Signal Processing and Machine Learning Techniques Towards Various Real-World Applications

by

Arindam Dutta

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved August 2018 by the
Graduate Supervisory Committee:

Daniel W. Bliss, Chair
Visar Berisha
Christ Richmond
Steven Corman

ARIZONA STATE UNIVERSITY

December 2018

ABSTRACT

Machine learning (ML) has played an important role in several modern technological innovations and has become an important tool for researchers in various fields of interest. Besides engineering, ML techniques have started to spread across various departments of study, like health-care, medicine, diagnostics, social science, finance, economics etc. These techniques require data to train the algorithms and model a complex system and make predictions based on that model. Due to development of sophisticated sensors it has become easier to collect large volumes of data which is used to make necessary hypotheses using ML. The promising results obtained using ML have opened up new opportunities of research across various departments and this dissertation is a manifestation of it. Here, some unique studies have been presented, from which valuable inference have been drawn for a real-world complex system. Each study has its own unique sets of motivation and relevance to the real world. An ensemble of signal processing (SP) and ML techniques have been explored in each study. This dissertation provides the detailed systematic approach and discusses the results achieved in each study. Valuable inferences drawn from each study play a vital role in areas of science and technology, and it is worth further investigation. This dissertation also provides a set of useful SP and ML tools for researchers in various fields of interest.

I dedicate this dissertation to my family with special gratitude to my grandparents for the love, support, and push for continuous personal improvement.

ACKNOWLEDGMENTS

I would like to express the deepest respect and appreciation to Dr. Daniel W. Bliss, my mentor, for his teaching, reflecting, encouraging, and most of all patience throughout the entire process. His knowledge, mindedness, and academic rigor made this dissertation possible and further have inspired me to be a better scholar.

I wish to thank my committee members who were more than generous with their expertise and precious time. A special thanks to Dr. Steven Corman and Dr. Visar Berisha for their contributions for technical guidance and inspiration in the entire process. Thank you, Dr. Christ Richmond for your valuable suggestions and insight that helped me shape this dissertation. I wish to thank all the professors with whom I had the opportunity to work in various projects, including Dr. Antonia Papandreou-Suppappola, Dr. Keith Holbert, Dr. Mathew P. Buman, Dr. Umit Ogras and Dr. Corrie Whisner.

To my friends and roommates, thank you for listening, offering me advice, and supporting me through this entire process. Special thanks to my friends: Alex Chiriyath, Harsha Rentala, Sanjeev Tarigopula, Praveen Kumar, Pradeep Balaji, Somya Singh, Karthik Rajan. The debates, dinners, game nights, jamming sessions as well as editing advice, and general help and friendship were all greatly appreciated. To all my labmates, especially Owen Ma, for all the help and support and mostly making it feel like a family.

This dissertation would not have been possible without funding from the National Science Foundation, USA.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER

1 INTRODUCTION 1

1.1 Contributions 4

2 SIGNAL PROCESSING AND MACHINE LEARNING TECHNIQUES . 10

2.1 Signal Feature Extraction Techniques 10

2.2 Dimensionality Reduction and Feature Selection Techniques 29

2.3 Machine Learning Techniques 32

2.4 Model Accuracy Evaluation Techniques 49

3 Study I: ACCELEROMETER SIGNAL PROCESSING TO MONITOR PHYSICAL ACTIVITY 52

3.1 Overview 52

3.2 Background 53

3.3 Lab-Based Study Protocol 56

3.4 Free-Living Study Protocol 64

3.5 Discussion 68

4 Study II: ANALYZING GUT-MICROBIOME DATA TO FIND THEIR RELATIONSHIP WITH BONE HEALTH 72

4.1 Overview 72

4.2 Background 72

4.3 Study Design and Data Collection 75

4.4 Problem Formulation 79

4.5 Prediction Analysis 80

CHAPTER Page

4.6 Results 81
4.7 Discussion 85
5 Study III: PROCESSING SPEECH TO UNDERSTAND HUMAN COMMUNICATION IN AN ORGANIZATION 87
5.1 Overview 87
5.2 Background 88
5.3 Organization Setting and Data Collection 89
5.4 Methods 90
5.5 Results 97
5.6 Discussion 99
6 Study IV: DECODING HUMAN INTENT USING MULTI-MODAL SENSOR DATA 102
6.1 Overview 102
6.2 Background 102
6.3 Multi-Modal Intent Decoding Methodology 105
6.4 Experimental Results 108
6.5 Discussion 111
7 Study V: DISCRIMINATION BETWEEN NEUTRON AND GAMMA RAYS EMITTED DURING RADIOACTIVE DECAY 113
7.1 Overview 113
7.2 Background 114
7.3 Experimental Setup and Data Acquisition 115
7.4 Neutron-Gamma Discrimination using Decay Time Algorithm 116
7.5 Results 118

CHAPTER Page

7.6 Neutron-Gamma Discrimination using Normalized Cross Correlation 120
7.7 Discussion . . . 125
8 OTHER STUDIES . . . 127
8.1 Identifying Behavioral Tasks using Local Field Potential Signals from Patients with Parkinson's Disease . . . 127
8.2 EEG Processing to Predict Postoperative Delirium . . . 135
8.3 Detection of Weak Signals in a SISO Wireless Communication Scenario . . . 142
9 SUMMARY . . . 152
REFERENCES . . . 155
APPENDIX
A List of Acronyms . . . 172

LIST OF TABLES

Table		Page
1.1	Summary of Each Study	8
2.1	Accelerometer Signal Features	12
3.1	Study I Lab-Based PA Dataset	57
3.2	PA Classification Accuracy using GMM/HMM	62
4.1	Study II Participant Demographics	76
4.2	Prediction Analysis of BMD using GM	81
4.3	Analysis Results of Significant Microbes	86
5.1	Prediction Analysis of Productivity using Communication Features	98
5.2	Communication Features Ranks	100
7.1	Study V Experimental Data	117
7.2	Radiation Pulse Parameters	121
8.1	Confusion Matrix using HMM-SVM	132
8.2	Confusion Matrix using Hybrid HMM-SVM	133
8.3	Confusion Matrix using DPGMM	134
8.4	Confusion Matrix for Delirium Prediction	141

LIST OF FIGURES

Figure		Page
2.1	A Simple Finite Graph	19
2.2	32 Channel EEG	21
2.3	Neutron and Gamma-ray Pulse Decay Time	27
2.4	SFS Block Diagram	31
2.5	SVM Classifier	36
2.6	HMM-SVM Classifier	38
2.7	Sequential Detection Process	45
2.8	Serial Hypothesis Tracking	47
2.9	Parallel Hypothesis Tracking	48
2.10	Confusion Matrix	49
2.11	ROC Curve	51
3.1	GeneActiv Accelerometer Signal	55
3.2	Lab-Based PA Classification Process	59
3.3	Class Merging using Confusion Matrix	61
3.4	Classifier Comparison between GMM and HMM	64
3.5	Convergence Characteristics of GMM/HMM	65
3.6	Free-Living PA Classification Process	66
3.7	Free-Living Analysis Results	69
4.1	Correlation Weights of GM	83
4.2	Positively Correlated Phyla	83
4.3	Proportion of Microbes in Participants	84
5.1	Communication Graph Analysis Process	92
5.2	Communication Graph	93
5.3	Communication Detection Confusion Matrix	94

Figure Page

5.4 Prediction Analysis Process 97
5.5 Correlation Coefficient of Communication Features 99
6.1 Multi-Modal Signal Analysis Process 105
6.2 Multi-Modal Sensor Units 106
6.3 Accelerometer Signal 107
6.4 Gyroscope Signal 108
6.5 Study IV Experimental Setup 109
6.6 Space-Delay Eigenspectra of Three Subjects 110
6.7 Eigenspectrogram of Three Subjects 111
7.1 PSD with AmBe Source 119
7.2 PSD without AmBe Source 119
7.3 Experimentally Obtained Pulse Models 121
7.4 PSD for Pileup Simulation 123
7.5 PSD at 13 kHz Event Rate 124
7.6 PSD at 1660 kHz Event Rate 126
8.1 DBS and LFP Signals 130
8.2 EEG Singal and Spectrogram during Surgery 138
8.3 Delirium Prediction Analysis Process 140
8.4 Detection Scenario 144
8.5 Channel Probing and Signaling Waveforms 145
8.6 ROC for Serial and Parallel Hypothesis Tracking 148
8.7 Latency Analysis Results 151

Chapter 1

INTRODUCTION

Machine Learning (ML) has grown in popularity since the term was initially coined by Arthur Samuel in 1959 [161], and now we have arrived at a point where it is being employed in various major fields of study. ML focuses primarily on the study and construction of algorithms that can learn and make predictions based on data, as well as overcome program limitations and make data-driven decisions. Thus, from any data driven study, we can model the data using a training set, and make predictions about future data. ML algorithms learn in different ways, and as they are exposed to new data sets, they adapt over time and improve their accuracy. These algorithms are employed to deal with the following problems,

- *Classification*: When the input data are divided into two or more classes, and the algorithm produces a model that assigns unseen inputs to these classes. In other words, when the aim is to assign each input vector to one of a finite number of categories.

- *Regression*: When the input data is used to predict new data values and the desired output consists of one or more continuous variables.

- *Clustering*: When the input data is to be divided into groups and but unlike in classification, the groups are not known beforehand.

- *Density Estimation*: When we need to find the distribution of the input data in some space.

- *Detection*: When we need to detect and interpret an event in real time.

- *Dimensionality reduction*: When we need to simplify the input data into a lower dimensional space.

These problems are associated with various fields of study, both in industrial and research applications. The science of learning plays a key role in the fields of statistics, data mining and artificial intelligence, intersecting with areas of engineering and other disciplines. Some of the common data-driven fields of research that has bolstered from the use of ML are pattern recognition [6, 13, 46, 104], clinical trials [36, 38, 66, 102], digital health-care [32, 135, 140, 200], natural language processing [203], computer vision [20, 27, 63] etc.

The learning problems that we consider can be roughly categorized as either supervised or unsupervised. In supervised learning, the goal is to predict the value of an outcome measure based on a number of input measures; in unsupervised learning, there is no outcome measure, and the goal is to describe the associations and patterns among a set of input measures. Both categories attempt to find the best model that fits the input data and makes accurate predictions. The entire process is accomplished in the following steps,

- *Data Pre-processing*: It involves tasks like cleaning the data, removing bad data, filtering to remove noise, removing artifacts and segmentation of data into small windows, if required. This step is extremely necessary, especially in cases of physiological and clinical data, where the data is affected by noise and artifacts.

- *Feature Extraction*: It involves identifying key features or identifiers that describe blocks of data. These features are extracted using various Signal Processing (SP) and statistical methods. They can be in the same space or in a

transformed space. It also serves a purpose to reduce the dimensionality of the data and avoids unnecessary of redundant information.

- *Feature Selection*: It involves selecting a subset from an entire set of features, that describes the data at best or in other words produces the most accurate model.

- *Classification/Prediction*: It involves finding the ML model that fits the data mostly accurately in its feature space. As a consequence of this, necessary inference can be made regarding classification or prediction, based on the problem.

SP techniques are used to pre-process the data and extract useful/dominant information hidden in the data. Classification/prediction is carried out using these features. Thus, the performance of the classifier depends on how good the features are. Also, an efficient feature set reduces the load on the classifier, thus making the choice of the classifier to be simple and computationally less demanding. In a lot of cases, the data has to be transformed into a different space to identify the efficient features. Thus, the feature extraction step is unique and specific to the type, origin and collection of data. ML algorithms are used for feature selection and classification/prediction purposes. These algorithms use methods and models borrowed from statistics and probability theory. Most of the learning problems are formulated as an optimization problem; minimization of some loss function on the training set. Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances. The identification of the right kind of model depends on feature properties and can be validated by accuracy estimation techniques like the k-fold-cross-validation, which randomly splits the data into k subsets where the $k-1$ instances of the data subsets are used to train the model while the kth subset instance is used to test the predictive ability of the training model. The overall

accuracy of the model is reported in various ways based on the type of the problem. Confusion matrix, Pearson's correlation coefficient, region of convergence are a few of such methods.

This work provides 8 unique research studies, where various real-world datasets were analyzed to find solutions to problems related to classification, prediction, clustering, detection and model estimation. Each dataset is investigated in a unique manner, exploring various SP algorithms to process the data and extract features from it and use various ML algorithms to make necessary inference based on the study. Each study has its own unique sets of motivation and relevance to the real world. This dissertation provides the detailed systematic approach and discusses the results achieved in each study. It also provides a set of useful tools for researchers in various fields of interest. Although each study might be appealing to specific set of audience, but all of them use similar ideology to address the problem. All these studies make use of various SP and ML algorithms to understand the real-world complex system derived from each dataset. We have presented 5 studies in details in five separate chapters and the other 3 studies were described briefly in Chapter 8.

1.1 Contributions

In this dissertation, various collaborative research works are presented, most of which were funded research projects. Each chapter discusses the purpose of each study along with the detailed process chain to achieve the desired goal. Each study presents a unique dataset which were collected for the purpose of the study, and approaches used to analyze the data. This dissertation presents a collection of novel, modified and existing algorithms that were used for analysis. In this section, the purpose and contribution of each study is presented briefly, mentioning the various collaborative teams who contributed to each study.

In Chapter 3, we presented a study where wearable sensors were used to collect accelerometer data. The data used in this analysis were gathered through two separate studies conducted by the College of Health Solutions, Arizona State University (ASU), lead by Dr. Matthew P. Buman. The first study provided accelerometer data on structured, lab-based physical activities which were classified and modeled using ML algorithms. The data from the free-living protocol were used to evaluate the developed algorithm in its capacity to estimate activity intensities in a free-living setting. The data analysis was done by the engineering team at Bliss Laboratory of Information, Signals, and Systems (BLISS) lab, lead by the author. This study was unique because both lab-based and free-living dataset were investigated, with one data co-dependent on the other. Also, it is an important advance from past studies, given its comparatively larger and diverse database and high classification accuracy (80-100%).

Chapter 4 introduces a study to investigate the relationship between gut microbiome and bone density and strength. Recent researches have shown that gut microbes (bacteria, viruses, fungi and other microscopic living things inside the intestines) affects the human body in various ways, metabolism, brain health, immune system etc. The data used in this study was collected from 28 healthy adolescent females, by the School of Nutrition and Health Promotions, ASU, lead by Dr. Corrie M. Whisner. We investigated 221 microbe classes to predict the measures of bone health and found high correlation of 0.8 between the actual and predicted bone density.

In Chapter 5, we analyzed speech data to understand the relationship between human communication and productivity in an organization. The data was collected from a software engineering organization called the Software Factory, operating from 3 years at ASU. The data collection and validation process was done by a team from the Hugh Downs School of Human Communication, ASU, lead by Dr. Steven

R. Corman. In this work, we construct inter-employee communication network using speech recordings and extract various speech features to predict the productivity of the organization. The engineering team at BLISS lab was responsible for the data analyses, lead by the author. This is a very unique empirical study, in terms of organizational social studies as it involves quantitative analysis showing the relation between human communication and productivity.

Chapter 6 introduces a simple multi-modal wearable system consisting of a motion processing unit (accelerometer and gyroscope) and a brain machine interface (Electroencephalogram (EEG) recording). This study was conducted to show that the system can successfully employ cognitive indicators to predict human activity. The motion processing unit was developed by the engineering team under Dr. Umit Ogras at ASU. Experiments were conducted for this purpose and data was collected using the wearable sensors. This work presented a methodology to predict future human action by analyzing EEG signal (15 out of 22).

Chapter 7 presents a study that provides techniques to discriminate between neutrons and gamma rays, which is an important field when we study about scintillators. Neutron and gamma rays are particles of mixed field radiations, which can be converted into electrical signals using scintillation detectors. In this work, we have proposed methods to distinguish between neutron and gamma ray pulses using experimental data. This work was a collaborative study between Radiation Monitoring Devices (RMD) Inc., who were responsible for data collection lead by Erik B. Johnson and ASU engineering team, lead by the author under the supervision of Dr. Keith E. Holbert. We implemented novel techniques to extract features and used clustering technique to perform neutron-gamma discrimination.

In chapter 8, we have presented our contribution in 3 other studies. The first one is the work from the author's Masters thesis. This work was a collaborative study

between Colorado Neurological Institute, who were responsible for data collection, lead by Dr. Adam O. Hebb and the engineering teams from University of Denver and ASU, lead by Dr. Jason J. Zhang and Dr. Antonia P. Suppappola. In this work, adaptive learning methods were proposed to identify different behavioral tasks of patients with Parkinson's disease. Local field potential (LFP) signals were collected from 12 subjects undergoing deep brain stimulation (DBS) as standard care for treatment of Parkinson's disease. We presented two classification algorithms, supervised and unsupervised and achieved 78-96% classification accuracy. The next study was based on a clinical trial, which was a collaboration work between Arizona State University and the Mayo Clinic. This study attempts to uncover bio-markers from intraoperative EEG to predict postoperative delirium, which is associated with long term health complications. The EEG data was recorded from 16 patients during their cardiac surgeries involving deep hypothermia circulatory arrest at Mayo Clinic, lead by Dr. Amy Z. Crepeau. The work presented in this chapter was lead by Owen Ma from BLISS lab. Postoperative delirium diagnoses were predicted by examining changes in the neural activity and we achieved correct predictions in 14 out of 16 cases. In the third and last part of this chapter we presented a study related to detection of weak signals in a single-input, single-output wireless communication scenario at high signal to noise ratio. This type of problem arises in time-sensitive applications like, remote surgery, vehicle to vehicle communication, radar and radio astronomy. Besides the author, Richard Gutierrez from BLISS lab made significant contributions to this work. We have introduced two hypothesis tracking algorithms built on the sequential detection framework and compared their performances using simulated data.

This dissertation presents various real world applications, that were studied in details and certain hypotheses were drawn from each study. Although the scope of

Table 1.1: Summary of each study

Chapter	Application	Data	Problem type	Tools used	Contributed publications
3	Identify and monitor physical activity	Tri-axial accelerometer	Classification Modeling	**FE**: accelerometer signal features modified wavelet features **FS**: Sequential forward selection **ML**: GMM, HMM, Gaussian model matching	[49–51]
4	Analyze relationship between the gut microbiome and bone health	GM profile data, BMD	Regression	**ML**: RF, LWL, Bagging, RC, Vote **FS**: CAE-RF	[52]
5	Analyze relationship between productivity and human communication in an organization	Speech	Regression	**FE**: Speech features Graph features **ML**: GPR, LR, SVR **FS**: CAE-GPR	[53]
6	Decode human intent using brain- machine interface	EEG, accelerometer, gyroscope	Detection	**FE**: EEG feature Spatio-temporal eigenspectra	[72]
7	Discriminate neutron and gamma-ray using radioactive scintillators	neutron-gamma pulses using CLYC scintillator	Classification Clustering	**FE**: Decay method, ESD, NCC **ML**: K-means	[47, 48]
8.2	Predict postoperative delirium in patients undergoing open heart surgery	EEG	Classification	**FE**: EEG feature Spatio-temporal eigenspectra	[113]
8.1	Identify behavioral task in patients with Parkinson's disease	LFP	Classification	**FE**: MPD **ML**: DHMM-SVM, DPGMM	[205]
8.3	Detect weak signals in a SISO wireless comm scenario	Simulated sequence	Detection	Serial hypothesis tracking Parallel hypothesis tracking	[76]

this topic is fairly large, each study poses specific set of problems belonging to a particular field of interest. We addressed these problems using various SP and ML techniques. In this work, we present an ensemble of techniques that were used in each study. Some of these techniques are novel and the others are either state-of-the-art or modified techniques. These techniques are briefly described in the next chapter under four sections; 2.1 signal feature extraction techniques, 2.2 feature selection techniques, 2.3 ML techniques and 2.4 model accuracy evaluation techniques. The ML techniques have been divided into three subsections; 2.3.1 Classification, 2.3.2 Regression, and 2.3.3 detection techniques. All the simulations and calculations were done in MATLAB (2014-2018) [119] and some ML techniques were employed using WEKA 3.8 [79]. The summary of each study, including the problem type and the tools

used, is shown in Table 1.1 along with contributed publications. In this dissertation, we first describe the various techniques in Chapter 2, that were used in the real-world application studies, described from Chapter 3 onward.

Chapter 2

SIGNAL PROCESSING AND MACHINE LEARNING TECHNIQUES

In this chapter, we present an ensemble of ML and SP techniques that has been used in later chapters. As mentioned before, most machine learning applications require extracting features, selecting best features, using a ML model and evaluating the accuracy of that model. These four functional blocks constitute a machine learning process chain. We have divided this chapter into four sections based on these functional blocks. In section 2.1, we describe various signals that have been investigated in later chapters. Various SP techniques have been described which were used to extract features from these signals. Some feature selection and dimensionality reduction techniques are presented in section 2.2. Section 2.3 presents various classification, regression and detection models that were used in the real-world studies in the later chapters. In the last section we mention a few techniques that are used to evaluate the accuracy of a ML model.

2.1 Signal Feature Extraction Techniques

In this section, we have described various SP techniques used to extract features from various measured data that have been investigated in this dissertation. These include accelerometer signal features in Chapter 3, speech and network graph features in Chapter 5, EEG signal features in Chapters 6 and 8.2, pulse shape discrimination features in Chapter 7, and local field potential (LFP) signal features in Chapter 8.1.

2.1.1 Accelerometer signal features

An accelerometer signal consists of three components; X, Y and Z. These signals have been used to classify physical activity, estimate energy expenditure, study gait patterns etc. Various time and frequency domain features are extracted from each components of acceleration. Features have also been extracted from the resultant acceleration of

- X, Y and Z:

$$R = \sqrt{X^2 + Y^2 + Z^2}$$

and

- Y and Z:

$$T = \sqrt{Y^2 + Z^2}$$

We have listed the features (Table 2.1) that were explored in Chapter 3.

- **Time-domain features**

  Time-domain features like mean, standard deviation, percentiles were previously used as important activity features. We have used some the same features and did a few extensions. We have employed seven sets of time-domain features namely mean, standard deviation, skewness, kurtosis, energy, squared sum of data under 25 percentile, squared sum of data under 75 percentile. The squared sum of data under 25 and 75 percentiles are extracted only from T. So, 22 different time domain features were used to represent each window.

