Human Behavioural Abnormality Detection With Home Radar Sensor

Younwoo Jeong

4th Year Project Report
Artificial Intelligence
School of Informatics
University of Edinburgh

2024

Abstract

This study explores the feasibility of detecting human behavioral abnormalities in a home setting using radar sensors, with a focus on enhancing healthcare monitoring and smart home technologies. Utilizing data collected from seven individuals in a controlled environment, we investigate the potential of radar sensors to recognize deviations in normal activities, such as changes in walking patterns and the frequency of restroom visits. The research employs a variety of machine learning models, including probabilistic models like Gaussian Mixture Models (GMMs) and Random Forest classifiers, as well as sequential models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformer models, to analyze the radar data for anomaly detection.

Our findings indicate that sequential models, particularly the hybrid CNN-LSTM and CNN-Transformer models, significantly outperform probabilistic models in identifying abnormal behaviors, achieving accuracies exceeding 90%. This highlights the effectiveness of machine learning techniques in processing and interpreting the unique data signatures provided by radar technology for anomaly detection in residential settings. The study underscores the promise of radar-based systems as a non-intrusive, privacy-preserving solution for monitoring individuals, especially the elderly or those with health conditions requiring supervision, within their homes. Future work will focus on enhancing model generalization to new subjects and integrating additional features to improve detection accuracy.

Research Ethics Approval

This project obtained approval from the Informatics Research Ethics committee.
Ethics application number: 671984
Date when approval was obtained: 2023-12-01
The participants' information sheet and a consent form are included in the appendix.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Younwoo Jeong*)

Acknowledgements

I firstly want to thank my supervisor Kia Nazarpour, for giving me a great opportunity to research with the Feather team. Secondly, I want to thank the Feather team for the feedback given every meeting. Thirdly, I would also like to mention Saber Mirzaee Bafti, who gave the most help along the way with amazing advice. If it wasn't for him I wouldn't have been able to finish this project properly. Lastly, I thank my family and friends for giving me support and advice.

Table of Contents

Chapter 1

Introduction

In healthcare monitoring and smart home technology, accurately identifying unusual behavior is highly desirable. Continuous tracking of health and behavior benefits from extensive historical data on daily activities, routines, and movement quality. Typically, this involves using sensors and devices to gather vast amounts of context information daily. The challenge lies in processing this data with effective machine learning techniques to identify anomalies. Various sensors can monitor an individual's activity, movement, or health status, each collecting different types of data and contributing to the broad field of human behaviour and activity recognition. These sensors include video cameras, door switches, motion detectors, and etc. The sensors could be used together to create a smart home environment which further improve the ability to detect activity.

In this research, we will focus on evaluating the radar sensor's capability for recognizing human behavior, particularly its potential to identify abnormalities in home settings. The choice of radar is motivated by its advantages over traditional health monitoring methods, including enhanced privacy, non-intrusive sensing, and robustness to lighting conditions (Amin et al., 2016). Should radar prove effective in detecting domestic abnormalities, it could offer a cost-effective and discreet safeguarding solution for individuals living alone who require monitoring, presenting a viable alternative to conventional smart home technologies.

Throughout our study, we characterize home abnormalities as alterations in walking patterns and the frequency of restroom visits. We employ radar data, including an individual's location, velocity, and acceleration within a room, to identify these changes. By applying feature extraction techniques to this data, we develop a machine learning model capable of detecting abnormalities. Moreover, we explore both probabilistic and sequential models to determine the most effective approach for this task.

1.1 Structure

The structure of this paper is divided into 4 sections: Background, Methodology, Results and Discussion, and Conclusion. The **background** chapter introduces the concept of human behavioural abnormality, and some relevant researches on it. The chapter also explains the concepts used in the methodology of the paper. **Methodology** entails the process of data collection to data processing and explains the model used for the study. The **results and discussion** chapter shows the performance of each model through tables and confusion matrices. Lastly, the **conclusion** chapter gives a brief summary of the work done as well as limitations of the study and directions for future work.

Chapter 2

Background

The task of identifying behavioral abnormalities through various sensors presents a notable challenge in the field of research, compounded by the subjective nature of what constitutes an ”abnormality.” Unlike targeting specific events, the aim here is to identify deviations in behavior, a goal that has escaped a definitive resolution thus far. However, the appearance of advanced methodologies, particularly those involving machine learning and neural networks, has significantly enhanced our capacity to pinpoint unusual occurrences with greater precision.

This section will delve into the diverse range of abnormalities that have been the focus of previous investigations, outlining the methodologies employed for data collection. It will then transition to an examination of data analysis techniques, with an emphasis on the application of machine learning and deep learning approaches tailored for the identification of behavioral anomalies. Additionally, the chapter will explore various data visualization techniques, offering insights into the behavioral patterns of subjects under observation.

2.1 Human Behavioural Abnormality

Previous investigations into abnormality detection in real-life scenarios have covered a wide range, from identifying atypical behaviors in laying hens using 3D cameras (Du and Teng, 2021) to detecting anomalies in crowds (Joshi and Patel, 2021). In the context of laying hens, abnormalities could include conditions such as heat stress, human disturbance, or failures in feeding and drinking systems. Conversely, crowd anomalies might encompass incidents like accidents, sudden movements, or conflicts. The definition of an anomaly, therefore, is highly context-dependent.

In the domain of vehicular behavior, abnormalities are identified as rapid accelerations, emergency braking, or abrupt lane changes (Jia et al., 2019), highlighting the diversity in abnormal behavior across different settings.

In the field of healthcare, detecting specific abnormal behaviors in individuals becomes crucial. This is especially true for conditions like dementia, where early detection of behaviors such as repetitive actions, sleep disruption, and confusion can be critical

(Arifoglu and Bouchachia, 2019). Early intervention, as noted, can significantly aid in the diagnosis and treatment of such conditions. Additionally, monitoring variables like sleep patterns and toilet visit frequency has been utilized to detect early signs of urinary tract infections in individuals with dementia (Enshaeifar et al., 2019), underscoring the tailored approach required in anomaly detection for effective and timely care.

2.2 Time Series Classification

Time Series Classification is an essential method in data analysis that helps spot patterns and sort data that changes over time. This method is very important for making predictions and spotting unusual patterns in various areas like the stock market, weather reports, and healthcare. Time series classification uses both simple statistical methods and more complex machine learning algorithms like neural networks. In the field of time series classification, one of the key tasks is to identify unusual data. This involves figuring out which data points or patterns don't fit the norm. To do this, experts might use historical data to know what's typical or turn to machine learning models that are trained to recognize what's normal and what's not (Liu et al., 2022). These models get better at spotting odd data by learning from lots of examples of both normal and abnormal behavior. This approach helps quickly highlight when something unusual pops up, making it easier to understand why it's happening and to take steps to address it (Johnson and Khoshgoftaar, 2019).

2.3 Data Collection and Structure

Like any multivariate data analysis, sensor data must undergo feature extraction. Various datasets have been developed to capture behavioral abnormalities in daily life, such as CASAS (Cook et al., 2013), ARAS (Alemdar et al., 2013), and the MIT Activity Recognition dataset (Tapia et al., 2004). CASAS leverages an array of sensors to simulate a smart home environment, gathering data from 18 apartments over a month. The MIT dataset, on the other hand, utilizes state-change sensors throughout the home, collecting data for two weeks with subjects logging activities every 15 minutes. ARAS follows a similar approach to CASAS, employing binary sensors in homes with two residents, collecting data over 30 days with resident annotations.

Recent studies on real-time detection of behavioral abnormalities often utilize camera data (Zhang et al., 2023), employing image processing to monitor the elderly in health-care settings. Other research, such as that by (Wang et al., 2022), uses smartphone sensors and Ultra-Wide Band (UWB) systems for activity and location tracking within a home. Similarly, another study utilizes passive infrared sensors (PIR) and pressure sensors in beds and sofas to gather data.

In exploring abnormal behavior prediction among the elderly, (Zerkouk and Chikhaoui, 2019) harness IoT and smart home technologies to observe residents' activities and health through embedded sensors, focusing on data from household appliances to ascertain what activities are being performed.

More specifically, our research on employing radar for indoor detection of behavioral abnormalities, particularly among seniors living alone, is part of an ongoing field of study. For instance, one study (Takabatake et al., 2019) utilized FMCW radar data to identify incidents of fainting in elderly individuals within confined spaces like bathrooms, focusing solely on the radar spectrum. However, unlike our research, their work was restricted to specific areas rather than more open spaces such as living rooms. Another study (Amin et al., 2016) explored radar's application in fall detection among the elderly, while (Hall et al., 2019) used radar to identify gait abnormalities, aligning more closely with our research's focus on abnormal behaviors. These studies primarily analyzed Doppler signatures from raw radar data, in contrast to our approach of examining individual movement features.

2.4 Data Analysis Methods

Analyzing sensory data involves several steps, starting with gathering data from sensors, processing this data to clean and reformat it, and finally using it to train a model for detecting behavioral abnormalities. During data pre-processing, tasks like eliminating duplicates, handling missing values, and organizing the data are crucial to ensure the model can effectively learn and make classifications.

This phase also includes filtering out irrelevant data, which necessitates comprehensive testing to identify the most critical data for detecting abnormalities. For instance, the study by (Weisenberg et al., 2008) focused solely on the subject's activity or inactivity levels for model training. Conversely, other research predetermined the subject's spatial position as essential for the training and classification process (Wang et al., 2022).

Identifying abnormalities has evolved to include statistical methods and, more recently, machine learning techniques. The study by (Spanos et al., 2019) employed both statistical (like Descriptive Statistics, PCA) and machine learning approaches for anomaly detection in smart homes. Various machine learning algorithms have been utilized across studies for classifying abnormal behaviors.

Moreover, some research has investigated behavioral abnormalities through activity sequences using Hidden Markov Models (HMM). For example, (Aliakbarpour et al., 2011) demonstrated the ability to detect ATM robberies using a concurrent HMM architecture. Similarly, (Lühr et al., 2004) improved detection accuracy in a kitchen scenario by employing modified HMMs, such as left-right HMMs and explicit state duration HMMs (ESD-HMMs), showing enhanced performance compared to standard HMMs.

2.4.1 GMMs

A Gaussian Mixture Model (GMM) is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Each distribution can be thought of as a class. GMMs are a type of mixture model that provide a method for representing an overall distribution as a combination of multiple Gaussian distributions. Each of these component Gaussian distributions contributes to the overall mixture model with a certain weight. Simply put, GMMs consider each data point as a mixture of possibilities from all of the distributions or classes.

In terms of abnormality detection, (Rojas and Tozzi, 2016) introduced implementing a Gaussian Mixture Model(GMM) for crowd abnormality detection through a video. The writers use the background subtraction algorithm and optical flow to convert the normal behavioural pattern of the crowd to the GMM.

