Anomaly Detection in Behavioural Patterns of Elderly Individuals with Mild Cognitive Impairment using Hidden Markov Model

A dissertation submitted to The University of Manchester for the degree of Master of Science in Health Data Science in the Faculty of Biology, Medicine and Health

2023

Student ID: 9734505

School of Health Sciences

Table of Contents

Word Count: 11,343

List of Figures

List of Tables

List of Algorithms

Abstract

Background: Mild Cognitive Impairment (MCI) represents a crucial transitional stage between normal aging and dementia. Early detection of behavioural anomalies in MCI individuals presents opportunities for early intervention. This dissertation explores the application of Hidden Markov Models (HMMs) for detecting MCI-related behavioural anomalies within smart home environments.

Objectives: The objectives of this study are threefold: (1) Develop a methodology to simulate MCI-related behavioural anomalies, (2) Apply HMM to detect these anomalies, and (3) Evaluate HMM's anomaly detection capabilities to other state-of-the-art methods.

Methods: A methodology was devised to simulate anomalies in daily activity sequences associated with the two subtypes of MCI. Anomalies included sleep disruptions, repetitive sequences, and fatigue, aligning with the gradual cognitive decline observed in MCI individuals. HMM was chosen as the primary anomaly detection method and its performance of was compared with benchmark models, Long Short-Term Memory (LSTM), and Gated Recurrent Unit Autoencoders (GRU-AE) using various evaluation metrics.

Results: Simulated anomalies effectively replicated common MCI-related behavioural patterns, demonstrating the success of the methodology. HMM consistently outperformed LSTM and GRU-AE in detecting aMCI and naMCI anomalies across all evaluation metrics. Statistical analysis revealed significant differences in favour of HMMs, highlighting both its high effectiveness and consistency in identifying MCI-related behavioural anomalies.

Discussion: While the anomaly simulation method successfully introduced MCI-related anomalies, there remains bias in the simulated dataset given that MCI individuals typically have huge variability in behaviour. For anomaly detection, analysing other contexts besides activity pattern such as activity duration could potentially enhance anomaly detection performance. Future studies could also research online learning techniques to adapt to changing habits and routines. This research has direct implications in the field of social care and neurology, with hopes of facilitating early disease interventions.

Declaration of Authorship

I, Tech Onn Ding, declare that the work presented in this dissertation, titled "Anomaly Detection in Behavioural Patterns of Elderly Individuals with Mild Cognitive Impairment using Hidden Markov Model" is my own. I confirm that no portion of the work referred to in this dissertation has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright and the ownership of intellectual property rights

Chapter 1: Introduction

The global population is experiencing a significant demographic shift, with a notable increase in the proportion of elderly individuals. To illustrate this, the proportion of the global population aged 65 years or older is anticipated to increase from 10% in 2022 to 16% in 2050 (United Nations, 2019). With this significant shift in demographic, one of the most pressing concerns that will grow alongside the aging population is the prevalence of neurodegenerative disorders such as Mild Cognitive Impairment (MCI) and dementia. Both disorders are part of a spectrum of cognitive disorders that impact memory, thinking, and reasoning abilities.

Dementia is a condition caused by the progressive loss of neuron function in the brain. It is characterised by gradual impairment of various cognitive functions such as memory, thinking, speech, and mobility (Patterson, 2018). The most common type of dementia is Alzheimer's disease which contributes to 60-80% of dementia cases (Arifoglu and Bouchachia, 2017, Yamasaki and Kumagai, 2021). As the elderly population increases at a rapid pace, the incidence of dementia will also grow. In 2018, there were approximately 50 million people globally living with dementia, a number forecasted to rise to 82 million by 2030 (Patterson, 2018). In this context, the significance of early detection of the disease cannot be underestimated. Timely detection plays an important role in slowing down disease progression and enhancing the quality of life for the affected individuals. However, 75% of dementia cases are often overlooked and only obvious when reaching a more advanced stage, resulting in a decline in cognitive abilities that severely impacts an individual's capacity to perform daily activities (Hodges et al., 2010).

Since an effective treatment for dementia does not exist to this date, a lot of research have focused on the detection of early symptom manifestation. At this critical juncture, understanding and detecting MCI provides a pivotal opportunity for early detection of dementia because it is commonly referred to as a precursor of dementia (Petersen et al., 2001; Knopman and Petersen, 2014). To put this into perspective, the annual conversion rate from MCI to dementia is 10 to 15% compared to 1 to 4% observed in the average elderly population (Csukly et al., 2016). In contrast to dementia, MCI is characterised by more subtle cognitive deficits and can be further classified into two distinct subtypes: amnestic and non-amnestic

MCI. In amnestic MCI (aMCI), memory loss is a mainstay symptom whereby the affected individual has a higher frequency of forgetting things compared to usual (Jekel et al., 2015; Yamasaki and Kumagai, 2021). On the other hand, non-amnestic MCI (naMCI) does not display significant memory loss, but other cognitive domains such as language and attention are impaired (Jekel et al., 2015; Yamasaki and Kumagai, 2021).

These MCI symptoms are often manifested in daily behavioural patterns such as sleep disturbances, unorganised lifestyle, and inability to complete tasks (Jekel et al., 2015, Tabira et al., 2020, Yamasaki and Kumagai, 2021). For example, an elderly person with aMCI may wake up in the middle of the night, sleep more during the day, or repeat certain activities multiple times throughout the day. Naturally, from a medical point of view, these deviations in daily activities offer the potential to serve as indicators of MCI.

Common assessment methods for behavioural patterns have traditionally relied on self-rated or informant-based questionnaires and direct task observation (Arifoglu and Bouchachia, 2019; Yamasaki and Kumagai, 2021). However, these methods are subject to bias, lack continuous monitoring, and may not be representative of the elderly individual's authentic performance within their home environment (Arifoglu and Bouchachia, 2019; Yamasaki and Kumagai, 2021). Consequently, it is important to recognise how recent advancements in technology, machine learning, and statistical methods can offer a promising avenue to enhance the accuracy in detection anomalous behaviours. More formally, anomaly detection refers to identifying patterns of data that deviate from the expected 'normal' pattern (Bakar et al., 2016).

With sensing devices becoming more accessible, there is enhanced capabilities for smart homes to offer continuous and objective monitoring of subtle behavioural changes in real-life settings. Moreover, prior studies have shown that elderly individuals typically prefer to reside in a familiar and private home environment due to rising costs and preference of personal space, independence, and autonomy (Wiles et al., 2012). By leveraging ambient sensors within smart homes and integrating machine learning methods to analyse the sensor data output, it becomes possible to create personalised interventions aimed at enhancing the quality of life for the elderly population. This approach holds promise as a potential novel biomarker-based diagnostic tool for MCI, thus driving the core motivation behind this dissertation.

Many approaches have been proposed to learn the behaviour of individuals through analysing smart home sensor data. One approach that has gained considerable attention Hidden Markov Model (HMM). HMM is a generative probabilistic modelling technique that is based on stochastic processes and utilises joint probabilities to model hidden states based on a set of observations. It is well-established in modelling sequential patterns and their temporal dependencies, making it suitable for the analysis of activity sequences derived from sensor data (Rabiner, 1989). The effectiveness of HMM have been demonstrated across diverse domains, including speech recognition and bioinformatics (Rabiner, 1989; Yu, 2010; Ramapatruni et al., 2019). In the context of human behaviour, HMM has the capacity to learn different types of behaviour given an observed sequence of daily activities. By training an HMM using sequential and longitudinal activity data, it has potential for identifying deviations from normal behaviour patterns that are potential indications of MCI.

Currently, there is no publicly available dataset with anomalous behavioural activity sequences, specifically those exhibited by individuals with MCI. Curating such a realistic dataset requires extensive time, effort, and expertise for experimental setup. Given the scarcity of MCI behavioural data, simulation of abnormal behaviour is a pragmatic solution to this. While a few existing studies have proposed methods to artificially generate abnormal behavioural activities reflecting that of an elderly person with dementia, they did not introduce MCI-related anomalies and did not showcase a gradual deterioration in cognitive status. This dissertation will build upon existing simulation works not only by introducing anomalies related to aMCI and naMCI but also focus on showing the gradual change in behaviour as a result of cognitive decline. Thus, a comprehensive method is proposed to artificially simulate abnormal activity patterns of an elderly individual with MCI that reflect real-life scenarios.

The main contributions of the present dissertation are outlined as follows:

1. Introduction of a novel methodology for generating synthetic data that reflects g the gradual manifestation of anomalous behaviour exhibited by elderly individuals with aMCI and naMCI.
2. This study is the first to apply HMM for anomaly detection of a simulated aMCI and naMCI dataset.

The rest of the dissertation is organised as follows. Chapter 2 provides an overview of existing literature work. Chapter 3 lists the aims and objectives. Chapter 4 describes the publicly accessible dataset used, data pre-processing steps, the simulation of MCI-related activity sequences patterns, HMM for building normal behaviour and anomaly detection, and evaluation metrics. Chapter 5 presents the simulation results and model evaluation. Chapter 6 provides a summary of the results, compares the results to related studies, discusses the limitations of the experiments, and talks about future works.

Chapter 2: Literature Review

The data produced by sensors in a smart home is inherently noisy, multidimensional, ordered temporally, lacks uniformity, and is generated in an endless sequence (Bakar et al., 2016). Naturally, due to these complicated features, sensory readings require further processing and analysis before it is ready to be used for learning elderly behavioural pattern. This is done at through low-level analysis whereby sensory readings are annotated with activity labels such as eating, sleeping, and going to the toilet. This dissertation does not focus on the activity recognition phase of human activities but rather utilised a publicly available smart home dataset that has already performed such analysis.

With the annotated sensor data, higher activity level analysis is performed on activity labels and the patterns are learned to model elderly behaviour. Subsequently, changes to the elderly's 'normal' behaviour pattern can be detected through anomaly detection methods. These changes could be deviations in different context domains such as location, time, duration, and sequence. Bakar et al. (2016) introduced two approaches to detect behavioural changes: Profiling and Discriminating. Profiling involves learning the normal behavioural pattern and then evaluate whether the new incoming sensor data matches this model or not. On the other hand, discriminating involves learning anomalies from historical data and actively search the new incoming data for similar anomalies.

Due to the scarcity of anomalous data in a smart home environment, the profiling approach is more practical to accomplish, and the rest of this chapter will focus on covering studies that have employed this approach.

2.1 Types of profiling approaches for abnormal behaviour detection

A range of machine learning techniques have been used in existing studies for modelling 'normal' pattern and then detecting anomalies from it. These include probabilistic, clustering, generative, discriminative, etc.

Probabilistic methods take advantage of the distribution of the sensor training data to determine boundaries of anomalies. Ordóñez et al. (2015) presented an approach to detect anomalies in smart home environments by extracting behavioural patterns of individuals using Bayesian statistics. The behaviour was statistically modelled by three probabilistic features: sensor activation likelihood, sensor sequence likelihood, and sensor event duration likelihood. The authors then utilised a Bayesian credible interval (BCI) approach for anomaly detection. BCI represents a range of plausible values for the various features and any values outside the BCI interval are classified as anomalous. Bayesian methods offer the advantage of incorporating prior knowledge of living patterns into the model, which can enhance its performance. However, these methods may encounter challenges in situations where data availability is limited and may struggle to distinguish routine behaviours across different periods, potentially leading to difficulties in accurately detecting anomalies.

Clustering methods have also gained attention due to their potential in effectively detecting deviations and modelling human behaviour in an unsupervised manner. These methods involve clustering the data into distinct groups and then identifying data points that do not belong to any of the clusters as anomalies. This method was employed by Lundstrom et al. (2016) whereby clusters of behaviour patterns were first modelled by random forest without prior labels. The output from the random forest is then mapped onto a 2D space to reveal data clusters using agglomerative clustering techniques. Additionally, transitions between clusters are modelled using a third-order Markov chain. After modelling the behaviour pattern, an online detection technique is applied to identify various types of deviations. This technique encompasses the detection of temporal deviations using z-scores of the time components, spatial deviations using Local Outlier Factor (LOF) values, and transitional deviations using transition probabilities.

Most studies utilise a generative model called HMM because it is a simple yet efficient algorithm for sequential smart home data which has a lot of uncertainties coming from the different behaviours of different people (Forkan et al., 2015; Bakar et al., 2016; Sánchez, Lysaker, and Skeie, 2020). It also works well on small datasets (Wang et al., 2023). Monekosso and Remagnino (2009) first learned the daily activities routine using clustering techniques and then utilised HMM to define normal behaviour. New incoming sequence is compared to the HMM-built model and the output is given in terms of log-probability indicating how likely the new sequence is a behaviour deviating from normal. However, many studies focus on analysing behaviour using only a single context, which can lead to a high false alarm rate when making decisions. For instance, identifying abnormal MCI-related behaviour based solely on activity data may not be sufficient; considering additional contextual data could improve accuracy. As a result, the Forkan et al. (2015) proposed an integrated system using different methods to profile normal pattern and detect anomalies: HMM for activity pattern, Gaussian distribution for routine behaviour, and exponential smoothing for physiological data. At the end, the authors propose the use of Fuzzy Logic to integrate all the outputs from the models and predict an outcome.

2.2 Behaviour profiling of individuals with MCI and Dementia

Compared to studies that have conducted anomaly detection on general elderly behaviour, a limited number of studies have performed it specifically on datasets with MCI and dementia-related anomalous behaviour.

