context and set of activities subsequently. This method, starts by putting every single beacon in Pa into a single cluster. This means that at the beginning, the number of clusters is equal to |Pa|, each of which contains a single BLE beacon. We use the words 'cluster' and 'pattern' interchangeably because once we finish clustering, we use each cluster as a context pattern. We consider a pattern to be satisfied if all beacons within its cluster are present in a data sample. To merge two clusters together at each step of AHAC, we use the intersect over union (IoU) between the two clusters as the measure of similarity.

ALGORITHM 4.1: AHAC

Input: all training records $\mathcal{B}^a$ for action a and its sample size n_a
Output: set of patterns (*i.e.*, clusters) P^a for action a

$P^a = [\,]$
for *(every beacon* $j = 1$ **to** $|M|$*)* **do:**
    if $\left(\frac{1}{n_a}\sum_{k=1}^{n_a} B_k^a[j]\right) > \Theta$
        $P^a \leftarrow j$
    end if
end for
$IoU = 1$
Assign each element of P^a to a cluster
while $(IoU > 0.75)$
    find *clusters i* and *j* with max $IoU(i,j)$
    $P^a \leftarrow$ [*cluster i*, *cluster j*]
    $IoU = IoU(i, j)$
end while

Again, note that we have a separate set of clusters for each possible activity label. Equation 4.3 shows how IoU is calculated for two patterns. This measure tells us about the proportion of co-occurrence of two patterns. Higher IoU means that the two patterns (i.e., set of BLE beacons) are consistently present simultaneously, which increases the likelihood that they belong to the same context.

$$IoU(p_1, p_2) = \frac{N_{p_{1,2}}}{N_{p_1} + N_{p_2} - N_{p_{1,2}}} \tag{4.3}$$

In Equation 4.3, N_{p_1} and N_{p_2} are the number of samples that pattern p_1 and p_2 apply to, respectively, and $N_{p_{1,2}}$ is the number of records that apply to both patterns, *i.e.*, both patterns are present at the same time.

At each iteration of clustering, the two patterns/clusters with the maximum IoU are chosen to be merged together. The two selected patterns are merged and added to the whole set of patterns. Here we make a modification to HAC by keeping all the original clusters and adding the merged cluster to the total set, meaning the cluster count increases by one with each iteration. In this way, we refer to our pattern extraction technique as Accumulative HAC (AHAC). We make this change because if we remove patterns p_1 and p_2 when they are merged, as is the case with unmodified HAC, we progressively create overly specific patterns such that it becomes more difficult to fit our dataset reliably, especially for new data. As opposed to the normal HAC where the number of clusters decreases by one at each iteration, this number increases by one in AHAC, meaning this process will not naturally terminate itself. In our approach, clustering is finished once the maximum IoU of any two previously unmerged patterns is below a threshold, which is set to 75% for our study. This process is done in offline training to create a set of important context patterns.

In fact, this approach can be seen as an iterative clustering algorithm. In first iterations, the BLE beacons are combined together, but as the algorithm proceeds, the clusters (patterns), which are a set of mutually observed beacons measured by Equation 4.3, are combined to create more specific patterns. Each pattern generated in this way could potentially be a combination of people and/or static devices in a particular location.

During the testing for a new sample, a single BLE scan/record is evaluated to find which patterns are satisfied; a pattern is satisfied when all beacons within that pattern are present in the record. Then, the action-context probability (ACP) estimation engine creates the probability of actions given the satisfied context patterns, selects the best classifier for the set of possible activities, and then sensor data is fed into that chosen classifier to recognize the action.

4.2.2.3. Action-Context Probability Estimation

This method aims to estimate the probability that a particular outcome (i.e. an activity) occurs given a particular context pattern p is seen. Algorithm 4.2 formalizes this procedure for generating action-context probability matrix; it populates this matrix using Bayes rule to measure the probability of some action given a context pattern. Algorithm 4.3, then formalizes the ACP estimation procedure in which the distribution of action-context probabilities for context patterns that are satisfied by a BLE record are used to decide which actions are probable such that the search space may be reduced. We use Bayes theorem to represent these action-context probabilities in Equation 4.4.

$$Pr(a|p) = \frac{(Pr(p|a)Pr(a))}{\sum_{\alpha \in A} Pr(p|\alpha)\, Pr(\alpha)} \tag{4.4}$$

where $Pr(a)$ is the prior probability of action a, for which we consider a uniform distribution, A is the set of all possible actions, and $Pr(p|a)$ is the probability of observing context pattern p while performing action a, which can be estimated from the training data as shown in Equation 4.5.

$$Pr(p|a) = Co_p^a = \frac{\#\ of\ appearance\ of\ p\ in\ a}{\#\ of\ appearance\ of\ a} \tag{4.5}$$

We call Co_p^a the 'coverage' of pattern p for action a, *i.e.*, the percentage of samples labeled a for which context pattern p was satisfied. Assume that for all the L actions possible in the chosen application domain, we have a total of K different context patterns. Then, we use the entire training dataset to generate an L by K matrix, Φ, which is a matrix containing all action-context probabilities. In the i^{th} row and j^{th} column of this matrix, we put the probability of action a_i given context pattern p_j as shown in Equation 4.6.

$$\Phi_{i,j} = Pr(a_i|p_j) \tag{4.6}$$

ALGORITHM 4.2: Creating ACP Matrix

Input: all training samples, set of user actions, set of all patterns,
Output: action-context matrix Φ
1- **for *(****each action a = 1* **to** *L****)*** **do:**
2- **for *(****each pattern p = 1* **to** *K****)*** **do:**
3- $\Phi_{a,p} = Co_p^a / \sum_\alpha^L Co_p^\alpha$ //calculated by Equation 4.4 and 4.5
4- **end for**
5- **end for**

ALGORITHM 4.3: ACP Estimation

Input: the observed record R for the current sample, Φ
Output: Set of mostly probable activities *Action_set*
1- *Action_set* = [ ]
2- $\phi^R = [\,]$
3- **for *(each*** *pattern* p_j***)*** **do:**
4- **if** ($p_j \in R$):
5- add column *j* of Φ to ϕ^R
6- **end if**
7- **end for**
8- **for *(each*** *action a = 1* **to** *L****)*** **do:**
9- $Prob_a$ = mean a^{th} row of ϕ^R
10- **if *(*** $Prob_a > \frac{1+\varepsilon}{L}$ ***):***
11- *Action_set* ← *a*
12- **end if**
13- **end for**

We retrieve a subset of this matrix's columns to generate a temporary matrix for the BLE record of the current sample based on the patterns P^R satisfied in the current record *R*. The retrieval of this matrix is shown in Equation 4.7 as well as Algorithm 4.3, and allows us to estimate the set of possible actions for the record *R* that may share context. These actions are called the 'Shared Context Set'.

$$\phi^R = \left(\mathbf{\Phi}_j : p_j \in P^R, j \in \{1, \dots, K\}\right) \tag{4.7}$$

For this goal, we calculate the mean probability for each possible outcome based on the probabilities of that action given each pattern in P^R. In other words, we take the average value over the columns in matrix ϕ^R for each row. We pick the actions for which this average is greater than a threshold as shown in Equation 4.8. The threshold in our study is one plus a parameter ε to reduce sensitivity (in our study $\varepsilon = 0.25$) over the total number of actions, L.

$$Shared\ Context\ Set^R = \left\{A_i : \frac{1}{|P^R|}\sum_{j=1}^{|P^R|}\phi_{ij}^R > \frac{1+\varepsilon}{L}\right\} \tag{4.8}$$

The set of possible actions are identified based on the outcome of Equation 4.8. In offline training phase, we identify possible sets of actions for all training samples and use them to train context-specific classifiers as explained in Section 4.2.3. In online testing phase, we first determine the set of possible actions for each new sample based on the context using Algorithm 4.3, and then pick appropriate context-specific classifier from the pool of classifiers.

4.2.3. General and Context-specific Classification

The context training needs to be personalized since the set of BLE devices around each user could be different. Moreover, for a user, the set of observed BLEs in a certain location may change over time. To account for this, the system should be able to learn the context for each user incrementally and the context learning should be a non-stop process. We assume that, at the beginning, there is a general activity classifier available that is trained merely on motion sensors' data with no contextual knowledge. This general classifier is trained solely on the motion data, but its accuracy obviously affects the performance of the system (see Section 4.3.3), so the best performance could be achieved if the motion-based classifier is trained or fine-tuned for the data of each user. The system starts learning the context when it is used by a new user and the context learning is updated periodically. Context training process begins with an empty set of patterns for each action a, and there is no context-specific classifier available when the user starts using the device for the very first time. For a couple of days, seven days in this study, the system only stores all the motion sensors as well as BLE data and uses the general motion-sensor-based classifier to assign labels to those samples. The samples about which the general classifier is confident, i.e., its confidence score is above a threshold, are selected as the samples for extracting context patterns as described in Section 4.2.2 and training the context-specific classifiers. Based on the output of ACP algorithm during this offline training, we end up with multiple lists of user actions, where all share context, i.e. Shared Context Sets. Consider an example: based on training data we realize that one subject's activities of 'eating', 'cooking', and 'cleaning' fall into one list as all of them happen in the same location. For each reduced set of probable outcomes, determined using the context, we train a distinct context-specific classifier. Since the context-specific classifiers are trained on a reduced set of activities and they also leverage contextual features, they are more accurate than the general classifier, which may confuse activities with similar motion signature.

After the first period of context learning, the context-specific classifiers are ready to be used to narrow down the list of activities to be classified. In the testing phase, for each sample i, we look at the observed BLE record R^i and retrieve the set of satisfied patterns. If there is any pattern p satisfied by the record R^i, the appropriate context-specific classifier is chosen using Algorithm 4.3 and used to detect the activity label. In this case, if the label generated by the general classifier is not in the list of probable activities that are determined by the context, and if the general classifier is confident about its decision, then that sample and the label generated by the general classifier will be stored for context retraining in future. This is important because there might be cases that the user performs a new activity in a previously known context. Please note that, the proposed ACP algorithm that maps the context to the probability of activities, considers the frequency of occurrence of the activity in each context; therefore, if the user randomly performs an activity in a specific context just for a very short time and non-consistently, that outlier will not affect the outcome of the algorithm. If there is no pattern satisfied by R^i, which means the user is visiting a previously unknown context, that sample along with the label generated by the general classifier will be stored for context retraining only if the confidence of the general classifier is high. These data that are collected during the testing time are fed into algorithms 1-3 to update patterns and action-context matrix. It should be noted that after every three days of data collection the context training is updated with the new set of labeled data.

Although we select the most confident samples labeled by the general classifier, the overall accuracy of this general classifier can impact the output of the algorithm. The required labels of the activities, however, can be provided by any external source such as the users themselves depending on the availability of that source. For example, the user could provide the labels for all the activities within the first week of data collection for context training and then the system will sporadically ask the user when it faces an unknown context or less confident predictions. In Section 4.3.3, we compare the effect of getting labels from a reliable source such as the user versus the general motion-based classifier.

For our case study on ADL recognition, we use SVM (with regularization parameter set to 1) as the classifier and use the probability scores as the measure of confidence of the classifier [127]. We extract a set of various time and frequency domain features from accelerometer and gyroscope sensors along with statistical features from BLE beacons as shown in Table 4.2 [128]. The features from motion sensors are calculated for each of X, Y, and Z axis along with the magnitude of the signals, calculated by Equation 4.9, which is orientation independent [109].

$$A = \sqrt{S_x^2 + S_y^2 + S_z^2} \tag{4.9}$$

In Equation 4.9, A is the magnitude of the signal, and S_u could be either acceleration or angular velocity signal around axis u measured by accelerometer and gyroscope respectively.

Table 4.2 - Context-aware ADL recognition input features

IMU Features					**BLE Features**
Mean	Min	Variance	Power in 0-0.5, 0.5-1, 1-2, and 2-5 Hz bins	Mean-cross-rate	Number of devices
Standard Variation	Max	Root mean square	Frequency of maximum FFT	Autoregressive coefficients of order 10	Turnover (Beacon Change Rate)

For the BLE statistical features, we consider the number of devices present and the ‘turnover’ of beacons. Turnover is effectively the change rate of beacons: by looking at how much the set of nearby beacons changes, we can get a rough idea of how much the environment is changing. If the subject is traveling or stays in a public place, there would be a high rate of changes. We calculate turnover as the *IoU* (Equation 4.3) between the current record and the previous one. In total we have 180 features that are normalized to range [0, 1]; 176 of these are calculated from motion signals.

It should be noted that the length and frequency of each activity was very different in this study depending on person’s routine. This results in a very unbalanced training dataset. Our strategy to mitigate the effect of unbalanced training data is to reweight the cost functions for each class. Explicitly, we calculate the weights w_i as shown in Equation 4.10, where c_i is the number of instances of activity i in the training set. These weights are used in the cost function for SVM classifier to mitigate the effect of unbalanced samples.

$$w_i = \frac{\max_j c_j}{c_i} \tag{10}$$

4.3. EXPERIMENTAL RESULTS

In this section we validate our proposed framework by showing how the environmental context helps recognize high level activities of daily living that could be nearly impossible to distinguish with merely wrist-worn motion sensors. Five subjects were equipped with smart watches and asked to collect data and labels on their daily activities. Each subject collected data for approximately 20 days. All data for this study were collected in the wild while users performed their normal daily activities. For data collection in this study, we used a commercial smart watch, a Polar M600 running the Android's smart watch operating system, Wear OS 2.0, based on Android 8.0. We collected accelerometer, gyroscope, and passive BLE data with a custom Android application. This app features a simple interface shown in Figure 4.4 for the user to start and stop data collection, as well as a list of labels that will log the label and time stamp when an entry in the list is tapped as shown by the pop-up in the figure.

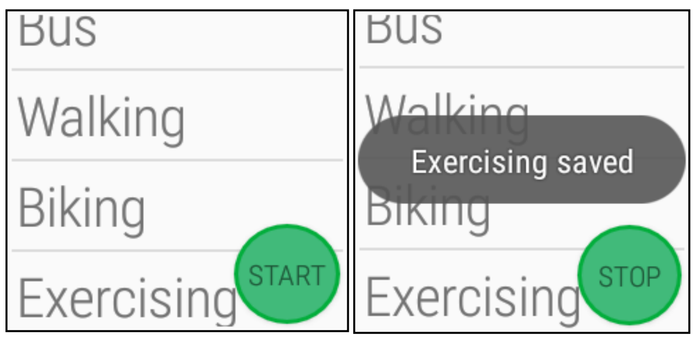


Figure 4.4 - Data collection app interface

We collected around 100+ days worth of data, which includes more than 580 hours, from all the subjects. The data from first two subjects were used for training the general motion-sensor-based classifier while the data of other subjects were used for context training as explained in Section 4.2.2. Table 4.3 shows the full list of all activities assessed in this study. In this study, the participants provided the labels for only the listed activities and we ignored the label of other activities not included in Table 4.3. It should be noted that this is an extensive set of ADLs compared to the existing literature in activity recognition. In fact, detecting such complex ADLs without leveraging contextual information and by merely using motion sensors would be very challenging and inaccurate. In the following, we first show the importance of context in recognizing these complex activities. We then compare our context modeling approach with baselines since there is no similar work in modeling the context from freely-available BLE beacons. Finally, we assess the effect of the general classifier's accuracy on the overall performance and the impact of having more reliable source of labeling such as the user. It should be noted that there is no similar study or publicly available dataset in the literature that includes all the BLE scans for a couple of days in uncontrolled environments.