2.4.2 Random Forest Classification

A Random Forest Classifier is an ensemble learning method that operates by constructing a multitude of decision trees during the training phase and outputting the class that is the mode of the classes of the individual trees. It is known for its high accuracy, ability to deal with unbalanced and missing data, and its feature of providing an estimate of variable importance. By combining multiple decision trees to make decisions, Random Forest reduces the risk of overfitting associated with single decision trees and enhances the predictive accuracy. Each tree in the Random Forest splits the data based on a feature, randomly selected at each split, making the ensemble robust against overfitting and capable of capturing complex patterns in the data.

In the context of classification, (Liaw and Wiener, 2002) demonstrated the effectiveness of Random Forest in classifying high-dimensional data. The study showcased Random Forest's ability to handle thousands of input variables without variable deletion, providing insights into which variables are important in the classification.

2.4.3 CNNs

Convolutional Neural Networks (CNNs) (O'Shea and Nash, 2015) have become invaluable in analyzing time series data, especially in spotting anomalies. Identifying anomalies means finding patterns or outliers that stand out from what's typical. While CNNs are traditionally known for their role in extracting spatial features in image processing, their use in time series analysis involves tailoring them to pick up on temporal features instead. (Zhao et al., 2017) Think of a time series as a linear sequence where CNN filters move along, pinpointing key elements such as trends and sudden changes. This feature enables CNNs to grasp intricate temporal relationships within the data, a vital aspect of detecting anomalies. Some recent work that involve neural networks to discriminate, report that they have been obtaining great results. (Joshi and Patel, 2021) use CNNs into practice to train and test crowd monitoring through video. They report that by using the CNN model they got an accuracy of up to 99.64% which was significantly higher than previous models they have researched.

2.4.4 RNNs

Recurrent Neural Networks (RNNs) (Rumelhart and McClelland, 1987) are designed to process sequences by retaining a memory of previous inputs, making them ideal for tasks like language translation and speech recognition. However, they struggle with long sequences due to the vanishing gradient problem, which affects their ability to learn long-term dependencies. Studies have reported that RNNs are also adequate in abnormal behaviour detection as daily activities routines can be recognised by RNN models. (Arifoglu and Bouchachia, 2017) The authors add that because of Vanilla RNN's vanishing gradient problem where they are not capable of capturing long term dependencies on sequences, they propose Long Short Term Memory (LSTM) RNNs. The study discloses that LSTM RNNs are able to mitigate the effects of the diminishing gradient issue.

LSTM (Hochreiter and Schmidhuber, 1997) or Long Short Term Memory models are a special type of Recurrent Neural Network (RNN) capable of learning long-term dependencies in sequence data. LSTMs are designed with a more complex computational unit called a cell, which includes three main components: an input gate, an output gate, and a forget gate. These gates regulate the flow of information into and out of the cell, and decide which parts of the long-term information to keep or discard, allowing LSTMs to effectively remember or forget information over long periods. Figure 2.1 (Hesaraki, 2023) shows the architecture of a LSTM. These characteristics makes LSTMs particularly useful for tasks involving sequences, such as time series classification.

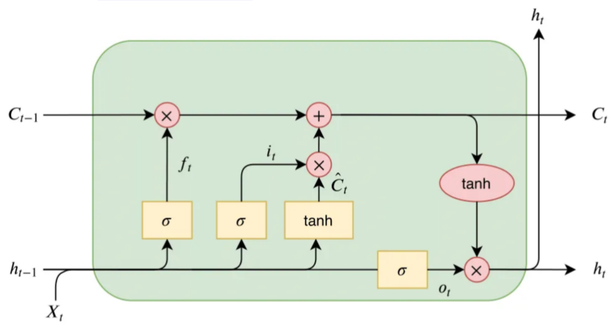


Figure 2.1: Architecture of a LSTM cell (Hesaraki, 2023)

2.4.5 Transformers

Transformers (Vaswani et al., 2017) are a model architecture in the field of machine learning that have revolutionized natural language processing and beyond, leveraging self-attention mechanisms to efficiently handle sequential data and capture complex dependencies across vast distances within the data. This architecture allows transformers to process entire sequences of data in parallel, significantly enhancing computational efficiency and enabling more sophisticated analysis of relationships within the data, compared to traditional sequential models.

In the area of time series, transformers take the approach to handling sequential data with their ability to model complex, long-term relationships. Unlike traditional models that may struggle with the dimension of the sequence and scale of data, transformers leverage their self-attention mechanism to directly compute relationships between all points in a time series, regardless of their distance in time. This capability allows for a more nuanced understanding of temporal dynamics, making transformers in theory effective for classification. Transformers have shown to outperform recurrent neural networks when training on long sequences, and convolutional neural networks, which are generally limited to capturing dependencies between adjacent elements. However, transformers come with a down side. Due to the nature of the self attention mechanism where each sequence is processed simultaneously, transformers are naturally weak on temporal sequence. (Zeng et al., 2023) Which means that transformers are unable to recognize the order of the data. To address this, a LSTM layer is commonly integrated prior to the transformer layer, enabling the model to better grasp temporal dynamics by combining LSTM's strength in handling time-series data with the transformer's capability in modeling complex dependencies. (Song et al., 2018)

2.4.6 Compound models

Some studies have used compound models that use a mix of many networks for abnormality detection. (Wang et al., 2022) use multiple techniques for their model including a CNN, a LSTM RNN, and a HMM. The study divides behaviour abnormality into two categories, specific behaviours including three types of falls and behavioral states which is the change in the long-term behaviour or routine. The CNN and LSTM is used together to classify the action of the subject. (i.e. Walking, Sitting down, Lying, Falling forward) The CNN is used to process the 30×9 data of the 9 different variables of measured activity (i.e. acceleration, orientation etc) by convolution and pooling operations. The high level features that have been outputted from the CNN is then fed into the LSTM to process the subject's behavioral classification. Then this information as well with the subject's positional information is then categorized into more specific states (i.e. Standing in the bedroom, Lying in the bedroom etc). The HMM is fed with this information which is used to find the abnormality in the daily behavior.

In addressing the complexities of surveillance video anomaly detection (SVAD), a previous study (Ullah et al., 2023), introduced a novel hybrid framework that merges convolutional neural networks (CNN) with transformers (Vaswani et al., 2017). This approach efficiently tackles the challenges posed by variations in object scale, the influence of the background, and the broad spectrum of anomalies encountered in surveillance contexts. The framework operates in a two-step process, beginning with spatial feature extraction via a CNN model, followed by the analysis of long-term temporal relationships through a transformer.

2.5 Data Visualisation

There are multiple ways of visualising a person's behaviour, this section will explain some of the previous techniques related to this study on behavioural visualisation.

(Vrotsou et al., 2007) notes a person's behaviour could be divided into various activities the person could be doing (i.e. Prepare Food, Eat Dinner, Wash up) and visualised using a 2D stacked bar chart where the activities are represented with different colors when the y-axis is time and x-axis is days. The study explains that when examining the behaviour of a person through multiple days, some of the general trend of behaviour could be revealed.

(Liu et al., 2017) uses heat maps to show driving behaviours of drivers when driving around circuits. The diagram is shaped like the actual circuit and colored with different colors for different driving behaviours for that particular part of tract. With parts that have a lower certainty mixed with different colors that also have probability. (Enshaeifar et al., 2019) also used heat maps to show the places where more or less activity was detected from the sensors. The six hour time window was also divided into six and shown as separate activity windows.

2.6 Evaluation metrics

The metrics used to test the performance of the models were as follows: **Accuracy**, **Precision**, **Recall**, and **F1 Score**. These metrics are used for classification tasks and calculated by four key benchmarks depending on the prediction of the model.

- **True Positives (TP)**: These are the instances where the model correctly predicts the positive class. In our case, if the model correctly labels normal data.
- **True Negatives (TN)**: These are the instances where the model correctly predicts the negative class. This is if the model correctly identifies the anomaly.
- **False Positives (FP)**: These occur when the model incorrectly predicts the positive class. This is if the model labels the normal data as an anomaly.
- **False Negatives (FN)**: These happen when the model incorrectly predicts the negative class. This is when anomaly is labelled as normal.

Then each of the metrics are calculated by the following equations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.3}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{2.4}$$

The accuracy gives a general score on the classification performance, precision means the effectiveness in avoiding false negatives, and recall indicates how well the model captures the positive data. The F1 score is the harmonic mean of precision and recall giving both metrics equal weight. A high F1 score suggests that the model has a good balance of precision and recall, making few false positive and false negative errors. F1 score is also useful when there is an imbalance between positive and negative data samples.(Aditsania et al., 2017) On top of this, we made use of macro F1 score which is the average F1 score for both classes, using this we made sure that the abnormal class wasn't underrepresented. Furthermore, the model trained based on maximising the macro F1 score rather than the loss or accuracy, as this was a better measure of how well the model was performing.

Chapter 3

Methodology

3.1 Overview

In the methodology section, we address the limitations of conventional video anomaly detection techniques within residential settings, primarily due to the significant privacy concerns they entail. The invasive nature of video surveillance precludes its widespread acceptance in homes, where privacy is highly valued.

Turning our focus to radar technology, we explore its potential as a less intrusive alternative that aligns more closely with privacy expectations in home environments. Radar's ability to monitor movements without capturing identifiable visual details positions it as a favorable tool for detecting anomalies within the privacy constraints of domestic spaces.

Central to our investigation is the evaluation of machine learning models' capacity to discern anomalies in home environments through radar data analysis. This inquiry is not rooted in home security, but rather in identifying irregularities in routine activities, such as changes in walking patterns. By analyzing how these models process and interpret the unique data signatures radar technology provides, we aim to ascertain the feasibility of employing machine learning for anomaly detection in residential settings, ensuring a delicate balance between effective monitoring and respect for individual privacy.

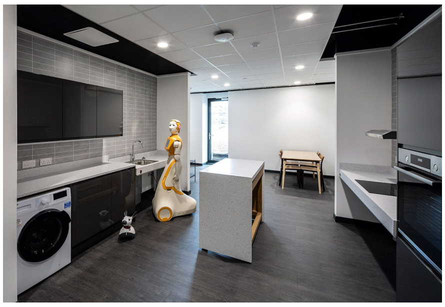
This section outlines the methodologies employed to investigate anomaly detection using a home radar system using machine learning models, aiming to provide a comprehensive understanding of the entire process from data collection, data cleaning, and anomaly classification.