Arifoglu, Charif, and Bouchachia (2020) used graph convolutional networks to learn normal behavioural pattern through granular-level sensor activations. For example, the activity of preparing coffee contains many sub-activities such as approaching the sink, turning the water on, filling up the coffee machine, etc. They then deployed the trained model on new data with simulated dementia-related abnormal behaviour and any data instance that has confidence probability lesser than a set threshold is flagged as an abnormal behaviour. The proposed method can identify abnormal behaviour from activity patterns and granular-level of activities in the context of cognitive decline. The same authors further extended their investigation into other methods to detect dementia-related abnormal behaviour (Arifoglu et al., 2021). In this study, they employed recursive auto-encoders (RAEs) to create a hierarchical tree structure that can detect abnormal behaviour of dementia patients at both activity and sub-activity levels. The reconstruction error was used to detect anomalies at both levels. The authors noted that while Long Short-Term Memory and Convolutional Neural Networks outperformed RAEs, they are supervised methods that require extensive labelled training data. In addition, behaviour of dementia patients change as the disease progresses and the abnormal data will need to be updated, making supervised methods impractical to implement in a real-life setting. While RAE and graph convolutional networks are unsupervised methods, the authors acknowledged that both methods do not consider the connections between different time segments and thus disregards the temporal relationships across them.

A study conducted by Riboni et al. (2015) aimed to support early detection of MCI for elderly people. This study used supervised learning and rule-based reasoning to model detailed information on MCI-related abnormal behaviour. The system processes raw data from ambient sensors to infer simple actions and passed them into a Markov Logic Network (MLN) to deduce the activities and correlate sequences of sensor events. One major limitation of this study is the rigid nature of supervised learning and non-probabilistic rules. Imposing pre-defined rules on abnormal behaviour will depend on the specific home environment and the specific habits of the elderly person, thus it is not generalisable to other home environments. Additionally, the quality of the inferred rules heavily relies on the expertise of the domain experts who annotate the training data. To overcome these limitations, the authors extended their research by proposing an approach to automatically learn rule-based definitions of behavioural anomalies for MCI detection (Janjua, Riboni, and Bettini, 2016). In this approach, a training set with normal and abnormal behaviours was used to infer rules based on features extracted from temporal sequences of sensor events. While the results were comparable to those achieved using manually-defined rules, the authors noted that this method is reliant on deterministic rules and makes it susceptible to anomaly mispredictions when there is noise in sensor data.

Given the limitations of existing studies in addressing the variable nature of abnormal behaviour detection of MCI and dementia, the choice of methodology is important. These cognitive conditions exhibit behaviour patterns characterised by temporal dependencies and progressive changes, necessitating an adaptable approach. HMM emerge as an appropriate tool for addressing these challenges. HMM is particularly well-suited to capture the inherent sequential nature of behaviour data, making it a natural fit for modelling the progression of activities and the potential deviations associated with cognitive decline. This exploration into HMM's application aligns with the necessity of unsupervised methods that can adapt to changing behaviour over time, as supervised approaches are often hindered by the scarcity of labelled data and the evolving nature of abnormal behaviours.

2.3 Anomaly Simulation

Due to lack of capabilities to perform a real-life experiment, publicly accessible data without any anomalies annotated was used for this dissertation. Thus, abnormal data simulation presents as a practical solution for this purpose. Only a few studies have artificially simulated their anomalies.

One study modified a real-world dataset to create health related abnormal behaviour (Forkan et al., 2015). Anomalies such as frequent toilet visit and sleeping without dinner were introduced. However, this study did not provide any details as to how the anomalies were created and the anomalies were not created with MCI and dementia in mind. In another study modelling smart home activity pattern artificially introduced abnormality by randomly shuffling the final 5 activities of a normal data instance (Poh et al., 2019). While it utilised HMM for abnormal behaviour detection, the anomalies were ambiguous and not representative of MCI and dementia behaviours.

Several other studies that were mentioned in the previous subsection simulated dementia-related behaviours (Ariflogu, Charif, and Bouchachia, 2020; Arifoglu et al., 2021). Specifically, these studies focused on detecting two distinct kinds of anomalies that emulate real-life scenarios of dementia: repeating activities and disruption in sleep. Repeating activities were simulated by introducing multiple occurrences of an activity such as 'preparing dinner' within a random area of a normal activity sequence. Disruption in sleep anomaly was generated by inserting activities such as 'eating' and 'bed to toilet' in normal night-time activity sequences. These anomalies were introduced by modifying the Aruba dataset which is a publicly accessible dataset provided by Washington State University's Centre for Advanced Studies in Adaptive Systems (CASAS) program (Cook, 2010).

While these studies have made progress in introducing dementia-related anomalies, there remains a notable gap in the diversity of simulated anomalies and simulation of MCI-related anomalies. Other types of anomalies prevalent in the daily lives of elderly individuals with cognitive decline have not received comparable attention. Furthermore, detecting MCI-related behavioural symptoms is arguably more important for early intervention. The gradual deterioration of cognition, a hallmark of these conditions, is also not effectively captured in existing anomaly simulation approaches.

Chapter 3: Aims and Objectives

Previously, HMM has been applied to model the behaviour of elderly individuals and detect anomalous behaviour with good results. While these efforts have yielded valuable insights, the existing literature appears to lack a focused evaluation of HMM to profile normal behaviour and detect abnormal behaviour specifically in the context of gradual cognitive decline observed in individuals affected by aMCI and naMCI. To bridge this gap, this study aims to assess the effectiveness of HMM in detecting deviations in activity patterns of elderly people afflicted with aMCI and naMCI.

To that end, the research objectives include:

1. Develop a robust methodology for simulating abnormal behaviour patterns that mimic the gradual cognitive decline observed in aMCI and naMCI. This involves introducing a diverse range of anomalous behaviours that closely resembles real-life MCI scenarios.
2. Assess the effectiveness of an HMM-based anomaly detection framework in detecting simulated behavioural deviations resembling varying levels of cognitive impairment seen in aMCI and naMCI.
3. Compare the anomaly detection capabilities of HMM to other state-of-the-art unsupervised methods, namely Long short-term memory (LSTM) and autoencoder, to discern their performance and limitations in detecting MCI-related behavioural deviations.

Chapter 4: Methods

4.1 Data source: Aruba CASAS smart home dataset

The dataset used is obtained from Washington State University's Centre for Advanced Studies in Adaptive Systems (CASAS) program. The goal of the CASAS dataset is to enable research and development of technologies that can understand and respond to human behaviour in real-world settings (Cook, 2010). To this end, real-time data was collected from sensors to analyse the resident's behaviour to improve future smart home living. For this study, one of the many public datasets curated by the CASAS programme called Aruba is used. Numerous other studies have also used this dataset for behavioural anomaly detection due to its large size, high level of activity annotation, diverse sensor types, and realistic home environment (Gupta et al., 2020; Bala Suresh and Nalinadevi, 2022; Fahad and Tahir, 2021). The Aruba dataset consists 220 days of continuous data (no missing days) with 11 activities annotated (Figure 1). Thirty-nine sensors, including thirty-four passive infrared sensors and five temperature sensors were used during data collection in a home of an elderly woman. A schematic of the sensor layout in the smart home is shown in Figure 2 while a sample of the collected and labelled sensor data is shown in Figure 3.

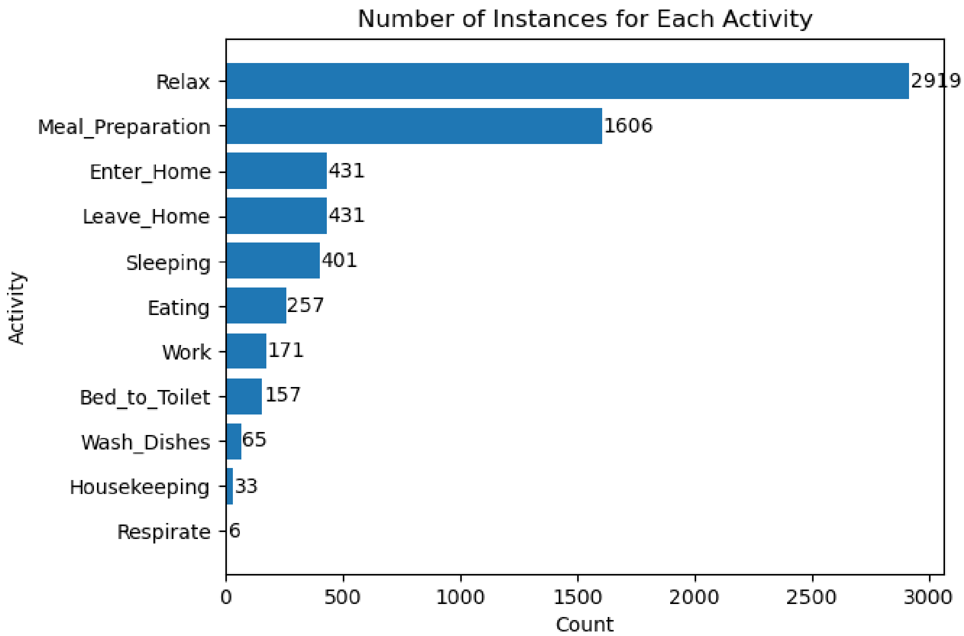


Figure 1: Instances of activity labels from Aruba dataset

These are the activities that form the daily activity sequence pattern of the dataset. Respirate only occurs 6 times and was removed during data pre-processing.

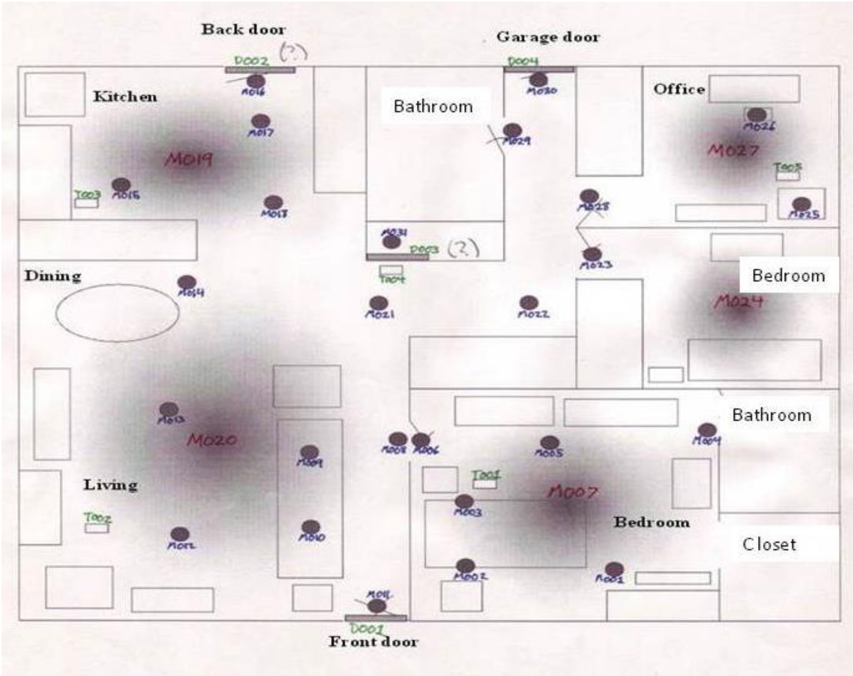


Figure 2: Schematic diagram of sensor layout in the smart home (Cook, 2010)

A total of 39 sensors were installed in the home of an elderly woman which are represented by the sensor IDs. These include 34 passive infrared sensors and 5 temperature sensors. The sensors are motion sensors if their IDs begin with "M", door closure sensors if their IDs begin with "D" and temperature sensors if their IDs begin with "T".

```
2010-11-04 05:40:43.642664 M003 OFF Sleeping end
2010-11-04 05:40:44.223548 M003 ON
2010-11-04 05:40:45.939846 M005 ON
2010-11-04 05:40:46.310862 M003 OFF
2010-11-04 05:40:51.303739 M004 ON Bed_to_Toilet begin
2010-11-04 05:40:52.342105 M005 OFF
2010-11-04 05:40:57.176409 M007 OFF
2010-11-04 05:40:57.941486 M004 OFF
2010-11-04 05:43:24.021475 M004 ON
2010-11-04 05:43:26.273181 M004 OFF
2010-11-04 05:43:26.345503 M007 ON
2010-11-04 05:43:26.793102 M004 ON
2010-11-04 05:43:27.195347 M007 OFF
2010-11-04 05:43:27.787437 M007 ON
2010-11-04 05:43:29.711796 M005 ON
2010-11-04 05:43:30.279021 M004 OFF Bed_to_Toilet end
2010-11-04 05:43:34.261135 M005 OFF
2010-11-04 05:43:35.941892 M007 OFF
2010-11-04 05:43:40.821615 M007 ON
2010-11-04 05:43:45.619681 M007 OFF
2010-11-04 05:43:45.7324 M003 ON   Sleeping begin
2010-11-04 05:43:52.044085 M003 OFF
2010-11-04 05:43:53.185335 M002 ON
```

Figure 3: Example raw smart home sensor data output with activity labels

This is just a snippet of raw data output from one day out of the 220 days of recorded data. The first column represents the date, the second column displays the time, the third column *lists sensor IDs, the fourth column shows sensor status, and the final column contains annotations for various activities.*

4.2 Data Pre-processing

The aim of data pre-processing is to convert the data into a suitable format for analysis and HMM modelling. While the steps involved closely align with previous studies that have performed HMM analysis on the Aruba dataset for activity pattern analysis (Poh et al., 2019; Tan et al., 2022), it's important to note that these studies employed different anomaly simulation methods that are not directly relevant to MCI. There are three stages including data cleansing, data transformation and data partitioning.