Table 4.3 - Set of activities

Activities				
Biking	Working	Eating	Cooking	Relaxing
Exercising	Class attending	Cleaning	Getting Ready	
Walking	Meeting	Shopping	Driving	

4.3.1. General versus context-specific activity classification

We start by demonstrating the need for context by comparing performance of ADL recognition using the general activity classifier that does not leverage the context with our system that uses the context. We first show that contextual information is essential for detecting complex ADLs. For many people, daily activities consist of long stretches of sedentary positions. For instance, during working or attending a class, there may be minor movements like typing on a keyboard, using a smartphone, or turning pages of a book, but the overall motion characteristics of largely immobile activities are likely to be similar if not indistinguishable. Attempting to recognize activities using only features from motion data, will perform poorly due to the similarity of the signals. Figure 4.5 shows the acceleration signal from five minutes of attending a class, working, and meeting activities. In our findings, most parts of signals for more sedentary activities, i.e., class, working, meeting, exhibit periods of inactivity with short, intermittent bursts of motion activity. If the duration and intensity of these bursts is to vary across instances of these activities, then there is little to distinguish said activities using motion alone.

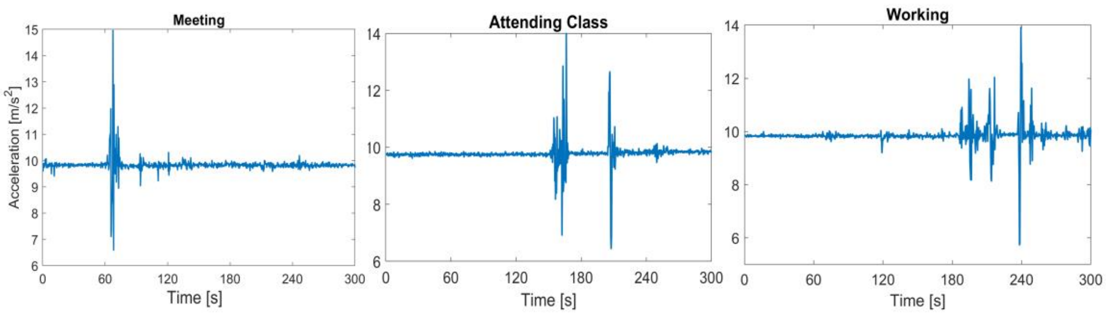


Figure 4.5 - Acceleration signal during three activities with extended periods of sedentary activity

In contrast motion signals during walking, running, and driving show more unique characteristics that are likely sufficient for classification with merely motion sensors. In truth, experiments have shown acceptable performance for high-mobility activities like walking, biking, and exercising, yet low-mobility ADLs perform poorly when only motion signals are used. Several other examples of such complex ADLs are investigated in [19] by placing pre-known BLE beacons in specific locations in user's home. This poor performance has a dramatic effect on the total accuracy of the system given that sedentary activities usually occur over long portions of time as discussed later in Section 4.3.2.

To remain fair in comparing our method using context identification with general activity classifier that does not leverage context, for the general classifier we use both the motion features as well as BLE statistical features as listed in Table 4.2. However, in the case of general activity classifier there is a single classifier that attempts to detect all the activities. In contrast, by modeling the context we have a series of context-specific classifiers each of which works on a specific set of activities. Figure 4.6 depicts the F1-score for each activity for the three subjects when context is or is not applied to reduce the search space of possible activities. It is clear that including context affects some activities much more than others. Generally, activities which have a greater degree of motion, such as 'biking', 'exercising' or 'walking', do not change much when context is included. In these cases, there is ample information in the motion features alone to confidently decide which activity the user is performing. However, we see this is not the case for sedentary activities such as attending 'class', 'meeting' or 'working'. In each of these, the motion and a few statistics from BLE scans are not enough to consistently choose the correct activity; however, an understanding of the context makes it much easier to detect the correct activity.

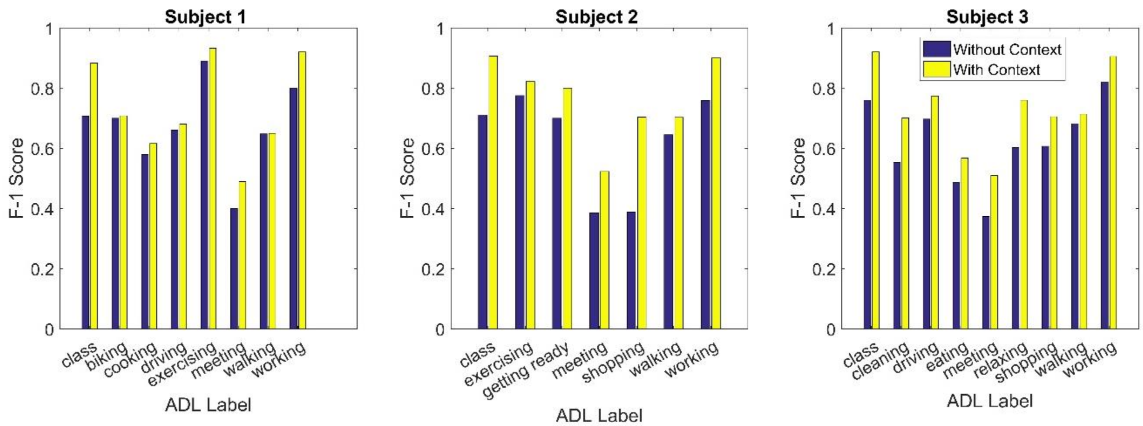


Figure 4.6 - Comparison of class-based F-1 score between the general activity classifier (without context) and context-specific classifier

For instance, including context results in F-1score improvement of 0.17 in attending 'class', 0.11 in 'working', 0.12 in 'meeting', 0.18 in 'cleaning' and 0.21 in 'shopping' on average. Note that the worst performing activity is 'meeting': this activity shares context with many instances of 'working' activity and has few distinguishing motion patterns of its own, making it extremely difficult to recognize this activity consistently, regardless of context knowledge. Also, modes of transportation such as 'driving' and 'walking' have lower performance (around 0.62 and 0.68, respectively) than would be expected given the amount of motion. The reason is two-fold: firstly, there were few instances of these activities in the dataset, so even a few misclassifications are highly damaging, and secondly, instances near the time of departure or arrival may detect context patterns from the previous activity, causing them to be incorrectly placed into a context model that does not consider the true activity as a possible outcome. Generally, these locomotive activities were found to have no detectable context.

Nonetheless, we see the weighted average of F-1 score increases from 0.72 to 0.80 on average. In addition to F1-score which considers both the precision and recall, the overall accuracy of activity classification is shown in Figure 4.7. By modeling the context and using it to narrow down the search space of activity recognition, the average accuracy increases from 73% to 82% compared to using a single classifier for activity recognition, although it is fed with both motion and BLE statistical features. It is without doubt we can claim context is advantageous to this application domain. Please note that the activities studied in this research are mostly complicated activities of daily living where detecting all of them with merely motion sensors is very challenging and nearly impossible [19].

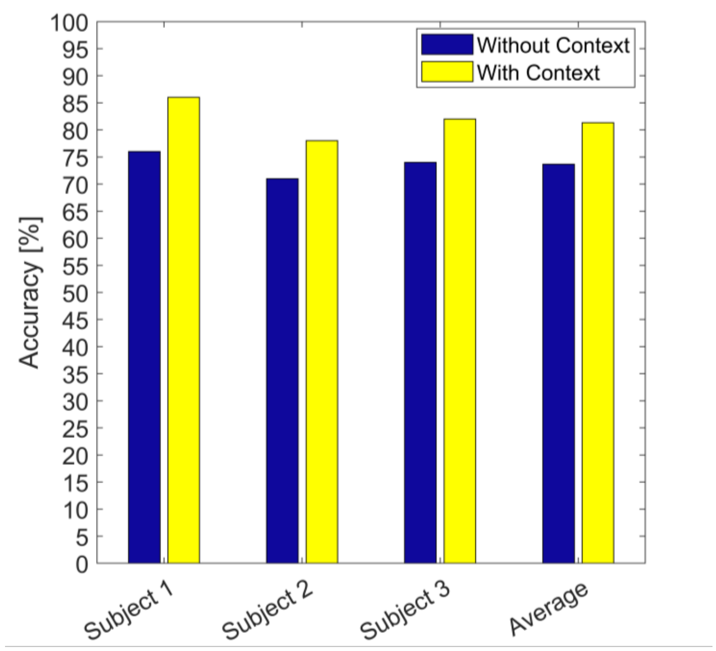


Figure 4.7 - Comparison of overall accuracy between the general activity classifier (without context) and context-specific classifier

Examples of such activities in this study are attending class, meeting, relaxing, working as well as cooking, shopping, cleaning, and getting ready. This is a significant improvement regarding the list of activities to be detected compared to the existing works in the field of activity recognition with wearable sensors. Moreover, it should be mentioned that there is no similar study and/or dataset in the literature that contains the BLE data from all the devices in the environment for detecting such high-level ADLs with a single wrist-worn sensor to compare with our method. In this study we did not include short-term activities such as toileting or hygiene, although these activities of daily living are very important in healthcare monitoring. It would be an important future direction to expand this work by collecting more data with a more exhaustive set of labeled activities. Since most of these activities happen in a certain location, the BLE would be helpful to narrow down the search space of activities so that motion sensors can easily differentiate them without confusing with other similar activities.

4.3.2. Context Learning Performance

In Section 4.3.1 we saw that using context for explicitly narrowing the search space achieves a higher accuracy than training a single model for the entire set of activities. In this section, we assess different components of our context modeling system and their impact on the performance of detecting the ADLs. It should be mentioned again that the studies that used BLE for complex ADL recognition all rely on setting up known BLE beacons in specific locations, which limits their scalability and convenience [19]. There is no study and/or dataset in the literature that contains BLE readings from all the devices without need to setting up specific hardware. As there are few related works whose methods can be applied here for a direct comparison, and also to demonstrate the importance of each component of our context training system, we compare our methods with respect to several more basic methods. In our approach, we build a set of patterns for each activity as described in Section 4.2.2.2. We refer to this method as Accumulative Hierarchical Agglomerative Clustering (AHAC), formally expressed in Algorithm 4.1. We relate patterns with activities probabilistically, i.e., the action-context probability (ACP), and use the distribution of these probabilities to infer context and narrow the set of possible activities as shown in Algorithms 4.2 and 4.3.

Therefore, we compare our methods to more basic ones based on the two aforementioned components of pattern creation and pattern usage to show the importance of each component of our framework. One basic approach is to use all singular beacons as patterns if they exceed the minimum coverage threshold in Equation 4.2. We refer to these as “Single-Beacon Patterns.” In fact, in this approach we ignore the AHAC module and assume each single beacon is associated with a context. Another basic approach is to ignore probabilistic association between the context and activities and use the patterns such that if a single pattern is discovered from an activity’s records, then we consider that activity as a candidate for recognition in that context. This is similar to building a look up table that assumes an activity a is probable in context p if there is even one observation of a at p. We refer to this as “Basic Pattern Usage.” One other approach uses a single classifier where each beacon’s presence is encoded as a binary feature in addition to the same set of IMU and BLE statistical features used in all the other classifiers. We also show results when patterns are extracted using the typical hierarchical agglomerative clustering (HAC) instead of our proposed modification (AHAC).

In order to account for probable noise in the detected nearables, mimic some real-world scenarios, and better evaluate our proposed method in response to challenging cases where BLE beacons can change drastically over time, we add or remove some beacons to/from the BLE records. To summarize these real-world scenarios, we simulate 1) a personally owned device by adding a specific beacon to 80% of all records, 2) beacons owned by friends or coworkers who span multiple contexts by adding specific beacons to 50% of instances of activities that may happen in the same context such as exercising and attending class, and 3) removal of some of the highly consistent beacons. We then compare our method with those described in the previous paragraph as shown in Figure 4.8.

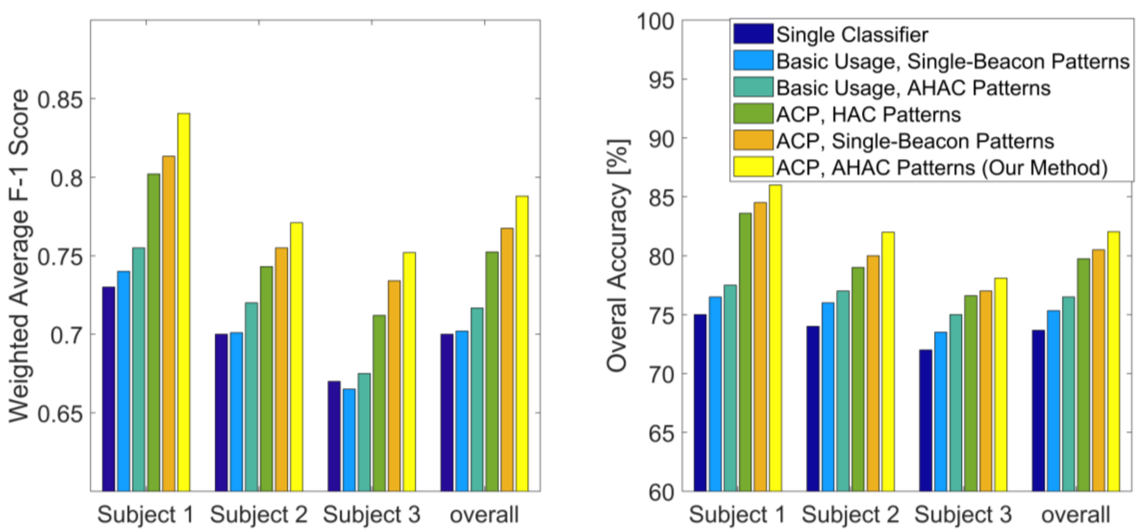


Figure 4.8 - Overall performance of different context detection approaches on noisy BLE data

According to Figure 4.8, our approach to pattern usage consistently outperforms the other approaches to context detection. Clearly, Basic Pattern Usage, in which only a single context pattern must be satisfied for an activity to be considered a possible outcome, performs most poorly as it is very susceptible to outliers. The mobile devices, such as subject's personal device causes different activities that do not naturally happen in the same environment to be considered probable, meaning that the model will perform little better than a single classifier using IMU and BLE statistics as input features. The difference in accuracy and F-1 score, for approaches that use ACP but different pattern sets are due to the more nuanced problems AHAC pattern extraction is designed to address. Patterns developed with AHAC can associate the mobile beacons with static beacons (or other consistent mobile beacons) from a particular context, thereby reducing the adverse effects that mobile beacons have on the distribution of ACPs when said beacon is also present in other contexts. However, AHAC will produce multi-beacon patterns composed of beacons that are frequently co-present. From these multi-beacon patterns, we get a more defined estimate of the set of feasible activities because these multi-beacon patterns will be more selective than their individual components. Even when important static beacons are not present, the combinations of multi-beacon patterns often still provide enough information to narrow the search space.