Table 2.1: Summary of features extracted

No. of features	**Feature domain**	**Description**	**Signals used**
4	Time	Mean	X, Y, Z, R
4	Time	Standard deviation	X, Y, Z, R
4	Time	Skewness	X, Y, Z, R
4	Time	Kurtosis	X, Y, Z, R
4	Time	Energy	X, Y, Z, R
1	Time	Squared sum of data under 25 percentile	T
1	Time	Squared sum of data under 75 percentile	T
4	Frequency	Maximum frequency	X, Y, Z, R
4	Frequency	Sum of heights of frequency components below 5Hz	X, Y, Z, R
4	Frequency	Number of peaks in spectrum below 5Hz	X, Y, Z, R
16	Principal components	PCA (1,2,3,4)	X, Y, Z, R
80	time-frequency	'Modified' Wavelet coefficients	X, Y, Z, R

- **Frequency-domain features**

  For frequency-domain feature fast Fourier transform (FFT) was applied on each window and three sets of features were extracted. It has been observed that

most of the signal strength lies between 0-15 Hz, with a bulk of peaks below 5 Hz. Keeping that in mind features like the frequency with maximum magnitude between 1-5 Hz, sum of height of frequency component below 5 Hz, and number of peaks in spectrum below 5 Hz.

- **Principal component features**

  Principal component analysis (PCA) is a method widely used for dimension reduction. It takes the matrix of n windows with $p = 1000$ samples which are correlated and summarizes it by uncorrelated components also known as principal components that are linear combinations of the original p variables. The first k components display as much as possible of the variation among objects. The new principal components have a variance equal to their corresponding eigen values with the first principal component having the maximum eigen value. We have used four principal components per window for the acceleration in each axis and the resultant.

- **Modified wavelet features**

  Wavelet transform provides a time-frequency representation of the signal as it gives an optimal resolution in both time and frequency domains [9]. It is defined as convolution between the signal and a wavelet function. With this approach, the original time-domain signal (maximum frequency f) is initially decomposed into a coarse approximation and detail information by low-pass filtering (bandpass $[0, f/2]$) and high-pass filtering (bandpass $[f/2, f]$), respectively. With wavelet decomposition, the half-band filters are designed to enable perfect reconstruction of the original signal and to avoid aliasing effects. In subsequent levels of decomposition, the approximation signal from the previous level is split into a second approximation and a detail coefficient. This process is repeated to

the desired decomposition level.In our case we have used a 3 level Haar wavelet decomposition, which has one of the simplest functions shown below:

$$\psi(t) = \begin{cases} 1, & \text{if } 0 \leq t < \frac{1}{2} \\ -1, & \text{if } \frac{1}{2} \leq t < 1 \\ 0 & \text{otherwise} \end{cases}$$

After decomposition we got 1000 coefficients for each window. Since all the coefficients are not useful features, 20 coefficients were selected using the Kolmogorov Smirnov (KS) test. The KS test quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples [172]. Here we have used the criteria which quantifies the deviation of a coefficient from normality. Given a coefficient x across all the windows, the test compares the cumulative distribution function of the data $F(x)$ with that of a Gaussian distribution with the same mean and variance $G(x)$. The deviation from normality is quantified by,

$$max(|F(x) - G(x)|)$$

2.1.2 Speech signal features

Speech feature extraction is a common method of reducing the dimension of a speech signal, while maintaining the perceptive power of the signal. We denote $x_i(n), n = 1, ..., W_L$ as the sequence of audio samples of the ith frame, where W_L is the length of the frame. A total of 12 types of features were investigated in Chapter 5 and can be subdivided as follows,

Pitch

Pitch is defined as the rate of periodic vibration of the vocal cords and is also referred

to as the fundamental frequency. Detailed analysis of fundamental frequency (f_0) and its harmonics helps in understanding of emotion dependent pitch modulation observed in expressive speech. The basic problem is to extract the f_0 from a sound signal, which is usually the lowest frequency component, or partial, which relates well to most of the other partials. In a periodic waveform, most partials are harmonically related, meaning that the frequency of most of the partials are related to the frequency of the lowest partial by a small whole-number ratio. The frequencies $\omega = \frac{2\pi}{k} f_0$ are referred to as the harmonics of wave, where k is the propagation constant. There are a lot of ways to estimate the pitch distribution, and in our case we calculated the autocorrelation function of the signal to estimate it [152]. The features, related to pitch distribution are described below.

- *Fundamental frequency (f_0) or pitch frequency*
- *Harmonics*: 1-12 harmonics can be used.
- *Hamonic ratio*: It is a measure of the harmonic to noise ratio, which provides an indication of the overall periodicity of the signal by quantifying the ratio between the periodic (harmonic) and the non-periodic (noise) part [130].

Temporal features

The time-domain features are

- *Zero crossing rate*: It is a measure of dominant frequency or the number of time domain zero-crossings within the speech frame, and is calculated as the rate of sign changes along each frame of the signal [148]. It is computed as,

$$Z(i) = \frac{1}{2W_L} \sum_{n=1}^{W_L} |sgn[x_i(n)] - sgn[x_i(n-1)]|,$$

  where $sgn(.)$ is the sign function.

- *Energy*: Short-time energy distinguishes voiced speech from unvoiced speech and evaluates the amplitude variation and power of the signal for each frame. It is calculated as,

$$E(i) = \sum_{n=1}^{W_L} |x_i(n)|^2$$

- *Energy entropy*: It can be interpreted as a measure of abrupt changes in the energy level [149]. We have used 100 ms of short term window to measure the entropy. It is calculated using the ratio of each sub frame (j) energy and the total energy of the frame.

$$e_j = \frac{E_{subFrame_j}}{E_{Frame_i}}$$

The energy entropy is given by,

$$H(i) = -\sum_{j=1}^{k} e_j \cdot \log_2(e_j)$$

Spectral features

These features are computed in the frequency-domain and they provide a convenient representation of the distribution of the frequency content of the signal. In order to proceed, let $X_i(k), k = 1, ..., W_{FL}$, be the magnitude of the DFT coefficients of the ith audio frame. The different frequency domain features are,

- *Spectral centroid*: It is a measure of the center of gravity of the spectrum of the signal frame. A higher value of spectral centroid correspond to a brighter sound. It is calculated as,

$$C_i = \frac{\sum_{k=1}^{W_{FL}} kX_i(k)}{\sum_{k=1}^{W_{FL}} X_i(k)}$$

- *Spectral spread*: It is the second central moment of the spectrum which measures how the spectrum is distributed around its centroid and calculated as,

$$S_i = \frac{\sum_{k=1}^{W_{FL}} (k - C_i)^2 X_i(k)}{\sum_{k=1}^{W_{FL}} X_i(k)}$$

- *Spectral entropy*: Similar to energy entropy, it computes the abrupt changes in the spectrum [127]. We first divide the spectrum of the short-term frame into L sub-bands (bins). The energy E_f of the fth sub-band, $f = 0, ..., L-1$, is then normalized by the total spectral energy, that is, $n_f = \frac{E_f}{\sum_{f=0}^{L-1} E_f}, f = 0, ..., L-1$. The entropy of the normalized spectral energy n_f is finally computed according to the equation,

$$H = -\sum_{f=0}^{L-1} n_f \cdot \log_2(n_f)$$

- *Spectral flux*: It measures the spectral change between two successive frames and is calculated as the squared difference between the normalized magnitudes of the spectra of the successive frames [159]. It is computed as,

$$Fl_{(i,i-1)} = \sum_{k=1}^{W_{FL}} (EN_i(k) - EN_{i-1}(k))^2$$

where, $EN_i(k)$ is the kth normalized DFT coefficient of the ith frame.

- *Spectral roll-off*: It is defined as the frequency below which 90% of the spectral distribution is concentrated [68]. It helps discriminating between voiced and unvoiced part of speech.

Cepstral features

Cepstral features are calculated by taking the inverse fourier transform of the logarithm of the estimated spectrum of the signal. The power cepstrum has been used widely for speech analysis.

- *Mel-frequency cepstrum coefficients (MFCC)*: These are perhaps the most popular features that has been used successfully in speech recognition problems [70, 133]. It is a type of cepstral representation of signal, where frequency bands are distributed according to the mel-scale. The coefficients are the discrete cosine transform of the mel-scaled log-power spectrum. We have used the first 13 MFCCs as they are considered to carry enough discriminative information.

We have calculated various statistics of these features over periods of time to study the changes in these features over time.

2.1.3 Network graph features

In Chapter 5, we constructed communication network graphs between employees in an organization using their speech recordings. Here we have described few features that can be extracted from a graph. A typical simple graph has been shown in Figure 2.1 with 7 nodes and 10 links. A finite graph can be represented in matrix forms. An adjacency matrix is a square matrix used to represent a finite graph, in which the elements of the matrix indicate whether pairs of vertices are adjacent or not in the graph. In graph theory, a degree of a node is defined as the number of edges incident on the node. A degree matrix is a diagonal matrix which contains information about the degree of each node. Several topological features that aim to describe the nature of daily interactions can be extracted from these graphs. A total of 11 graph features were investigated in this work.

Basic graph descriptors

- *Number of edges*: The total number of edges or links present in the graph.

- *Number of nodes*: The total number of active nodes present in the graph.

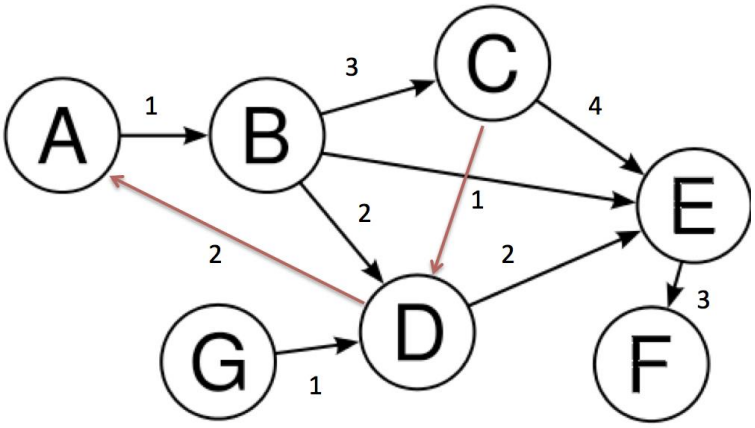


Figure 2.1: Finite graph. (A, B, C ... are nodes and the numbers connecting the nodes denote the weights of the links)

- *Average degree*: It is the number of links per node, and defined as $\frac{2m}{n}$, where n is the number of nodes and m is the number of edges.

- *Number of connected triples*: This routine counts the number of connected triples of nodes. Here it can be defined as a subgraph of three nodes such that there is at least one node among the three which is adjacent to both of the other two nodes.

- *Number of cycles*: This routine calculates the number of independent loops of cycles. It is defined as $m - n + c$, where m is the number of edges, n is the number of nodes and c is the number of connected components.

Graph centrality measures

- *Degrees*: The average number of edges adjacent to a node.

- *Average neighbor degree*: It is a measure of the average degree of adjacent or neighboring nodes for every vertex. In our work we computed took the average of this measure across all nodes.

- *Eigen centrality*: It is the eigenvector corresponding to the largest eigenvalue of the adjacency matrix. The i-th component of this eigenvector gives the centrality score of the i-th node of the network. The average eigen centrality across all nodes was computed for this study.

Laplacian features

These functions concern mostly the spectrum of the adjacency matrix, for which the Laplacian of the graph has to constructed. The Laplacian graph is defined as the difference between the degree and the adjacency matrix. The Laplacian matrix $L_{n\times n}$, is defined as,

$$L = D - A,$$

where, D is the degree matrix and A is the adjacency matrix.

- *Graph spectrum*: It is the list of all the eigenvalues of the Laplacian of the graph. An average over all eigen values was used for this study.
- *Algebraic connectivity*: It is the second smallest eigenvalue of the Laplacian graph.
- *Graph energy*: It is defined as the sum of the absolute values of the real components of the eigenvalues of the graph.

2.1.4 EEG signal features

Electroencephalography (EEG) is used to examine activity in the brain using a net of sensors placed over the scalp to measure electric potential. Modern event related potential studies that intend topographic analysis, often employ 32 to 256 electrodes. Location of such electrodes is given in Figure 2.2. Spatiotemporal information is extracted to study the interaction between these multichannel signals. One method to

study ensemble interaction is through spatiotemporal correlation. EEG signals are not statistically stationary, so the data are analyzed in windows of reasonable lengths under which they can be approximated as a stationary process. Any changes in latent probabilistic models can still be observed through windowing. One way to succinctly capture this is through spatiotemporal eigenspectra, which show fluctuations in correlation within the ensemble over time. This method was particularly useful in seizure prediction [201].

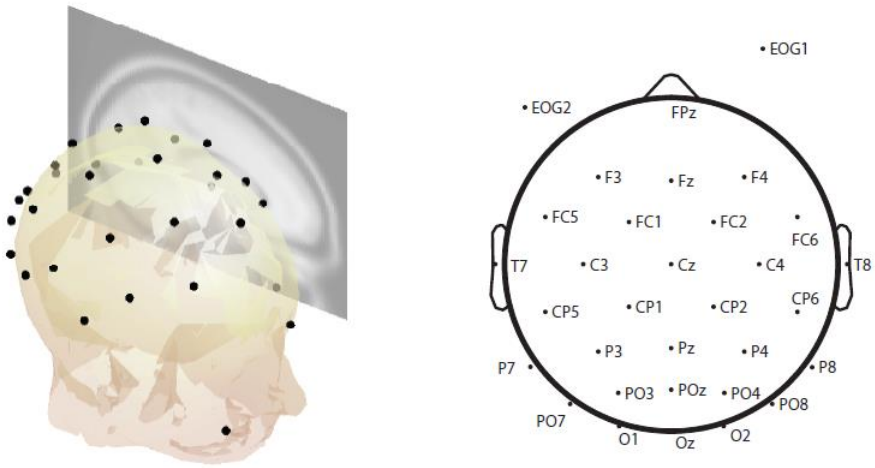


Figure 2.2: 32 channel EEG sensor locations

To compute a spatiotemporal eigenspectrum, space-delay data matrices must be constructed. We first window the data into matrices denoted as $\mathbf{X}_l[j]$, with each row indexed by m containing the m^{th} channel data vector of the form shown in 2.1. l denotes the particular N_w sample long window under consideration that has been delayed by $j\tau$ samples, where j is a multiplier and τ is an incremental delay length.

$$\mathbf{x}_{m,l}[j] = \begin{bmatrix} x_{m,l}[j\tau + 0] & \dots & x_{m,l}[j\tau + N_w - 1] \end{bmatrix} \tag{2.1}$$

Now space-delay data matrices, $\mathbf{Y}_l$, are formulated using 2.2. N_τ is the maximum number of delays considered. N_w must be large enough to properly estimate the spa-

tiotemporal correlation matrix, which is dependent on its dimensionality determined by the number of channels and N_τ. This tends to result in long windows, and poor time resolution. To capture any abrupt and tenuous activity, Because all EEG channels are utilized and N_τ was taken to be 5, and 10 s window was used to maintain as much temporal detail possible. Windows were overlapped by 50%. There is no prior knowledge about the proper delay length to use, so several different lengths were tested. The shortest incremental delay, τ, allowed is 10 ms, which was empirically deemed suitable to compute for our calculations.

$$\mathbf{Y}_l = \begin{bmatrix} \mathbf{X}_l[0] \\ \mathbf{X}_l[1] \\ \vdots \\ \mathbf{X}_l[N_\tau] \end{bmatrix} \tag{2.2}$$

An estimate of the correlation matrix $\mathbf{R}_l$ is constructed by first computing $\mathbf{Z}_l$, a matrix containing centered and normalized rows of $\mathbf{Y}_l$ respective to row means and standard deviations, and then evaluating 2.3.

$$\mathbf{R}_l = \frac{1}{N_w}\mathbf{Z}_l\mathbf{Z}_l^T \tag{2.3}$$

An eigenvalue distribution is formed by extracting eigenvalues from the spatiotemporal correlation matrix, $\mathbf{R}_l$, and ordering them from largest to smallest. Eigenvalue distributions are computed from all windows of data considered. Arranging them in chronological order produces a spatiotemporal eigenspectrum. Typically, the eigenvalues decay in strength quickly, so to see details at the lower ranks, the natural log of the eigenvalues is evaluated.

2.1.5 LFP signal features (Matching pursuit decomposition)

Electrical events at deeper locations in the brain which can be recorded by inserting metal or glass electrodes, or silicon probes into the brain are called LFP (also known as micro-EEG). LFP differs from normal EEG or electrocorticography (ECOG) and ranges between 5–1,000 μV with frequency less than 200 Hz. The LFP signal is the most informative brain signal as it contains action potentials and other membrane potentials-derived fluctuations in a small neuron volume [23].

As the frequency content of the LFP signals changes with time, we apply the matching pursuit decomposition (MPD) algorithm to extract time-frequency based features [116], [26]. Using the MPD, a signal is decomposed into a linear expansion of Gaussian basis functions that are selected from a redundant basic dictionary. Each dictionary element is a Gaussian signal that is a time-shifted, frequency-shifted and scaled version of a basic Gaussian signal at the time-frequency origin. The feature obtained from each extracted dictionary element is a four-dimensional (4-D) vector consisting of the extraction weight coefficient, time shift, frequency shift and scale change parameters.

The MPD is a well known technique for sparse signal representations. It is a greedy algorithm that expands a signal into a linear approximation of basis functions by iteratively projecting the signal over a redundant, possibly non-orthogonal set of signals called dictionary. Since it is a greedy algorithm, the approximation may be sub-optimal. The dictionary functions are iteratively selected to best match the signal structure, resulting in a sub-optimal expansion.

In general any basis function can be used as a dictionary to decompose the required signal, can be shown that the only signal that achieves the lower bound is the Gaussian signal. However, the Gaussian signal is most often selected since it attains

Heisenberg's uncertainty principle. According to this principle, a signal cannot simultaneously achieve high temporal resolution and high frequency resolution; the signal that achieves the best trade-off in both time and frequency resolution is the Gaussian signal. This can be shown by computing the time-bandwidth product of a Gaussian signal,

$$x(t) = e^{-bt^2} \implies T_x F_x = \frac{1}{4\pi}$$

where T_x and F_x are the duration and bandwidth of the Gaussian signal, respectively. For all other signals it can be shown that $T_x F_x > (1/4\pi)$. The Gaussian dictionary element $g_\gamma(\mathrm{t})$, $\gamma = 1, \ldots, \Gamma$ is given by

$$g_\gamma(t) = \kappa e^{2\pi\kappa(t-\tau)^2} e^{j2\pi t\nu}$$

This forms a dictionary $\mathcal{D}$ with Γ independent Gaussian waveforms. The MPD algorithm begins by projecting the signal on each dictionary signal $g_\gamma(t)$ and computing the residue after every iteration. After P iterations, the MPD results in a linear weighted expansion of P selected Gaussian elements, $g_p(t)$ and their corresponding weight coefficients α_p, $p = 1, \ldots, P$. This is given by,

$$x(t) = \sum_{p=1}^{P} \alpha_p g_p(t) + r_P(t) \tag{2.4}$$

where $r_P(t)$ is the residual signal after P iterations. The iterations start by setting $r_0(t) = x(t)$; at the Pth iteration, the best matched Gaussian signal $g_p(t)$ is selected that results in the maximum correlation with the remainder signal $r_p(t)$. In particular,

$$g_p(t) = \underset{g_\gamma(t)\in\mathcal{D},\ \gamma=1,\ldots,\Gamma}{\arg\max} \left| \int r_{p-1}(t) g_\gamma^*(t) dt \right|$$

The weight coefficient is then obtained using,

$$\alpha_p = \left| \int r_{p-1}(t) g_p^*(t) dt \right| = |\langle r_{p-1}(t),\ g_p(t) \rangle \tag{2.5}$$

With this choice $r_{p-1}(t)$ is projected onto $g_p(t)$ and decomposed as follows:

$$r_{p-1}(t) = r_p(t) + \alpha_p g_p(t), \tag{2.6}$$

From Equation (2.6), we can see that the decomposition of $x(t)$ is given by (2.4) where α_p is given by (2.5). It can be shown that $r_P(t)$ converges exponentially to 0 when p tends to ∞

$$\lim_{p\to\infty} \|r_P(t)\| = 0;$$

Hence,

$$x(t) = \sum_{k=0}^{\infty} |\langle r_k(t),\ g_k(t)\rangle| g_k(t),$$

and,

$$|x(t)|^2 = \sum_{k=0}^{\infty} |\langle r_k(t), g_k(t)\rangle|^2,$$

Thus, the original vector $x(t)$ is decomposed into a sum of the dictionary signals that matches best the signal and its residuals at each iterations. It can also be seen that the decomposition preserves the signal energy assymptotically. After feature extraction, the pth feature vector, $p = 1, \ldots, P$, is given by $F_p = [\alpha_p\ \tau_p\ \nu_p\ \sigma_p]^T$, corresponding to the MPD amplitude, time-shift, frequency-shift and scale change parameters.

2.1.6 Features for Pulse Shape Discrimination

In Chapter 7, we use various features to discriminate between neutron and gamma-ray pulses. The motivation behind the study is discussed in details in that chapter. For this subsection, we describe few features that were used to discriminate the shape of the two radiation pulses.

- **Decay time extraction**

  A neutron pulse has a significantly longer decay time than that of a gamma-ray pulse [74]. This algorithm has been used to discriminate between neutron and gamma-ray pulses. It captures the decay time of every pulse with fairly simple calculations. In this case, the time taken for the pulse to decline from its peak to 64% of its amplitude is defined as the decay time. The reason for choosing 64% is the fact that it showed the best result while discriminating between neutron and gamma ray pulses. The algorithm is explained as follows,

  - For a given pulse, the maximum amplitude is recorded.
  - Since the base of the pulse is not zero, its minimum amplitude is recorded. Then 36% of the span (Max–Min) is calculated.
  - The time taken from the Max to a reduction by 36% of the span is defined as the decay time.