4.2.1 Data Cleansing

During the data cleansing stage, noisy data segments are removed to enhance the quality of the dataset. Noisy data refers to information containing characteristics that are different compared to the rest of the data (Poh et al., 2019). Three categories of noise were identified and removed. Firstly, the activity “Respirate” was removed because it is an activity that only appeared 6 times throughout the dataset’s 220 days (see Figure 1). Secondly, the sensor ID, sensor status, and time columns were removed because they are not needed when building an HMM model for activity pattern analysis. The third type of noise removed includes data in between activity start and end times, which can affect the HMM model by introducing empty strings. Lastly, the same activities that were recorded multiple times consecutively were removed. Closely spaced repetitions do not reflect realistic behaviour and may bias the model's perception of activity transitions by assigning a higher importance to repeated activities.

4.2.2 Data Transformation

In the activity column, the rows indicating the end of an activity are removed. Additionally, to fulfil the integer input requirement of the HMM library used, each of the unique activities was mapped to a numeric symbol, as detailed in Table 1. An example of the cleansed and transformed data is shown in Figure 4. The encoded activity column will be used as input for HMM.

Activities	**Numerical Labels**
Bed_to_Toilet	0
Eating	1
Enter_Home	2
Housekeeping	3
Leave_Home	4
Meal_Preparation	5
Relax	6
Sleeping	7
Wash_Dishes	8
Work	9

Table 1: Mapping of original activity labels to encoded numerical values for input into HMM

Date	Activity	Encoded_Activity
2010-11-04	Sleeping	7
2010-11-04	Bed_to_Toilet	0
2010-11-04	Sleeping	7
2010-11-04	Meal_Preparation	5
2010-11-04	Relax	6
2010-11-04	Housekeeping	3
2010-11-04	Meal_Preparation	5
2010-11-04	Eating	1
2010-11-04	Wash_Dishes	8
2010-11-04	Housekeeping	3
2010-11-04	Leave_Home	4
2010-11-04	Enter_Home	2
2010-11-04	Leave_Home	4
2010-11-04	Enter_Home	2
2010-11-04	Relax	6
2010-11-04	Meal_Preparation	5
2010-11-04	Eating	1
2010-11-04	Work	9
2010-11-04	Relax	6
2010-11-04	Meal_Preparation	5
2010-11-04	Relax	6
2010-11-04	Meal_Preparation	5
2010-11-04	Relax	6
2010-11-04	Eating	1
2010-11-04	Wash_Dishes	8
2010-11-04	Work	9

Figure 4: Example sensor data format after data cleansing and transformation

In comparison to the raw data depicted in Figure 3, certain columns that are not necessary for constructing the HMM were excluded. These columns consist of time, sensor ID, and sensor status. Additionally, labels ending with 'end' were omitted to avoid redundancy, as the HMM would interpret them as repetitions of the same activity.

4.2.3 Data Partitioning

As previously mentioned, the Aruba dataset consists of 220 days of data. While existing studies allocated majority of days as for training the normal behaviour profile, it can be argued that one month's worth of data can adequately capture the daily routine of an elderly individual. This potentially also allows more time to detect abnormal behaviour. Furthermore, to simulate gradual cognitive decline representative of the transition from MCI to dementia, emphasis was placed on retaining majority of the days for testing. Accordingly, the data was split into a 31-day training set, a 31-day validation set, and a 158-day test set.

4.3 Synthesis of abnormal activities associated with MCI

One of the objectives of this study is to create a robust methodology that realistically replicates the behavioural patterns observed by elderly individuals affected by aMCI and naMCI. To achieve this, the types of behavioural anomalies to introduce were selected based on insights from existing literature on MCI-related behavioural changes, coupled with considerations regarding the nature of the Aruba dataset, which will be elaborated in the subsections below. Given the dataset's limitations, notably its composition of only 11 distinct activities, with just a subset relevant for MCI-related behavioural analysis, this study focused on introducing three specific types of MCI-related behavioural anomalies: 1) Repetitive activities, 2) Sleep disruption, and 3) Fatigue throughout the day.

4.3.1 Simulation methods for different types of behavioural anomalies

To simulate the different types of behavioural anomalies listed above, specific sets of activities are inserted into the daily sequence of activities in the test set. The placement of each type of behavioural anomaly will depend on the severity and subtype of MCI and the specifics are detailed in the subsequent subsection 4.3.2.

(1) Repetitive activities

A hallmark trait of aMCI is the potential for elderly people to forget whether they have performed certain daily activities, leading to repetition of the activities (Saives, Pianon, and Faraut, 2015). This becomes medically significant when the frequency of occurrence increases. For example, an elderly individual with forget to eat dinner before sleeping or forget that he or she has already eaten dinner, leading to the event of meal preparation late at night.

To replicate this behaviour as accurate as possible, a repetitive set of actions were manually inserted within a random segment of daily activity sequence. This generates multiple repetitions of the chosen activity happening at unusual times of the day. The activities introduced include eating, meal preparation, and housekeeping. Consider an activity sequence of a day, denoted as $S = d_1, d_2, d_3, \dots, d_x$, where each d_i represents a specific activity. To fabricate the anomalous version, a repetitive anomalous sequence A was inserted within the sequence S, resulting in a modified sequence with the following pattern: $S = d_1, d_2, d_3, \ldots, A, \ldots, d_x$ (see Algorithm 1). The anomalous sequence A, in this case, encompasses, for instance, the activities 'eating, meal preparation, eating, meal preparation,' replicating the repetitive characteristic associated with MCI.

INITIALISE $S = < d_1, d_2, d_3, \ldots, d_x >$ **AND** $A = < a_1, a_2, a_1, a_2, \ldots, a_x >$

CALCULATE insertion_index **AS** the integer division of length S by 2

INSERT A into S at position insertion_index

OUTPUT $S = < d_1, d_2, d_3, \ldots, a_1, a_2, a_1, a_2, \ldots, a_x, \ldots, d_x >$

Algorithm 1: Pseudocode for simulation of repetitive activities

Repetitive sequences were introduced to reflect MCI-related memory impairment. Instead of choosing a random position to insert the sequence, the insertion index was chosen as the integer division of length S by due to the chosen activities to include. It makes more sense to repeat eating, meal preparation, and housekeeping activities during the day.

(2) Sleep disruption

Another type of simulated abnormal behaviour is disruption in sleep. Many studies have indicated that individuals with MCI have poorer sleep quality characterised by difficulty falling asleep, early awakening, and night-time awakenings (Hayes et al., 2014; Alberdi et al., 2018; Raewter et al., 2020).

This disruption was simulated by infusing a normal night-time sequence with additional activities like 'eating' and 'bed to toilet'. This is to capture the elderly person waking up in the middle of the night for various reasons such as drinking water and going to the toilet. Consider the standard daily activity sequence $S = d_1, d_2, d_3, \ldots, d_x$, where d_1 may or may not be the activity 'Sleeping'. To replicate sleep disruption, the anomalous sequence A was incorporated either immediately after d_1 if d_1 is 'Sleeping' or as the first activity if d_1 is not 'Sleeping'. This process transforms S into two possible outcomes $S = d_1, A, d_2, d_3, \ldots, d_x$ or $S = A, d_1, d_2, d_3, \ldots, d_x$ (see Algorithm 2). The specific sequence A in this case, could be 'bed to toilet, drinking, sleeping' illustrating the night-time disturbances observed in both aMCI and naMCI.

```
INITIALISE S = < d1, d2, d3, …, dx > AND A = < a1, a2, …, ax >

IF d1 IN S is 'Sleeping':
    INSERT A into S after d1
ELSE:
    INSERT A into S before d1

OUTPUT S = < d1, a1, a2, …, ax, d2, d3, …, dx > OR < a1, a2, …, ax, d1, d2, d3, …, dx >
```

Algorithm 2: Pseudocode for simulation of sleep disruption to reflect poor sleep quality observed in elderly individuals with MCI.

(3) Fatigue during the day

Fatigue during the day can serve as an important indicator of cognitive health, particularly when considering its associations with conditions such as depression. Within a smart home environment, daytime fatigue can manifest in the forms of increased frequency of resting or daytime napping. These behaviours have consistently been identified by existing research as early indicators of cognitive decline (Leng et al., 2019; Rawtaer et al., 2020; Chimamiwa et al., 2022). Findings from MRC Cognitive Function and Ageing Study suggested that reported napping was associated with an increased risk of cognitive impairment in 10 years (Keage et al., 2012).

To simulate these specific behavioural anomalies, instances of resting and napping were introduced into the daily activity sequence, with variations depending on the cognitive stage being modelled. More instances of resting activities were inserted throughout the day to model early-stage MCI while resting and daytime naps were incorporated to model the later stages of MCI. In a typical daily activity sequence denoted as $S = d_1, d_2, d_3, \dots, d_x$, the modified sequence for increased relaxation activities could appear as $S = d_1, d_2, d_3, r_1, d_4, r_2, d_5, \dots, d_x$, where r_1 and r_2 represent 'Relaxing' activities (see Algorithm 3). On the other hand, for increased daytime napping, the sequence might take the form of $S = d_1, d_2, d_3, \dots, s_1, \dots, d_x$, where $s_1$correspond to activity 'Sleeping' (see Algorithm 4).

```
INITIALISE S = < d1, d2, d3, …, dx > AND A = < r1, r2, …, rx >

CALCULATE num_insertion_slots AS (length of S minus length of A + 1)

CALCULATE insertion_indices by randomly sampling from num_insertion_slots

FOR item in A:
    SELECT a random index from insertion_indices
    INSERT item into S at the selected index

OUTPUT S = < d1, d2, d3, …, r1, d4, r2, d5, …, dx >
```

Algorithm 3: Pseudocode for simulation of increased relaxing activities to reflect fatigue and exhaustion symptoms observed by elderly individuals with MCI.

```
INITIALISE S = < d1, d2, d3, …, dx > AND A = < s1 >

CALCULATE insertion_index AS the integer division of length S by 2

INSERT A into S at position insertion_index

OUTPUT S = < d1, d2, d3, …, s1, …, dx >
```

Algorithm 4: Pseudocode for simulation of daytime napping observed by elderly individuals with MCI.

4.3.2 Simulating method for gradual cognitive decline in aMCI and naMCI

The simulation methodology implemented in this study systematically captures the progression of cognitive decline in both aMCI and naMCI individually, starting from day 80 of the test dataset. This is designed to enable a thorough assessment of the HMM's effectiveness in detecting the gradual changes in abnormal behaviours associated with these two subtypes of MCI.

Individuals affected by aMCI often experience subtle cognitive changes, characterised by slight memory loss (Jekel et al., 2015; Yamasaki and Kumagai, 2021). To represent the progressive changes in cognitive capabilities observed in aMCI, the simulation is as follows:

- From day 80 to 99, sleep disruption anomalies were introduced in 10 random days. Notably, repetitive activities are excluded during this phase, as they align with more advanced memory decline (Chimamiwa et al., 2022).
- From day 100 to 119, sleep disruption and repetitive sequences were introduced in 10 random days.
- From day 120 to 139, sleep disruption, repetitive sequences, and an increased frequency of relaxing activities were introduced in 10 random days.
- From day 140 to 157, sleep disruption, repetitive sequences, relaxing activities, and daytime napping were introduced in 9 random days.

For naMCI, individuals affected by this disease subtype do not exhibit significant memory loss but instead display worse sleep quality (Jekel et al., 2015; Leng et al., 2019; Yamasaki and Kumagai, 2021). Consequently, the simulation slightly varies compared to aMCI as follows:

- From day 80 to 99, sleep disruption anomalies are introduced in 10 random days.
- From day 100 to 119, worsened sleep quality, characterised by a higher frequency of sleep disruption than in aMCI, is introduced in 10 random days.
- From day 120 to 139, the simulation includes worsened sleep disruption and an increased frequency of relaxing activities in 10 random days.
- From day 140 to 157, the simulation replicates worsened sleep disruption, relaxing activities, and daytime napping in 9 random days.

In total, 39 anomalous days were simulated for each MCI subtype. The stepwise introduction of these various artificial anomalies illustrates the gradual cognitive changes observed in elderly individuals affected by MCI. Since the anomalous days were chosen at random, the simulations for both MCI subtypes were repeated ten times. This repetition serves to assess the general effectiveness and prediction reliability of the HMM in detecting these MCI-related anomalies and to compare its relative effectiveness in distinguishing between the subtypes.

4.4 Hidden Markov Model for normal behaviour profiling and anomaly detection

4.4.1 Normal behaviour profiling

HMM is the technique used for both normal behaviour profiling and anomaly detection in this study. HMM is a generative probabilistic model made up of two key components: observations and hidden states (Rabiner, 1989; Forkan et al., 2015). The observations represent tangible data points collected from ambient sensors, which are the various activities in this case. The sequence of observations is represented by $O = o_1, o_2, \ldots, o_M$, where M is the number of observations. On the other hand, the hidden states are not directly observable but can be inferred based on the observations. In this case, the hidden states are the different types of behaviour observed by the elderly individual. These can be shown as $Q = q_1, q_2, \ldots, q_N$, where N is the number of states.