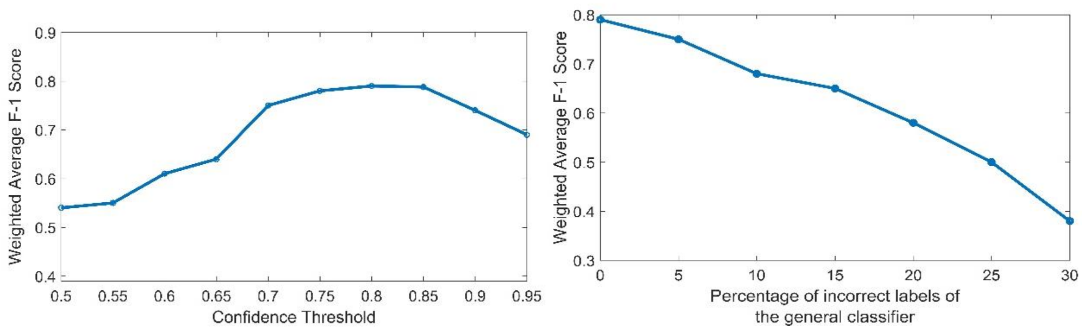
Using the single classifier, for subject 1, not a single instance of 'cooking' is recognized. Here, there was likely much reliance on specific beacons such that features from motion signals were entirely ignored. Moreover, for the single classifier, 'class' and 'working' are commonly confused with each other. We also observed that numerous instances of ‘working’ were misclassified as ‘walking’. This is a surprising result due to the sedentary nature of working. Upon close inspection, we found that the subject walked enough within the building they work in to have a meaningful effect on the context models. This means that some beacons from their typical ‘working’ contexts also contain beacons considered important to walking. When we remove some highly consistent beacons from activities labeled ‘working’, some of those central to a context for walking still remain, causing the wrong context model to be chosen. However, when we use AHAC to extract patterns, that beacon which has a significant connection to walking forms multi-beacon patterns with beacons consistent with the ‘working’ activity. This more often causes the model to select the classifier trained on ‘working’ or both ‘working’ and ‘walking’. In any case, the weighted-average F1-score of AHAC compared to Single-Beacon patterns is 0.78 compared to 0.70, suggesting AHAC has a substantial benefit.

To generalize these results, we see that noisier context data necessitates a clever usage of patterns. Our ACP estimation approach is effective at ignoring much of this noise by focusing on the overall set of patterns and how they relate to individual outcomes in the application domain. The means by which patterns are extracted is also important but plays a lesser role than their exact usage; we note that this is highly dependent on the data set. To quantify this, Figure 4.8 shows that Basic Pattern Usage has an average weighted F1-score of 0.70 and accuracy of 73% on average across all subjects, while applying ACP to those same patterns results in an increase to 0.76 F-1 score and 80% accuracy; this further increases to 0.78 and 82% when AHAC is used for pattern extraction, suggesting ACP to be the more influential part of our approach.

It should be noted that the use of BLE devices for context identification in the literature is limited to deploying pre-known BLE beacons in specific places for localization purposes. First, this approach is burdensome for users as they need to deploy the hardware in their living environment. Second, extra infrastructure is needed to integrate the beacons. Third, it is limited to localization in specific places such as user's home. On the other hand, our proposed method leverages all the freely available BLE devices in user's locale, which could be very noisy information, but can account for both the location and the people and/or devices around the user. For example, when a user attends a class, the context may be identified as other students in the same class rather than the physical location of the classroom. Since this study is one of the first attempts towards modeling the context from freely available BLE devices broadcasted by any device around a user, we could not provide quantitative comparison with other context identification models. However, this study shows the feasibility of modeling context based on freely available BLE devices, which can enable new paradigms in detecting complex activities of daily living with wearable sensors.

4.3.3. Impact of Activity Labeling on Context Learning

As explained in Section 4.2, our proposed framework does not require to use any direct label of the context for the purpose of context learning. This is a realistic assumption since it is not feasible to get such labels from users about the details of their surrounding context all the time. Instead our algorithm leverages the labels of the activities to learn the context in order to narrow down the search space. To achieve this goal, in Section 4.2.3, we described how the system can use the most confident labels generated by the general motion-based classifier. In this section, in Figure 4.9, we show how the accuracy of that general classifier can impact the overall performance of context learning. Moreover, we demonstrate the improvement that can be achieved by getting the correct activity labels from the user in Figure 4.10.



(a) (b)

Figure 4.9 - Impact of activity labeling on context learning; (a) The effect of threshold for choosing most confident samples for context learning on the performance of the system; (b) The impact of general classifier's accuracy on the performance of the system

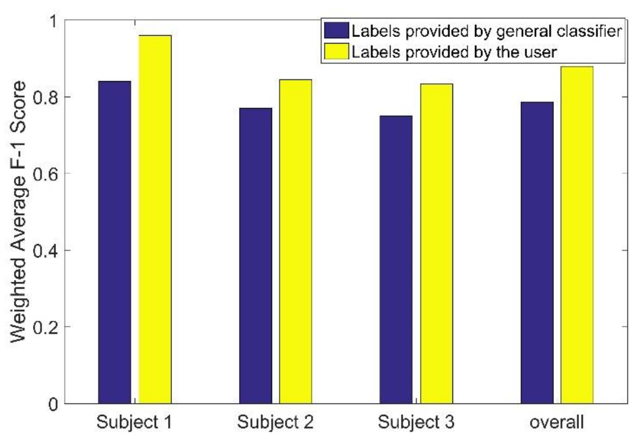


Figure 4.10 - The impact of getting label from the user on the overall performance of the system

Figure 4.9-a shows that varying the threshold, which is used to select the most confident decisions from the general classifier for context learning (see Section 4.2.3), affects the overall accuracy of the model. The Y axis in this figure is the weighted F-1 score averaged over all the subjects. As the figure shows, very low threshold reduces the accuracy because the system ends up using samples that can be misclassified by the general classifier, which results in learning inaccurate context. Similarly, higher threshold causes choosing very small number of biased samples which results non-sufficient context learning. Based on this figure, we chose threshold of 0.8 as a reasonable number to keep the confident samples for context training given our collected dataset. Figure 4.9-b shows that as the error of the general classifier increases, the accuracy of the context-based activity recognition decreases because the system ends up associating context to incorrect activities. In this figure, the X axis shows the percentage of labels to which we assigned a randomly incorrect label to mimic the error in general classifier, and Y axis is the F-1 score averaged over all the subjects. Therefore, more accurate motion-based classifier can help to better context learning, which in turn helps to more accurate recognition of activities.

A possible limitation of this approach is that when BLE devices are added or removed to/from the subject's surroundings, the system should relearn the changed context. This is a challenge since the BLE devices around a user can dynamically change over time as the person visits new places and people. Therefore, this system must always be updated via an online learning approach. It could also be combined with a supervised online learning in which the annotations can be queried from the user for more accurate training. The need for user intervention, however, could be minimized by detecting consistently observed BLE patterns and then query the user for annotation. Hence, most part of this online learning can be automated without the need for extensive user interference. This inspires an excellent future work in this area.

Figure 4.10 shows the performance when the labels of activities for context learning are provided by the user, which can be interpreted as the upper bound of the accuracy. As this figure shows, the overall accuracy of the system can be increased to 0.88 when the reliable labels are provided, which is 0.08 more than the accuracy when the labels are supplied by the general motion-based classifier.

4.4. Conclusion

We explored data-driven context detection on passively observable devices in the user's daily environment. Our approach relies on knowledge within a chosen application domain for training the context model. In this work, the application is recognition of ADLs, and context is built from passively-sensed BLE devices. Our approach builds an offline context model by leveraging the consistency at which individual BLE devices are observed near the user when a particular action of the user, i.e., activity, is being performed. This consistency is used to generate a set of context patterns which are then mapped probabilistically back to the actions to aid performance via a reduction in the set of probable outcomes. Using this method, we achieved an average F-1 score of 0.80 for detecting ADLs from data collected in real-world as the users went about their daily lives. This method offers a significant improvement over a single classifier approach for the same input features and has been shown to be more robust to realistic types of noise than similar methods. Important advantages of our proposed method compared to the state-of-the-art environmental context detection approaches are that it does not require one to deploy additional hardware and infrastructure, reduces users' burden, and can potentially infer the context related to both physical location and people and/or devices around a user. The proposed technique enhances the capabilities of wearable sensors by helping them understand their working environment. It also enables detecting complex ADLs with a single wrist-worn motion sensor.

5. PERSONALIZING ACTIVITY RECOGNITION MODELS THROUGH QUANTIFYING DIFFERENT TYPES OF UNCERTAINTY*

Data gathered by wearable sensors could be analyzed by rigorous machine learning models, such as context-aware models, to build a system to recognize ADLs [129]. However, a given model trained on a specific user may not generalize well to new users due to variation in how people perform specific activities [130]. Therefore, it is necessary to personalize the underlying machine learning models to new users.

5.1. Adaptable Machine Learning Models

Supervised and unsupervised learning approaches contain promising methods to design adaptable machine learning models with personalization capabilities [63], [131], [132]. Supervised learning requires the gathering of annotated data from a new user to retrain the machine learning models. However, this collection process is time consuming and burdensome. In fact, it is shown that user compliance to the wearable devices decay over time, particularly, when the collection system requires constant interaction [133]. Therefore, it is vital to personalize the machine learning models for new users with minimum labeled data available from the users in order to minimize the burden on the users. Unsupervised retraining approaches attempt to assign pseudo labels to unlabeled data by leveraging cross-user similarities, and then use them for retraining the model

*Reprinted with permission from "Personalizing Activity Recognition Models through Quantifying Different Types of Uncertainty using Wearable Sensors" by Ali Akbari and Roozbeh Jafari, 2020. in IEEE Transaction on Biomedical Engineering (TBME), vol. 67, issue 9, pp. 2530-2541, ©2020 by IEEE.

parameters [76]. Consequently, training and adapting the models in this manner is often less accurate than supervised approaches, especially, in the presence of significant inter-subject variability. However, in the field of ADL recognition, there is typically an abundance of unlabeled data available; thus, leveraging these algorithms effectively could improve and accelerate the personalization process [134].

Another promising solution known as active learning may be oriented in a supervised fashion by identifying the most critical data samples and soliciting their labels from the users to retrain the machine learning models [118]. The uncertainty of the classification models is one of the mostly used metrics for identifying those critical samples. These methods, however, suffer from a few limitations: 1) They do not consider different sources of uncertainty in their models. For example, the model may experience high uncertainty due to temporary noise in sensor measurements. In this case, the model will make an unnecessary request for labels while it is not reasonable to use it for training [111]. 2) They also do not limit the number of interactions with the user and so do not consider their limited capacity in responding to external prompts for labeling, which introduces aforementioned burden. This becomes a prevalent issue when the model experiences noisy measurements for an extended period of time [118]. 3) Lastly, they often solely rely on the labels acquired from the user and ignore all other unlabeled data which could potentially be further leveraged in the retraining process [117].

Typically, there are two types of uncertainties surrounding wearable recognition systems. The first is referred to as aleatoric uncertainty, which is data dependent and is related to any noise present in sensor measurements such as sensor displacement and sensor movements with respect to the body. Such uncertainty cannot be mitigated by increasing the training data. The second type is known as epistemic uncertainty which is related to the inability of the model to recognize certain variations of data samples due to the lack of sufficient diversity in the training data. This increases as the model receives input samples with unfamiliar characteristics. This can be mitigated by enhancing the training data, therefore, epistemic uncertainty should be strongly considered for active learning tasks.

In this study, we propose an ADL recognition system with personalization capability. We leverage both active learning and unsupervised learning methods to facilitate the retraining process. For the active learning component, we leverage the uncertainty of the model on its decisions to identify the critical samples where annotations should be requested and used for retraining. We propose a unified Bayesian deep learning framework to model the aforementioned types of uncertainties (aleatoric and epistemic).

5.2. Methodology

5.2.1. Overview

In this study we propose a framework for personalization of machine learning models for the applications of ADL recognition using wearable sensors. This unified framework can quantify different types of uncertainty of wearable sensors in order to provide an effective active learning model. Moreover, it allows for leveraging unlabeled data samples within an unsupervised autoencoder-based model to boost the personalization process. Figure 5.1 illustrates the overall proposed framework. Figure

5.1-a shows the training phase of the model using the labeled training data available from certain users, while Figure 5.1-b shows the personalization procedure.

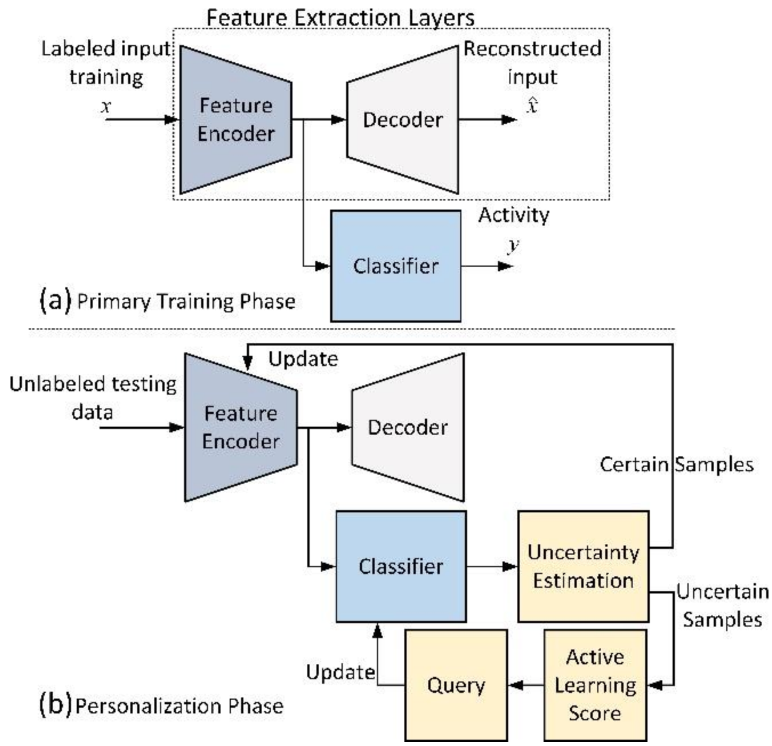


Figure 5.1 - The overall flow of the proposed model for personalizing ADL recognition models

In Figure 5.1-a, the encoder extracts features from the motion signals automatically through multiple convolution layers. The decoder is then responsible to ensure that the features generated by the encoder learn the intrinsic structure of the input data regardless of the associated activity labels. These two components create a variational autoencoder (VAE) framework that is necessary for estimation of data-dependent uncertainty (i.e., aleatoric) (see Section 5.2.3.1). We add a classifier network in addition to this VAE framework to ensure that the extracted features remain discriminative for the ADL recognition task. It is important since the learned features should not only be task specific (discriminative), but also be able to retain the intrinsic structure of the data regardless of their task-specific labels [135].