3.2 Data Collection

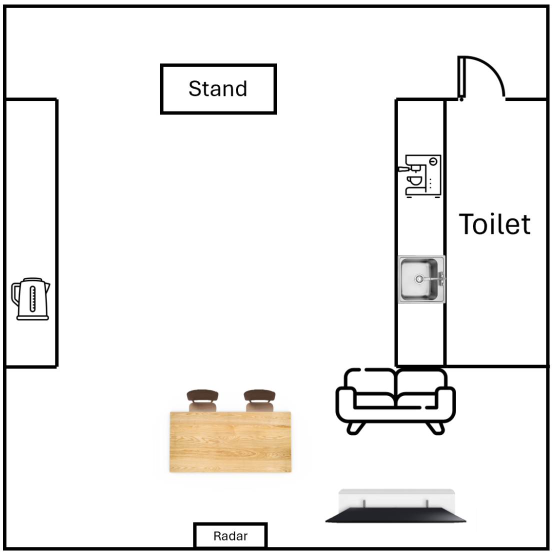
The dataset for the research was autonomously gathered from a smart home environment situated within the National Robotarium[1], hosted by Heriot-Watt University. The objective of the experimental study was to accumulate data pertaining to typical and

[1] https://thenationalrobotarium.com/

atypical behaviors within a domestic setting. An illustrative depiction of the room is presented in figure 3.1a((TheFIS, 2023)). The radar apparatus was strategically positioned at the rear of the room, oriented towards the living area.



(a) Smart home located at the National Robotarium (TheFIS, 2023)



(b) Room layout

Figure 3.1: Photo and layout of the smart home inside the National Robotarium

The data acquisition for our study was facilitated by the use of Texas Instrument's AWR6843AOP mmWave radar sensor[2]. We engaged seven individuals([C,D,E,F,G,H,I]) to take part in the research, which was structured into three distinct sessions, each lasting 50 minutes. The initial and concluding sessions were dedicated to capturing data reflective of standard behavior, while the intermediary session focused on documenting deviations from this norm.

Participants were instructed to engage in a series of predetermined activities during each session, which included preparing tea, making a sandwich, using the restroom, watching television, reading, and engaging in a puzzle. The activities involving reading and puzzle-solving were designated to a table setting, whereas television viewing could occur either at the table or on the sofa. The remaining tasks were distributed throughout the living room space.

To ensure variability, the sequence of activities was randomized each session and informed to the participants. A distinctive aspect of the abnormal behavior session involved instructing participants to simulate discomfort in their movement and to alter their frequency of restroom visits compared to their typical behavior. Uncomfortable movement would be limping while walking, forward leaning, etc.

[2]https://www.ti.com/product/AWR6843AOP

3.3 Dataset Characteristics

This section will introduce the characteristics of the data collected from the radar for the 3 sessions from the 7 subjects. The radar is equipped with a localization system embedded within its firmware, enabling it to determine the x, y, and z positions of individuals in the room, along with their acceleration and velocity on the x and y axes. Furthermore, the radar documents the exact timestamp at which each data point is collected. The default interval for data gathering is established at 0.05 seconds, though this may fluctuate contingent upon the activity engaged in. While the standard interval is used to during active movements such as walking, it is adjusted to a lower frequency during less active movements, like when an subject is seated. The full set of features collected by the radar is:

- **X**: The position of the subject on the x axis
- **Y**: The position of the subject on the y axis
- **Z**: The position of the subject on the z axis
- **AccX**: The acceleration of the subject on the x axis
- **AccY**: The acceleration of the subject on the y axis
- **VelX**: velocity of the subject on the y axis
- **VelY**: The velocity of the subject on the y axis
- **Timestamp**: The exact time when the data sample was collected

An example of the x,y position detected by the radar can be seen from the heatmap of subject D's position inside the room during a session at figure 3.2.

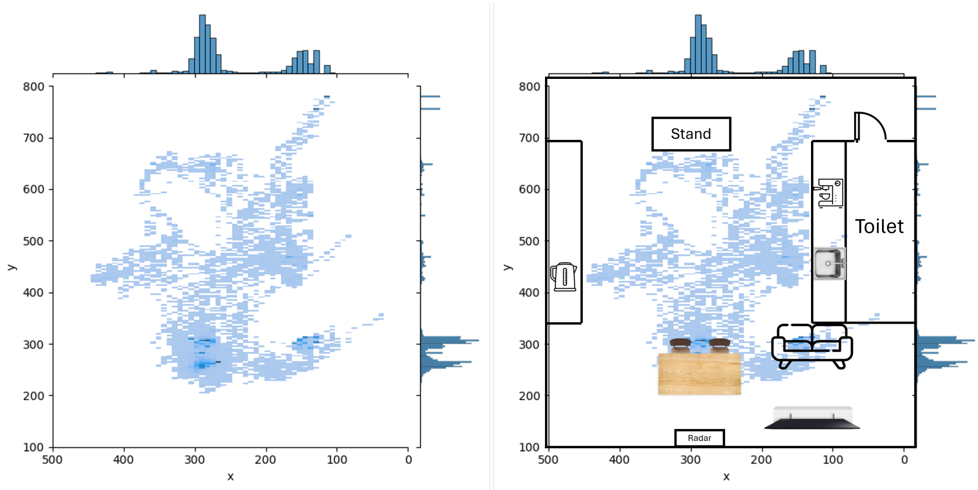


Figure 3.2: Heatmap of x y position of subject D during a normal session

3.4 Data Pre-processing

We've reviewed the data collection and its characteristics. Next, we'll examine feature extraction and segmentation to effectively train machine learning models. Due to the presence of walls, furniture, and kitchen appliances, the room's interior often introduces interference to the radar's signal. This occasionally lead to the localization system's inability to accurately track the subject, thus generating noisy data. To mitigate this issue, we established a boundary corresponding to the room's dimensions, treating any detection beyond these parameters as unwanted noise. Moreover, we addressed additional disturbances such as lag(instances where the radar fails to promptly track the subject's movement) through refinement processes during feature extraction.

In the context of feature extraction, our focus was narrowed to analyzing the walking patterns of participants and the frequency of their toilet visits, pivotal for anomaly detection. Furthermore, we disregarded data from the z-axis, as it provided no substantive insights into the horizontal (x,y) movement dynamics of the subjects, thereby streamlining our analysis towards more relevant parameters. Furthermore, during the machine learning model training phase, it became evident that utilizing the absolute x and y coordinates of the subject inadvertently led to overfitting based on their positions, rather than on the characteristics of their movements. This complication stemmed from the observation that subjects followed varied paths and occupied different locations within the room throughout the experiment, which inadvertently encoded session-specific patterns into the positional data. To address this, we shifted our approach from absolute positioning to relative displacement, focusing on the change in position (the difference between consecutive x and y coordinates) to better capture the essence of movement without the bias of specific location data. These were noted as dX and dY.

Two more features of the motion were then added, the scalar magnitude of x, y velocity and acceleration. Using the following equations:

$$v_{scalar} = \sqrt{v_x^2 + v_y^2}$$

$$a_{scalar} = \sqrt{a_x^2 + a_y^2}$$

Finally, to encode the toilet visits into the data, we added a toilet visit frequency metric which is calculated by the toilet visits over the time elapsed. The final set of features used for the sequential models were: **AccX**, **AccY**, **VelX**, **VelY**, **dX**, **dY**, **Scalar Velocity**, **Scalar Acceleration**, and **Toilet frequency**. Figure 4.3 show a sample of normalised walking pattern data. Note that the data was normalised only to help visualisation of the data. The actual data used for the model was not normalised. For the sequential models, we refined the data to create samples spanning 40 time steps, equivalent to 2 seconds, for both training and testing purposes. This involved reviewing the pre-processed data to isolate consecutive movements. We defined consecutive movement based on a set threshold for dx and dy values. Movements where the subject did not exceed a specific distance were classified as non-consecutive. We then segmented each consecutive movement sequence into 40-time-step samples, using a stride of 10. This way, we obtained 2365 sequences of walking data across all subjects which is a relatively small

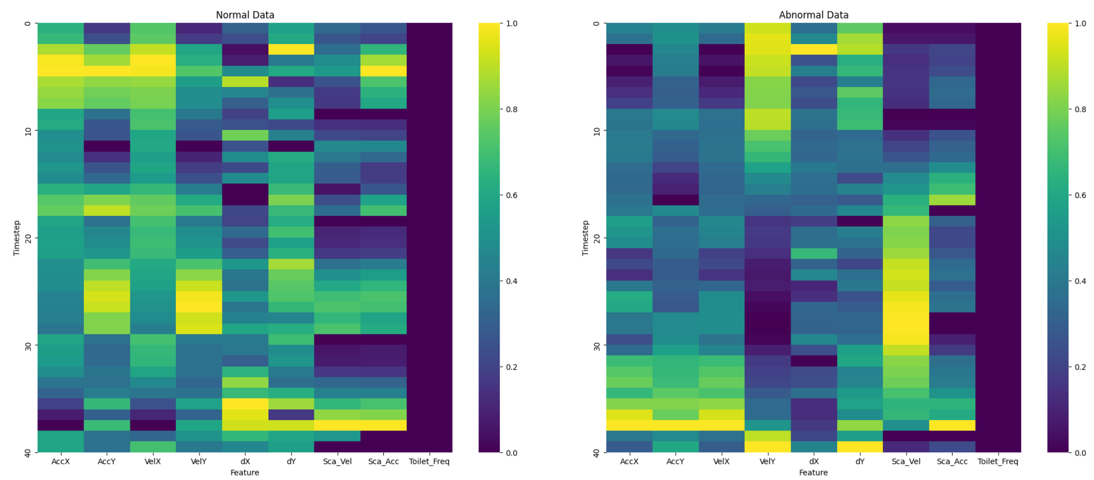


Figure 3.3: Normalised features matrix of walking for 40 time steps. Left is the normal walking pattern, right is the abnormal walking pattern

dataset for a time series classification model. This process was applied uniformly across all subjects, with each sample subsequently categorized as either normal or abnormal for the training and testing phases. The entire process can be seen at figure 3.4.

3.5 Problem Description

After refining the dataset and splitting the train and test data set, we now how to create a model to be able to classify each time series as normal and abnormal. The details of the time series classification can be defined as follows: A multivariate time series sequence can be defined as $X = \{x_1, x_2, \ldots, x_L \mid x_t \in R^M\}$, where each x_t at time t is of dimension M (representing the number of feature at each data point in the multivariate series), L is the length of the sequence. The dataset D can be defined as $D = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)\}$ where X_i is a multivariate time series and Y_i is the class label of the series. The time series classification problem consists of training a classifier on a dataset D in aiming to establish a mapping from the space of potential input variables to a probability distribution across possible values of the class variable.

3.6 Sequential models

As talked about in the background, previous studies have shown that supervised sequential learning models have proven effective in detecting anomalies in time series data, adept at distinguishing normal patterns from irregularities. Inspired by this, we aim to explore the efficacy of these models on radar data, a less conventional medium for anomaly detection. Our objective is to test various sequential learning architectures to assess their performance in identifying anomalies within radar datasets. By comparing different models of CNNs, LSTMs, and transfomers, we seek to identify the most suitable approaches for radar data analysis. On top of this, we tested on compound models of a combination of layers.