Together, HMM requires that there be an observable process O whose outcomes are influenced by the outcomes of hidden states Q. In other words, the types of behaviours influence the activity sequence of an elderly person. Given the above definition, the problem of interest now is how do we build an HMM to model the observed sequence of activity sequence performed by an elderly person.

The way the HMM algorithm models the observed sequence is by learning three probability matrices (Rabiner, 1989; Jurafsky and Martin, 2023):

- Transition probability matrix $A = a_{ij}$ captures the probability of one hidden state Q_i transitioning to another state Q_j. It is important to note that the HMM makes a very strong assumption in which the current state is only influenced by the state immediately preceding it and not any other states. This is represented by Equation 1:

$$a_{ij} = P\left[q_{t+1} = Q_j \mid q_t = Q_i\right], \text{ each } a_{ij} \geq 0 \text{ } and \text{ } \sum_{j=1}^{N} a_{ij} = 1 \tag{1}$$

* t is the time instant associated with state changes

- Emission probability matrix $B = b_i(o_k)$ (Equation 2) contains the probability of an observation o_t being generated from hidden state j:

$$b_j(o_k) = P[o_k\ at\ t \mid q_t = Q_j], \text{ each } b_{ij} \geq 0\ and\ \sum_{j=1}^{N} b_{ij} = 1$$

(2)

- Initial state probability matrix π (Equation 3) represents the probabilities associated with starting in each hidden state:

$$\pi_i = P[q_1 = Q_i]$$

(3)

Together, an HMM model can be represented by $\lambda = (A, B, \pi)$

These matrices of probabilities are learned from historical data through a process called parameter estimation.

One popular method for parameter estimation in HMMs is the Baum-Welch algorithm, which falls under the umbrella of Expectation-Maximization (EM) algorithms (Rabiner, 1989; Jurafsky and Martin, 2023). The Baum-Welch algorithm (Equation 4) iteratively updates the parameters of the HMM to maximise the likelihood of the observation sequence given the model. In other words, the algorithm adjusts the model parameters $\lambda = (A, B, \pi)$ to maximise $P(O|\lambda)$.

$$\lambda * = argmax_{\lambda \in \Lambda} P(O|\lambda) \text{ where } \Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_k\} \text{ set of available models}$$

(4)

One hyperparameter to set is the number of hidden states which can vary according to the design of the model (Forkan et al., 2015). Since the different types of behaviour are unknown, a range of hidden state numbers were selected to explore and assess each model's performance. This was done by using the validation set and relied on a combination of log-likelihood, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The best performing number of hidden states is indicated by the AIC and BIC minimum and log-likelihood reaching a plateau. This was then adopted as the optimal configuration for modelling the observed activities.

The systematic procedure undertaken to select the optimal number of hidden states followed the tutorial from the Python hmmlearn library. The steps are as follows along with a more detailed explanation with pseudocode in Algorithm 5:

1. The range of hidden states was varied from 2 to 15
2. For each distinct number of hidden states, the HMM was initiated using 10 different random initialisations of parameters, which was done to ensure reproducibility of results across different runs.
3. The model was trained on the training data.
4. Given the trained HMM, the log-likelihood, AIC and BIC values were obtained by testing the model on the validation set.

```
INITIALISE lists for AIC, BIC, and log-likelihood

FOR hidden_state_value IN range(2,16):

    best_log_likelihood = None
    best_model = None

    FOR random_seed IN range(10):
        INITIALISE HMM model WITH hidden_states, n_iterations, and random_seed
        TRAIN HMM model on train_data WITH encoded_activities
        CALCULATE log_likelihood FOR validation_data USING HMM model
        IF best_log_likelihood IS None OR log_likelihood > best_log_likelihood:
            SET best_log_likelihood TO log_likelihood
            SET best_model TO HMM model

    CALCULATE AIC value FOR best_model USING validation_data
    CALCULATE BIC value FOR best_model USING validation_data
    STORE best_log_likelihood, AIC, and BIC values FOR hidden_states_value

END
```

Algorithm 5: Pseudocode for optimal HMM hidden state selection

The log-likelihood is computed by the Forward-Backward algorithm (Equation 5), which calculates the probability that a given sequence of observations O is generated by the HMM model λ (Rabiner, 1989; Forkan et al., 2015; Jurafsky and Martin, 2023). This algorithm employs dynamic programming to compute these probabilities efficiently.

$$P(O|\lambda) = \sum\nolimits_{s\in valid(Q)} \prod\nolimits_{t=1}^{T} a_{q_{t-1}q_t} b_{q_t}(o_t) \tag{5}$$

After the optimal number of hidden states is determined, the training set activity sequences were used to train the HMM and build a normal activity pattern profile of the elderly individual (see Algorithm 6).

```
best_log_likelihood = None
best_model = None

FOR random_seed IN range(10):
    INITIALISE HMM model WITH hidden_states, n_iterations, and random_seed
    TRAIN HMM model WITH encoded_activities from train_data
    CALCULATE log_likelihood FOR validation_data USING HMM model
    IF best_log_likelihood IS None OR log_likelihood > best_log_likelihood:
        SET best_log_likelihood TO log_likelihood
        SET best_model TO HMM model

END
```

Algorithm 6: Pseudocode for training HMM with normal activity pattern

4.4.2 Anomaly detection

To detect abnormal days, the HMM model built was used to classify daily activity sequences within the test set as possible abnormalities based on the log-likelihood values. However, the initial step involves determining an appropriate log-likelihood threshold. To identify the optimal threshold for anomaly detection, the validation set served as a representation of 'normal' daily behaviour. By subjecting each day's activity sequence from the validation set to the trained HMM, the minimum and maximum log-likelihood values were extracted to represent the log-likelihood range indicative of ‘normal’ behaviour. Subsequently, 3000 threshold choices were linearly sampled from this range, with each threshold’s performance evaluated using the F1-score and the one with the best score was chosen for anomaly detection. The F1-score metric was chosen because it balances the trade-off between precision and recall, making it suitable to find a threshold that achieves a balance between minimising false positives and false negatives. Once the optimal threshold was selected, any daily data instance from the test set that outputs a log-likelihood lower than the threshold was classified as potentially anomalous.

```
LINEARLY SAMPLE 1000 threshold_choices BETWEEN min AND max OF val_log_likelihoods

best_threshold = None
best_f1_score = None

FOR threshold IN threshold_choices:
    SET predicted_labels WHEN test_log_likelihoods > threshold
    CALCULATE the f1_score WITH true_labels and predicted_labels
    IF best_f1_score IS None OR f1_score > best_f1_score:
        best_threshold = threshold
        best_f1_score = f1_score
```

Algorithm 7: Pseudocode for sampling a range of thresholds from the validation log-likelihood and selecting the optimal threshold based on the best F1-score.

```
INITIALISE list FOR test_log_likelihoods

FOR date IN test_data:
    CALCULATE log_likelihood FOR encoded_activities from test_data USING trained HMM model
    STORE log_likelihood IN test_log_likelihoods list

FOR log_likelihood IN test_log_likelihoods:
    SET anomaly IF log_likelihood < best_threshold

END
```

Algorithm 8: Pseudocode for calculating log-likelihood of each day in test dataset when compared to trained HMM and performing anomaly detection based on best threshold from Algorithm 7.

4.5 Benchmark models

4.5.1 Long short-term memory

LSTM is a type of recurrent neural network (RNN) architecture designed to address the challenges of modelling sequential data effectively (Zerkouk and Chikhaoui, 2020). One of the distinctive characteristics of LSTM is its ability to learn temporal features over a long period

(Zerkouk and Chikhaoui, 2020). This trait allows the model to decide whether to keep or discard information based on weighted importance, making it suitable to understand human behaviour. The LSTM architecture consists of three layers: an input layer, a hidden layer, and an output layer. These layers are fully interconnected, with the hidden layer playing a crucial role in capturing and processing sequential dependencies.

The reason why LSTM was selected as a benchmark method for HMM is because it is an established state-of-the-art method for both activity recognition and anomaly detection of smart home data. Many prior studies have either implemented LSTM as a primary anomaly detection tool or used it as a benchmark in their research (Poh et al., 2019; Zerkouk and Chikhaoui 2019; Zerkouk and Chikhaoui 2020; Arifoglu et al., 2021; Shahid et al., 2023).

The LSTM algorithm implemented in this dissertation closely follows the methodology described in the study by Poh et al. (2019). The only difference lies in the hyperparameters; this dissertation utilised 4 layers with 128 hidden units per layer, whereas Poh et al. (2019) employed 1 and 2 layers with 32 and 64 hidden units per layer. This change was made based on better overall performance observed when testing on the simulated dataset in this dissertation. Algorithm 9 demonstrates the Python code for creating the LSTM architecture, which utilizes the 'torch.nn' library. Additionally, it's worth noting that 'Keras' offers a built-in LSTM library, but it couldn't be executed in this study due to software and technical constraints.

```
CLASS LSTMModel WITH input_size, hidden_size, num_layers, output_size:
  FUNCTION to initialise LSTMModel:
    CALL super(LSTMModel, self).__init__()
    SET self.hidden_size TO hidden_size
    SET self.num_layers TO num_layers
    SET self.lstm TO nn.LSTM(input_size, hidden_size, num_layers, batch_first=True)
    SET self.fc TO nn.Linear(hidden_size, output_size)
    SET self.softmax TO nn.Softmax(dim=2)

  FUNCTION forward(self, x):
    SET hidden_state TO torch.zeros(self.num_layers, x.size(0), self.hidden_size).to(x.device)
    SET cell_state TO torch.zeros(self.num_layers, x.size(0), self.hidden_size).to(x.device)
    SET lstm_output, _ TO self.lstm(x, (hidden_state, cell_state))

    SET fc_output TO self.fc(lstm_output)
    RETURN self.softmax(fc_output)
```

Algorithm 9: Pseudocode for the LSTM model architecture.

4.5.2 Gated recurrent unit-based autoencoder

Another benchmark anomaly detection method used in this dissertation is the autoencoder, which is a type of artificial neural network used for unsupervised learning. They are made up of an encoder network which encodes input data into a lower-dimensional representation, and a decoder network which attempts to reconstruct the encoded input data. Similar to LSTM, they consist of an input layer, hidden layer, and an output layer. Autoencoders are designed to capture the patterns and features of data and have been shown to have good performance in anomaly detection (Cowton et al., 2018; Arifoglu et al., 2021).

Because the traditional autoencoders do not intrinsically take into account temporality and sequential data, this study utilised gated recurrent unit-based autoencoders (GRU-AE) which is also a type of RNN. While GRU-AE has not been employed for smart home behaviour anomaly detection specifically, it was demonstrated to have good ability in detecting sequential data with temporal dependencies (Cowton et al., 2018). Thus, it is believed to be a good benchmark model in the context of smart home activity pattern data.

The GRU-AE architecture used in this study followed standard designs, with the encoder consisting of GRU layers that reduce the input dimensionality while retaining important features. The specifics of the GRU-AE architectures and their hyperparameters are shown in Algorithm 10.

```
CLASS GRU_Autoencoder WITH input_size, hidden_size:
  FUNCTION to initialise GRU_Autoencoder:
    CALL super(GRU_Autoencoder, self).__init__()
    SET self.embedding TO nn.Embedding(input_size, hidden_size)
    SET self.encoder TO nn.GRU(hidden_size, hidden_size)
    SET self.decoder TO nn.GRU(hidden_size, hidden_size)
    SET self.output_layer TO nn.Linear(hidden_size, input_size)
    SET self.relu TO nn.ReLU()

  FUNCTION forward(self, x):
    SET embedded TO self.embedding(x)
    SET embedded TO embedded.permute(1, 0)
    SET encoded, _ TO self.encoder(embedded)
    SET decoded, _ TO self.decoder(encoded)
    SET reconstructed TO self.output_layer(decoded)
    SET reconstructed TO self.relu(reconstructed)
    RETURN reconstructed
```

Algorithm 10: Pseudocode for the GRU-based autoencoder model architecture.

4.6 Evaluation Metrics

In the real-world application of the profiling and anomaly detection method presented in this study, the challenge of evaluating its effectiveness arises due to its unsupervised nature. This challenge becomes more apparent given the complexity and variability of MCI symptoms. Anomalies can vary significantly between individuals affected by these disorders, making what is considered anomalous for one person potentially appearing as a coincidental outlier for another. Nonetheless, since this study introduced artificial anomalies, ground truth data is available for evaluating the model's performance.

To assess how effective HMM and the benchmark models are at detecting MCI-related anomalies, the following metrics are used: Accuracy, Precision, Recall, F-measure, Sensitivity, Specificity, and Receiver Operating Characteristic (ROC) curve. For each evaluation metric, the mean performance over 10 iterations of simulation were calculated, and the standard deviation was determined to determine the model's variability under different scenarios.