In Figure 5.1- b, during the personalization phase, the model-dependent uncertainty (i.e., epistemic) is used to identify the vital samples for which we should acquire labels from the user. This uncertainty is used in conjunction with a score function to limit the amount of user interaction while also considering other parameters into account (see Section 5.2.3). Essentially, the critical samples for which the model is highly uncertain are identified by this module and labels are queried. However, the labels are not queried for the samples about which the model is confident. Those unlabeled samples are used to retrain the feature extraction layers in an unsupervised manner (see Section 5.2.4), while the labeled samples are used to fine-tune the classification layers.

5.2.2. CNN for Automated Feature Extraction and Classification

Convolutional neural network (CNN) is a widely used model for different classification tasks due to its ability to extract features automatically. The network is fed with raw signal x, and it maps x to a latent variable z, which serve as the features that are extracted from a raw input signal. In a typical neural network f, we can write $f = h \circ g$ where $g : x \rightarrow z$ maps raw inputs to a higher level feature space z and $h : z \rightarrow y$ is a discriminative function that maps the features to desired class labels. A deep CNN with multiple layers extracts features from raw input signals by using kernels that can be interpreted as filters applied to the signal via a convolution operation. The trainable weights of CNN kernels (W_g) and the weights of the classifier (W_h) are learnt through the training of the neural network.

In classification setting, the network outputs a vector of unaries, where each unary corresponds to a class, and the vector obtained by the concatenation of all the unaries would be passed through a Softmax function to yield an estimation of probability distribution over classes. It has been shown that the Softmax does not provide a reliable and precise estimation of the actual model uncertainty [27], [136]. Equation 5.1 shows the Softmax function where $f_i(x)$ is the network output for i^{th} class before the Softmax layer and N is the total number of classes.

$$softmax\big(f_i(x)\big) = \frac{e^{f_i(x)}}{\sum_{j=1}^{N} f_j(x)} \tag{5.1}$$

5.2.3. Supervised Active Learning

Sd Supervised retraining of classification models can end up with a higher accuracy than unsupervised retraining. However, it creates a huge burden on the user to provide lots of labeled training data for the system. On the other hand, a fully unsupervised retraining paradigm does not require such an extensive data collection and annotation but it does not have an ideal performance. To leverage both advantages, we propose a supervised active learning paradigm to select the most important samples and query the user only for those samples to minimize user's burden; in addition, our proposed system can make use of unsupervised retraining (Section VI). This can reduce the interaction of the system with the user drastically while the performance improvement after personalization would be still significant.

In order to select the most important samples to query from the user, the model needs to understand the uncertainty/confidence about its decision. To be more specific, samples of which the model is not confident need to be identified to fine-tune the classifier parameters [111]. However, it is also important that the model understands if it is uncertain due to lack of training or due to the noisy sensor data. The former, known as epistemic uncertainty, is essential for designing an effective active learning model while the latter is not. In fact, the model should query the label for the samples of which it is uncertain due to the lack of training. However, when the model is uncertain due to temporary noise in sensor data, it is not helpful to use those samples for retraining. Using those noisy samples for retraining could even degrade the performance of the classifier and lead to overfitting. In this section, we first explain our methodology for quantifying different types of uncertainty and then explain how the critical samples are identified to be labeled by the user for model retraining.

5.2.3.1. Uncertainty Quantification

In this section, we propose a unified framework to quantify both aleatoric and epistemic uncertainties in the classification model introduced in Section 5.2.2. To take into account the epistemic uncertainty, which is a measure of the model's uncertainty, we treat the weights of the discriminative function (function h in Section 5.2.2) as random variables instead of deterministic ones. Moreover, to capture the aleatoric, which is a data-dependent uncertainty, we consider the latent variable z as random variables instead of deterministic values. By considering the randomness on the weights and latent variables and treating them as random values and considering their distribution, the final label inference can be written as Equation 5.2.

$$p(y_i|x) = \int p(y_i|W_h, x, z)p(W_h, z|x)dzdW_h \tag{5.2}$$

where $p(y_{i|W,x,z)}$is the likelihood function of i^{th} class, which is calculated as the output of the neural network with a Softmax function in our classification task using Equation 5.1. $p(W_h)$ are $p(z|x)$ is the posterior distribution of weights and features given input data, respectively. The latent variable z in Equation 4.2 is independent of the weights W_h and also W_h is independent of the input data x, so Equation 4.2 can be written as Equation 4.3.

$$p(y_i |x) = p(y_i |W_h, x, z)p(W_h |D)p(z| x)dzdW_h \tag{4.3}$$

where D is the whole training dataset. Calculating the integral in Equation 3 is challenging but it can be approximated through Monte Carlo estimation as shown in Equation 4:

$$p(y_i|x) = \frac{1}{n}\sum p\left(y_i\middle|\widehat{W_h}, x, \hat{z}\right) \tag{4.4}$$

$$\widehat{W_h} \sim p(W_h) \qquad \hat{z} \sim p(z|x)$$

Here, z is used to model the aleatoric uncertainty of the data. W_h distribution, on the other hand, is used to model the epistemic uncertainty.

In order to sample from the weight distribution, Dropout variational inference is a practical approach for approximation inference [27]. In this approach a dropout layer is used after every dense layer, and the dropout is applied at test time to sample from the approximate posterior of the weights (stochastic forward passes, referred to as Monte

Carlo dropout [27]). It has been shown that dropping the weight randomly in the testing time is equivalent to sampling from the distribution of the weights [27]. To accomplish the sampling, in the testing phase, each sample is passed through the network multiple times (*n* times) and in every pass the weights of network are dropped randomly with the probability of p_{drop}. The output of the network for all *n* passes are calculated and the average is interpreted as the Monte Carlo estimation for Bayesian inference.

Estimating the posterior distribution of latent variables/features $p(z|x)$ is then required to complete the calculation in Equation 4.3 and estimate the uncertainty of the model. Note that *z* is an unobserved latent variable while our observation is *x*. The Bayesian analysis can be directly used to calculate $p(z|x)$ as; however, this leads to an intractable integral for calculating the denominator. In order to address this problem, we approximate $p(z|x)$ with a variational distribution $q(z|x)$ from a Gaussian distribution family and try to find the parameters of $q(z|x)$ such that it closely estimates the $p(z|x)$. To find the parameters of the variational estimation $q(z|x)$, we could minimize the Kullback-Leibler divergence (D_{KL}) between the two distributions [137]:

$$\min D_{KL}\{q(z|x)||p(z|x)\}$$

This minimization can then be written as follows:

$$D_{KL}\{q(z|x)||p(z|x)\} = E_{z\sim q}[\log q(z|x) - \log p(z|x)] = \tag{4.5}$$

$$E_{z\sim q}[\log q(z|x) - \log p(x|z) - \log p(z) + p(x)]$$

where $p(x)$ is derived from the expectation as it does not depend on *z*. By rearranging Equation 4.5 we have:

$$\log p(x) - D_{KL}\{q(z|x)||p(z|x)\} = E_{z_i\sim q}[\log p(z|x)] - D_{kl}\{q(z|x)||p(z)] \tag{4.6}$$

To minimize the KL divergence between $q(z|x)$ and $p(z|x)$ we can minimize the right hand side of Equation 4.6. This is exactly the objective function of a variational autoencoder [137]. The first term on the right hand side of Equation 4.6 is the loss of reconstructing input x from the latent variable z and the second term is the divergence between the variational approximation with the prior distribution of z. For this prior distribution we use a standard Gaussian distribution with the mean of zero and variance of one. Therefore, by training a variational autoencoder and leveraging the latent variable z as the features that are provided to a classifier, we can approximate the posterior distribution of the features to accomplish the inference in Equation 4.4. In a typical VAE, the encoder estimates the variational approximation $q(z|x)$ and the decoder estimates the first term on the right-hand side of Equation 4.6. The output of the encoder in a VAE is the mean and standard deviation that serve as the parameters of a Gaussian distribution that models the posterior distribution of z given x. Given the posterior distribution of features, we can sample from it to calculate the Monte Carlo estimation of Equation 4.4.

A typical VAE containing convolutional layers in the encoder can extract informative features from a signal in an unsupervised manner. In a typical VAE framework, the only concern is to extract features that can retain the structure of the input data. However, in a supervised classification problem the features should be discriminative regarding the labels given in the training data. Based on this intuition, we propose a new framework of deep neural network as shown in Figure 5.2 by modifying the typical VAE objective function as follows:

$$L = p(y|x) + E_{z \sim q}[\log p(z|x)] - D_{kl}\{q(z|x)||p(z)] \quad (4.7)$$

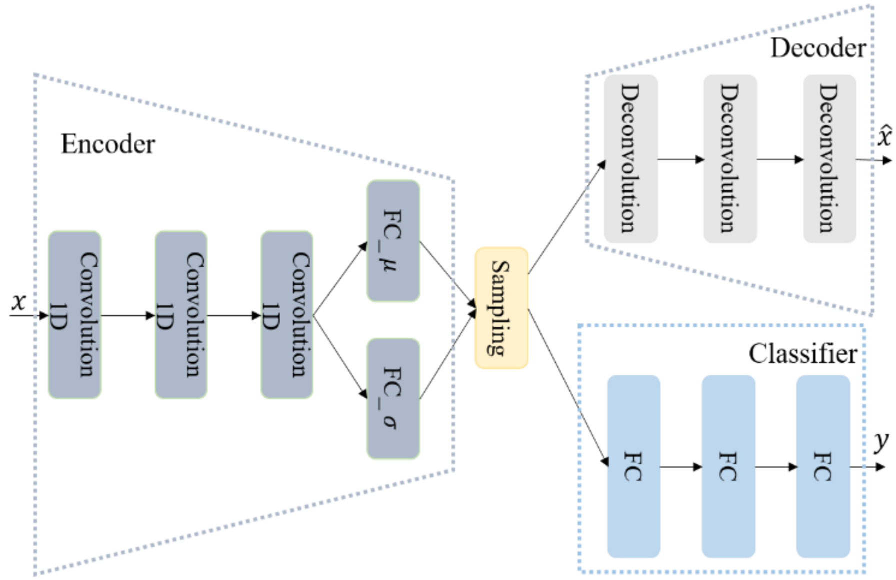


Figure 5.2 - The architecture of the proposed neural network for feature extraction and uncertainty estimation

In fact, maximizing the *p(y|x)* is added to the typical VAE objective function in Equation 4.6 to produce the new objective function in Equation 4.7. This objective function guides the VAE to produce latent features that not only can reproduce the input data but also discriminate between different class labels. The encoder, which serves as a feature extractor, estimates the parameters of the posterior distribution of the features. The decoder ensures that the latent variable *z* is able to retain the structure of the input data, and is discarded after the training. The classifier samples from the distribution of the features, which is approximated by the encoder, and maps those samples to the class labels.

It should be mentioned that this framework can be considered in line with the idea of pre-training an autoencoder in an unsupervised manner and replacing the decoder with a classifier to improve the performance [138]. The authors in [138] share this observation that features created by the autoencoder are a good representative for training datasets that support better generalization of the trained models [139]. What distinguishes our work is that we embed the two processes of the classifier learning and the data-dependent feature extraction in a single framework, which improves the discriminative power of our features compared to the case of unsupervised pre-training.

The procedure for estimating the label and the confidence of the classifier is shown in Algorithm 5.1 for the trained neural network as shown in Figure 5.2. In Algorithm 5.1, *OneHotEncoding(.*) is a function that returns the one hot encoding of a vector and *std(.)* calculates standard deviation. To predict the class label for each input data, the *n* samples are acquired from the distribution of the features, the weights of the classifier are dropped randomly, and labels are generated by the classifier. *n* is a hyperparameter of the model that is determined empirically through cross-validation in the training phase. In our datasets *n*=100 ended up with the best cross-validation accuracy. The final decision of the classifier is the average of the outputs. Moreover the standard deviation of the generated outputs is the measure of uncertainty. Intuitively, the classifier would generate more consistent labels for the samples that it is confident on, while for non-confident samples it would generate distinct labels that leads to higher standard deviations. The uncertainty calculated by Algorithm 5.1 is called the combined uncertainty as it contains both aleatoric and epistemic uncertainties. To only consider the epistemic uncertainty, we use the mean of *z* in step 4 of Algorithm 5.1, instead of sampling from the posterior distribution of the features. In fact, in this uncertainty we only consider the randomness of the weights of the classifier network and do not care about the randomness of *z*, which explains the aleatoric (i.e., data-dependent) uncertainty. On the other hand, to only consider the aleatoric uncertainty, we do not drop the weights during the testing time as shown in step 3 of Algorithm 5.1, which leads to ignoring the randomness on the model's weights. Leveraging this approach, we establish a method where we discard the model uncertainty and we only take into account the uncertainty over latent variables *z* which is data dependent.

Algorithm 5.1 Label and uncertainty estimation

Input: test data x, encoder network g, classifier network h, p_{drop}, number of labels N, parameter n
Initialize *prediction = zeros(n,N)*
Output: classifier decision $\hat{y}$, *Uncertainty*
1- ***for*** *j = 1* **to** *n* **do:**
2- Take a sample $z_j \sim g(x)$
3- Drop weights of h with the probability of p_{drop}
4- $y^j = h(z_j)$ // the output of Softmax function
5- *prediction [j,:] = OneHotEncoding(y^j)*
6- **end for**
7- $\hat{y} = \frac{1}{n}\sum_{j=1}^{n} y^j$
8- *Uncertainty = std(prediction)*

5.2.3.2. Deep Neural Network Implementation

In this section, we discuss the details of all neural networks for various components used in this study. The detail of all encoder and classification layers are presented in Table 5.1. For the encoder network as shown in Figure 5.2, we use three layers of CNN followed by one fully connected (FC) layer for each of mean and standard deviation estimation. Based on our experiments using less number of layers did not offer an acceptable accuracy and using more layers does not offer significant improvement in the performance of the system while it increases model complexity. For the classifier network we used three fully connected layers. The decoder contains three deconvolution layers. Re-parametrization trick is used for t handling the sampling from a Gaussian distribution when training the network with backpropagation algorithm.

In the preprocessing phase, the data is normalized to retain zero mean and single variance (centered and scaled) and segmented prior to supplying it into the CNN. We utilize a fixed-size window with a length of one seconds and overlap of 50%.

Table 5.1 - Characteristics of the proposed neural networks

	Layer	**# of kernels/ neurons**	**Activation function**
Encoder	Conv2d_1	32	ReLU
	Conv2d_2	64	ReLU
	Conv2d_3	100	ReLU
	FC_mean	20	Sigmoid
	FC_std	20	Sigmoid
Classifier	FC_1	64	ReLU
	FC_2	128	ReLU
	FC_3	200	ReLU
	FC_classifier	Same as the # of classes	Softmax

5.2.3.3. Identifying Critical Samples

Our primary goal is to determine the most important samples that can contribute to improving the retraining process for model personalization. Here, we define such samples as the ones of which the model is uncertain due to the lack of training data.