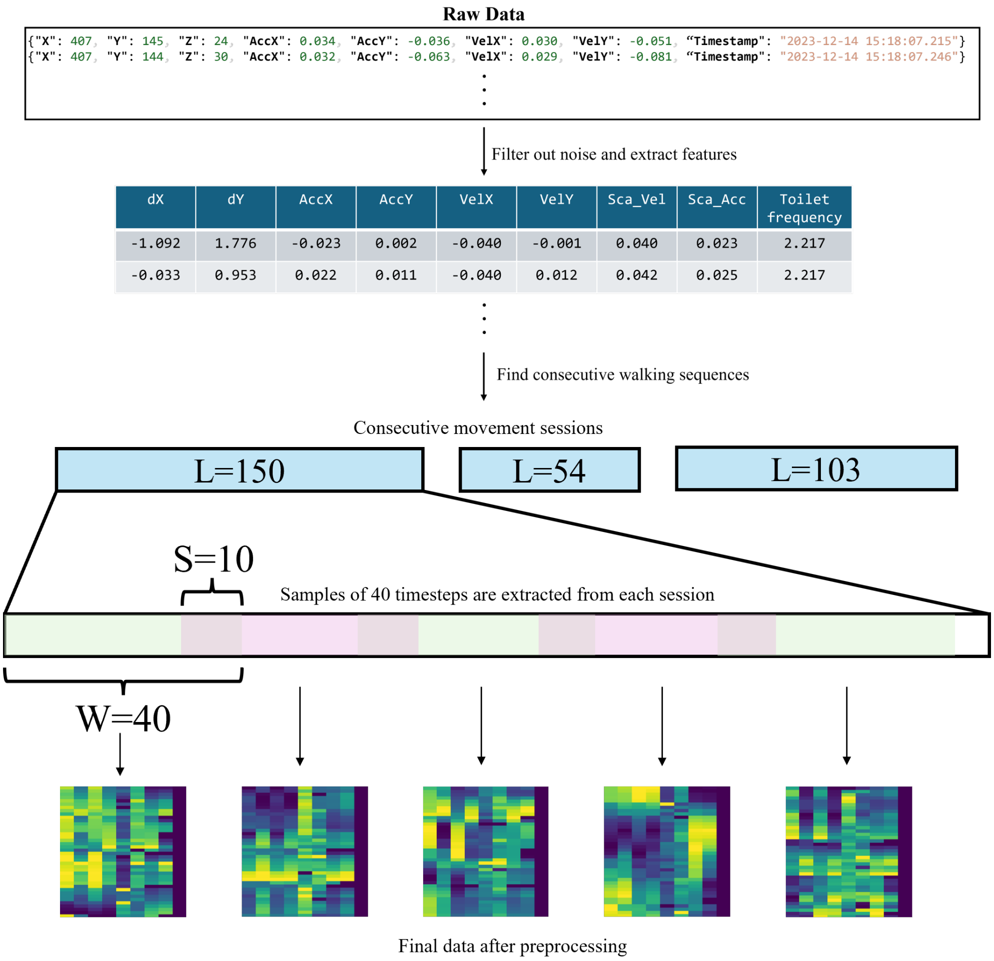


Figure 3.4: Overview of the data processing stage. L is sequence length, W being the window size, S is the stride. The final data is marked normal/abnormal depending on which session it was extracted from

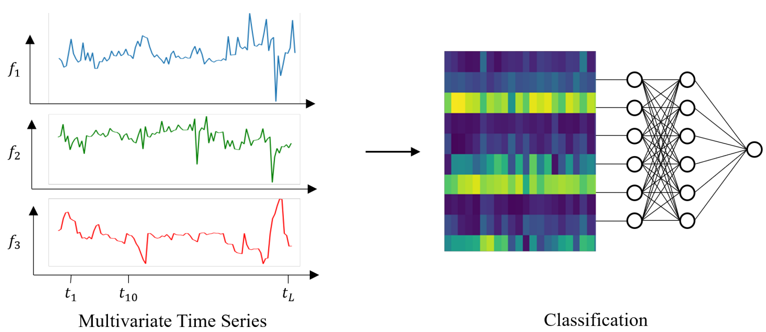


Figure 3.5: Overview of the Multivariate time series classification problem, f_i is the features of X

3.6.1 CNN

In this study, we began by experimenting with 1D convolutional layers, which are well-suited for analyzing time series data. Our foundational CNN architecture comprises three convolutional layers, incorporating ReLU activation to introduce non-linearity, enabling the model to detect and represent intricate data patterns. The absence of ReLU would render multiple layers no more effective than a single layer due to linear dependency. To enhance the learning of complex features, the output dimensions of the second and third convolutional layers were reduced by half. For the purpose of classification, average pooling was employed towards the model's end. A depiction of the base CNN architecture is provided in figure 3.6. Beyond the initial model, this research also explores combining 1D convolutional layers with LSTM and Transformer layers, forming hybrid models that will be discussed subsequently. The CNN model was also used for hyperparameter tuning and these parameters were also used for the other models as well. This will be further explored in the next section, results and discussion.

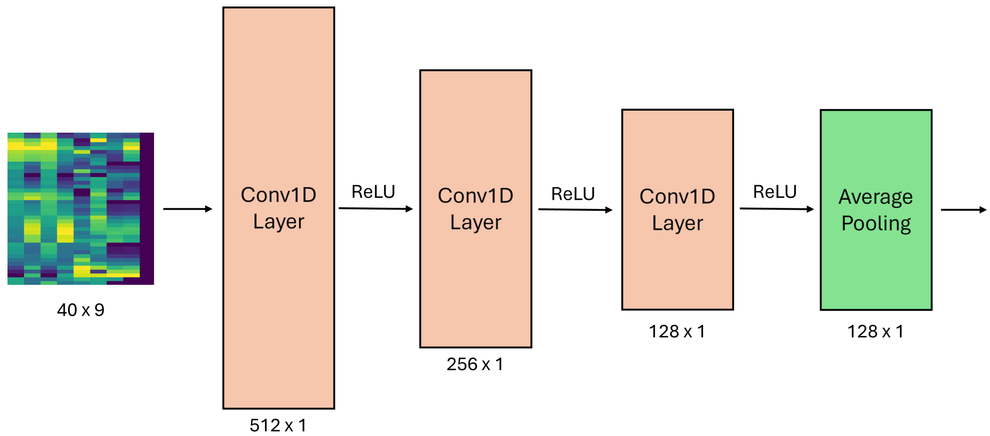


Figure 3.6: Base CNN model with 3 layers and ReLU activation

3.6.2 LSTM

Subsequently, our exploration extended to LSTM (Long Short-Term Memory) models, recognizing their proficiency in handling sequential data. The foundational LSTM setup for our experiments incorporated three LSTM layers, each equipped with hyperparameters used for the CNN, to assess their capability in processing temporal sequences through a purely LSTM-based architecture. Building on this, we also experimented with a hybrid model that integrates LSTM with CNN layers, inspired by the approach in (Livieris et al., 2020), where such a combination was applied to analyze gold price time series data, yielding superior outcomes. Echoing the methodology in the referenced study, our hybrid model begins with two 1D convolutional layers, followed by max pooling, and culminates in an LSTM layer, aiming to harness the strengths of both LSTM and CNN for enhanced performance in time series analysis. This dual-layered approach is designed to capture both the temporal dynamics via LSTM and the spatial feature extractions through CNN, potentially offering a more robust model for sequential data interpretation.

3.6.3 Transformer

This research assessed the performance of two distinct transformer-based models in detecting abnormalities within our dataset. Mentioned in section 2.4.5, both models incorporate an LSTM layer preceding a transformer encoder layer, compensating for the transformer's inherent limitation in recognizing sequence order, as depicted in figure 3.7. The rationale behind testing each model, however, differed. One model employed dual transformer encoder layers to explore the transformer's capability in addressing our specific task. Conversely, the second model integrated a CNN and LSTM layer prior to the transformer encoder, reminiscent of the CNN-LSTM framework, aiming to evaluate the synergistic effect of combining all three types of layers. Consistent with the base LSTM model, each layer in both models was configured with hyperparameters from the CNN, striving to maintain uniformity in the model's complexity and capacity for processing.

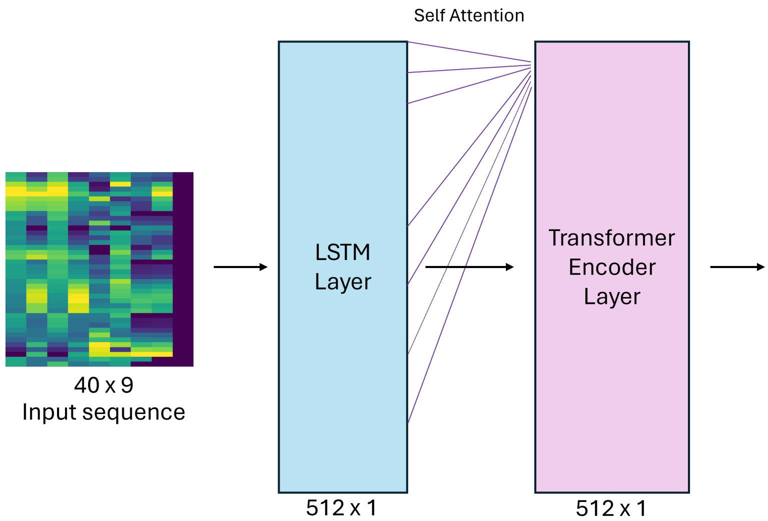


Figure 3.7: Transformer model architecture with 512 hidden units

For this study two different types of transformer models were used, one with one layer of LSTM followed by two transformer layers, and the other one with a convolutional layer followed by a LSTM, transformer layer.

3.6.4 Stratified K-fold Cross Validation

After pre-processing the data in section 3.4 and creating the models, we assessed the performance of these sequential models. Given that there were two sessions for the normal condition and one for the abnormal, the dataset exhibits a class imbalance with a ratio of approximately 2:1. Furthermore, the dataset is fairly small, with only 7 subjects and 2365 walking samples. In such cases, the standard approach of splitting data into training, validation, and test sets might not be the most efficient, as it's crucial to utilize all available data for both training and testing purposes. K-fold cross validation addresses this by dividing the dataset into a predetermined number of folds, or subsets. In each iteration, one fold is reserved for testing while the remaining folds are used for training, allowing every data point to be used for validation exactly once, ensuring comprehensive model evaluation. (Wieczorek et al., 2022) In addition to the standard K-fold, stratified K-fold (T R et al., 2023) is able to deviate from the class imbalance as it ensures that each fold maintains the same proportion of class samples as the entire dataset. Figure 3.8 (Anello, 2022) shows an example of stratified K-fold.