Accuracy quantifies the percentage of correctly identified cases (Equation 6). While accuracy will be calculated and reported, its significance is influenced by the prevalence of classes in the dataset, often resulting in the frequently occurring class having more weight (Fahim and Sillitti, 2019; Arifloglu and Bouchachia, 2019). Given the imbalanced nature of the datasets in anomaly detection systems, this metric should be interpreted with caution.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative

Precision (Equation 7) is a measure that accurately identifies positive cases from all the predicted positive instances, making it particularly useful when the cost of False Positive is high. In contrast, recall (Equation 8) evaluates the correct identification of positive cases among all the actual positive instances, making it suitable when the cost of False Negatives is high. Since these two metrics are computed by averaging across classes, they are more appropriate for interpreting models tested on imbalanced datasets (Fahim and Sillitti, 2019). High precision and recall are desirable for building a good anomaly detection system (Fahim and Sillitti, 2019). F1-score (Equation 9) is also a useful evaluation metric because it balances precision and recall.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{9}$$

The ROC curve plots the true positive rate against the false positive rate for a range of threshold values. In tandem with this, the Area Under the Curve (AUC) provides a value that summarises the model's performance across different thresholds. A high AUC value (close to 1) typically indicates better performance in minimising both False Positive Rate (FPR) and False Negative Rate (FNR).

Furthermore, to assess the significance of differences in model performance within each MCI subtype, a one-way ANOVA with a Tukey's HSD post hoc test was conducted. Additionally, to compare the model performance between MCI subtypes, an independent t-test was performed.

4.7 Python code

The full code used for this study is available on Github at https://github.com/techonn/HMM.

Chapter 5: Results

To reiterate, the aim of this study is to assess the effectiveness of HMM in detecting deviations in activity patterns among elderly individuals affected by the two subtypes of MCI: amnestic MCI and non-amnestic MCI. To achieve this, a robust methodology was developed to simulate abnormal behaviour patterns that closely mimic the gradual cognitive decline observed in individuals with aMCI and naMCI. Subsequently, HMM was applied to this simulated data to evaluate its capacity in detecting these anomalies, and these capabilities were further assessed by comparing them with those of existing state-of-the-art methods. The following sections of this chapter will present the results of the anomaly simulation and delve into the performance of HMM in anomaly detection.

5.1 Simulation of MCI-related activity pattern anomalies

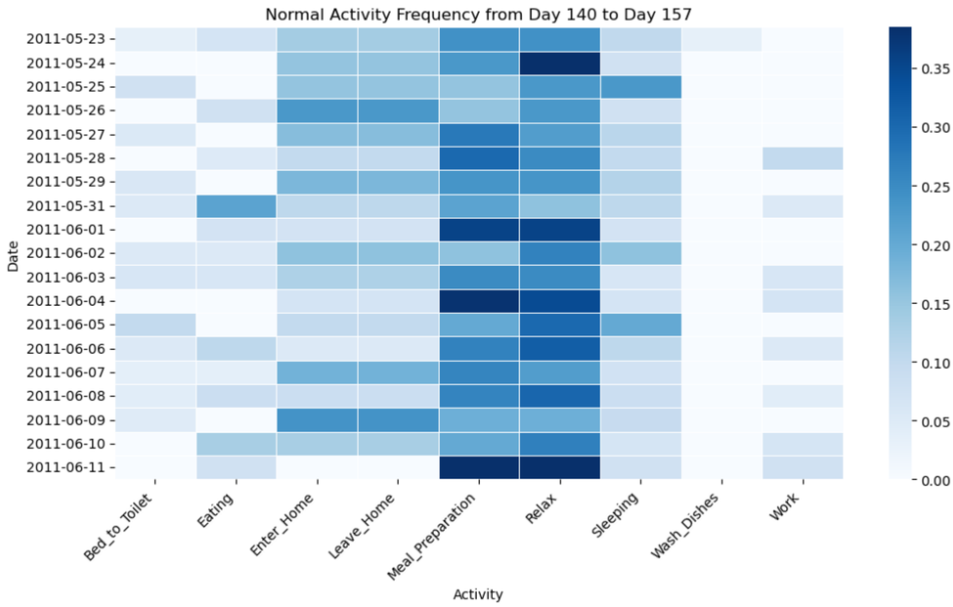
The goal of this methodology was to introduce a diverse range of anomalous behaviours that resemble real-life scenarios of gradual cognitive impairment. To ensure the simulated anomalies are realistic and accounted for the progressive nature of cognitive decline, the proposed methodology incorporated insights from existing literature on MCI-related behavioural changes. To visually observe the changes introduced by the anomalies, the activity frequencies of the original data and simulated anomalies are presented in Figure 5. It should be noted that these that these visualizations showcase an excerpt from days 140 to 157 of the test set, representing only the tenth iteration of the anomaly simulation process. Moreover, it's worth emphasising that the simulated behavioural symptoms during this specific period are among the most severe compared to all other days.

To simulate aMCI-related behavioural patterns, anomalies introduced included disruptions in sleep patterns, introduction of repetitive sequences, instances of napping, and an elevated frequency of relaxation activities. These changes were made to reflect the memory impairment, sleep disturbances, and the increased susceptibility to fatigue or mild depressive symptoms observed in individuals with aMCI. On the other hand, naMCI does not typically involve memory impairment symptoms. Therefore, repetitive sequences were not introduced in this context. Instead, an elevated frequency of sleep disruptions was simulated, while other anomalies remained consistent with those observed in aMCI.

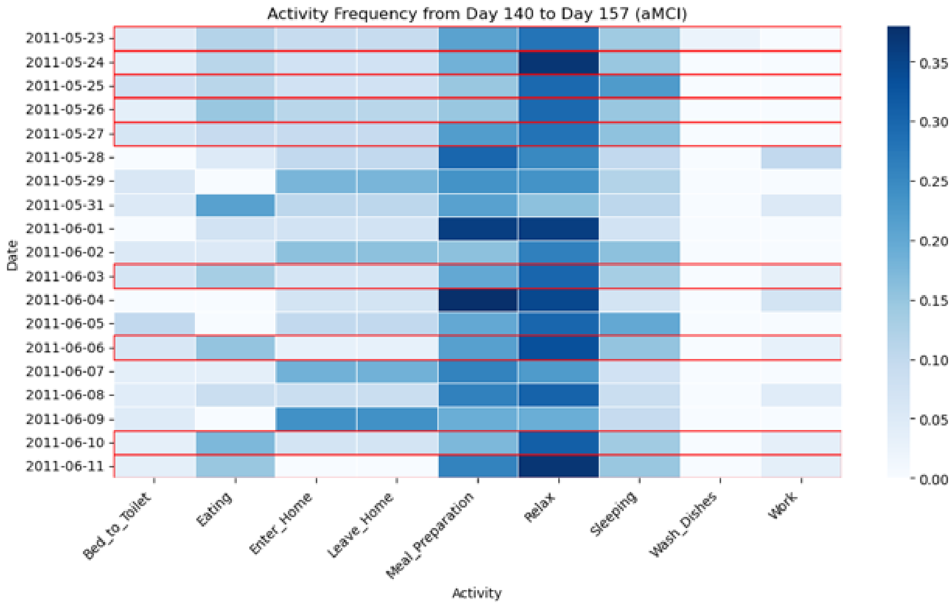
As can be observed in the heatmap representing aMCI anomalies, there is a noticeable increase in the frequency of sleeping activities compared to the original data. Additionally, a higher frequency of eating activities is apparent, mimicking instances where an individual repeatedly eats due to forgetting prior meals. In contrast, the naMCI anomalies heatmap depicts a greater frequency of sleeping activities compared to both the original data and the aMCI-simulated data.

The observed differences between the aMCI and naMCI anomalies and their respective normal activity patterns underscore the significance of this anomaly simulation methodology in evaluating the effectiveness of anomaly detection methods in identifying deviations in MCI-related behaviour.

(a)



(b)



(c)

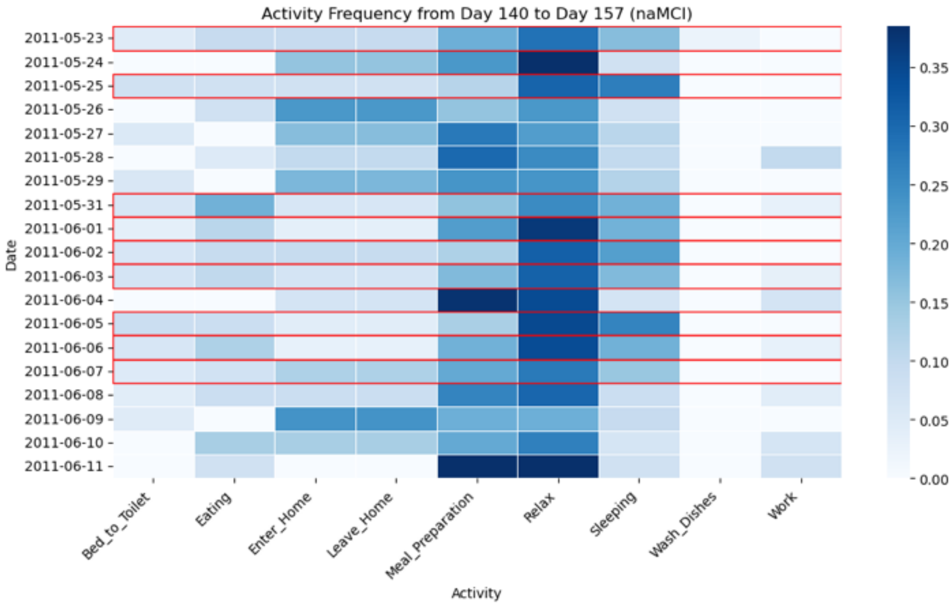


Figure 5: Activity Frequency and Anomalies Comparison

*Heatmaps illustrating the normalized frequency of various activities performed on the last 18 days of the test dataset (day 140 to 157) by **(a)** a normal elderly individual, **(b)** an elderly individual affected by aMCI, and **(c)** an elderly individual affected by na-MCI. Each row corresponds to a specific date, while each column represents a different activity. The shades of blue reflect the frequency of each activity relative to each day, with darker shades indicating higher frequency. For the anomalous heatmaps, the rows with red outlined borders represent the dates with introduced anomalies.*

*Note: The anomalies shown in heatmaps **(b)** and **(c)** are from the tenth iteration of anomaly simulation. During this period, the simulated behavioural symptoms are the most pronounced compared to all other preceding days.*

5.2 Assessment of HMM-Based Anomaly Detection Framework

5.2.1 Optimal hidden state selection

As described in the Methods chapter, an important prerequisite before training an HMM is to select the optimal number of hidden states. This was achieved using a combination of log-likelihood, AIC and BIC. The results after running Algorithm 6 is shown in Figure 6. A clear minimum can be observed for 5 hidden states and the log-likelihood also stabilises at this point. This implies that 5 hidden states sufficiently capture the diverse forms of 'normal' behaviour represented by the activity sequences. Thus, this hyperparameter value was used for subsequent HMM training.

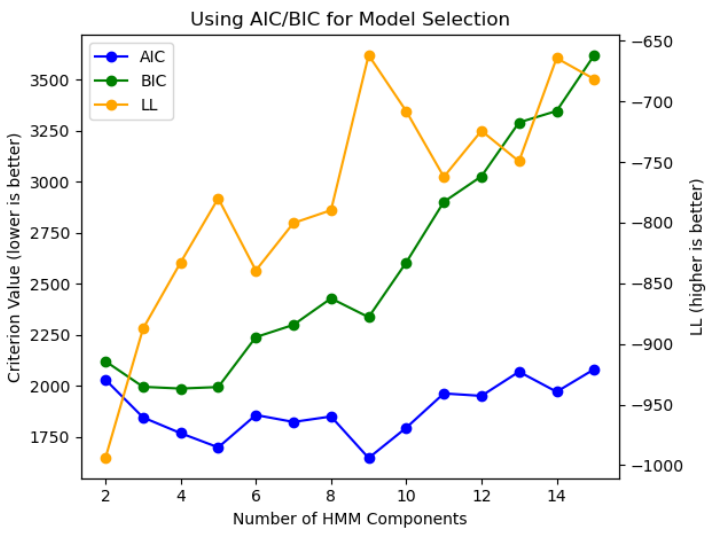


Figure 6: Model Selection using AIC, BIC, and Log-likelihood

This figure illustrates a graph for selecting the optimal number of hidden states to train the HMM using AIC, BIC, and log-likelihood values. It displays the values of AIC, BIC, and log-likelihood for varying numbers of HMM hidden states, ranging from 2 to 15. The AIC and BIC values are shown on the left y-axis while the log-likelihood score is shown on a the right y-axis. The lower the AIC and BIC values and the higher the log-likelihood values, the better the model fits the data.

5.2.2 Anomaly detection with HMM, LSTM, and GRU-AE

Due to the randomisation of anomalous days in the anomaly simulation algorithm, each model was evaluated 10 times to assess its performance across varying scenarios and ensure consistency of the results. Mean and standard deviations values were calculated to summarise the overall performance of each method on both the simulated aMCI and naMCI datasets, and the results are displayed in Tables 2a and 2b, respectively. For a comprehensive evaluation of model performance, a focus on AUC and F1-score, as presented in Figure 7, was deemed more informative for assessing models in the context of imbalanced datasets. The term "imbalanced dataset" in this context implies a significant difference in the number of days with normal activity patterns compared to those with anomalies (normal: 118 days vs. anomaly: 39 days).

