Automatic Detection of Mild Cognitive Impairment in Older Adults Using Unobtrusive Sensing Technologies

by

Ahmad Akl

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Institute of Biomaterials and Biomedical Engineering
University of Toronto

© Copyright 2016 by Ahmad Akl
Abstract
Automatic Detection of Mild Cognitive Impairment in Older Adults Using Unobtrusive Sensing Technologies
Ahmad Akl
Doctor of Philosophy
Graduate Department of Institute of Biomaterials and Biomedical Engineering
University of Toronto
2016

The public health implications of growing numbers of older adults at risk for dementia place pressure on identifying dementia at its earliest stages so as to develop proactive management plans. However, this capability of early detection is very challenging with the contemporary detection process in the form of conventional clinician visits, which are typically not sensitive to detecting mild cognitive or functional decline. Fortunately, some studies have documented that changes in motor capabilities precede and may be indicative of cognitive impairment. A major motivation for this thesis was: can we utilize changes in motor capabilities and activity patterns to automatically predict the onset of dementia in older adults through continuous monitoring using unobtrusive sensors?

Many studies have demonstrated that the prodromal dementia phase commonly identified as mild cognitive impairment is an important target for this early detection of impending dementia amenable to treatment. Consequently, in this thesis we explored different approaches for the automatic detection of mild cognitive impairment using unobtrusive sensing technologies. We started off by exploring the feasibility of detecting mild cognitive impairment in older adults using a number of predefined measures associated with their in-home walking speed. We were able to achieve this goal with an
area under the ROC curve and an area under the precision-recall curve of 0.97 and 0.93, respectively, using a time frame of 24 weeks.

We were also very interested in exploring the feasibility of detecting mild cognitive impairment in older adults using changes in their activity patterns. We used inhomogeneous Poisson processes to build generalized linear models of older adults’ activity that would model their presence in different rooms throughout the day. Using these models, we extracted very interesting insights and visualized significant changes in older adults’ activity patterns as they started experiencing cognitive impairment. Finally, we devised a method for automatic detection of mild cognitive impairment using changes in older adults’ activity models, and we achieved this goal with an $F_{0.5}$ score of 85.6 percent. We also investigated the two subtypes of mild cognitive impairment, namely the amnestic and the non-amnestic subtypes, and made very interesting observations.
Dedication

To my Parents - Ziad and Aida
To my Wife - Rebecca
To my Sons - Abdullah and Ziad
Acknowledgements

At times our own light goes out and is rekindled by a spark from another person. Each of us has cause to think with deep gratitude of those who have lighted the flame within us.

Albert Schweitzer

I was recently asked by a friend who I would choose as a role model, and I immediately answered my Ph.D. supervisor Dr. Alex Mihailidis. Since I started my Ph.D. back in 2010, Dr. Mihailidis has been more of a friend than a supervisor. Dr. Mihailidis has been extremely understanding in times that other supervisors have shown no understanding at all. I remember how supportive and understanding he was when my dear mother-in-law passed away, when I had my first son, and when I had my second son. Furthermore, my Ph.D. thesis turned out to be a very challenging one, and in the inevitable times of tribulation, it would only take a visit to Dr. Mihailidis’s office to be motivated again and to get reassured that I could do it. I will always be indebted to Dr. Mihailidis for accepting to supervise my Ph.D., for believing in me, and for his guidance and unwavering support—both academically and financially. Dr. Mihailidis has been a great leader and mentor to me, and if I ever become a professor, I will always look to be like him. This is why Dr. Mihailidis is my role model.

I would like to thank the group at the Oregon Center for Aging and Technology for providing me with a unique collaboration opportunity by letting me use their data set in my work. I was overwhelmed with their hospitality when I visited Portland, Oregon in 2011, and they did not hesitate to provide me with all the help and support that I needed to understand their data set. I am grateful in particular to Dr. Jeffrey Kaye, Dr. Tamara Hayes (may her soul rest in peace), Dr. Daniel Austin, and Ms. Nora Mattek. It was a pleasure working with them.
I would also like to thank my committee: Dr. Tom Chau, Dr. Gary Naglie, Dr. Babak Taati, and Dr. Craig Boutilier for their guidance and comments during the annual committee meetings. Their feedback was very helpful in keeping me on the right track and their guidance played a key role in the fruition of this thesis. In addition, I would like to give special thanks to Dr. Jasper Snoek for guiding me through the process of building generalized linear models of older adults’ in-home activity. This work was very exciting to me and was the basis for two of my journal publications. Furthermore, special thanks to Dr. Elaine Biddiss, my departmental defense examiner, and Dr. Rafik Goubran, my Senate defense examiner, for all their constructive comments and feedback. I also thank the coordinators at the department’s graduate office, in particular Mr. Jeffrey Little and Ms. Rhonda Marley, for their help with the administrative work associated with my Ph.D. and other scholarship applications.

Most of all, I would like to express my deepest gratitude to my parents, brothers, and sister for their unconditional love and support throughout my life. I am completely indebted to my father, Ziad, for his tremendous generosity, care, and encouragement, and how he always worked his hardest to provide me with the best life. As for my mother, Aida, I could not have asked more of her. Her everlasting love, constant support, and friendship made me who I am and brought meaning to my achievements. I have always been proud of them as my parents, and today I hope they are proud of me. Finally, I would like to express my sincerest appreciation to my wife, Rebecca, whose unceasing love, relentless patience, and unyielding support supplied me with the energy and confidence to keep going. I am so proud to have her by my side as I end this chapter of my life and embark on a new endeavor.
Contents

Abstract ............................................................. ii

Dedication ........................................................... iv

Acknowledgements .................................................... v

List of Figures ......................................................... xi

List of Tables .......................................................... xiv

Mathematical Conventions ........................................ xvi

1 Introduction ......................................................... 1

1.1 Dementia ........................................................ 1

1.1.1 Definition and Symptoms .................................. 2

1.1.2 Causes ..................................................... 2

1.1.3 Risk Factors ............................................... 4

1.1.4 Complications ............................................. 6

1.2 Concerning Statistics ........................................... 7

1.2.1 Contemporary Detection Process ......................... 7

1.2.2 Shortcomings of Contemporary Detection Process ...... 8

1.2.3 Consequences of Shortcomings .......................... 9

1.3 Motivation ....................................................... 10
1.3.1 Interviewing Family Members ........................................ 10
1.3.2 Assessing Cognitive Ability ........................................ 11
1.3.3 Conclusion .......................................................... 11
1.4 Research Questions ...................................................... 13
1.5 Outline ................................................................. 13
1.6 Relationship to Published Work ....................................... 15

2 Background ................................................................. 17
  2.1 Motor Capabilities and Cognition .................................... 17
  2.2 Smart Systems and Related Work .................................. 20
    2.2.1 Commercial Systems ............................................ 20
    2.2.2 Prototype Systems .............................................. 21
    2.2.3 Related Studies .................................................. 23
    2.2.4 Limitations of Current Smart Systems and Studies .......... 25
  2.3 Mild Cognitive Impairment .......................................... 26
    2.3.1 Types of Mild Cognitive Impairment .......................... 28
    2.3.2 Single Domain and Multiple Domain MCI ..................... 28
  2.4 Affinity Propagation ................................................ 28
    2.4.1 Similarity Function ............................................. 29
    2.4.2 Message Passing: Responsibility and Availability .......... 30
    2.4.3 Cluster Decisions .............................................. 31

3 Oregon Center for Aging and Technology Data ...................... 32
  3.1 Data from the Oregon Center for Aging and Technology ........ 33
    3.1.1 Subjects ...................................................... 33
    3.1.2 Sensing System ................................................. 34
    3.1.3 Weekly Self-Report ........................................... 35
  3.2 Annual Assessments ................................................ 37
### 3.3 Summary of Data Received from ORCATECH

- 3.3.1 Sets of Data

### 3.4 Relevant Study by ORCATECH

### 4 Data Labeling & Cleaning

- 4.1 Introduction
- 4.2 Data Labeling
  - 4.2.1 CDR Scale
  - 4.2.2 Neuropsychological Tests
- 4.3 Preprocessing & Cleaning Process
- 4.4 Summary of Clean Data

### 5 Performance Evaluation Metric

- 5.1 Introduction
- 5.2 Case Study
- 5.3 Conclusion

### 6 Predefined Measures of Speed and Activity

- 6.1 Introduction
- 6.2 Problem Setup
- 6.3 Data, Measures, and Features
  - 6.3.1 Measures and Features
- 6.4 Results and Discussion
  - 6.4.1 Average of Activity Measures
  - 6.4.2 Probability Densities of Activity Measures
  - 6.4.3 Trajectories of Activity Measures
  - 6.4.4 Average of Walking Speed Measures
  - 6.4.5 Probability Densities of Walking Speed Measures
  - 6.4.6 Trajectories of Walking Speed Measures
6.4.7 Trajectories of Activity & Walking Speed Measures  79
6.4.8 Trajectories of Walking Speed Measures with More Data  80
6.4.9 Feature Ranking  84
6.5 Conclusion  89

7  Generalized Linear Models of Home Activity  91
7.1 Introduction  91
7.2 Data and Data Labeling  93
7.3 Problem Setup  93
7.4 Results and Discussion  96
  7.4.1 Generalized Linear Models of Activity  97
  7.4.2 Kullback–Leibler Divergence  102
  7.4.3 Detecting MCI  108
7.5 Conclusion  110

8  Clustering of Home Activity GLMs  112
8.1 Introduction  112
8.2 Data and Data Labeling  113
8.3 Problem Setup  113
  8.3.1 Room Activity GLM Estimator  117
  8.3.2 Relabeling Using $k$-means  118
  8.3.3 Kullback–Leibler Divergence  120
  8.3.4 Affinity Propagation  121
8.4 Results and Discussion  122
  8.4.1 80 percent Training Data and 20 percent Test Data  124
  8.4.2 Leave-One-Subject-Out Cross-Validation  128
  8.4.3 a-MCI & na-MCI Subtypes  128
8.5 Conclusion  131
9 Conclusions & Future Work

9.1 Translation of Work ........................................ 139
9.2 Limitations and Future Work .............................. 139
   9.2.1 General Limitations .................................. 140
   9.2.2 Chapter 6 .............................................. 140
   9.2.3 Chapter 7 .............................................. 141
   9.2.4 Chapter 8 .............................................. 141

Appendices .................................................... 143

A Examples of Labeling Data Using CDR .................... 144
   A.1 Example 2 ............................................... 144
   A.2 Example 3 ............................................... 145
   A.3 Example 4 ............................................... 145

B Examples of Labeling Data Using Neuropsychological Assessments 147
   B.1 Example 2 ............................................... 147
   B.2 Example 3 ............................................... 148
   B.3 Example 4 ............................................... 149

Bibliography .................................................. 151
List of Figures

2.1 Cognitive continuum showing the overlap between normal aging and mild
cognitive impairment, and mild cognitive impairment and early dementia
or Alzheimer’s disease [56]. ........................................ 27

3.1 A home map with sensor layout [59]. ................................ 35
3.2 Subject walking speed data. ........................................ 39
3.3 Subject activity sensor data. ....................................... 39
3.4 Subject clinical data - set 1. ....................................... 40
3.5 Subject clinical data - set 2. ....................................... 41

4.1 A high-level labeling process implemented in this thesis. ............ 44
4.2 An example of an excel file containing health questionnaire data with
disordered columns. ............................................... 53
4.3 The transformed excel file. Note how the columns are ordered and the
entries are sorted. .................................................. 53
4.4 Console showing a subset of sensors and their firings for Home 162. ... 55
4.5 Console showing a period of inactivity. ............................ 55
4.6 Console showing replacement of broken sensor. ..................... 56

6.1 Trajectories of weekly measures pertaining to subject 1 in the database.
Each asterisk represents a weekly measure. Features are extracted using
a window of size $\omega$, that slides one week at a time. ................ 64
7.4 Cumulative distribution function of normalized KL-divergences of “TR” GLMs from “CIN” GLMs, “MCI” GLMs from “CIN” GLMs, and “MCI” GLMs from “TR” GLMs for all rooms.

7.5 Estimating reference GLM using “CIN” data, and estimating GLMs of activity using a sliding time window of size $\omega$ weeks.

7.6 a) Room activity reference GLM. b) GLMs estimated using a sliding window of size $\omega$ weeks as it slid through the “CIN” data, the “TR” data, and the “MCI” data. The result is a trajectory of $n$ normalized KL-divergences.

7.7 Example of trajectories of normalized KL-divergences of sliding window GLMs from reference GLM for a) bedroom, b) bathroom, c) kitchen, and d) living room.

7.8 Estimating reference GLM using “CIN” activity sensor data, and estimating GLMs using a sliding time window of size $\omega$ weeks.

8.1 General overview of the cognitive status recognition process using activity GLMs corresponding to room $r$. (a) Training stage. (b) Test stage.

8.2 The process of building a data set of room activity GLMs for room $r$ for subject $n$. 

xiv
List of Tables

3.1 Weekly self-report questions [59] .................................................. 36
3.2 Six neuropsychological tests administered for cognitive assessment .... 37
3.3 Summary of subject demographics .............................................. 38
3.4 Statistics of monitoring periods of subjects .................................. 38

4.1 Example of labeling protocol rule 1 using CDR ............................. 45
4.2 Example of labeling protocol rule 2 using CDR ............................. 45
4.3 Example of labeling protocol rule 3 using CDR ............................. 46
4.4 Example of labeling protocol rule 4 using CDR ............................. 46
4.5 Example 1 of labeling subject data using CDR labeling protocol ........ 47
4.6 Example of labeling protocol rule 1 using neuropsychological tests .... 48
4.7 Example of labeling protocol rule 2 using neuropsychological tests .... 49
4.8 Example of labeling protocol rule 3 using neuropsychological tests .... 50
4.9 Example of labeling protocol rule 4 using neuropsychological tests .... 51
4.10 Example 1 of labeling subject data using the neuropsychological tests la-
       beling protocol ............................................................................. 51
4.11 Summary of subjects using CDR as ground truth .......................... 57
4.12 Summary of subjects using neuropsychological tests as ground truth ... 58

6.1 Baseline model using only age and gender as features for the data set of
       sixty-eight subjects ........................................................................ 72
6.2 Performance of SVM and RF with and without PCA for $\omega = 1$ week

6.3 Performance of SVM and RF using average of activity measures for $\omega = 1, 2, 3, \text{ and } 4$ weeks

6.4 Performance of RF and SVM using probability densities of activity measures for $\omega = 4, 8, 12, 16, 20, \text{ and } 24$ weeks

6.5 Performance of SVM and RF using trajectories of activity measures for $\omega = 4, 8, 12, 16, 20, \text{ and } 24$ weeks

6.6 Performance of SVM and RF using average of walking speed measures for $\omega = 1, 2, 3, \text{ and } 4$ weeks

6.7 Performance of SVM and RF using probability densities of walking speed measures for $\omega = 4, 8, 12, 16, 20, \text{ and } 24$ weeks

6.8 Performance of SVM and RF using trajectories of walking speed measures for $\omega = 4, 8, 12, 16, 20, \text{ and } 24$ weeks

6.9 Performance of SVM and RF using trajectories of combined measures of walking speed and home activity for $\omega = 4, 8, 12, 16, 20, \text{ and } 24$ weeks

6.10 Baseline model using only age and gender as features for the data set of ninety-seven subjects

6.11 Performance of RF and SVM using trajectories of walking speed measures calculated from the data set of ninety-seven homes

6.12 Ranking features by removing one feature at a time using SVM and $\omega = 24$ weeks

7.1 Areas under curves generated by thresholding trajectories of normalized KL-divergences of sliding window GLMs from reference GLMs. A sliding window of size $\omega = 12$ weeks was used. The performance of this simple thresholding was compared to the approach in Chapter 6 for $\omega = 12$ weeks and $\omega = 24$ weeks
8.1 $F_{0.5}$ scores with $\omega = 4$ weeks and slide sizes of 2 weeks, 3 weeks, and 4 weeks .......................................................... 125
8.2 $F_{0.5}$ scores with $\omega = 8$ weeks and slide sizes of 2 weeks, 3 weeks, and 4 weeks .......................................................... 126
8.3 $F_{0.5}$ scores with $\omega = 20$ weeks and slide sizes of 2 weeks, 3 weeks, and 4 weeks .......................................................... 126
8.4 $F_{0.5}$ scores with $\omega = 24$ weeks and slide sizes of 2 weeks, 3 weeks, and 4 weeks .......................................................... 127
8.5 $F_{0.5}$ scores corresponding to window of size $\omega = 20$ weeks and slide size = 4 weeks associated with leave-one-subject-out cross-validation .......... 128
8.6 $F_{0.5}$ scores corresponding to window of size $\omega = 20$ weeks and slide size = 4 weeks with na-MCI GLMs only .................................................. 129
8.7 $F_{0.5}$ scores for $\omega = 20$ weeks and slide size = 4 weeks with a-MCI GLMs only .................................................. 130

A.1 Example 2 of labeling subject data using CDR labeling protocol ........ 144
A.2 Example 3 of labeling subject data using CDR labeling protocol ........ 145
A.3 Example 4 of labeling subject data using CDR labeling protocol ........ 146

B.1 Example 2 of labeling subject data using the neuropsychological tests labeling protocol .......................................................... 147
B.2 Example 3 of labeling subject data using the neuropsychological tests labeling protocol .......................................................... 148
B.3 Example 4 of labeling subject data using the neuropsychological tests labeling protocol .......................................................... 149
Mathematical Conventions

In this thesis, lower case letters, e.g. $d$, will denote variables, bold lower case letters, e.g. $\mathbf{d}$, will denote vectors, bold upper case letters, e.g. $\mathbf{D}$, will denote matrices, and calligraphic upper case letters, e.g. $\mathcal{D}$, will denote sets.
Chapter 1

Introduction

I want to tell you how much I miss my mother. Bits of her are still there. I miss her most when I am sitting across from her.

Candy Crowley

In this chapter, we introduce the problem by giving an overview of dementia: its symptoms, causes, risk factors, and complications. Then, we report concerning statistics associated with its prevalence and high under-recognition rate. Following that, we discuss the motivation behind this thesis. Finally, we conclude this chapter by giving an overview of the rest of the chapters and their relationship with published work.

1.1 Dementia

Although memory loss generally occurs in dementia, individuals with memory loss alone do not necessarily have dementia. There is a certain extent of memory loss that is part of normal aging.
1.1.1 Definition and Symptoms

Dementia is a neurocognitive disorder that results in significant cognitive decline which impacts an individual’s ability, mostly older adults, to live independently and perform everyday activities [1]. Dementia describes a group of symptoms such as memory loss, decline in language, and impaired judgment and other cognitive skills caused by the permanent damage or death of brain nerve cells or neurons. Dementia significantly impacts an individual’s ability to carry out activities of daily living (ADLs) such as washing their hands or brushing their teeth, and other instrumental activities of daily living (IADLs) such as paying bills or making phone calls.

Symptoms and complications of dementia vary depending on the cause. However, many studies have identified common cognitive and personality changes. Among the many cognitive changes are memory loss, difficulty communicating or finding words, difficulty completing complex tasks, compromised planning and organizing ability, feeling disoriented, and exhibiting psychological changes. Similarly, personality changes include inability to reason, getting easily agitated, and hallucinating [2].

1.1.2 Causes

Dementia is primarily caused by the damage of neurons in parts of the brain responsible for cognitive functioning. Since different parts of the brain are responsible for different functions and behaviors, different types of dementia exist. In some cases, individuals do not have dementia, but instead have a condition whose symptoms mimic those of dementia. In general, two groups of dementia types exit: progressive types and dementia-like types [3].

Progressive Types of Dementia

Progressive types of dementia worsen with time and are irreversible. These include [3]:
1. **Alzheimer’s Disease:** Alzheimer’s disease is the most common cause of dementia and accounts for 60 – 80 percent of the cases. Early symptoms of Alzheimer’s disease include difficulty remembering recent conversations, names, or events as well as apathy and depression. Later symptoms include impaired communication, disorientation, poor judgment, behavior changes, and ultimately, difficulty speaking, swallowing, and walking. Alzheimer’s disease occurs most commonly due to the progressive accumulation of the protein fragment beta-amyloid (plaques) outside neurons in the brain and twisted strands of the protein tau (tangles) inside neurons. These changes eventually result in the damage and death of neurons.

2. **Dementia with Lewy Bodies (DLB):** Individuals with dementia with Lewy bodies (DLB) have some of the symptoms common in Alzheimer’s disease, but are more predisposed to exhibit early symptoms of sleep disturbances, visual hallucinations and slowness, and gait imbalance. These symptoms may occur in the absence of significant memory impairment. Lewy bodies are abnormal aggregations of the protein alpha-synuclein that accumulate in neurons. When they develop in a part of the cortex, dementia results.

3. **Vascular Dementia:** Vascular dementia as the sole cause of dementia accounts for about 10 percent of the cases. However, it is very common in older adults with other types of progressive dementia, appearing in about 50 percent of the cases. Symptoms of Vascular dementia include impaired judgment or ability to make decisions and plan. Vascular dementia occurs most commonly from blood vessel blockage or damage leading to strokes or bleeding in the brain. The location and size of the brain injuries determine how the individual’s thinking and physical functioning are affected.

4. **Parkinson’s Disease (PD) Dementia:** Symptoms of Parkinson’s disease (PD) commonly include problems with movement such as slowness, rigidity, tremor, and
changes in gait. Its incidence rate is about 6 – 8 percent. PD is commonly caused by alpha-synuclein aggregates appearing in an area deep in the brain called the substantia nigra. The aggregates are thought to cause degeneration of the nerve cells that produce dopamine.

Dementia-like Symptoms

Common causes of dementia-like symptoms are depression, delirium, side effects from medications, thyroid problems, certain vitamin deficiencies, and excessive use of alcohol. Unlike the progressive types of dementia, if these dementia-like symptoms are detected early, they can be reversed through treatment, otherwise permanent cognitive impairment results [3].

1.1.3 Risk Factors

There are many factors that increase the likelihood of an individual to have dementia. Some of the risk factors cannot be changed while others can be addressed and moderated to lower the risk of dementia.

Risk Factors that cannot be Changed

There are two main factors that cannot be changed [4]:

1. **Age:** It is well-established within the clinical community that age is the most significant known risk factor of dementia [5]. As an individual ages, the risk of Alzheimer’s disease, Vascular dementia, and several other dementias greatly increases, especially after the age of 65. However, dementia can still occur in younger people.

2. **Family History:** If an individual has a family history of dementia, then they are at a greater risk of developing the condition [6]. However, many individuals with a
family history never end up having dementia, and many other individuals without a family history end up having dementia.

Risk Factors that can be Addressed

On the other hand, some risk factors can be addressed in order to lower the likelihood of having dementia. These include:

1. **Heavy Alcohol Use:** Some of the detrimental effects of heavy alcohol consumption on brain function are similar to those observed with Alzheimer’s disease. Therefore, individuals who consume large amounts of alcohol are at a higher risk of having dementia [7].

2. **Atherosclerosis:** Consuming large amounts of fatty substances can lead to buildup of fats and other substances in and on artery walls. This in turn can result in reduced blood flow to the brain, leading to stroke [4]. As we discussed earlier, reduced blood flow to the brain can cause Vascular dementia.

3. **Cholesterol:** The two types of lipoprotein that carry cholesterol to and from cells are high-density lipoprotein (HDL) and low-density lipoprotein (LDL). A relatively high level of HDL protects against heart disease. On the other hand, high levels of LDL eventually reduce blood flow to the heart, which in turn reduce blood flow to the brain. As with atherosclerosis, high levels of LDL increase the risk of developing Vascular dementia or Alzheimer’s disease [8]. Similarly, obesity [9], smoking [10], and diabetes [11] also increase the risk of developing Vascular dementia. Management strategies include quitting smoking, reduced intake of saturated fats, increased physical activity, weight reduction, and certain types of drugs.
1.1.4 Complications

Progressive types of dementia significantly impact an individual’s ability to carry out daily activities, leading to several complications, including [12]:

1. **Inadequate Nutrition**: Individuals with dementia eventually reduce or stop eating and drinking. They start forgetting to eat or thinking that they have already eaten. Similarly, taking medications can be very challenging to individuals with dementia.

2. **Reduced Hygiene**: Individuals having moderate to severe dementia cannot live and function independently, and therefore, eventually lose their ability to bathe or dress themselves, brush their hair, or even use the toilet on their own.

3. **Deterioration of Emotional Health**: Dementia leads to depression, aggression, confusion, frustration, anxiety, and disorientation.

4. **Difficulty Communicating**: As dementia progresses, impaired individuals start forgetting names of people, places, and other information leading to communication challenges. Progressive deterioration in communication can lead to feelings of agitation, isolation, and depression.

5. **Hallucinations**: Individuals with certain progressive types of dementia may experience delusions in which they have false ideas about another person or situation. Individuals with dementia with Lewy bodies, in particular, may have visual hallucinations.

6. **Sleep Difficulties**: Many individuals with dementia start experiencing disturbed sleep patterns such as rapid eye movement (REM) sleep behavior disorder, where they act out their dreams. They physically move limbs or engage in sleep talking, shouting, screaming, hitting, or punching endangering themselves, their bed partners, or any other person they encounter during an REM episode.
All these complications make dementia, especially one caused by Alzheimer’s disease, ultimately fatal [13].

1.2 Concerning Statistics

People aged 65 and older constitute the fastest growing population segment in North America, Europe, and Asia. According to the US Census Bureau, the global number of adults over the age of 60 is expected to reach 1.2 billion by the year 2025 [14]. In Canada, the proportion of Canadians aged 65 and over is expected to represent 26.4 percent of the total population by the year 2031 [15]. Alzheimer’s disease has been reported in the 2015 Alzheimer’s Disease Facts and Figures annual report as the fifth-leading cause of death for people aged 65 and older in North America [16] [17]. Currently, reports show that one in nine Americans, and one in eleven Canadians, aged 65 and older have Alzheimer’s disease [18], and according to data from the Chicago Health and Aging Project (CHAP), an estimated 700 thousand people in the United States aged 65 and older will die with Alzheimer’s in 2015 [19]. As the “baby boomer” generation ages, these figures are projected to increase dramatically thus posing a serious challenge to the health-care infrastructure. Consequently, detection of the cognitive decline that precedes Alzheimer’s disease, or dementia in general, becomes vital so as to develop proactive management plans.

1.2.1 Contemporary Detection Process

However, this capability of early detection is very challenging with the contemporary detection processes in the form of conventional clinician visits which are typically not sensitive to detecting mild or subtle cognitive or functional decline. Current detection processes start by general practice physicians referring patients to memory clinics for cognitive assessment after repeated reports of memory problems by the patients themselves,
family members, or caregivers. In memory clinics, cognition of patients is assessed using questionnaires, screening tools, and episodic examinations of cognitive capacity such as the Montreal Cognitive Assessment (MoCA) [20], the Folstein Mini-Mental State Examination (MMSE) [21], and the Clinical Dementia Rating (CDR) [22].

MoCA was designed as a rapid screening instrument for mild cognitive impairment. It assesses different cognitive domains: attention and concentration, executive functions, memory, language, visuospatial skills, conceptual thinking, calculations, and orientation. The total possible score is 30 points, where a score lower than 26 indicates cognitive impairment. The Folstein MMSE or MMSE is an assessment tool that has been widely used in clinics and by some research centers to systematically and thoroughly assess mental status. It consists of eleven questions and tests five areas of cognitive function: orientation, registration, attention and calculation, recall, and language. The maximum score on an MMSE is 30, and a score of 23 or lower is indicative of cognitive impairment. Similarly, CDR is used in both research and clinical settings and characterizes six domains of cognitive and functional performance: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care by rating each domain on a 5-point scale. A CDR score of 0 implies no impairment, a score of 0.5 indicates mild cognitive impairment, and a score of 1, 2, and 3 indicates mild, moderate, and severe dementia, respectively.