  If P is a pulse and S is the magnitude, $|max(P) - min(P)|$, then the decay time is the time taken for the pulse from its peak to 36% of S.

  Figure 2.3 shows an illustration of the decay time calculation for a typical neutron and gamma ray event. It is seen that the decay time for a gamma ray is 76 ns and that of a neutron is 216 ns. Therefore, it is clear that there is a significant difference in the decay times for a gamma ray pulse and a neutron pulse. Thus, it can be said that the decay time can serve as an important parameter to discriminate neutrons and gamma rays.

- **Energy spectral density**

  The energy spectral density (ESD) for a signal is a measure of its energy distribution as a function of frequency [187]. It is a useful concept that allows us

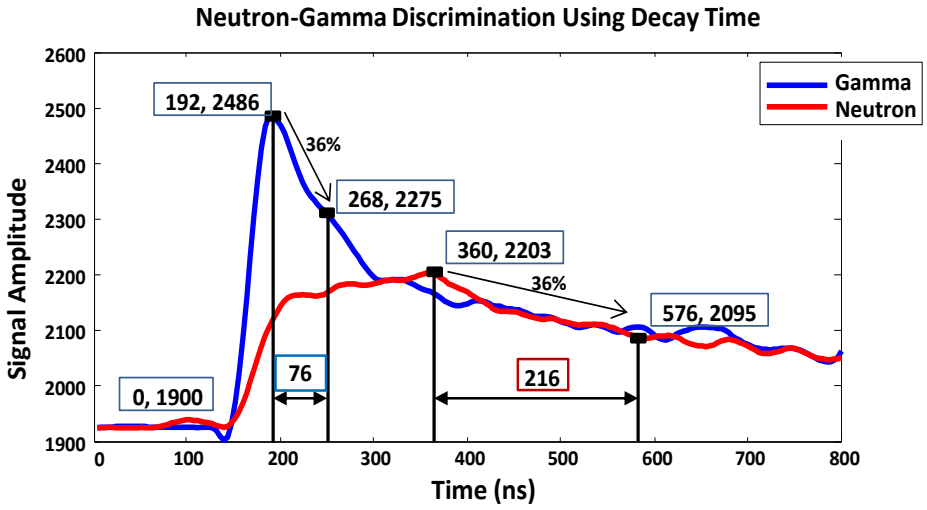


Figure 2.3: Comparison of decay time determination for neutron and gamma ray pulses

to determine the bandwidth associated with each neutron or gamma ray pulse. It is also a measure of the signal energy $E = \int \|f(t)\|^2$, where $f(t)$ is the input signal. A neutron or a gamma-ray pulse can be viewed as a signal that decays exponentially and hence has a finite energy. So it can be called as an energy signal. The power of an energy signal is zero, because it has finite energy in infinite time. The Fourier transform of the autocorrelation function of an energy signal is equal to the energy spectral density of the signal. So the ESD of a pulse (neutron or gamma-ray) possesses important information about the pulse, which might be a useful parameter to discriminate between a neutron and a gamma ray.

If $f(t)$ is the given pulse (gamma ray or neutron) the energy dissipated is,

$$E_f = \int_{-\infty}^{\infty} f^2(t)dt$$

Using Parseval's theorem for Fourier transforms we get,

$$\Delta E_f = \frac{1}{2\pi}\|F(\omega)\|^2 \Delta\omega$$

If we add up all the energy increments, we obtain the total energy. The function $\|F(\omega)\|^2$ is referred to as the signal ESD, and it conveys where the energy is distributed in frequency.

- **Normalized cross-correlation**

In signal processing, the cross correlation is a measure of similarity between two series, expressed as a function of the time lag (τ) relative to one another. If x and y are two continuous functions, the cross correlation is defined as

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t+\tau)y(t)dt = \int_{-\infty}^{\infty} x(t)y(t+\tau)dt$$

The cross correlation is such that $R_{xy}(-\tau) = R_{yx}(\tau)$. If x and y are the same real signal, then R_{xx} is termed the autocorrelation function. We shall refer to the second signal as the reference, i.e., in R_{xy}, y is the model of the pulse template. Numerically, the cross correlation between sampled signals x_i and y_i is calculated by,

$$R_{xy}(k) = \begin{cases} \sum_{i=1}^{N-k} x_{i+k}y_i, & k = 0, 1, ... \\ \sum_{i=1}^{N-k} y_{i-k}x_i, & k = 0, 1, ... \end{cases}$$

where N is the series length, and k is the lag. In our work, we systematically analyze different cases of cross correlation between the reference gamma-ray and neutron pulses.

2.2 Dimensionality Reduction and Feature Selection Techniques

Feature selection is different from dimensionality reduction. Both methods seek to reduce the number of features/attributes in the dataset, but a dimensionality reduction method do so by creating new combinations of features, where as feature selection methods include and exclude attributes present in the data without changing them.

2.2.1 Principal component analysis

Principal component analysis (PCA) is a method widely used for dimension reduction [93]. PCA takes the matrix of n windows with p samples which are correlated and summarizes it by uncorrelated components, also known as principal components, which are linear combinations of the original p variables. The first k components selected reveal the largest variation among objects. One can express each of the p variables as a linear combination of m factors (F_i)

$$X_j = A_{1j}F_1 + A_{2j}F_2 + ... + A_{mj}F_m + U_j$$

where X_j is the jth dimension of the vector and U_j is the variance that is unique to variable j which cannot be explained by any of the common factors. Each component is estimated as a weighted sum of the p variables. The ith component is thus

$$F_i = w_{i1}X_1 + w_{i2}X_2 + ... + w_{ip}X_p$$

The first weight vector $\mathbf{w}_{(1)}$ is calculated using the following condition,

$$\mathbf{w}_{(1)} = \arg\max_{\|\mathbf{w}\|=1}\{\sum_i (\mathbf{x}_{(i)} \cdot \mathbf{w})^2\}$$

Or equivalently, writing this in matrix form gives,

$$\mathbf{w}_{(1)} = \arg\max_{\|\mathbf{w}\|=1}\{\|\mathbf{X}\mathbf{w}\|^2\} = \arg\max_{\|\mathbf{w}\|=1}\{\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w}\}$$

since $\mathbf{w}_{(1)}$ has been defined to be a unit vector, it equivalently also satisfies,

$$\mathbf{w}_{(1)} = \arg\max\{\frac{\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w}}{\mathbf{w}^T\mathbf{w}}\}$$

A standard result for a symmetric matrix such as $\mathbf{X}^T\mathbf{X}$ is the quotient's maximum possible value or the largest eigenvalue of the matrix, which occurs when $\mathbf{w}$ is the corresponding eigenvector. With $\mathbf{w}_{(1)}$ found, the first component of a data vector $X_{(i)}$ can then be given as a score $F_{1(i)} = X_{(i)}$. The kth component can be found by subtracting the first $k-1$ principal components from $\mathbf{X}$:

$$\hat{\mathbf{X}}_k = \mathbf{X} - \sum_{s=1}^{k} \mathbf{X}\mathbf{w}_{(s)}\mathbf{w}_{(s)}^T$$

In this manner, the components are calculated, with the first component associated with the maximum weight and the last component with the minimum.

2.2.2 Principal component analysis with KS test

In this subsection, we present a modified version of the PCA algorithm. Here we first select points from each window of signal using KS test, after which the principal component analysis is applied on those points. Using KS test we select the samples which are most deviant from the standard normal distribution of the all the samples.

2.2.3 Sequential forward selection

Sequential Forward Selection (SFS) is a popular and simple algorithm for feature selection. It is a greedy search algorithm, ideal for cases where it is desirable to minimize the number of features. The algorithm starts with an empty set of features and iteratively adds new features to the optimal set, in order to maximize a certain objective function. It works in conjunction with the classification or clustering method, updating the optimal subset of features after testing each model generated by a given

subset of features. A common objective function is given by:

$$\mathrm{rr}(\boldsymbol{P}) = \frac{\sum_{c=1}^{C} \mathrm{rr}_c(\boldsymbol{P})}{C},$$

where rr_c is the recognition rate of the cluster c. In other words, the objective function is given by the mean of the main diagonal of the confusion matrix. A classifier is required to compute the confusion matrix at every point. Figure 2.4 shows the block diagram for SFS.

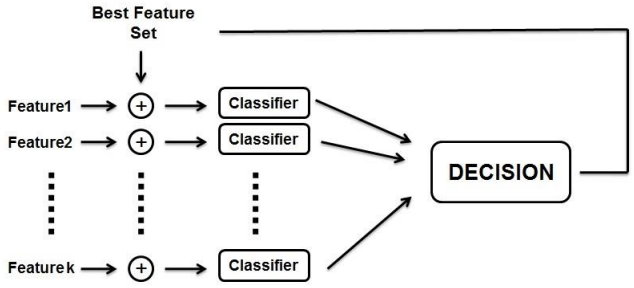


Figure 2.4: Block Diagram for SFS

2.2.4 Classifier attribute evaluation (CAE)

The CAE is a technique by which each feature in the dataset is evaluated in the context of the output variable. The search method is the technique by which to try or navigate different combinations of attributes in the dataset in order to arrive on a short list of chosen features. A ranker search method is used which evaluates each feature and lists them in a rank order. In case of regression, the root mean square error (RMSE) of the correlation coefficient (CC) is used to evaluate each feature and rank them. A base classifier is required to evaluate the RMSE or CC.

2.3 Machine Learning Techniques

In this section, we have described various ML modeling techniques used in classification, regression and detection problems. This section contains mostly state-of-the-art techniques along with some modified and novel techniques.

2.3.1 Classification Algorithms

(I) **K-means algorithm**

Probably the most famous clustering formulation is k-means. k-Means is in the family of assignment based clustering [95]. Each cluster is represented by a single point, to which all other points in the cluster are "assigned". Consider a set X, and distance $\mathbf{d} : X \times X \to \mathbb{R}_+$, and the output is a set $Z = \{z_1, z_2, ..., z_k\}$. This implicitly defines a set of clusters where $\phi_C(x) = \arg\min_{c \in C} \mathbf{d}(x, c)$. Then the k-means clustering problem is to find the set C of k clusters (often, but always as a subset of X) to

$$\sum_{x \in X} \mathbf{d}(\phi_C(x), x)^2$$

So we want every point assigned to the closest center, and want to minimize the sum of the squared distance of all such assignments.

(II) **Gaussian mixture model**

Gaussian mixture models (GMM) are one of the most commonly used classifier, which are statistical models where the distribution of data takes the form of a mixture, i.e, a linear combination of several probability density functions. In a GMM, the basic probability functions are multivariate Gaussians, where the

probability of the mixture is given by:

$$P(x) = \sum_{m=1}^{k} a_m \mathcal{N}(\mu_m, \Sigma_m),$$

where M is the number of Gaussians in the mixture and a_m, μ_m, Σ_m denote the weight, mean and covariance matrix for each gaussian m.

GMMs have been extensively used for supervised classification problems, in which a single GMM can model a class. To create a model, one needs to find the optimal values for a, μ and Σ for all Gaussians in the mixture, and this task is accomplished here by using the Expectation-Maximization (EM) procedure [129].

One shortcoming of this method is assuming that each class has a Gaussian distribution, which is seldom the case. Besides, this method is very sensitive to a good initialization for the Gaussian parameters which is usually performed by first applying the K-means algorithm. K-means is one the simplest unsupervised learning algorithms and it is a fast, robust, easier to understand and relatively efficient algorithm that gives best result when data is distinct or well separated. However, it has some disadvantages like its inability to handle noisy data, outliers and non-linear data. Also there's always the need to specify the number of clusters apriori and initialize them properly. These centers should be placed in a cunning way because different locations generate different results. This means random intialization cannot be used if the goal is to compare different approaches, like what is done on the SFS algorithm. One way to overcome random initialization is to use the Linde-Buzo-Gray (LBG) algorithm [145], which is a slight modification of the generic K-means algorithm, in which initially just one center (the sample mean) is selected, and centers are iteratively divided until the desired number of centroids is achieved. For this reason, this

algorithm only works with M center values which are a power of 2. In this study , we used a modified version of the LBG algorithm, where the center division is performed differently for each number of gaussians M. For M values that are a power of 2, the center division is given by:

$$\mu_1 = \mu_0 + \epsilon.\sigma_0$$

$$\mu_2 = \mu_0 - \epsilon.\sigma_0,$$

where μ_0 is the split cluster's center and σ_0 is the cluster's standard deviation. ϵ was empirically set to 0.4 in this study. For M values that are a power of 3, the expression is modified to:

$$\mu_1 = \mu_0$$

$$\mu_2 = \mu_0 + \epsilon.\sigma_0$$

$$\mu_3 = \mu_0 - \epsilon.\sigma_0.$$

For other M values, only the center with the highest standard deviation is chosen to be split, and this operation is repeated until the desired number M is achieved.

(III) **Hidden Markov model**

Hidden Markov Models (HMM) are statistical models related to Markov processes. In this model, the occurrence of a sequence of events is due to the events having gone through a sequence of (hidden) states, each emitting an observation. This way, a HMM is defined by a model $\lambda = [Q, A, \pi, b_q(x)]$, where Q is the number of states, A is the state transition matrix, π is the initial state probability vector and $b_q(x)$ is the observation pdf for each state q. The optimal values for these parameters are estimated from the data using

the Baum-Welch algorithm [151], which is a modification of the Expectation maximization (EM) algorithm. Like GMM, HMM can be used for supervised and unsupervised classification problems, where the former models each class as a single λ and, the latter, as a state. HMM gives the ability to model each class as Gaussian mixtures (GMs) for unsupervised case.

For input $x_1, ..., x_n$, the labels corresponding to x_i, $\hat{z_1}, ..., \hat{z_n}$ are estimated for which the joint distribution of x_i and z_i is maximum, given by,

$$\{\hat{z_1}, ..., \hat{z_n}\} = \underset{\{z_1,...,z_n\}}{\arg\max}\, p(x_1, ..., x_n, z_1, ..., z_n)$$

where joint probability is given by,

$$p(x_1, ..., x_n, z_1, ..., z_n) = \prod_1^n q(z_i|z_{i-2}, z_{i-1}) \prod_1^n e(x_i|z_i)$$

Here, we have used the GMs as the probability distribution function of the HMM output. So the output distribution is a single multivariate Gaussian with mean μ_j and covariance matrix Σ_j.

$$e(x|z_i) = N(x; \mu_j, \Sigma_j)$$

As discussed in the previous section, the Expectation Maximization algorithm is used to estimate the parameters of the above distribution. We have used the Viterbi approximation process [65] to calculate,

$$\underset{\{z_1,...,z_n\}}{\arg\max}\, p(x_1, ..., x_n, z_1, ..., z_n),$$

which just considers the most likely path, instead of summing over all possible state sequences.

(IV) **Support vector machine**

A Support Vector Machine (SVM) is a classifier formally defined by a separating hyper-plane [11]. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper-plane which categorizes new examples. In two dimensional space this hyper-plane is a line dividing a plane in two parts where in each class lay in either side.

Let x be a feature vector (i.e., the input of the SVM), where $x \in \mathbb{R}_n$ and n is the dimension of the feature vector; z be the class (i.e., the output of the SVM), where $y \in \{-1, 1\}$, i.e. the classification task is binary. $\mathbf{w}$ and b be the parameters of the SVM and we need to learn them using the training set. $(x^{(i)}, y^{(i)})$ be the ith sample in the dataset. Let's assume we have N samples in the training set. With $n = 2$, one can represent the SVM's decision boundaries as shown in Figure 2.5.

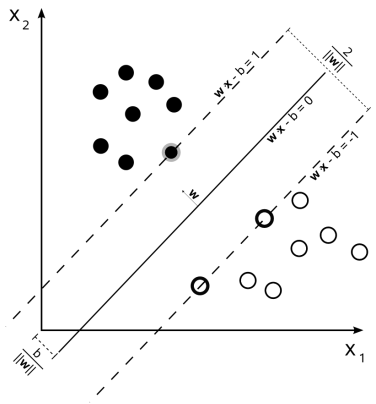


Figure 2.5: Support vector machine classifier

The class z is determined as follows,

$$z^{(i)} = \begin{cases} -1, & \text{if} \mathbf{w}^T x^{(i)} + b \leq -1 \\ 1, & \text{if} \mathbf{w}^T x^{(i)} + b \geq 1 \end{cases}$$

While training, we tune and regularize the SVM parameters to maximize the distance between the two decision boundaries.

(V) **Hidden Markov model- support vector machine**

We consider discrete HMMs (DHMMs), where the state space of the hidden variables is discrete. The observation symbol to train the HMMs for detecting and identifying the behavioral tasks is the feature β. Different HMMs are trained corresponding to different types of behaviors. For the $\Lambda^{(m)}$ DHMM, we specify the initial state distribution vector, hidden state transition matrix, and state-dependent observation density matrix as $\lambda^{(m)} = [\pi^{(m)}, A^{(m)}, B^{(m)}]$, respectively. The maximum-likelihood estimate for $\lambda^{(m)}$ is obtained based on the Baum-Welch algorithm

$$\lambda_{\mathrm{ML}}^{(m)} = \arg\max_{\lambda^{(m)}} \log P(\beta|\lambda^{(m)}, \Lambda^{(m)}) \,.$$

After i iterations,

$$\lambda_{(i+1)}^{(m)} = \arg\max_{\lambda^{(m)}} \sum_{H} P(H|\beta, \lambda^{(m)}, \Lambda^{(m)}) \cdot \log P(H, \beta|\lambda_{(i)}^{(m)}, \Lambda^{(m)}) \,.$$

Here, H indicates the hidden states. The sigma (sum) over H indicates the overall possible state sequences in the HMM. The probability of the test is calculated as

$$P(\beta|\lambda, \Lambda) = \sum_{H} \pi H_1 \prod_{n=1}^{NA_{\mathrm{obs}}} aH_n, H_{n+1} \prod_{n=1}^{NA_{\mathrm{obs}} bH_n(\beta_n)}$$

where πH_1 is defined as the initial state probability of state H_1. Also, the state transition probability from sate H_n to H_{n+1} is defined as aH_n, H_{n+1}; $b_{H_n}(\beta_n)$ is the probabilty from observation β_n to state b_{H_n}.

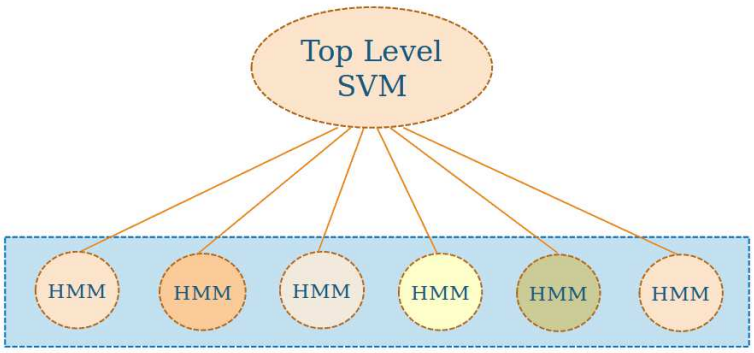


Figure 2.6: Independent DHMMs fused by the top level SVM classifier

Fusing HMMs by Top Level SVM Classifier

Since the features have temporal patterns, we used HMM models as a good candidate to model and classify cognitive tasks patterns. Based on the probability nature of HMMs and the random assignment of prior probabilities, the HMMs may lead to different classification results in different runs of the model. In our first approach, to avoid the randomness and improve the performance of the classifier, we trained several DHMM classifiers for each behavior and put the log-likelihood returned by all DHMMs in a new vector. The resulted vector is given as an input of a SVM to decide between the outputs of the DHMMs. According to [186], the objective function of the SVM classifier is given by

$$\min_{w,b} \frac{\|w\|^2}{2}, \quad \text{subject to: } \forall i,\ y_i(w\varphi_i - b) \geq 1$$

The top-level SVM fuses all the low-level DHMMs and returns one value as demonstrated in Figure 2.6. This approach improved the recognition result for seven percent comparing to the previous work [91].

Hybrid SVM-HMM Model

In the second alternative, we trained a hybrid SVM-HMM model which is an

implementation of structural SVMs for sequence tagging based on a combination of SVM and HMM [3]. It handles dependencies between neighboring labels by Viterbi decoding. In contrary to basic HMM, the learning procedure is based on a maximum soft margin criterion but it also shares the major advantages with other discriminative methods, specially the potential to deal with overlapping features. In the SVM-HMM, models that are isomorphic to an kth-order HMM are discriminatively trained utilizing the structural SVM concept. Considering an input sequence $\{\mathbf{x}_1, \ldots, \mathbf{x}_l\}$ of feature vectors $\mathbf{x}_i$, $i = 1, \ldots, l$, the SVM-HMM tries to train a model that predicts a tag sequence $\{y_1, \ldots, y_l\}$ by applying the linear discriminant function

$$y = \arg\max_{y} \sum_{i=1}^{l} \sum_{j=1}^{k} \mathbf{x}_i \mathbf{w}_{y_{i-j},\ldots,y_i} + \boldsymbol{\phi}_{\mathrm{tr}}(y_{i-j}, \ldots, y_i)\mathbf{w}_{\mathrm{tr}}$$

where $\mathbf{w}_{y_{i-j},\ldots,y_i}$ is the emission weight vector learned for each different jth-order tag sequence $\{y_{i-j}, \ldots, y_i\}$, and $\mathbf{w}_{\mathrm{tr}}$ is the transition weight vector for the transition weights between adjacent tags. The indicator vector $\boldsymbol{\phi}_{\mathrm{tr}}(y_{i-j}, ..., y_i)$ has an entry set to 1 for the sequence $\{y_{i-j}, \ldots, y_i\}$. If the training entries are $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})$, the feature vectors are $\mathbf{x}_i^{(j)}$, $i = 1, \ldots, l$ and the training sequence tags are $y_i^{(j)}$, $i = 1, \ldots, l$, then for a model with first-order transitions and zero-oder emissions the optimization problem is

$$\min \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{C}{n}\sum_{i=1}^{n}\xi_i$$

such that

$$\forall y : \left[\sum_{i=1}^{l}(x_i^1 W_{y_i^1}) + \varphi_{\mathrm{tr}}(y_{i-1}^1, y_i^1)W_{\mathrm{tr}}\right] \geq \left[\sum_{i=1}^{l}(x_i^n W_{y_i}) + \varphi_{\mathrm{tr}}(y_{i-1}, y_i)W_{\mathrm{tr}}\right] + \Delta(y^n, y) - \xi_n$$

Here, C is the trade-off parameter for the margin size and training error.