The HMM consistently outperformed the deep learning models of LSTM and GRU-AE across all evaluation metrics. This is confirmed using one-way ANOVA with Tukey's HSD post hoc test which revealed that HMM's better performance over the benchmark models (LSTM and GRU-AE) were statistically significant for the aMCI subtype across all evaluation metrics. Similarly, for the naMCI subtype, HMM also exhibited significantly better performance compared to both LSTM and GRU-AE, except for recall where no significant difference was observed between HMM and GRU-AE. Notably, LSTM and GRU-AE also had slightly greater variation in their performance metrics, suggesting less reliability in anomaly detection. Coupled with HMM's more consistent performance, these findings suggest that HMM's underlying concept of modelling normal behaviour through its hidden states is effective in detecting slight deviations in MCI-related behaviour.

(a) Amnestic MCI

Model	AUC	F1-score	Precision	Recall	Accuracy
HMM	89.83 ± 2.33	78.05 ± 2.54	83.54 ± 4.67	73.59 ± 4.84	89.75 ± 1.10
LSTM	77.54 ± 3.19*	61.39 ± 3.03*	57.52 ± 5.39*	66.67 ± 6.94*	79.17 ± 2.36*
GRU-AE	81.30 ± 3.75*	69.35 ± 4.39*	72.95 ± 6.13*	66.41 ± 5.33*	85.41 ± 2.28*

(b) Non-Amnestic MCI

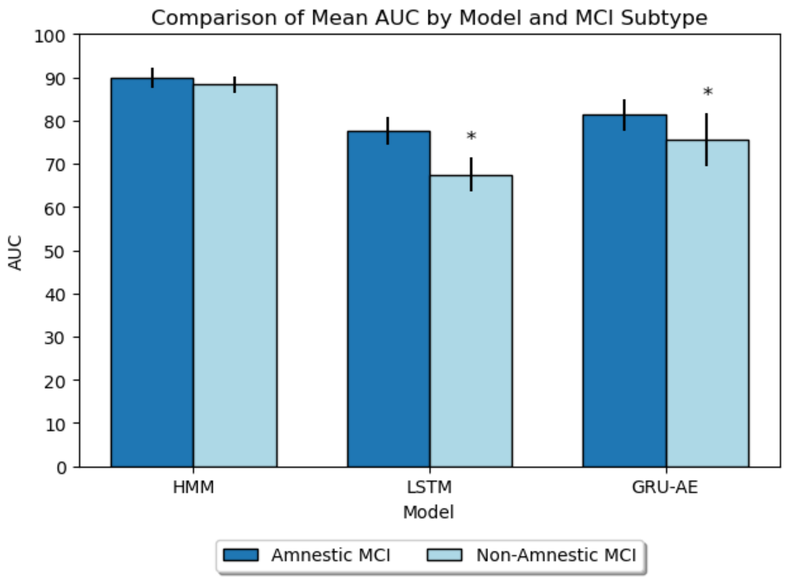
Model	AUC	F1-score	Precision	Recall	Accuracy
HMM	88.32 ± 1.84	74.92 ± 3.64	84.96 ± 4.85	67.43 ± 6.40	88.85 ± 1.35
LSTM	67.48 ± 3.96*	51.82 ± 3.44*	50.07 ± 6.47*	54.87 ± 6.96*	74.65 ± 3.25*
GRU-AE	75.64 ± 6.19*	61.61 ± 6.92*	64.18 ± 12.24*	62.05 ± 12.2	80.77 ± 5.13*

Table 2: Comparison of Evaluation Metrics of Different Machine Learning Models for Anomaly Detection in Different MCI Subtypes

*The tables provide a comprehensive comparison of evaluation metrics for anomaly detection in (a) aMCI and (b) naMCI. The evaluation involves HMM, LSTM and GRU-AE. Mean values along with standard deviations are presented for AUC, F1-socre, precision, recall, and accuracy. The table offers insights into the models' respective abilities to detect behavioural deviations associated with varying levels of MCI subtypes. Statistical analysis was conducted using one-way ANOVA with Tukey's HSD post hoc test; * indicates a significant difference (p<0.05) between the benchmark models and HMM.*

Interestingly, when comparing HMM's performance between the MCI subtypes, HMM showcased a slightly more pronounced effectiveness in detecting anomalies within the aMCI subtype. However, based on results from the independent t-test, statistical significance was only achieved for the F1-score and not for AUC. This implies HMM's potential in identifying certain types of behavioural deviations that are associated with memory-related anomalies, which are represented by repetitive activity sequences in this study. Conversely, naMCI does not exhibit repetitive sequence anomalies. This could explain why HMM's performance is slightly reduced in detecting anomalies related to naMCI, despite the presence of more pronounced sleep disruptions in this subtype.

(a)



(b)

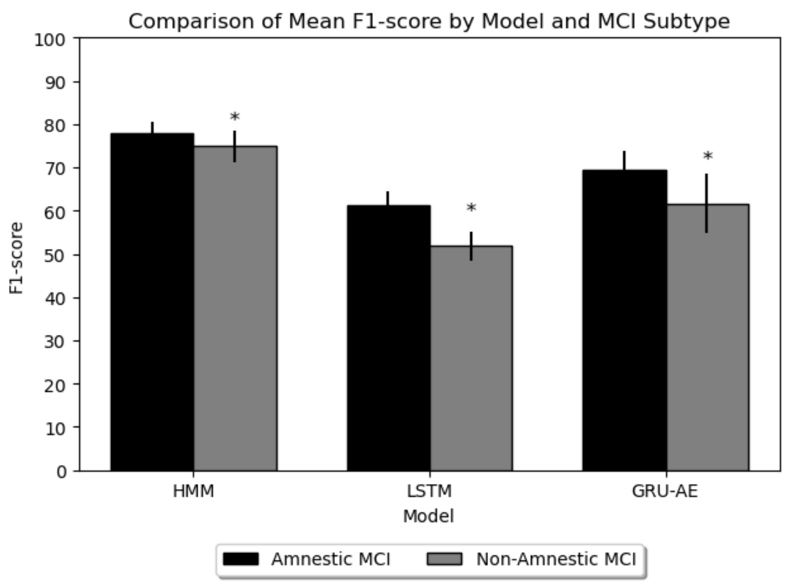


Figure 7: Bar Chart Comparing AUC and F1-score of Different Machine learning Models for Anomaly Detection in Different MCI Subtypes

*The bar charts illustrate the comparison of (a) mean AUC and (b) F1-score values for different anomaly detection models HMM, LSTM, and GRU-AE between aMCI and naMCI. The error bars represent standard deviation. Statistical analysis was conducted using independent t-test; * indicates a significant difference (p<0.05) between the aMCI and naMCI subtypes.*

5.2.3 Visualising anomaly detection results

To provide more insight into the performance of the anomaly detection models, visualisations were utilised to showcase their effectiveness in identifying MCI-related behavioural deviations. As an example, ROC curves, confusion matrices, and an anomaly detection graph are presented for each model in Figures 8, 9, and 10, respectively, based on the tenth experimental iteration for aMCI.

Firstly, the ROC curves in Figure 8 are consistent with earlier results, affirming that HMM outperforms the other benchmark models. It displayed the highest true positive rate while minimising false positive rates. To further illustrate HMM's higher efficacy in identifying MCI-related behaviours, the confusion matrices in Figure 9 highlight that HMM has the highest true positive and lowest false negative rates. While the GRU-AE appears to have a lower false positive rate for this iteration, this is not consistent across multiple iterations when considering the mean precision value, where HMM significantly has a higher value.

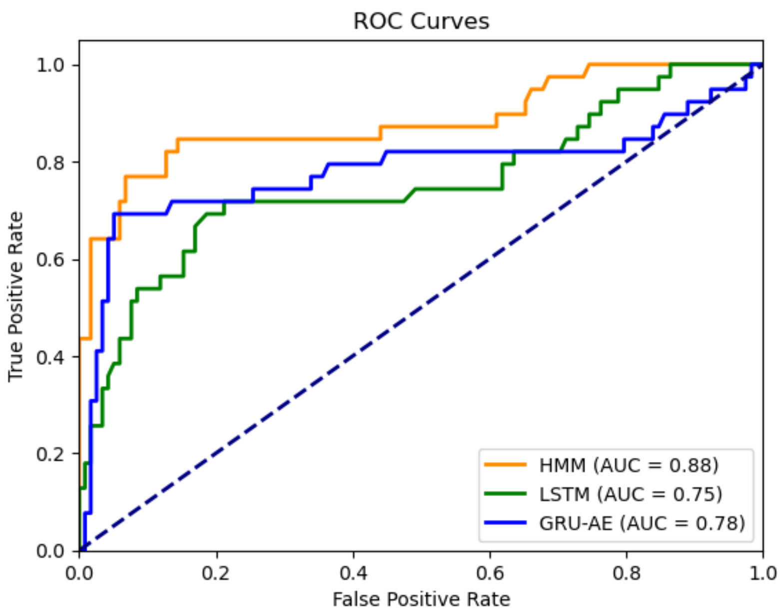


Figure 8: Comparison of ROC Curves of Different Machine Learning Models in Detecting aMCI-Related Anomalies

The ROC plot shown is from the tenth iteration of anomaly simulation. It illustrates the comparison of AUC values for different anomaly detection models HMM, LSTM, and GRU-AE across aMCI. AUC values closer to 1 is indicative of a better classification model. The diagonal *line spanning from coordinate (0,0) to (1,1) represents a model that is no better than random guessing.*

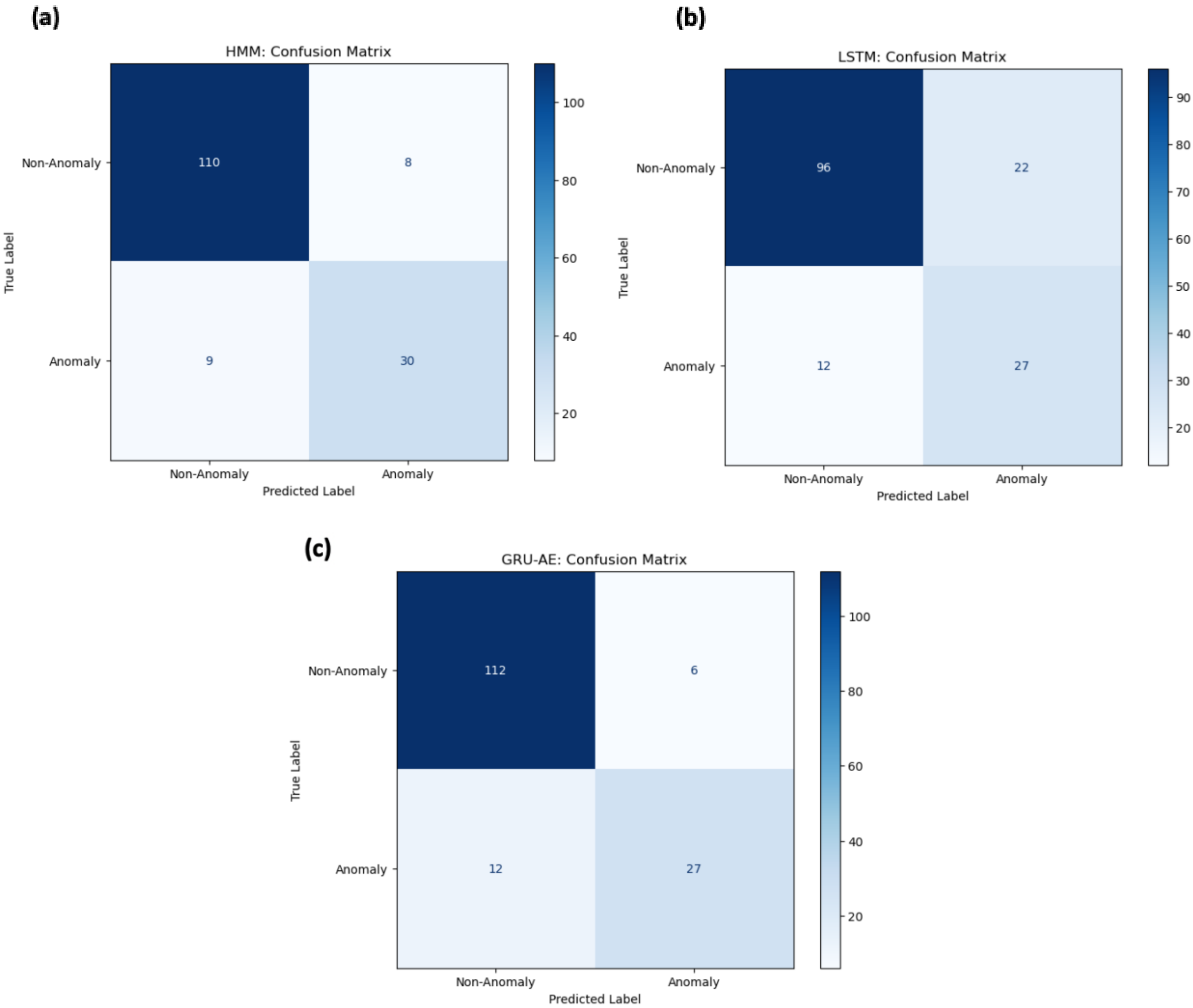


Figure 9: Comparison of Confusion Matrices for Different Machine Learning Models in Detecting aMCI-Related Anomalies

The confusion matrices presented here are derived from the tenth iteration of the anomaly simulation for aMCI detection. These matrices provide a detailed breakdown of true positive, false positive, true negative, and false negative classifications, offering insights into the performance of (a) HMM, (b) LSTM, and (c) GRU-AE in identifying aMCI-related behavioural anomalies. The blue shaded bar represents density data points that fall into each category with a higher density represented by dark blue.