1.2.2 Shortcomings of Contemporary Detection Process

Although these screening tools and examinations have been very accurate in identifying dementia, the detection process itself suffers from inherent shortcomings. First, some studies have found that older adults or families reported memory problems in only a small percentage of cases in which the older adult has been clinically labeled as cognitively impaired [23]. This could be because older adults might be unaware of their impairment, or if noted, might be uncomfortable discussing their concerns. Also, people cannot re-
call with high fidelity meaningful changes that are infrequent and brief in duration or subtle and evolving slowly over time. Accordingly, older adults may fail to sufficiently identify key transient events which because of their infrequent occurrence may be easily forgotten [24]. As for questionnaires and episodic in-person examinations, they depend on a snapshot observation of function and assume that observations recorded during the examination represent the person’s typical state of function and cognition for relatively long periods of time prior to the assessment [25].

1.2.3 Consequences of Shortcomings

These shortcomings have resulted in a high under-recognition rate of dementia. A recent study has reported that in more than 50 percent of the cases that were investigated in the study, it was the family members who served as the primary source of recognition and not the family doctors [26]. This delay in detecting cognitive decline can be detrimental especially to older adults with reversible forms of dementia, who form up to 11 percent of the cognitively impaired population [27]. Early detection of remediable causes of cognitive impairment, such as medication complications or nutritional deficiencies, facilitates timely intervention, possibly increasing the chances of reversing the condition. Even for older adults with irreversible decline, early detection of their cognitive decline still provides them and their families with an opportunity to proactively plan for their future. It provides them with ample time to seek the appropriate interventions that can maintain their quality of life and daily functioning, and that can reduce any emotional stress or behavioral symptoms such as depression, apathy, wandering, sleep disturbances, agitation, and aggression [28].
1.3 Motivation

Soon after I started my Ph.D. and while I was trying to decide on my thesis, I completed an observership at the Toronto Rehabilitation Institute (TRI), which was a requirement towards the National Science and Engineering Research Council (NSERC) CREATE Academic Rehabilitation Engineering (CARE) scholarship that I had received. I completed my observership under the supervision of Dr. Ron Keren, the medical director of the geriatric rehabilitation program at TRI. I accompanied Dr. Keren for one month as he administered cognitive assessments to patients. At TRI, a patient is normally asked to be accompanied by family members who reside at the same residential address, or by caregivers if the patient lives alone, since the cognitive assessment is composed of two parts: (i) interviewing family members or caregivers, and (ii) interviewing the patient and assessing his or her cognition.

1.3.1 Interviewing Family Members

At TRI, the first part of the assessment represents interviewing the family members who live with or take care of the patient in order to elicit from them as much relevant information as possible about noticeable changes in the patient’s living patterns and changes in his or her cognitive ability. This interview is conducted in the absence of the patient. Questions related to changes in memory include whether the patient has been forgetting conversations, or whether he or she has been repeating questions or the same stories again, or whether he or she has been forgetting days, dates, anniversaries, birthdays, or appointments. Family members or caregivers are also asked about the patient’s daily behavior in terms of the frequency of misplacing items, forgetting to pay the bills, or paying the same bill more than once. In addition, family members or caregivers are also asked about possible changes to the patient’s executive functioning in terms of any noticeable decline in handwriting or spelling, any difficulty operating
machines, computers, or TV, or any noticeable changes in his or her ability to plan and organize. Finally, family members are asked about any possible changes in the patient’s mood such as being more irritable, less social, or suspicious.

1.3.2 Assessing Cognitive Ability

The second part of the assessment represents administering an assessment of the patient’s cognition in order to screen the patient for mild cognitive impairment (MCI). At TRI, screening for MCI is performed primarily using the Montreal Cognitive Assessment (MoCA), which consists of eight sections and assesses several cognitive domains such as the patient’s visuospatial ability, executive functioning, language ability, short-term memory recalling ability, and his or her attention, concentration, and working memory ability. MoCA also assesses the patient’s delayed recall ability, and evaluates his or her orientation to time and place.

In addition to the MoCA tool, the patient’s memory is further assessed using word retention exercises, where ten words are shown on a computer screen after which the patient is asked to repeat as many words as they can recall.

1.3.3 Conclusion

Surprisingly, only about 10 percent of the patients who were assessed by Dr. Keren during my observership were assessed as cognitively intact or experiencing mild cognitive impairment. The remaining 90 percent of the patients were assessed as having mild to moderate dementia. Frustration and confusion were clear on the patients’ faces as Dr. Keren communicated the results to them. Some of them broke into tears as they were informed that they could not live independently anymore, or that their driver’s license was going to be revoked. Evidently, the experience was shocking, horrendous, and completely unpleasant to those patients. This explains why the majority of older adults are forced into crisis management soon after they are informed that they have
dementia [29].

This personal experience inspired and motivated me to work on designing a smart system that can detect dementia in its early stages through continuous monitoring of older adults in their homes, especially after many family members interviewed by Dr. Keren reported that they first started observing changes in the patient’s cognitive ability years before the assessment. A smart system that can unobtrusively detect early changes in cognition can provide older adults and their families as well as physicians with numerous advantages.

As for older adults, they would be better able to report symptoms and concerns and understand their own impairment. They would also have the opportunity to seek community resources and care that could potentially prolong their independence. In addition, they would have more treatments available as the condition is still in its early stages. They could also choose to participate in clinical studies that would help advance the clinical understanding of cognitive impairment and its progression. Finally, such a system would provide families with the opportunity to learn about Alzheimer’s disease and dementia and plan for their future together, which would result in reduced stress, anxiety, and feelings of burden.

As for physicians, such a system would facilitate easier identification of the treatable or reversible types. It would also enable physicians to more effectively manage complications from co-existing medical conditions and to better understand the patient’s abilities to manage their own care, including taking medications. Furthermore, it would enable physicians to encourage patients to take part in clinical studies in their residential area and to discuss with them the significance of their participation in further understanding their impairment and its progression. Finally, given that older adults would have prolonged periods of intact cognition, such as system would enable physicians to continue exercising giving their patients the right for self-determination and the right to make their own health-care decisions, whenever possible.
1.4 Research Questions

This dissertation revolves around answering the following three main research questions:

1. How accurately can machine learning algorithms detect mild cognitive impairment in older adults using a number of predefined measures associated with in-home walking speed and general activity, calculated from unobtrusive sensing technologies?

2. Can we build statistical models of older adults’ in-home activity that provide intuitive statistical analysis, and how would these models of activity of older adults when cognitively intact compare to their models when experiencing mild cognitive impairment?

3. How can we use models of older adults’ in-home activity to automatically detect mild cognitive impairment in older adults using unobtrusive sensing technologies?

1.5 Outline

We conclude this chapter by giving an overview of the rest of the chapters in this dissertation and their relationship with published work in peer-reviewed conferences and journals.

Chapter 2 lays the foundation for all of the subsequent chapters. It describes other related works in the literature and highlights their major limitations. Chapter 2 also introduces mild cognitive impairment and discusses how detecting it facilitates detecting dementia in its early stages. Furthermore, Chapter 2 describes a clustering technique, affinity propagation, that forms a building block of the algorithm described in Chapter 8.

Chapter 3 describes the data that we received from the Oregon Center for Aging and Technology and that we utilized in our analysis and used to build the algorithms described in Chapters 6 – 8. Chapter 3 also describes the inclusion criteria defined by
Chapter 1. Introduction

the Oregon Center for Aging and Technology to recruit older adults and describes the sensing technologies deployed into the homes of older adults for data acquisition.

Chapter 4 discusses the labeling protocols that we defined and implemented for two ground truths: the Clinical Dementia Rating and neuropsychological assessments. Chapter 4 also describes the three-stage cleaning process that we completed to clean the data, and provides a summary of the clean data available for analysis.

Chapter 5 presents a case study and explains the importance of using certain evaluation metrics, such the \( F \)-score or the area under the precision-recall curve, to evaluate the performance of the algorithms built in Chapters 6 – 8 using imbalanced data sets.

Chapter 6 explores the feasibility of autonomously discriminating older adults with mild cognitive impairment from their cognitively intact counterparts using a number of predefined measures associated with their in-home walking speed such as the median weekly walking speed, the coefficient of variation of the weekly walking speed, the coefficient of variation of the morning walking speed, and other measures. Different window sizes are used to extract features from these predefined measures. The extracted features are then used to train and test two machine learning algorithms, namely support vector machines and random forests. Chapter 6 concludes with a ranking of the extracted features and an interesting discussion of their importance to the detection process.

Chapter 7 explores the feasibility of detecting mild cognitive impairment in older adults by monitoring changes in room-level activity instead of longitudinal walking speed trajectories. Independent inhomogeneous Poisson processes are used to build generalized linear models to model the presence of older adults within different rooms throughout the day. Very interesting insights are extracted about the behavior of older adults during different cognitive states. In addition, Chapter 7 discusses interesting changes in the estimated generalized linear models of older adults’ activity models as they start experiencing cognitive impairment.
Chapter 8, motivated by the promising results in Chapter 7, describes a clustering-based method to automatically detect mild cognitive impairment in older adults using the estimated generalized linear models of their room activity. Affinity propagation is employed to search for exemplars to represent clusters of the room activity models belonging to different states of cognition. These exemplars are then used to determine the cognitive status of an older adult with unknown cognitive status. Chapter 8 also investigates the two subtypes of mild cognitive impairment: the amnestic and the non-amnestic subtypes, and reports very interesting findings about trying to automatically detect each subtype.

Chapter 9 concludes this thesis by summarizing the main contributions, discussing the main limitations of this dissertation, and proposing interesting potential directions for future work.

1.6 Relationship to Published Work

The algorithms built and described in Chapter 6, Chapter 7, and Chapter 8 have been published in peer-reviewed conferences and journals. We next outline the relevant peer-reviewed publications for each of these chapters.

Chapter 6


Chapter 7

Chapter 1. Introduction


Chapter 8

Chapter 2

Background

What ‘Philadelphia’ did for AIDS,
‘Still Alice’ may do for Alzheimer’s.

Maria Shriver

In this chapter, we discuss the background material relevant to the subsequent chapters. We start off by summarizing several studies that demonstrated a relationship between changes in motor capabilities and cognitive impairment. Then, we review key smart systems that have been designed to continuously monitor individuals in their home environment, and we address their main limitations. Finally, we conclude the chapter by discussing affinity propagation which is the main clustering algorithm employed in Chapter 8.

2.1 Motor Capabilities and Cognition

Many studies have identified several risk factors that predict the onset of Alzheimer’s disease and dementia in older adults. The strongest include age, sex, educational level, genetic makeup, mild cognitive impairment, and gait impairment [30].

Marquis et al. conducted a study to examine integrating several individually vali-
dated predictors of cognitive decline to determine if they independently predict decline in a group of cognitively intact individuals [31]. The study involved one hundred eight older adults and ran for a mean period of six years. Each recruited older adult underwent evaluation at six-month intervals using standardized assessment tools and annual neurologic and neuropsychological assessments and magnetic resonance imaging. Among the many findings of the study was that impaired gait, independent of significant medical, orthopedic, or rheumatological disease, was a significant predictor of the development of persistent cognitive impairment (PCI), which is a persistent state of cognitive decline close to dementia. Interestingly, the study also reported that motor impairment could be observed prior to the development of cognitive decline in older adults.

In a similar study, Tabbarah et al. tested the relationship between patterns of change in cognition and change in performance on various physical tasks such as standing on a single leg, tapping foot, turning in a circle, and walking at a fast or normal pace [32]. The study was based on data from four hundred eighty-eight older adults and ran for a mean period of seven years. Assessment of the recruited older adults included detailed measurement of physical performance, cognitive performance, health status, and social and psychological characteristics. Recruited older adults were assessed at baseline and assessed again after seven years resulting in two sets of scores to be compared: baseline and follow-up. Analysis of the acquired data showed that declines in cognitive performance were associated with decline in performance on physical tasks. These findings about the interrelation between patterns of change in cognitive and physical performance suggested that cognition plays an integral role in the execution of most physical tasks.

Another study that aimed at exploring the independent predictors of cognitive impairment was designed by Camicioli et al. [33]. This study involved eighty-five older adults who were functionally independent at baseline, without complaints of memory impairment or clinical evidence of dementia or depression. Physical assessment of the recruited older adults included number of finger tappings in ten seconds, hand opening
and closing in ten seconds, timed one-leg standing, and steps and time to walk thirty feet. The recruited older adults were assessed at six-month intervals. Cognitive impairment was defined by a Clinical Dementia Rating (CDR) score of 0.5 or more based on the presence of cognitive or functional decline either by self-report, mental status evaluation, or collateral report. The study deduced that motor slowing was evident on clinical examination before or coincident with the development of cognitive impairment in cognitively intact older adults. More concretely, steps to walk thirty feet declined progressively in the recruited older adults who became cognitively impaired.

Finally, Atkinson et al. conducted a study, along the same line as the study completed by Camicioli et al., with the objective of determining the incidence and correlates of combined declines in cognitive and physical performance [34]. Analyses were performed on data pertaining to five hundred fifty-eight older adults and the study ran for over three years with in-home study assessments by trained study nurses every six months. The recruited older adults were assessed cognitively using the Folstein MMSE, and were assessed physically by measuring the time it took them to walk a distance of four meters. Cognitive decline was defined as a decrease in MMSE score to less than 24. Physical decline was defined as a decrease in walking speed to less than 0.4 meters per second since a walking speed lower than 0.42 meters per second has been demonstrated to be predictive of functional dependence and is considered to represent severe walking disability [35] [36]. From the assessment of statistical dependence of cognitive and physical decline, the study demonstrated that physical decliners were 1.98 times as likely to have cognitive decline as non-physical decliners, and cognitive decliners were 1.80 times as likely to develop physical decline.

All these studies have demonstrated that changes in motor capabilities precede and may be indicative of cognitive impairment, and that changes in walking speed would be good predictors of progression to dementia [37]. Accordingly, and with the advancement of technology and the proliferation of smart systems, a good alternative to the traditional
clinical paradigm is to bring assessment into the daily activity of a person in their home environment via unobtrusive sensors and systems.

2.2 Smart Systems and Related Work

The literature is rich with smart systems and studies designed to monitor the health and well-being of older adults and to support their independence. Few systems are commercially available whereas others are still in the prototype stage being further developed in research labs.

2.2.1 Commercial Systems

Among the commercially available systems are the GE QuietCare system [38] and the GrandCare System [39]. Both QuietCare and GrandCare systems employ wireless motion sensors in key areas including the bedroom, kitchen, bathroom, and meal preparation and medication areas for motion detection. GrandCare, however, also employs magnetic door switches, temperature sensors, light switch sensors, and other sensors placed on the bed to monitor wellness, blood pressure, and weight. These sensors communicate with a central base station, which in turn periodically transmits the information to central computers over standard telephone lines. Systems are tailored according to the occupant’s living patterns. The objective of these systems is to monitor the occupant’s activity and alert caregivers to possible problems, particularly missing events, such as the occupant not being present in bed when it is bed time or at the medication kit when it is time to take their medication. Caregivers can monitor the occupant’s general activity levels and check whether they are following normal patterns of living or not.
2.2.2 Prototype Systems

Among the numerous “prototype” systems is the Microsoft’s EasyLiving project [40]. The Microsoft’s EasyLiving project, based on “context aware computing,” uses tracking video to monitor residents [41]. Images from the input video are analyzed and processed using distributed computing. The system identifies people-shaped clusters of blobs in real time, allowing the system to follow individuals through the house. EasyLiving project uses geometric models to define geometric relationships between entities including people and devices. In an example scenario, Tom, the resident, wishes to start playing music. The smart home uses its geometric world knowledge to select those speakers and other components which are most suitable for the task, based on Tom’s current location. Thus, Tom is able to focus only on the decisions that require his input: the music itself. Current development is focusing on integrating more devices to provide a coherent user experience. Research is still on-going on a variety of fronts, including middleware development to facilitate distributed computing, geometric world modeling to provide location-based context, computer perception to gather data about world state, and better service description to support decomposition of device control, internal logic, and user interface.

Another smart house is the Gator Tech Smart House (GTSH) [42]. GTSH’s goal is to create assistive environments that can provide special services to the residents to compensate for cognitive, mobility, health, and other age-related impairments [43]. GTSH provides several services and features including entry assistant where the resident is granted automatic entrance by means of passive RFID (Radio-Frequency Identification) tags attached to their keys. Another feature of GTSH is location tracking and activity monitoring. A smart floor consisting of residential-grade raised platform represents the location tracking system through force sensors. GTSH uses the smart floor in addition to computer vision and RFID technologies to recognize residents and learn their activities. If significant changes in activity are registered, a caregiver is immediately notified.

Similar to GTSH is the Aware Home Research Initiative at the Georgia Institute
of Technology [44]. The Aware Home is a three-story home that functions as a living laboratory for the design, development, and evaluation of future domestic technologies [45] [46]. The smart floor senses an individual’s footsteps, which allows the home to build a model based on user habits and behavior. The main goal of the Aware Home Initiative is to enable older adults to remain in familiar surroundings as they age, helping them maintain their quality of life. A very intriguing feature of the Aware Home is a system of tracking and sensing technologies to help find frequently lost objects such as wallets, glasses, and remote controls. Through small radio-frequency (RF) tags, older adults interact with the system via LCD touch panels placed strategically throughout the house to guide them to the lost object using audio cues.

Another smart house is The House.n at MIT or “the house of the future” [47]. The House.n proposes a smart service-delivery system that conducts qualitative and quantitative studies on the relationships between environmental factors and the behavior of an inhabitant [48]. The system consists of three components: a set of state-change sensors used to collect data about the use of objects, a context-aware experience sampling tool (ESM) used by inhabitants to label their activities, and pattern recognition and classification algorithms for recognizing activities. The inhabitant model is built using a training data set. In one of the studies carried out by the House.n group, the sensors were installed on pieces of furniture, kitchen appliances, bathroom appliances, and the washing machine. Recruited older adults were given a personal digital assistant (PDA) running the ESM software at the start of the study and were asked to collect their activity data. A naive Bayesian network approach was then used to train the recruited older adults’ models and predict their activities.

Unlike the aforementioned smart homes, CarerNet is a smart system that deploys various “telecare” and “hospital at home” services such as a home emergency alarm, access to community health information, and ambulatory monitoring [49]. In addition to therapy units, CarerNet includes a sensor set, a sensor bus, an intelligent monitoring
Chapter 2. Background

system, and a control unit. CarerNet collects physiological data such as ECG, photo-
plethysmograph, spirometry, temperature, galvanic skin response, colorimetry, and pulse,
and determines the patient’s lifestyle through passive infra-red sensors, accelerometers,
inductive badges, smart information technology (IT) badges, and piezoelectric sensors.
CarerNet senses and is aware of the environment through thermometers, microphones,
and infra-red smoke alarms.

A similar home health system is TERVA [50]. The TERVA system monitors physiolog-
cal and psychological health through vital sign measurements, including arterial blood
pressure, heart rate, body temperature, and the patient’s long-term behavioral diary.
Long-term behavior includes data on mood and emotional responses, and use of tobacco,
alcohol, coffee, and tea. The system is controlled through a laptop computer, and the
measurement modules are controlled via a dedicated user interface. Data from the mea-
surement devices are transmitted to the computer via a serial interface (RS232). The
system includes a blood pressure monitor, a thermometer, an electrocardiogram (ECG)
and activity monitoring device, a scale to measure body weight, a static charge-sensitive
bed, a diary, and a laptop software used to initiate and manage bedside measurements
of the patient.

The most recent published smart system is the Center for Advanced Studies in Adap-
tive Systems (CASAS) at the Washington State University. CASAS is an on-campus
smart home that is instrumented with motion sensors on the ceiling, door sensors on
cabinets and doors, item sensors, temperature sensors in each room, and many other
sensors.

2.2.3 Related Studies

In addition to the above smart systems, many groups carried out studies on continuous
monitoring of older adults in their home environment and tried to detect changes in their
living patterns that were indicative of a change in their health or cognition.
In one of their recent studies, the CASAS group used a machine learning approach to discriminate cognitively impaired older adults based on their ability to complete a number of daily activities [51] [52]. More specifically, one hundred seventy-nine older adults were recruited to complete a “Day Out Task” that consisted of carrying out eight Instrumental Activities of Daily Living (IADL) that might be interwoven. Data acquisition was completed in the CASAS testbed. Machine Learning algorithms were employed to discriminate the participating older adults based on several features that were computed from the sensor data.

In another study, Sixsmith performed a trial to test the ability of an intelligent monitoring system, composed of sensors in the home, to identify emergencies by detecting deviations from normal activity patterns [53]. Twenty-two older adults were recruited and monitored for three months. The older adults’ ages ranged from 60–85 years and lived in four differed localities in the UK–Ipswich, Northumberland, Merseyside, and Nottingham. The sensing technologies consisted of passive infra-red sensors, contact sensors, and temperature sensors.

In another study, similar to the one conducted by Sixsmith, Skubic et al. implemented a simple alert algorithm to generate health alerts to clinicians via continuous monitoring of older adults in their homes using unobtrusive sensing technologies [54]. Twenty-one older adults were recruited and monitored for nine months. The sensing technologies consisted of passive infra-red motion sensors, bed sensors, and temperature sensors.

Finally, Hayes et al. conducted a study by deploying sensing technologies into the homes of fourteen older adults, aged 65 years and older, and monitored them for an average period of three hundred fifteen days [25]. High-level sensing technologies were used including passive infra-red motion sensors to detect general movement in the house, and contact switches placed at the front door to detect entry and exit from the house. Using a total of 108 thousand person-hours of continuous activity, Hayes et al. used wavelet analysis to analyze the variance of activity of the fourteen older adults at multiple time
scales. It was found that, for some older adults, the 24-hour wavelet variance was greater in older adults as they started having mild cognitive impairment, which potentially indicated that the day-to-day patterns of activity of older adults in the impaired group were more variable than those of their cognitively healthy counterparts.

2.2.4 Limitations of Current Smart Systems and Studies

While these smart systems can provide continuous monitoring of older adults and the aforementioned studies extracted very interesting insights, they still suffer from several major limitations:

1. They are currently very expensive and laborious to implement in people’s homes. Many of the above systems rely heavily on installing many sensors around the house and in the environment. Furthermore, some of them necessitate tagging objects with sensors. These require a significant undertaking both in installation and maintenance which render them unadaptable to the goal of unobtrusive monitoring.

2. The majority of the aforementioned smart homes and systems are not designed specifically for the objective of detecting changes in cognition. They focus on learning living patterns of inhabitants in order to automate repetitive tasks and patterns. It was only the group at CASAS who published a study on using the CASAS testbed to discriminate older adults with cognitive impairment from their cognitively healthy counterparts.

3. The contemporary performance of many of the smart homes and other related studies is not competitive. For example, the House_n group at MIT acknowledges that the performance of House_n is not as high as expected. The main problems were the low quality and number of activity labels, and the small training set of two weeks. The House_n group hopes to enhance the performance by collecting more training data, generating higher quality activity labels by video observation or other meth-
ods, and improving the information collection boards and sensors. As for the study by Sixsmith [53], of the total sixty-one alerts that were recorded in three months, forty-six alerts were classified as false alerts and only fifteen alerts were classified as genuine. In other words, the approach suffered from a high false positive rate, approximately 75 percent. Similarly, the alert algorithm described by Skubic et al. in [54] generated one hundred eighty-three poor alerts out of a total of two hundred ninety-nine alerts in nine months. This translated into an accuracy of 38.8 percent and a false positive rate of 61.2 percent.

4. All the results reported by the majority of the aforementioned systems and studies were based on data acquired in a laboratory environment and not in a real world setting. An approach that would be more reflective of older adults’ actual performance would be to continuously monitor them in their own homes, where they are more likely to exhibit their true unbiased performance.

2.3 Mild Cognitive Impairment

In recent years, more attention is being paid to the mild end of the cognitive spectrum spanning normal aging to dementia, mainly caused by Alzheimer’s disease. There likely is a transitional period between normal aging and the diagnosis of clinically very early dementia, and this transitional period has been described using a variety of terms such as mild cognitive impairment, dementia prodrome, incipient dementia, isolated memory impairment, and some others [55]. In this thesis, we refer to this period as mild cognitive impairment or MCI.

Mild cognitive impairment (MCI) is a concept that describes the transitional period between normal aging with intact cognition and early Alzheimer’s disease and dementia as depicted in Figure 2.1 [56]. Individuals with MCI have measurable changes in their memory and thinking abilities but with minor impact on their ability to carry out ac-
Figure 2.1: Cognitive continuum showing the overlap between normal aging and mild cognitive impairment, and mild cognitive impairment and early dementia or Alzheimer’s disease [56].

Activities of daily living. An individual with MCI would typically start forgetting things more often, forgetting important events such as appointments or social engagements, losing their train of thought or the thread of conversations, movies, or books. Furthermore, an individual with MCI would feel increasingly overwhelmed by making decisions or interpreting instructions, and would become more impulsive or show increasingly poor judgment.

Although not all individuals with MCI progress to develop Alzheimer’s or other dementias, the proposed criteria and guidelines for diagnosis of Alzheimer’s disease published in 2011 have documented that individuals with MCI are at a higher risk of developing dementia and Alzheimer’s disease [57]. Therefore, detecting MCI serves the objective of detecting cognitive decline early enough for individuals at risk of developing dementia to seek the right intervention and treatment. In this thesis, we use MCI as an indicator of the onset of dementia or Alzheimer’s disease in older adults.
2.3.1 Types of Mild Cognitive Impairment

As with dementia, because different parts of the brain are responsible for different functions and behaviors, different subtypes of MCI exist depending on what cognitive skills are compromised. The two main subtypes of MCI are amnestic MCI (a-MCI) and non-amnestic MCI (na-MCI).

In amnestic MCI, memory is significantly impaired while other cognitive functions remain intact. In non-amnestic MCI on the other hand, memory remains intact but one or more other cognitive functions, such as language, visuospatial, or executive functioning skills, are significantly impaired.

2.3.2 Single Domain and Multiple Domain MCI

In single domain MCI, only one cognitive domain, such as memory or executive functioning, is impaired. In multiple domain MCI, one or more cognitive abilities are compromised. Impairment only in memory results in amnestic MCI - single domain whereas impairment in memory and other cognitive functions results in amnestic MCI - multiple domain. Similarly, impairment in a single cognitive domain, other than memory, results in non-amnestic MCI - single domain whereas impairment in more than one cognitive domain, other than memory, results in non-amnestic MCI - multiple domain.

Older adults with amnestic MCI, single or multiple domain, are at an increased risk for Alzheimer’s disease. Older adults with non-amnestic MCI are at an increased risk for other dementias, such dementia with Lewy bodies, and Parkinson’s disease dementia.

2.4 Affinity Propagation

In this thesis, we use a new novel technique for data clustering called affinity propagation to cluster activity models [58]. This algorithm is mainly utilized in Chapter 8. In this section, we describe the affinity propagation algorithm and explain how it works.
Affinity propagation is an algorithm that simultaneously considers all datapoints as potential exemplars and recursively transmits real-valued messages until a good set of exemplars and clusters emerges. Affinity propagation does not require the number of clusters be known prior to clustering. Instead, the clusters emerge naturally.

### 2.4.1 Similarity Function

Affinity propagation takes as input a collection of real-valued similarities between datapoints, where the similarity $s(i, j)$ indicates how well suited the datapoint with index $j$ is suited to be the exemplar for datapoint $i$. Affinity propagation aims to maximize the similarity $s(i, j)$ for every datapoint $i$ and its chosen exemplar $j$. For example, the similarity function could be the negative Euclidean distance between the datapoints. Negative Euclidean distance is used so that a maximum similarity corresponds to the closest datapoint.