Therefore, we need to identify the source of uncertainty for test samples and pick the ones with maximum epistemic uncertainty. During the training phase, we calculate the average combined, aleatoric, and epistemic uncertainties over all the correctly and misclassified samples shown as σ_{cor}^{comb}, σ_{cor}^{ale}, σ_{cor}^{epi}, σ_{mis}^{comb}, σ_{mis}^{ale}, and σ_{mis}^{epi} respectively. For a test sample we also calculate the three uncertainties calling them as σ_{test}^{comb}, σ_{test}^{ale}, σ_{test}^{epi} . The process of identifying samples starts by picking samples that have combined uncertainty above a threshold. This threshold is determined empirically. Afterwards, we calculate the ratio between the increase in epistemic to the aleatoric uncertainty compared to the correctly classified samples as shown in Equation 4.8.

$$\sigma_{ratio} = \frac{\sigma_{test}^{epi} - \sigma_{cor}^{epi}}{\sigma_{test}^{ale} - \sigma_{cor}^{ale}} \tag{4.8}$$

However, the uncertainty is not the only parameter to consider to determine when labels need to be solicited. Using merely the uncertainty might lead to querying the user too frequently. To further control this process, we define a parameter q_{limit} that indicates the total number of questions allowed to be asked within a certain period of time. The system keeps track of how many questions it has asked so far as $q_{inquired}$. The model becomes stricter in querying the user as the number of questions previously asked increases. In fact, the likelihood of asking more questions should decrease as the model asks more question. We model this behavior as a linear function of the number of remaining queries. The final score function to select samples to be queried from the user is shown in Equation 4.9.

$$s = \sigma_{ratio} + \beta \frac{q_{limit} - q_{inquired}}{q_{limit}} \tag{4.9}$$

where s is the score assigned to each sample, and β is a tuning constant that is determined empirically during the training phase. In addition to the aforementioned parameters, a good sample to query from the user is the one with higher occurrence. Note that for ADL recognition we need to segment the sensor data into windows of a fixed length. Therefore, a label queried from the user can be typically assigned to one segment of data. However, we argue that if there are multiple consecutive segments of the data of which the system is consistently uncertain, then the queried annotation can be used to label all those segments with a higher chance that the label applies to all segments due to the expected consistently in temporal distribution of ADLs. To incorporate this assumption and observation, at each time step, we calculate the proposed score s for all segments within the last five minutes and query the user only if the value is higher than the threshold for all of them. The threshold can be set empirically through cross validation in the training phase for online active learning.

Lastly, the labeled data queried from the user is used to fine tune the weights of the classifier. In fact, when using these labeled data, we freeze all the weights of the encoder and decoder networks, and only retrain the weights of the classification layers. This is important because the amount of the labeled data is very small and is not enough to retrain all the feature extraction layers. Moreover, it is shown that usually the first layers, which are responsible for feature extraction, are more generalizable and transferable, while the last layers are more task specific and less transferable between different users.

5.2.4. Unsupervised Retraining

Scarce-labeled data is not sufficient to retrain all feature extraction layers including the encoder and the decoder networks since they require to be extensively trained to learn appropriate features of motion signals. However, due to our autoencoder structure, we can leverage the huge amount of unlabeled data to retrain the encoder component which produces the hidden state that represents the patterns and features of the new data. When this occurs, we keep the weights of the classification layers unchanged as the encoder/decoder weights are updated. This unsupervised retraining allows the feature extraction layers to adapt to the patterns and morphology of the signals of a new user.

When combining this unsupervised retraining of the feature extraction layers, with the supervised fine-tuning of the classification layers through the active learning process, the whole network is adapted effectively to the data of a new user. So, feature extraction layers are updated to capture the patterns of the new signal, while the classification layers are updated to learn how to map those features to the desired activity classes. Consequently, we must first accomplish the unsupervised retraining to update the encoder-decoder weights before we perform the supervised fine-tuning of the classifier's weights using the labels gathered by the active learning module.

5.3. Results

S We evaluate the effectiveness of our methods for personalization of deep learning ADL recognition system on the new users. In this section, we start by introducing the datasets used for the evaluation and then presenting the performance of our personalization method on detecting ADLs and compare it to baselines and state-of-the-art approaches. Afterwards, we deeply analyze our uncertainty modeling by investigating its behavior in response to different sources of uncertainty to understand how it can improve the effectiveness of active learning designs. Finally, we investigate how personalization accuracy is affected upon presence/absence of each component of the proposed framework, including the supervised fine tuning of classification layers through active learning and the unsupervised retraining of the feature extraction layers. The purpose of this investigation is to show the importance of each of those components in the personalization process.

To demonstrate the effectiveness of our proposed framework, we used two publicly available datasets including PAMAP2 [140], and MoST [141]. PAMAP2 comprises of 18 physical activities measured by three wearable inertial measurement units (IMUs) with the sampling frequency of 100 Hz performed by 9 different subjects. IMUs were worn on three different body parts: the wrist of the dominant hand, the chest, and the ankle of the dominant foot. In this study, we used eight out of 18 activities, which have the greatest number of samples. We used 3D acceleration and gyroscope sensors that results in 18 axes of data. MoST dataset, collected by our own group, contains 23 daily activities captured by six IMUs working at the frequency of 200 Hz placed on the arm, wrist, chest, ankle, and both legs. The data was collected from 20 healthy subjects. Since, several activities in this dataset are similar, we grouped them as one activity and once again, removed the classes with small training data. Table 5.2 represents the list of activities used in this study.

Table 5.2 - Set of activities

	PAMAP2	MoST
	Biking	Sit-to-stand
	Sitting	Sitting
	Standing	Standing
	Walking	Walking
	Stair climbing	Grasping floor
	Lying down	Lying down
	Running	Turning 90°
	Rope jumping	Jumping
# of samples	17000	7500

5.3.1. Personalization Results

We assume that there is a large amount of labeled training data available for certain subjects, called primary subjects, that can be used to train the initial ADL recognition model. We aim to personalize the model that is trained on the primary subjects to a new subject to achieve the highest performance possible. For each dataset, we exclude one subject as the new subject and use the remaining subjects as the primary subjects to train the initial model. For the excluded subject, i.e., the test subject, we used 50% of the data for personalization (i.e., retraining) and used the remaining 50% for testing the accuracy of the personalized model. We repeat the experiments by changing the excluded subject to cover all the subjects. In other words, in our experiments, each subject has been treated as the new subject once. All reported results in this section are the average over all repetitions.

Table 5.3 shows the results of comparing our model to other baseline and existing methods regarding personalization accuracy. The first row in Table 5.3 represents the accuracy of the ADL recognition when the model is trained on primary subjects and it is tested on the same subjects. This shows the upper bound of training, when the testing and training data come from the same subjects. The second row shows the performance of the system when it is trained on primary subjects but it is tested on a new subject with no retraining and/or personalization of the model. This shows how the performance drops when the model is used for a new subject and also emphasizes the necessity of personalization in such systems. The third row shows the results of using all the labeled data of the new user for personalization through retraining of the whole neural networks. This case represents another upper bound, which achieving that in real-world scenarios is impossible due to high burden on the users. The fourth row represents the performance of personalization when the labels are queried for random samples. This is a baseline to show why smart active learning is required. The fifth row represents another baseline for active learning by using the output of Softmax function as a measure of uncertainty. In this approach, a query for a label is submitted to the user when the output of the Softmax is less than a threshold. Based on our experiments setting that threshold to 0.7 obtains the best results in our study. The sixth row of the table shows the results of a state-of-the-art active learning method using entropy. Based on this model, the samples with the highest entropy are the most informative samples for the classifier and their label are acquired from the user [117]. The seventh row shows another state-of-the-art method that uses the agreement between the labels generated by different trees in a random forest classifier as a measure of uncertainty to select most uncertain samples for label solicitation [117]. Finally, the last row shows the performance of our proposed method for personalization. It should be noted that in this experiment, the total number of questions allowed to be asked is set to 200 for all the algorithms, which means that we query the labels for 200 samples of new data.

As Table 5.3 shows, our method outperforms all the baseline methods by 14.3% and existing methods by 8%. This shows the superiority of our method that is due to more effectively quantifying the uncertainty and the unsupervised retraining of feature encoder weights. In fact, we distinguish between different types of uncertainty and try to ask questions when the system is uncertain due to lack of training rather than facing noisy sensor measurements. Figure 5.3 reveals the importance of differentiating between different types of uncertainties for more effective active learning. For this analysis we add a synthetic Gaussian noise to our sensor measurements to understand the importance of the uncertainty decomposition. We add different level of noise to different data samples using Equation 4.10.

$$x_{noisy} = x_{clean} + \alpha.\epsilon \tag{4.10}$$

$$\epsilon \sim N(0,1) \qquad \alpha \sim Uniform(0,0.1)$$

where α is the amplitude of the noise that is different for each data sample, and s is a white noise. As Figure 5.3 shows, querying the label for the samples that are chosen based off of epistemic (i.e., model dependent) uncertainty provides highest improvement compared to using the combined or aleatoric (i.e., data dependent) uncertainty. Moreover, the ratio proposed in our method in Equation 4.8 achieves the highest personalization performance. The reason for this is as follows: the two uncertainty metrics have a complementary nature to some extent meaning that in presence of novel data, both epistemic and aleatoric uncertainties increase. However, the increase in epistemic uncertainty in this case is much more significant than the aleatoric uncertainty as the novel data is associated with model dependent uncertainty (see Section 5.3.2).

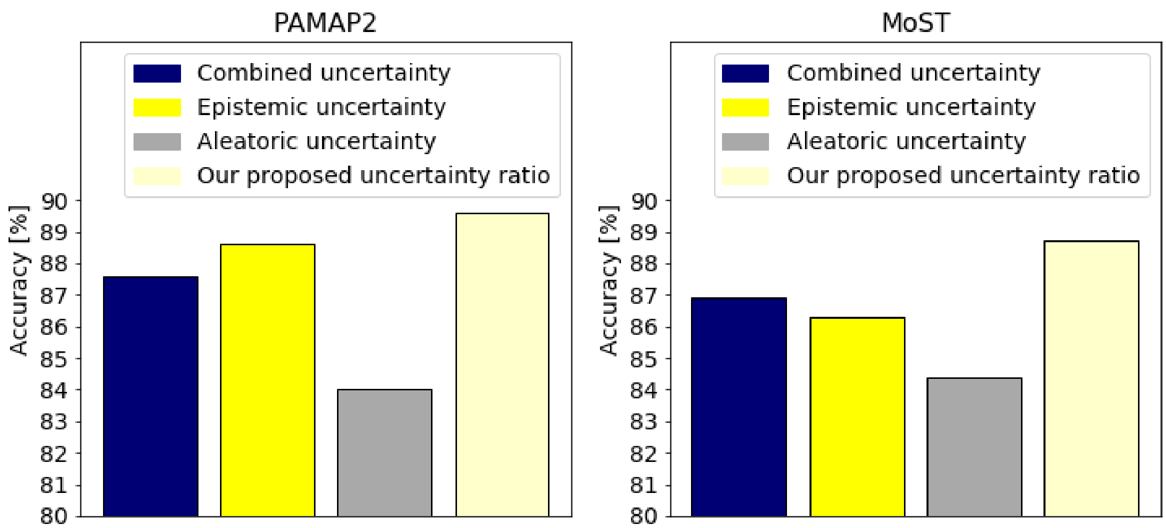


Figure 5.3 - Effect of using different uncertainty metrics on personalization

The uncertainty ratio proposed in Equation 4.8 takes this observation into account and helps the model to better understand if it is facing uncertainty due to lack of training rather than noisy measurements. In conclusion, Figure 5.3 provides the indication on the importance of distinguishing between the types of uncertainty for active learning tasks.

Table 5.3 - Personalization accuracy [%]

Method	PAMAP2	MoST
Primary subjects	94.2	91.8
New subjects with no retraining	53.7	64.6
Use all labeled data for retraining	93.8	91.9
Query labels for random samples	76.2	70.0
Active learning with Softmax	74.9	77.2
Entropy-based active learning [117]	79.3	82.3
RF-based active learning	80.0	81.4
Our personalization method	**89.6**	**88.7**

Figure 5.4 depicts the performance of personalization vs. the number of queries form the user averaged over all the subjects within each dataset. As the figure shows, our proposed method achieves the highest accuracy compared to the other methods, especially when the number of questions is very small. There are two reasons: first, our method attempts to choose the most important samples by taking into account the limitation on the number of questions; second, it uses unlabeled data along with the queried labels to adapt the feature extraction layers to the new data. Based on Figure 5.4, by querying at least 60 and 140 data points from the user, in MoST and PAMAP2 datasets, respectively, our method can achieve more than 20% improvement in the performance of the system compared to the case of using no personalization for the new users. Moreover, the acquired accuracy from the personalization by querying only 200 data points from the new user is only 3.7% less than the upper bound of using all labeled data from the new user (i.e., 1700 datapoints in PAMAP2 and 400 data points in MoST dataset on average).

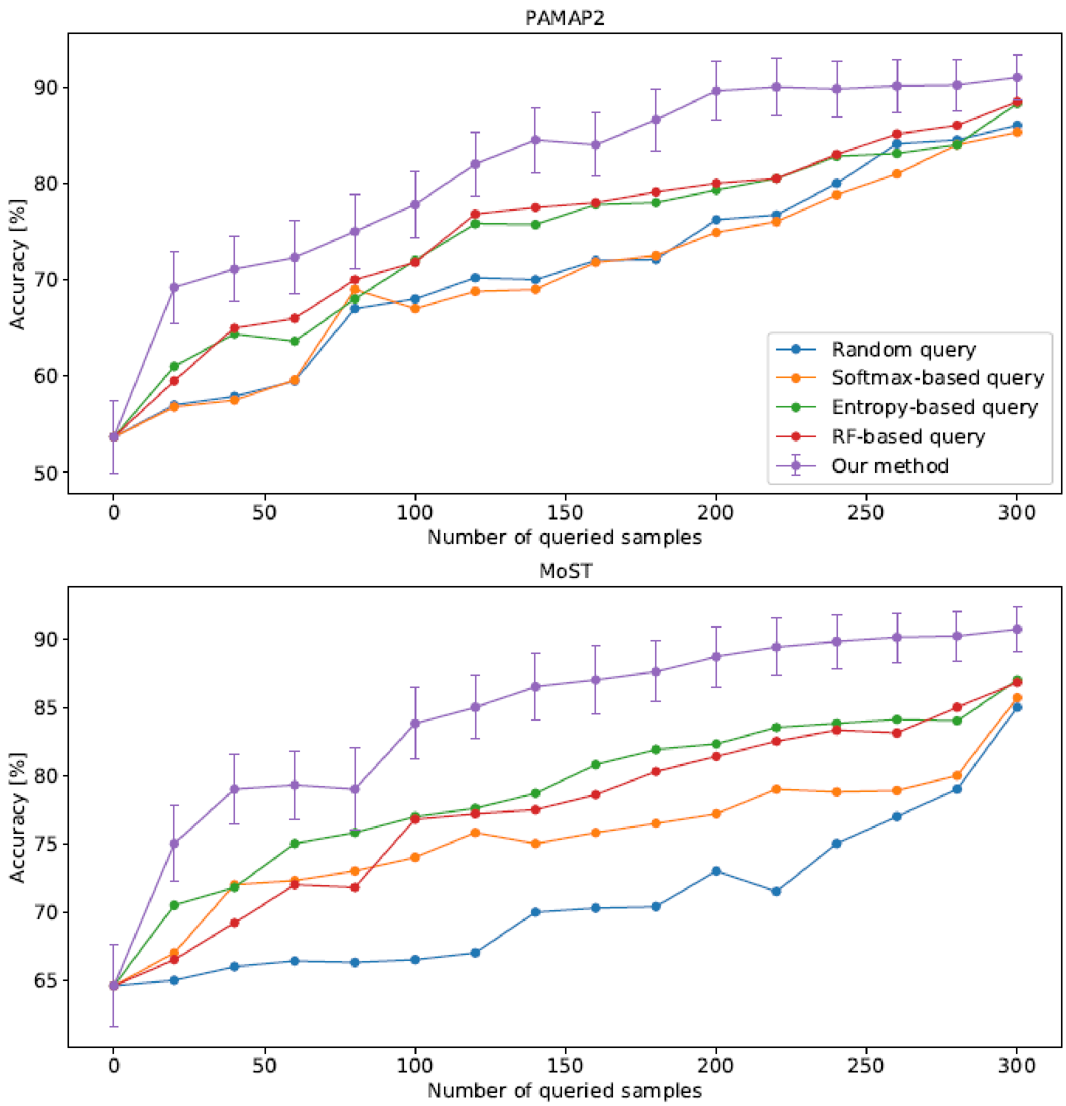


Figure 5.4 - Personalization performance vs. number of queried samples through

5.3.2. Uncertainty Metrics

To assess the quality of our uncertainty measurement technique and to better understand how they are associated with different sources of uncertainty in sensor data, we investigate three questions:

- How do epistemic and aleatoric uncertainties change when the system is dealing with novel data?
- How do these types of uncertainty behave when dealing with noisy sensor measurements? With last two questions, we investigate whether these two types of uncertainties are separable.
- Is there a relationship existing between these uncertainties and misclassifications?