Although not directly indicative of the model performances, the anomaly detection graph shown in Figure 10 serves a practical purpose. It visually displays the presence of anomalous days within the dataset given a chosen threshold. This visualisation translates complex data and model outputs into a user-friendly format, making it a valuable tool for caretakers looking after the elderly individual. It offers an easily interpretable representation of the abnormal behavioural patterns exhibited by elderly individuals affected by MCI. For instance, a notable shift in behaviour patterns can be observed starting from day 80 in the example anomaly detection graph. This timely information can alert caretakers to potential issues, facilitating timely support for the elderly individual. However, in a real-life setting, detected anomalies in smart home data can originate from diverse sources such as technical issues, guests visiting, or actual MCI indicators. Thus, it is essential to recognise that identified anomalies do not equate disease manifestation and require additional assessment from healthcare professionals.

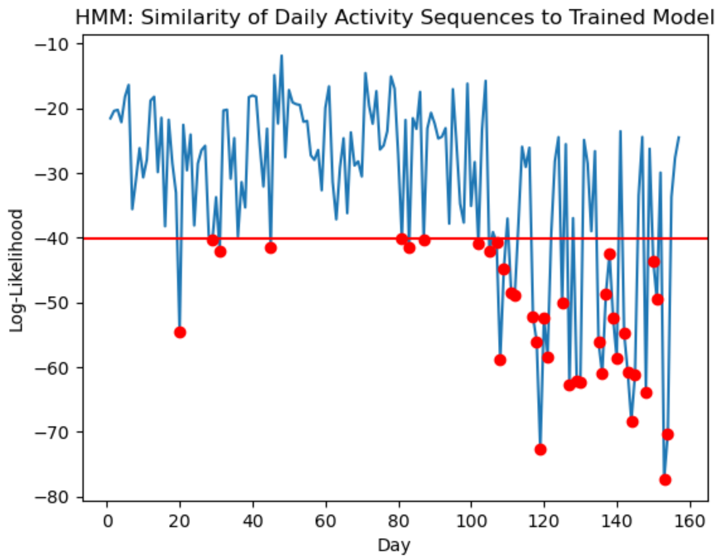


Figure 10: Anomaly Detection Graph for HMM

This graph visually presents the detection of anomalous days within the dataset based on log-likelihood values and a predefined threshold. Each point on the graph represents a day, with the y-axis indicating the log-likelihood of the daily activity sequence when tested on the trained HMM model. The red line signifies the selected threshold, while red dots highlight the detected anomalies below this threshold.

Chapter 6: Discussion

6.1 Summary of the results

To reiterate, the research objectives of this study were threefold: first, to simulate MCI-related behavioural anomalies as realistically as possible; second, to evaluate HMM's ability to detect these anomalies; and lastly, to compare HMM's performance with current state-of-the-art methods. The purpose of these experiments was to establish a foundation for the exploration of simulating MCI-related behavioural anomalies and highlight the potential of machine learning models, particularly HMM, in identifying these anomalies. This research holds direct clinical implications in the fields of social care and neurology, with the potential to facilitate the development of personalised care plans and early interventions in the future.

Firstly, the viability of the simulated aMCI and naMCI datasets was considered. The heatmaps in Figure 5 suggest that the anomaly simulation methodology successfully introduced diverse behavioural anomalies associated with the typical symptoms of MCI. Importantly, these anomalies were tailored to align with the distinctive characteristics of each disease subtype. For aMCI, the results indicated a noticeable increase in the frequency of sleeping, eating, and meal preparation activities. This suggests the presence of symptoms related to sleep disruption and memory impairment, which are well-recognised features of the aMCI subtype. In comparison, the simulated naMCI dataset exhibited more sleep disruptions while remaining consistent with other aspects of aMCI-related anomalies. However, repetitive sequences were deliberately omitted in this context as memory is not affected in naMCI individuals. These differences in activity patterns between the original dataset, aMCI and naMCI dataset highlight the methodology's viability in evaluating anomaly detection methods for diverse MCI-related behaviours.

In terms of anomaly detection, HMM's performance in detecting the simulated anomalies was compared with LSTM and GRU-AE. The results consistently favoured HMM in both the aMCI and naMCI datasets, demonstrating superior efficacy in detecting MCI-related anomalies across all the evaluation metrics when compared to LSTM and GRU-AE. To add on, the performance of LSTM and GRU-AE had greater variation, suggesting HMM's higher reliability in anomaly detection within this context. When comparing between the two MCI subtypes, HMM displayed slightly higher efficacy in detecting anomalies within aMCI, although this was only significant when comparing F1-score and not AUC. These results suggest HMM's potential in identifying specific behavioural deviations associated with memory-related anomalies. Nevertheless, it should be reiterated that the aim of this study was not to differentiate between MCI subtypes but rather to assess whether HMM can effectively detect abnormal behaviour in both MCI subtypes.

6.2 Comparison with existing studies

To contextualise the findings and assess the significance of this study, a mixture of existing studies that evaluated anomaly detection methods on both MCI and dementia-related abnormal behaviours were compared. However, it should be prefaced that every study employs different algorithms, sets up that own experiment to obtain real data, or simulate anomalies is a differently. Naturally, this makes it very challenging to directly compare results. Nevertheless, by examining these studies, valuable insights can be gained into the evolving landscape of anomaly detection in MCI or, more generally, neurodegenerative disease research.

To start off the comparison, the study that inspired this dissertation's simulation methodology will be examined. Arifoglu, Charif, and Bouchachia (2020) used graph convolutional networks (GCN) to detect potential indicators of dementia. However, instead of using activity sequences as their model input, the researchers utilised granular-level sensor activities because they hypothesise that the order in which sub-activities are executed provides clues to dementia. This study also performed simulation of abnormal behaviour but did so in two levels: activity-level and sub-activity level. In this dissertation, the anomalies stimulated closely resemble the activity-level anomalies simulated in Arifoglu and researchers' (2020) study, with the addition of fatigue anomalies. While they detailed their methods for simulating anomalies, there was no indication of the number of anomalies generated, the temporal changes in anomaly severity, or other related aspects. In contrast, this dissertation delved into these aspects. In terms of anomaly detection focused on activity-level anomalies, the study results show that GCN, LSTM, and HMM achieved an AUC of 66%, 59%, and 42%, respectively. It was noted that the poor performance of LSTM and HMM was attributed to the 'Bag of Sensor' representation, which disregards the temporal order of sensor data. In comparison, this dissertation highlights HMM's superior performance across various evaluation metrics, revealing its potential in identifying diverse MCI-related behaviours, even within imbalanced datasets.

In another study by Riboni et al. (2015), the aims and objectives aligned with the focus of this dissertation which is the early detection of MCI in elderly individuals. Nevertheless, there were notable distinctions between their approach and the one undertaken in this dissertation. The researchers gathered data from an instrumented smart home environment and enlisted volunteers to replicate the daily routines of both healthy seniors and elderly individuals displaying early symptoms of MCI. Additionally, their dataset included activities of daily living that were more related to MCI, such as medicine intake, compared to the dataset used in this dissertation. For anomaly detection, the methodology employed in Riboni et al.'s (2015) study relied on supervised learning and rule-based reasoning to model MCI-related abnormal behaviour. This approach entailed collaboration with cognitive neuroscience experts, leveraging their domain knowledge to enhance the anomaly detection process. The results from their study were reported in terms of precision, recall, and F1 score, yielding high values of 89.8%, 96.4%, and 93.0%, respectively. These metrics significantly outperformed the HMM results presented in this dissertation which is expected when employing supervised learning techniques in collaboration with domain experts. However, the researchers acknowledged that rigid user-defined anomalies are not practical in real-world scenarios as individual behaviour of MCI individuals can vary considerably.

Lastly, another study that closely aligned with the aims and objectives of this dissertation is by Enshaeifar et al. (2018). In their research, they conducted a real-world study to perform anomaly detection in participants clinically diagnosed with dementia residing in residential homes. Anomaly detection of daily activity sequences was executed using a Markov chain model and entropy rate. Interestingly, their method achieved a sensitivity or recall of 74% which was similar to the HMM results presented in this dissertation. However, this study did not present any other evaluation metric and did not compare its results to other benchmark models in the field. This raises the question on whether a more robust method like HMM would perform better than a basic Markov chain. The researchers also acknowledged the challenge of validating detected anomalies. While the anomalies were validated by clinical notifications and technical monitoring, it's not entirely clear how accurately these labels reflect true behavioural anomalies related to dementia.

6.3 Limitations and Future Research

6.3.1 Challenges of simulating MCI-related behaviour

Although this dissertation achieved its aims and objectives, the system created is far from real-world clinical use and several limitations should be acknowledged for future research. Firstly, the simulation of MCI-related anomalies in activity patterns is a major challenge on its own due to the variability of behaviour among individuals with MCI. Unlike many medical conditions that follow predictable trajectories, MCI can manifest very differently from one individual to another (Knopman and Petersen, 2014, Rawtaer et al., 2020). Some individuals may experience rapid cognitive decline and present with many types of behavioural anomalies, while others with MCI may maintain stable cognitive conditions with subtle behavioural anomalies over a long period of time (Jekel et al., 2015). This variability raises questions regarding the validity of the simulated MCI datasets and their representation of real-life MCI behaviour.

Furthermore, the scope of behavioural symptoms captured by the simulated dataset in this dissertation is somewhat limited by the publicly accessible dataset used. Smart home environments have the potential to capture a wider range of relevant activities, such as medicine intake, which should be considered in future research (Riboni et al., 2015).

To address these challenges, a real-world study would ideally be conducted. However, such studies are complex and incur substantial costs, ethical considerations, and privacy concerns. Conducting research directly with individuals affected by MCI in their daily lives requires careful consideration of ethical guidelines and privacy regulations. Therefore, initial exploration with a simulated dataset serves as a valuable initial step.

6.3.2 Considerations to improve anomaly detection of MCI-related behaviour

While HMM demonstrated higher efficacy in anomaly detection in this dissertation compared to benchmark models like LSTM and GRU-AE, it's important to note that the benchmark models were not fully optimised due to time limitations. Future research could explore more advanced configurations and hyperparameter tuning for LSTM and GRU-AE to assess whether their performance can be improved. This could potentially close the performance gap with HMM.

Another major limitation of the anomaly detection methodology of this dissertation is that it only focused on a single context for detecting MCI-related behavioural anomalies within a smart home environment. This only provides a partial perspective on individual behaviour. Incorporating various contexts, as demonstrated by Forkan et al. (2014), such as activity patterns, activity routines, and physiological data, has the potential to enhance the accuracy of anomaly detection. Besides that, Luhr et al. (2004) showcased the advantages of using an extension of HMM called explicit state duration HMM to incorporate activity durations alongside the activity sequences for anomaly detection. While this study was not conducted in the context of MCI or dementia, abnormal activity durations is a feature directly relevant to MCI and dementia. For example, elderly individuals with MCI may exhibit altered sleep patterns or longer activity completion times due to memory or attention deficits. Another extension of HMM worth exploring is hierarchical HMM (H-HMM). This model involves two layers: the top layer comprises high-level activities, and the bottom layer consists of sub-activities that form high-level activities. Kang et al. (2010) demonstrated that H-HMM outperformed standard HMM in detecting behavioural anomalies. Therefore, one potential direction for future research is to explore these HMM extensions within simulated MCI datasets.

Lastly, considering how habits and routines change sometimes, investigation into online learning techniques could prove to be valuable. While existing systems predominantly concentrate on activity recognition and anomaly detection, they often fall short in adapting to change in behaviour as individuals with MCI transition through different stages of cognitive impairment. To this end, habit recognition emerges as an important future work to consider for valuable insights into disease progression. Recognising habits can help identify the shift from one stage of the disease to another (e.g., from MCI to dementia). However, this is a difficult task given the lack of real-world datasets and challenges associated with artificial data simulation mentioned above. Nonetheless, determining when to update the definition of 'normal' behaviour is crucial in allowing the anomaly detection system to remain sensitive to subtle changes in behaviour while reducing the risk of false alarms or missed alarms.

To conclude, this study has laid the foundation for simulating MCI-related behavioural anomalies and showcased the potential of machine learning models, particularly HMM, in detecting these anomalies. However, there remains substantial work ahead to refine this anomaly detection system for real-world clinical applications, offering promising prospects in the fields of social care and neurology.

Chapter 7: References

Alberdi, A., Weakley, A., Schmitter-Edgecombe, M., Cook, D.J., Aztiria, A., Basarab, A., Barrenechea, M., 2018. Smart home-based prediction of multi-domain symptoms related to Alzheimer's Disease. IEEE J Biomed Health Inform 22, 1720–1731. https://doi.org/10.1109/JBHI.2018.2798062

Allen, N.E., Schwarzel, A.K. and Canning, C.G. (2013). 'Recurrent Falls in Parkinson's Disease: A Systematic Review', *Parkinson's Disease*, 2013, p. e906274. Available at: https://doi.org/10.1155/2013/906274.

Aran, O. *et al.* (2016). 'Anomaly Detection in Elderly Daily Behavior in Ambient Sensing Environments'. *Human Behavior Understanding*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 51–67. Available at: https://doi.org/10.1007/978-3-319-46843-3_4.

Arifoglu, D. and Bouchachia, A. (2017). 'Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks', *Procedia Computer Science*, 110, pp. 86–93. Available at: https://doi.org/10.1016/j.procs.2017.06.121.