In addition to the measure of similarity, affinity propagation takes as input a set of real numbers, known as self-similarity $s(i, i)$, or preference ($p$), for each datapoint. The preference ($p$) influences the number of exemplars that emerge. Initializing a datapoint with a larger or smaller self-similarity, respectively increases or decreases the likelihood of the datapoint becoming an exemplar. If all the datapoints are initialized with the same constant self-similarity, then all datapoints are equally likely to become exemplars. The preference ($p$) also controls how many clusters are generated. By increasing and decreasing this common self-similarity input, the number of clusters produced is increased and decreased, respectively. If all datapoints are assigned the median of the input similarities, a moderate number of clusters emerges, and if they are assigned the minimum of the input similarities, the smallest number of clusters emerges. Affinity propagation can generate better clusters, compared to other clustering techniques, like $K$-means clustering, because of its initialization-independent property [58].
2.4.2 Message Passing: Responsibility and Availability

Clustering is based on the exchange of two types of messages: the “responsibility” message to decide which datapoints are exemplars, and the “availability” message to decide to which cluster a datapoint belongs.

- The responsibility message, \( r(i, j) \), sent from datapoint \( i \) to candidate exemplar \( j \), reflects the accumulated evidence for how well-suited datapoint \( j \) is to serve as the exemplar for datapoint \( i \), taking into account other potential exemplars, \( j' \), for datapoint \( i \). Mathematically, this is equivalent to

\[
    r(i, j) = s(i, j) - \max_{j' \text{ s.t. } j' \neq j} \left\{ a(i, j') + s(i, j') \right\} ,
\]

where \( i \neq j \), and \( s(i, j) \) is the similarity between datapoint \( i \) and datapoint \( j \) and \( a(i, j) \) is the availability message defined below.

- The availability message, \( a(i, j) \), sent from candidate exemplar \( j \) to datapoint \( i \), reflects the accumulated evidence for how appropriate it would be for datapoint \( i \) to choose datapoint \( j \) as its exemplar, taking into account the support from other datapoints, \( i' \), that datapoint \( j \) should be an exemplar. Mathematically, this is equivalent to

\[
    a(i, j) = \min \left\{ 0, r(j, j) + \sum_{i' \text{ s.t. } i' \neq i, j} \max\{0, r(i', j)\} \right\} .
\]

- The self-responsibility, \( r(j, j) \), and self-availability, \( a(j, j) \), are two additional messages calculated for each datapoint \( j \). Both of these messages reflect accumulated evidence that datapoint \( j \) is an exemplar, and they are used to find the clusters. The self-responsibility bases exemplar suitability on input preference and the maximum availability received from surrounding datapoints whereas the self-availability bases
exemplar suitability on the number and strength of positive received responsibilities. Mathematically,

\[ r(j, j) = s(j, j) - \max_{j', s.t. j' \neq i} \left\{ a(j, j') + s(j, j') \right\}, \quad (2.3) \]

and,

\[ a(j, j) = \sum_{i', s.t. i' \neq j} \max\{0, r(i', j)\}. \quad (2.4) \]

### 2.4.3 Cluster Decisions

With affinity propagation, the exemplar of each datapoint \( i \) is found using the following equation,

\[ \text{exemplar}_i = \arg \max_j \{ a(i, j) + r(i, j) \}. \quad (2.5) \]

This clustering procedure may be performed at any iteration of the algorithm, but final clustering decisions should be made once the algorithm stabilizes. The algorithm can be terminated once exemplar decisions become constant for some number of iterations, indicating that the algorithm has converged. It should be noted that the algorithm possesses another useful feature; it is possible to determine when a specific datapoint has converged to exemplar status for a specific iteration. When a datapoint’s self-responsibility plus self-availability becomes positive, that datapoint has become the exemplar for its cluster.
Chapter 3

Oregon Center for Aging and Technology Data

The data may not contain the answer. The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.

John Tukey

In this chapter, we describe the data that we received from the Oregon Center for Aging and Technology and that we utilized to answer the main research questions of this thesis, listed in Chapter 1. We start off by summarizing the inclusion criteria that was implemented by the Oregon Center for Aging and Technology to recruit older adults. Next, we describe the sensing technologies that were deployed into the homes of the recruited older adults for continuous monitoring. Finally, we provide a summary of the data that we received from the Oregon Center for Aging and Technology, and we discuss one study by the Oregon Center for Aging and Technology that is relevant to the work.
3.1 Data from the Oregon Center for Aging and Technology

The Oregon Center for Aging and Technology (ORCATECH) at the Oregon Health and Science University (OHSU) developed and pilot-tested a community-wide, scalable home-based assessment platform and protocol. They deployed unobtrusive sensing technologies into the homes of approximately three hundred healthy older adults for at least thirty-six months, resulting in a huge database of sensor data. A detailed description of ORCATECH’s sensing system, inclusion criteria of older adults, and their approach of data acquisition are discussed next.

3.1.1 Subjects

In this thesis, we use the term “subjects” to refer to the older adults who were recruited by ORCATECH. All subjects were recruited from the Portland, Oregon metropolitan area and provided written informed consent before participating in the study activities. Eligibility criteria included [59]:

- being a man or a woman aged 80 years or older (or 70 years or older for non-Whites and for individuals residing with a participant aged 80 years or older),

- living independently (cohabitation with a companion or a spouse was allowed but not with a formal caregiver),

- residing in a larger than one-room or “studio” apartment,

- cognitively healthy, i.e., Clinical Dementia Rating (CDR) score < 0.5, Mini-Mental State Examination (MMSE) score > 24, and,
• in average health for age, i.e., well-controlled chronic diseases and comorbidities or none at all.

Medical illnesses with the potential to limit physical participation (e.g. wheelchair bound) or likely to lead to ultimately death over 35 months were exclusions. Subjects lived in a variety of settings from apartments in organized retirement communities to single-family detached homes.

### 3.1.2 Sensing System

Collection of continuous activity data was achieved by installing a network of wireless sensors in the home of each subject. This network consisted of the following [60]:

- passive infra-red motion sensors (type MS16A from X10.com),
- contact sensors (type DS10A from X10.com), and,
- a transceiver, W800, attached to the serial port of a computer.

In order to detect movement, wireless passive infra-red motion sensors (MS16A) were placed in rooms frequently visited by subjects (bedroom, bathroom, kitchen, living rooms, and hallway-entry areas). Walking speed was estimated using data from sensors positioned sequentially on the ceiling approximately 61 centimeters apart in areas such as corridors or other hallways. These sensors were modified so that they had a restricted field view of $\pm 4^\circ$, corresponding to $\pm 6.5$ centimeters at a distance of 90 centimeters from the sensor, only firing when someone passed directly within their path. Wireless magnetic contact sensors, or reed switches, (DS10A) were placed on each door of the home to track visitors and absences from the home. All sensors sent their data wirelessly to a transceiver, W800, attached to the serial port of a dedicated research laptop computer placed in the subject’s home. The data received by the transceiver were then time-stamped and stored in an SQL database. All data were automatically uploaded daily to a central database in the project data center called “console.”
Floor plans of each home were drawn to provide a layout of the place and a map of placement of sensors in the house. Figure 3.1 shows an example of a home layout with sensors map superimposed on it [59]. The boxes labeled ‘S’ indicate the locations of the passive infra-red motion detectors used to detect the presence of the subject in proximity of the sensor. The boxes labeled ‘D’ indicate the locations of the contact sensors placed on exit/entry door and refrigerator door. The boxes labeled ‘W’ indicate the position of the sensor line for measuring walking speed. Finally, the box labeled ‘HC’ indicates the location of the home computer.

Figure 3.1: A home map with sensor layout [59].

3.1.3 Weekly Self-Report

Each subject received a desktop computer, and internet broadband services were provided in order to facilitate data acquisition and participation in all study activities. Subjects
received computer training based on their computer familiarity as determined by an assessment of computer experience administered at the first training session. Subjects were considered computer literate once they were able to reliably send and receive e-mails. This computer literacy was essential because subjects received a weekly online health questionnaire to complete questions about behaviors that could affect activity patterns. Questions covered nine areas concerning medication changes, falls, injuries, health changes, emergency room visits, depression, changes to living space, vacations, and visitors. Table 3.1 shows all the questions that were asked in the questionnaire [59].

In case a subject failed to complete the questionnaire on a certain week, they received a phone call from the research team at ORCATECH to inquire about the reason or just to remind them to complete the questionnaire.

### Table 3.1: Weekly self-report questions [59]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self Report Date</td>
<td>Date Form Submitted</td>
</tr>
<tr>
<td>Home</td>
<td>Been away from home overnight?</td>
</tr>
<tr>
<td>Visit</td>
<td>Had visitors that have stayed for more than one day in your home?</td>
</tr>
<tr>
<td>Meds</td>
<td>Had a change in your medication?</td>
</tr>
<tr>
<td>Fall</td>
<td>Have you had a fall, including a slip or a trip, in which you lost your balance and landed on the floor, the ground, or a lower level?</td>
</tr>
<tr>
<td>Hurt</td>
<td>Had any other injuries?</td>
</tr>
<tr>
<td>ER</td>
<td>Had any hospitalizations or emergency room visits?</td>
</tr>
<tr>
<td>Health</td>
<td>Has your physical health limited you more than usual?</td>
</tr>
<tr>
<td>Space</td>
<td>Had any changes in home-space or living situation?</td>
</tr>
<tr>
<td>Blue</td>
<td>Have you felt downhearted or blue for more than three days?</td>
</tr>
</tbody>
</table>
3.2 Annual Assessments

Each subject was assessed in-home at baseline, and during annual in-home visits by research personnel who administered standardized health and function questionnaires and physical and neurological examinations, including the Mini-Mental State Examination (MMSE), the Clinical Dementia Rating (CDR), and six neuropsychological tests, summarized in Table 3.2, considered to be representative of five cognitive domains.

Table 3.2: Six neuropsychological tests administered for cognitive assessment

<table>
<thead>
<tr>
<th>Cognitive Domain</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>Logical Memory Delayed Recall [61]</td>
</tr>
<tr>
<td>Executive Function</td>
<td>Category Fluency Animals [62] or Trail Making B [63]</td>
</tr>
<tr>
<td>Attention</td>
<td>Wechsler Adult Intelligence Scale (WAIS) Digit Symbol [64]</td>
</tr>
<tr>
<td>Language</td>
<td>Boston Naming Test [65]</td>
</tr>
<tr>
<td>Visuospatial Function</td>
<td>WAIS Revised Block Design [66]</td>
</tr>
</tbody>
</table>

3.3 Summary of Data Received from ORCATECH

We received data from ORCATECH pertaining to one hundred fourteen homes. In this thesis, we focused on homes with single occupants. Therefore, all one hundred fourteen homes had single occupants only. Table 3.3 shows a summary of the subject demographics.

Subjects were monitored for different periods of time for reasons such as subjects passing away, moving out of the metropolitan area, or feeling overwhelmed by study procedures. Table 3.4 presents statistics associated with the monitoring periods (in
Table 3.3: Summary of subject demographics

<table>
<thead>
<tr>
<th>Statistics (N = 114)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ±SD)</td>
</tr>
<tr>
<td>Education (mean ±SD in years )</td>
</tr>
<tr>
<td>Gender (% women)</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Asian</td>
</tr>
</tbody>
</table>

weeks) of the one hundred fourteen subjects.

Table 3.4: Statistics of monitoring periods of subjects

<table>
<thead>
<tr>
<th>monitoring period (in weeks)</th>
<th>minimum</th>
<th>maximum</th>
<th>mean</th>
<th>median</th>
<th>std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>43.5</td>
<td>258</td>
<td>172.6</td>
<td>185.6</td>
<td>62.4</td>
</tr>
</tbody>
</table>

3.3.1 Sets of Data

We received four sets of data:

1. **Walking Speed Data:** For each of the one hundred fourteen subjects, we received an excel file that contained the walking speeds estimated each time the subject walked under the line of the sensors, labeled ‘W’ in Figure 3.1, as shown in Figure 3.2.

   As Figure 3.2 shows, the walking speed data pertaining to each subject included the ‘Home ID,’ the ‘Subject ID,’ the ‘Walking Speed’ in centimeters per second, and the ‘Date’ and ‘Time’ at which the subject’s walking speeds were estimated.

2. **Home Activity Sensor Data:** For eighty-five subjects, we also received an excel
Figure 3.2: Subject walking speed data.

file that contained the data pertaining to the boxes labeled ‘S’ and ‘D’ in Figure 3.1, as shown in Figure 3.3.

Figure 3.3: Subject activity sensor data.
As Figure 3.3 shows, the home activity sensor data pertaining to each of the eighty-five subjects included the ‘Home ID,’ the ‘Subject ID,’ the ‘Sensor ID,’ and the ‘Date’ and ‘Time’ at which the sensor firing was received by the transceiver.

3. **Answers to Weekly Questionnaires:** In addition to the walking speed data and the home activity sensor data, we received an excel file for each of the one hundred fourteen subjects that contained their answers to the weekly health questionnaires. Each entry or row represented the subject’s answers to a weekly questionnaire. Processing of the weekly questionnaires document is discussed in details in Section 4.3.

4. **Clinical Data:** Finally, we also received clinical data pertaining to each of the one hundred fourteen monitored subjects. The clinical data, as presented in Figure 3.4 and Figure 3.5, included the dates on which the annual assessments were administered and the subject’s scores on MMSE, CDR, and the neuropsychological tests listed in Table 3.2. Furthermore, the subject’s gender, race, years of education, and age were also available.

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Date of Assessment</th>
<th>Gender</th>
<th>Years of Education</th>
<th>Race</th>
<th>Age at Visit</th>
<th>MMSE Score</th>
<th>CDR Score</th>
<th>Logical Memory Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>381</td>
<td>8/16/07</td>
<td>Female</td>
<td>12</td>
<td>Caucasian</td>
<td>90.1</td>
<td>23</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>381</td>
<td>8/28/08</td>
<td>Female</td>
<td>12</td>
<td>Caucasian</td>
<td>91.2</td>
<td>23</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>381</td>
<td>7/23/09</td>
<td>Female</td>
<td>12</td>
<td>Caucasian</td>
<td>92.1</td>
<td>28</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>381</td>
<td>8/25/10</td>
<td>Female</td>
<td>12</td>
<td>Caucasian</td>
<td>93.2</td>
<td>27</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>381</td>
<td>7/20/11</td>
<td>Female</td>
<td>12</td>
<td>Caucasian</td>
<td>94.1</td>
<td>24</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

Figure 3.4: Subject clinical data - set 1.

Furthermore, for each subject, their scores on the individual neuropsychological tests, given in Table 3.2, were transformed into z-scores based on normative values generated in the past by the Alzheimer’s Disease Centers (ADC) program for
subject-specific age, sex, and educational levels [67]. The z-scores corresponding to the five cognitive domains are shown in Figure 3.5.

### 3.4 Relevant Study by ORCATECH

We conclude this chapter by discussing a study that was published by the ORCATECH group that is relevant to the work described in Chapter 6.

Dodge et al. used the walking speeds of eighty-five older adults, monitored for an average period of three years, to generate latent trajectory models of weekly mean walking speed and walking speed variability, in the form of the coefficient of variation [68]. The walking speeds were estimated using sensors installed in the homes of the recruited older adults. Fifty-four older adults had intact cognition at baseline whereas thirty-one older adults had non-amnestic MCI (na-MCI) at baseline. The built trajectory models identified three distinct trajectories: fast, moderate, and slow of weekly mean walking speed. The proportion of older adults with na-MCI among the fast, moderate, and slow trajectories was found to be 16.7 percent, 34.6 percent, and 66.7 percent, respectively.

As for the coefficient of variation of walking speed, four distinct trajectories were identified: group 1, which represented the highest baseline and increasing coefficient of variation followed by a sharply declining coefficient of variation; groups 2 and 3, which represented fairly stable coefficient of variation; and group 4, which represented the lowest
baseline and decreasing coefficient of variation. Interestingly, the proportion of recruited older adults with na-MCI among each trajectory was found to be 53.3 percent, 33.3 percent, 16.7 percent, and 61.5 percent for groups 1, 2, 3, and 4, respectively.

Although the study made very interesting observations, a significant percentage of the recruited older adults with na-MCI was found in each of the identified groups that emerged from both the walking speed trajectory models and the coefficient of variation trajectory models. Therefore, such an approach would perform poorly when trying to discriminate older adults with mild cognitive impairment from other cognitively intact older adults.
Chapter 4

Data Labeling & Cleaning

I often hear people say that a person suffering from Alzheimer’s is not the person they knew. I wonder to myself - Who are they then?

Bob DeMarco

4.1 Introduction

In this chapter, we explain the two labeling protocols that we defined and implemented to label the data that we received from ORCATECH; one for using the Clinical Dementia Rating as our ground truth and another one for using neuropsychological assessments as our ground truth. After that, we describe the three-stage cleaning process that we used to clean the data. We conclude this chapter by providing a summary of the data available for analysis for each ground truth.

4.2 Data Labeling

Figure 4.1 shows a high-level labeling process that we implemented. Since subjects were
assessed annually, labeling of data fell into three categories:

1. cognitively intact ("CIN"),
2. transitioning to mild cognitive impairment ("TR"), and,
3. experiencing mild cognitive impairment ("MCI").

In this thesis, we use two cognitive assessments to serve as our ground truth. In Chapter 6 and Chapter 7, we use the Clinical Dementia Rating (CDR) as our ground truth. In Chapter 8, we use the neuropsychological tests listed in Table 3.2 as our ground truth. In the next subsections, we discuss how we used the CDR and the neuropsychological tests to assign the above labels to subject data.

4.2.1 CDR Scale

Using CDR as our ground truth, mild cognitive impairment (MCI) was defined as a score of 0.5 on the CDR scale whereas a score of 0 indicated cognitive intactness. In order to ensure consistency in labeling subjects based on their annual assessment scores, we developed the following labeling protocol:

1. An impairment, i.e., a score of 0.5 on the CDR scale, should last for at least two consecutive annual assessments to be considered. An example is shown in Table 4.1.
As shown in Table 4.1, the subject exhibited impairment on the 1st annual assessment, and this impairment continued to appear on the next two subsequent annual assessments. Therefore, the subject is considered impaired starting from the 1st annual assessment onward.

2. If a subject exhibits impairment on an annual assessment, but this impairment disappears on all of the subsequent annual assessments, then the impairment is overlooked and the subject is considered cognitively intact throughout the monitoring period. An example is shown in Table 4.2.

As shown in Table 4.2, on the 1st annual assessment, the subject scored 0.5 on the CDR scale. However, this impairment disappeared on all of the subsequent annual assessments. Therefore, this impairment is overlooked and the subject is considered cognitively intact for the entire monitoring period.

3. If a subject exhibits impairment on an annual assessment, and this impairment disappears on the following annual assessment, but then reappears on all of the subsequent annual assessments, then the assessment on which impairment is absent is
overlooked and the subject is considered impaired starting from the first assessment that shows impairment. An example is shown in Table 4.3.

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>baseline</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR score</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

As shown in Table 4.3, on the 1st annual assessment, the subject scored 0.5 on the CDR scale. However, this impairment disappeared on the 2nd annual assessment, but then reappeared on all the subsequent assessments. Therefore, the missing impairment on the 2nd annual assessment is overlooked and the subject is considered impaired from the 1st annual assessment onward.

4. If a subject has \( n \) annual assessments, and exhibits impairment one the \((n - 1)^{th}\) annual assessment, which is equivalent to the last annual assessment, with no previous impairment or instance of bouncing, then the subject is considered impaired from the \((n - 1)^{th}\) annual assessment onward although no subsequent assessments are available. An example is shown in Table 4.4.

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>baseline</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR score</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

As shown in Table 4.4, the subject exhibited impairment on the 3rd annual assessment, after which no subsequent assessments are available. In this case, the subject is considered cognitively impaired from the 3rd annual assessment onward.
Here is an example of how we used the above labeling protocol to assign the labels depicted in Figure 4.1 to the data of the subject whose assessment scores are shown in Table 4.5. According to Table 4.5, the subject exhibited impairment on the 3\textsuperscript{rd} annual assessment, and the impairment appeared again on the subsequent annual assessment. According to rule 1, since the impairment appeared on two consecutive annual assessments then it is considered. Also, according to rule 4, the subject is considered impaired from the 4\textsuperscript{th} annual assessment onward. Accordingly:

1. The subject data from baseline until the date of the 2\textsuperscript{nd} annual assessment will be assigned the label “CIN.”

2. The subject data from the date of the 2\textsuperscript{nd} annual assessment up to the date of the 3\textsuperscript{rd} annual assessment will be assigned the label “TR.” This is because the conversion to cognitive impairment is not a point event but a gradual process. Accordingly, the subject’s cognition would be in flux between the 2\textsuperscript{nd} and the 3\textsuperscript{rd} annual assessments and would belong to neither “CIN” nor “MCI.”

3. The subject data from the date of the 3\textsuperscript{rd} annual assessment onward will be labeled “MCI.”

Three more examples are discussed in Appendix A.
4.2.2 Neuropsychological Tests

As for using neuropsychological tests as our ground truth, diagnosis of MCI was made using the Petersen criteria operationalized as objective impairment on one or more of the neuropsychological tests listed in Table 3.2, which was equivalent to a $z$-score of -1.5 or more [56]. Neuropsychological assessments equipped us with the ability to investigate different subtypes of MCI, namely amnestic MCI (a-MCI) and non-amnestic MCI (na-MCI).

- a-MCI was defined as impairment on the neuropsychological testing in the memory domain only.

- na-MCI was defined as impairment on the neuropsychological testing in any of the four domains - executive function, language, attention, and visuospatial but not in the memory domain.

Similar to CDR, we developed a labeling protocol in order to ensure consistency in labeling subject data based on their annual assessment scores. However, it was more challenging to build this protocol since several cognitive domains were involved:

1. An impairment in a domain should last at least two consecutive annual assessments to be considered. An example is shown in Table 4.6.

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>MMSE</th>
<th>CDR</th>
<th>Executive Function</th>
<th>Language</th>
<th>Attention</th>
<th>Memory</th>
<th>Visuospatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>30</td>
<td>0</td>
<td>-0.03</td>
<td>-1.94</td>
<td>-0.43</td>
<td>-0.91</td>
<td>-0.43</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>27</td>
<td>0</td>
<td>-0.56</td>
<td>-1.49</td>
<td>-0.72</td>
<td>-1.33</td>
<td>-0.75</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>28</td>
<td>0</td>
<td>-0.57</td>
<td>-1.73</td>
<td>-0.83</td>
<td>-0.45</td>
<td>-0.31</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>29</td>
<td>0</td>
<td>-0.20</td>
<td>-0.64</td>
<td>-0.88</td>
<td>-0.82</td>
<td>-0.06</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt;</td>
<td>27</td>
<td>0</td>
<td>-0.49</td>
<td>-2.45</td>
<td>-0.42</td>
<td>-0.91</td>
<td>-0.08</td>
</tr>
</tbody>
</table>
As shown in Table 4.6, the subject exhibited impairment in the language domain at baseline. However, this impairment disappeared on the subsequent annual assessment, but reappeared on the 2nd annual assessment. Then, it disappeared again on the 3rd annual assessment. Although it appeared again on the 4th annual assessment, we were not able to claim with high confidence that the subject exhibited symptoms of MCI on the subsequent assessment. Therefore, the subject is labeled as “bouncing,” and such subjects are excluded from our analysis.

2. If a subject exhibits impairment in one domain on an annual assessment, but this impairment disappears on all of the subsequent assessments, then this impairment is overlooked and the subject is considered cognitively intact throughout the monitoring period. An example is shown in Table 4.7.

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>MMSE</th>
<th>CDR</th>
<th>Executive Function</th>
<th>Language</th>
<th>Attention</th>
<th>Memory</th>
<th>Visuospatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>29</td>
<td>0</td>
<td>-0.59</td>
<td>-0.65</td>
<td>-1.17</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>1st</td>
<td>29</td>
<td>0</td>
<td>-1.20</td>
<td>-1.28</td>
<td>-0.93</td>
<td>0.12</td>
<td>-1.05</td>
</tr>
<tr>
<td>2nd</td>
<td>28</td>
<td>0</td>
<td>-0.78</td>
<td>-1.07</td>
<td>-2.01</td>
<td>0.20</td>
<td>0.11</td>
</tr>
<tr>
<td>3rd</td>
<td>29</td>
<td>0</td>
<td>-0.92</td>
<td>-1.55</td>
<td>-1.06</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>4th</td>
<td>27</td>
<td>0</td>
<td>-0.69</td>
<td>0.27</td>
<td>-0.66</td>
<td>-0.12</td>
<td>0.03</td>
</tr>
</tbody>
</table>

As shown in Table 4.7, on the 2nd annual assessment, the subject scored -2.01 for attention exhibiting impairment in this domain. However, this impairment disappeared on the two subsequent annual assessments. Similarly, on the 3rd annual assessment, the subject exhibited impairment in the language domain by scoring -1.55, but this impairment disappeared on the following annual assessment. Therefore, these two instances of impairment are overlooked and the subject is considered cognitively intact for the entire period of four years.
3. If a subject exhibits impairment in one domain on an annual assessment, and this impairment disappears on the following annual assessment, but then reappears on all of the subsequent annual assessments in the same domain or a different domain of the same MCI subtype, then the assessment on which impairment is absent is overlooked and the subject is considered impaired starting from the first annual assessment that shows impairment. An example is shown in Table 4.8.

**Table 4.8: Example of labeling protocol rule 3 using neuropsychological tests**

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>MMSE</th>
<th>CDR</th>
<th>Executive Function</th>
<th>Language</th>
<th>Attention</th>
<th>Memory</th>
<th>Visuospatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>30</td>
<td>0.5</td>
<td>-2.32</td>
<td>-1.52</td>
<td>-1.67</td>
<td>-0.81</td>
<td>-1.47</td>
</tr>
<tr>
<td>1\textsuperscript{st}</td>
<td>30</td>
<td>0</td>
<td>-1.47</td>
<td>-1.31</td>
<td>-1.47</td>
<td>-0.87</td>
<td>-1.48</td>
</tr>
<tr>
<td>2\textsuperscript{nd}</td>
<td>29</td>
<td>0.5</td>
<td>-1.75</td>
<td>-2.45</td>
<td>-1.97</td>
<td>-1.01</td>
<td>-1.19</td>
</tr>
<tr>
<td>3\textsuperscript{rd}</td>
<td>28</td>
<td>0</td>
<td>-1.49</td>
<td>-1.55</td>
<td>-1.52</td>
<td>-1.12</td>
<td>-1.77</td>
</tr>
</tbody>
</table>

As shown in Table 4.8, at baseline, the subject exhibited impairment in more than one domain. Attention was one of them. However, this impairment in the attention domain disappeared on the 1\textsuperscript{st} annual assessment, but then reappeared on the 2\textsuperscript{nd} and the 3\textsuperscript{rd} annual assessments. Therefore, the disappearance of the impairment on the 2\textsuperscript{nd} annual assessment is ignored and the subject is considered cognitively impaired for the entire monitoring period.

4. If a subject undergoes \( n \) annual assessments, and exhibits impairment in one domain on the \((n - 1)\textsuperscript{th}\) annual assessment, which is equivalent to the last annual assessment, with no previous impairment or instance of bouncing, then the subject is considered impaired from the \((n - 1)\textsuperscript{th}\) annual assessment onward although no subsequent assessments are available. An example is shown in Table 4.9.