First, we analyze the behavior of the quantified uncertainties when the system is presented with novel data. The novel data is the data drawn from a different distribution compared to the training data, and it could come from new activities for which the system has not been trained, or new users that perform activities differently. A good measure of uncertainty must increase when the system is provided with novel unfamiliar data. Herein we compare the uncertainty in three different cases including testing on data same as the training data, data of new subjects, and data of new activities, for which the results are shown in Figure 5.5. In the first case (the navy bar in Figure 5.5), we test the model with data from the same distribution as the training (non-novel data). In other words, the model is tested on the data of the same subjects and with the same activities as in the training set. In this figure, it is seen that all combined (Figure 5.5-a), epistemic (Figure 5.5-b) and aleatoric (Figure 5.5-c) uncertainties are smaller compared to when the model is tested on the data of new subject or new activities (novel data). To test the model on novel data, in the second case (the yellow bar), the model is trained on all but one subject and is tested upon the data of the excluded (new) subject. In the last case (the gray bar), we test the model with the data of new activities that were not used in the training.

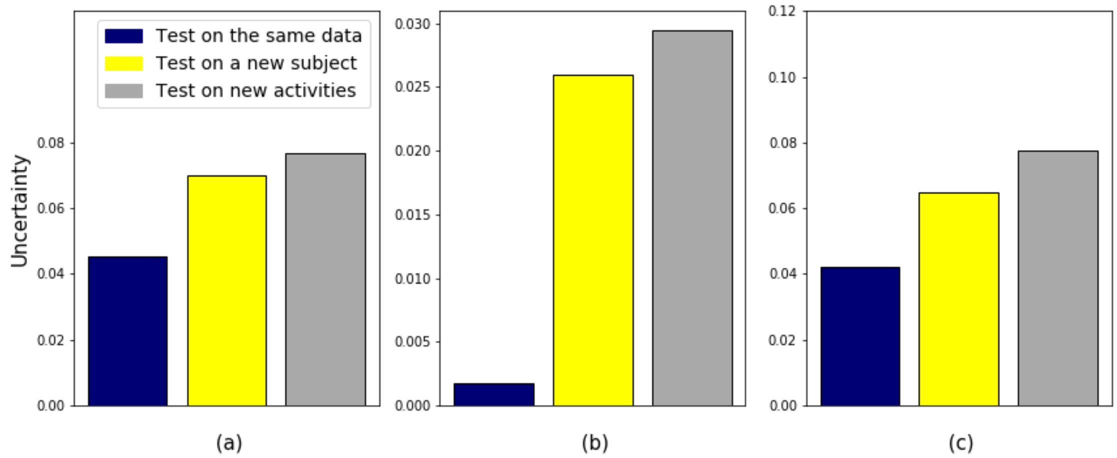


Figure 5.5 - The uncertainty when the model is tested on novel data (averaged over all the subjects). (a) combined uncertainty; (b) epistemic uncertainty; (c) aleatoric uncertainty

Expectedly, as Figure 5.5 shows, all combined (Figure 5.5-a), epistemic (Figure 5.5-b) and aleatoric (Figure 5.5-c) uncertainties increase when the model faces novel data (yellow and gray bars in comparison to the navy bar). However, this increase is much more significant in the epistemic uncertainty (Figure 5.5-b) compared to the aleatoric uncertainty (Figure 5.5-c). This confirms the initial hypothesis about the type of uncertainties and it is in line with prior reports [142]. In fact, it shows that the epistemic uncertainty, which considers the model uncertainty, is much more sensitive to the lack of training data compared to the aleatoric uncertainty. By measuring this uncertainty, we can realize if the input data is not familiar for the model. In such cases, it is reasonable to solicit the user to get more information about the activity label for retraining the model

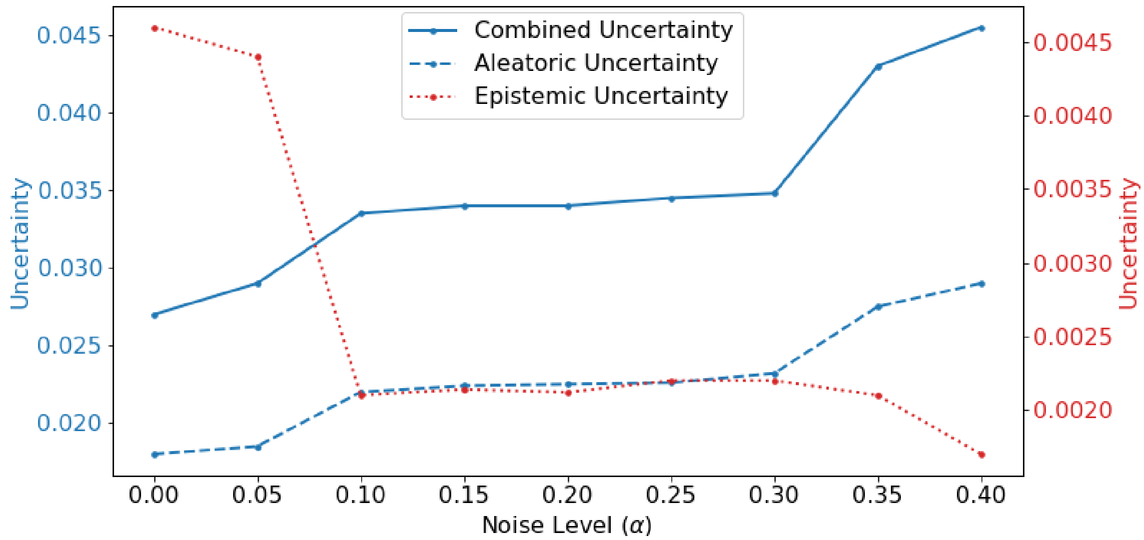


Figure 5.6 - The uncertainty of the model in presence of different noise levels averaged over all the subjects). X axis corresponds to values of α in Equation 4.10

Second, we expect the uncertainty to increase in response to noisy sensor measurements too. Thus, the second experiment is devoted to the analysis of the effect of sensor noise on each type of uncertainties quantified in this study as shown in Figure 6. We seek to understand whether the uncertainty sourced by the noisy data is distinguishable from the uncertainty of the model sourced by lack of sufficient training data. We add synthetic noise to data where varying level of noise is represented in the X axis of Figure 5.6. Each axis of raw sensor data is corrupted with a Gaussian noise, as described in Equation 4.10, with the only difference here being that we use various constant values for α as depicted on the X-axis of Figure 5.6. It should be noted that we chose the range [0,0.4] for α to better show how the uncertainty changes in response to different levels of noise in the data. We did not use larger values for α because the noise dominates the signal and the data becomes non-informative. As Figure 5.6 shows, the aleatoric uncertainty (dashed navy line), which corresponds to the data, increases consistently as the noise magnitude (and power) increases. Contrarily, the epistemic uncertainty (dotted red in Figure 5.6) does not show an increasing pattern as consistent as the aleatoric uncertainty. However, we naturally expect to see an increase in the uncertainty when the system is fed with noisy data, and that is what is observed with the aleatoric as well as the combined uncertainty, which is dominated by the aleatoric. Hence, the decreasing pattern in the epistemic uncertainty, as shown in Figure 5.6, indicates that this type of uncertainty is not measuring the true data-dependent uncertainty due to noisy signals. Combined uncertainty (solid navy line in Figure 5.6), similar to aleatoric uncertainty, shows a steady increase which is desired since the noisy data introduces challenges for the predictions. This shows that the uncertainty increases when the system is presented with noisy sensor measurements while it can still capture the fact that this uncertainty is not caused by the model not knowing the data but because the data is noisy.

Figure 5.5 and Figure 5.6 show that the proposed measures of uncertainty are sensitive to different sources of uncertainty and this allows the system to understand which type of uncertainty is being observed. Now we investigate the change in the uncertainties when the output of the model is correct versus when the predictions are incorrect. We expect to observe an increase in the uncertainty when the classifier makes a mistake. If this appears to be the case, then the system will be able to detect potential errors and appropriately intervene. Figure 5.7 depicts the aleatoric, epistemic and combined uncertainties, averaged for both correctly and incorrectly classified data samples. According to the figure, the uncertainties in the correctly classified data samples (the navy bar) are steadily lower than the ones that are misclassified (the yellow bar). This indicates that the model makes more mistakes on the data samples on which it is less certain. Therefore, the uncertainties developed here can serve their true purpose.

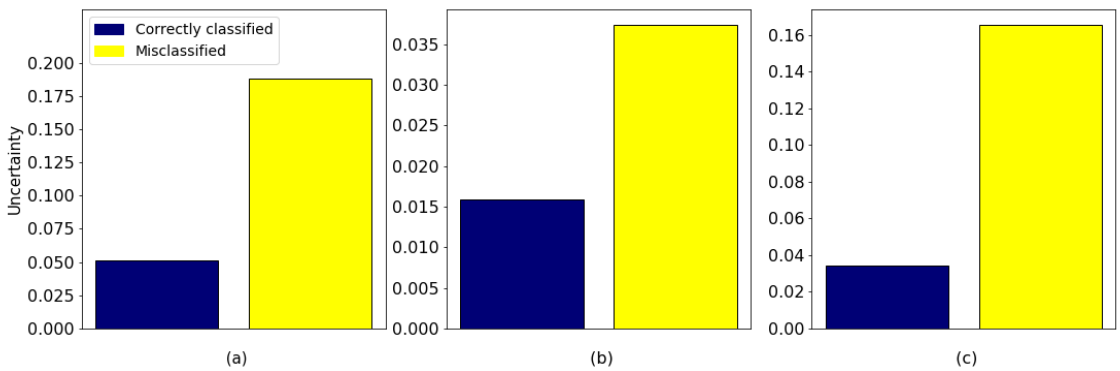


Figure 5.7 - Comparing the uncertainty for the samples that are correctly classified by the classifier and the samples that are misclassified (a) combined uncertainty; (b) epistemic uncertainty; (c) aleatoric uncertainty

To further investigate the effectiveness of the proposed uncertainty metrics, we compare it to the Softmax output. To accomplish this comparison, we first determine 0.075 as a threshold on our combined uncertainty for distinguishing between certain vs. uncertain data samples based on Figure 5.7-a. Surprisingly, 45% of all misclassified samples that are labeled as uncertain by our system (i.e., their uncertainty is above the threshold) have a Softmax output of higher than 0.95. This shows that, even the misclassified samples that are not located close to the decision boundary of the neural network (i.e., Softmax is very certain about them) can be detected by the proposed uncertainty metric. In other words, the proposed uncertainty is capable of capturing the uncertainty even for the samples that are far from the decision boundaries of the classifier. Moreover, from all the misclassified samples, only 13% have the Softmax output of below 0.95, which shows the inability of Softmax in detecting uncertainty corresponding to misclassification, while 47% of them have been detected as uncertain by our model. This shows the superiority of our algorithm over Softmax regarding detecting the uncertainty to mark misclassified samples.

5.3.3. Evaluating Components of the Proposed Model

The proposed personalization framework for human ADL recognition consists of unsupervised retraining of the feature extraction layers and supervised fine-tuning of the classification layers using the labels that are acquired by the active learning. In this section, we aim to analyze the effect of each of those two components on the personalization performance. Figure 5.8 compares the personalization performance of our method when various components are disabled. As the figure shows, in one hand, the accuracy of the ADL recognition for new user (i.e., personalized model) drops by 5.4% when we do not retrain the feature extraction layers in an unsupervised manner, and we only rely on supervised fine-tuning of the classification layers. On the other hand, the performance drops by 9.5% when we ignore the supervised fine-tuning with the labels acquired by active learning, and we only rely on unsupervised retraining. This demonstrates the importance of supervised fine-tuning in personalization. Overall, Figure 5.8 shows that the unsupervised retraining of feature extraction layers and the supervised fine-tuning of the classification layers using the labels acquired by active learning are complementary, and both techniques are required to obtain effective personalization.

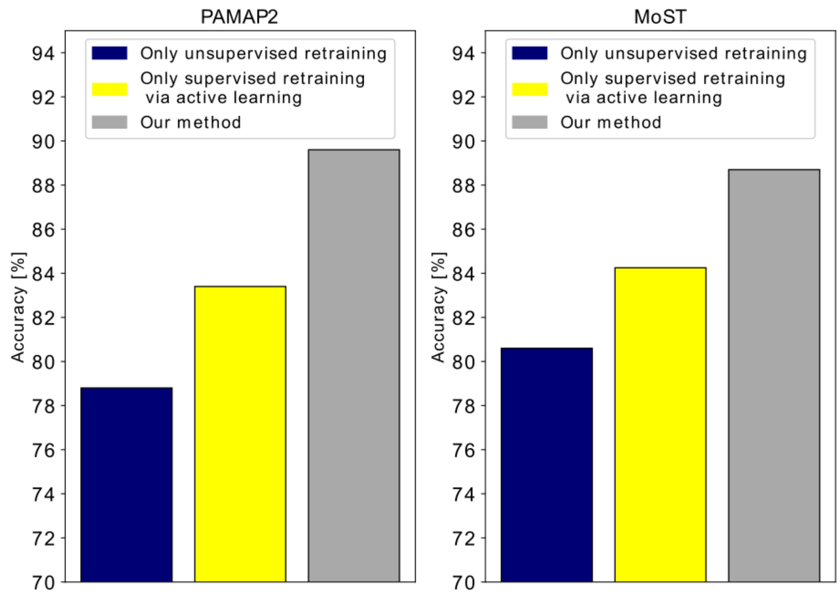


Figure 5.8 - The effect of different components of the proposed method on personalization performance

5.4. Conclusion

We proposed a personalization framework for ADL recognition using deep learning. The proposed method consists of supervised fine-tuning of classification layers as well as unsupervised retraining of feature extraction layers. For the supervised fine-tuning we proposed an active learning technique to acquire labels for most important samples. To achieve a more effective supervised active learning, we designed a unified deep Bayesian neural network to detect different types of uncertainties. Through experimental analysis, we demonstrated how data-dependent and model-dependent uncertainties could be distinguished and measured by the proposed method. We leveraged the model dependent uncertainty to identify the samples that are important to gain maximum accuracy through fine-tuning the classification model. Our experiments showed that in general, personalization is critical when an ADL recognition system is used for a new user. Moreover, using unlabeled data as well as labeled data acquired by active learning, understanding the source of uncertainty, and limiting the amount of interaction with the user to solicit labels while designing the active learning method are vital components to achieve the maximum personalization performance while minimizing the burden on the users. The proposed method is important to improve usability of ADL recognition systems that provide important contextual information for many mobile health and wellness service applications such as patient monitoring, assisted living, dietary and fitness monitoring.