(VI) **Dirichlet process Gaussian mixture model**

Given a data or feature vector $\mathbf{x}$, a mixture model is described by the PDF:

$$p(\mathbf{x}|\theta_1, ..., \theta_M) = \sum_{m=1}^{M} \omega_m p(\mathbf{x}; \theta_M),$$

where $\omega_1, \ldots, \omega_M$ are the individual mixture component weights, θ_m is the parameter space representing the PDF $p(x|\theta_m)$ and M is the number of mixture components. The goal of the mixture model is to estimate M, ω_m and θ_m, $m = 1, \ldots, M$ that best fit the data. In this case, $p(\mathbf{x}; \theta_m) \sim \mathcal{N}(\mathbf{x}; \mu_m, \Sigma_m)$ is a Gaussian PDF with a parameter vector consisting of the mean and the covariance of the PDF, $\theta_m = [\mu_m \ \Sigma_m]^\intercal$. Therefore, the complete representation of clustering in the space is:

$$p(\mathbf{x}|w, \mu, \Sigma) = \sum_{m=1}^{M} \omega_m \mathcal{N}(\mathbf{x}; \mu_m, \Sigma_m).$$

In a Bayesian sense, a prior PDF must be selected in order to determine the non parametric statistics related to a data-set. The DP is a prior in which a base distribution, G_0, and an innovation parameter α, fully describe the DP. The innovation parameter effectively describes how likely a new data point is to be placed within a cluster of previous data or in a newly formed cluster. In the stick-breaking algorithm of the DP [88], this corresponds to how fine the breaks or cluster divisions are made. Consider the distribution:

$$G \sim \ DP(\alpha, G_0) \quad \theta_n|G \sim \ G, n = 1, \ldots, N\,.$$

Theoretically, this is an infinite number of weights, with many tiny weights, but practically, the model is truncated to a sufficient number of components to achieve a certain error, and some classes are empty. The tendency to have large weights is controlled by α.

The final model of the data, $\mathbf{x}_n$, $n = 1, \ldots, N$ is:

$$\begin{aligned}
&\boldsymbol{\theta}_m \sim\ G_0, \quad v_i \sim\ \text{Beta}(1, \alpha)m, i = 1, \ldots, \infty \\
&w_m = v_m \prod_{i=1}^{m-1} (1 - v_i), m = 1, ..., \infty. \\
&c_n|\mathbf{w} \sim\ \text{Categorical}(\mathbf{w}), n = 1, \ldots, N \\
&\mathbf{x}_n|c_n \sim\ p(\mathbf{x}_n|\boldsymbol{\theta}_{c_n}), n = 1, \ldots, N
\end{aligned}$$

where $\boldsymbol{\theta}_m$ are the Gaussian parameters, w_m are the weights that sum to 1, forming a discrete distribution on the counting numbers up to a truncation limit M, which c_n is drawn from to specify the Gaussian for each data point $\mathbf{x}_n$ to be drawn from.

As we do not have access to the true posterior distribution, we employ blocked Gibbs sampling [88]. After deciding the truncation limit M and using the conjugate prior for Gaussian parameters, the Normal-Wishart, we cyclically sample blocks of parameters: all of the $\boldsymbol{\theta}_m$, all of the c_n, and all of the w_m, which explores the space an iteratively trends toward drawing samples from the posterior distribution, with appropriate burn-in iterations

$$p(\boldsymbol{\theta}_m|c_1, \ldots, c_M, \mathbf{x}_1, \ldots, \mathbf{x}_N) \propto G_0(\boldsymbol{\theta}_m)$$

$$\begin{aligned}
c_n^{(i)} \sim\ & p(c_n|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M, c_1, \ldots, c_M, \mathbf{x}_1, \ldots, \mathbf{x}_N) \\
& = \sum_{m=1}^{M} w_m\, p(\mathbf{x}_n|\boldsymbol{\theta}_m)\delta(c_n - m),\ n = 1, \ldots, N
\end{aligned}$$

$$p(v_m|c_1^{(i)}, \ldots, c_M^{(i)}) = \text{Beta}(1 + N_m^*, \alpha + \sum_{m'=m+1}^{M} N_{m'}^*),\ m = 1, \ldots, M.$$

$$w_m = v_m \prod_{j=1}^{m-1} (1 - v_j),\ m = 1, \ldots, M.$$

where $n : c_n = m$ is the set of n such that $c_n = m$, N_m^* is the number of such n for that m, and G_0 is Normal-Wishart.

2.3.2 Regression Techniques

We have used various regression techniques in chapters 4 and 5. They are described below.

(I) **Linear regression**

Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables. One variable, denoted x, is regarded as the predictor, explanatory, or independent variable, and the other variable, denoted y, is regarded as the response, outcome, or dependent variable. If the goal is prediction, or forecasting, or error reduction, linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. Thus the model takes the form,

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^{\mathsf{T}} \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where ε is the error term. The matrix notation can be given by

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

(II) **Gaussian process regression**

It implements Gaussian processes with polynomial kernel for regression without hyperparameter-tuning [114].

(III) **Support vector regression**

It implements Alex Smola and Bernhard Scholkopf's sequential minimal optimization algorithm for training a support vector regression model [167, 173].

(IV) **Decision tree**

Decision tree learning uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modelling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.

(V) **Random forest**

The RF model is one of the most effective machine learning models for predictive analytics [17]. The RF model makes predictions by combining decisions from a sequence of base models. More formally we can write this class of models as:

$$g(x) = f_0(x) + f_1(x) + f_2(x) + ... \tag{2.7}$$

where the final model g is the sum of simple base models f_i. Here, each base model is a simple decision tree. This broad technique of using multiple models to obtain better predictive performance is called model ensembling. In random forests, all the base models are constructed independently using a different subsample of the data.

(VI) **Bagging**

Bagging [16] is also an ensemble base classifiers, and the RF model has been used in our work.

(VII) **Locally weighted learning**

It is a form of memory-based learning method, in which a collection of simple local models are used to model the data [8]. A local base model and a nearest neighbor search algorithm are required to implement the LWL. In this problem, we have used the RF model as the base model and a linear nearest neighbor search algorithm.

(VIII) **Random committee**

It is an ensemble of randomizable base classifiers. Each base classifiers is built using a different random number seed and the final prediction is a straight average of the predictions generated by the individual base classifiers. Here the RF model is used as the base model.

(IX) **Vote** This method combines all the above-mentioned four models.

2.3.3 Detection

In this subsection, we present two binary detection algorithms involving log-likelihood ratio test which has been used in section 8.3, to detect a signaling sequence in a wireless communication scenario.

We consider a binary detection problem of the form

$$\mathcal{H}_0 : \mathbf{z} = \mathbf{s}_0 * \mathbf{h} + \mathbf{w} \tag{2.8}$$

$$\mathcal{H}_1 : \mathbf{z} = \mathbf{s}_1 * \mathbf{h} + \mathbf{w}, \tag{2.9}$$

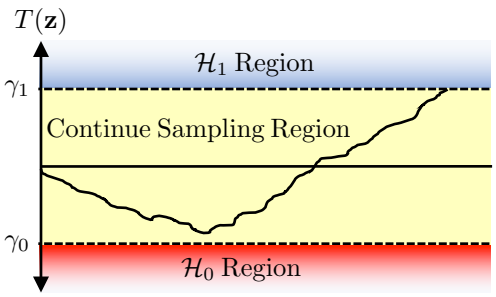


Figure 2.7: Illustration of the sequential detection process. The cumulative log-likelihood ratio is computed for each new sample and compared to the thresholds γ_0 and γ_1. If no decision is made the receiver will continue to sampling new points until a decision is made.

where '*' denotes the convolution operator, $\mathbf{s}_0 \in \mathbb{C}^{1\times N}$ denotes a channel probing sequence, $\mathbf{s}_1 \in \mathbb{C}^{1\times N}$ denotes a signaling sequence, $\mathbf{h} \in \mathbb{C}^{1\times L}$ denotes the complex channel attenuation coefficients, and $\mathbf{w} \in \mathbb{C}^{1\times(N+L-1)}$ denote zero mean thermal noise drawn from the complex Gaussian distribution ($\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\,\mathbf{I})$). We assume that the channel coefficients, $\mathbf{h}$, and noise variance, σ^2, are estimated before detection begins at the receiver. Given these assumptions, we form the log-likelihood ratio test [97]

$$\log\{L(\mathbf{z})\} = \log\left\{\frac{p(\mathbf{z};\mathcal{H}_1)}{p(\mathbf{z};\mathcal{H}_0)}\right\} \tag{2.10}$$

$$p(\mathbf{z};\mathcal{H}_1) = \frac{1}{(\pi\,\sigma^2)^N}\,\exp\left\{-\frac{1}{\sigma^2}\,\|\mathbf{z}-\tilde{\mathbf{s}}_1\|^2\right\} \tag{2.11}$$

$$p(\mathbf{z};\mathcal{H}_0) = \frac{1}{(\pi\,\sigma^2)^N}\,\exp\left\{-\frac{1}{\sigma^2}\,\|\mathbf{z}-\tilde{\mathbf{s}}_0\|^2\right\}, \tag{2.12}$$

where $\tilde{\mathbf{s}}_i = \mathbf{s}_i * \mathbf{h}$ for $i \in \{0,1\}$.

The channel coefficients and noise variance are computed using the maximum

likelihood estimates (MLE) [14]

$$\mathbf{h} = \mathbf{z}\,\mathbf{S}_0^{\dagger}\,(\mathbf{S}_0\,\mathbf{S}_0^{\dagger})^{-1} \tag{2.13}$$

$$\sigma^2 = \frac{1}{N}\|\mathbf{z}\,\mathbf{P}_{\mathbf{s}_0}^{\perp}\|^2 \tag{2.14}$$

$$\mathbf{P}_{\mathbf{s}_0}^{\perp} = \mathbf{I} - \mathbf{S}_0^{\dagger}\,(\mathbf{S}_0\,\mathbf{S}_0^{\dagger})^{-1}\,\mathbf{S}_0 \tag{2.15}$$

$$\mathbf{S}_0 = \begin{bmatrix} \mathbf{s}_0(\delta_0) & \mathbf{s}_0(\delta_1) & \ldots & \mathbf{s}_0(\delta_{\text{L-1}}) \end{bmatrix}^T, \tag{2.16}$$

where † and ^{T} denote the Hermitian and non-Hermitian transpose, and $\mathbf{P}_{\mathbf{s}_0}$ denotes the orthogonal projection of matrix $\mathbf{S}_0$. The rows of matrix $\mathbf{S}_0 \in \mathbb{C}^{L\times N}$ consist of channel probing waveform samples at delays $\{\delta_0, \delta_1, \ldots, \delta_{L-1}\}$. After simplifying 2.10 the test statistic is [97]

$$T(\mathbf{z}) = \Re\left\{\mathbf{z}\,\tilde{\mathbf{s}}_{10}^{\dagger}\right\} \tag{2.17}$$

$$= \Re\left\{\sum_{n=0}^{N-1} \{\mathbf{z}\}_n\ \{\tilde{\mathbf{s}}_{10}\}_n^*\right\}, \tag{2.18}$$

where $\{\cdot\}_n$ denotes the n^{th} element of the vector inside the brackets, $\tilde{\mathbf{s}}_{10} = (\tilde{\mathbf{s}}_1 - \tilde{\mathbf{s}}_0) * \mathbf{h}$ and the sample index $n \in \{0, \ldots, N-1\}$. Inherently, 2.18 is the cumulative log-likelihood ratio with a fixed sample size N.

We employ the SPRT [189] to determine the least number of samples needed to make a positive detection of the signaling sequence. To facilitate this process, the SPRT accumulates the log-likelihood ratios on a per sample basis (as in 2.18) and partitions the decision space into two regions

$$\ell = \begin{cases} 1, & T(\mathbf{z}) \geq \gamma_1 \\ 0, & T(\mathbf{z}) \leq \gamma_0, \end{cases} \tag{2.19}$$

such that the receiver decides $\mathcal{H}_1$ if $\ell = 1$ and $\mathcal{H}_0$ if $\ell = 0$. If $\gamma_0 < T(\mathbf{z}) < \gamma_1$ the receiver will continue to sample data until a decision is made. The idea of performing detection in this manner is that a correct decision can be made in N^* samples where

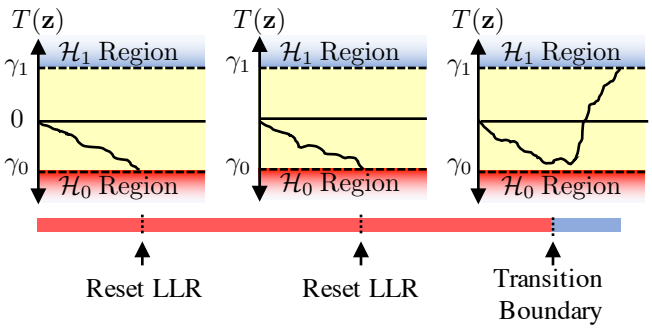


Figure 2.8: Illustration of Serial Hypothesis Tracking. The cumulative detection statistic is computed and compared against the threshold pairs. It is reset if a probing sequence is detected and continued till it detects a signaling sequence.

$N^* < N$. However, this is not always the case. We illustrate the detection process in 2.7.

(a) **Serial hypothesis tracking**

In this algorithm, the cumulative test statistic in 2.18 is calculated and compared against the thresholds for each received sample. If the cumulative test statistic value exceeds the upper threshold, a positive detection $\ell = 1$ is signaled and the algorithms stops. If the cumulative test statistic value is less than the lower threshold, a negative detection $\ell = 0$ is signaled and the cumulative test statistic is reset to zero. The receiver will continue to process new samples until a positive detection is made.

(b) **Parallel hypothesis tracking**

For the parallel tracking method, we assume that each new sample is potentially the start of the signaling sequence. Thus, we compute a new test statistic starting with the k^{th} sample, where $k = 1, \ldots, K$. This algorithm lends itself naturally to a parallel receiver architecture implementation because the receiver can compute K cumulative

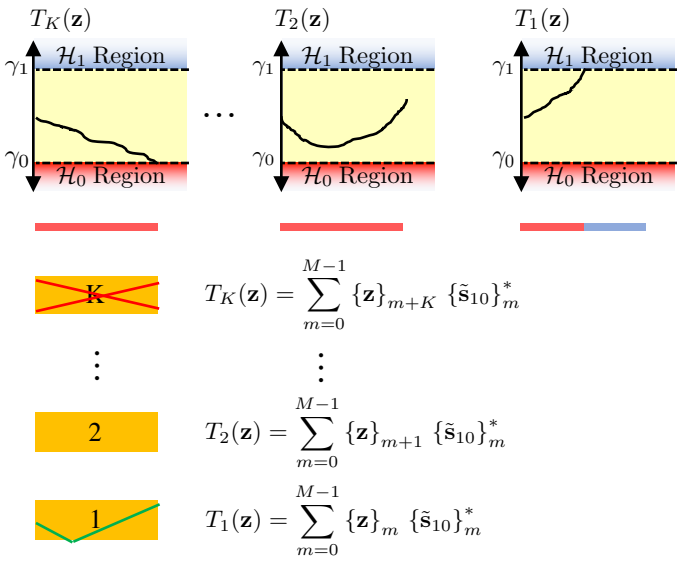


Figure 2.9: Illustration of the Parallel Hypothesis Tracking. Multiple cumulative detection statistics are computed assuming the signaling sequence begins at the $k = 1, 2, \ldots, K$ sample. Assuming a parallel software architecture is employed at the receiver the cumulative test statistics are computed simultaneously and checked against the thresholds after M test statistics are computed. If the signaling sequence is not detected then the k^{th} thread begins processing new samples starting at the $k + K + M - 1$ sample. Once the signaling sequence is detected the algorithm stops.

test statistics simultaneously, where K represents the number of threads/processors we employ for computation.

The k^{th} thread compares the cumulative test statistic to each threshold after M samples, where $M \leq N$. If no signaling sequence is detected after M samples, the k^{th} thread will process new samples starting at the $k + K + M - 1$ sample. This process continues until the signaling sequence is detected. We illustrate these two hypothesis tracking methods in 2.8 and 2.9 respectively.

2.4 Model Accuracy Evaluation Techniques

Here, we have presented some techniques used to evaluate the accuracy of the machine learning models.

2.4.1 Confusion matrix

By definition a confusion matrix C is such that $C_{i,j}$ is equal to the number of observations known to be in group i but predicted to be in group j. Thus, in binary classification, the count of true negatives is $C_{0,0}$, false negatives is $C_{1,0}$, true positives is $C_{1,1}$ and false positives is $C_{0,1}$. An example of a confusion matrix is illustrated in Figure 2.10. This technique is used mostly to evaluate classification accuracy.

		Prediction	
		0	**1**
Actual	**0**	**TN**	**FP**
	1	**FN**	**TP**

Figure 2.10: Confusion Matrix (**TN**: True Negative, **FN**: False Negative, **FP**: False Positive, **TP**: True Positive)

2.4.2 Correlation coefficients and errors

These parameters are used to evaluate accuracy during regression/prediction analysis. These measures can also be used to analyze relationship between variables. Let us denote x_i and y_i as the actual and predicted values respectively. Then the different measures are computed as follows,

- **Correlation coefficient (CC)**

$$CC = r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- **Maximum absolute error (MAE)**

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|$$

- **Root mean square error (RMSE)**

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$$

- **Relative absolute error (RAE)**

$$RAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{\sum_{i=1}^{n}|\bar{x} - x_i|}$$

- **Root relative squared error (RRSE)**

$$\sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(\bar{x} - x_i)^2}}$$

2.4.3 Receiver operating characteristic (ROC)

The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The TPR is also known as *sensitivity*, *recall* or *probability of detection* in machine learning. The FPR is also known as the *fall-out* or *probability of false alarm* and can be calculated as (1 - *specificity*). This measure is generally used in detection problems. An illustration is shown in Figure 2.11.

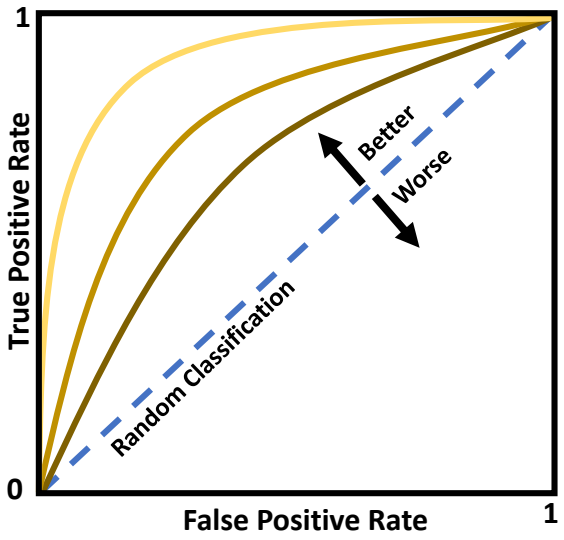


Figure 2.11: An example of ROC curve. The area under the curve (AUC) is also used as an important measure to evaluation detection efficiency

2.4.4 Figure of merit (FOM)

The FOM is used as a parameter to evaluate the performance of the algorithm during binary classification problems and in clustering problems. First, two Gaussian distributions are drawn to fit the clusters after classification. Then the FOM is calculated as the ratio of the distance (S) between the two distribution peaks and the sum of the full width half maximums (FWHM) of the peaks.

$$FOM = \frac{S}{\delta_1 + \delta_2}$$

Where δ_1 and δ_2 are the FWHM of the peaks.

Chapter 3

STUDY I: ACCELEROMETER SIGNAL PROCESSING TO MONITOR PHYSICAL ACTIVITY

3.1 Overview

The purpose of this study was to classify and model various physical activities performed by a diverse group of participants equipped with a single wrist-worn accelerometer in a supervised lab-based protocol and utilize the model to estimate physical activity in a free-living setting. The lab-based study included participants (N=251) divided in two distinct groups: adults, age 18-64 years (n=152) and older adults, age 65-94 years (n=99). Wrist-worn accelerometer data were collected and processed to identify 24 unique physical activities performed in controlled settings. The Gaussian mixture model (GMM) and the hidden Markov model (HMM) algorithms were used to classify activities with time and frequency based features. The most accurate model was then used to estimate physical activity intensity of 20 participants, each recorded for two free-living sessions of approximately 6 hours each. In the lab-based study, 24 physical activities were classified with a classification accuracy of 92.7% and 94.7% using GMM and HMM, respectively, for the adult group. Walking activities performed by older adults with and without various gait impairments were identified with an accuracy of 99.5%. In the free-living study, a prediction accuracy of 75% was achieved while predicting the free-living activity intensity levels. The estimated activity intensities showed the dominance of sedentary and light intensity activities in 36 out of 40 recorded sessions. We propose a novel activity recognition process to identify supervised, lab-based physical activities and utilize the adult clas-

sification model to estimate free-living activity intensity. We achieved high accuracy while classifying the diverse set lab-based activities and predicting the free-living activity intensities, as validated from direct observations. This study provides both supervised and free-living analysis on one of the largest and diverse population and physical activity database. The work presented in this chapter was a collaboration between the College of Health Solutions and the engineering team at ASU, lead by the author.