As shown Table 4.9, the subject exhibited impairment in the visuospatial domain on the 3\textsuperscript{rd} annual assessment, after which no subsequent assessments are available.
Table 4.9: Example of labeling protocol rule 4 using neuropsychological tests

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>MMSE</th>
<th>CDR</th>
<th>Executive Function</th>
<th>Language</th>
<th>Attention</th>
<th>Memory</th>
<th>Visuospatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>27</td>
<td>0.5</td>
<td>-1.09</td>
<td>-0.37</td>
<td>-0.34</td>
<td>-0.59</td>
<td>-0.52</td>
</tr>
<tr>
<td>1st</td>
<td>26</td>
<td>0</td>
<td>-0.73</td>
<td>-0.22</td>
<td>-0.43</td>
<td>-0.82</td>
<td>-0.84</td>
</tr>
<tr>
<td>2nd</td>
<td>28</td>
<td>0.5</td>
<td>-0.59</td>
<td>0.27</td>
<td>-0.22</td>
<td>-1.37</td>
<td>-0.92</td>
</tr>
<tr>
<td>3rd</td>
<td>25</td>
<td>0</td>
<td>-0.45</td>
<td>-0.64</td>
<td>-0.30</td>
<td>-0.90</td>
<td><strong>-2.23</strong></td>
</tr>
</tbody>
</table>

In this case the subject is considered cognitively impaired from this point onward.

Here is an example of how we used the above labeling protocol to assign the labels depicted in Figure 4.1 to the subject whose assessment scores are shown in Table 4.10. According to Table 4.10, the subject exhibited impairment in the language domain at baseline. However, this impairment disappeared on all of the subsequent assessments. Therefore, according to rule 1, this impairment is ignored since it did not appear on two consecutive assessments. On the other hand, the subject exhibited impairment in the attention and memory domains on the 3rd assessment, which is the last available assessment. According to rule 4, the subject is considered impaired in these domains from the 3rd annual assessment onward. Accordingly:

1. The subject data from baseline until the date of the 2nd annual assessment will be
assigned the label “CIN.”

2. The subject data from the date of the 2\textsuperscript{nd} annual assessment up to the date of the 3\textsuperscript{rd} annual assessment will be assigned the label “TR.” This is because the conversion to cognitive impairment is not a point event but a gradual process. Accordingly, the subject’s cognition would be in flux between the 2\textsuperscript{nd} and the 3\textsuperscript{rd} annual assessments and would belong to neither “CIN” nor “MCI.”

3. The subject data from the date of the 3\textsuperscript{rd} annual assessment onward will be labeled “MCI,” in particular “a-MCI multiple domain” because impairment is in the memory domain as well as the attention domain.

Three more examples are discussed in Appendix B.

4.3 Preprocessing & Cleaning Process

Recall that in addition to the walking speed data, the home activity sensor data, and the clinical data, we also received an excel file for each subject that contained their answers to the weekly health questionnaires. Each entry or row represented their answers to a weekly questionnaire. However, the health questionnaires were in a totally unordered fashion as shown in Figure 4.2. The entries were not arranged in chronological order, and the columns were disarrayed which made it very challenging to extract any information from them. Therefore, before any processing of the health questionnaires could be done, we had to order the columns and arrange the entries in chronological order.

However, ordering the columns and sorting the entries chronologically would have been incredibly time-consuming to do manually. Therefore, we wrote a program that read in each excel file and transformed it into an ordered and a sorted version as shown in Figure 4.3. Obviously, the transformed version of the excel files was much easier to use to extract information concerning the subjects’ answers to the queries in the health questionnaires.
Figure 4.2: An example of an excel file containing health questionnaire data with disordered columns.

Figure 4.3: The transformed excel file. Note how the columns are ordered and the entries are sorted.
After completing this preprocessing step, we proceeded to clean the data through three main stages:

1. The first stage consisted of discarding the days on which the subjects had their annual assessments. Since assessments were conducted in-home, then research personnel were present in the home along with the subject. Given that high-level sensing technologies were utilized by ORCATECH, it was not possible to differentiate the research personnel’s activity from the subject’s activity. Therefore, these days were discarded in order to ensure that any subject activity was not confounded with activity pertaining to the research personnel.

2. The second stage consisted of using the weekly health questionnaires to discard days on which subjects had any visitors over, days which subjects spent away from home, days spent in ER, days on which subjects had maintenance people over, or days on which people reported health problems that limited their activity.

3. The third and final stage of cleaning consisted of discarding days on which sensors failed to fire or malfunctioned. Recall that all the sensor firings, that were collected by the transceiver, were timestamped and uploaded to an online database called “console.” Each sensor was assigned a unique item number and an ID corresponding to a room. Figure 4.4 shows a screenshot of console displaying the firings of some sensors for one of the homes. The first sensor on the list is O15 which corresponds to the bedroom according to the homemap, and the number (1010) represents the sensor’s unique item number. This stage was the most challenging of all three stages and consisted of three main steps:

   (a) The first step represented mapping the list of sensors in console to the sensor layout on the homemap for each home.

   (b) The second step represented checking the sensor firings in console for any periods of inactivity as depicted in Figure 4.5. This step was completed in time frames
Figure 4.4: Console showing a subset of sensors and their firings for Home 162.

Figure 4.5: Console showing a period of inactivity.
of ninety days only per home as console displays only a maximum of ninety days of sensor firings at a time.

(c) The final step represented checking the sensor firings in console for any periods of sensor malfunctioning or replacement as shown in Figure 4.6. As shown in

Figure 4.6, the bedroom sensor 95 died at some point and was replaced by the new sensor 1010. However, because the room did not change, the new sensor was still assigned the same room ID, O15. As with the second step, this step was completed in time frames of ninety days for each home.

The clean data were used to answer the research questions listed in Chapter 1 as described in the remaining chapters of this dissertation.
## 4.4 Summary of Clean Data

We received walking speed data for all one hundred fourteen homes and activity sensor data for eighty-five homes. However, according to our labeling protocols, we only included subjects who either remained cognitively intact throughout the monitoring period or transitioned to MCI. Subjects who “bounced” between cognitively intact and experiencing MCI were excluded from our analysis. This “bouncing” was dependent on the ground truth. Therefore, different number of subjects were used in different chapters as shown in Tables 4.11 and 4.12.

<table>
<thead>
<tr>
<th>Type of Sensor Data</th>
<th>Bouncers</th>
<th>Available</th>
<th>Remained Cognitively Intact</th>
<th>Transitioned to MCI</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking Speed</td>
<td>17</td>
<td>97</td>
<td>79</td>
<td>18</td>
<td>10</td>
<td>87</td>
</tr>
<tr>
<td>Activity</td>
<td>17</td>
<td>68</td>
<td>53</td>
<td>15</td>
<td>7</td>
<td>61</td>
</tr>
</tbody>
</table>

According to Table 4.11, using CDR as ground truth, seventeen subjects were identified as “bouncers.” Accordingly, ninety-seven subjects had data available for analysis of walking speed. Of the ninety-seven subjects, seventy-nine subjects remained cognitively intact throughout the monitoring period and eighteen subjects transitioned to MCI. Similarly, sixty-eight subjects had data available for analysis of home activity. Of the sixty-eight subjects, fifty-three subjects remained cognitively intact throughout the monitoring period and fifteen subjects transitioned to MCI.

According to Table 4.12, using neuropsychological assessments as ground truth, five subjects were identified as “bouncers.” Accordingly, one hundred nine subjects had data available for analysis of walking speed. Of the one hundred nine subjects, seventy-
nine subjects remained cognitively intact throughout the monitoring period and thirty subjects transitioned to MCI. Twelve subjects exhibited symptoms of a-MCI whereas eighteen subjects exhibited symptoms of na-MCI. Similarly, eighty subjects had data available for analysis of home activity. Of the eighty subjects, fifty-four subjects remained cognitively intact throughout the monitoring period and twenty-six subjects transitioned to MCI. Eleven subjects exhibited symptoms of a-MCI whereas fifteen subjects exhibited symptoms of na-MCI.

Table 4.12: Summary of subjects using neuropsychological tests as ground truth

<table>
<thead>
<tr>
<th>Type of Sensor Data</th>
<th>Bouncers</th>
<th>Available</th>
<th>Remained Cognitively Intact</th>
<th>Transitioned to MCI</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking Speed</td>
<td>5</td>
<td>109</td>
<td>79</td>
<td>12</td>
<td>94</td>
</tr>
<tr>
<td>Activity</td>
<td>5</td>
<td>80</td>
<td>54</td>
<td>11</td>
<td>70</td>
</tr>
</tbody>
</table>
Chapter 5

Performance Evaluation Metric

People think it is a terrible tragedy when somebody has Alzheimer’s. But in my mother’s case, it is different. My mother has been unhappy all her life. For the first time in her life, she is happy.

Amy Tan

5.1 Introduction

Although the majority of clinical studies use sensitivity and specificity scores as evaluation metrics, in this chapter we justify why these scores can be misleading, especially with an imbalanced database like the one that was utilized in this thesis.

5.2 Case Study

Suppose we have a database of one hundred subjects, eighty of which are cognitively intact and the remaining twenty subjects have MCI. Let’s assume we are able to detect
MCI in this database with equal sensitivity and specificity scores of 0.8. At first glance, one is tempted to believe that the system is performing very well. However, if we analyze these numbers, we will discover that the system is suffering from a false positive rate as high as 50 percent. Recall that,

\[
\text{sensitivity} = \text{recall} = \frac{TP}{TP + FN}, \tag{5.1}
\]

and,

\[
\text{specificity} = \frac{TN}{TN + FP}, \tag{5.2}
\]

where TP stands for true positives, FN stands for false negatives, TN stands for true negatives, and FP stands for false positives. A sensitivity score of 0.8 means that we are able to correctly classify sixteen of the twenty subjects with MCI, i.e., TP = 16. Similarly, a specificity score of 0.8 means that we are able to correctly classify sixty-four of the eighty cognitively intact subjects, i.e., TN = 64, but it also means that FP = 16. In other words, in order to correctly classify sixteen of the twenty MCI subjects, we incorrectly classified as many subjects to be exhibiting symptoms of MCI. However, because the database is imbalanced - having many more cognitively intact subjects than subjects with MCI, this high false positive rate did not significantly impact the sensitivity and specificity scores. On the other hand, if we compute the precision score using the given TP and FP,

\[
\text{precision} = \frac{TP}{TP + FP}, \tag{5.3}
\]

we find that precision = 0.5. The F-score, defined as,

\[
F = 2 \cdot \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \tag{5.4}
\]

is equal to 0.62. Evidently, these scores better reflect the true performance of systems with imbalanced databases. Furthermore, if one is interested in putting more emphasis
on precision than recall, i.e., making sure we penalize the system for false positives, we can use the $F_{0.5}$ score as the evaluation metric, defined as,

$$F_{0.5} = (1 + 0.25) \frac{\text{precision} \times \text{recall}}{(0.25 \times \text{precision} + \text{recall})},$$

which, using the sensitivity and specificity scores given in the example above, is equal to 0.54. Note that a high $F_{0.5}$ score means high $F$, sensitivity, and specificity scores.

### 5.3 Conclusion

In conclusion, although in some chapters we report scores such as the area under the ROC curve, note that the performances of algorithms developed in this thesis are optimized to yield the best area under the precision-recall curve, $F$ score, or $F_{0.5}$ score, depending on which evaluation metric is being reported.
Chapter 6

Predefined Measures of Speed and Activity

You do not just wake up one day with dementia or Alzheimer’s; these conditions are developmental.

6.1 Introduction

As we discussed in the previous chapters, walking speed and variability in walking speed have been found to be good measures to differentiate older adults with mild cognitive impairment and cognitive decline syndromes. Accordingly, in this chapter, we explore the following research questions:

1. Can we use signal processing along with machine learning techniques to autonomously detect older adults with mild cognitive impairment (MCI) using predefined measures associated with their walking speed and general activity in the home, calculated from unobtrusive sensing technologies?
2. What time span of predefined measures results in the highest areas under the ROC curve and the precision-recall curve?

3. How do features extracted from these predefined measures rank in terms of their importance for detecting mild cognitive impairment?

### 6.2 Problem Setup

Suppose that a database consists of $N$ subjects being continuously monitored in their homes using unobtrusive sensing technologies, and for each subject, $p$ weekly measure vectors exist. A weekly measure vector $v$ is basically a vector of values, such as median walking speed and coefficient of variation of walking speed, calculated over a period of one week from the sensing technologies. Given this database of measure vectors, we are interested in autonomously labeling these vectors as belonging to subjects who are cognitively intact or to subjects exhibiting symptoms of mild cognitive impairment (MCI). We are formulating the problem as a classification problem with two classes: ‘cognitively intact’ and ‘experiencing MCI.’ In this chapter, we refer to the ‘experiencing MCI’ class as ‘positive’ class and the ‘cognitively intact’ class as ‘negative’ class, and will be using them interchangeably depending on the context. Mathematically, the cognitive status recognition problem can be formulated as follows: the database consists of $N$ subjects, each having $p$ measure vectors, tabulated as the following sets,

\[
\mathcal{M}_1 = \{v_{1,1}, v_{1,2}, \ldots, v_{1,p} \}, \\
\mathcal{M}_2 = \{v_{2,1}, v_{2,2}, \ldots, v_{2,p_2} \}, \\
\vdots \\
\mathcal{M}_N = \{v_{N,1}, v_{N,2}, \ldots, v_{N,p_N} \}.
\] (6.1)

Each $v_{i,j}$ is an $R$-dimensional vector, where $R$ is the number of measures calculated from the sensor data as shown in Figure 6.1. Note that $R$ is the same for all vectors among
Figure 6.1: Trajectories of weekly measures pertaining to subject 1 in the database. Each asterisk represents a weekly measure. Features are extracted using a window of size $\omega$, that slides one week at a time.

all subjects but the number of weekly measure vectors $p$ can be different among subjects since subjects were monitored for different periods, as mentioned in Section 3.3. Hence, the subscript $i$ in $p_i$ indicates the subject number.

Figure 6.2 depicts the general overview of the proposed approach in recognizing a subject’s cognitive status. Using a sliding window of size $\omega$ (in weeks), the measure vectors are transformed into feature vectors. The resulting feature space can be tabulated as the following sets,

$$
\mathcal{F}_1 = \{ \mathbf{w}_{1,1}, \mathbf{w}_{1,2}, \ldots, \mathbf{w}_{1,q_1} \},
$$

$$
\mathcal{F}_2 = \{ \mathbf{w}_{2,1}, \mathbf{w}_{2,2}, \ldots, \mathbf{w}_{2,q_2} \},
$$

$$
\vdots
$$

$$
\mathcal{F}_N = \{ \mathbf{w}_{N,1}, \mathbf{w}_{N,2}, \ldots, \mathbf{w}_{N,q_N} \}.
$$

Note that for $\omega > 1$ week, the resulting number of feature vectors, $q_i$, will be less than the corresponding number of measure vectors, $p_i$. In this chapter, each feature vector represents a datapoint. Note that the number of feature vectors, or datapoints, for
Figure 6.2: General overview of the cognitive status recognition process.

Each subject is different for different ω’s. Also, note that each \( w_{i,j} \) can have different dimensions depending on the feature type. For example, if the extracted feature type is the average of the individual measures in a window, then each \( w_{i,j} \) would be an \( R \)-dimensional vector. On the other hand, if the extracted feature type is a concatenation of the trajectories of the individual measures in a window, then \( w_{i,j} \) would be an \((R \times ω)\)-dimensional vector.

The feature vectors generated are then used to train and test a machine learning algorithm using the following procedure. The database of feature vectors is first divided into three groups of subjects, each group containing approximately the same number of feature vectors pertaining to each class: positive class and negative class. Since subjects have different numbers of datapoints due to being monitored for different periods, then each group does not necessarily contain the same number of subjects. However, all three groups are created such that each group has approximately the same total number of datapoints pertaining to each class. Then the performance of the machine learning algorithm is evaluated through a 3-fold cross-validation process, that consists of three runs. In each run, two groups are used to train the algorithm and the third group is
used to test it. Accordingly, in each run, the algorithm is tested on datapoints it has not seen in the training phase. Using this methodology, the algorithm eventually yields a prediction for each datapoint. As depicted in Figure 6.2, performance is quantified by generating the ROC curve and the precision-recall curve and calculating the areas under them, $AUC_{SS}$ and $AUC_{PR}$ respectively, where,

$$\text{sensitivity} = \text{recall} = \frac{TP}{TP + FN},$$ \hspace{1cm} (6.3)

$$\text{specificity} = \frac{TN}{TN + FP},$$ \hspace{1cm} (6.4)

and,

$$\text{precision} = \frac{TP}{TP + FP},$$ \hspace{1cm} (6.5)

where TP stands for true positives, FN stands for false negatives, TN stands for true negatives, and FP stands for false positives. Since our ROC curve displays sensitivity versus $(1 - \text{specificity})$, then an algorithm with a good performance yields a point close to the upper left corner of the ROC space, representing high sensitivity and specificity scores. A completely random guess would give a point along the diagonal line from the bottom left corner to the top right corner.

As we discussed in Chapter 5, because the data set we are dealing with is imbalanced–having significantly more instances of the negative class than the positive class, then using sensitivity and specificity scores only could lead to overly optimistic results. Therefore, we also compute precision and recall scores. On a precision-recall curve, a good performing algorithm would yield a point close to the upper right corner, which represents a high average precision–indicating a low false positive rate.
6.3 Data, Measures, and Features

In this chapter, we use CDR as our ground truth and we employ the corresponding labeling protocol to label the subject data. Recall that labeling of data fell into three categories:

1. cognitively intact (“CIN”),
2. transitioning to MCI (“TR”), and
3. having MCI (“MCI”).

6.3.1 Measures and Features

From the sensor data, a number of predefined measures is computed for each week of monitoring. In order to compute these measures, we make the following definitions:

1. Measures are computed for each week, where a week is defined from Monday to Sunday.
2. Weekly walking speed of a subject is computed as the median of all the walking speeds registered within a week.
3. Morning period is defined from 6AM - 3PM, and evening period is defined from 3PM - 12AM.
4. Because we are interested in variability of measures more than increase or decrease of absolute values of measures, difference between two variables is computed as the square difference between the medians of the variables, e.g. difference between \( x \) and \( y \) is computed as

\[
\Delta_{x,y} = (median(x) - median(y))^2. \tag{6.6}
\]
5. A walk is defined as walking under the line of sensors labeled ‘W’ in Figure 3.1, which is used to measure the walking speed.

6. An outing is defined as a firing of any of the exit-door sensors followed by a period of inactivity for at least 15 minutes. Exit-doors include front door, back door, garage door, or any other door from which a subject can exit their living unit.

7. Activity is defined as the total number of sensor firings averaged by the total time spent inside the home.

Based on these definitions, we compute two sets of measures: one associated with in-home walking speed and another one associated with general activity in the home. List 1 below presents the measures associated with walking speed:

1. weekly walking speed ($ws$),
2. coefficient of variation of weekly walking speed ($cv_{ws}$),
3. coefficient of variation of weekly morning walking speed ($cv_{mws}$),
4. coefficient of variation of weekly evening walking speed ($cv_{ews}$),
5. difference between morning and evening speeds ($\Delta_{mws}$), and,
6. coefficient of variation of number of walks ($cv_{w}$).

Accordingly, the number of walking speed measures, $Sm_{ws}$, is equal to 6. List 2 below presents the measures associated with in-home activity:

1. coefficient of variation of weekly number of outings ($cv_{o}$),
2. coefficient of variation of daily activity ($cv_{da}$),
3. coefficient of variation of morning activity ($cv_{ma}$),
4. coefficient of variation of evening activity ($cv_{ea}$), and,
5. difference between morning and evening activities ($\Delta_{mea}$).

Accordingly, the number of activity measures, $Sm_a$, is equal to 5.

Buracchio et al. recently conducted a study that aimed at comparing the trajectory of motor decline exhibited by older adults who developed MCI and those who remained cognitively intact [69]. Different change points – times at which the change in gait or finger-tapping speed accelerates – were reported between men and women. The study also reported significant difference in baseline gait speed between those who transitioned to MCI and those who did not, only in women. So gender seems to be an important clinical measure to include as a feature. Furthermore, it is well-established within the clinical community that age is the most significant known risk factor of dementia. As a result, in addition to the above measures enumerated in Lists 1 and 2, age and gender are included as clinical measures, $Cm$. Accordingly, the total number, $R$, in Eqn. (6.1) is equal to $Sm_{ws} + Sm_a + Cm$, which is equal to 13.

A sliding window of size $\omega$ is used to extract features from the aforementioned sensor measures. Note that age is computed as the mean age in a window of size $\omega$. In this chapter, we experiment with three types of features:

1. Average of measures, where features are extracted by taking the average of the individual measures in the window of size $\omega$. Each $w_{i,j}$ in Eqn. (6.2) is an $R$-dimensional vector.

2. Probability densities of measures, where features are extracted by estimating the probability densities of the individual measures in the window of size $\omega$. By treating each measure as a random variable, we use kernel density estimation to estimate the probability density for each measure using a Normal kernel. These densities are then concatenated into one vector. Accordingly, each $w_{i,j}$ in Eqn. (6.2) is a $(K \times (Sm_{ws} + Sm_a) + C_m)$-dimensional vector, if all measures from Lists 1 and 2 are included. $K$ here indicates the dimension of the probability density, chosen as
3. Trajectories of measures, where features are extracted by concatenating the trajectories of the individual measures, as they appear in each window, into one vector. In addition, with this type of features, we keep track of how the walking speed changes within a window by computing the difference between the current walking speed and the window baseline walking speed ($\Delta_{ws,bws}$), as indicated in Figure 6.1. This results in an additional measure which would increase the total dimension by $\omega - 1$ since the difference between the walking speed at baseline and the baseline walking speed would always be 0. Accordingly, each $w_{i,j}$ in Eqn. (6.2) is an $(((Sm_{ws} + Sm_{a}) \times \omega + \omega - 1 + C_m)-$dimensional vector.

### 6.4 Results and Discussion

Since CDR served as our ground truth, then according to Table 4.11, sixty-eight subjects had data available for analysis of both home activity and walking speed. Among those sixty-eight subjects, seven subjects were males, two of which had MCI at baseline or transitioned to MCI during the monitoring period. The remaining sixty-one subjects were females, thirteen of which had MCI at baseline or transitioned to MCI during the monitoring period. Note that baseline here corresponded to the beginning of the data that we received from ORCATECH, which for some subjects did not necessarily represent the time when they were recruited, since the inclusion criteria required all the subjects to be cognitively intact.

The age densities of the sixty-eight subjects are shown in Figure 6.3. Although the age density of the cognitively intact subjects peaks at approximately 87 years, the age density of the impaired subjects covers a wider range of ages resulting in a substantive overlap between the two groups. Evidently, age alone does not suffice in discriminating older adults with MCI from their cognitively intact counterparts.
Using the clean data of the sixty-eight subjects, the measures in Lists 1 and 2 were calculated for each subject and the sets in Eqn. (6.1) were generated. Given that we formulated the cognitive status recognition problem as a classification problem, only the “CIN” and “MCI” data were used. In this chapter, we experimented with two machine learning algorithms: support vector Machines (SVM) and random forests (RF). For implementation of the machine learning algorithms, we used the SVM library, LIBSVM, developed by Chang and Lin [70], to train and test SVM with a Radial Basis Function (RBF)-kernel. As for RF, we used the TreeBagger class that is part of the Statistics Toolbox in Matlab 2009b.

As a baseline model, we trained and tested SVM and RF using only age and gender as features. The results are summarized in Table 6.1. Both algorithms performed poorly using only age and gender. This result was expected given the substantive overlap in the age densities of the two groups and the gender imbalance of the subjects with a majority of them being females.
Table 6.1: Baseline model using only age and gender as features for the data set of sixty-eight subjects

<table>
<thead>
<tr>
<th></th>
<th>$AUC_{SS}$</th>
<th>$AUC_{PR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>S0.47</td>
<td>0.13</td>
</tr>
<tr>
<td>RF</td>
<td>0.42</td>
<td>0.10</td>
</tr>
</tbody>
</table>

6.4.1 Average of Activity Measures

We started off by analyzing the activity data separately. The first type of features with which we experimented was the average of the individual measures in the sliding window of length $\omega$ weeks. All the measure vectors as well as the feature vectors were of length $R = Sm_a + Cm = 7$. Although a dimension of 7 was not considered a high dimension relative to the size of the data, we anticipated that the dimension of the vectors could pose a problem especially with the next two types of features. Accordingly, we decided to introduce a stage of dimensionality reduction alongside the “training and testing” block in Figure 6.2. Principal component analysis (PCA) was chosen to project the data onto a lower dimensional subspace while preserving 95 percent of the variance in the data. Table 6.2 shows the results obtained from trying to recognize the subjects’ cognitive status with and without PCA using $\omega = 1$ week. Recall that all reported results were optimized for best area under the precision-recall curve, $AUC_{PR}$. So the parameters

Table 6.2: Performance of SVM and RF with and without PCA for $\omega = 1$ week

<table>
<thead>
<tr>
<th></th>
<th>$AUC_{SS}$</th>
<th>$AUC_{PR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.48</td>
<td>0.18</td>
</tr>
<tr>
<td>RF</td>
<td>0.33</td>
<td>0.16</td>
</tr>
<tr>
<td>With PCA</td>
<td>0.45</td>
<td>0.27</td>
</tr>
<tr>
<td>Without PCA</td>
<td>0.45</td>
<td>0.13</td>
</tr>
</tbody>
</table>
associated with SVM, mainly the soft margin parameter, $C$, and the standard deviation of the Gaussian RBF-kernel, $\lambda$, were optimized to yield the best $AUC_{PR}$. Similarly, the parameter associated with RF, mainly the number of trees, was optimized to yield the best $AUC_{PR}$. As Table 6.2 shows, using PCA led to a 5 percent increase in $AUC_{PR}$ with SVM and a 3 percent increase in $AUC_{SS}$. As for RF, using PCA enhanced the algorithm’s performance with a 6 percent increase in $AUC_{SS}$ and a 4 percent increase in $AUC_{PR}$. Accordingly, all results reported, from this point onward, are based on 3-fold cross-validation with each run accompanied by a dimensionality reduction step using PCA.

Table 6.3 provides a summary of the areas under the curves obtained by extracting features as the average of the individual activity measures for $\omega = 1, 2, 3,$ and 4 weeks. We observed that although both algorithms performed slightly better than a random classifier in terms of $AUC_{SS}$, they still performed worse than a random classifier based on $AUC_{PR}$. With $\omega = 4$ weeks, RF had a false positive rate of 80 percent. Also, note that as $\omega$ increased, the number of datapoints associated with both classes decreased. Finally, although $\omega = 4$ weeks yielded the largest $AUC_{PR}$ using SVM, an increase in $\omega$ did not necessarily lead to an increase in the areas under the curves. Intuitively, this was

<table>
<thead>
<tr>
<th>$\omega$ (in weeks)</th>
<th>Feature Vector Length</th>
<th># Negative Datapoints</th>
<th># Positive Datapoints</th>
<th>$AUC_{SS}$</th>
<th>$AUC_{PR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SVM RF</td>
<td>SVM RF</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>3787</td>
<td>482</td>
<td>0.62 0.45</td>
<td>0.33 0.11</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>3168</td>
<td>391</td>
<td>0.56 0.49</td>
<td>0.12 0.14</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>2689</td>
<td>326</td>
<td>0.35 0.53</td>
<td>0.25 0.12</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>2311</td>
<td>278</td>
<td>0.56 0.55</td>
<td>0.39 0.20</td>
</tr>
</tbody>
</table>
expected because by averaging the measures, useful discriminative information could be lost. This gave rise to the second type of features.

### 6.4.2 Probability Densities of Activity Measures

Instead of averaging and potentially losing discriminative information, with the second type of features, we tried to estimate the probability densities of the individual activity measures in the sliding window of size $\omega$. By assuming that each measure was a random variable, we used kernel density estimation to estimate the probability density function that generated the measure samples. The features were represented by the estimated density functions. All density functions were computed at 16 points, resulting in $K = 16$ and feature vectors of dimension $82$ ($5 \times 16 + 2 = 82$) for all $\omega$.