Arifoglu, D. and Bouchachia, A. (2019). 'Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks', *Artificial Intelligence in Medicine*, 94, pp. 88–95. Available at: https://doi.org/10.1016/j.artmed.2019.01.005.

Arifoglu, D., Charif, H.N. and Bouchachia, A. (2020). 'Detecting indicators of cognitive impairment via Graph Convolutional Networks', *Engineering Applications of Artificial Intelligence*, 89, p. 103401. Available at: https://doi.org/10.1016/j.engappai.2019.103401.

Arifoglu, D., Wang, Y. and Bouchachia, A. (2021). 'Detection of Dementia-Related Abnormal Behaviour Using Recursive Auto-Encoders', *Sensors*, 21(1), p. 260. Available at: https://doi.org/10.3390/s21010260.

Bakar, U.A.B.U.A. *et al.* (2016). 'Activity and Anomaly Detection in Smart Home: A Survey', in Subhas Chandra Mukhopadhyay (ed.) *Next Generation Sensors and Systems*. Cham: Springer International Publishing (Smart Sensors, Measurement and Instrumentation), pp. 191–220. Available at: https://doi.org/10.1007/978-3-319-21671-3_9.

Bala Suresh, P.V. and Nalinadevi, K. (2022). 'Abnormal Behaviour Detection in Smart Home Environments', in J.S. Raj, K. Kamel, and P. Lafata (eds) *Innovative Data Communication Technologies and Application*. Singapore: Springer Nature (Lecture Notes on Data Engineering and Communications Technologies), pp. 289–300. Available at: https://doi.org/10.1007/978-981-16-7167-8_22.

Bargiotas, I. *et al.* (2023). 'Preventing falls: the use of machine learning for the prediction of future falls in individuals without history of fall', *Journal of Neurology*, 270(2), pp. 618–631. Available at: https://doi.org/10.1007/s00415-022-11251-3.

Burns, A. and Iliffe, S. (2009). 'Alzheimer's disease', *BMJ*, 338, p. b158. Available at: https://doi.org/10.1136/bmj.b158.

Chimamiwa, G. *et al.* (2022a). 'Are Smart Homes Adequate for Older Adults with Dementia?', *Sensors (Basel, Switzerland)*, 22(11), p. 4254. Available at: https://doi.org/10.3390/s22114254.

Cook, D.J., (2010). 'Learning Setting-Generalized Activity Models for Smart Spaces.' *IEEE Intelligent Systems*, 2010(99), 1. https://doi.org/10.1109/MIS.2010.112

Cowton, J. *et al.* (2018). 'A Combined Deep Learning GRU-Autoencoder for the Early Detection of Respiratory Disease in Pigs Using Multiple Environmental Sensors', *Sensors (Basel, Switzerland)*, 18(8), p. 2521. Available at: https://doi.org/10.3390/s18082521.

Csukly, G. *et al.* (2016). 'The Differentiation of Amnestic Type MCI from the Non-Amnestic Types by Structural MRI', *Frontiers in Aging Neuroscience*, 8. Available at: https://www.frontiersin.org/articles/10.3389/fnagi.2016.00052 (Accessed: 28 August 2023).

Dokuzlar, O. *et al.* (2020). 'Factors that Increase Risk of Falling in Older Men according to Four Different Clinical Methods', *Experimental Aging Research*, 46(1), pp. 83–92. Available at: https://doi.org/10.1080/0361073X.2019.1669284.

Enshaeifar, S. *et al.* (2018). 'Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques', *PLoS ONE*, 13(5), p. e0195605. Available at: https://doi.org/10.1371/journal.pone.0195605.

Fahad, L.G. and Tahir, S.F. (2021). 'Activity recognition and anomaly detection in smart homes', *Neurocomputing*, 423, pp. 362–372. Available at: https://doi.org/10.1016/j.neucom.2020.10.102.

Fahim, M. and Sillitti, A. (2019). 'Anomaly Detection, Analysis and Prediction Techniques in IoT Environment: A Systematic Literature Review', *IEEE Access*, 7, pp. 81664–81681. Available at: https://doi.org/10.1109/ACCESS.2019.2921912.

Forkan, A.R.M. *et al.* (2015). 'A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living', *Pattern Recognition*, 48(3), pp. 628–641. Available at: https://doi.org/10.1016/j.patcog.2014.07.007.

Gupta, P., McClatchey, R. and Caleb-Solly, P. (2020). 'Tracking changes in user activity from unlabelled smart home sensor data using unsupervised learning methods', *Neural Computing and Applications*, 32(16), pp. 12351–12362. Available at: https://doi.org/10.1007/s00521-020-04737-6.

Hayes, T.L. *et al.* (2014). 'Sleep Habits in Mild Cognitive Impairment', *Alzheimer disease and associated disorders*, 28(2), pp. 145–150. Available at: https://doi.org/10.1097/WAD.0000000000000010.

hmmlearn — hmmlearn 0.3.0.post3+g23c0f13 documentation (no date). Available at: https://hmmlearn.readthedocs.io/en/latest/index.html (Accessed: 1 September 2023).

Hodges, M.R. *et al.* (2010). 'Automatic Assessment of Cognitive Impairment through Electronic Observation of Object Usage', in P. Floréen, A. Krüger, and M. Spasojevic (eds) *Pervasive Computing*. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), pp. 192–209. Available at: https://doi.org/10.1007/978-3-642-12654-3_12.

Patterson, C. (2018). 'World Alzheimer Report 2018: The state of the art of dementia research: New frontiers'. Available at: https://www.alzint.org/resource/world-alzheimer-report-2018/ (Accessed: 13 August 2023).

Janjua, Z.H., Riboni, D. and Bettini, C. (2016). 'Towards automatic induction of abnormal behavioral patterns for recognizing mild cognitive impairment', in *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. New York, NY, USA: Association for Computing Machinery (SAC '16), pp. 143–148. Available at: https://doi.org/10.1145/2851613.2851687.

Jekel, K. *et al.* (2015). 'Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review', *Alzheimer's Research & Therapy*, 7(1), p. 17. Available at: https://doi.org/10.1186/s13195-015-0099-0.

Jurafsky, D. and Martin, J.H. (2023). *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. 3rd edition. Stanford University.

Kang, W., Shin, Dongkyoo and Shin, Dongil (2010). 'Detecting and predicting of abnormal behavior using hierarchical Markov model in smart home network', in *2010 IEEE 17Th International Conference on Industrial Engineering and Engineering Management*. *2010 IEEE 17Th International Conference on Industrial Engineering and Engineering Management*, pp. 410–414. Available at: https://doi.org/10.1109/ICIEEM.2010.5646583.

Keage, H.A.D. *et al.* (2012). 'What sleep characteristics predict cognitive decline in the elderly?', *Sleep Medicine*, 13(7), pp. 886–892. Available at: https://doi.org/10.1016/j.sleep.2012.02.003.

Knopman, D.S. and Petersen, R.C. (2014). 'Mild Cognitive Impairment and Mild Dementia: A Clinical Perspective', *Mayo Clinic proceedings*, 89(10), pp. 1452–1459. Available at: https://doi.org/10.1016/j.mayocp.2014.06.019.

Leng, Y. *et al.* (2019). 'Objective napping, cognitive decline and risk of cognitive impairment in older men', *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 15(8), pp. 1039–1047. Available at: https://doi.org/10.1016/j.jalz.2019.04.009.

Lundström, J., Järpe, E. and Verikas, A. (2016). 'Detecting and exploring deviating behaviour of smart home residents', *Expert Systems with Applications*, 55, pp. 429–440. Available at: https://doi.org/10.1016/j.eswa.2016.02.030.

Monekosso, D., Remagnino, P. and Kuno, Y. (eds) (2009). 'Anomalous Behavior Detection: Supporting Independent Living', in *Intelligent Environments: Methods, Algorithms and Applications*. London: Springer (Advanced Information and Knowledge Processing), pp. 33–48. Available at: https://doi.org/10.1007/978-1-84800-346-0_3.

Ordóñez, F.J., de Toledo, P. and Sanchis, A. (2015). 'Sensor-based Bayesian detection of anomalous living patterns in a home setting', *Personal and Ubiquitous Computing*, 19(2), pp. 259–270. Available at: https://doi.org/10.1007/s00779-014-0820-1.

Petersen, R.C. *et al.* (2001). 'Current Concepts in Mild Cognitive Impairment', *Archives of Neurology*, 58(12), pp. 1985–1992. Available at: https://doi.org/10.1001/archneur.58.12.1985.

Poh, S.-C. *et al.* (2019). 'LSTM and HMM Comparison for Home Activity Anomaly Detection', in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 1564–1568. Available at: https://doi.org/10.1109/ITNEC.2019.8729168.

Rabiner, L.R. (1989). 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proceedings of the IEEE*, 77(2), pp. 257–286. Available at: https://doi.org/10.1109/5.18626.

Ramapatruni, S. *et al.* (2019). 'Anomaly Detection Models for Smart Home Security', *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pp. 19–24. Available at: https://doi.org/10.1109/BigDataSecurity-HPSC-IDS.2019.00015.

Rawtaer, I. *et al.* (2020). 'Early Detection of Mild Cognitive Impairment With In-Home Sensors to Monitor Behavior Patterns in Community-Dwelling Senior Citizens in Singapore: Cross-Sectional Feasibility Study', *Journal of Medical Internet Research*, 22(5), p. e16854. Available at: https://doi.org/10.2196/16854.

Riboni, D. *et al.* (2015). 'Fine-grained recognition of abnormal behaviors for early detection of mild cognitive impairment', *2015 IEEE International Conference on Pervasive Computing* *and Communications (PerCom)*, pp. 149–154. Available at: https://doi.org/10.1109/PERCOM.2015.7146521.

Saives, J., Pianon, C. and Faraut, G. (2015). 'Activity Discovery and Detection of Behavioral Deviations of an Inhabitant From Binary Sensors', *IEEE Transactions on Automation Science and Engineering*, 12(4), pp. 1211–1224. Available at: https://doi.org/10.1109/TASE.2015.2471842.

Sánchez, V.G., Lysaker, O.M. and Skeie, N.-O. (2020). 'Human behaviour modelling for welfare technology using hidden Markov models', *Pattern Recognition Letters*, 137, pp. 71–79. Available at: https://doi.org/10.1016/j.patrec.2019.09.022.

Shahid, Z.K., Saguna, S. and Åhlund, C. (2023). 'Forecasting and Detecting Anomalies in ADLs in Single-Resident Elderly Smart Homes', *Proceedings of the International Conference on Research in Adaptive and Convergent Systems*. *RACS '23: International Conference on Research in Adaptive and Convergent Systems*, Gdansk Poland: ACM, pp. 1–8. Available at: https://doi.org/10.1145/3599957.3606216.

Tabira, T. *et al.* (2020). 'Age-Related Changes in Instrumental and Basic Activities of Daily Living Impairment in Older Adults with Very Mild Alzheimer's Disease', *Dementia and Geriatric Cognitive Disorders EXTRA*, 10(1), pp. 27–37. Available at: https://doi.org/10.1159/000506281.

Tan, Y.-F. *et al.* (2022). 'Home Activities Sequence Pattern Analysis', in R. Alfred and Y. Lim (eds) *Proceedings of the 8th International Conference on Computational Science and Technology*. Singapore: Springer (Lecture Notes in Electrical Engineering), pp. 459–467. Available at: https://doi.org/10.1007/978-981-16-8515-6_36.

Truong, C. *et al.* (2019). 'A Data Set for the Study of Human Locomotion with Inertial Measurements Units', *Image Processing On Line*, 9, pp. 381–390. Available at: https://doi.org/10.5201/ipol.2019.265.

United Nations (2019). 'World Population Prospects 2019'. Available at: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf (Accessed: 4 September 2023).

Wang, Y. *et al.* (2023). 'A Survey on Ambient Sensor-Based Abnormal Behaviour Detection for Elderly People in Healthcare', *Electronics*, 12(7). Available at: https://doi.org/10.3390/electronics12071539.

Wiles, J.L. *et al.* (2012). 'The Meaning of "Aging in Place" to Older People', *The Gerontologist*, 52(3), pp. 357–366. Available at: https://doi.org/10.1093/geront/gnr098.

Yamasaki, T. and Kumagai, S. (2021). 'Nonwearable Sensor-Based In-Home Assessment of Subtle Daily Behavioral Changes as a Candidate Biomarker for Mild Cognitive Impairment', *Journal of Personalized Medicine*, 12(1), p. 11. Available at: https://doi.org/10.3390/jpm12010011.

Yu, S.-Z. (2010). 'Hidden semi-Markov models', *Artificial Intelligence*, 174(2), pp. 215–243. Available at: https://doi.org/10.1016/j.artint.2009.11.011.

Zerkouk, M. and Chikhaoui, B. (2019). 'Long Short Term Memory Based Model for Abnormal Behavior Prediction in Elderly Persons', in J. Pagán et al. (eds) *How AI Impacts Urban Living and Public Health*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 36–45. Available at: https://doi.org/10.1007/978-3-030-32785-9_4.

Zerkouk, M. and Chikhaoui, B. (2020). 'Spatio-Temporal Abnormal Behavior Prediction in Elderly Persons Using Deep Learning Models', *Sensors (Basel, Switzerland)*, 20(8), p. 2359. Available at: https://doi.org/10.3390/s20082359.