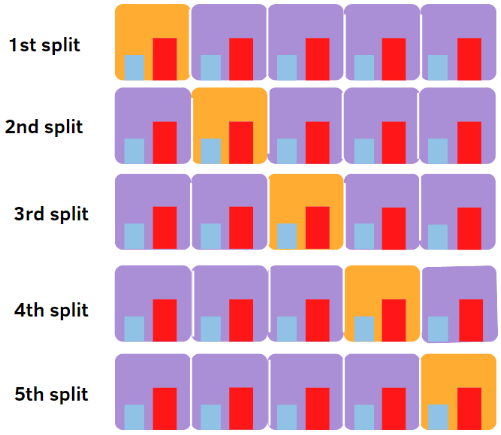


Figure 3.8: Example of Stratified K-fold Cross Validation. Yellow is test set, purple is train set (Anello, 2022)

3.6.5 Experiment Setup

The hyper parameters and details of the setup of the experiment are stated in this section. The process of obtaining the hyper parameters are explained in the results section. Every model in our study utilized the Adam optimizer (Kingma and Ba, 2014) aiming to optimize the macro F1 score, incorporating a dropout rate of 0.4 to prevent overfitting. Batch learning(Rostamizadeh et al., 2021) was applied, with batches comprising 32 samples each, enhancing the speed of convergence, improving generalization, and optimizing memory utilization by processing 32 instances simultaneously. The dataset was divided in an 80-20 ratio for training and testing, respectively. The larger portion was subjected to Stratified K-Fold cross-validation for training purposes, while the smaller segment was set aside to evaluate the final performance of the models. For K-fold, the data was split into 5 portions and trained with 40 epochs each.

3.7 Probabilistic Models

Following sequential models are the effectiveness of probabilistic models. This part will cover the further data processing that is needed for the probabilistic models. As the previous study mentioned in the background (Rojas and Tozzi, 2016), the data was tested for abnormality detection using probabilistic models. For this experiment, a Gaussian Mixture Model(GMM) and a Random Forest Classifier was used. The gaussian mixture model representing unsupervised learning, and random forest representing supervised learning.

These models were given the same data as the sequential models. However, as these models are probabilistic, they rely on 1 dimensional features rather than the sequential change of the data. As a result, the data with 40 time steps needs to be flattened through further feature extraction. Inspired by (Nabriya, 2021), the features used were the mean, standard deviation (std), min, max values of each feature of the dataset except toilet frequency. As 40 time steps is equal to 2 seconds, min, max, and standard deviation did not have much meaning for toilet frequency. Consequently, only the mean value was used. The process of extracting these features from the time series data can be seen from figure 4.3. Some of the features were dropped later on as it didn't represent a Gaussian distribution. Figure 3.10 depicts the distributions of the different features and the red rectangles show the two features that did not show a proper Gaussian distribution and were dropped. After dropping the two features, the GMM and the random forest was fit with data consisting of 31 features.

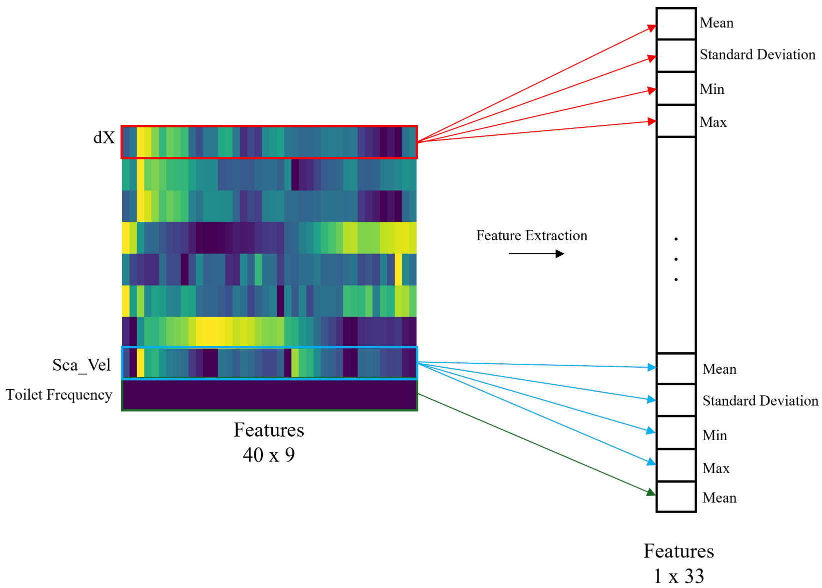


Figure 3.9: Flattening of the time series data to 33 features

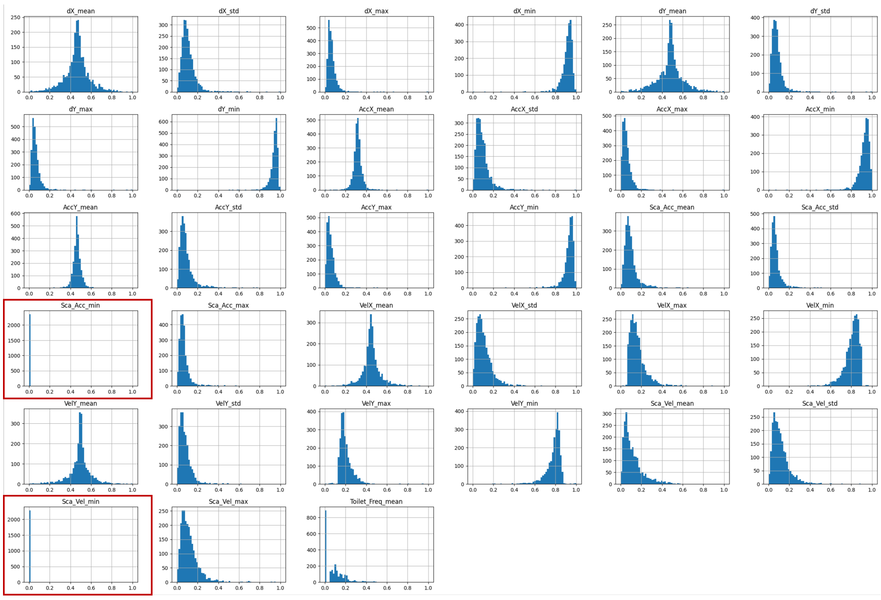


Figure 3.10: Distribution of the different features. Two red rectangles are the dropped features

Chapter 4

Results and Discussion

4.1 Task Description

We set out experiments to verify whether machine learning models can detect anomaly of walking posture from data from a home radar system. The task is to identify if the sequence of 40 time steps is normal or abnormal.

4.2 Probabilistic Models

First as stated in the methodology, we test with probabilistic models. The data we use is explained in section 3.7. The extracted 31 features from the 40 steps time series data were used to train the GMM and Random Forest model. The results are presented as tables with the Precision, Recall, and F1 score for each classification. The Accuracy, and Macro F1 score for the model is also shown.

4.2.1 GMM

The GMM was trained on the dataset unsupervised, meaning that it was trained without the label of 'normal' or 'abnormal'. Only one gaussian distribution was made hoping that everything outside the distribution is abnormal. The test set used was separated before training. The results can be seen in table 4.1.

Class	Precision	Recall	F1 Score	Accuracy	Macro F1
Normal	0.62	0.60	0.61	0.52	0.49
Abnormal	0.35	0.38	0.37		

Table 4.1: Results on the test set for the GMM

4.2.2 Random Forest Classifier

The random forest classifier, a supervised model, was trained on the same data as the GMM, using labels for 'normal' and 'abnormal'. Table 4.1 shows the performance of the model.

Class	Precision	Recall	F1 Score	Accuracy	Macro F1
Normal	0.73	0.89	0.80	0.72	0.67
Abnormal	0.69	0.43	0.53		

Table 4.2: Results on the test set for the random forest classifier

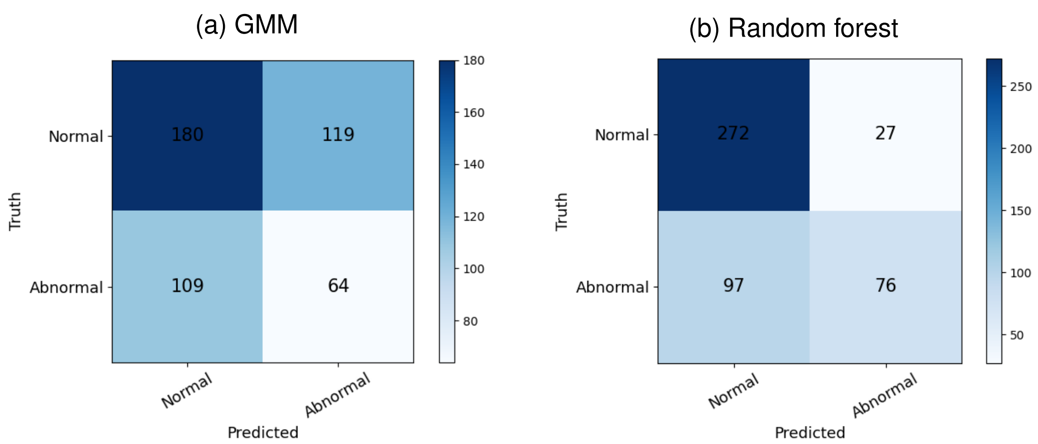


Figure 4.1: Confusion matrix of the predicted values vs the true values for probabilistic models

4.2.3 Discussion

Overall, probabilistic models were not great in classifying our dataset. The GMM's performance in detecting abnormalities is notably weak, evidenced by an accuracy of 0.52 and a macro F1 score of 0.49. The correlation matrix (figure 4.1) suggests the GMM classifies data as normal or abnormal seemingly at random. This limitation likely stems from the GMM's unsupervised nature, where it defines class boundaries based on data distribution rather than clear, predefined categories. Consequently, in cases where normal and abnormal data are not distinctly separable, the GMM struggles to classify effectively.

The random forest classifier outperformed the GMM, showing a 20% increase in accuracy, as expected from a supervised model. This improvement indicates detectable patterns differentiating abnormal from normal data within the dataset. However, as Table 4.2 reveals, the model is more effective at identifying normal data, with an F1 Score of 0.8 for normal and only 0.53 for abnormal classes. This implies that accurately classifying abnormal samples remains challenging.

4.3 Sequential Models

The probabilistic models under performed, but findings suggest supervised learning models could yield better outcomes. We then moved on to supervised sequential models to test the efficacy on classification of the time series data obtained by radar. The data used was the processed time series data of 40 steps from section 3.4. As introduced in section 3.6, five models were tested and like the probabilistic models, the results were shown in a table with the Precision, Recall, and F1 score, Accuracy, and Macro F1 score.

4.3.1 Sequential Model Hyperparameter tuning

To be able to depict the performance of the sequential models properly, it is crucial to tune the hyperparameters. Hyperparameters are the configuration settings used to structure the model, which significantly influence the model's ability to learn from the data. Common hyperparameters throughout all the models were obtained from testing on the CNN. These hyperparameters are **number of hidden nodes**, **number of layers**, and **batch size**. Other layer specific parameters were optimized during the testing of the base models. (e.g. Base LSTM model, LSTM-Tran-Tran model) The performance was tested by Accuracy and Macro F1.