Table 6.4 presents a summary of the areas under the curves obtained using features in the form of the probability densities of the activity measures for $\omega = 4, 8, 12, 16, 20,$ and 24 weeks. The performance of both algorithms in terms of $AUC_{SS}$ was comparable. In

<table>
<thead>
<tr>
<th>$\omega$ (in weeks)</th>
<th>Feature Vector Length</th>
<th># Negative Datapoints</th>
<th># Positive Datapoints</th>
<th>$AUC_{SS}$</th>
<th>$AUC_{PR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SVM RF</td>
<td>SVM RF</td>
</tr>
<tr>
<td>4</td>
<td>82</td>
<td>3274</td>
<td>409</td>
<td>0.64 0.40</td>
<td>0.26 0.10</td>
</tr>
<tr>
<td>8</td>
<td>82</td>
<td>3368</td>
<td>420</td>
<td>0.50 0.49</td>
<td>0.13 0.11</td>
</tr>
<tr>
<td>12</td>
<td>82</td>
<td>3384</td>
<td>410</td>
<td>0.51 0.51</td>
<td>0.23 0.15</td>
</tr>
<tr>
<td>16</td>
<td>82</td>
<td>3314</td>
<td>359</td>
<td>0.50 0.38</td>
<td>0.12 0.08</td>
</tr>
<tr>
<td>20</td>
<td>82</td>
<td>3262</td>
<td>335</td>
<td>0.43 0.50</td>
<td>0.13 0.11</td>
</tr>
<tr>
<td>24</td>
<td>82</td>
<td>3207</td>
<td>311</td>
<td>0.43 0.53</td>
<td>0.13 0.12</td>
</tr>
</tbody>
</table>
terms of $AUC_{PR}$, SVM outperformed RF, especially for smaller $\omega$. However, the algorithms still performed worse than a random classifier since $AUC_{PR}$ for both algorithms was much less than 0.5 for all $\omega$. As with averaging the measures, the performance of the algorithms did not necessarily improve with an increase in $\omega$, and the number of datapoints decreased as $\omega$ increased.

6.4.3 Trajectories of Activity Measures

We also experimented with features in the form of trajectories of the individual activity measures as they appeared in the sliding window. One big challenge with this type of features was that the dimension of the data grew with $\omega$. Another challenge was that the measure vectors had to exist for all weeks in a given window. If a week was missing, then the whole window of data had to be discarded, and because the sensing technologies pertaining to the home activity were very noisy, a large amount of data ended up being discarded. For a given window of size $\omega$, the extracted feature vectors had a dimension of $(5\omega + 2)$. Table 6.5 presents a summary of the results associated with features in the form of trajectories of the individual activity measures for $\omega = 4, 8, 12, 16, 20,$ and 24 weeks. By examining the areas under the curves obtained, we found that SVM outperformed RF for all $\omega$ and yielded higher areas under the curves especially in terms of $AUC_{PR}$. As expected, as $\omega$ increased, the number of datapoints associated with each class decreased. This type of features resulted in the best performance using measures of activity. SVM performed the best, yielding an $AUC_{SS}$ of 0.47 and an $AUC_{PR}$ of 0.33 with $\omega = 24$ weeks. Although SVM performed slightly better a random classifier in terms of $AUC_{SS}$, it still performed worse than a random classifier in terms of $AUC_{PR}$. These consistent results of performing slightly better than a random classifier, at best, implied that the aforementioned activity measures do not include enough discriminative information to discriminate older adults with MCI from their cognitively healthy counterparts.
Table 6.5: Performance of SVM and RF using trajectories of activity measures for $\omega = 4, 8, 12, 16, 20,$ and 24 weeks

<table>
<thead>
<tr>
<th>$\omega$ (in weeks)</th>
<th>Feature Vector Length</th>
<th># Negative Datapoints</th>
<th># Positive Datapoints</th>
<th>$AUC_{SS}$ SVM</th>
<th>$AUC_{SS}$ RF</th>
<th>$AUC_{PR}$ SVM</th>
<th>$AUC_{PR}$ RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>22</td>
<td>2311</td>
<td>278</td>
<td>0.53</td>
<td>0.51</td>
<td>0.34</td>
<td>0.12</td>
</tr>
<tr>
<td>8</td>
<td>42</td>
<td>1342</td>
<td>158</td>
<td>0.50</td>
<td>0.42</td>
<td>0.28</td>
<td>0.16</td>
</tr>
<tr>
<td>12</td>
<td>62</td>
<td>872</td>
<td>113</td>
<td>0.34</td>
<td>0.31</td>
<td>0.21</td>
<td>0.10</td>
</tr>
<tr>
<td>16</td>
<td>82</td>
<td>588</td>
<td>90</td>
<td>0.60</td>
<td>0.47</td>
<td>0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>20</td>
<td>102</td>
<td>417</td>
<td>70</td>
<td>0.48</td>
<td>0.42</td>
<td>0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>24</td>
<td>122</td>
<td>312</td>
<td>51</td>
<td>0.47</td>
<td>0.46</td>
<td>0.39</td>
<td>0.18</td>
</tr>
</tbody>
</table>

6.4.4 Average of Walking Speed Measures

We then analyzed the walking speed measures separately. The first type of features that we explored was the average of the individual measures in the sliding window of size $\omega$ weeks. All the measure vectors as well as the feature vectors were of length $R = Sm_{ws} + Cm = 8$. Table 6.6 provides a summary of the areas under the curves obtained by extracting features as the average of the individual walking speed measures for $\omega = 1, 2, 3,$ and 4 weeks. Both algorithms yielded much better results compared to measures of home activity. RF yielded an $AUC_{SS}$ of 0.62 with $\omega = 4$ weeks which was much better than a random classifier. However, its performance did suffer from a high false positive rate. Similarly, SVM yielded very promising results: $AUC_{SS} = 0.64$ and $AUC_{PR} = 0.45$. As with home activity measures, increasing $\omega$ did not necessarily result in an increase in the areas under the curves since useful discriminative information could have gotten lost by averaging the measures.
Table 6.6: Performance of SVM and RF using average of walking speed measures for $\omega = 1, 2, 3, \text{ and } 4$ weeks

<table>
<thead>
<tr>
<th>$\omega$ (in weeks)</th>
<th>Feature Vector Length</th>
<th># Negative Datapoints</th>
<th># Positive Datapoints</th>
<th>$AUC_{SS}$</th>
<th>$AUC_{PR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>3787</td>
<td>482</td>
<td>0.69</td>
<td>0.55</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>3168</td>
<td>391</td>
<td>0.64</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>2689</td>
<td>326</td>
<td>0.44</td>
<td>0.59</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>2311</td>
<td>278</td>
<td>0.64</td>
<td>0.62</td>
</tr>
</tbody>
</table>

6.4.5 Probability Densities of Walking Speed Measures

Moving to the second type of features, instead of averaging the measures, we used kernel density estimation to estimate the probability density functions that generated the measures’ samples in each sliding window. All density functions were also computed at 16 points, resulting in $K = 16$ for all $\omega$ and feature vectors of dimension $98 (6 \times 16 + 2 = 98)$ for all $\omega$.

Table 6.7 presents a summary of the areas under the curves obtained using features in the form of the probability densities of the measures for $\omega = 4, 8, 12, 16, 20, \text{ and } 24$ weeks. The performance of the algorithms in terms of both $AUC_{SS}$ and $AUC_{PR}$ was comparable. Similar to previous observations with the first type of features, both algorithms still performed worse than a random classifier in terms of $AUC_{PR}$. As with averaging the measures, the performance of the algorithms did not necessarily improve with an increase in $\omega$, and the number of datapoints decreased with $\omega$. 
Table 6.7: Performance of SVM and RF using probability densities of walking speed measures for $\omega = 4, 8, 12, 16, 20,$ and 24 weeks

<table>
<thead>
<tr>
<th>$\omega$ (in weeks)</th>
<th>Feature Vector Length</th>
<th># Negative Datapoints</th>
<th># Positive Datapoints</th>
<th>$AUC_{SS}$ SVM</th>
<th>$AUC_{SS}$ RF</th>
<th>$AUC_{PR}$ SVM</th>
<th>$AUC_{PR}$ RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>98</td>
<td>3274</td>
<td>409</td>
<td>0.69</td>
<td>0.49</td>
<td>0.29</td>
<td>0.18</td>
</tr>
<tr>
<td>8</td>
<td>98</td>
<td>3368</td>
<td>420</td>
<td>0.58</td>
<td>0.53</td>
<td>0.23</td>
<td>0.15</td>
</tr>
<tr>
<td>12</td>
<td>98</td>
<td>3384</td>
<td>410</td>
<td>0.59</td>
<td>0.55</td>
<td>0.30</td>
<td>0.21</td>
</tr>
<tr>
<td>16</td>
<td>98</td>
<td>3314</td>
<td>359</td>
<td>0.54</td>
<td>0.45</td>
<td>0.18</td>
<td>0.12</td>
</tr>
<tr>
<td>20</td>
<td>98</td>
<td>3262</td>
<td>335</td>
<td>0.53</td>
<td>0.53</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>24</td>
<td>98</td>
<td>3207</td>
<td>311</td>
<td>0.51</td>
<td>0.56</td>
<td>0.19</td>
<td>0.17</td>
</tr>
</tbody>
</table>

### 6.4.6 Trajectories of Walking Speed Measures

Finally, we also explored the third type of features in the form of trajectories of the walking speed measures. For a given window of size $\omega$, the extracted feature vectors had a dimension of $(6\omega + \omega - 1 + 2)$. Table 6.8 presents a summary of the results associated with features in the form of trajectories of the individual walking speed measures for $\omega = 4, 8, 12, 16, 20, $ and 24 weeks. By examining the areas under the curves obtained, we found that trajectories of walking speed measures resulted in the best performance, especially using SVM. Similar to the first two types of features, SVM outperformed RF for all $\omega$. SVM yielded the best scores with $\omega = 24$ weeks: $AUC_{SS} = 0.79$ and $AUC_{PR} = 0.53$. For the first time, $AUC_{PR}$ increased monotonically, and the performance of SVM improved with $\omega$. These results were very promising.
Table 6.8: Performance of SVM and RF using trajectories of walking speed measures for \( \omega = 4, 8, 12, 16, 20, \) and 24 weeks

<table>
<thead>
<tr>
<th>( \omega ) (in weeks)</th>
<th>Feature Vector Length</th>
<th># Negative Datapoints</th>
<th># Positive Datapoints</th>
<th>( AUC_{SS} )</th>
<th>( AUC_{PR} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SVM</td>
<td>RF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SVM</td>
<td>RF</td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td>2624</td>
<td>350</td>
<td>0.69</td>
<td>0.71</td>
</tr>
<tr>
<td>8</td>
<td>57</td>
<td>1541</td>
<td>219</td>
<td>0.69</td>
<td>0.68</td>
</tr>
<tr>
<td>12</td>
<td>85</td>
<td>997</td>
<td>160</td>
<td>0.70</td>
<td>0.51</td>
</tr>
<tr>
<td>16</td>
<td>113</td>
<td>673</td>
<td>130</td>
<td>0.67</td>
<td>0.50</td>
</tr>
<tr>
<td>20</td>
<td>141</td>
<td>474</td>
<td>106</td>
<td>0.68</td>
<td>0.53</td>
</tr>
<tr>
<td>24</td>
<td>169</td>
<td>349</td>
<td>83</td>
<td>0.79</td>
<td>0.46</td>
</tr>
</tbody>
</table>

6.4.7 Trajectories of Activity & Walking Speed Measures

After observing that the features in the form of trajectories of the measures resulted in the best performance, especially in the walking speed measures case, we were curious to find out if using trajectories of all 11 (\( Sm_{ws} + Sm_a = 6 + 5 = 11 \)) measures would result in a better performance. So, for a given sliding window of size \( \omega \) weeks, we extracted feature vectors of length \( (11 * \omega + \omega - 1 + 2) \). Table 6.9 presents a summary of the areas under the curves that were obtained for \( \omega = 4, 8, 12, 16, 20, \) and 24 weeks.

The best performance noted was yielded by SVM: \( AUC_{SS} = 0.57 \) and \( AUC_{PR} = 0.43 \) with \( \omega = 24 \) weeks. This performance was worse than the one obtained using trajectories of the walking speed measures only; recall that an \( AUC_{SS} \) of 0.79 and an \( AUC_{PR} \) of 0.53 were obtained with \( \omega = 24 \) weeks using walking speed measures only. Therefore, adding trajectories of activity measures to trajectories of walking speed measures resulted in a worse performance. This result confirmed our earlier observation that the defined measures of activity did not contain enough discriminative information to detect MCI in
Table 6.9: Performance of SVM and RF using trajectories of combined measures of walking speed and home activity for \( \omega = 4, 8, 12, 16, 20, \) and 24 weeks

<table>
<thead>
<tr>
<th>( \omega ) (in weeks)</th>
<th>Feature Vector Length</th>
<th># Negative Datapoints</th>
<th># Positive Datapoints</th>
<th>( AUC_{SS} ) SVM</th>
<th>( AUC_{SS} ) RF</th>
<th>( AUC_{PR} ) SVM</th>
<th>( AUC_{PR} ) RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>49</td>
<td>2311</td>
<td>278</td>
<td>0.58</td>
<td>0.54</td>
<td>0.35</td>
<td>0.16</td>
</tr>
<tr>
<td>8</td>
<td>97</td>
<td>1342</td>
<td>158</td>
<td>0.54</td>
<td>0.46</td>
<td>0.30</td>
<td>0.17</td>
</tr>
<tr>
<td>12</td>
<td>145</td>
<td>872</td>
<td>113</td>
<td>0.40</td>
<td>0.38</td>
<td>0.26</td>
<td>0.12</td>
</tr>
<tr>
<td>16</td>
<td>193</td>
<td>588</td>
<td>90</td>
<td>0.67</td>
<td>0.51</td>
<td>0.28</td>
<td>0.27</td>
</tr>
<tr>
<td>20</td>
<td>241</td>
<td>417</td>
<td>70</td>
<td>0.58</td>
<td>0.50</td>
<td>0.30</td>
<td>0.16</td>
</tr>
<tr>
<td>24</td>
<td>289</td>
<td>312</td>
<td>51</td>
<td>0.57</td>
<td>0.49</td>
<td>0.43</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Another potential reason that could explain the worse performance was the high dimension of the extracted feature vectors. Notice how when combining trajectories of activity and walking speed measures, the dimension almost doubled. This most likely resulted in overfitting of the training data. In other words, with the high dimension, the algorithms were learning the training data very well, and thus failed to generalize their learning to data they had not seen before.

6.4.8 Trajectories of Walking Speed Measures with More Data

After achieving promising results with trajectories of walking speed measures, we were interested in exploring the effect of adding more data on the performance. Subsequently, we studied the effect of adding more data by using the data pertaining to the ninety-seven homes instead of just sixty-eight homes. We repeated the analysis of recognizing the subjects’ cognitive status using walking speed measures only in addition to age and
gender. Figure 6.4 displays the age densities for both classes: ‘cognitively intact’ and ‘experiencing MCI’ for the ninety-seven subjects. Adding more subjects did not result in a considerable change in the age density corresponding to the cognitively intact subjects. On the other hand, the age density pertaining to the subjects with MCI exhibited a negative skew or was skewed towards older ages. However, there was still a significant overlap between the two classes. Similar to our analysis of the data set of sixty-eight subjects, we created a baseline model by training and testing SVM and RF on age and gender only. Table 6.10 shows a summary of the performance of both algorithms. Although the performance of both algorithms improved due to the changes in the age densities, both algorithms still performed poorly, especially in terms of $AUC_{PR}$. This

Table 6.10: Baseline model using only age and gender as features for the data set of ninety-seven subjects

<table>
<thead>
<tr>
<th></th>
<th>$AUC_{SS}$</th>
<th>$AUC_{PR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.57</td>
<td>0.16</td>
</tr>
<tr>
<td>RF</td>
<td>0.54</td>
<td>0.13</td>
</tr>
</tbody>
</table>
supports our observation that age and gender alone do not suffice in discriminating older adults with MCI from their cognitively healthy counterparts.

Table 6.11 shows a summary of the areas under the curves that were obtained using more subjects. A great enhancement was achieved in the performance of RF registering

Table 6.11: Performance of RF and SVM using trajectories of walking speed measures calculated from the data set of ninety-seven homes

<table>
<thead>
<tr>
<th>$\omega$ (in weeks)</th>
<th>Feature Vector Length</th>
<th># Negative Datapoints</th>
<th># Positive Datapoints</th>
<th>$AUC_{SS}$ SVM</th>
<th>$AUC_{PR}$ SVM</th>
<th>$AUC_{SS}$ RF</th>
<th>$AUC_{PR}$ RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>29</td>
<td>5093</td>
<td>593</td>
<td>0.70</td>
<td>0.66</td>
<td>0.37</td>
<td>0.16</td>
</tr>
<tr>
<td>8</td>
<td>57</td>
<td>3206</td>
<td>408</td>
<td>0.67</td>
<td>0.73</td>
<td>0.53</td>
<td>0.27</td>
</tr>
<tr>
<td>12</td>
<td>85</td>
<td>2113</td>
<td>305</td>
<td>0.85</td>
<td>0.79</td>
<td>0.54</td>
<td>0.32</td>
</tr>
<tr>
<td>16</td>
<td>113</td>
<td>1477</td>
<td>245</td>
<td>0.80</td>
<td>0.69</td>
<td>0.57</td>
<td>0.32</td>
</tr>
<tr>
<td>20</td>
<td>141</td>
<td>1072</td>
<td>211</td>
<td>0.79</td>
<td>0.56</td>
<td>0.62</td>
<td>0.24</td>
</tr>
<tr>
<td>24</td>
<td>169</td>
<td>801</td>
<td>183</td>
<td>0.81</td>
<td>0.66</td>
<td>0.71</td>
<td>0.30</td>
</tr>
</tbody>
</table>

an $AUC_{SS}$ of 0.8 for $\omega = 12$ weeks. However, RF still suffered from a high false positive rate, reflected as low $AUC_{PR}$ for all $\omega$. This is most likely because RF requires large amounts of data to perform well since it has been documented that RF is susceptible to overfitting problems.

On the other hand, SVM yielded very good scores: $AUC_{SS}$ of 0.81 and $AUC_{PR}$ of 0.71 for $\omega = 24$ weeks, and $AUC_{PR}$ increased monotonically with $\omega$. Figure 6.5 and Figure 6.6 show the ROC curves and precision-recall curves, respectively, corresponding to SVM with $\omega = 24$ weeks, since $\omega = 24$ weeks resulted in the best performance. Two curves are plotted in each figure. The solid line, shown in Figure 6.5, depicts the ROC curve for the case when we used trajectories of walking speed measures calculated from the data set of sixty-eight homes. The dashed line on the other hand, depicts the ROC curve
Figure 6.5: ROC curves corresponding to using SVM to classify trajectories of walking speed measures for $\omega = 24$ weeks.

Figure 6.6: Precision-recall curves corresponding to using SVM to classify trajectories of walking speed measures for $\omega = 24$ weeks.
for the case when we used trajectories of walking speed measures calculated from the data set of ninety-seven homes. The ROC curve corresponding to the data set of ninety-seven homes is closer to the top left corner of the figure indicating an enhanced performance in terms of sensitivity and specificity scores. Similarly, Figure 6.6 shows the precision-recall curves for both data sets generated by using SVM and \( \omega = 24 \) weeks. Again, the curve corresponding to the data set of ninety-seven homes in Figure 6.6 exhibited a big shift to the top right corner of the figure indicating an improved performance in terms of precision and recall scores.

### 6.4.9 Feature Ranking

After achieving very satisfactory results with SVM and \( \omega = 24 \) weeks, we executed a remove-one-feature process in order to rank the features in terms of their importance for the discrimination process. We repeated the analysis of recognizing the subjects’ cognitive status nine times, where in each time we removed a feature to see how its absence would affect the overall performance of SVM. Table 6.12 presents a summary of the areas under the curves that were obtained by removing one feature at a time. The areas under the curves obtained when all the features were present served as our reference: \( AUC_{SS} = 0.811 \) and \( AUC_{PR} = 0.709 \).

Table 6.12: Ranking features by removing one feature at a time using SVM and \( \omega = 24 \) weeks

<table>
<thead>
<tr>
<th>Feature Removed (in trajectories except for age and gender)</th>
<th>none</th>
<th>gender</th>
<th>age</th>
<th>( \Delta_{w_s,bws} )</th>
<th>( cv_w )</th>
<th>( \Delta_{mws,ews} )</th>
<th>( cv_{ews} )</th>
<th>( cv_{mws} )</th>
<th>( cv_{ws} )</th>
<th>( w_s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( AUC_{SS} )</td>
<td>0.811</td>
<td>0.810</td>
<td>0.781</td>
<td>0.826</td>
<td>0.929</td>
<td>0.812</td>
<td>0.796</td>
<td>0.762</td>
<td>0.751</td>
<td>0.687</td>
</tr>
<tr>
<td>( AUC_{PR} )</td>
<td>0.709</td>
<td>0.599</td>
<td>0.685</td>
<td>0.730</td>
<td>0.779</td>
<td>0.709</td>
<td>0.668</td>
<td>0.656</td>
<td>0.668</td>
<td>0.531</td>
</tr>
</tbody>
</table>

As Table 6.12 shows, removing the gender did not have a significant impact on \( AUC_{SS} \).
but still lead to a reduction of approximately 10 percent in $AUC_{PR}$ indicating an increased false positive rate. Evidently, gender plays an important role in detecting older adults with MCI since it accounts for potential sex-specific physiological differences between male and female subjects. As for age, removing it resulted in a reduction of 3 percent in $AUC_{SS}$ and a reduction of approximately 2 percent in $AUC_{PR}$. The resulting small change was due to the big overlap between the age densities of the two classes, as depicted in Figure 6.4.

Interestingly, removing $\Delta_{ws,bws}$ and $cv_w$ lead to an enhanced performance, with removing $cv_w$ resulting in the best scores. This is most likely because these features were not providing any additional discriminative information. Since we defined SVM with a RBF-kernel, then it would be expected that the algorithm would indirectly learn the difference between the walking speeds and the corresponding windows’ baseline walking speeds. Similarly, $cv_w$ was an indication of the level of activity that the subjects were exhibiting. SVM was most likely able to infer this information from the other features, rendering $\Delta_{ws,bws}$ and $cv_w$ unnecessary. Similarly, removing $\Delta_{mws,ews}$ barely resulted in any noticeable change in performance.

The features that resulted in a deteriorated performance were $ws$, $cv_{ws}$, $cv_{mws}$, and $cv_{ews}$. Removing trajectories of $ws$ and $cv_{mws}$ resulted in greater deterioration in terms of $AUC_{PR}$ compared to removing trajectories of $cv_{ws}$ and $cv_{ews}$. Therefore, variability of walking speed in the morning period was more conducive to detecting older adults with MCI as opposed to variability of the walking speed in the evening period. This is potentially due to the fact that the population of older adults, whether cognitively intact or cognitively impaired, tend to be less active in the evening. Consequently, List 3 below presents the features ranked in descending order of importance:

1. trajectories of weekly walking speed,
2. trajectories of coefficient of variation of morning walking speed,
3. trajectories of coefficient of variation of walking speed,

4. trajectories of coefficient of variation of evening walking speed,

5. age,

6. gender,

7. trajectories of difference between morning and evening walking speeds,

8. trajectories of difference between walking speed and window baseline speed, and finally,

9. trajectories of coefficient of variation of number of walks.

By excluding the features that lead to an improved performance when removed, mainly trajectories of $\Delta_{w,bws}$, $cv_w$, and $\Delta_{mws,ews}$, we repeated the analysis of recognizing the subjects’ cognitive status one final time using SVM with $\omega = 24$ weeks, and we were able to achieve an $AUC_{SS} = 0.97$ and an $AUC_{PR} = 0.93$. The reason for the enhanced performance is that SVM with an RBF kernel is susceptible to the problem of overfitting especially with a leave-one-out cross-validation process. As mentioned earlier, the removed features did not add additional discriminative information and SVM was most likely able to learn the information provided by these features indirectly from the other features. For example, the difference between the window-baseline walking speed and the current walking speed could be learned from the trajectory of the walking speed in the window. Furthermore, note that removing these features resulted in a tremendous reduction in the dimension of the feature vectors from 169 to 98, which is equivalent to a 42 percent reduction in dimension, and with the comparatively small data size, a great enhancement in performance is expected.

The new ROC curve and the new precision-recall curve are shown in Figure 6.7 and Figure 6.8 respectively, and are represented by the lines with x-markers. By comparing
Figure 6.7: ROC curves corresponding to using SVM for three different cases: data set of sixty-eight homes, data set of ninety-seven homes, and data set of ninety-seven homes + best features.

Figure 6.8: Precision-recall curves corresponding to using SVM for three different cases: data set of sixty-eight homes, data set of ninety-seven homes, and data set of ninety-seven homes + best features.
these curves to the curves that were obtained earlier, we observe that the line with x-markers in Figure 6.7 exhibited a substantial jump towards the top left corner indicating a great improvement in the performance of SVM. Similarly, the line with x-markers in Figure 6.8 exhibited a substantial jump towards the top right corner indicating a great improvement in the precision of SVM.

Finally, we were interested in investigating the cause of the dip in the precision-recall curve using the best features with data from the ninety-seven homes, which corresponds to the line with x-markers in Figure 6.8. Recall that the performance of the algorithms was evaluated using a 3-fold cross-validation process. In other words, the data set was divided into three groups, where two groups were used for training and the third one was used for testing. Accordingly, all the ROC curves shown in Figure 6.5 and Figure 6.7 were each the average of three ROC curves: one obtained by testing SVM on group 1 after being trained on groups 2 and 3, another one obtained by testing SVM on group 2 after being trained on groups 1 and 3, and a third one obtained by testing SVM on group 3 after being trained on groups 1 and 2. Similarly, all the precision-recall curves shown in Figure 6.6 and Figure 6.8 were each the average of three precision-recall curves: one obtained by testing SVM on group 1 after being trained on groups 2 and 3, another one obtained by testing SVM on group 2 after being trained on groups 1 and 3, and a third one obtained by testing SVM on group 3 after being trained on groups 1 and 2.

Figure 6.9 shows the individual precision-recall curves obtained by testing SVM on each group. The dotted line represents the precision-recall curve obtained by testing SVM on group 1, the dashed line represents the precision-recall curve obtained by testing SVM on group 2, the line with x-markers represents the precision-recall curve obtained by testing SVM on group 3, and the solid line represents the mean of all three curves. As revealed by Figure 6.9, the dip in the mean precision-recall curve is a result of two misclassified datapoints in group 3. The way these precision-recall curves were estimated was by ranking the datapoints in the group by decreasing scores, starting from rank
Figure 6.9: Individual precision-recall curves corresponding to testing SVM on the three groups using the data set of ninety-seven homes + best features.

1. Then precision and recall scores were computed at different ranks by calculating the percentage of datapoints belonging to the positive class, or the ‘experiencing MCI’ class, among the subset of datapoints. For example, precision(3) was computed as the percentage of datapoints belonging to the positive class among the two datapoints with the largest scores. As shown in Figure 6.9, the two datapoints with the highest scores belonged to the negative class, or the ‘cognitively intact’ class, and therefore, precision was 0 for ranks 2 and 3. This resulted in the dip shown in the mean precision-recall curve when all three precision-recall curves were averaged.

### 6.5 Conclusion

In conclusion, we demonstrated the ability of signal processing along with machine learning algorithms to autonomously detect MCI in older adults. Several measures were calculated from the sensor and clinical data pertaining to ninety-seven homes with single occupants. A sliding time window was then used to generate features to train and test two machine learning algorithms namely SVM and RF. We found that measures associ-
ated with general activity did not contain enough discriminative information to detect MCI in older adults. However, measures associated with in-home walking speed yielded very promising results. Using seven measures associated with in-home walking speed in addition to age and gender, we were able to detect MCI in older adults using unobtrusive sensing technologies with an area under the ROC curve of 0.79 and an area under the precision-recall curve of 0.53. This answered our first research question.