3.2 Background

Engaging in sufficient amounts of physical activity (PA) is associated with decreased risk of premature mortality from all sources and cardiovascular diseases [30, 34, 55]. The 2008 physical activity guidelines recommend engaging in at least 150 minutes per week of moderate-vigorous physical activity [185]. However, there are still gaps in our current understanding of the associations between physical activity and health. This is partly due to our inability to quantify accurately these behaviors, especially in free-living environments. Without an accurate measurement tool, our ability to determine the relationship between physical activity and health, develop novel and effective interventions to promote these healthy behaviors, and evaluate the effectiveness of these interventions, is severely limited. Physical activities and sedentary behaviors can be of many forms including daily household activities, walking, aerobics, and strength training, as examples with physical activities can be performed at various intensities (i.e., light, moderate or vigorous). These intensity levels are usually measured in terms of a metabolic equivalent (MET) value [90], defined as the ratio between energy expenditure during an activity and energy expenditure at rest. With the advent of wearable systems, it has become easier to monitor the PA levels due to their objectivity, minimal participant burden and rich data that can

be collected for a long period. Accelerometers have become inexpensive, small and lightweight, can gather high-frequency data and can be used by the population across all demographics. Accelerometer-based systems have been used to gain insights about physical activity all age groups, adolescents [31, 118], young adults [39, 40, 42, 83], old adults [44, 80, 117, 171, 195, 206]and seniors [80, 157, 162, 180]. There are two major research foci in PA monitoring studies: (a) energy expenditure (EE) estimation and (b) activity classification. The traditional approach to EE estimation using accelerometer data is to estimate the intensity (MET value) of an activity through simple linear regression modeling [67]. Another approach is to attempt to identify the type of activity performed, and calculate EE using knowledge of the activity's intensity. Current methodological development, especially in signal processing and machine learning techniques, have led to researchers implementing different frameworks to estimate EE. Of which, methods such as artificial neural network [128], novel estimation framework based on statistical estimation theory [193] and piecewise linear regression model [169], deserve special mention due to their high prediction accuracy. Activity classification studies have been undertaken using both supervised laboratory and free-living protocols. The classification approach involves extracting features from raw or processed accelerometer data and using them to identify unique physical activities using machine learning or deep learning based classifiers. Various studies have investigated the different types of features, classifiers and physical activities with high accuracy ranging from 68 to 99% [44, 49, 50, 56, 118, 179, 183, 195, 206]. The purpose of this study was to classify and model various physical activities performed by a diverse group of participants equipped with a single wrist-worn accelerometer in a supervised lab-based protocol and utilize the model to estimate the physical activity level of a sub-sample of participants in free-living settings. In the first part of this chapter, we train and test our classifiers to model physical activities using accelerom-

eter data from lab-based settings. The second part of the chapter uses free-living accelerometer data and applies an unsupervised classification from the classification models from the supervised lab protocol results to estimate activity intensity levels.

The data used in this analysis were gathered through two separate studies conducted in a southwestern university in the USA. The first study provided accelerometer data on structured, lab-based activities that were used to train and validate the proposed machine learning method. The data from the free-living protocol were used to evaluate the developed algorithm in its capacity to estimate activity intensities in a free-living setting. The details of each study (recruitment, participant characteristics, and data collection methods) are described in the following sections.

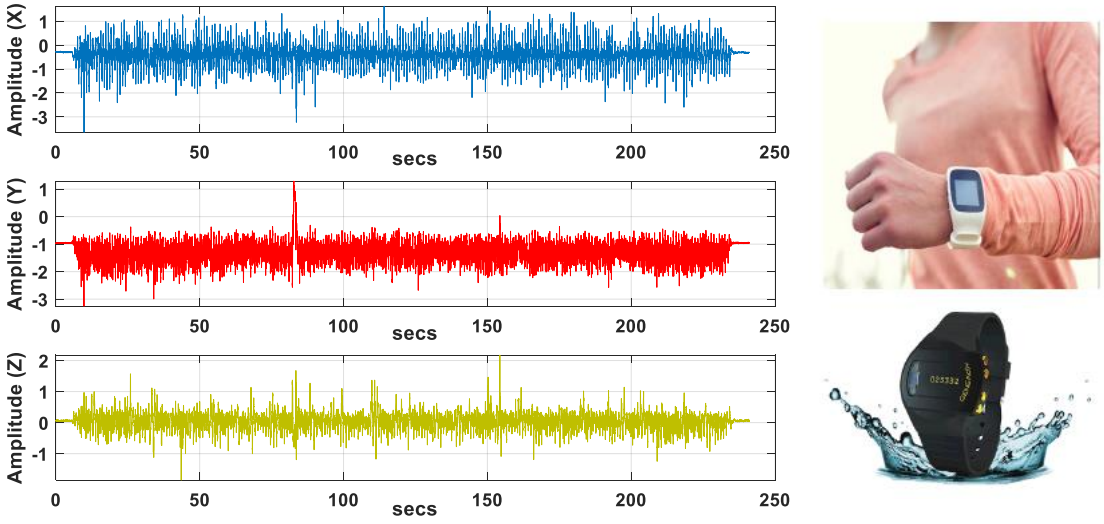


Figure 3.1: (a) Tri-axial accelerometer signal during walking activity, (b) Wrist-worn GeneActiv accelerometer

3.3 Lab-Based Study Protocol

3.3.1 Data Collection

A total of 152 adult participants (48% male, age 18-64 years old) and 99 older adults (75% using assistive devices, age 65-94) were recruited in the lab-based protocol and 20 participants (50% male, age 21-46 years old) participated in the free-living protocol. The recruitment method and participant eligibility criteria were similar in both lab-based and free-living protocol and accomplished via fliers, emails, and social networks (e.g., Twitter, Facebook). Interested participants completed an online screener and scheduled a lab visit to determine eligibility by performing a wide array of physical activities. Participants were screened for conditions that could limit their physical activity (e.g., cardiovascular disease, high blood pressure) as well as completed a physical activity readiness questionnaire. For both studies, informed consent was obtained from each participant prior to enrollment. The university's institutional review board approved all study materials and procedures.

After obtaining consent, each participant was scheduled for a 2-hour laboratory visit. They were instructed to wear comfortable clothing and were fitted with a GENEActiv accelerometer (Activinsights Ltd., Kimbolton, UK) on their non-dominant wrist along with other activity monitors. The GENEActiv is a lightweight, waterproof, wrist-worn sensor that collects raw acceleration data (see Figure 3.1). Adult participants performed a set of ambulatory and lifestyle activities randomly selected from a predetermined pool of activities (see Table 3.1). The older adults performed a 'walking' task with or without using assistive devices based on their gait impairments. Participants were video-recorded completing each of the activities using a custom-designed Android application developed by our research team. It provided automated start and stop times for each activity and an video file of each activity.

Table 3.1: Laboratory dataset and physical activity details, showing the unique physical activities performed by the participants, with the associated MET values and intensities.

Dataset (Adults)	PA No	PA Class	MET value	Intensity Label
Non-ambulatory	1	Seated, folding/ stacking laundry	2.0	L
	2	Standing/ fidgeting with hands while talking	1.8	L
	3	1 minute brushing teeth + 1 minute brushing hair	2.0	L
	4	Driving a car	2.5	L
Walking	5	Treadmill at 1 mph	2.0	L
	6	Treadmill at 2 mph	2.8	L
	7	Treadmill at 3 mph	3.5	M
	8	Treadmill at 3 mph, 5% grade	5.3	M
	9	Treadmill at 4 mph	4.9	M
	10	Hard surface walking	2.8	L
	11	Hard surface, hand in pocket	3.5	M
	12	Hard surface, while carrying 8 lb. object	5.0	M
	13	Hard surface, holding cell phone	4.5	M
	14	Hard surface, holding filled coffee cup	3.5	M
	15	Carpet with high heels or dress shoes	2.8	L
	16	Grass barefoot	4.8	M
	17	Uneven dirt	4.5	M
	18	Uphill with high heels or dress shoes, 5% grade	5.3	M
	19	Downhill with high heels or dress shoes, 5% grade	3.3	M
Running	20	Treadmill at 5 mph	8.3	V
	21	Treadmill at 6 mph	9.8	V
	22	Treadmill at 6 mph, 5% grade	12.3	V
Stair climbing	23	Walking upstairs (5 floors)	4.0	M
	24	Walking down stairs (5 floors)	3.5	M
Dataset (Seniors)				
Walking	25	No disability, 0.25 miles on hard surface	2.5-4.0	L/M
	26	SPPB Pos, 0.25 miles on hard surface	2.5-4.0	L/M
	27	Cane, 100 m hard surface	2.5-4.0	L/M
	28	Walker, 100 m hard surface	2.5-4.0	L/M

Table 3.1 shows the major groups of PA, the MET values associated to each PA, according to the 2011 Adult Compendium of PA [1], and their corresponding intensity levels.

3.3.2 Data Processing

The tri-axial accelerometer data (X, Y and Z axes) were collected at a sampling rate of 100 Hz, during which the participants had to stay still for the first 5 seconds, and then perform the specific PA for a fixed period of time. As a part of pre-processing, the resultant acceleration, $R = \sqrt{X^2 + Y^2 + Z^2}$, was calculated and used as a fourth signal to process along with the X, Y and Z- direction acceleration signals. After the activity transitions were identified from observed labels, the four acceleration signals were divided into windows of 1 sec (1000 samples) without overlaps, which is enough to capture both stationary properties of the signal and change of the signal over time. The proposed classification method involves three major steps after windowing, which are, feature extraction, feature selection and activity classification, as shown in 3.2.

Feature Extraction

In this study, we have investigated some state-of-the-art features and introduced some novel ones to represent each window of accelerometer signals. 130 such features were extracted from every 1 sec window from different combinations of the four acceleration signals. We have briefly explained some essential backgrounds of the features investigated in this study. These features are described in details in section 2.1.1.

- *Time-Domain Features*: Mean, standard deviation, skewness, kurtosis, energy and the squared sum of the Y and Z acceleration signals under the 25^{th} and 5^{th} percentile.

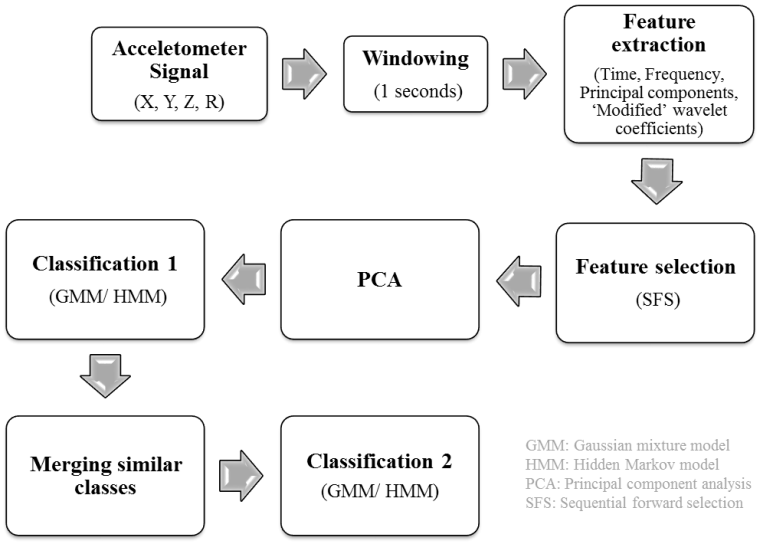


Figure 3.2: Lab-based activity classification process chain

- *Frequency-Domain Features*: Maximum magnitude between 1 $\rightarrow$ 5 Hz, sum of frequency component heights below 5 Hz and number of peaks in spectrum below 5 Hz.
- *Principal Component Features*: First four principal components.
- *'Modified' Wavelet Coefficient Features*: In our case we used a three level 'Haar' wavelet decomposition to extract 20 wavelet coefficients from each 1 sec window.

Feature Selection

Machine learning algorithms always present problems when dealing with high dimensional inputs, so we selected the most 'efficient' features out of 130 features by reducing the feature dimension. For this purpose, we used the sequential forward selection (SFS) method (described in section 2.2.3). The SFS is a greedy search algorithm that works in tandem with classifiers and compares classification accuracy at each step. We used the SFS algorithm for training on a random subset of 40 adult

participants' laboratory accelerometer data, using GMM classifier. Results showed that the 'modified' wavelet coefficients extracted from the resultant acceleration signal (R) were the most 'efficient' features. We performed principal component analysis (PCA) (see section 2.2.1) to make use of all the 20 wavelet coefficients.

Classification

For classification of physical activities, we explored two classification algorithms, GMM and HMM, with the first 10 principal components used as the input feature vector. One major property of principle components is that the components are orthogonal vectors; hence, the most probable distribution to model unique PA features is a Gaussian distribution. Thus, the choice of GMM and HMM with Gaussian distribution as their output distribution is justifiable. We executed two levels of classification; the first level was used to combine similar activities and reduce the number of activity classes and the second level was used extract the models for each activities (unique or combined).

- *Gaussian Mixture Model*: GMM is one of the most commonly used classifiers, which models the probability distribution of data as a linear combination of multiple Gaussian distributions. We have described the GMM in details in section 2.3.1(II) along with the modifications we made.

- *Hidden Markov Model*: HMM was used with the GMM as the probability distribution function of the HMM output. The Viterbi approximation path algorithm was used to estimate the new labels for which the joint distribution of X (feature vector) and Z (observed PA labels) is maximized. The algorithm considers the most likely path instead of summing over all possible state sequences, which saves a lot of computation time. (See section 2.3(III) for details)

- *Merging Similar Classes*: The confusion matrix (see section 2.4.1) was constructed after first level of classification using the predicted and actual classes as its rows and columns, respectively. Since we were working with many activities with similar MET values, there was a high possibility that certain classes would be predicted as another class with similar MET values. We employed a thresholding technique to combine similar classes into one larger class. For any class, if more than 50% of the class was predicted as another class, we combined the classes together into a single class (see Figure 3.3). This method helped us to find similar PA's in an unsupervised manner. After combining the similar classes, we performed second level of classification to construct the final confusion matrix.

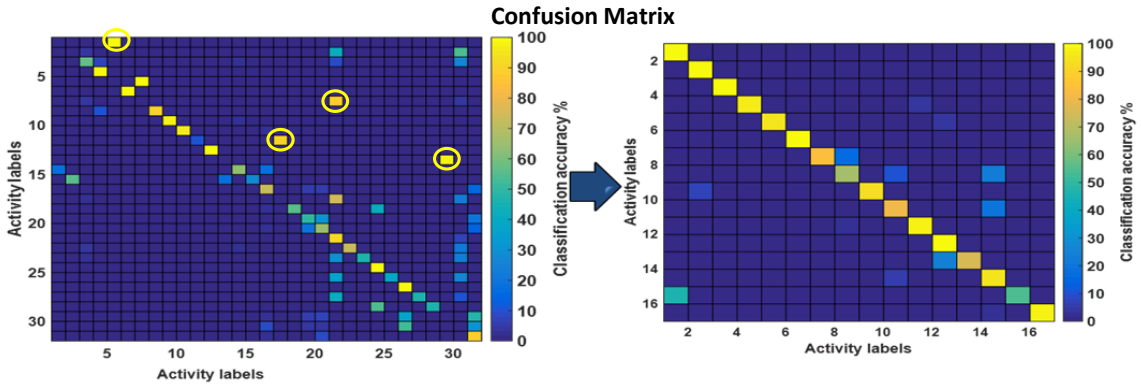


Figure 3.3: An example showing how similar classes were merged. The left confusion matrix is obtained after first level of classification, the confused classes are circled and are merged. The final confusion matrix after second level classification is shown on the right hand side.

3.3.3 Results

We trained the classification process with a random subset of 40 adult participants using SFS and GMM. The rest (112 adults and 99 older adults) were used as the test

data. The classification results for the adult and senior samples are shown in Table 3.2.

Table 3.2: Classification accuracy for various lab-based physical activities using GMM and HMM (showing the initial and the final number of PA classes after merging)

Dataset (Adult)	Original PA classes	PA classes after merging (GMM)	PA classes after merging (HMM)	Classification accuracy % (GMM)	Classification accuracy % (HMM)
Non-ambulatory	4	4	4	99.91	89.32
Walking	15	10	10	79.57	84.87
Running	3	2	3	91.4	99.86
Stair-climbing	2	2	2	100	99.8
Dataset (Seniors)					
Walking	4	4	4	93.94	96.37
Total Dataset					
All activities	28	21	17	78.54	90.2

Adult Sample

We have shown classification results for the four major classes of activities; non-ambulatory, walking, running and stair climbing in Table 3.2. The table shows the number of initial activity classes, the number of combined classes after merging and the final classification accuracy using GMM and HMM. Out of 15 walking activities, 5 classes (PA classes 5, 10, 12, 14 and 16 from 3.1) and two classes (PA classes 15 and 19 from 3.1) were merged to two single PA classes using GMM. With HMM, two classes (PA classes 6 and 9), four classes (PA classes 10, 12, 14 and 16) and two classes (PA class 15 and 19) were merged into three single PA classes. Similar group

of classes were merged with both classifiers. In all other activities, the classes mostly remained unmerged. Comparing the two classifiers, the final classification accuracies for all the major classes were 99.91% (GMM) for non-ambulatory, 84.87% (HMM) for walking, 99.86% (HMM) for running and 100% (GMM) for stair climbing activities.

Senior Sample

A total of 4 walking PA classes were identified based on the type of assistive devices they carried. An overall classification accuracy of 99.47% and 99.17% was achieved with GMM and HMM respectively. This dataset has been used only to test our classification algorithms, and to show how well they perform with a diverse dataset. This dataset played no part in the estimation of the best classification model, which is discussed below.

Best Classification Model Selection

The best classification model was estimated from the adult dataset of 24 PA. Three parameters were used to compare the performances of the classifiers; trace of the confusion matrix after the two levels of classification, final classification accuracy (mean of the trace of the final confusion matrix) and total number of classes identified after merging as shown in Figure 3.4. Given the input feature vector of 10 principal components (of 'modified' wavelet features), we tested the classifiers in multiple feature spaces (2-D to 10-D, each dimension representing a principal component) to find the best classification model and feature space. The GMM gave best classification performance in 10-D feature space with an accuracy of 78.5% (21 final classes), and HMM gave the best performance in 6-D feature space with accuracy 90.5% (17 final classes). Upon comparing both classifiers, the GMM achieved higher average trace with fewer classes combined, while the HMM achieved higher classification accuracy

with more classes combined.

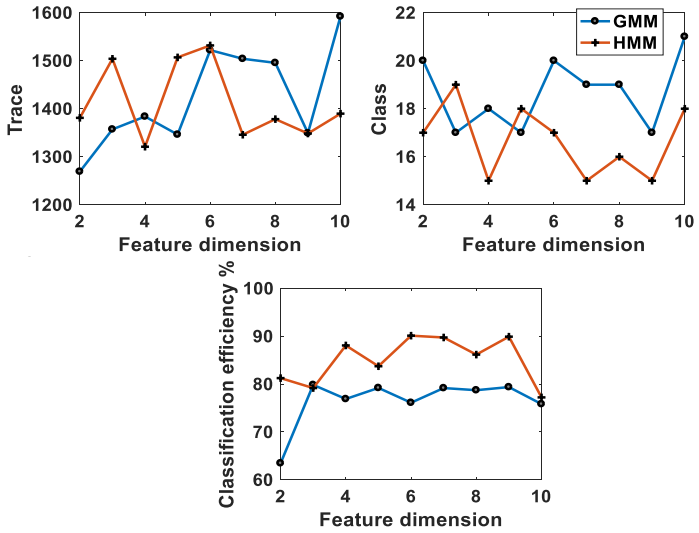


Figure 3.4: Comparison between GMM and HMM performance

To decide the better classifier, we made a comparison between the two best cases of GMM and HMM by combining one class at a time and comparing the accuracy after each step, until all classes were merged to one whole class (3.5). As seen in 3.5 , as we kept combining PA classes, HMM showed a better convergence than GMM. Thus, the classification model of the HMM in 6-D feature space was selected as the best classification model.

3.4 Free-Living Study Protocol

3.4.1 Data Collection

The 20 participants (50% male, age 21-46 years old) who participated in the free-living protocol were instructed to indicate two typical days (one weekday and one weekend day) for data collection. On those selected days, they were instructed to maintain their usual daily activity pattern while two researchers were independently

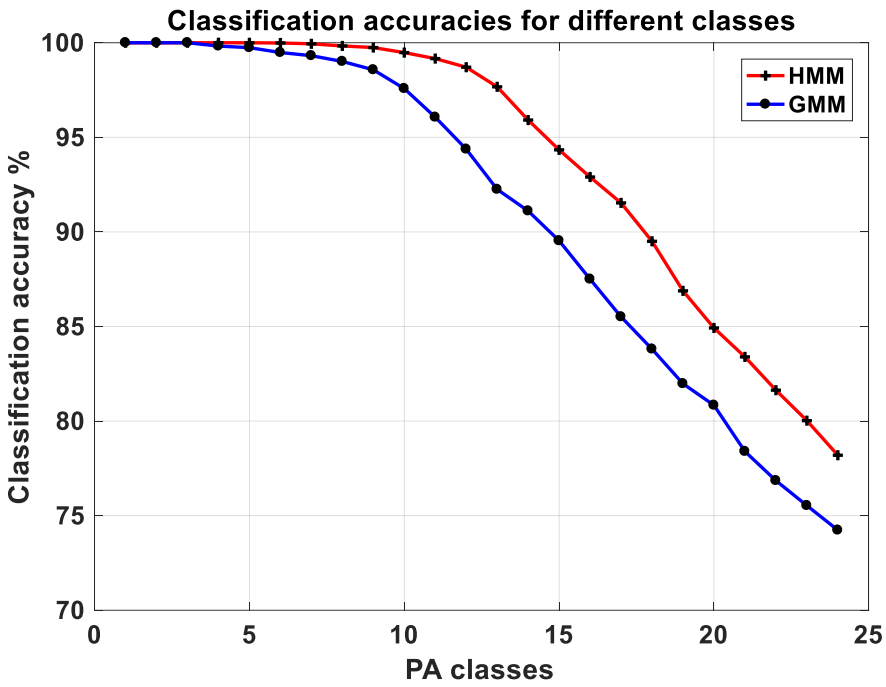


Figure 3.5: Convergence characteristics of each classifier (GMM and HMM), comparing the classification accuracies, as a PA class is merged at every step.

classifying their activities through direct observation [64]. The researchers continuously classified a participant's activity over a 6-8 hours' period using a researcher-developed mobile app that allowed for continuous activity classification. Activities were labeled based on the type and context of activities. The six physical activity type labels were walking, sitting, jogging, reclining, standing and squatting whereas the context labels were sports/exercise, household chores, transportation, transportation, occupation and leisure. About 8% of the data were classified as unobserved, when participants required private time (e.g. restroom use) or were out of sight of the researchers. On both days, participants were asked to wear the GENEActiv continuously along with other activity monitors.