6. REAL-WORLD EXPERIMENTATION: DISEASE ONSET PREDICTION WITH WEARABLES

The details of this study cannot be revealed per Texas A&M Engineering Experiment Station (TEES) instructions due to confidentiality concerns.

7. CONCLUSIONS

This dissertation document described the research that tackled three practical challenges associated with remote health and wellness monitoring through wearables including context measurement, personalization, and large-scale machine learning deployment for disease prediction.

To address the need for measuring context for health monitoring, we used freely available nearables, i.e., passively observable Bluetooth-enabled devices in the user's daily environment, to estimate users' location and interaction opportunistically. We leveraged that contextual information for detecting complex activities of daily living that cannot be detected merely with motion sensors. We designed an offline context model training method by leveraging BLE devices that are observed consistently in the vicinity of the user when a particular action of the user, i.e., activity, is being performed. This consistency was used to generate a set of context patterns that were then mapped probabilistically back to the actions to aid performance via a reduction in the set of probable outcomes. We collected data from several participants over more than one month to train and validate our model. A significant improvement was achieved in detecting complex activities such as shopping, meeting and cleaning by leveraging context to narrow down the search space of activities. An important feature of our proposed approach is the fact that it requires no extra infrastructure and no user intervention for deployment of new sensors and devices. The proposed technique enhances the capabilities of wearable sensors by helping them understand their working environment, which is a vital information to better interpret physiological data.

Personalization is another vital factor in machine learning algorithms designed for healthcare data due to inter-subject variabilities. These variabilities exist in both physiological as well as behavioral and contextual parameters. We studied the notion of personalization for detecting ADLs with motion sensors. We designed a Bayesian deep convolutional neural network with stochastic latent variables to estimate both aleatoric (data-dependent) and epistemic (model-dependent) uncertainties in recognition task. We leveraged these uncertainties to design an efficient active learning model that chooses the most important and informative samples to solicit the user to label them. These labeled samples are used to fine-tune the activity classifier. In addition, the feature extraction is performed automatically in this framework through the stochastic latent variable. The specific design of this module based on variational autoencoder enables opportunity for leveraging unsupervised/unlabeled data to update the model too. These two components, namely active learning for acquiring labeled data for supervised retraining and the capability for unsupervised retraining, significantly improved the performance of the model to adapt to new users with minimum number of labels solicited from the users (in comparison to the state-of-the-art). We also showed how distinguishing between the two aforementioned sources of uncertainty leads to more effective active learning paradigms. The proposed method is important not only to improve the performance of the machine learning models but also to enhance user compliance since it allows them to adopt new devices with minimum intervention required. Therefore, the proposed method improves usability of ADL recognition systems that provide important contextual information for many mobile health and wellness service applications such as patient monitoring, assisted living, dietary and fitness monitoring.

With addressing context-awareness and personalization needs for wearable sensing, to show effectiveness of wearables in a real-world remote health monitoring application, we investigated the changes in continuous physiological data collected by wearables in day-to-day life and showed that wearables can effectively predict onset of certain diseases. *Further details of this work cannot be revealed per Texas A&M Engineering Experiment Station (TEES) instructions due to confidentiality concerns.*

Findings of this work can help alleviate the most important roadblocks to continuous, pervasive remote health and wellness monitoring through wearables. It can also unlock new sensing and data analytics paradigms that improve quality of healthcare services and makes them accessible to underserved population.

REFERENCES

[1] E. Mynatt *et al.*, "Research opportunities and visions for smart and pervasive health," *arXiv Prepr. arXiv1706.09372*, 2017.

[2] S. S. Thomas, V. Nathan, C. Zong, K. Soundarapandian, X. Shi, and R. Jafari, "BioWatch: A noninvasive wrist-based blood pressure monitor that incorporates training techniques for posture and subject variability," *IEEE J. Biomed. Heal. informatics*, vol. 20, no. 5, pp. 1291–1300, 2015.

[3] "Polar M600 | Sport smart watch for fitness | Polar Global." .

[4] "Oura Ring: Accurate Health Information Accessible to Everyone." https://ouraring.com/ (accessed Mar. 30, 2021).

[5] M. Esfahani, M. Iman, and M. A. Nussbaum, "Classifying Diverse Physical Activities Using Smart Garments," *Sensors*, vol. 19, no. 14, p. 3133, 2019.

[6] "Garmin International | Home." https://www.garmin.com/en-US/ (accessed Mar. 30, 2021).

[7] A. Aygun and R. Jafari, "Robust Heart Rate Variability and Interbeat Interval Detection Algorithm in the Presence of Motion Artifacts," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2019.

[8] V. Nathan and R. Jafari, "Particle filtering and sensor fusion for robust heart rate monitoring using wearable sensors," *IEEE J. Biomed. Heal. informatics*, vol. 22, no. 6, pp. 1834–1846, 2017.

[9] V. Nathan and R. Jafari, "Reducing the noise level of EEG signal acquisition

through reconfiguration of dry contact electrodes,” in *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, 2014, pp. 572–575.

[10] G. Sprint, D. Cook, R. Fritz, and M. Schmitter-Edgecombe, “Detecting health and behavior change by analyzing smart home sensor data,” in *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2016, pp. 1–3.

[11] H. Ghasemzadeh, R. Fallahzadeh, and R. Jafari, “A hardware-assisted energy-efficient processing model for activity recognition using wearables,” *ACM Trans. Des. Autom. Electron. Syst.*, vol. 21, no. 4, p. 58, 2016.

[12] E. Šabić, D. Keeley, B. Henderson, and S. Nannemann, “Healthcare and anomaly detection: using machine learning to predict anomalies in heart rate data,” *AI \& Soc.*, vol. 36, no. 1, pp. 149–158, 2021.

[13] A. Akbari, J. Wu, R. Grimsley, and R. Jafari, “Hierarchical signal segmentation and classification for accurate activity recognition,” in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 2018, pp. 1596–1605.

[14] A. Kumar *et al.*, “Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock,” *Crit. Care Med.*, vol. 34, no. 6, pp. 1589–1596, 2006, doi: 10.1097/01.CCM.0000217961.75225.E9.

[15] J. Martinez, A. Akbari, K. Sel, and R. Jafari, “Strategic Attention Learning for Modality Translation,” in *EEE International Conference on Acoustics, Speech,*

and Signal Processing (ICASSP), 2020.

[16] A. Akbari and R. Jafari, “Transition-Aware Detection of Modes of Locomotion and Transportation Through Hierarchical Segmentation,” *IEEE Sens. J.*, vol. 21, no. 3, pp. 3301–3313, 2020.

[17] B. Ibrahim and R. Jafari, “Cuffless Blood Pressure Monitoring from an Array of Wrist Bio-Impedance Sensors Using Subject-Specific Regression Models: Proof of Concept,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1723–1735, 2019.

[18] R. P. Hirten *et al.*, “Use of Physiological Data From a Wearable Device to Identify SARS-CoV-2 Infection and Symptoms and Predict COVID-19 Diagnosis: Observational Study,” *J. Med. Internet Res.*, vol. 23, no. 2, p. e26107, 2021.

[19] P. Bharti, D. De, S. Chellappan, and S. K. Das, “HuMAn: complex activity recognition with multi-modal multi-positional body sensing,” *IEEE Trans. Mob. Comput.*, vol. 18, no. 4, pp. 857–870, 2018.

[20] R. Solis, A. Pakbin, A. Akbari, B. J. Mortazavi, and R. Jafari, “A human-centered wearable sensing platform with intelligent automated data annotation capabilities,” in *Proceedings of the International Conference on Internet of Things Design and Implementation*, 2019, pp. 255–260.

[21] A. Akbari and R. Jafari, “Transferring activity recognition models for new wearable sensors with deep generative domain adaptation,” in *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*,

2019, pp. 85–96.

[22] E. R. Sykes, S. Pentland, and S. Nardi, “Context-aware mobile apps using iBeacons: towards smarter interactions,” in *Proceedings of the 25th Annual International Conference on Computer Science and Software Engineering*, 2015, pp. 120–129.

[23] J. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, “Deep convolutional neural networks on multichannel time series for human activity recognition,” in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[24] H. Ghasemzadeh, V. Loseu, and R. Jafari, “Collaborative signal processing for action recognition in body sensor networks: a distributed classification algorithm using motion transcripts,” in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2010, pp. 244–255.

[25] S. A. Rokni, M. Nourollahi, and H. Ghasemzadeh, “Personalized human activity recognition using convolutional neural networks,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[26] Z. E. Ashari and H. Ghasemzadeh, “Mindful active learning,” *arXiv Prepr. arXiv1907.12003*, 2019.

[27] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning,” in *international conference on machine learning*, 2016, pp. 1050–1059.

[28] M. Sandberg, J. Kristensson, P. Midlöv, C. Fagerström, and U. Jakobsson,

"Prevalence and predictors of healthcare utilization among older people (60+): Focusing on ADL dependency and risk of depression," *Arch. Gerontol. Geriatr.*, vol. 54, no. 3, pp. e349--e363, 2012.

[29] D. Wang, J. Zheng, M. Kurosawa, Y. Inaba, and N. Kato, "Changes in activities of daily living (ADL) among elderly Chinese by marital status, living arrangement, and availability of healthcare over a 3-year period," *Environ. Health Prev. Med.*, vol. 14, no. 2, p. 128, 2009.

[30] L. G. Clift, J. Lepley, H. Hagras, and A. F. Clark, "Autonomous computational intelligence-based behaviour recognition in security and surveillance," in *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II*, 2018, vol. 10802, p. 108020L.

[31] K. Davis *et al.*, "Activity recognition based on inertial sensors for ambient assisted living," in *2016 19th international conference on information fusion (fusion)*, 2016, pp. 371–378.

[32] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, 2006, pp. 4--pp.

[33] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 8, pp. 1082–1090, 2008.

[34] N. Győrbíró, Á. Fábián, and G. Hományi, "An activity recognition system for mobile phones," *Mob. Networks Appl.*, vol. 14, no. 1, pp. 82–91, 2009.

[35] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, 2011.

[36] T. Gu, S. Chen, X. Tao, and J. Lu, "An unsupervised approach to activity recognition and segmentation based on object-use fingerprints," *Data \& Knowl. Eng.*, vol. 69, no. 6, pp. 533–544, 2010.

[37] P. Palmes, H. K. Pung, T. Gu, W. Xue, and S. Chen, "Object relevance weight pattern mining for activity recognition and segmentation," *Pervasive Mob. Comput.*, vol. 6, no. 1, pp. 43–57, 2010.

[38] M. Philipose *et al.*, "Inferring activities from interactions with objects," *IEEE pervasive Comput.*, vol. 3, no. 4, pp. 50–57, 2004.

[39] B. Li *et al.*, "A wearable IoT aldehyde sensor for pediatric asthma research and management," *Sensors Actuators B Chem.*, vol. 287, pp. 584–594, 2019.

[40] G. Cappon, G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, "Wearable continuous glucose monitoring sensors: a revolution in diabetes treatment," *Electronics*, vol. 6, no. 3, p. 65, 2017.

[41] A. Almogren, "An automated and intelligent Parkinson disease monitoring system using wearable computing and cloud technology," *Cluster Comput.*, vol. 22, no. 1, pp. 2309–2316, 2019.

[42] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *2013 Humaine association conference on affective computing and intelligent interaction*, 2013, pp. 671–676.

[43] K. Vandecasteele *et al.*, “Automated epileptic seizure detection based on wearable ECG and PPG in a hospital environment,” *Sensors*, vol. 17, no. 10, p. 2338, 2017.

[44] R. Mc Ardle, S. Del Din, B. Galna, A. Thomas, and L. Rochester, “Differentiating dementia disease subtypes with gait analysis: feasibility of wearable sensors?,” *Gait \& posture*, vol. 76, pp. 372–376, 2020.

[45] D. M. Morens, G. K. Folkers, and A. S. Fauci, “The challenge of emerging and re-emerging infectious diseases,” *Nature*, vol. 430, no. 6996, pp. 242–249, 2004.

[46] A. S. Fauci, “Infectious diseases: considerations for the 21st century,” *Clin. Infect. Dis.*, vol. 32, no. 5, pp. 675–685, 2001.

[47] K. Subbarao and S. Mahanty, “Respiratory virus infections: understanding COVID-19,” *Immunity*, vol. 52, no. 6, pp. 905–909, 2020.

[48] I. Astuti and others, “Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response,” *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 4, pp. 407–412, 2020.

[49] X. Jiang, S. Rayner, and M.-H. Luo, “Does SARS-CoV-2 has a longer incubation period than SARS and MERS?,” *J. Med. Virol.*, vol. 92, no. 5, pp. 476–478, 2020.

[50] C. Leung, “The difference in the incubation period of 2019 novel coronavirus (SARS-CoV-2) infection between travelers to Hubei and nontravelers: The need for a longer quarantine period,” *Infect. Control Hosp. Epidemiol.*, vol. 41, no. 5, pp. 594–596, 2020.

[51] T. Mishra *et al.*, “Pre-symptomatic detection of COVID-19 from smartwatch data,” *Nat. Biomed. Eng.*, vol. 4, no. 12, pp. 1208–1220, 2020.

[52] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," *J. Autoimmun.*, vol. 109, p. 102433, 2020.

[53] P. G. Katona and F. Jih, "Respiratory sinus arrhythmia: noninvasive measure of parasympathetic cardiac control," *J. Appl. Physiol.*, vol. 39, no. 5, pp. 801–805, 1975.

[54] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Front. public Heal.*, vol. 5, p. 258, 2017.

[55] N. Marinsek *et al.*, "Measuring COVID-19 and influenza in the real world via person-generated health data," *medRxiv*.

[56] D. K. Ming *et al.*, "Continuous physiological monitoring using wearable technology to inform individual management of infectious diseases, public health and outbreak responses," *Int. J. Infect. Dis.*, vol. 96, pp. 648–654, 2020.

[57] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *International workshop on ambient assisted living*, pp. 216–223.

[58] H. Guo, L. Chen, L. Peng, and G. Chen, "Wearable sensor based multimodal human activity recognition exploiting the diversity of classifier ensemble," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 1112–1123.

[59] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," *arXiv Prepr*.

arXiv1604.08880.