By varying the size of the time window, mainly from 4 weeks to 24 weeks with a step size of 4 weeks, and by using features in the form of trajectories of measures, we observed an increasing trend in the areas under the curves. By adding more data, we were able to enhance the performance of the machine learning algorithms. The best performance was obtained using SVM and a sliding window of size 24 weeks, where an area under the ROC curve of 0.81 and an area under the precision-recall curve of 0.71 were obtained. This answered our second research question.

Finally, by carrying out a remove-one-feature process to determine the most important features for detecting older adults with MCI, we found that trajectories of weekly walking speed, coefficient of variation of the walking speed, coefficient of variation of the morning and evening walking speeds, and the subjects' age and gender were the most important features for the process of detecting MCI in older adults. Running SVM on these top ranking features detected MCI in older adults with an area under the ROC curve of 0.97 and an area under the precision-recall curve of 0.93. This answered our third research question.
Chapter 7

Generalized Linear Models of Home Activity

The ordinary acts we practice everyday at home are of more importance to the soul than their simplicity might suggest.

Thomas Moore

7.1 Introduction

In Chapter 6, we demonstrated that the approach of engineering features and feeding them into a machine learning algorithm to solve a classification problem does not always work. So while this approach of engineering features worked greatly with measures associated with in-home walking speed, it failed with measures associated with general home activity. However, as we described in Section 3.4, Hayes et al. reported preliminary results on potential variability in the day-to-day pattern of activities of older adults with MCI when compared to those who were cognitively intact [25]. Therefore, in this
chapter we seek an alternative approach to discriminate older adults with MCI from their cognitively intact counterparts based on changes in their home activity.

Generalized linear models (GLMs) are commonly used in computational neuroscience to characterize the functional relationship between external sensory stimuli and neural spike trains. GLMs are a generalization of standard linear regression models that can handle any of the predictive distributions belonging to the exponential family such as Gaussian, Bernoulli, Poisson, and others. Although single neuron spiking data could be potentially modeled as a homogeneous Poisson process, where a scalar rate parameter is used to estimate the probability of spiking, several studies have demonstrated that the assumption of homogeneity is unrealistic and the inhomogeneous Poisson process is a superior model [71]. Paninski investigated the definition of the rate parameter as a non-linear warping to a linear weighting of the inputs and the use of maximum likelihood to estimate the parameters of the GLM [72]. Pillow et al. extended the model developed by Paninski and demonstrated that the model of individual neuron spiking activity was significantly improved by including coupling filters that capture dependencies on spiking in other neurons [73]. GLMs have proven to be a useful tool for exploratory data analysis in computational neuroscience, and analysis of the model parameters fit to spiking data has helped develop a stronger understanding of neural spiking behavior. In this chapter, we use an analogous formulation to explore the following research questions:

1. Can we build statistical models of the subjects' general activity in their homes using an approach that facilitates detecting MCI in older adults?

2. What measures of difference can be used to reveal statistical differences between models pertaining to cognitively intact and cognitively impaired older adults?

3. How does this approach compare to the approach of using a predefined set of activity measures that we implemented in Chapter 6?
7.2 Data and Data Labeling

In this chapter, we use CDR as our ground truth and we employ the corresponding labeling protocol to label the subject data. Recall that labeling of data fell into three categories:

1. cognitively intact (“CIN”),
2. transitioning to MCI (“TR”), and,
3. having MCI (“MCI”).

Since CDR served as our ground truth, then according to Table 4.11, sixty-eight subjects had data available for analysis of home activity.

7.3 Problem Setup

Suppose a database consists of $N$ subjects, each subject residing in a living unit with $R$ rooms, and we are interested in estimating the probability of a subject being present in room $r$ within a fixed time interval throughout the day, where $1 \leq r \leq R$. Accordingly, the problem can be well modeled as a Poisson process, because a Poisson process models the number of occurrences of an event in a fixed period of time. In our case, the event is “being present in a room.” By defining a variable, $y_r^{(t)}$, that captures the number of times a subject visits room $r$ during time interval $t$, the probability of the subject being present in room $r$ during time interval $t$ can be given by

$$p(y_r^{(t)}) = \frac{e^{-\lambda} \lambda^{y_r^{(t)}}}{y_r^{(t)}!},$$

(7.1)

where $\lambda$ is the Poisson distribution parameter. However, after analyzing the parameter $\lambda$ for many subjects, it was found that $\lambda$ associated with each room was not constant and varied each day. Furthermore, note that $\lambda$ would most likely vary throughout the day.
too, as the presence in each room is highly time dependent. Therefore, an inhomogeneous Poisson process would serve as a better model.

To estimate the parameter $\lambda$ of the inhomogeneous Poisson process, we propose the following approach. We start by dividing the day into $K$ equal intervals and define a $\lambda_t$ for each interval where $1 \leq t \leq K$. Therefore, a day represented by a matrix $X$ would look like the following:

$$X = [x^{(1)}, x^{(2)}, \ldots, x^{(t)}, \ldots, x^{(K)}], \tag{7.2}$$

where,

$$x^{(t)} = x_j \text{ for } 1 \leq j \leq K \text{ s.t. } \begin{cases} x_j = 1 & t = j \\ x_j = 0 & t \neq j. \end{cases} \tag{7.3}$$

In other words, $X$ takes the form of an identity matrix of size $K \times K$, where each column represents a time interval, and each interval is associated with a $\lambda_t$. The reason why we choose to binarize the day instead of treating it as a continuous variable is, in the case of the continuous variable, the probability of a person being in a room can either increase or decrease linearly throughout the day, which does not reflect a realistic situation. For example, we do not expect the likelihood of a subject to be in a room to only increase throughout the day and reach a maximum at the end of the day. On the other hand, we would expect the likelihood of a subject to be in the kitchen, for example, to be high at lunch time but to be low before and after that. Binarizing the data in the form shown in Eqn. (7.2) and Eqn. (7.3) takes care of this non-linearity in the likelihood throughout the day.

For a subject who was monitored for 900 days for example, the subject’s input space consists of a total number of $z = (900 \times K)$ vectors for each of the $R$ rooms, and each vector has a corresponding label $y_r$ indicating the number of times the subject visits
room \( r \) during the corresponding time interval. The problem can then be formulated as estimating the probability of the subject being present in room \( r \) given a time interval \( x^{(t)} \) as input. To estimate this probability, the \( \lambda_{rv} \)'s corresponding to each interval for room \( r \), denoted by the vector \( \lambda_r \), need to be found. Given a data set of \( z \) time intervals along with the corresponding labels represented by the vector \( y_r = \{ y_r^{(v)} \}_{v=1}^z \), the goal is to find \( \lambda_r \) that maximizes the likelihood function \( L(\lambda_r) \).

\[
L(\lambda_r) = p(\mathbf{y}_r | \mathbf{X}; \lambda_r) = \prod_{v=1}^z \frac{e^{-\lambda_{rv}} \lambda_{rv} y_r^{(v)}}{y_r^{(v)!}}.
\]  

(7.4)

However, maximizing Eqn. (7.4) is equivalent to maximizing its log. Therefore,

\[
\ell(\lambda_r) = \log(L(\lambda_r)) \\
= \log \prod_{v=1}^z \frac{e^{-\lambda_{rv}} \lambda_{rv} y_r^{(v)}}{y_r^{(v)!}} \\
= \sum_{v=1}^z \log \frac{e^{-\lambda_{rv}} \lambda_{rv} y_r^{(v)}}{y_r^{(v)!}} \\
= \sum_{v=1}^z -\lambda_{rv} + y_r^{(v)} \log \lambda_{rv} - \log(y_r^{(v)!}). 
\]

(7.5)

According to the work by Paninski in [72], to estimate \( \lambda_r \) originating from an inhomogeneous Poisson process, \( \lambda_r \) should be defined as a function that is monotonic, grows at least linearly, decays exponentially, and has a derivative. \( \lambda_r \) of the form \( \exp (\mathbf{X}^T \mathbf{w}_r) \) meets these constraints, where \( \mathbf{w}_r \) is a \( K \times 1 \) vector of weights. The problem of estimating \( \lambda_r \) becomes equivalent to estimating the weight vector \( \mathbf{w}_r \). Substituting the definition of \( \lambda_{rv} = e^{\mathbf{x}^{(v)^T} \mathbf{w}_r} \) into Eqn. (7.5) yields the following,

\[
\ell(\mathbf{w}_r) = -\sum_{v=1}^z e^{\mathbf{x}^{(v)^T} \mathbf{w}_r} + \sum_{v=1}^z y_r^{(v)} \mathbf{x}^{(v)^T} \mathbf{w}_r - \sum_{v=1}^z \log(y_r^{(v)!}). 
\]

(7.6)
Maximizing Eqn. (7.6) is equivalent to minimizing its negative, i.e., minimizing,

\[
\ell'(w_r) = -\ell(w_r) = \sum_{v=1}^{z} e^{x^{(v)T}w_r} - \sum_{v=1}^{z} y^{(v)}_r x^{(v)T}w_r + \sum_{v=1}^{z} \log(y^{(v)}_r!),
\]

(7.7)

whose derivative is given by

\[
\frac{d\ell'(w_r)}{w_{rt}} = \sum_{v=1}^{z} x^{(v)T}e^{x^{(v)T}w_r} - \sum_{v=1}^{z} y^{(v)}_r x^{(v)T}e^{x^{(v)T}w_r}.
\]

(7.8)

Once \( w_r \) is found, \( \lambda_r = [\lambda_{r1}, \lambda_{r2}, \ldots, \lambda_{rK}] \) is computed as \( \lambda_{rt} = e^{w_{rt}} \). The room activity generalized linear model (GLM), \( d_r = [d_{r1}, d_{r2}, \ldots, d_{rt}, \ldots, d_{rK}] \), is then estimated as:

\[
d_{rt} = p(y^{(t)}_r = 1|x^{(t)}) = e^{-\lambda_{rt}} for \ 1 \leq t \leq K.
\]

(7.9)

For a subject residing in a living unit with \( R \) rooms, \( R \) GLMs of activity would be estimated resulting in a set of models \( D = [d_1, \ldots, d_r, \ldots, d_R] \). In order to address changes in cognition, we will add the superscript \( c \) to indicate label: \( D^c \) where \( c \in \{\text{CIN, TR, MCI}\} \). So for a subject who remained cognitively intact, all their data would be used to estimate \( D^{\text{CIN}} \). On the other hand, for a subject who transitioned to MCI, all their “CIN” data would be used to estimate \( D^{\text{CIN}} \), all their “TR” data would be used to estimate \( D^{\text{TR}} \), and all their “MCI” data would be used to estimate \( D^{\text{MCI}} \).

### 7.4 Results and Discussion

Of the sixty-eight subjects, seven were males and two transitioned to MCI during the monitoring period. The remaining sixty-one subjects were females, thirteen of which transitioned to MCI during the monitoring period.
7.4.1 Generalized Linear Models of Activity

Given that sensing technologies were deployed into the homes of the participating subjects, in order to standardize the comparison across subjects, we based our analysis on the subjects’ activity in four rooms: the main bedroom, the main bathroom, the kitchen, and the living room. Therefore, we used $R = 4$ in the formulation above.

Initially, we estimated GLMs of activity for a number of subjects in order to determine the best $K$, the number of time intervals per day. Figure 7.1a) shows the estimated $\lambda$ vector for the bedroom for a cognitively intact subject using hourly intervals ($K = 24$) whereas Figure 7.1b) shows the estimated $\lambda$ vector for the bedroom for the same subject using 30-minute intervals ($K = 48$). As Figure 7.1 shows, the $\lambda$ vector estimated using 30-minute intervals appears smoother than the one estimated using hourly intervals.

Intuitively, the shorter the interval length the smoother the estimated $\lambda$ vector but the higher the computational cost. However, the smoothness portrayed by Figure 7.1b) was satisfactory and therefore, the rest of the results reported in this chapter are based on $K$
= 48 intervals.

Figure 7.2 depicts the estimated GLMs of activity pertaining to a female subject, residing in a one-bedroom apartment, and who remained cognitively intact for the entire period of monitoring. These GLMs were created by calculating the probability defined in Eqn. (7.9) for each interval defined in Eqn. (7.2) and Eqn. (7.3) for $K = 48$ using the estimated $\lambda$ vectors. Because the motion sensors utilized in this study were passive infra-red sensors, they do not detect passive activity such as sound sleeping, recuperating, or reading. Accordingly, although we have set out to estimate the probability of a subject being present in a room at different time intervals throughout the day, Figure 7.2 instead shows the probability of the sensing technologies firing in response to the subject’s activity in each room.

Interestingly, using our approach we were able to visualize and elicit patterns of activity specific to each subject, a feature that was very difficult to have using models built from a number of predefined measures. According to Figure 7.2, nighttime activity, especially between 10PM and 6AM, is dominated by the bedroom with frequent trips to the bathroom during the night between 1:30AM and 6AM. The subject is most likely to wake up at 6AM before she proceeds to the kitchen and living room area to prepare breakfast. Occasionally, she seems to wake up before 6AM and rest in the living room for sometime before proceeding to the kitchen to prepare her breakfast. From the kitchen activity GLM, we are able to identify two distinct peaks that most likely correspond to breakfast, between 6AM and 7AM, and to lunch, between 12PM and 1PM. The subject usually has dinner between 5:30PM and 7:30PM. In addition, it seems that the subject frequently visits the kitchen for a late night snack or drink between 9PM and 10PM. Generally, the subject goes to bed shortly after 10PM.

Furthermore, by comparing subjects’ GLMs of activity when cognitively intact to their GLMs of activity as they started transitioning to MCI and when experiencing MCI, we were able to extract important differences among the three sets of GLMs. One main
Figure 7.2: Estimated GLMs of activity pertaining to a female subject residing in a one-bedroom apartment. a) Bedroom GLM. b) Bathroom GLM. c) Kitchen GLM. d) Living room GLM.
observation was that GLMs pertaining to the cognitive intactness stage were smoother than the GLMs pertaining to the transitioning to MCI stage and the experiencing MCI stage. Another important observation was that we were able to extract an important activity-related MCI symptom, namely disturbed sleep patterns from many subjects who transitioned to MCI. In Figure 7.3, for example, we show the bedroom activity GLMs of a subject who transitioned to MCI during the monitoring period.

As stated earlier, because the motion sensors utilized in this study were passive infrared sensors, then deep sound sleeping was not be detected by these sensors despite the presence of the subject in the bedroom. Accordingly, when cognitively intact, as shown in Figure 7.3a), the high probability at 10PM implied that the subject would most likely be in the bedroom at this time to go to sleep. Once the subject had laid down and was engaged in sound sleeping, the probability started decreasing despite the fact that the subject was still in the bedroom. However, the probability did not go down all the way to 0 due to natural movements during sleep such as rolling over from one side to another or the occasional visits to the bathroom or sometimes to the kitchen. As the subject started transitioning to MCI as shown in Figure 7.3b), the probability of the subject being present in the bedroom was high at 11PM, indicating that this was their new time to go to sleep. However, as time passed, the probability did not go down. Instead, it remained high for the whole night which implied that the subject was continuously moving and triggering the motion sensors. A similar pattern is observed in Figure 7.3c) when the subject was exhibiting symptoms of MCI except that the probability was even higher. This change of activity pattern in the bedroom during the night was most likely the result of disturbed sleep patterns which have been found to be associated with MCI [74]. In addition, what was also interesting was that the subject was most likely compensating for their disturbed sleep by napping in the afternoon. This explains the distinct peak between 12PM and 1PM in the transitioning to MCI and the having MCI bedroom activity GLMs.
Figure 7.3: a) Bedroom activity GLM of a subject when cognitively intact. b) Bedroom activity GLM of the same subject when transitioning to MCI. c) Bedroom activity GLM of the same subject when exhibiting symptoms of MCI.
7.4.2 Kullback–Leibler Divergence

In order to quantify the statistical difference between GLMs of the subjects’ activity when cognitively intact, when transitioning to MCI, and when exhibiting symptoms of MCI, we took an information theoretic approach by computing the Kullback–Leibler (KL)-divergence, which is a measure of difference, between the estimated GLMs of activity [75]. For example, the KL-divergence of a GLM of activity pertaining to a subject when in a certain cognitive status, denoted by $d_1$, from the GLM of activity pertaining to the same subject when in a different or the same cognitive status, denoted by $d_b$, is computed as,

$$D_{KL}(d^n_b, d^n_1) = \sum_{t=1}^{K} d^n_b(t) \ln \frac{d^n_b(t)}{d^n_1(t)}, \quad (7.10)$$

where $t$ and $K$, as defined in Eqn. (7.2) and Eqn. (7.3), are the interval number and the number of intervals per day, respectively, and $d^n_b = [d^n_{b1}, \ldots, d^n_{bK}]$ and $d^n_1 = [d^n_{11}, \ldots, d^n_{1K}, \ldots, d^n_{1K}]$ are the probability distributions of the GLMs of activity, estimated as:

$$d^n_{bl} = \frac{d^n_{bl}}{\sum_i^K d^n_{bi}}, \quad (7.11)$$

and,

$$d^n_{1l} = \frac{d^n_{1l}}{\sum_i^K d^n_{1i}}. \quad (7.12)$$

A KL-divergence of 0 implies that the two GLMs of activity are identical. The higher the KL-divergence the greater the difference between the two GLMs. In this work, we used a normalized version of the KL-divergence, defined as,

$$D_{KL}^{norm}(d^n_b, d^n_1) = 1 - \exp (-D_{KL}(d^n_b, d^n_1)). \quad (7.13)$$

This normalization limits the values of the KL-divergence to be between 0 and 1.

For the fifteen subjects who transitioned to MCI during the monitoring period, we
used Eqn. (7.13) to compute the normalized KL-divergences between their room activity GLMs when they were cognitively intact and when they were transitioning to MCI, to compute the normalized KL-divergences between their room activity GLMs when they were cognitively intact and when they were experiencing MCI, and to compute the normalized KL-divergences between their room activity GLMs when they were transitioning to MCI and when they were experiencing MCI.

According to our formulation, cognitive changes occur in the following order: cognitive intactness (“CIN”) → transitioning to MCI (“TR”) → having MCI (“MCI”). Because the KL-divergence is asymmetric, the KL-divergences were computed as the divergence of the GLMs pertaining to a cognitive status from the GLMs pertaining to the preceding cognitive status. For example, the KL-divergences between the “CIN” and the “TR” GLMs were computed as the KL-divergences of the “TR” GLMs from the “CIN” GLMs. Similarly, the KL-divergences between the “CIN” and the “MCI” GLMs were computed as the KL-divergences of the “MCI” GLMs from the “CIN” GLMs. Finally, the KL-divergences between the “TR” and the “MCI” GLMs were computed as the KL-divergences of the “MCI” GLMs from the “TR” GLMs.

The computed normalized KL-divergences were then used to estimate the cumulative distribution function (cdf) of the normalized KL-divergences of the “TR” GLMs from the “CIN” GLMs, to estimate the cdf of the normalized KL-divergences of the “MCI” GLMs from the “CIN” GLMs, and to estimate the cdf of the normalized KL-divergences of the “MCI” GLMs from the “TR” GLMs, for all rooms. The estimated cdfs are shown in Figure 7.4.

An increasing average of normalized KL-divergence was portrayed as we moved from cognitive intactness to having MCI. This supported our hypothesis that GLMs of activity of subjects when having MCI are different from their GLMs of activity when cognitively intact. A very interesting observation was that 50 percent of the subjects who transitioned to MCI had a normalized KL-divergence of their “MCI” GLMs from their “TR” GLMs.
Figure 7.4: Cumulative distribution function of normalized KL-divergences of “TR” GLMs from “CIN” GLMs, “MCI” GLMs from “CIN” GLMs, and “MCI” GLMs from “TR” GLMs for all rooms.

of 0.05 or less. Furthermore, 20 percent of the subjects who transitioned to MCI had a normalized KL-divergence of their “MCI” GLMs from their “TR” GLMs between 0.05 and 0.1, and 30 percent of the subjects who transitioned to MCI had a normalized KL-divergence of their “MCI” GLMs from their “TR” GLMs between 0.1 and 0.5. This meant that the majority of the subjects, when transitioning to MCI, exhibited behaviors and patterns closer to their “MCI” patterns than to their “CIN” patterns.

In order to investigate this finding, we completed a small experiment that is depicted in Figure 7.5. For each of the fifteen subjects who transitioned to MCI, we used their “CIN” data to estimate their “CIN” GLMs for the same aforementioned four rooms. These models served as “reference” models. Then, a sliding time window of size $w$ weeks was defined and used to estimate the room activity GLMs using a subset of the data. Then, the GLMs estimated using the sliding window were compared to the reference model by computing their normalized KL-divergences from the reference model. The result was a trajectory of $n$ normalized KL-divergences, where $n$ was the number of
Figure 7.5: Estimating reference GLM using “CIN” data, and estimating GLMs of activity using a sliding time window of size \( \omega \) weeks.

Slides it took the time window to slide through the entire data for a particular subject. Since subjects were monitored for different periods, \( n \) was different for different subjects. Figure 7.6 depicts the process for one room.

Figure 7.6: a) Room activity reference GLM. b) GLMs estimated using a sliding window of size \( \omega \) weeks as it slid through the “CIN” data, the “TR” data, and the “MCI” data. The result is a trajectory of \( n \) normalized KL-divergences.
Figure 7.6a) shows the room activity reference model that was estimated using all the “CIN” data. Figure 7.6b) shows the GLMs that were estimated using a sliding time window of size $\omega$ weeks as it slid through the “CIN” data, the “TR” data, and the “MCI” data. The hypothesis was that the normalized KL-divergence would start increasing as the time window started sliding over the “TR” and the “MCI” data.

Figure 7.7 shows the trajectories of normalized KL-divergences for a subject who confirmed our hypothesis with respect to all four rooms, for whom $n = 197$. The dotted lines represent the “CIN” normalized KL-divergences, which represent the normalized KL-divergences of the “CIN” sliding window GLMs from the reference GLM. Similarly, the dashed lines represent the “TR” normalized KL-divergences, which represent the normalized KL-divergences of the “TR” sliding window GLMs from the reference GLM. Finally, the solid lines represent the “MCI” normalized KL-divergences, which represent the normalized KL-divergences of the “MCI” sliding window GLMs from the reference GLM. According to Figure 7.7, the “CIN” normalized KL-divergences fluctuated between 0.1 and 0.2. However, as the time window started sliding over the “TR” and the “MCI” data, the normalized KL-divergences increased to reach approximately 0.8 for all four rooms. Not all subjects showed the same pattern for all four rooms, but our hypothesis was confirmed for at least one room for the majority of the subjects.

What was more interesting was that for twelve of the fifteen subjects, the KL-divergences started increasing significantly several months prior to the beginning of the transitioning period. The trajectories in Figure 7.7 are a clear example of this observation. This most likely explains why the majority of the subjects had very small normalized KL-divergences of their “MCI” GLMs from their “TR” GLMs. Yet, these subjects still scored 0 on the CDR scale on the annual assessment that marks the beginning of the transitioning period, indicating that they were still cognitively intact, and it was not until the subsequent year that they were assessed as having MCI by scoring 0.5 on the CDR scale. This delay in detecting MCI in the participating subjects could be very well attributed to the inherent
Figure 7.7: Example of trajectories of normalized KL-divergences of sliding window GLMs from reference GLM for a) bedroom, b) bathroom, c) kitchen, and d) living room.
shortcoming of episodic cognitive assessments and examinations in that they depend on a snapshot observation of cognitive function and assume that observations recorded during the assessment represent the person’s typical state of cognition for relatively long periods of time prior to the assessment. Evidently, this was not the case for the majority of the subjects here neither before the annual assessment nor after. This finding reinforces the need for an alternative approach in which assessment is brought into the daily activity of a person through continuous monitoring in their home environment.

7.4.3 Detecting MCI

After demonstrating significant statistical difference in the activity GLMs of the majority of the subjects when cognitively intact and when exhibiting symptoms of MCI, we were interested in comparing the performance of this approach to the approach we implemented in Chapter 6 using a set of predefined activity measures to detect MCI in older adults. We formulated a classification problem, composed of two classes: ‘cognitively intact’ with the label ‘−1,’ which consisted of the “CIN” data, and ‘experiencing MCI’ with the label ‘1,’ which consisted of the “TR” and the “MCI” data. We repeated a similar process of building trajectories of normalized KL-divergences as shown in Figure 7.8, but this time using data from all sixty-eight subjects. Figure 7.8 also shows that the GLMs of activity estimated using “TR” and “MCI” data were assigned the same label to indicate that they both belonged to the same class: ‘experiencing MCI.’

By thresholding the resulting trajectories of normalized KL-divergences, we generated two curves: the ROC curve by plotting sensitivity versus (1 - specificity) and the precision-recall curve by plotting precision versus recall. A normalized KL-divergence that was below the threshold was classified as ‘cognitively intact,’ and a normalized KL-divergence that was above the threshold was classified as ‘experiencing MCI.’ The area under the ROC curve was denoted by $AUC_{SS}$, and the area under the precision-recall curve was denoted by $AUC_{PR}$. 
Figure 7.8: Estimating reference GLM using “CIN” activity sensor data, and estimating GLMs using a sliding time window of size $\omega$ weeks.

Table 7.1 presents a summary of the areas under the curves obtained using a window of size $\omega = 12$ weeks. Average areas under the ROC curve and the precision-recall curve of 0.716 and 0.706, respectively, were obtained. We were able to identify three main sources of error. The first main source of error was the innate erroneous labeling of the data. The normalized KL-divergences prior to the beginning of the transitioning period
were labeled as belonging to the ‘cognitively intact’ class when they clearly belonged to the ‘experiencing MCI’ class. The second main source of error was, for several subjects who transitioned to MCI during the monitoring period, no increasing “TR” and “MCI” normalized KL-divergences were detected for some of the rooms especially the bathroom. This explains why the areas under the curves corresponding to the bathroom were the smallest. Finally, the third main source of error was, for three of the subjects who transitioned to MCI during the monitoring period, no increasing “TR” and “MCI” normalized KL-divergences were detected for any of the rooms. This could be associated with the subjects most likely experiencing amnestic mild cognitive impairment only, where it was only the memory skills of the subjects that were compromised while their thinking and executive functioning skills were still intact. This source of error is a consequence of the inherent shortcoming of the CDR assessment as its definition of MCI lacks a distinction between the two MCI subtypes: amnestic mild cognitive impairment (a-MCI) and non-amnestic mild cognitive impairment (na-MCI).

Despite these sources of error, our approach significantly outperformed the best performance reported using a set of predefined activity measures. The best scores reported in Chapter 6 using activity measures were obtained using trajectories of the activity measures. $AUC_{SS} = 0.47$ and $AUC_{PR} = 0.39$ were obtained for $\omega = 24$ weeks. In other words, with only half the size of the window frame, we were able to outperform the best performance reported in Chapter 6. Overall, using generalized linear models of home activity, we were able to increase the area under the ROC curve by 37 percent and the area under the precision-recall curve by 49 percent for $\omega = 12$ weeks when compared to the performance obtained in the Chapter 6 using activity measures only.

### 7.5 Conclusion

In conclusion, we proposed and presented an alternative approach to that of existing studies on smart systems developed to monitor cognitive decline. Using the sensor and
clinical data pertaining to sixty-eight subjects, fifteen of which transitioned to MCI during the monitoring period, we demonstrated that older adults’ home activity could be well modeled as independent inhomogeneous Poisson processes. Using this approach, we were able to visualize and elicit patterns of activity specific to each subject, a feature that is very difficult to have using other works where a number of predefined measures is used. This answered our first research question.

In order to quantify any statistical differences between GLMs of activity pertaining to older adults when cognitively intact and when transitioning to MCI or when experiencing MCI, we took an information theoretic approach and used the KL-divergence to demonstrate statistical differences between the activity GLMs corresponding to the different states of cognition. This answered our second research question.