Number of hidden nodes

The number of hidden nodes can be translated the number of nodes/neurons inside layers of the neural network. It can also be seen as the complexity of each layer. Depending on the complexity of the data set, the number of hidden nodes have to adjusted. If the model lacks nodes, meaning that the underlying patterns in the data are not being adequately captured by the current model architecture, the model will decrease in performance. On the other hand, if there are too many nodes in each layer, the model will overfit and thus will do poorly on unseen data. The hidden nodes tested for this study is 128, 256, and 512. Anything more than 512 was deemed computationally expensive. From table 4.3,

No. of hidden nodes	Accuracy	Macro F1
128	0.78	0.73
256	0.85	0.83
512	0.95	0.94

Table 4.3: Hyperparameter results for hidden nodes

there was a significant improvement of the model when using more nodes suggesting that the data is complex enough to make use of the additional nodes. As a result, for all models, 512 nodes were used for each layer.

Number of hidden layers

The number of layers in a neural network represents the depth of the model, with each layer potentially capturing different levels of abstraction in the data. Similar to the concept of hidden nodes, the layers' depth can be viewed as the overall complexity of the network architecture. The adequacy of the number of layers is contingent on the dataset's complexity and the specific problem at hand. A model with insufficient layers might lack the depth required to identify and learn the hierarchical patterns present in the data, leading to under performance. Conversely, an excessively deep model, with too many layers, risks overfitting, where it memorizes the training data's noise, impairing its generalization to new, unseen data. For this study, the neural network architectures explored include depths of 2, 3, and 4 layers.

No. of hidden layers	Accuracy	Macro F1
2	0.82	0.79
3	0.95	0.94
4	0.94	0.94

Table 4.4: Hyperparameter results for number of hidden layers

Table 4.3 shows that using 3 hidden layers retrieves the best results. Using 2 layers seems insufficient for the complexity of our dataset. Consequently, all the models used a total of 3 layers.

Batch size

The batch size in a neural network training process refers to the number of training examples utilized in one iteration of the model's update. It is a crucial hyperparameter that balances the model's learning efficiency and computational resource demands. A smaller batch size means that the model updates its weights more frequently, potentially leading to faster convergence but with more noise in the update process. This noise can sometimes be beneficial by providing a regularizing effect and helping the model escape local minima. However, too small a batch size may result in unstable training and require more iterations to converge.

Conversely, a larger batch size provides a more accurate estimate of the gradient, but with less frequent updates, it can lead to smoother convergence. This comes at the cost of increased memory usage and computational power. Additionally, very large batch sizes may lead to poorer generalization as the model might settle into sharper minima of the loss landscape. For this study, batch sizes of 32, 64, and 128 were evaluated. The decision to cap the batch size at 128 was driven by the trade-off between computational efficiency and the quality of the gradient estimate.

Batch size	Accuracy	Macro F1
32	0.95	0.94
64	0.93	0.94
128	0.94	0.93

Table 4.5: Hyperparameter results for batch size

From table 4.5, it seems that batch size does not make a significant difference to the performance of the results. However, batch size of 32 was used for the models as it did do marginally better than the other batch sizes.

4.3.2 Base CNN Model

As mentioned previously the architecture of the CNN base model could be seen in figure 3.6. The hyperparameter used was retrieved from section 4.3.1, with 512 hidden nodes, 3 hidden layers, and batch size of 32. Table 4.6 shows the performance of the model.

Class	Precision	Recall	F1 Score	Accuracy	Macro F1
Normal	0.95	0.97	0.96	0.95	0.94
Abnormal	0.94	0.91	0.92		

Table 4.6: Results on the test set for the 3 layer Convolutional model

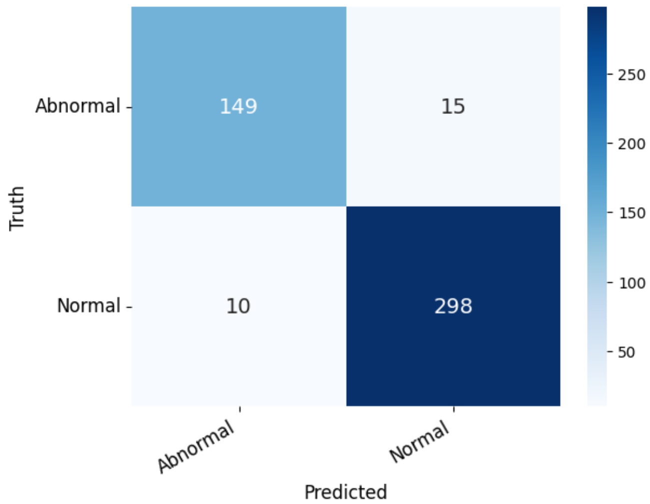


Figure 4.2: Confusion matrix of the predicted values vs the true values of the Convolutional model

4.3.3 Base LSTM Model

The base LSTM model consisted of 3 LSTM layers with the same hyperparameters as the CNN. No additional hyperparameters needed to be tuned. Table 4.7 shows the performance of the model.

Class	Precision	Recall	F1 Score	Accuracy	Macro F1
Normal	0.95	0.96	0.95	0.94	0.93
Abnormal	0.92	0.90	0.91		

Table 4.7: Results on the test set for the 3 layer LSTM model

4.3.4 CNN-LSTM Model

The CNN-LSTM model's layers were ordered by one CNN layer followed by two LSTM layers. The results can be seen in table 4.8.

Class	Precision	Recall	F1 Score	Accuracy	Macro F1
Normal	0.96	0.97	0.97	0.96	0.95
Abnormal	0.95	0.92	0.93		

Table 4.8: Results on the test set for the CNN-LSTM model

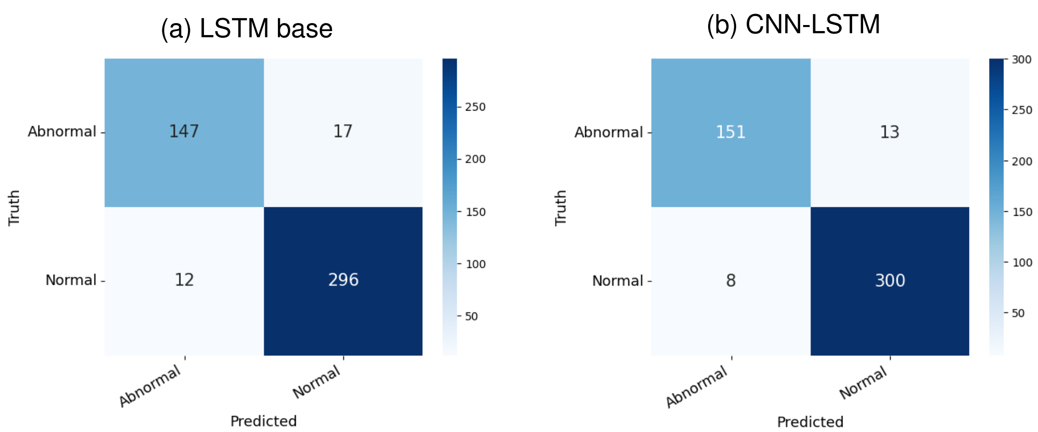


Figure 4.3: Confusion matrix of the predicted values vs the true values for LSTM models

4.3.5 LSTM-Trans-Trans Model

This model used a LSTM layer followed by 2 transformer encoder layers. The hyper-paramer of the transformer encoder layer was the number of parallel attention heads. More heads are usually better for more complex tasks. The other hyperparameters remain unchanged from the base CNN model. We tested the number of heads with 2, 4, 8, and 16 to find the value that gives the best results. In the end we chose 2 as our number of heads for the LSTM-Trans-Trans model and the CNN-LSTM-Trans model of the next section. The full results of the test can be seen from table 4.9.

No. of heads	Accuracy	Macro F1
2	0.96	0.95
4	0.95	0.95
8	0.94	0.94
16	0.93	0.93

Table 4.9: Hyperparameter results for the number of heads

Class	Precision	Recall	F1 Score	Accuracy	Macro F1
Normal	0.96	0.97	0.97	0.96	0.95
Abnormal	0.94	0.93	0.94		

Table 4.10: Results on the test set for the LSTM-Trans-Trans model

4.3.6 CNN-LSTM-Trans Model

The last model to be tested was a combination of 3 different layers. The results for this model can be seen from table 4.11. Number of heads was set to 2 and the other parameters followed the base CNN.

Class	Precision	Recall	F1 Score	Accuracy	Macro F1
Normal	0.96	0.97	0.96	0.95	0.95
Abnormal	0.95	0.91	0.93		

Table 4.11: Results on the test set for the CNN-LSTM-Trans model

4.3.7 Discussion

In our comparative analysis of sequential models, each demonstrated significant improvement over the earlier examined probabilistic models, achieving an accuracy and macro F1 score of higher than 0.90 accross all models. This remarkable performance underscores the advanced capability of these models in handling the complexity and nuances of the time series data.

The base LSTM model, showed a relatively lower effectiveness, with its accuracy of 0.94 and macro F1 of 0.93. This indicates that while LSTMs are adept at managing

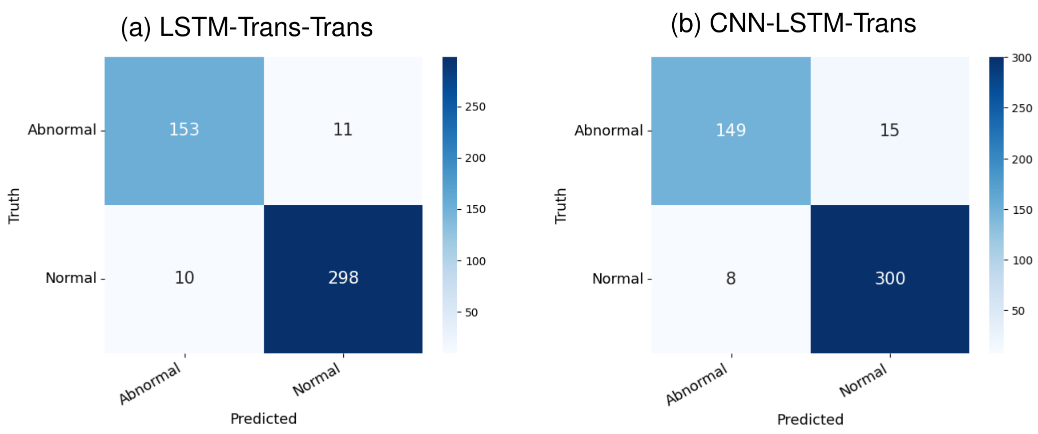


Figure 4.4: Confusion matrix of the predicted values vs the true values for transformer models

sequence data, just using LSTM layers might not fully capture the intricate patterns present in the dataset. Particularly in the context of abnormality detection where nuance in the sequence can be critical.