3.4.2 Data Processing

In the lab-based settings, we identified unique PA by classification models of Gaussian mixtures. From the lab-based results it was shown that the best supervised classification model for the adult dataset was accomplished using HMM with the first 6 principal components of the 'modified' wavelet features. We only used the adult group PA classification model to identify PA in the free-living settings. The primary goal of the free-living data analysis was to estimate the PA intensity associated to the identified activity types. The entire process was divided into the following sub-sections based on the order they were performed: pre-processing, feature extraction, unsupervised classification and Gaussian model matching (see 3.6).

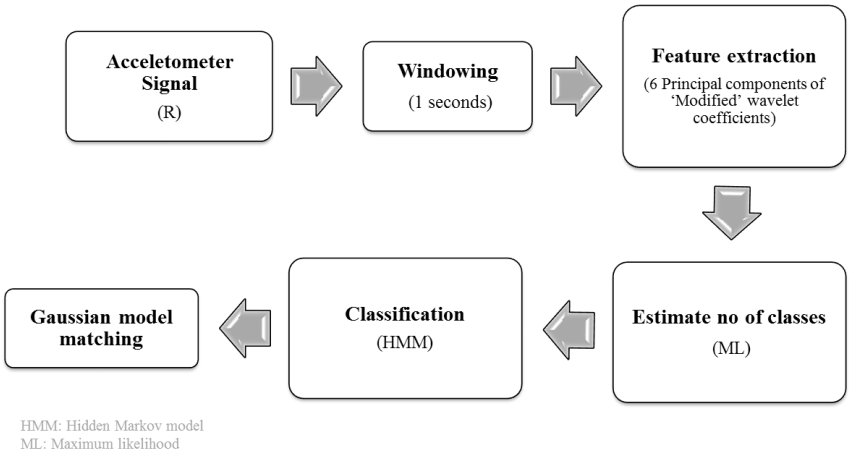


Figure 3.6: Free-living activity identification process chain

Pre-processing

The resultant acceleration, R, was calculated from the tri-axial acceleration signal and was used as the main signal from which features were extracted from windows of 1 sec.

Feature Extraction

The 'modified' wavelet coefficients proved to be the most efficient feature choice for the lab-based settings study. The first 6 principal components of 20 'modified' wavelet coefficients were used as feature vector.

Unsupervised Classification

Since the number of activities performed during a session was unknown, we first estimated the total number of activities performed using maximum likelihood estimation. Then we performed classification using HMM to find the Gaussian distribution corresponding to each activity.

Gaussian Model Matching

Using an unsupervised classification, we managed to estimate the total number of activities and modeled them by a mixture of Gaussians in the 6-D feature space. However, we still needed to identify the activities. We identified an unsupervised activity as the lab-based PA that had the 'minimum distance' in the feature space. We defined this 'distance' as the distance between the means of the Gaussians of the unsupervised model and the best supervised model. The predicted unsupervised PA is given by,

$$x_c = \underset{L_c, c=1,\ldots,24}{\arg\min} |\mu_x - \mu_{L_c}|$$

μ_x and μ_{L_c} are the means of the x^{th} free-living PA class from the unsupervised model and c^{th} lab-based PA class from the best supervised model Gaussian distributions.

3.4.3 Results

We first estimated the total number of activities performed by a participant in each session by calculation the Gaussian mixture maximum likelihood. HMM and Gaussian model matching was used to identify the estimated activities. We estimated the intensity levels of the activities using the corresponding MET values of the estimated classes. A correlation coefficient of 0.75 was achieved between the estimated and the actual intensity level, which was calculated from direct observations. Figure 3.7 shows the estimated number of activities performed by a participant (total 20) in each session (total 2) and the estimated intensity levels along with the recorded activity and context labels. Results are shown in terms of percentage of time spent on different activity labels (estimated and observed).

3.5 Discussion

This study systematically classified 24 lab-based supervised PAs and used the best classification model to estimate activity intensity levels of participants in free-living settings. The adult participants (n = 152) performed activities from a pool of 4 sedentary, 15 walking, 3 running and 2 stair climbing activities, while the senior participants (n = 99) performed a simple walking task with or without an assisted device. We achieved fairly high accuracy, identifying classes from each activity group with both GMM (79-100%) and HMM (85-99%) for both dataset (adults and seniors). We then tested both the classifiers with the entire adult dataset in different feature spaces to find out the best classification space. The HMM in a 6-D feature space proved to be the best classification model and this model was used to identify unsupervised activities in the free-living settings. We estimated the total number of activities and the activity intensity levels of a participant in each session with high accuracy (ap-

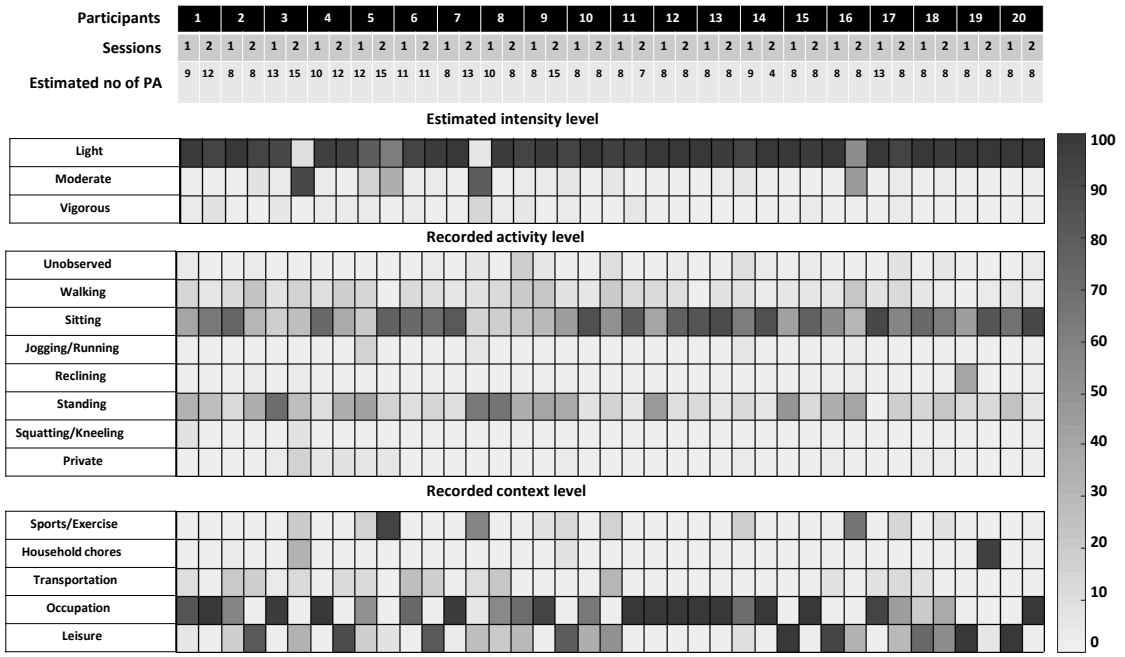


Figure 3.7: Free-living analysis results, showing the estimated number of PA and intensity levels (estimated), the observed activity types and contexts for every participant in each session. Results are shown in form of percentage of time spent in each type of activity in a session. For e.g., participant 1 performed more that 90% of his 1st session performing sedentary or light intensity physical activities (estimated), and from the observed labels, it can be seen that he performed sitting and standing activities for more than 90% of the time which was spent mostly during his occupation.

proximately 80%). Nearly all participants spent most of their time doing sedentary and light intensity tasks. The recorded activity levels showed that participants spent most of their time doing stationary (i.e., sitting and standing) activities. With the addition of analysis of data from 99 older adults, we have a better understanding of gait patterns in older adults with and without assistive devices and possible long-term

monitoring approaches can be applied.

This study is unique because both lab-based and free-living dataset were investigated, with one data co-dependent on the other. This study is an important advance from past studies given its comparatively larger lab-based database with a diverse population group of 274 participants and diverse range of activities. Twenty-four different activities, including sedentary and light-to-vigorous PA intensity were identified in this study with high accuracy. We used novel features ('modified' wavelet coefficients) from the acceleration signals and modeled each activity with a Gaussian distribution using GMM and HMM. The most accurate model was tested with a free-living dataset of 20 participants PA intensity, enabling us to estimate free-living energy expenditure. Of note, our results suggest less degradation in activity classification accuracy from laboratory to free-living settings than previous studies. We posit that this may be because of the robust activity classifier that was developed given the large sample size and diverse set of laboratory-based activities.

This study has important implications for physical activity researchers. First, because these data were collected using raw waveform accelerometry on the wrist, these computations can be replicated across a large range of wearable sensors that capture and make available to the researcher raw accelerations, and are not limited to the GENEactiv sensors used here. Second, while ample detail are provided here for data scientists to replicate this approach, the methods are not computationally intensive and more user-friendly tools are currently being prepared to make this approach available for physical activity researchers. The manufacturers of wearable sensor that are commonly used by physical activity researchers are encouraged to include this algorithm in data analysis packages they make available to their customers.

In summary, this study contributes to the use of wearable sensors to identify physical activities and estimate energy expenditure in free-living settings by applying

state-of-the-art machine learning approaches to a diverse set of laboratory-based, supervised activities. This study has demonstrated success in transferring lab-based validation techniques to the estimation of free-living activities that can be applied to future studies that wish to estimate physical activity in cohort or intervention studies.

Chapter 4

STUDY II: ANALYZING GUT-MICROBIOME DATA TO FIND THEIR RELATIONSHIP WITH BONE HEALTH

4.1 Overview

The human gut-microbiome (GM) have been associated with regulation of various biological processes and recent studies have made significant contributions in investigating various effects of the GM community. There have been a lot of evidences that suggest that the GM plays an important role in bone health regulation. In this chapter, we present a study analyzing the relationship between the GM and bone health using data from (n = 26) adolescent females in a controlled feeding study. Various machine learning models were used to predict measures of bone mineral density (BMD) and calcium absorption using cross validation. Results showed that there exists a high positive correlation (0.81) between GM and bone health. We further identified the microbe classes that affected this relationship and observed that they were consistent with previously published results. Results show that *Clostridiaceae*, *Bifidobacterium*, *Faecalibacterium* and *Prevotella* were positively correlated and were found in more quantity in participants with higher BMD. Therefore, this study identifies the important GM communities, and validates their importance in bone regulation.

4.2 Background

The human gut (gastrointestinal tract) is home to a large number of microbe organisms (around 40 trillion). Most of these microbes belong to 30 to 40 species, but there can be up to 1000 different species in total [138]. The gut microbiome

(GM) plays a fundamental role in human health and certain diseases [137, 174, 182]. While many studies link microbiome composition to phenotypes [33, 175, 190], we lack understanding of the boundaries of bacterial diversity within the human population and how they affect health conditions, lifestyle, and diet. Some of the known facts are as follows; certain classes of GM benefit us by breaking dietary fiber down into shot-chain fatty acids [194], aiding in metabolizing of complex compounds [184], bsorbing calcium [198], synthesizing vitamins B and K [160] etc. Recent research also infer that dysregulation of the GM might be an important factor in inflammatory [10] and autoimmune conditions [98]. Other studies conclude that the GM might play a significant role on human psychology and behavior. It has been found that there exists a strong connection between mental health and the GM, regardless of gender, age or geographical location [166]. Although, researches in this field have recently started to develop, results are still inconclusive and requires more experimental studies to get a better understanding of the human GM and to educate the broader community about this key aspect of human health [121]. Detailed knowledge about the effects of the GM community might lead to the development of precision medicine and probiotics and even diagnostic tools for disease detection.

In this work, we analyze the relationship between the human GM and bone health. Previous studies suggest that modulation of the GM can affect the bone density and strength [28, 86, 141, 194]. Experiments in [164] and [170] have produced some evidence of a relationship between GM and bone health in germ-free animals. Other studies showing intestinal bacterial overgrowth during bone loss [41, 176], suggest that it may contribute to occurrence of metabolic bone disease. It has conjointly been reported that the GM influence in absorption of varied nutrients [194, 198], thus helping in bone development. Other studies on prebiotics, probiotics and antibiotics [85] suggest that they have important positive health effects on the host.

Several studies have reported GM as a regulator of bone mass [120, 170, 194] through mediation of the immune system (e.g., osteoclastogenesis) [191], intestinal calcium absorption and the release of neurotransmitters (e.g., serotonin) [191]. Other effects involve complex mechanisms including modulation of $CD4^{+}$T cell activation, control of osteoclastogenic cytokine production and modifications in hormone levels [86]. Despite there being a few reports of individual bacteria impacting multiple aspects of bone physiology, we are far from understanding the complex interplay between the GM and bone health. Current research efforts, although limited, clearly indicate that the GM may be implicated in bone metabolism, and therefore, further exploration of this relationship is a promising area of focus in bone health and osteoporosis research. Most existing studies investigate this relationship using animal models and needs more human studies.

In this paper, we study the relationship between the human GM and different measures of bone health using data collected from 26 adolescent females during a randomized crossover dose-response study. Two fecal samples were collected from each participant, for each dosage of soluble corn fiber (SCF; 0, 10, 20 g fiber/d) added to their normal diet. We predicted different measures of bone health using 221 classes of GM using machine learning models. Our results reveal a high positive correlation ($r \approx 0.8$) between the predicted and measured bone mineral density (BMD). We further studied some of the important GM classes that affected this relationship. Various microbes which have been studied before for their contribution to bone regulation, appeared to have high correlation with the BMD measures. Further, it was seen that some of these microbes are present in higher proportion in participants with higher BMD.

4.3 Study Design and Data Collection

4.3.1 Study Participants

A total of 28 Caucasian, female adolescents (aged $11 - 14$ years) were recruited from local schools, neighborhoods, and community establishments around the Purdue University campus to participate in this randomized crossover dose-response study. Eligible participants were identified as healthy adolescents with calcium intakes between the 5th and 95th percentile of usual intake for this age group ($550 - 1500$ mg/d). The participants were provided a list of foods and beverages that contained prebiotics and probiotics. This study was conducted according to the guidelines in the Declaration of Helsinki, and all procedures that involved human subjects were approved by the Institutional Review Board of Purdue University. Written informed consent was obtained from all subjects and parents or guardians.

A final sample size of 26 participants data was used for the prediction analysis, as baseline fecal sample was not collected from two participants. Table 4.1 shows the demographic statistics of the participants who were enrolled in this study.

4.3.2 Study Design

Study participants in this 3-phase, double-blind, crossover study were assigned 0, 10, and 20 g fiber/d from PROMITOR SCF 85 (provided by Tate and Lyle) in randomized order. The SCF ingredient was a fermentable, nondigestible carbohydrate that contained a minimum of 85% soluble dietary fiber, $< 2\%$ sugar with a caloric content of 1.2 kcal/g. For 4 wk, participants consumed one-half of the daily SCF dose (0, 5, and 10 g fiber/d, which is 0, 6.6%, and 13.37% PROMITOR 85) in a muffin, and the second half in a fruit-flavored beverage. Control (0 g fiber/d from SCF) beverages contained a maltodextrin placebo, whereas muffins were prepared per

Table 4.1: Baseline age and pubertal and anthropometric characteristics of the participants

Characteristics	Value
Age, y	13.46 ± 0.9690 $(15.25 - 11.59)$
Height, cm	161.17 ± 53.25 $(173.13 - 147.68)$
Weight, kg	55.27 ± 105.56 $(76 - 35.75)$
BMI, kg/m^2	21.10 ± 6.41 $(26.29 - 15.77)$
Total body BMD, g/cm^2	1.04 ± 0.01 $(1.17 - 0.83)$

Values are means ± variance (ranges), n = 26. BMI: body mass index; BMD: bone mineral density

the recipe with no placebo. The 4 week consumption period was followed by a 3 day clinical visit and a 3 to 4 week washout phase.

Habitual dietary intake was estimated with 6 day diet records that were completed before the start of the study and between each subsequent intervention, during the washout phase. Diet record data were analyzed with the Nutrition Data System for Research (software version 2013; Nutrition Coordinating Center University of Minnesota). For each of the 3 interventions, baseline and 4 week (end of intervention) fecal samples were collected for high-throughput sequencing of the intestinal microbiome. These samples were stored in a walk-in refrigerator and processed within 48 hours of collection.

4.3.3 Bone Health Measures

We have used BMD and bone mineral content (BMC) and the measures of bone health. A dual-energy X-ray absorptiometry (DXA) is used to measure bone health.

It is a quantitative method, proposed by the WHO which allows accurate estimation of BMD and BMC. During one of the clinical visits, DXA scans (iDXA; GE Lunar) were taken of the total skeleton, dual hip, and lumbar spine. The average BMD is measured as,

$$BMD = \frac{BMC}{W}[\text{g/cm}^2] \tag{4.1}$$

where W is the width at the scanned line and BMC is measured in g/cm.

4.3.4 Fecal Processing and Phylogenetic Diversity Analysis

The pH of each fecal sample was assessed by inserting an electrode pH probe (pHSpear; Eutech Instruments, Thermo Fisher Scientific) into 3 different locations of each stool. The fecal samples were weighed, and twice this weight in ultra-pure (Milli-Q) sterile water was added to each sample. After homogenization, $5 - 10$ mL of the fecal slurry was stored in 15 mL sterile centrifuge tubes at -20 C and used for all further microbiologic analyses. DNA was extracted and quantified as described in [197]. Extracted DNA was subjected to PCR in 2 phases with the use of Q5 High Fidelity DNA Polymerase (New England Biolabs). The first phase amplified the 16S rRNA gene with the use of primers specific to the V3– V4 region (forward TAC GGR AGG CAG and reverse CTA CCR GGG TAT CTA ATC C primers) as described in [111, 139]. Unincorporated primers and nucleotides were separated from PCR amplicons with the use of Agencourt AMPure XP kit (Becker Coulter, Inc.). The second PCR phase allowed for incorporation of 8-base pair forward and reverse primer tags to allow for sample differentiation after sequencing. The Agencourt AmPure XP Kit (Becker Coulter, Inc.) was used to purify second-phase PCR products, and purified amplicons were quantified by fluorometry after staining with the PicoGreen

DNA Assay Kit.

The phylogenetic diversity of bacterial communities was determined after amplicons from each sample were combined in equivalent quantities and sent to the Purdue Genomics facility for sequencing by high-throughput, paired-end, MiSeq technology (Illumina). After removing primer tags and low quality sequences, paired-end reads were merged and analyzed with the QIIME pipeline [25]. Reported results were limited to known genera in the Greengenes database, version 13.5 [122]. These sequences were first prefiltered with the Greengenes core sequences with the use of a 60% threshold value, and operational taxonomic unit (OTU) assignments were made with the uclust method [54]. Finally, representative OTU sequences were obtained after sequence alignment with the use of PyNast [24] to filter out sequences that did not align with the Greengenes core sequences. Taxonomic assignments were made with the RDP classifier at 80% confidence and the Greengenes database. Rarefaction analysis was used to obtain an estimation of sequence coverage of the community. To obtain a rarefied data set, 10 iterations of randomly choosing 28800 sequences (lowest number of sequences in a sample) from each data set was performed, then data sets were merged to obtain a set of 28800 sequences that were representative of each sample.

4.3.5 Gut-Microbiome Diversity

A total of 221 classes of the GM were investigated in this work. The community contained species from 14 bacteria and 1 Archaea kingdoms (k) and 1 unassigned group. *Firmicutes*, *Proteobacteria*, *Actinobacteria* and *Bacteroidetes* were the major phyla (p). The taxonomy order used in this study is, kingdom (k)$\rightarrow$ phylum (p)$\rightarrow$ class (c)$\rightarrow$ order (o)$\rightarrow$ family (f)$\rightarrow$ genus (g).

4.3.6 Calcium Absorption

A calcium absorption test was performed with dual-stable calcium isotope technique. Urine was collected in 12-hour pools for up to 48 hours and analyzed for isotope enrichment, as described in [199]. In brief, modifications to this protocol included the use of 10 mg of orally consumed ^{44}Ca in milk and 3 mg of intravenously delivered ^{43}Ca. Blood draws occurred at baseline and at 3, 24, and 36 hours after intravenous infusion, of which the first 2 were 15-mL samples (taken from intravenous catheter) and the last 2 were 5-mL samples. During each intervention phase, weekly questionnaires were administered to assess gastrointestinal symptoms with the use of a Likert scale from 0 (no symptoms) to 5 (severe symptoms). Symptoms assessed included flatulence, bloating, abdominal pain, diarrhea, and stomach noises.

4.4 Problem Formulation

Our aim in this work is to find the relationship between GM and bone strength. For each interventions (0, 10 and 20 g fiber/d of SCF), 2 fecal samples were collected for GM sequencing. So, a total of 6 ($fs = 2 \times 3$) GM sequences were collected from each participants. To evaluate this relationship, we used the GM sequence from 26 ($n = 26$) participants to predict their bone strength using machine learning. We can denote the feature vector by, $Feat_{GM} \in \mathbb{R}^1 \times GM_{seq}$, where GM_{seq} is the total number of microbiomes. The dependent variable (bone strength) can be denoted by $V_{BMD} \in \mathbb{R}^{(n \times fs) \times 1}$. We have used the DXA measured of BMD as the measure of bone health. Since there was hardly any change in the value of BMD over the 3 intervention studies, we have used the same measure for all 6 instances of GM sequences.

4.5 Prediction Analysis

We have investigated five regression techniques to predict the dependent variable. The prediction process included the input features ($Feat_{GM}$) and a regression model. We predicted various measures of bone strength to analyze the relationship between the GM and bone health. We also predicted the fractional calcium absorption, using the same models. Further we used a feature selection to rank the microbes that show high correlation with bone strength and calcium absorption.

4.5.1 Regression Techniques

We used 5 regression learning methods, which are also called ensemble learning methods, as they require a base model. They are briefly described the below.