Finally, by using a simple thresholding approach of trajectories of normalized KL-divergences, we were able to detect mild cognitive impairment in older adults with average areas under the ROC curve and the precision-recall curve of 0.716 and 0.706, respectively, using GLMs estimated with a sliding time window of 12 weeks. Our proposed approach outperformed the approach that we implemented in Chapter 6, where a number of predefined measures associated with older adults’ general activity was used to detect MCI. This answered our third research question.
Chapter 8

Clustering of Home Activity GLMs

Alzheimer’s is the cleverest thief, because she not only steals from you, but she steals the very thing you need to remember what has been stolen.

Jarod Kintz

8.1 Introduction

In the Chapter 7, we built generalized linear models (GLMs) of older adults’ home activity, and used the KL-divergence to quantitatively demonstrate significant statistical differences between activity GLMs of older adults when cognitively intact and when exhibiting symptoms of MCI. Building on Chapter 7, we explore the following research questions:

1. Can we use GLMs of home activity to automatically detect MCI in older adults using unobtrusive sensing technologies?

2. What size of time frame results in the highest $F_{0.5}$ score when detecting MCI in older adults?
3. How do GLMs of home activity of older adults when cognitively intact compare with their activity GLMs when exhibiting symptoms of amnestic MCI (a-MCI) or non-amnestic MCI (na-MCI)?

8.2 Data and Data Labeling

In this chapter, we use neuropsychological assessments as our ground truth and we employ the corresponding labeling protocol to label the subject data. Recall that labeling of data fell into three categories:

1. cognitively intact ("CIN"),

2. transitioning to MCI ("TR"), and,

3. having MCI ("MCI").

Since neuropsychological tests served as our ground truth, then according to Table 4.12, eighty subjects had data available for analysis of home activity.

8.3 Problem Setup

Suppose that a database consists of $N$ subjects, each residing in a living unit of $R$ rooms, who were monitored in their homes for a few years using unobtrusive sensing technologies. For each subject, a sliding time window of size $\omega$ weeks is used to build a data set of room activity GLMs. These activity models estimate the probability of the subject being present in room $r$ during a time interval $t$ of the day, where $r = 1, \ldots, R$ and $1 \leq t \leq K$, where $K$ is the number of intervals per day as defined in Section 7.3. Data sets from all subjects are used to create a database of room activity GLMs. Given this database, we are interested in autonomously classifying subjects as belonging to subjects who are cognitively intact or having MCI. In other words, we are formulating the problem as a
classification problem with two classes: ‘cognitively intact’ corresponding to the “CIN” data, and ‘experiencing MCI’ corresponding to the “TR” and “MCI” data.

Mathematically, the cognitive status recognition problem can be formulated as follows: the database consists of \( N \) subjects, each residing in a living unit of \( R \) rooms, tabulated as the following sets,

\[
\mathcal{D}_1 = \{D_{1,1}, D_{1,2}, \ldots, D_{1,R}\}, \\
\mathcal{D}_2 = \{D_{2,1}, D_{2,2}, \ldots, D_{2,R}\}, \\
\vdots \\
\mathcal{D}_N = \{D_{N,1}, D_{N,2}, \ldots, D_{N,R}\}.
\]

Each \( D_{n,r} \) is a \( s_n \times K \) matrix, where each row represents an estimated activity GLM of room \( r \) pertaining to subject \( n \), and \( s_n \) is the total number of GLMs estimated for each room for subject \( n \). Each of the estimated GLMs represents a datapoint. Note that \( s_n \) is different across subjects because subjects could have been monitored for different periods of time. Therefore, subjects do not necessarily have the same number of activity GLMs for all rooms. Hence, the subscript \( n \).

The database of room activity GLMs in Eqn. (8.1) is used to extract GLMs, referred to as “exemplars,” that best represent the database. These exemplars are extracted using an off-line procedure. This constitutes the training stage. In a test stage, for a subject with \( R \) room activity GLMs, the objective of the cognitive status recognition system is to assign a label to each room activity GLM and then use these labels to determine the cognitive status of the subject. Figure 8.1 depicts the general overview of the proposed approach in assigning labels to room activity GLMs corresponding to room \( r \). Note that the block diagram represents a two-stage system. Figure 8.1(a) represents the first stage being the training stage whereas Figure 8.1(b) represents the second stage being the test stage.

The training stage comprises two parts. For each subject, a sliding time window of
Clustering raw sensor data for room $r$ for all subjects

sliding window of size $\omega$ weeks

room activity GLM estimator

database of room activity GLMs
$
\{D_{1,r}, D_{2,r}, \ldots, D_{N,r}\}$

Clustering

Affinity Propagation

relabeling using $k$-means

exemplars for room $r$
$
\{e_1, e_2, \ldots, e_m\}$

(a)

(b)

raw sensor data for room $r$
over $\omega$ weeks for test subject

room activity GLM estimator

d$_{test}$

KL-divergence

label of $d_{test}$ =

label of exemplar with smallest KL-divergence

Figure 8.1: General overview of the cognitive status recognition process using activity GLMs corresponding to room $r$. (a) Training stage. (b) Test stage.
size $ω$ weeks is used to divide the sensor data for room $r$ into frames, each of size $ω$ weeks. Then for each frame, the “room activity GLM estimator” block estimates the corresponding room activity GLM. This results in a data set of room activity GLMs for each room per subject. A training set for room $r$ is formed by combining all data sets of room activity GLMs from all subjects. This training set is then processed through a clustering stage that consists of two sub-blocks. The first clustering sub-block performs high level clustering using $k$-means to deal with possible mislabeling of the activity models. Since subjects were assessed annually, then it was not possible to determine when exactly the transition to MCI occurred. Accordingly, as per our labeling protocol, some GLMs could potentially be mislabeled. For example, some GLMs could be labeled “CIN” when they should actually be labeled “TR” or the other way around. The relabeled GLMs are then fed into the affinity propagation (AP) sub-block to decompose the training set into multiple clusters. Clustering in essence represents the core of the training stage. Each cluster is represented by one of its members, referred to as the “exemplar.” The output of the clustering stage, and in turn the training stage, is a set of exemplars $E$, where each exemplar represents a cluster of room activity GLMs, and each cluster contains GLMs belonging to one state of cognition only—either “CIN,” “TR,” or “MCI.”

As for the test stage, a room activity GLM coming from a subject with unknown cognitive status is compared to the set of exemplars by computing the normalized KL-divergence of the test GLM from the set of exemplars. The test GLM is then assigned the label of the exemplar that yields the smallest normalized KL-divergence. The cognitive status of the test subject based on room $r$ activity GLM is then determined as the class to which the label of the activity GLM belongs, either ‘cognitively intact’ or ‘experiencing MCI.’ Note that the overview portrayed by the block-diagram in Figure 8.1 corresponds to one room only. Therefore, the training described above should be repeated $R$ times. The details of each sub-block are discussed in the following subsections.
8.3.1 Room Activity GLM Estimator

The first step in the cognitive status recognition process is creating a training set, $D_r$, for each of the $R$ rooms using the data sets of room activity GLMs built for each subject. This is equivalent to taking a vertical slice of the database defined in Eqn. (8.1). In other words,

$$D_r = \begin{bmatrix} D_{1,r} \\ D_{2,r} \\ \vdots \\ D_{N,r} \end{bmatrix}.$$  \hfill (8.2)

The process of building a data set of room activity GLMs for room $r$ for subject $n$ using a sliding window of size $\omega$ weeks is summarized in Figure 8.2. Each frame of the sensor data, extracted using a sliding window of size $\omega$ weeks, is used to build a $K \times 1$ GLM denoted by $d^{(c)}$, where $c \in \{-1, 1\}$ with ‘−1’ representing the ‘cognitively intact’ class, and ‘1’ representing the ‘experiencing MCI’ class. The estimated room activity GLM provides the probability of the subject being present in room $r$ during some time interval during the day. The process of building these GLMs has been discussed in details in Figure 8.2: The process of building a data set of room activity GLMs for room $r$ for subject $n$. 

![Diagram](image-url)
Section 7.3.

The resulting data set, $D_{n,r}$, of activity GLMs for room $r$ pertaining to subject $n$ is created as,

$$
D_{n,r} = \left[ \begin{array}{c}
\leftarrow d_{n,1}^{(-1)} \rightarrow \\
\leftarrow d_{n,2}^{(-1)} \rightarrow \\
\vdots \\
\leftarrow d_{n,i-4}^{(-1)} \rightarrow \\
\leftarrow d_{n,i-3}^{(-1)} \rightarrow \\
\leftarrow d_{n,i-2}^{(-1)} \rightarrow \\
\leftarrow d_{n,i-1}^{(-1)} \rightarrow \\
\leftarrow d_{n,i}^{(1)} \rightarrow \\
\leftarrow d_{n,i+1}^{(1)} \rightarrow \\
\vdots \\
\leftarrow d_{n,s_n}^{(1)} \rightarrow 
\end{array} \right].
$$

(8.3)

A subject who transitioned to MCI during the monitoring period would have $(i-1)$ GLMs belonging to the ‘cognitively intact’ class and $(s_n-i+1)$ GLMs belonging to the ‘experiencing MCI’ class. A subject who did not transition to MCI would have all $s_n$ GLMs belonging to the ‘cognitively intact’ class.

8.3.2 Relabeling Using $k$-means

Because subjects were assessed annually for MCI, it is not possible to determine at what point exactly the transition took place. Consequently, for subjects who transitioned to MCI, some activity GLMs could potentially be mislabeled. This was demonstrated in Chapter 7, where we found out that the majority of the subjects started transitioning to MCI several months prior to the beginning of the “TR” period. Therefore, several GLMs labeled “CIN” should be actually labeled “TR.” As for subjects who started transitioning
after the beginning of the transitioning period, some of the activity GLMs labeled “TR” should most likely be labeled “CIN.”

In order to address this issue, we add a preprocessing step of high-level clustering using $k$-means. We are not much concerned about the activity GLMs bordering the “TR” period and the “MCI” period since we are formulating the problem as a classification problem and both of these periods are lumped into one class, the ‘experiencing MCI’ class. Therefore, for each subject who transitioned to MCI, the room activity GLMs for each room are clustered using $k$-means into two clusters: one cluster for the ‘cognitively intact’ class, and another cluster for the ‘experiencing MCI’ class. As a result, subjects who started transitioning prior to the beginning of the “TR” period would have some of their “CIN” GLMs assigned the label ‘1’ instead of ‘$-1$,’ indicating that they belong to the ‘experiencing MCI’ class. This change is reflected in the data sets of room activity GLMs for each subject. For subjects who did not transition, no change occurs in their data sets of activity GLMs. For subjects who transitioned prior to the beginning of the “TR” period, possible changes are shown in Eqn. (8.4).

By comparing Eqn. (8.4) to Eqn. (8.3), we see that GLMs bordering the “CIN” and the “TR” periods such as $(i - 3), (i - 2),$ and $(i - 1),$ which were previously labeled ‘$-1$,’ are now labeled ‘1,’ meaning that they belong to the ‘experiencing MCI’ class. Similarly, for subjects who started transitioning after the beginning of the “TR” period, in comparison with Eqn. (8.3), GLMs such as $(i), (i + 1),$ and $(i + 2),$ would be labeled
‘−1’ instead of ‘1,’ meaning that they belong to the ‘cognitively intact’ class instead.

\[ D_{n,r} = \begin{bmatrix}
- d_{n,1}^{(-1)} & \rightarrow \\
- d_{n,2}^{(-1)} & \rightarrow \\
\vdots & \\
- d_{n,i-4}^{(-1)} & \rightarrow \\
- d_{n,i-3}^{(1)} & \rightarrow \\
- d_{n,i-2}^{(1)} & \rightarrow \\
- d_{n,i-1}^{(1)} & \rightarrow \\
- d_{n,i}^{(1)} & \rightarrow \\
- d_{n,i+1}^{(1)} & \rightarrow \\
\vdots & \\
- d_{n,s_n}^{(1)} & \rightarrow \\
\end{bmatrix} \quad (8.4) \]

### 8.3.3 Kullback–Leibler Divergence

As in Chapter 7, we use the Kullback–Leibler (KL)-divergence to quantify statistical differences between the GLMs of the subjects’ activity when cognitively intact, when transitioning to MCI, and when exhibiting symptoms of MCI. One useful property of the KL-divergence is asymmetry. So \( D_{KL}(d_1^n, d_2^n) \neq D_{KL}(d_2^n, d_1^n) \). Accordingly, we are able to quantify the divergence of activity GLMs pertaining to a cognitive state from the GLMs pertaining to the preceding cognitive state. For example, we can quantify the divergences of the “TR” GLMs from the “CIN” GLMs, the divergence of the “MCI” GLMs from the “CIN” GLMs, and the divergence of the “MCI” GLMs from the “TR” GLMs.
8.3.4 Affinity Propagation

The next sub-block in the clustering step is using affinity propagation (AP) to divide the training set of GLMs into different clusters and to appoint a representative to each cluster, referred to as the “exemplar” [58]. AP is used as the clustering algorithm to accomplish this objective. As discussed in Section 2.4.1, one of the input arguments that AP takes is a set of real numbers, known as self-similarity or preference ($p$) for each activity GLM, so that GLMs with larger values of $p$ are more likely to be chosen as exemplars. For the proposed cognitive status recognition system, the self-similarity $p$ is proportional to the median of the input similarities, that is,

$$p = \beta \ast \text{median}\{D_{KL}(d^n_i, d^n_j)\}, \; \forall \; i, j \in \{1, 2, ..., M\},$$

(8.5)

where $M$ is the total number of activity GLMs in the training set, and $\beta$ is a constant that controls the number of clusters to be generated in an inversely proportional manner. In other words, as the value of $\beta$ decreases, more clusters will be generated.

AP is chosen as the clustering technique because it handles better the fact that activity GLMs belonging to the same class should not be forced to have only one exemplar or center as is the case with $k$-means for example. Ideally, it would be perfect if we are able to generate only one cluster for the ‘cognitively intact’ class and another cluster for the ‘experiencing MCI’ class. However, practically this is not the case given that different individuals are more likely to exhibit different patterns of daily activity. Accordingly, with AP, we are able to divide the training set of activity GLMs into different clusters, each represented by an exemplar while ensuring that members of each cluster belong to the same class.

The output of AP is a matrix of exemplars, $E_r$, represented as,

$$E_r = \{e_1, e_2, \ldots, e_j, \ldots, e_{m_r}\},$$

(8.6)
along with a vector of exemplar labels, $c_r$, represented as,

$$c_r = \{c_1, c_2, \ldots, c_j, \ldots, c_{m_r}\},$$

(8.7)

for each room $r$. Each column, $e_j$, of the matrix $E_r$ represents an exemplar, and each label, $c_j$, in $c_r$ indicates whether the corresponding exemplar represents a cluster of ‘cognitively intact’ GLMs or a cluster of ‘experiencing MCI’ GLMs. So $c_j \in \{-1, 1\}$ with ‘$-1$’ corresponding to the ‘cognitively intact’ class and ‘$1$’ corresponding to the ‘experiencing MCI’ class, where $1 \leq j \leq m_r$. Note that the number of exemplars $m_r$ is room dependent because different subjects can potentially exhibit more variability in their activity in some rooms than other rooms. For example, subjects, or older adults in general, normally have a high probability of being in the bedroom early in the morning and late at night. You would expect to see more variability in other rooms such as the living room where older adults spend more time and thus potentially exhibit more variability. Therefore, it would be intuitive to expect fewer exemplars to represent GLMs of bedroom activity than GLMs of living room activity. Also, note that clustering is done per room. Therefore, clustering should be done $R$ times, one for each room. A summary of our proposed cognitive status recognition methodology is presented in Algorithm 1.

### 8.4 Results and Discussion

Recall that the cognitive status of the eighty subjects based on annual neuropsychological assessments during the monitoring period were as follows:

1. fifty-four subjects remained cognitively intact throughout the monitoring period,

2. eleven subjects had a-MCI at baseline or transitioned to a-MCI during the monitoring period, and,

3. fifteen subjects had na-MCI at baseline or transitioned to na-MCI during the mon-
Algorithm 1 Cognitive status recognition algorithm using clustering of room activity GLMs

Input: - Sets of estimated room activity GLMs \( \{D_1, D_2, \ldots, D_N\} \) for \( R \) rooms for all subjects

Output: - Cognitive status of an older adult: ‘cognitively intact’ or ‘experiencing MCI’

Training Stage:
1: for room \( r = 1, \ldots, R \) do
2: For subjects who transitioned to MCI, relabel GLMs using \( k \)-means
3: Run affinity propagation to extract a set of exemplars, \( E_r \), and vector of exemplar labels, \( c_r \)
4: end for
- Output of Training Stage: \( \{E_1, c_1, E_2, c_2, \ldots, E_R, c_R\} \)

Classification Stage:
For an older adult with room activity GLMs \( \{d_{1}^{test}, d_{2}^{test}, \ldots, d_{R}^{test}\} \):
1: for each \( d_{r}^{test} \) do
2: compute normalized KL-divergences \( \{D_{KL_{1}}, D_{KL_{2}}, \ldots, D_{KL_{mr}}\} \) of \( d_{r}^{test} \) from exemplars in \( E_r \)
3: \( c_{d_{r}^{test}} = \) label of exemplar corresponding to \( \min \{D_{KL_{1}}, D_{KL_{2}}, \ldots, D_{KL_{mr}}\} \)
4: end for
- Output of Classification Stage: \( \{c_{d_{1}^{test}}, c_{d_{2}^{test}}, \ldots, c_{d_{R}^{test}}\} \)

- Return cognitive status of older adult as the class that corresponds to the majority of \( \{c_{d_{1}^{test}}, c_{d_{2}^{test}}, \ldots, c_{d_{R}^{test}}\} \)
Baseline here corresponded to the beginning of the data that we received from ORCATECH, which for some subjects did not necessarily represent the actual date when subjects were recruited, because inclusion criteria required all subjects to be cognitively intact. In this chapter, we continue to focus our analysis on four rooms: main bedroom, main bathroom, kitchen, and living room. In other words, $R$ in Eqn. (8.1), is equal to 4.

8.4.1 80 percent Training Data and 20 percent Test Data

Since not all the subjects transitioned to MCI during the monitoring period, it was not possible to test our approach using a leave-one-subject-out cross-validation process by including all 80 subjects. Alternatively, we randomly divided our database of subjects into a training set and a test set of subjects, such that the training set contained 80 percent of the GLMs belonging to each cognitive state—“CIN,” “TR,” and “MCI,” and each MCI subtype—“a-MCI,” and “na-MCI,” and the test set contained 20 percent of the GLMs belonging to each cognitive state and MCI subtype. Note that the creation of these sets was based on full subject data in order to ensure that no activity GLMs pertaining to the same subject existed in both the training and the test sets.

We experimented with different time frames ranging from 4 - 24 weeks as well as different slide sizes ranging from 2 - 4 weeks. We empirically found that a value of $\beta$, defined in Eqn. (8.5), of 0.5 resulted in the best performance. The results obtained using a time frame of 4 weeks, i.e., $\omega = 4$ weeks, and slide sizes of 2 weeks, 3 weeks, and 4 weeks are summarized in Table 8.1. The ‘Bedroom,’ ‘Bathroom,’ ‘Kitchen,’ and ‘Living Room’ $F_{0.5}$ scores reported in Table 8.1 corresponded to detecting MCI using individual room activity GLMs separately. The ‘Combined’ $F_{0.5}$ scores corresponded to detecting MCI using majority voting among the four rooms. At least three rooms should have declared the subject as having MCI for the subject to belong to the ‘experiencing MCI’ class. If a tie existed, where two rooms declared the subject as exhibiting symptoms of
Table 8.1: $F_{0.5}$ scores with $\omega = 4$ weeks and slide sizes of 2 weeks, 3 weeks, and 4 weeks

<table>
<thead>
<tr>
<th>Slide Size (in weeks)</th>
<th>Bedroom</th>
<th>Bathroom</th>
<th>Kitchen</th>
<th>Living Room</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.625</td>
<td>0.573</td>
<td>0.615</td>
<td>0.439</td>
<td>0.760</td>
</tr>
<tr>
<td>3</td>
<td>0.696</td>
<td>0.567</td>
<td>0.477</td>
<td>0.570</td>
<td>0.762</td>
</tr>
<tr>
<td>4</td>
<td>0.683</td>
<td>0.596</td>
<td>0.441</td>
<td>0.525</td>
<td>0.762</td>
</tr>
</tbody>
</table>

MCI and the other two rooms declared the subject as still cognitively intact, the subject was classified as belonging to the ‘experiencing MCI’ class.

According to Table 8.1, we were able to detect MCI in older adults with an $F_{0.5}$ score of 0.625 based on the bedroom activity GLMs only, an $F_{0.5}$ score of 0.573 based on the bathroom activity GLMs only, an $F_{0.5}$ score of 0.615 based on the kitchen activity GLMs only, an $F_{0.5}$ score of 0.439 based on the living room activity GLMs only, and an $F_{0.5}$ score of 0.760 based on majority voting among the four rooms, as described in Algorithm 1, all using a slide size of 2 weeks. Interestingly, using majority voting among the four rooms, a 13.5 percent improvement in performance was achieved compared to the best performance achieved using individual room activity GLMs, which was using the bedroom in this case. As the slide size increased, we continued to observe that majority voting among the four rooms resulted in an improvement of approximately 6 percent compared to using the bedroom activity GLMs only. Another interesting observation is that, using individual room activity GLMs only, although higher $F_{0.5}$ scores were achieved as the slide size increased, we did not note a significant change in the $F_{0.5}$ scores using majority voting among the four rooms.

Similarly, Table 8.2 presents the $F_{0.5}$ scores using a time frame of 8 weeks and slide sizes of 2 weeks, 3 weeks, and 4 weeks. Interestingly, using the individual room activity GLMs we were able to achieve a performance that was comparable to the majority voting
Table 8.2: $F_{0.5}$ scores with $\omega = 8$ weeks and slide sizes of 2 weeks, 3 weeks, and 4 weeks

<table>
<thead>
<tr>
<th>Slide Size (in weeks)</th>
<th>$F_{0.5}$ Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bedroom</td>
</tr>
<tr>
<td>2</td>
<td>0.799</td>
</tr>
<tr>
<td>3</td>
<td>0.794</td>
</tr>
<tr>
<td>4</td>
<td>0.726</td>
</tr>
</tbody>
</table>

performance obtained using a time frame of 4 weeks. This was true for the bedroom, bathroom, and living room activity GLMs but not the kitchen. Despite the low $F_{0.5}$ scores corresponding to using the kitchen activity GLMs, with majority voting, a 2 percent improvement in performance was recorded compared to the time frame of 4 weeks. According to Table 8.2, for a time frame of 8 weeks, we found that a slide size of 2 weeks resulted in the best performance.

As the window size continued increasing, better performance and higher $F_{0.5}$ scores were noted until a time frame of 20 weeks, after which the performance started dropping, as shown in Table 8.3 and Table 8.4. The best performance was achieved using a time frame of 20 weeks and a slide size of 4 weeks as shown in Table 8.3. An $F_{0.5}$ score of 0.815
Table 8.4: \(F_{0.5}\) scores with \(\omega = 24\) weeks and slide sizes of 2 weeks, 3 weeks, and 4 weeks

<table>
<thead>
<tr>
<th>Slide Size (in weeks)</th>
<th>Bedroom</th>
<th>Bathroom</th>
<th>Kitchen</th>
<th>Living Room</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.812</td>
<td>0.788</td>
<td>0.490</td>
<td>0.747</td>
<td>0.780</td>
</tr>
<tr>
<td>3</td>
<td>0.813</td>
<td>0.758</td>
<td>0.492</td>
<td>0.744</td>
<td>0.783</td>
</tr>
<tr>
<td>4</td>
<td>0.806</td>
<td>0.783</td>
<td>0.533</td>
<td>0.705</td>
<td>0.777</td>
</tr>
</tbody>
</table>

was achieved using the bedroom activity GLMs only, an \(F_{0.5}\) score of 0.789 was achieved using the bathroom activity GLMs only, an \(F_{0.5}\) score of 0.791 was achieved using the kitchen activity GLMs only, an \(F_{0.5}\) score of 0.747 was achieved using the living room activity GLMs only, and an \(F_{0.5}\) score of 0.789 was achieved using majority voting among the four rooms.

One reason that could explain the worsening performance for windows larger than 20 weeks is that as the window size increased, the number of estimated activity GLMs per room for each subject decreased. Many subjects who transitioned to MCI during the monitoring period also had a long period of being cognitively intact. This was in addition to subjects who remained cognitively intact throughout the monitoring period. As a result, the database clearly contained more activity GLMs belonging to the ‘cognitively intact’ class than the ‘experiencing MCI’ class, which in turn meant that the set of exemplars created from the training stage most likely contained more exemplars representing activity GLMs belonging to the ‘cognitively intact’ class. If exemplars representing one class dominate the set of exemplars, then there is a higher probability for any given room activity GLM to be classified as belonging to the ‘cognitively intact’ class.
8.4.2 Leave-One-Subject-Out Cross-Validation

Out of the twenty-six subjects who were labeled as having MCI, twenty-two of them were cognitively intact for some period and then transitioned to either a-MCI or na-MCI. So in order to test the validity of the results obtained using the 80 percent – 20 percent division of the database, using the features that resulted in the best performance, i.e., \( \omega = 20 \) weeks and slide size = 4 weeks, we completed a leave-one-subject-out cross-validation process using the twenty-two subjects who transitioned to MCI during the monitoring period. The corresponding \( F_{0.5} \) scores achieved are summarized in Table 8.5. As Table 8.5 shows, comparable performance was noted which confirmed the results obtained using the 80 percent – 20 percent division of the database. The majority voting \( F_{0.5} \) score was higher by approximately 6 percent. This was most likely because the activity GLMs pertaining to the twenty-two subjects were more balanced. So there was not a big difference between the number of room activity GLMs belonging to the ‘cognitively intact’ class and the number of GLMs belonging to the ‘experiencing MCI’ class as was the situation with the 80 percent – 20 percent split case. As a result, the number of exemplars representing each class was balanced as well.

<table>
<thead>
<tr>
<th>Room</th>
<th>( F_{0.5} ) Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bedroom</td>
<td>0.842</td>
</tr>
<tr>
<td>Bathroom</td>
<td>0.829</td>
</tr>
<tr>
<td>Kitchen</td>
<td>0.813</td>
</tr>
<tr>
<td>Living Room</td>
<td>0.826</td>
</tr>
<tr>
<td>Combined</td>
<td>0.856</td>
</tr>
</tbody>
</table>

8.4.3 a-MCI & na-MCI Subtypes

We were also interested in investigating the different subtypes of MCI and how the room activity GLMs of subjects exhibiting symptoms of each of the subtypes differed from their
room activity GLMs when they were cognitively intact. Of the fifteen subjects who were labeled as having na-MCI, fourteen of them transitioned to na-MCI during the monitoring period. Table 8.6 summarizes the $F_{0.5}$ scores corresponding to the data pertaining to these fourteen subjects obtained using leave-one-subject-out cross-validation. The performance of the algorithm was impressive; the $F_{0.5}$ scores were all over 0.9 using the individual room activity GLMs, and with majority voting, the $F_{0.5}$ score was increased to 0.958. This result was very interesting since it indicated that subjects exhibiting symptoms of na-MCI have room activity GLMs that were significantly different from their room activity GLMs when they were cognitively intact. According to the definition of na-MCI, it is a subtype of MCI that impacts any of the cognitive domains listed in Table 3.2 except for the memory domain. Accordingly, an individual with na-MCI would have their executive functioning, attention, and thinking skills compromised, rendering them less able to make sound decisions, or to judge the sequence of steps needed to complete daily tasks in an efficient manner [56]. Accordingly, it is expected for changes to exist in the way older adults complete certain activities as they start exhibiting symptoms of na-MCI, which are reflected as changes in their room activity GLMs.