Conversely, the CNN-LSTM hybrid model and the CNN-trans-trans model performed equally as well in our study, both achieving an accuracy of 0.96 and a macro F1 of 0.95. The CNN-LSTM model leveraged the strengths of both CNNs in feature extraction and LSTMs in sequence memory, achieving a balanced approach for this task. Surprisingly, the compound model of all three layers did worse than than the CNN-LSTM and CNN-trans-trans model. This could be because of model complexity where the addition of more layers and mechanisms might introduce redundancy or dilute the model's focus, making it less effective at distinguishing the nuanced patterns crucial for abnormality detection in our dataset. The outcome highlights that sequential models are effective when it comes to classifying walking into normal and abnormal from the time series data from the radar of just 40 time steps. Also, careful architectural design and the need for a distinct approach when combining multiple advanced modeling techniques.

Chapter 5

Conclusions

5.1 Summary

In summary, this research successfully shows that radar, coupled with machine learning models, can identify abnormal human behavior. Radar sensors offer a privacy-conscious alternative to traditional video surveillance or smart home systems, eliminating concerns about privacy breaches and the high costs of installation. Data was gathered from seven individuals at the National Robotarium, focusing on indicators of abnormal behavior such as walking discomfort and altered restroom visit frequency. By analyzing the subjects' location and movement data captured by the radar, and after processing and feature extraction, we applied probabilistic and sequential machine learning techniques to assess their effectiveness in detecting these abnormalities.

For probabilistic models, we tested on GMMs and random forests. We've tried using a GMM to see if unsupervised learning can detect the patterns that outline abnormality. The random forest was used to see the effectiveness of a supervised probabilistic model. The GMM failed to find patterns that differentiate the two classes resulting in an accuracy of 0.52. The random forest on the other hand did somewhat better with an accuracy of 0.72. Even if the results were not very high, this suggested that it is possible for to identify abnormality using the features.

Subsequent analysis using sequential models significantly improved accuracy, exceeding 0.90 across all models. Notably, the CNN-LSTM and CNN-trans-trans models performed exceptionally well, indicating that a 1D convolutional layer effectively processes our data. Furthermore, the success of LSTMs and transformers within our specific framework underscores their utility for this type of data analysis.

Our radar-based model has various potential applications, including monitoring elderly individuals living alone. It can serve as a safeguard, alerting caregivers in case of unusual activity. Additionally, it could be valuable for patients with Urinary Tract Infections (UTI), where early detection of acute symptoms is critical. This system offers continuous monitoring while ensuring privacy.

5.2 Future work

In conclusion, while this study has resulted in an accuracy of over 0.95 when it comes to identifying abnormality for the seven subjects, it also opens the door to several avenues for future research. One immediate area of exploration involves addressing the limitations in generalizing what constitutes abnormal behavior. With only seven participants simulating abnormality, the models could easily distinguish between normal and abnormal sequences by learning each subject's specific walking patterns during the sessions. Yet, applying the model to new subjects not included in the training set proves difficult, as the model lacks prior data on their normal and abnormal behavior patterns. This issue suggests that abnormalities are highly individualistic, making broad generalizations challenging even with extensive training data from thousands of subjects.

One potential solution to this challenge would be fine tuning (Tian et al., 2023). Fine tuning is process of taking a pre-trained model and adjusting it slightly to make it suitable for a specific task. For our task, the model will be pre-trained with the big dataset of many subjects, and fine tuned with the data of just the subject in concern. This way. we have enough data to create a general enough model while being specific enough for each individual subject's data.

Enhancements to the probabilistic models could be explored in future work. Boosting the effectiveness of these models might entail improving feature selection by removing unhelpful features and adding more relevant ones. Additional features could be derived using the Fast-Fourier transform, and their importance determined through principal component analysis (PCA).

Similarly, the performance of sequential models could benefit from incorporating a broader set of features. Further, examining the correlations between features to identify and eliminate those that have little impact on distinguishing between normal and abnormal sequences could also prove beneficial.

The experimental design could be enhanced by considering the sequence of activities as an indicator of abnormal behavior. This approach is supported by background research indicating that individuals often alter their typical activity patterns under abnormal circumstances. To address these changes, we could employ a Hidden Markov Model (HMM), similar to the methodology discussed in the background section, to determine the normality or abnormality of activity sequences. Integrating HMMs with our existing neural networks could provide a more comprehensive model for detecting abnormal human behavior with the capability to detect both abnormal activity and walking patterns.

Bibliography

Annisa Aditsania, Adiwijaya, and Aldo Lionel Saonard. Handling imbalanced data in churn prediction using adasyn and backpropagation algorithm. In *2017 3rd International Conference on Science in Information Technology (ICSITech)*, pages 533–536, 2017. doi: 10.1109/ICSITech.2017.8257170.

Hande Alemdar, Halil Ertan, Ozlem Durmaz Incel, and Cem Ersoy. Aras human activity datasets in multiple homes with multiple residents. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 232–235, 2013.

Hadi Aliakbarpour, Kamrad Khoshhal, João Quintas, Kamel Mekhnacha, Julien Ros, Maria Andersson, and Jorge Dias. *HMM-Based Abnormal Behaviour Detection Using Heterogeneous Sensor Network*, page 277–285. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. URL `http://dx.doi.org/10.1007/978-3-642-19170-1`$_3$0.

Moeness G. Amin, Yimin D. Zhang, Fauzia Ahmad, and K.C. Dominic Ho. Radar signal processing for elderly fall detection: The future for in-home monitoring. *IEEE Signal Processing Magazine*, 33(2):71–80, 2016. doi: 10.1109/MSP.2015.2502784.

Eugenia Anello. K-fold cross validation for machine learning models, Nov 2022. URL `https://pub.towardsai.net/k-fold-cross-validation-for-machine-learning-models-918`

Damla Arifoglu and Abdelhamid Bouchachia. Activity recognition and abnormal behaviour detection with recurrent neural networks. *Procedia Computer Science*, 110:86–93, 2017. doi: 10.1016/j.procs.2017.06.121.

Damla Arifoglu and Abdelhamid Bouchachia. Detection of abnormal behaviour for dementia sufferers using convolutional neural networks. *Artificial Intelligence in Medicine*, 94:88–95, 2019. doi: 10.1016/j.artmed.2019.01.005.

Diane J. Cook, Aaron S. Crandall, Brian L. Thomas, and Narayanan C. Krishnan. Casas: A smart home in a box. *Computer*, 46(7):62–69, 2013. doi: 10.1109/MC.2012.328.

Xiaodong Du and Guanghui Teng. An automatic detection method for abnormal laying hen activities using a 3d depth camera. *Engenharia Agrícola*, 41(3):263–270, May 2021. doi: 10.1590/1809-4430-eng.agric.v41n3p263-270/2021.

Shirin Enshaeifar, Ahmed Zoha, Severin Skillman, Andreas Markides, Sahr Thomas Acton, Tarek Elsaleh, Mark Kenny, Helen Rostill, Ramin Nilforooshan, and Payam Barnaghi. Machine learning methods for detecting urinary tract infection and analysing daily living activities in people with dementia. *PLOS ONE*, 14(1), 2019. doi: 10.1371/journal.pone.0209909.

Donald L. Hall, Tyler D. Ridder, and Ram M. Narayanan. Abnormal gait detection and classification using micro-doppler radar signatures. In Kenneth I. Ranney and Armin Doerry, editors, *Radar Sensor Technology XXIII*. SPIE, May 2019. doi: 10.1117/12.2519663. URL `http://dx.doi.org/10.1117/12.2519663`.

Saba Hesaraki. Long short-term memory (lstm), Oct 2023. URL `https://medium.com/@saba99/long-short-term-memory-lstm-fffc5eaebfdc`.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, November 1997. ISSN 1530-888X. doi: 10.1162/neco.1997.9.8.1735. URL `http://dx.doi.org/10.1162/neco.1997.9.8.1735`.

Shuo Jia, Fei Hui, Shining Li, Xiangmo Zhao, and Asad J. Khattak. Long short-term memory and convolutional neural network for abnormal driving behaviour recognition. *IET Intelligent Transport Systems*, 14(5):306–312, 2019. doi: 10.1049/iet-its.2019.0200.

Justin M. Johnson and Taghi M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1), March 2019. ISSN 2196-1115. doi: 10.1186/s40537-019-0192-5. URL `http://dx.doi.org/10.1186/s40537-019-0192-5`.

Kinjal Joshi and Narendra Patel. A cnn based approach for crowd anomaly detection. *International Journal of Next-Generation Computing*, 12(1), Mar 2021. doi: https://doi.org/10.47164/ijngc.v12i1.185.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL `https://arxiv.org/abs/1412.6980`.

Andy Liaw and Matthew Wiener. Classification and Regression by randomForest. *R News*, 2(3):18–22, 2002. URL `http://CRAN.R-project.org/doc/Rnews/`.

HaiLong Liu, Tadahiro Taniguchi, Yusuke Tanaka, Kazuhito Takenaka, and Takashi Bando. Visualization of driving behavior based on hidden feature extraction by using deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 18(9): 2477–2489, 2017. doi: 10.1109/TITS.2017.2649541.

Pengfei Liu, Xiaoming Sun, Yang Han, Zhishuai He, Weifeng Zhang, and Chenxu Wu. Arrhythmia classification of lstm autoencoder based on time series anomaly detection. *Biomedical Signal Processing and Control*, 71:103228, January 2022. ISSN 1746-8094. doi: 10.1016/j.bspc.2021.103228. URL `http://dx.doi.org/10.1016/j.bspc.2021.103228`.

Ioannis E. Livieris, Emmanuel Pintelas, and Panagiotis Pintelas. A cnn–lstm model for gold price time-series forecasting. *Neural Computing and Applications*, 32(23): 17351–17360, April 2020. ISSN 1433-3058. doi: 10.1007/s00521-020-04867-x. URL `http://dx.doi.org/10.1007/s00521-020-04867-x`.

Sebastian Lühr, Svetha Venkatesh, Geoff West, and Hung H. Bui. *Explicit State*

Duration HMM for Abnormality Detection in Sequences of Human Activity, page 983–984. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. URL http://dx.doi.org/10.1007/978-3-540-28633-2$_1$25.

Pratik Nabriya. Feature engineering on time-series data, Jul 2021. URL https://towardsdatascience.com/feature-engineering-on-time-series-data-transforming-signal-data-of-a-smartphone-accelerometer-for-72cbe34b8a60.

Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks, 2015. URL https://arxiv.org/abs/1511.08458.

Oscar Ernesto Rojas and Clesio Luis Tozzi. *Abnormal Crowd Behavior Detection Based on Gaussian Mixture Model*, page 668–675. Springer International Publishing, Cham, 2016. URL http://dx.doi.org/10.1007/978-3-319-48881-3$_4$7.