[60] Y. Iwasawa, K. Nakayama, I. Yairi, and Y. Matsuo, “Privacy Issues Regarding the Application of DNNs to Activity-Recognition using Wearables and Its Countermeasures by Use of Adversarial Training.,” in *IJCAI*, pp. 1930–1936.

[61] L. C. Jatoba, U. Grossmann, C. Kunze, J. Ottenbacher, and W. Stork, “Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity,” in *2008 30th annual international conference of the ieee engineering in medicine and biology society*, pp. 5250–5253.

[62] W.-X. Li and N. Vasconcelos, “Complex activity recognition via attribute dynamics,” *Int. J. Comput. Vis.*, vol. 122, no. 2, pp. 334–370.

[63] L. Yao, F. Nie, Q. Z. Sheng, T. Gu, X. Li, and S. Wang, “Learning from less for better: semi-supervised activity recognition via shared structure discovery,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 13–24.

[64] H. Zhao and Z. Liu, “Human action recognition based on non-linear SVM decision tree,” *J. Comput. Inf. Syst.*, vol. 7, no. 7, pp. 2461–2468, 2011.

[65] Z. Zhao, Y. Chen, J. Liu, Z. Shen, and M. Liu, “Cross-people mobile-phone based activity recognition,” in *Twenty-second international joint conference on artificial intelligence*.

[66] D. J. Cook, “Learning setting-generalized activity models for smart spaces,” *IEEE Intell. Syst.*, vol. 2010, no. 99, p. 1.

[67] B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. Intille, “A long-term evaluation of sensing modalities for activity recognition,” in *International conference on Ubiquitous computing*, pp. 483–500.

[68] L. Wang, T. Gu, X. Tao, and J. Lu, “Sensor-based human activity recognition in a multi-user scenario,” in *European Conference on Ambient Intelligence*, pp. 78–87.

[69] M. Zeng *et al.*, “Convolutional neural networks for human activity recognition using mobile sensors,” in *6th International Conference on Mobile Computing, Applications and Services*, 2014, pp. 197–205.

[70] W. Jiang and Z. Yin, “Human activity recognition using wearable sensors by deep convolutional neural networks,” in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 1307–1310.

[71] T. Zebin, P. J. Scully, and K. B. Ozanyan, “Human activity recognition with inertial sensors using a deep learning approach,” in *2016 IEEE SENSORS*, 2016, pp. 1–3, doi: 10.1109/ICSENS.2016.7808590.

[72] F. Ordóñez and D. Roggen, “Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition,” *Sensors*, vol. 16, no. 1, p. 115, 2016.

[73] X. Li *et al.*, “Concurrent Activity Recognition with Multimodal CNN-LSTM Structure.” 2017.

[74] S. Song *et al.*, “Multimodal Multi-Stream Deep Learning for Egocentric Activity Recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2016.

[75] S. Matsui, N. Inoue, Y. Akagi, G. Nagino, and K. Shinoda, "User adaptation of convolutional neural network for human activity recognition," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 753–757, doi: 10.23919/EUSIPCO.2017.8081308.

[76] R. Fallahzadeh and H. Ghasemzadeh, "Personalization without user interruption: Boosting activity recognition in new subjects using unlabeled data," in *Proceedings of the 8th International Conference on Cyber-Physical Systems*, 2017, pp. 293–302.

[77] S. A. Rokni and H. Ghasemzadeh, "Synchronous dynamic view learning: a framework for autonomous training of activity recognition models using wearable sensors," in *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2017, pp. 79–90.

[78] V. F. Rey and P. Lukowicz, "Label propagation: An unsupervised similarity based method for integrating new sensors in activity recognition systems," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–24.

[79] H. Hachiya, M. Sugiyama, and N. Ueda, "Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition," *Neurocomputing*, vol. 80, pp. 93–101.

[80] M. Kurz, G. Hölzl, A. Ferscha, A. Calatroni, D. Roggen, and G. Tröster, "Real-time transfer and evaluation of activity recognition capabilities in an opportunistic system," *Mach. Learn.*, vol. 1, no. 7, p. 8.

[81] G. Singla, D. J. Cook, and M. Schmitter-Edgecombe, "Recognizing independent

and joint activities among multiple residents in smart environments," *J. Ambient Intell. Humaniz. Comput.*, vol. 1, no. 1, pp. 57–63.

[82] D. J. Cook, N. C. Krishnan, and P. Rashidi, "Activity discovery and activity recognition: A new partnership," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 820–828.

[83] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "CASAS: A smart home in a box," *Computer (Long. Beach. Calif).*, vol. 46, no. 7, pp. 62–69.

[84] N. C. Krishnan and D. J. Cook, "Activity recognition on streaming sensor data," *Pervasive Mob. Comput.*, vol. 10, pp. 138–154.

[85] G. Singla, D. J. Cook, and M. Schmitter-Edgecombe, "Incorporating temporal reasoning into activity recognition for smart home residents," in *Proceedings of the AAAI workshop on spatial and temporal reasoning*, pp. 53–61.

[86] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *International conference on pervasive computing*, pp. 158–175.

[87] T. L. M. Van Kasteren, G. Englebienne, and B. J. A. Kröse, "Activity recognition using semi-Markov models on real world smart home datasets," *J. Ambient Intell. Smart Environ.*, vol. 2, no. 3, pp. 311–325.

[88] G. W. Musumba and H. O. Nyongesa, "Context awareness in mobile computing: A review," *Int. J. Mach. Learn. Appl.*, vol. 2, no. 1, p. 5, 2013.

[89] M. Weiser, "The computer for the 21st century," *IEEE pervasive Comput.*, vol. 1, no. 1, pp. 19–25, 2002.

[90] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the internet of things: A survey," *IEEE Commun. Surv. tutorials*, vol. 16, no. 1, pp. 414–454, 2014.

[91] L. Herranz, S. Jiang, and R. Xu, "Modeling restaurant context for food recognition," *IEEE Trans. Multimed.*, vol. 19, no. 2, pp. 430–440, 2016.

[92] R. Reichle *et al.*, "A comprehensive context modeling framework for pervasive computing systems," in *IFIP International Conference on Distributed Applications and Interoperable Systems*, 2008, pp. 281–295.

[93] D. De, P. Bharti, S. K. Das, and S. Chellappan, "Multimodal wearable sensing for fine-grained activity recognition in healthcare," *IEEE Internet Comput.*, vol. 19, no. 5, pp. 26–35, 2015.

[94] A. Filippoupolitis, W. Oliff, B. Takand, and G. Loukas, "Location-enhanced activity recognition in indoor environments using off the shelf smart watch technology and BLE beacons," *Sensors*, vol. 17, no. 6, p. 1230, 2017.

[95] L. Aalto, N. Göthlin, J. Korhonen, and T. Ojala, "Bluetooth and WAP push based location-aware mobile advertising system," in *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, 2004, pp. 49–58.

[96] J. Zheng and L. M. Ni, "An unsupervised learning approach to social circles detection in ego bluetooth proximity network," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, 2013, pp. 721–724.

[97] M. Han, J. Bang, C. Nugent, S. McClean, and S. Lee, “A lightweight hierarchical activity recognition framework using smartphone sensors,” *Sensors*, vol. 14, no. 9, pp. 16181–16195, 2014.

[98] A. Subramanya, A. Raj, J. A. Bilmes, and D. Fox, “Recognizing activities and spatial context using wearable sensors,” *arXiv Prepr. arXiv1206.6869*, 2012.

[99] J. Chon and H. Cha, “Lifemap: A smartphone-based context provider for location-based services,” *IEEE Pervasive Comput.*, vol. 10, no. 2, pp. 58–67, 2011.

[100] N. Bulusu, J. Heidemann, D. Estrin, and others, “GPS-less low-cost outdoor localization for very small devices,” *IEEE Pers. Commun.*, vol. 7, no. 5, pp. 28–34, 2000.

[101] T.-B. Nguyen, T. Nguyen, W. Luo, S. Venkatesh, and D. Phung, “Unsupervised inference of significant locations from wifi data for understanding human dynamics,” in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia*, 2014, pp. 232–235.

[102] A. Alvarez-Alvarez, J. M. Alonso, and G. Trivino, “Human activity recognition in indoor environments by means of fusing information extracted from intensity of WiFi signal and accelerations,” *Inf. Sci. (Ny).*, vol. 233, pp. 162–182, 2013.

[103] H. Hong, C. Luo, and M. C. Chan, “Socialprobe: Understanding social interaction through passive wifi monitoring,” in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2016, pp. 94–103.

[104] Z. Chen, Y. Chen, S. Wang, J. Liu, X. Gao, and A. T. Campbell, “Inferring social

contextual behavior from bluetooth traces," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, 2013, pp. 267–270.

[105] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey," in *23th International conference on architecture of computing systems 2010*, 2010, pp. 1–10.

[106] S. L. Lau, I. König, K. David, B. Parandian, C. Carius-Düssel, and M. Schultz, "Supporting patient monitoring using activity recognition with a smartphone," in *2010 7th International Symposium on Wireless Communication Systems*, 2010, pp. 810–814.

[107] J. Pansiot, D. Stoyanov, D. McIlwraith, B. P. L. Lo, and G.-Z. Yang, "Ambient and wearable sensor fusion for activity recognition in healthcare monitoring systems," in *4th international workshop on wearable and implantable body sensor networks (BSN 2007)*, 2007, pp. 208–212.

[108] J. Wu, A. Akbari, R. Grimsley, and R. Jafari, "A decision level fusion and signal analysis technique for activity segmentation and recognition on smart phones," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 2018, pp. 1571–1578.

[109] J. Wu and R. Jafari, "Orientation independent activity/gesture recognition using wearable motion sensors," *IEEE Internet Things J.*, 2018.

[110] G. M. Weiss and J. Lockhart, “The impact of personalization on smartphone-based activity recognition,” in *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[111] H. M. Hossain, M. D. Al Haiz Khan, and N. Roy, “DeActive: scaling activity recognition with active deep learning,” *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 2, p. 66, 2018.

[112] O. D. Lara and M. A. Labrador, “A Survey on Human Activity Recognition using Wearable Sensors,” *IEEE Commun. Surv. Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013, doi: 10.1109/SURV.2012.110112.00192.

[113] P. Siirtola, H. Koskimäki, and J. Röning, “Personalizing human activity recognition models using incremental learning,” *arXiv Prepr. arXiv1905.12628*, 2019.

[114] P. Siirtola, H. Koskimäki, and J. Röning, “Importance of user inputs while using incremental learning to personalize human activity recognition models,” *arXiv Prepr. arXiv1905.11775*, 2019.

[115] S. Sabato and T. Hess, “Interactive algorithms: from pool to stream,” in *Conference on Learning Theory*, 2016, pp. 1419–1439.

[116] T. Diethe, N. Twomey, and P. A. Flach, “Active transfer learning for activity recognition.,” in *ESANN*, 2016.

[117] S. Bagaveyev and D. J. Cook, “Designing and evaluating active learning methods for activity recognition,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014,

pp. 469–478.

[118] A. Mannini and S. Intille, “Classifier personalization for activity recognition using wrist accelerometers,” *IEEE J. Biomed. Heal. informatics*, 2018.

[119] T. Sztyler and H. Stuckenschmidt, “Online personalization of cross-subjects based activity recognition models on wearable devices,” in *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2017, pp. 180–189.

[120] A. Akbari, R. S. Castilla, R. Jafari, and B. J. Mortazavi, “Using intelligent personal annotations to improve human activity recognition for movements in natural environments,” *IEEE J. Biomed. Heal. informatics*, vol. 24, no. 9, pp. 2639–2650, 2020.

[121] A. Subramanya, S. Srinivas, and R. V. Babu, “Confidence estimation in Deep Neural networks via density modelling,” *CoRR*, vol. abs/1707.0, 2017, [Online]. Available: http://arxiv.org/abs/1707.07013.

[122] D. R. Seshadri *et al.*, “Wearable sensors for COVID-19: a call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments,” *Front. Digit. Heal.*, vol. 2, p. 8.

[123] S. Hassantabar *et al.*, “Coviddeep: Sars-cov-2/covid-19 test based on wearable medical sensors and efficient neural networks,” *arXiv Prepr. arXiv2007.10497*.

[124] G. Quer *et al.*, “Wearable sensor data and self-reported symptoms for COVID-19 detection,” *Nat. Med.*, vol. 27, no. 1, pp. 73–77.

[125] A. Natarajan, H.-W. Su, and C. Heneghan, “Assessment of physiological signs

associated with COVID-19 measured using wearable devices," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–8, 2020.

[126] S. G. Holt *et al.*, "Monitoring skin temperature at the wrist in hospitalised patients may assist in the detection of infection," *Intern. Med. J.*, vol. 50, no. 6, pp. 685–690, 2020.

[127] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, no. Aug, pp. 975–1005, 2004.

[128] J. Wu, L. Sun, and R. Jafari, "A wearable system for recognizing American sign language in real-time using IMU and surface EMG sensors," *IEEE J. Biomed. Heal. informatics*, vol. 20, no. 5, pp. 1281–1290, 2016.

[129] E. Guenterberg, H. Ghasemzadeh, V. Loseu, and R. Jafari, "Distributed continuous action recognition using a hidden markov model in body sensor networks," in *International Conference on Distributed Computing in Sensor Systems*, 2009, pp. 145–158.

[130] H. Ghasemzadeh, E. Guenterberg, K. Gilani, and R. Jafari, "Action coverage formulation for power optimization in body sensor networks," in *Proceedings of the 2008 Asia and South Pacific Design Automation Conference*, 2008, pp. 446–451.

[131] E. Hoque and J. Stankovic, "AALO: Activity recognition in smart homes using Active Learning in the presence of Overlapped activities," in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare*

(PervasiveHealth) and Workshops, 2012, pp. 139–146.

[132] D. Riboni, T. Sztyler, G. Civitarese, and H. Stuckenschmidt, "Unsupervised recognition of interleaved activities of daily living through ontological and probabilistic reasoning," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 1–12.

[133] R. Wang *et al.*, "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, 2014, pp. 3–14.

[134] M. A. Alsheikh, D. Niyato, S. Lin, H.-P. Tan, and Z. Han, "Mobile big data analytics using deep learning and apache spark," *IEEE Netw.*, vol. 30, no. 3, pp. 22–29, 2016.

[135] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," in *European Conference on Computer Vision*, 2016, pp. 597–613.

[136] A. Subramanya, S. Srinivas, and R. V. Babu, "Confidence estimation in Deep Neural networks via density modelling," *CoRR*, vol. abs/1707.0, 2017.

[137] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv Prepr. arXiv1312.6114*, 2013.

[138] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. Dec, pp. 3371–3408,

2010.

[139] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?," *J. Mach. Learn. Res.*, vol. 11, no. Feb, pp. 625–660, 2010.

[140] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *Wearable Computers (ISWC), 2012 16th International Symposium on*, 2012, pp. 108–109.

[141] T. R. Bennett, H. C. Massey, J. Wu, S. A. Hasnain, and R. Jafari, "Motionsynthesis toolset (MoST): An open source tool and data set for human motion data synthesis and validation," *IEEE Sens. J.*, vol. 16, no. 13, pp. 5365–5375, 2016.

[142] A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 5574–5584.