As for a-MCI, of the eleven subjects who were labeled as having a-MCI, eight of them transitioned to a-MCI during the monitoring period. Table 8.7 summarizes the $F_{0.5}$ scores corresponding to the data pertaining to these eight subjects obtained using leave-one-subject-out cross-validation. The observations based on these results were very interesting as well. First, the $F_{0.5}$ scores dropped drastically except for the bedroom,

<table>
<thead>
<tr>
<th>$F_{0.5}$ Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bedroom</td>
</tr>
<tr>
<td>0.941</td>
</tr>
</tbody>
</table>
Table 8.7: $F_{0.5}$ scores for $\omega = 20$ weeks and slide size = 4 weeks with a-MCI GLMs only

<table>
<thead>
<tr>
<th></th>
<th>Bedroom</th>
<th>Bathroom</th>
<th>Kitchen</th>
<th>Living Room</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{0.5}$</td>
<td>0.784</td>
<td>0.524</td>
<td>0.498</td>
<td>0.574</td>
<td>0.539</td>
</tr>
</tbody>
</table>

even majority voting yielded an $F_{0.5}$ score of 0.539, which was slightly better than random classification. This indicated that there were no significant differences between the subjects’ room activity GLMs when they were cognitively intact and when they were experiencing a-MCI.

According to the definition of a-MCI, it a subtype of MCI that primarily affects only the memory domain in Table 3.2. So individuals experiencing a-MCI are less able to remember information such as appointments, conversations, or recent events [56]. However, they are still able to carry out tasks in the same efficient manner they were able to when cognitively intact. This means that their executive functioning and attention skills are still intact, and thus they maintain their ability to judge the sequence of steps needed to complete daily tasks. Consequently, changes in their room activity GLMs are less likely to occur.

However, the majority of the subjects with a-MCI did have significant differences in their bedroom activity GLMs after they transitioned to MCI. This is mostly attributed to disturbed sleep patterns, mainly in the form of waking up at night. This finding is supported by a recent study by McGee et al., who concluded that sleep disturbance is common in subjects with a-MCI, and that many individuals with a-MCI have more than one symptom of sleep disturbance [76] [77]. Besides waking up at night, other symptoms include initial insomnia, snoring, breathing difficulties, nightmares, or excessive daytime sleepiness. In another study, Hita-Yañez et al. investigated the existence of sleep disturbances in MCI subjects via polysomnographic recordings in cognitively healthy older
adults and in patients with MCI [74]. Hita-Yañez et al. concluded that sleep disruptions were evident in MCI patients. These differences in results between a-MCI and na-MCI may be related to differences in the pathology underlying these MCI subtypes.

8.5 Conclusion

In conclusion, we proposed a home-based clustering technique for automatic detection of MCI in older adults. We formulated the problem as a classification problem with two classes: ‘cognitively intact’ and ‘experiencing MCI.’ A sliding window was used to create a database of room activity GLMs from the sensor data pertaining to eighty subjects. Affinity propagation was then applied to cluster the room activity GLMs in the database and to extract exemplars to represent the GLMs belonging to the different cognitive states. This represented the training stage. In the test stage, the room activity GLMs belonging to a test subject were compared to the exemplars and were assigned the labels of the exemplars that yielded the smallest normalized KL-divergence. The labels of the room activity GLMs were then used to determine to which class the test subject belonged. This answered our first research question.

By varying the size of the time frame, mainly from 4 weeks to 24 weeks, and the slide size from 2 weeks to 4 weeks, we observed an overall improvement of performance, in the form of higher $F_{0.5}$ scores as the window size increased. We did not really detect significant differences between the different slide sizes. The best performance was obtained using a time frame of 20 weeks and a slide size of 4 weeks. An $F_{0.5}$ score of 0.815 was obtained using the bedroom activity GLMs only, an $F_{0.5}$ score of 0.789 was obtained using the bathroom activity GLMs only, an $F_{0.5}$ score of 0.791 was obtained using the kitchen activity GLMs only, an $F_{0.5}$ score of 0.747 was obtained using the living room activity GLMs only, and an $F_{0.5}$ score of 0.789 was achieved using majority voting among the four rooms. This answered our second research question.

Finally, we investigated the different subtypes of MCI, namely a-MCI and na-MCI.
Using leave-one-subject-out cross-validation, we found that subjects exhibiting symptoms of na-MCI had room activity GLMs that were significantly different from their room activity GLMs when there were cognitively intact. This was reflected as very high $F_{0.5}$ scores, with a majority voting $F_{0.5}$ score of 0.958. As for a-MCI, we did not detect significant differences between the subjects’ room activity GLMs when exhibiting symptoms of a-MCI and when cognitively intact, except for the bedroom activity GLMs. This interesting finding was most likely attributed to disturbed sleep patterns which have been shown to be prevalent in individuals with a-MCI. This answered our third research question.
Chapter 9

Conclusions & Future Work

Of all the things I’ve lost, I miss my mind the most.

______________________________
Mark Twain

Alzheimer’s disease, the most common cause of dementia, is a fatal disease that eventually affects an older adult’s ability to think, make decisions, and perform simple daily tasks such as eating, bathing, and getting dressed. As the “baby boomer” generation ages, both the proportion and number of older adults with dementia is projected to increase dramatically, thus greatly increasing the burden on the health-care infrastructure. Consequently, detection of the cognitive decline that precedes Alzheimer’s disease, or dementia in general, becomes vital so as to appropriately apply current best practices for clinical evaluations and management as well as to be available for rationale administration of anticipated new therapies.

However, this ability of early detection is very challenging with the contemporary detection processes in the form of traditional doctor visits since many of the subtle clues are difficult to spot. This corroborates what has been reported in some studies that in more than 50 percent of cases that were investigated, it was the family members who served as the primary source of recognition and not the family doctors. This delay in
detecting cognitive decline can be detrimental especially to older adults with reversible forms of dementia, who form up to 11 percent of the cognitively impaired population. Early detection of remediable causes of cognitive impairment, such as medication complications or nutritional deficiencies, facilitates timely intervention, possibly increasing the chances of reversing the condition. Even for older adults with irreversible decline, early recognition still provides them and their families with an opportunity to proactively plan for their future by seeking the appropriate interventions and to avoid being forced into crisis management.

Mild Cognitive Impairment (MCI) is a concept that describes the transitional period between normal aging with intact cognition and early Alzheimer’s disease or other dementias. Individuals with MCI have measurable changes in their memory and thinking abilities but with minor impact on their ability to carry out activities of daily living. Although not all older adults with MCI progress to develop Alzheimer’s or other dementias, the proposed criteria and guidelines for diagnosis of Alzheimer’s disease published in 2011 have documented that individuals with MCI are at a higher risk of developing dementia and Alzheimer’s disease. Therefore, detecting MCI serves the objective of detecting cognitive decline early enough for older adults to seek the right treatment or intervention. Different subtypes of MCI exist such as amnestic MCI (a-MCI), where the memory of an individual is compromised, and non-amnestic MCI (na-MCI), where other cognitive domains, excluding memory, of an individual are impaired.

Several studies have demonstrated that changes in motor capabilities precede and may be indicative of cognitive impairment, and that changes in walking speed and patterns of daily activity could be good predictors of progression to dementia. Accordingly, and with the advancement of technology and the proliferation of smart systems, a good alternative to the traditional clinical paradigm is to bring assessment into the daily activity of a person in their home environment via unobtrusive sensors and systems. The literature is rich with smart systems and studies designed to monitor the health and well-being of older
adults and to support their independence. However, the majority of these smart homes and systems are designed for general health monitoring and not specifically for cognitive monitoring. Furthermore, all the results reported by these systems and studies have been based on data acquired in a laboratory environment and not in a real world setting. An approach that would be more reflective of the actual performance of individuals would be to continuously monitor them, perhaps in their own homes where they are more likely to exhibit their true performance.

The Oregon Center for Aging and Technology (ORCATECH) recruited more than three hundred cognitively healthy older adults and monitored them for an average period of three years using unobtrusive sensing technologies, resulting in a large database of sensor data and clinical data. Recently, ORCATECH published work on generating trajectories of home-based daily walking speeds and their variability in recruited older adults. In addition, the group at ORCATECH also compared the resulting trajectories between the cognitively intact older adults and those with non-amnestic mild cognitive impairment (na-MCI). Participating older adults with na-MCI were characterized by a slowing of walking speed and exhibited the highest and lowest variability in their walking speeds in comparison with their cognitively intact counterparts.

In Chapter 6, we built on the work by ORCATECH and explored the feasibility of autonomously discriminating older adults with MCI from their cognitively intact counterparts using a number of predefined measures associated with their in-home walking speed such as the weekly median walking speed, the coefficient of variation of the weekly walking speed, and the coefficient of variation of the morning walking speed. Different window sizes were used to extract features from these predefined measures which were then used to train and test two machine learning algorithms, namely support vector machines and random forests. We also explored using a number of predefined measures associated with the older adults’ home activity. Activity was measured as the total number of sensor firings. We experimented with measures such as the coefficient of variation
of total activity, coefficient of variation of morning activity, and coefficient of variation of evening activity. By completing the work described in Chapter 6, we answered the first research question and we made the following contributions:

1. We demonstrated how a signal processing approach equipped with a machine learning paradigm can be used to discriminate older adults with MCI from their cognitively healthy counterparts using several predefined measures associated with their walking speed.

2. We demonstrated that by analyzing a time window of only 24 weeks we could detect older adults with MCI with areas under the ROC curve and the precision-recall curve of 0.97 and 0.93, respectively.

3. We further analyzed the walking speed measures and their respective features and ranked them in descending order based on their importance for detection of cognitive impairment. We found that trajectories of weekly walking speed, coefficient of variation of walking speed, coefficient of variation of morning and evening walking speeds, and the older adults’ age and gender were the most important features for the process of detecting MCI in older adults.

4. We took the lead in reporting these promising results using real-world data suffering from too much noise and missing many datapoints. Almost all of the other related work in the literature, by contrast, have reported results obtained by carrying out activities and analyzing data acquired in a laboratory setting which might not have necessarily reflected the real performance of the recruited older adults.

5. Finally, we demonstrated that activity measures corresponding to the number of sensor firings did not carry sufficient discriminative information to discriminate older adults with MCI from their cognitively healthy counterparts.

In Chapter 7, we were motivated to seek an alternative approach to discriminate
older adults with MCI from their cognitively healthy counterparts using changes in their activity patterns. We built statistical models of the recruited older adults’ home activity by using inhomogeneous Poisson processes to model their presence within different rooms throughout the day. The resulting generalized linear models (GLMs) of the older adults’ home activity provided intuitive statistical analysis, and using the Kullbak–Leibler (KL)-divergence measure we demonstrated an existing statistical difference between activity models pertaining to older adults when cognitively intact and when exhibiting symptoms of MCI. By completing the work described in Chapter 7, we answered the second research question and we made the following contributions:

1. We demonstrated that older adults’ home activity can be well modeled as independent inhomogeneous Poisson processes.

2. We showed how the resulting GLMs of the older adults’ in-home activity provided intuitive statistical analysis and equipped us with the ability to visualize a older adult’s pattern of activity, a feature that was very challenging to obtain with models built using a number of predefined measures.

3. We took an information theoretic approach and used the KL-divergence measure to demonstrate an existing statistical difference between activity models pertaining to cognitively intact and impaired older adults.

4. By using a simple thresholding approach of the KL-divergence measure, we were able to detect mild cognitive impairment in older adults with average areas under the ROC curve and the precision-recall curve of 0.716 and 0.706, respectively, using activity models estimated over sliding time windows of 12 weeks only.

Finally, in Chapter 8, we built on the promising results discussed in Chapter 7 and devised a clustering-based method to automatically detect MCI in older adults using the estimated GLMs of their home activity. Exemplars were created using a combination of
Chapter 9. Conclusions & Future Work

$k$-means and affinity propagation (AP) to represent clusters of the room activity GLMs belonging to different states of cognition. These exemplars were then used to label room activity GLMs belonging to a test subject with unknown cognitive status. The GLMs labels were then used to determine the cognitive status of the test subject. We also used neuropsychological assessments as our ground truth to investigate potential differences between the two subtypes of MCI, namely amnestic MCI (a-MCI) and non-amnestic MCI (na-MCI). By completing the work described in Chapter 8, we answered the third research question and we made the following contributions:

1. We proposed a clustering-based method to automatically detect MCI in older adults using estimated GLMs of their home activity. Exemplars were created using a combination of $k$-means and affinity propagation (AP) to represent clusters of the room activity GLMs belonging to different states of cognition. These exemplars were then used to classify room activity GLMs belonging to a test subject. These classifications were then employed to determine the cognitive status of the test subject.

2. We were able to detect MCI in older adults with an $F_{0.5}$ score of 0.856 using a time frame of 20 weeks, which is approximately 5 months. Accordingly, a time frame of 20 weeks is empirically reported as the duration that generates room activity GLMs that are most conducive to detecting MCI in older adults.

3. Interestingly, when including the recruited older adults exhibiting symptoms of na-MCI only, we were able to detect MCI in older adults with an $F_{0.5}$ score of 0.958. On the other hand, we were able to detect significant differences only in the bedroom activity GLMs of older adults as they started experiencing a-MCI.
9.1 Translation of Work

In this thesis, we described two methodologies and approaches to automatically detect MCI in older adults— one using longitudinal trajectories of in-home walking speed and another one using clusters of generalized linear models of room activity. Ideally, using both approaches is supposed to provide us with a robust detection of MCI in older adults. If both approaches classify an older adult as experiencing MCI, then it is with high confidence that this detection is a true positive.

One advantage of using trajectories of in-home walking speed is the simplicity of the sensing system. Only a line of four or five sensors with restricted field of view is needed to be installed in a long hallway or corridor in the living unit. The long hallway provides older adults with ample distance to walk naturally and thus enables the sensing system to measure their walking speed. However, despite its simplicity, such a system is not always feasible especially in living units that lack such a long hallway.

In the event that such a long hallway does not exist, then the approach using clusters of generalized linear models of room activity is still feasible. This approach can be easily implemented as well given the unsophisticated sensors utilized. Wireless security systems are becoming prevalent. These security systems utilize passive infra-red motion sensors which are very similar to the ones used in this thesis. Therefore, our approach of using clusters of generalized linear models of room activity can be amalgamated with the security system, and then the unified system can be used simultaneously as a security system and to detect changes in cognition.

9.2 Limitations and Future Work

Next, we discuss some of the main limitations of the thesis and we propose directions for future work.
9.2.1 General Limitations

1. The algorithms and methods described in this thesis were based on clean data. Recall that the eligibility criteria included older adults being in average health for age. This meant that the recruited older adults either had no chronic diseases or in the event that they had chronic diseases or comorbidities, then they were well-controlled and had insignificant impact on their level of activity and walking speed. This important criterion was maintained by utilizing the information that was submitted by the recruited older adults in the weekly online questionnaires. The questionnaires were leveraged to discard days on which diseases that impacted overall health and walking speed, such as arthritis or chronic congestive heart failure, were reported to have worsen. Maintaining clean data would ensure that any changes in motor capabilities would correlate with changes in cognition only and not with other chronic diseases or conditions.

2. In building the algorithms and methods described in this thesis, we focused on homes with single occupants only. If the system were to be implemented in living units with more than one occupant, then the algorithms need to be extended to handle the additional occupants. One potential solution would be to incorporate additional sensors such as radio-frequency identification (RFID) tags and readers in order to facilitate differentiating the multiple occupants.

3. Finally, although very interesting and promising results were obtained, we believe it is still very important to test our proposed algorithms and methods on a larger population size in order to corroborate these results and confirm their validity.

9.2.2 Chapter 6

One limitation of the work we completed in Chapter 6 is that with features in the form of trajectories of the individuals measures, data have to be present for every week. If a
gap exists in the form of a discarded week, then the whole window would be discarded. This poses a serious challenge since older adults are likely to have visitors over more frequently, and with the sensing technologies employed in this study, many weeks could potentially end up being discarded since there is no way to differentiate the activity pertaining to the monitored older adult from the activity pertaining to the visitors. As we have demonstrated, discarding too much data could result in the problem of overfitting, especially for windows of large sizes such as 24 weeks, due to the high dimension of the resulting feature vectors. Therefore, one potential direction for future work is to build generalized linear models of in-home walking speed. Linear regression with a Gaussian likelihood can very well model trajectories of walking speed since this would be equivalent to minimizing the mean squared error between the actual walking speeds and the estimated ones. As we demonstrated with generalized linear models of home activity in Chapter 7, such an approach is indisposed to the problem of overfitting and provides intuitive statistical analysis.

9.2.3 Chapter 7

One limitation of the work we completed in Chapter 7 is that we built the generalized linear models (GLMs) of home activity under the assumption that the Poisson processes that modeled an older adult’s presence in different rooms throughout the day were independent. However, this assumption is not realistic because an older adult cannot be in two different rooms at the same time. Therefore, one very interesting direction for future work is to incorporate this dependency factor in the formulation of estimating the activity models and analyze any differences that emerge in the the estimated models.

9.2.4 Chapter 8

One limitation of the proposed work in Chapter 8 is the prerequisite of having sufficient data to build a database of room activity GLMs in order to extract exemplars of the dif-
different states of cognition. Accordingly, an alternative method has to be designed in order to address cases where such a database of room activity GLMs cannot be created. One direction for future work is to investigate the feasibility of using changepoint detection for automatic detection of MCI in older adults. If feasible, such an approach would not require a database of room activity GLMs. Instead, it would track each individual separately and perform intra-individual comparison by comparing each older adult’s activity GLMs to his/her past GLMs, and assess his/her cognition accordingly [78]. In order to successfully accomplish this, it is necessary to determine the best window size needed to build the baseline GLM for each older adult for each room in their living unit. Furthermore, we need to determine how much change in the model parameters would reflect a significant change in cognition and based on which an older adult would be referred to a memory clinic for a comprehensive cognitive assessment. This comprehensive cognitive assessment is necessary in order to clinically confirm any cognitive impairment associated with Alzheimer’s disease or dementia.

Finally, the work described in this dissertation is of great significance as it contributes to the overarching objective of detecting dementia and Alzheimer’s disease early. Early detection of cognitive impairment is an important source of help for families as it enables them to gain a measure of control over the situation, provides a basis for accepting and communicating with others about the reality of the impairment and its effects, and sets the stage for planning for financial and future health care needs. Even if no effective treatment for dementia is available, interventions that enhance daily functioning and safety and reduce emotional distress, familial burden, and individual fear and uncertainty about the future, are certainly available. Enabling families to seek community resources and sound care management for their impaired members early in the process, through early detection of cognitive impairment, can go a long way toward quality patient care for this vulnerable population.
Appendices
Appendix A

Examples of Labeling Data Using CDR

In this Appendix, we describe three additional examples of how we used the labeling protocol, that we defined for CDR, to label subject data using their annual assessment scores.

A.1 Example 2

Table A.1: Example 2 of labeling subject data using CDR labeling protocol

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>baseline</th>
<th>1&lt;sup&gt;st&lt;/sup&gt;</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt;</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR score</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

According to Table A.1, the subject exhibited impairment on the 3<sup>rd</sup> annual assessment, after which no subsequent assessments are available. Therefore, according to rule 4, the subject is considered impaired from the date of this assessment onward. Accordingly:
1. The subject data from baseline until the date of the 2\textsuperscript{nd} annual assessment will be assigned the label “CIN.”

2. The subject data from the date of the 2\textsuperscript{nd} annual assessment up to the date of the 3\textsuperscript{rd} annual assessment will be assigned the label “TR.” This is because the conversion to cognitive impairment is not a point event but a gradual process. Accordingly, the subject’s cognition would be in flux between the 2\textsuperscript{nd} and the 3\textsuperscript{rd} annual assessments and would belong to neither “CIN” nor “MCI.”

3. The subject data from the date of the 3\textsuperscript{rd} annual assessment onward will be labeled “MCI.”

\section*{A.2 Example 3}

\begin{table}[h]
\centering
\caption{Example 3 of labeling subject data using CDR labeling protocol}
\begin{tabular}{llll}
\hline
Annual Assessments & baseline & 1\textsuperscript{st} & 2\textsuperscript{nd} & 3\textsuperscript{rd} \\
\hline
CDR score & 0 & 0.5 & 0 & 0.5 \\
\hline
\end{tabular}
\end{table}

According to Table A.2, the subject exhibited impairment on the 1\textsuperscript{st} annual assessment, but the impairment disappeared on the subsequent assessment. The impairment reappeared on the 3\textsuperscript{rd} annual assessment. However, because of the bouncing instance, we cannot consider this impairment. In fact, this subject is considered a “bouncer” and is excluded from our analysis.

\section*{A.3 Example 4}

According to Table A.3, the subject exhibited impairment on the 1\textsuperscript{st} annual assessment. Although the impairment disappeared on the 2\textsuperscript{nd} annual assessment, it reappeared on the
all of the subsequent assessments. Therefore, according to rule 3, the absent impairment on the 2nd annual assessment is ignored and the subject is considered impaired from the 1st assessment onward. Accordingly:

1. None of the subject data will be assigned the label “CIN.”

2. The subject data will be labeled “TR” from baseline until the date of the 1st annual assessment.

3. The subject data will be labeled “MCI” from the date of the 1st annual assessment onward.
Appendix B

Examples of Labeling Data Using Neuropsychological Assessments

In this Appendix, we describe three additional examples of how we used the labeling protocol, that we defined for the neuropsychological tests, to label subject data using their annual assessment scores.

B.1 Example 2

Table B.1: Example 2 of labeling subject data using the neuropsychological tests labeling protocol

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>MMSE</th>
<th>CDR</th>
<th>Executive Function</th>
<th>Language</th>
<th>Attention</th>
<th>Memory</th>
<th>Visuospatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>25</td>
<td>0</td>
<td>-0.86</td>
<td>-1.31</td>
<td>-1.33</td>
<td>-0.93</td>
<td>-2.01</td>
</tr>
<tr>
<td>1st</td>
<td>24</td>
<td>0</td>
<td>-0.60</td>
<td>-0.64</td>
<td>-0.96</td>
<td>-0.66</td>
<td>-1.32</td>
</tr>
<tr>
<td>2nd</td>
<td>26</td>
<td>0</td>
<td>-0.31</td>
<td>-1.55</td>
<td>-1.23</td>
<td>-0.47</td>
<td>-0.82</td>
</tr>
</tbody>
</table>

According to Table B.1, the subject exhibited impairment in the visuospatial domain at baseline. However, this impairment disappeared on all of the subsequent annual as-
Appendix B. Examples of Labeling Data Using Neuropsychological Assessments

assessments. Therefore, according to rule 1, this impairment is ignored since it did not appear on two consecutive annual assessments. On the other hand, the subject exhibited impairment in the language domain on the 2\textsuperscript{nd} annual assessment, which is the last available assessment. According to rule 4, the subject is considered impaired from the 2\textsuperscript{nd} annual assessment onward. Accordingly:

1. The subject data from baseline until the date of the 1\textsuperscript{st} annual assessment will be assigned the label “CIN.”

2. The subject data from the date of the 1\textsuperscript{st} annual assessment up to the date of the 2\textsuperscript{nd} annual assessment will be assigned the label “TR.” This is because the conversion to cognitive impairment is not a point event but a gradual process. Accordingly, the subject’s cognition would be in flux between the 2\textsuperscript{nd} and the 3\textsuperscript{rd} annual assessments and would belong to neither “CIN” nor “MCI.”

3. The subject data from the date of the 2\textsuperscript{nd} annual assessment onward will be labeled “MCI,” in particular “na-MCI” because impairment is in the language domain.

### B.2 Example 3

Table B.2: Example 3 of labeling subject data using the neuropsychological tests labeling protocol

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>MMSE</th>
<th>CDR</th>
<th>z-scores</th>
<th>Executive Function</th>
<th>Language</th>
<th>Attention</th>
<th>Memory</th>
<th>Visuospatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>25</td>
<td>0.5</td>
<td>-1.63</td>
<td>0.27</td>
<td>-0.08</td>
<td>-0.66</td>
<td>-1.36</td>
<td></td>
</tr>
<tr>
<td>1\textsuperscript{st}</td>
<td>26</td>
<td>0.5</td>
<td>-0.61</td>
<td>-1.31</td>
<td>-0.21</td>
<td>-0.61</td>
<td>-1.64</td>
<td></td>
</tr>
<tr>
<td>2\textsuperscript{nd}</td>
<td>29</td>
<td>0.5</td>
<td>-1.02</td>
<td>-0.64</td>
<td>-0.08</td>
<td>-0.65</td>
<td>-1.04</td>
<td></td>
</tr>
<tr>
<td>3\textsuperscript{rd}</td>
<td>29</td>
<td>0</td>
<td>-0.37</td>
<td>0.27</td>
<td>-0.23</td>
<td>-0.94</td>
<td>-1.04</td>
<td></td>
</tr>
</tbody>
</table>

According to Table B.2, the subject exhibited impairment in the executive function domain at baseline, and exhibited impairment in the visuospatial domain on the 1\textsuperscript{st}
Appendix B. Examples of Labeling Data Using Neuropsychological Assessments

annual assessment. However, both impairments disappeared on all of the subsequent assessments. According to rule 1, these impairments are ignored because they did not appear on two consecutive assessments. Accordingly:

1. All of the the subject data will be assigned the label “CIN.”

2. None of the subject data will be assigned the label “TR.”

3. None of the subject data will be assigned the label “MCI.”

B.3 Example 4

Table B.3: Example 4 of labeling subject data using the neuropsychological tests labeling protocol

<table>
<thead>
<tr>
<th>Annual Assessments</th>
<th>MMSE</th>
<th>CDR</th>
<th>Executive Function</th>
<th>Language</th>
<th>Attention</th>
<th>Memory</th>
<th>Visuospatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>28</td>
<td>0</td>
<td>-0.06</td>
<td>-0.22</td>
<td>-2.09</td>
<td>-0.90</td>
<td>-1.25</td>
</tr>
<tr>
<td>1st</td>
<td>24</td>
<td>0.5</td>
<td>-0.47</td>
<td>-0.86</td>
<td>-1.67</td>
<td>-1.78</td>
<td>-1.25</td>
</tr>
<tr>
<td>2nd</td>
<td>20</td>
<td>0.5</td>
<td>-0.61</td>
<td>-0.40</td>
<td>0.18</td>
<td>-2.03</td>
<td>0.00</td>
</tr>
<tr>
<td>3rd</td>
<td>19</td>
<td>0.5</td>
<td>-1.63</td>
<td>1.18</td>
<td>0.18</td>
<td>-3.33</td>
<td>0.00</td>
</tr>
</tbody>
</table>

According to Table B.3, the subject exhibited impairment in the attention domain on two consecutive annual assessments. Therefore, according to rule 1, impairment in the attention domain is considered. Similarly, the subject exhibited impairment in the memory domain on three consecutive assessments, including the 3rd annual assessment, which is the last available assessment. Therefore, this impairment is also considered and according to rule 4, the subject is considered impaired in the memory domain from the 1st annual assessment onward. In addition, the subject exhibited impairment in the executive function domain on the 3rd annual assessment, which was the last available assessment. Therefore, according to rule 4, the subject is considered impaired in the
executive function domain from the 3rd annual assessment onward. However, according to rule 3, the impairments in the attention domain and the executive function domain are combined since they result in the same type of MCI, and the gap on the 2nd annual assessment is ignored. Accordingly:

1. None of the subject data will be assigned the label “CIN.”

2. None of the subject data will be assigned the label “TR.”

3. All of the subject data will be labeled “MCI.” In particular, the subject’s data will be labeled “na-MCI” from baseline to the 1st annual assessment, and will be labeled “a-MCI multiple domain” from the 1st annual assessment onward.
Bibliography