Afshin Rostamizadeh, Anand Rajagopalan, Claudio Gentile, Giulia DeSalvo, Gui Citovsky, Laz Karydas, and Sanjiv Kumar. Batch active learning at scale. In *NeurIPS 2021*, 2021.

David E. Rumelhart and James L. McClelland. *Learning Internal Representations by Error Propagation*, pages 318–362. 1987.

Hui Song, Jiejie Dai, Lingen Luo, Gehao Sheng, and Xiuchen Jiang. Power transformer operating state prediction method based on an lstm network. *Energies*, 11(4):914, April 2018. ISSN 1996-1073. doi: 10.3390/en11040914. URL http://dx.doi.org/10.3390/en11040914.

Georgios Spanos, Konstantinos M. Giannoutakis, Konstantinos Votis, and Dimitrios Tzovaras. Combining statistical and machine learning techniques in iot anomaly detection for smart homes. In *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pages 1–6, 2019. doi: 10.1109/CAMAD.2019.8858490.

Mahesh T R, Vinoth Kumar V, Dhilip Kumar V, Oana Geman, Martin Margala, and Manisha Guduri. The stratified k-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification. *Healthcare Analytics*, 4:100247, December 2023. ISSN 2772-4425. doi: 10.1016/j.health.2023.100247. URL http://dx.doi.org/10.1016/j.health.2023.100247.

Wataru Takabatake, Kohei Yamamoto, Kentaroh Toyoda, Tomoaki Ohtsuki, Yohei Shibata, and Atsushi Nagate. Fmcw radar-based anomaly detection in toilet by supervised machine learning classifier. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2019. doi: 10.1109/GLOBECOM38437.2019.9014123.

Emmanuel Munguia Tapia, Stephen S. Intille, and Kent Larson. Activity recognition in the home using simple and ubiquitous sensors. *Lecture Notes in Computer Science*, page 158–175, 2004. doi: 10.1007/978-3-540-24646-6$_1$0.

TheFIS. The national robotarium, September 2023. URL https://www.thefis.org/project-library/national-robotarium/.

Katherine Tian, Eric Mitchell, Huaxiu Yao, Christopher D. Manning, and Chelsea Finn. Fine-tuning language models for factuality, 2023. URL https://arxiv.org/abs/2311.08401.

Waseem Ullah, Tanveer Hussain, Fath U Min Ullah, Mi Young Lee, and Sung Wook Baik. Transcnn: Hybrid cnn and transformer mechanism for surveillance anomaly detection. *Engineering Applications of Artificial Intelligence*, 123:106173, August 2023. ISSN 0952-1976. doi: 10.1016/j.engappai.2023.106173. URL http://dx.doi.org/10.1016/j.engappai.2023.106173.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. 2017. URL https://arxiv.org/pdf/1706.03762.pdf.

Katerina Vrotsou, Kajsa Ellegard, and Matthew Cooper. Everyday life discoveries: Mining and visualizing activity patterns in social science diary data. In *2007 11th International Conference Information Visualization (IV '07)*, pages 130–138, 2007. doi: 10.1109/IV.2007.48.

Lingling Wang, Ying Zhou, Rao Li, and Lieyun Ding. A fusion of a deep neural network and a hidden markov model to recognize the multiclass abnormal behavior of elderly people. *Knowledge-Based Systems*, 252:109351, Sep 2022. doi: 10.1016/j.knosys.2022.109351.

Jenny Weisenberg, Paul Cuddihy, and Vrinda Rajiv. Augmenting motion sensing to improve detection of periods of unusual inactivity. In *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, HealthNet '08, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605581996. doi: 10.1145/1515747.1515751. URL https://doi.org/10.1145/1515747.1515751.

Jerzy Wieczorek, Cole Guerin, and Thomas McMahon. K-fold cross-validation for complex sample surveys. *Stat*, 11(1), May 2022. ISSN 2049-1573. doi: 10.1002/sta4.454. URL http://dx.doi.org/10.1002/sta4.454.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 37 (9):11121–11128, June 2023. ISSN 2159-5399. doi: 10.1609/aaai.v37i9.26317. URL http://dx.doi.org/10.1609/aaai.v37i9.26317.

Meriem Zerkouk and Belkacem Chikhaoui. *Long Short Term Memory Based Model for Abnormal Behavior Prediction in Elderly Persons*, page 36–45. Springer International Publishing, Cham, 2019. URL http://dx.doi.org/10.1007/978-3-030-32785-9$_4$.

Yinlong Zhang, Wei Liang, Xudong Yuan, Sichao Zhang, Geng Yang, and Ziming Zeng. Deep learning based abnormal behavior detection for elderly healthcare using consumer network cameras. *IEEE Transactions on Consumer Electronics*, pages 1–1, 2023. doi: 10.1109/TCE.2023.3309852.

Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1):162–169, 2017. doi: 10.21629/JSEE.2017.01.18.

Appendix A

Participants' information sheet

Participant Information Sheet

Project title:	**Facilitating health and wellbeing by developing systems for early recognition of urinary tract infections - Feather**
Principal investigator (PI):	Kia Nazarpour
Researcher(s):	Lynda Webb; Saber Mirzaee, Emilyann Nault, Bhavith Manapoty, Jeong Younwoo, Aidan McConnell-Trevillion
PI contact details:	kianoush.nazarpour@ed.ac.uk

This study is certified according to the Informatics Research Ethics Process, RT number **671984**. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers? The research team are members of the Feather Project from The University of Edinburgh, Heriot-Watt University and research partners. Kia Nazarpour is Principal Investigator. Nigel Goddard (UoE), Steve Leung (NHS Lothian), Lynne Baillie (HW) and Mauro Dragone (HW) are Co-Investigators. Saber Mirzaee Bafti, Lynda Webb, and Emilyann Nault who are researchers in the team. They may be accompanied by PhD and BSc students, Aidan McConnell-Trevillion, Bhavith Manapoty and Younwoo Jeong, while conducting this project.

What is the purpose of the study? The objective of this study is to explore how non-invasive sensing technologies, for example wearables and RADAR, can register movement patterns during the activities of daily living in a simulated home environment.

Do I have to take part? No – participation in this study is entirely up to you. You may decide to stop being a part of the research study at any time without explanation. You have the right to ask that any data you have supplied to that point be withdrawn or destroyed. You have the right to refuse to answer or respond to any question that is asked of you. You have the right to have your questions about the procedures answered (unless answering these questions would interfere with the study's outcome). If you have any questions as a result of reading this information sheet, you should ask the researcher before the study begins. You will have the option of taking part in the longer or shorter experiments.

What will happen if I decide to take part? The study will be conducted at the National Robotarium in the Laboratory of Robotic Assistive Living – LARA, which is an accessible home, based on the "Concept Blackwood House".

In this experiment, participants will be monitored over a three-hour period. The participants will engage in various everyday activities such as reading, eating, drinking, and more. During some periods they will be asked to simulate the experience of pain and the impact this pain may have on how you sit/move/walk.

The extent of this simulation is entirely under your control. Researchers will monitor your actions from a separate room, and our sensing technology including, for example a contactless radar device, a camera and a wrist-worn wearable device will track and record your movements throughout the experiment.

At the start of the experiment, you will receive instructions about the activities and their duration. You will be provided with all the equipment for these activities which include reading a magazine, watching TV, paying video games on tablets, completing a jigsaw puzzle. Making and consuming tea/coffee, preparing and eating a sandwich, and visits to an adjacent room/facility within the flat, where you will be requested to wait for a set amount of time.

Time Commitment The experiment will not take more than three hours and 15 minutes including the introduction and breaks.

Are there any risks associated with taking part? There are no risks associated with participation.

What will happen to the results of this study? The results of this study will be used to evaluate and iterate our sensing technology and inform the development of data analysis methods to detect changes in people's movement and behavioural patterns. The results may be summarised in published articles, reports and presentations. Data may also be used for future research. Raw data will be anonymised and archived on a public data repository as per the requirement of the funding agency.

Data protection and confidentiality Your movement records from our movement tracking device and the wearable, along with videos from the camera will be electronically stored on a secure hard drive. Your data will be processed in accordance with Data Protection Act (2018). All information collected about you will be kept strictly confidential. Your data will only be viewed by the research team, as described above. All electronic data will be stored on a password-protected encrypted computer, at The University of Edinburgh's School of Informatics' secure file servers and all paper records will be stored in a locked filing cabinet in the Principal Investigators office. Your consent information will be kept separately from your responses to minimise risk. The data will be kept as long as is required by Government Regulation, typically 4 years.

What are my data protection rights? The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection

Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

Who can I contact? Kia Nazarpour will be glad to answer your questions about this study at any time. You may contact him at kianoush.nazarpour@ed.ac.uk. If you want to find out about the final results of this study, you can contact Dr Nazarpour directly. If you wish to make a complaint about the study, please contact inf-ethics@inf.ed.ac.uk. When you contact, please provide the study title and detail the nature of your complaint.

Signature

By signing below, you are agreeing that:

- you have read and understood the Participant Information Sheet,
- questions about your participation in this study have been answered satisfactorily,
- you are aware of the potential risks (if any), and
- you are taking part in this research study voluntarily (without coercion).

Participant's Name (Printed)*: ____________________
Participant's signature*: ____________________
Date: ____________________
**Participants wishing to preserve some degree of anonymity may use their initials.*

Appendix B

Participants' consent form

THE UNIVERSITY of EDINBURGH
informatics

Participant Consent Form

Project title:	**Facilitating health and wellbeing by developing systems for early recognition of urinary tract infections - Feather**
Principal investigator (PI):	Kia Nazarpour
Researcher(s):	Lynda Webb; Saber Mirzaee, Emilyann Nault, Bhavith Manapoty, Jeong Younwoo, Aidan McConnell-Trevillion
PI contact details:	kianoush.nazarpour@ed.ac.uk

By participating in the study, you agree that:

- I have read and understood the participant information sheets (PIS) and I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction.
- My participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights. In addition, should I not wish to answer any particular question or questions, I am free to decline.
- I consent to my anonymised data being used in academic publications and presentations.
- I understand that my anonymised data will be stored for the duration outlined in the Participant Information Sheet.

Please tick yes or no for each of these statements.

		Yes	No
1.	I agree to being video recorded.	☐	☐
2.	I allow my data to be used in future ethically approved research.	☐	☐
3.	I understand that my responses will be kept strictly confidential. I give permission for members of the research team to have access to my anonymised data and responses. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research.	☐	☐
4.	I agree to take part in this study.	☐	☐

Name of Participant	Date	Signature
Name of person taking consent	Date	Signature
Name of Chief Investigator	Date	Signature

To be signed and dated in presence of the participant (appears under "Name of person taking consent")

Once this has been signed by all parties the participant can receive a copy of the signed and dated consent form, participant information sheets and any other written information provided to the participants. A copy of the signed and dated consent form will be kept with the project's main documents in a secure location.