- *Random Forest* (RF): (see section 2.3.2 (V) for more details)
- *Locally Weighted Learning* (LWL): (see section 2.3.2 (VII) for more details)
- *Random Committee* (RC): (see section 2.3.2 (VIII) for more details)
- *Bagging*: (see section 2.3.2 (VI) for more details)
- *Voting*: (see section 2.3.2 (IX) for more details)

4.5.2 Feature Selection Technique

We used feature selection to identify the most relevant GM genera for measuring bone health. The feature selection method was used based on ranker search protocol, that ranked these microbes, while predicting the measures of bone strength.

- *Classfier attribute evaluator (CAE-RF)*: This algorithm ranks features based on the accuracy of prediction, measured using correlation weights (see section 2.2.4

for more details). For this method, we used the RF model as the prediction algorithm and the correlation coefficient as the evaluation measure, with 10-fold cross validation.

4.6 Results

Table 4.2: Correlation coefficient while predicting BMD using the GM (using each regression techniques); CC: correlation coefficient; MAE: maximum absolute error

	RF		LWL		RC		Bagging		Voting	
Methods	CC	MAE	CC	MAE	CC	MAE	CC	MAE	CC	MAE
V_1 (TBBMD)	**0.81**	0.06	0.79	0.06	**0.80**	0.06	0.78	0.06	**0.81**	0.06
V_2 (SPBMD)	0.76	0.08	**0.80**	0.08	0.79	0.08	0.76	0.09	0.79	0.08
V_3 (HBMD)	**0.83**	0.07	**0.80**	0.07	**0.82**	0.07	**0.81**	0.07	**0.82**	0.07

The machine learning algorithms were implemented with the data mining software Weka 8.0 [79], and plots were created using MATLAB 2018a. The results are shown in 3 subsections.

4.6.1 The Gut-Bone Relationship

As mentioned before, we have used the DXA measures of BMD as the measure of bone health. We compared six different metrics as dependent variables for prediction,

- V_1 = Total body BMD (TBBMD),
- V_2 = Spine BMD (SPBMD), and
- V_3 = Hip BMD (HBMD).

For the RF model, 1000 iterations and a batch size of 100 was used. The LWL, RC, bagging and voting algorithms were trained using a batch size of 100. Ten

iterations were used in bagging and RC. The LWL, RC and bagging models used the RF model as their base models. To evaluate the accuracy of prediction we used Pearson's correlation coefficient (CC) and the mean absolute error (MAE) (see section 2.4.2).

The CC obtained after predicting different measures of BMD, using five regression techniques are shown in Table 4.2. There exists a high positive correlation between the predicted and actual BMD, which implies that the GM has a strong relationship with bone health. We have highlighted the cases where correlation was > 0.8. Maximum correlations of 0.81, 0.80 and 0.83 were achieved while predicting the TBBMD, SPBMD and HBMD, respectively. The maximum absolute error (MAE) was between 0.06 and 0.08. We also predicted three measures of BMC, also measured using DXA; total body BMC, spine BMC and hip BMC, and predicted them with CCs of 0.78, 0.76 and 0.77 respectively.

4.6.2 Microbe Analysis

Feature selection method was used to rank the microbes based on their correlation weights. Figure 4.1 shows the correlation weights of all 221 GM genera, while predicting the TBBMD. We ranked all the genera while predicting the different measures of BMD and BMC. A total of 55 out of 221 different genera of the GM were identified with a correlation weight > 0.3. Figures 4.2 shows the broader phyla to which these 55 GM genera belong. *Actinobacteria*, *Bacteriodetes*, *Firmicutes* and *Proteobacteria* phyla had the most number of positively correlated genera.

The mean TBBMD of the 26 participants was 1 g/cm^2. We calculated the average unit of all the top 55 positively correlated GM found in the fecal samples from all participants. We also calculated the difference in the amount of each genus found in participants with TBBMD above and below the mean (1 g/cm^2). Figure 4.3

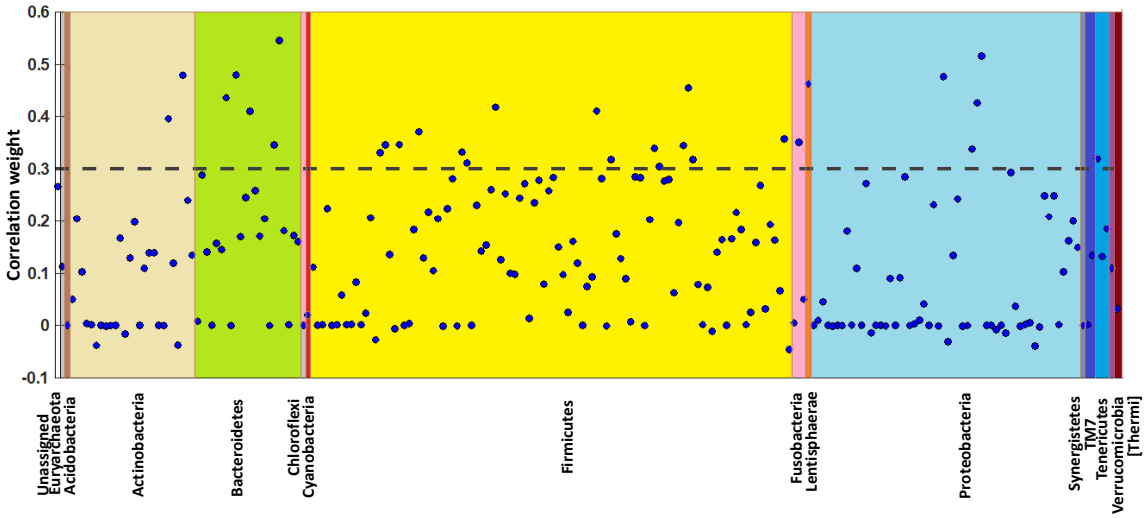


Figure 4.1: Correlation weights of 221 genera of GM while predicting the TBBMD

shows average units of some microbes (positively correlated, unit$>$ 0.005) found in participants with TBBMD above and below 1 g/cm^2.

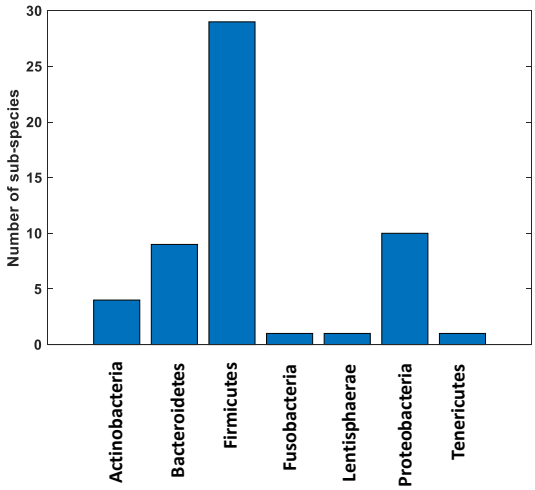


Figure 4.2: Positively correlated GM and number of sub-classes for each phylum

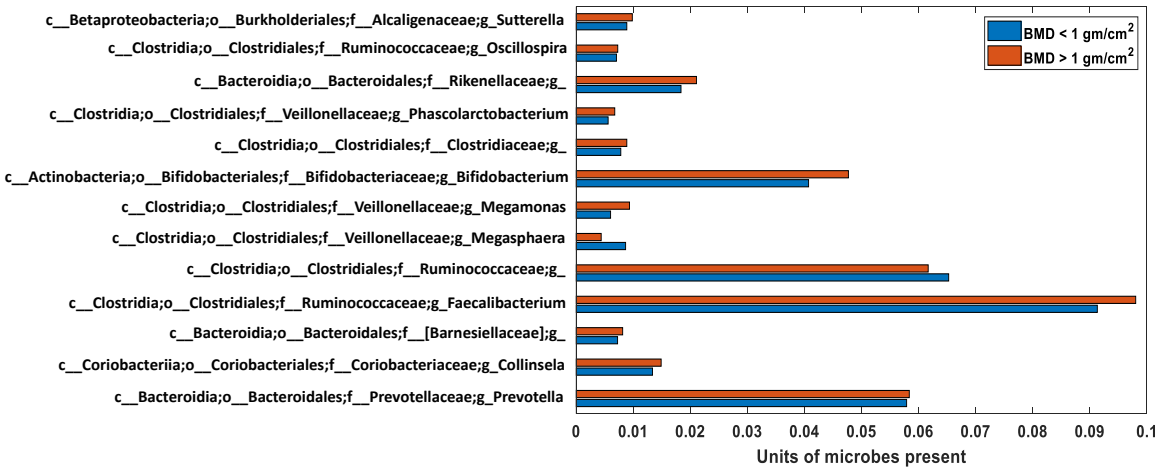


Figure 4.3: Average unit of significant bacteria present in two groups of participants discriminated based on their TBBMD (unit> 0.005)

4.6.3 Calcium Absorption Analysis

The prediction results showed CCs ranging from 0.51 to 0.58 and MAE of 0.05. We used the same feature selection method to identify microbes that were highly correlated with fractional calcium absorption. The five most highly correlated genera were,

- *f_Rikenellaceae* (0.36)
- *f_Mogibacteriaceae* (0.32)
- *f_Ruminococcaceae* (0.31)
- *f_Lachnospiraceae* (0.31)
- *f_Christensenellaceae* (0.31)

4.7 Discussion

Through this study, we observe a strong relationship between the GM community and the overall bone strength. Results showed that there exists a positive high correlation between the microbes and the six DXA measures of BMC and BMD. Five regression techniques were used to predict these measures. We also identified the microbe genera which were responsible for such high correlations. *Firmicutes*, *Bacteroidetes*, *Proteobacteria* and *Actinobacteria* were the major phyla, positively correlated with the bone health. Twenty nine genera from the *Firmicutes* phylum were identified as positively correlated (> 0.3) GM. We have further shown the proportion of some positively correlated microbes present in two groups of participants, based on their TBBMD. From Figure 4.3, microbes *Clostridiaceae*, *Bifidobacterium*, *Faecalibacterium* and *Prevotella* were present in higher proportion in participants with TBBMD $> 1g/cm^2$. On the contrary, *Megasphaera* was found more in participants with TBBMD $< 1g/cm^2$.

Many of the 55 microbes, which were identified as positively correlated microbes, have been identified in previous studies as microbes that were responsible for bone development in a direct or indirect manner. *Lactobacillales* and *Bifidobacteriales* having previously identified as probiotics that help increase bone mass in mice [94, 170, 191, 194]. *Clostridiaceae* has been found to be beneficial for fractional calcium absorption [198]. The calcium absorption analysis showed high correlation with *Rikenellacae*, *Mogibacteriaceae* and *Christensenellaceae*, which have been found to be significantly correlated with bowel movement frequency in healthy Japanese adults (n = 516) [142]. *Lachnospiraceae* and *Bifidobacterium* are suggested to be involved in SCF fermentation in [194] .

In Table 4.3, we present some statistics about some of these important microbe

classes, which have been previously reported to affect bone health. These microbes have high correlation weights and are present in greater quantity in participants with higher TBBMD.

Table 4.3: Analysis results of some important GM communities

	Correlation weights				Units present	
Microbes	HBMD	SPBMD	TBBMD	CalAbs	TBBMD < 1	TBBMD > 1
o_Lactobacillales	**0.30**	**0.38**	**0.39**	0.24	0.0037	0.0038
o_Bifidobacteriales	**0.31**	**0.42**	**0.31**	0.22	0.0407	0.0477
f_Clostridiaceae	0.26	**0.30**	**0.33**	0.20	0.008	0.009
f_Ruminococcaceae	0.27	**0.37**	**0.42**	**0.31**	0.0984	0.1054
g_Prevotella	**0.40**	**0.38**	**0.54**	0.02	0.0579	0.0584
g_Faecalibacterium	0.22	**0.33**	**0.32**	0.15	0.0617	0.0653
g_Eubacterium	**0.34**	0.27	**0.38**	0.24	$7e^{-5}$	$9e^{-5}$

Therefore, this study identifies the important GM communities, and provides validation that they can play a major role in bone regulation. This is one of the few studies where the relationship between GM and bone health has been presented using clinical trials of human participants (adolescent females). It is important to note that while this study reveals high correlation between the GM and bone health, it does not mean that this relationship is causal. In summary, we can infer that there may exist a strong relationship between the GM and bone health and we identified several microbes that can affect this relationship. Further experimental studies are required to learn more about the role of GM in bone regulation in larger and diverse population. More studies are required to investigate how individual microbes affect bone health. Furthermore, more human trials are required to study the effects of the GM on various biological studies and how they can be used as markers for diseases.

Chapter 5

STUDY III: PROCESSING SPEECH TO UNDERSTAND HUMAN COMMUNICATION IN AN ORGANIZATION

5.1 Overview

This chapter presents a study where the relation between human communication and productivity in an organization is analyzed by processing speech recordings from the employees in small scale software company. Though it is a truism that communication contributes to organizational productivity, there are surprisingly few empirical studies documenting a relationship between observable interaction and productivity. This is because comprehensive, direct observation of communication in organizational settings is notoriously difficult. Seventy-nine employees working within a software engineering organization wore microphones and had their speech recorded during working hours for a period of approximately 3 years. From the speech data, we infer when any two individuals are talking to each other and use this information to construct a communication graph for the organization for each week. We use the spectral and temporal characteristics of the produced speech and the structure of the resultant communication graphs to predict the productivity of the group, as measured by the number of lines of code produced. The results indicate that the most important speech and graph features for predicting productivity include those that measure the number of unique people interacting within the organization, the frequency of interactions, and the topology of the communication network. This work was a collaboration between Hugh Downs School of Human Communication and the school of Electrical, Computer, and Energy Engineering at ASU.

5.2 Background

The "structural imperative" in network research [196] suggests that we can represent any organization as a network and look at the network as a determinant of behavior, culture, and the individuals within the organization. Organizational networks are generated and populated by human beings who are active agents with intentions, knowledge, and the ability to rationalize their actions. From interactions between individuals in an organization we can derive certain qualitative aspects like behavior, intentions, emotions and inter-employee relations of a workplace. These aspects play a large role in the effectiveness and productivity of an organization. In this paper we aim to objectively study this relationship between productivity and communication.

While productivity is relatively straightforward to measure, existing studies measure communication indirectly, either through member self-reports of communication on rating scales [124, 150], through external raters' evaluation using global scales that assess communication behavior [124], as communication technology investment [103], or through questionnaires measuring more distal constructs such as communication satisfaction or perceived effectiveness [158]. While these studies are useful, they can be challenged on the grounds that perceptions of communication do not correspond to actual communication behavior [99]. Direct observation is the "gold standard" for measuring communication and provides the most rigorous test of the communication-productivity relationship. Studies based on direct observation are needed to supplement and validate current literature.

In this work, we estimate inter-employee communication networks in a software engineering organization using speech recordings. All employees were recorded during their working hours for a period of approximately 3 years, and weekly communica-

tion graphs were estimated based on the detected speech. We use a simple speech activity detector, combined with inter-recorder correlations, to detect interactions between individuals and to construct daily communication graphs. In addition, we also measure several speech features that describe the speaking style of each individual. These include, pitch, temporal features (energy, zero crossing rate), spectral features (spectral centroid, spectral flux etc), and cepstral features (MFCCs). Together with the communication graphs and speech analysis we analyze the relationship between productivity and human communication in an organization.

We construct the communication graphs on a weekly basis and study the relationship between these graphs and the productivity of the software engineering organization, as measured by the effective lines of code (LOC) produced by the employees that week. Other studies in the literature have found that LOC is an effective measure of productivity in software organizations [177, 188]. Our results reveal a moderate correlation between the predicted productivity and measured productivity.

5.3 Organization Setting and Data Collection

We collected data from a software engineering organization called the Software Factory (SF), operating for three years at Arizona State University. The SF data is a unique, existing dataset that aimed to accomplish, as nearly as possible, ubiquitous observation of a set of 79 employees and clients of the organization. The dataset contains recorded audio data from participants between September-2002 and June-2005. Whenever they entered the dedicated facility, participants attached a digital recorder and lapel microphone, and logged in to a server which placed a time stamp on the recording. When leaving, they uploaded the recorded audio to a server for storage. The resultant dataset contains daily recordings of all SF employees and visitors (primarily clients) comprising 7000 hours of time synchronized recordings.

The participant recordings were created in digital speech standard (DSS) file formats, a compressed proprietary format optimized for speech. They were converted to an uncompressed WAV format using the Switch Sound File Converter software. The files were stored using a 6kHz sampling rate with 8-bits/sample.

In addition to the recordings, we analyzed the code written by employees at the SF. All codes were stored and managed using a Visual Source Safe (VSS) 6.0 repository. We used the VSS API to extract records from the repository. Each record included the filename, date, user, version, and changes, insertions, and deletions at check-in. From this information we were able to compute the number of lines of code at each check-in. In particular, we computed the total number of inserted, deleted and changed lines of code per employee per week. A total of 11276 entries of LOC were recorded staring from the first week of March-2003.

The SF dataset affords a unique opportunity to obtain a holistic picture of work activity and communication in a small organization over an extended period. In this work, we have used the audio recording from March-2003 to June-2005 (124 weeks), to build communication networks and extract speech features to predict the effective lines of codes obtained using VSS analysis.

5.4 Methods

In this section, firstly we have formulated the problem of predicting productivity, and then discussed the methods used to approach and solve the problem.

5.4.1 Problem Formulation

Our aim in this work is to find the relationship between productivity and the communication network. All analyses were done on a weekly basis. In case of communication graphs, individual interactions between any two individuals were detected

using a simple cross-correlation scheme (see section 5.4.2). Individual interactions were converted to a communication graph representing the frequency of interactions between any two individuals over the course of a week. From this graph , we extract a set of features that describe the topology of the resultant network and denote that by, $G_w \in \mathbb{R}^{1\times f_g}$, where f_g is total number of graph features. In addition, we also extract several speech features from the daily recordings and calculate two statistics (mean and variance) for these features across the whole week for all participants. These are defined as, $S_w \in \mathbb{R}^{1\times(2\times f_s)}$, where f_s is total number of speech features. Thus, we had a total communication feature space defined by $C_w : (G_w \| S_w) \in \mathbb{R}^{1\times(f_g+2\times f_s)}$ (where $\|$ is the concatenation operator).

We describe the details of how we estimate the communication graph and the feature extraction in the sections below. We then describe how we predict productivity using these features.

5.4.2 Communication Graph Analysis

To construct the communication graphs, we use cross channel signal analysis. The entire process of graph analysis can be subdivided into two main blocks, the construction of speech cross-correlation graphs and graph feature extraction as shown is Figure 5.1.

Speech Cross Correlation Graph

As a pre-processing step we normalized the data by the mean to remove direct current offset. We investigated preliminary conversation detection performance on the SF data by using a two-stage approach. The first stage identified continuous segments of speech using a simple energy-based detector; in the second stage, we use a pair-wise cross-correlation between one speaker's channel and the remaining channels

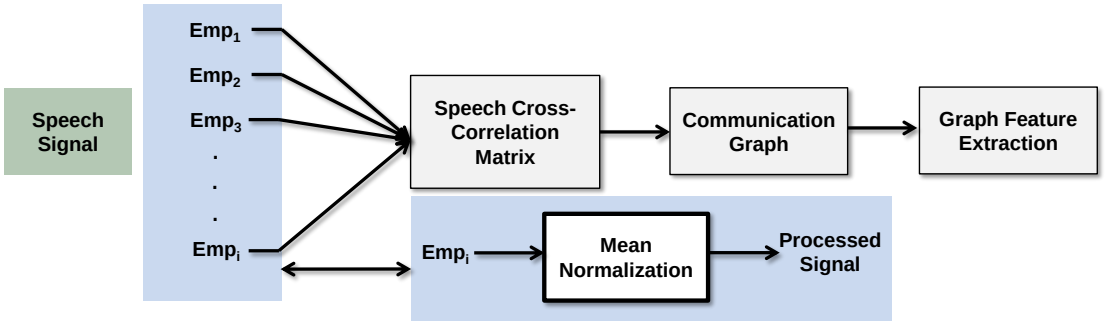


Figure 5.1: Process chain for communication graph analysis

to detect with whom that person was speaking. This done using a 36 seconds window. The basic idea behind this approach is that, if two individuals are speaking, their microphone will pick up each-other's speech and cross correlation will be high. A cross-correlation matrix was constructed using mean correlation weights between participant pairs across day. This matrix represents a proxy for the frequency of interactions between any two individuals. The same data can also be used to detect individual interactions and compare against manually coded data. Pairwise conversations between two speakers were detected by the algorithm and were presented to a listener for manual coding. The daily cross-correlation matrices, which represent a proxy for frequency of interaction between two speakers, were averaged over the week to construct weighted communication graphs, with participants as nodes and the correlation weights as edges. Figure 2 shows the communication network on a typical Monday. There were 9 participants present on that day, from which the communication graph was estimated, each node represents the participant ID and the link distance represents the frequency of interactions between participant pairs.

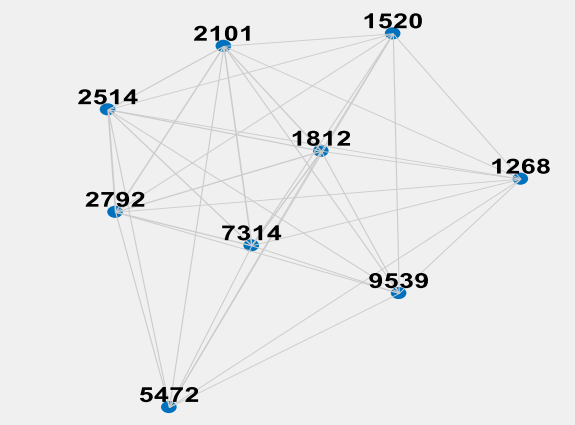


Figure 5.2: Communication graph on a typical Monday (Nodes: Participant Id, Edges: weighted based on average cross correlation)

We validated pair-wise communication using external raters. Communication were labelled for every 15 secs based on,

- Silence/noise (**0**)
- Employee 1 speaking (**1**)
- Employee 2 speaking (**2**)
- Both employees speaking (**12**)

A total of 28890 seconds of pairwise recording were labeled. We computed the accuracy of communication detection using confusion matrix as shown below. Although, variable weights were used for each label, we achieved an overall accuracy of 87%, while detecting pair-wise communication.

Graph Feature Extraction

After the graph is constructed, we extract several topological features that aim to describe the nature of daily interactions. A total of 11 graph features (see section

Detected (sec) \ Actual (sec)	Silence or Noise (**0**)	Emp1 Speaking (**1**)	Emp2 Speaking (**2**)	Both Speaking (**12**)
0	**22470**	240	285	120
1	1065	**915**	0	240
2	690	45	**495**	210
12	210	630	75	**1200**

Figure 5.3: Confusion matrix evaluating the accuracy of communication detection

2.1.3) were investigated in this work:

Basic graph descriptors

- *Number of edges*: The total number of edges or links present in the graph.
- *Number of nodes*: The total number of active nodes present in the graph.
- *Average degree*: Defined as $\frac{2m}{n}$, where n is the number of nodes and m is the number of edges.
- *Number of connected triples*: A count of the number of connected triples in the graph.
- *Number of cycles in a graph*: Defined as $m - n + c$, where m is the number of edges, n is the number of nodes and c is the number of connected components.
- *Graph energy*: The sum of the absolute values of the real components of the eigenvalues of the graph.

Graph centrality measures

- *Degrees*: The average number of edges adjacent to a node.

- *Average neighbor degree*: The average degree of adjacent or neighboring nodes for every vertex. We took the average of this measure across all nodes.

- *Eigen centrality*: The i-th component of the eigenvector of the adjacency matrix gives the centrality score of the i-th node of the network. The average eigen centrality across all nodes was computed for this study.

Laplacian features

We extract the following features from the graph Laplacian,

- *Graph spectrum*: The average over all Laplacian eigen values was used for this study.

- *Algebraic connectivity*: The second smallest eigenvalue of the Laplacian graph.

These features are estimated based on daily graphs. We average over the week to compute a weekly graph feature vector, $G_w \in \mathbb{R}^{1\times 11}$, where 11 is total number of graph features investigated.

5.4.3 Speech Analysis

In addition to the graph features, we extract speech features from every speaker from the data. These features carry information about speaker identity and various aspects of affect, which are important characteristics for predicting productivity. We describe the details below.

Speech Feature Extraction

Speech features are extracted independently for every speaker (e.g. every recording channel). Prior to feature extraction, we remove the DC offset, apply hamming

window, and split the data into 1-sec speech segments. All features are extracted at this scale. These features are described in details in section 2.1.2.

A total of 35 different features were obtained from speech. They are shown below,

- **Pitch**: We calculated the autocorrelation function of the signal to estimate the pitch features. They are the *fundamental pitch frequency*, *12 harmonics* and *harmonic ratio*.
- **Temporal**: *Zero-crossing rate*, *energy* and *energy entropy*.
- **Spectral features**: *Spectral centroid*, *spectral spread*, *spectral entropy*, *spectral flux* and *spectral rolloff*.
- **Cepstral features**: *13 Mel-frequency cepstrum coefficients (MFCC)*: We estimate these on 20 ms frames and average over 1 sec window.

We calculated the mean and variance of these features over the working days of a week to compute weekly speech feature vectors defined as, $S_w \in \mathbb{R}^{1\times(35\times2)}$, where 35 is total number of graph features investigated and 2 is the number of statistics computed for each speech feature.

Thus, together with the graph and speech features we had a combined communication feature set defined by, $C_w \in \mathbb{R}^{1\times(11+70)} = \mathbb{R}^{1\times81}$.

5.4.4 Prediction Technique

We have investigated three regression techniques with feature selectors to predict the measure of productivity. The prediction process included normalization of the communication features (C_w), a feature selection method and a regression model, as shown in Figure 5.4. The entire process chain was carried out with the data mining software Weka 8.0 [79].

Feature selection was used to identify the most relevant communication features for measuring productivity. The combined feature set of communication graph features and speech features were normalized before the feature selection process. The feature selection method was used based on ranker search protocol.

- *Classfier attribute evaluator* (CAE-GPR): We used the CAE feature selector with Gaussian process regression (GPR) model as the prediction algorithm and the correlation coefficient as the evaluation measure, with 10-fold cross validation. Features with correlation weights above 0 were selected for prediction analysis.

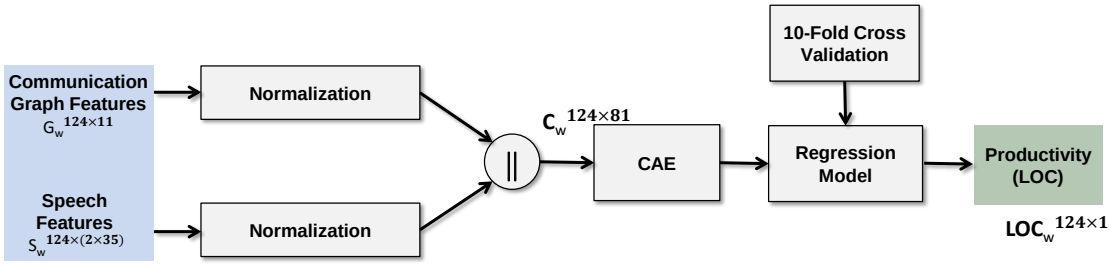


Figure 5.4: Prediction process chain

Three well known regression models were used, each with 10-fold cross validation,

- *Linear regression (LR)*: see 2.3.2 (I)
- *Gaussian process regression (GPR)*: see 2.3.2 (II)
- *Support vector regression (SVR)*: see 2.3.2 (III)

5.5 Results

As mentioned before, we have used the total lines of codes (LOC) per week as the measure of productivity in the SF. We compared three different metrics as dependent variables for prediction,

- V_1 = Changed + inserted + deleted LOC,
- V_2 = Changed + inserted - deleted LOC, and
- V_3 = Changed + inserted LOC.

These variables were normalized, based on, (a) number of '*hours*' spent on coding and, (b) number of '*coders*' involved each time the LOC were checked in. Therefore, a total of 9 variables were identified as measures of productivity with and without normalization. All the measures were converted to log scale to reduce dynamic range. We used the correlation coefficient between predicted productivity and measured productivity to measure model accuracy.

The cross correlation coefficients obtained after predicting the dependent variables using three regression techniques are shown in Table 5.1.

Table 5.1: Correlation coefficients between productivity measures and communication features using 3 regression techniques

Variables	**GPR**	**LR**	**SVR**
V_1	**0.42**	**0.41**	**0.49**
V_1 norm '*coders*'	0.36	0.31	0.36
V_1 norm '*hours*'	0.37	0.27	0.38
V_2	**0.40**	**0.43**	**0.42**
V_2 norm '*coders*'	0.28	0.36	0.26
V_2 norm '*hours*'	0.40	0.23	**0.43**
V_3	**0.44**	**0.41**	**0.48**
V_3 norm '*coders*'	0.38	**0.44**	0.38
V_3 norm '*hours*'	**0.40**	0.34	**0.40**

We computed the correlation weights for each communication feature while predicting different measures of productivity. We also computed the correlation coefficients between the predicted and the actual measures of productivity for each selected feature. Figure 5.5 shows the correlation coefficient achieved while predicting V_1.

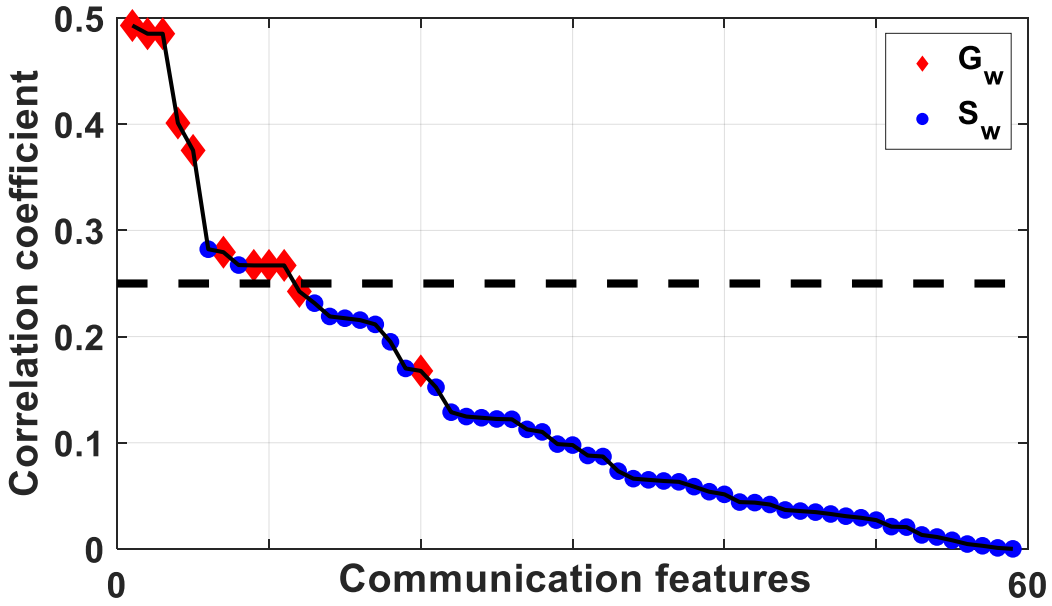


Figure 5.5: Correlation coefficients for each selected communication feature

The graph features had more correlation weights compared to the speech features. The top 12 features ranked based on correlation weights using CAE-GPR are shown in Table 5.2. From Table 5.2, we can see that 10 graph features and 2 speech features played major roles to predict measures of productivity.

5.6 Discussion

The best choice of a productivity measure can be argued here. Both changed and inserted lines of codes are important measures that cannot be neglected, when it comes to programmer productivity. The inclusion of deleted lines of codes is debatable, as

Table 5.2: Communication feature ranks (G:graph feature, S:speech feature)

Ranks	**Features**	**Feature type**
1	Graph energy	G
2	Number of nodes	G
3	Average neighbor degree	G
4	Algebraic connectivity	G
5	Eigen centrality	G
6	MFCC 1st coefficient (mean)	S
7	Number of cycles	G
8	Pitch frequency (var)	S
9	Average degree	G
10	Degree	G
11	Number of edges	G
12	Graph spectrum	G

those can be errors or bugs in previously-written codes, that can said to be counter-productive. But at the same time, it can argued that deletion mean shortening of code or making it more compact using improved logic, which is an important aspect of productivity. This is why we introduced 3 dependent variables. From Table 5.1, the average correlation coefficient lies between 0.375 and 0.45, which suggests that there exists a moderate relationship between the communication graphs and productivity. Correlation coefficients above 0.45 have been shown in bold.

The results from Table 5.2 indicate the communication graph features played a more important role than speech features in predicting the dependent variables. Among the top graph features, graph energy tells us about the structural complexity

of the network; algebraic connectivity, number of nodes and average neighbor degree signify the total number of employees and frequency of interactions between them. From the speech features, the mean MFCC coefficients are likely tapping into the number of speakers in the graph; the spectral roll-off and the zero-crossing rate variability are likely measuring the number of speakers and frequency of interactions. It is interesting that pitch variability is a measure of productivity - this could be a proxy for gender diversity in the organization, although this most certainly requires additional study.

It is important to note that while this study reveals some relationship between communication and productivity, it does not mean that this relationship is causal. It is unknowable from out data whether it is the productivity that induces a change in the network or whether the network induces a change in productivity.

This study is very unique in terms of organizational social studies as it involves quantitative analysis showing the relationship between human communication network and productivity in an organization. We have used speech recordings from employees in a software organization to estimate communication networks and extract speech features over a period of 3 years. Effective lines of code was used as the measure of productivity which we attempted to predict using both communication network and speech features. It was found that there exists a moderate relationship between human communication and productivity and it depends on the number of employees, the frequency of conversation between them and the topography of the network. Thus, with the help of such unique features and ML, we were able to make certain assumptions about a complex system. Further investigation can be done by including other forms of communication like, email, texts etc. Besides that, more complex graphs with multiple modules (employee, project, task)can be investigated, which can be a better representative of an organizational setting model.

Chapter 6

STUDY IV: DECODING HUMAN INTENT USING MULTI-MODAL SENSOR DATA

6.1 Overview

The field of human-computer interface has been evolving rapidly. Despite the phenomenal advances in the computational power of electronic systems, human-machine interaction has been largely limited to simple control panels, such as keyboards and mice, which only use physical senses. Consequently, these systems either rely critically on close human guidance or operate almost independently. A richer experience can be achieved if cognitive inputs are used in addition to the physical senses. Towards this end, this chapter introduces a simple signal processing and classification technique to predict human activity using data collected from a wearable system that consists of a motion processing unit and brain-machine interface. We show that our system can successfully employ cognitive indicators to predict human activity.

6.2 Background

The way in which humans interact with machines is rapidly evolving. The complexity of human involvement is sometimes hidden behind the seemingly simple interaction methods itself, which can use physical, cognitive or affective inputs [96]. The physical aspect is related to the mechanics of interaction. It is the most commonly used method today, since physical interaction allows for simple devices, such as keyboards, mice and displays. However, in this form of interaction, machines are passive agents that can only respond to physical inputs. Cognitive inputs can significantly

improve the user experience by facilitating intellectual interaction, e.g., household electronics controlled by brain activity. Similarly, affective inputs can help machines understand emotional state of the user, e.g., a music player selecting the tracks based on user's mood. Consequently, we can enable machines to understand what their user wants by enriching current physical interaction with cognitive and affective inputs.

Pervasive use of cognitive and affective inputs for human-machine interaction relies critically on two technologies. First, there is a need for methodologies and algorithms that can process physiological signals, such as electroencephalogram (EEG), to decode the users' intentions and needs. To enable a seamless interaction, this processing has to be in situ and real-time. Therefore, the second requirement is a wearable system that can monitor and process the relevant physiological signals. This capability can enable a symbiotic human-machine relationship by providing a continuous interaction and real-time feedback between the human and machines.

This paper first presents a simple wearable system prototype capable of sensing, processing and communicating user motion. Together with a commercial brain machine interface (BMI) [58], this system enables us to analyze user gestures and brain activity reflected by EEG signals. In order to do so, we designed an experiment involving a repetitive task, i.e., moving small boxes from one side of a table to the other side. Before completing this task, the user may be given a second task to stimulate cognitive activity, e.g., drawing a picture. If the second task is given, the user starts performing it upon completing the repetitive task. Our objective is to determine whether the second task is given or not by monitoring cognitive activity. To achieve this objective, we developed a methodology to analyze the EEG signals with the help of timing information extracted from the hand movement data. We successfully show that EEG signals can be used as an indicator of future user activity. Therefore, this work serves as a first step towards a wearable system that can decode human intent

using multi-modal sensor data.

The term human intent has been used to express the partial intent sufficient enough in the context of a specific target application. For example, in a car disassembly task, the intent is the desired direction and magnitude change in the position of the end-effector [59]. Similarly, a driver's turn intent [29, 45] is sufficient to represent the intent in driver assistance applications. In our context, human intent refers to whether the user is preparing to perform a second task upon completing the current one or not. It has been shown that the human intent can be predicted by low-level limb movements [136], vision processing [131], and using physiological signals [105]. There have been numerous attempts at human decoding intent by classification of single modalities, such as speech [37], body movement [100], head movement [2], gestures [154], facial expression [146], eye movement [21], hand pressure [123] and brain-activity [12]. However, single modality is not sufficient for robust human-machine communication [132], which is the first step towards symbiosis. Indeed, intuitive *human-human* communication leverages multiple modalities, such as speech, gestures, mimics, and body language [165]. In this work, we employ two modalities. The main modality is EEG processing with the help of the Emotiv EPOC+ EEG headset. To improve the effectiveness of the EEG signal processing, we also use a tri-axis gyroscope and accelerometer data which encodes the hand movements.

Brain machine interfaces (BMI) have become a promising technology to interface directly with computers [109]. Impressive progress in novel BMI technologies is discussed in [126]. The first symbiotic BMI was introduced by Mahmoudi and Sanchez [115]. This approach has an adaptive actor interface to the motor cortex and a critic that evaluates the value of actor's actions using Nucleus Accumbens. The authors show that a continuous perception-action-reward cycle enables operation in changing environments by adapting the BMI decoder.

6.3 Multi-Modal Intent Decoding Methodology

The proposed multi-modal intent decoding methodology is illustrated in Figure 6.1. The flow consists of hand motion processing and EEG signal processing. Hand motion is monitored using a wearable prototype described in Section 6.3.1. The purpose of hand motion processing is to provide timing information about the user activity. More precisely, we segment the EEG data non-uniformly into multiple blocks with the help of hand motion processing. EEG processing itself consists of conditioning the raw data, segmentation and feature extraction using spatiotemporal eigenspectrum construction. The following subsections detail each step of this flow.

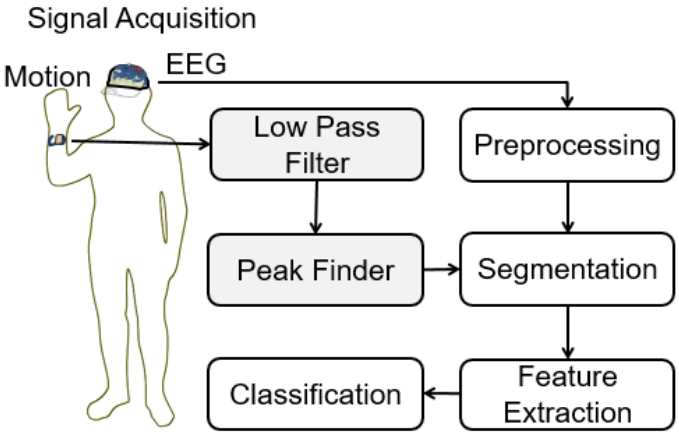


Figure 6.1: Multi-modal intent decoding methodology flow.

6.3.1 Wearable System Prototype

We record the brain activity using Emotiv EPOC+ EEG headset [58] shown in Figure 6.2. The Emotiv headset uses 14 channels and two reference electrodes The first reference is an absolute voltage reference called common mode sense (CMS). The second one, referred to as driven right leg (DRL), is a feedback cancellation system to float the reference level on the common mode body potential. The raw EEG data

is stored using wireless connection at a sampling rate of $128Hz$.



Figure 6.2: (a) Emotiv EPOC+ headset, (b) motion processing unit.

The user hand motion was recorded at a sampling rate of $100Hz$ using a motion processing unit that integrates a tri-axis gyroscope and accelerometer, and a TI-CC2650 microcontroller [181] on a flexible polyimide substrate. It senses motion, processes the raw data locally, and supports both Bluetooth low energy and Zigbee wireless communication protocols. The maximum supported data rate is $192kbps$, and the average power consumption is $12.2mW$ at $1.44kpbs$ throughput. Our prototype device occupies $3.8cm \times 3.8cm$ and weighs $50mg$ including the programming interfaces. Its area and low weight allow us to attach it directly to the skin or clothing. Detailed design specification can be found in [75].

6.3.2 Hand Motion Processing

As a reminder, the repetitive task in our experiment is moving a set of boxes on a table from one side to the other side. The repetitive task is said to be complete, if all the boxes are moved from one side to the other. To infer the start time of each new round, we employed accelerator and gyroscope measurements. The accelerator gives acceleration along $x-$, $y-$ and $z-$ axis in terms of gravitational constant g.

Gyroscope measures the rate of rotation, i.e., angular velocity, in three components called roll, pitch and yaw. The raw tri-axial accelerometer and gyroscope readings are plotted in Figure 6.4(a) and Figure 6.4(b), respectively. Both gyroscope and accelerometer readings are preprocessed using a 5-point moving average filter to suppress the measurement noise. Then, the filtered data is used to find the angular velocity of the hand. The local minima of the angular velocity indicate the start time of each new round. We find the local minima using a peak finder algorithm, and identify the start times, as illustrated by the vertical lines in Figure 6.4.

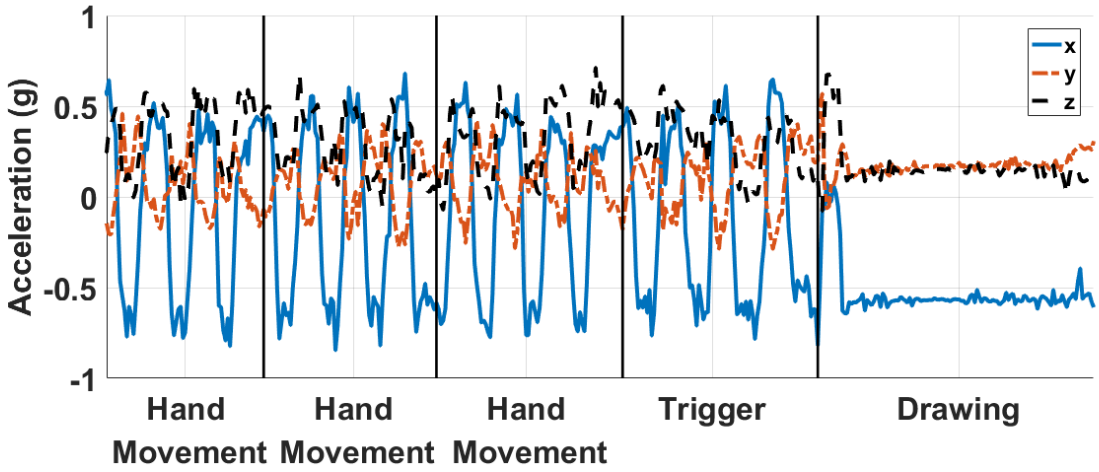


Figure 6.3: Tri-axis accelerometer data.

6.3.3 EEG Signal Processing

To remove the irrelevant artifacts from the measure data, we first filter the raw EEG data using a high pass filter with a pass band starting at 4Hz and a notch filter at 50Hz. Then, we use the time instances from motion processing to segment the input data into multiple blocks. We used spatiotemporal correlation to extract features from multi-channel EEG data [202], to predict the trigger event. One way to capture this is through spatiotemporal eigenspectra (described in section 2.1.4),

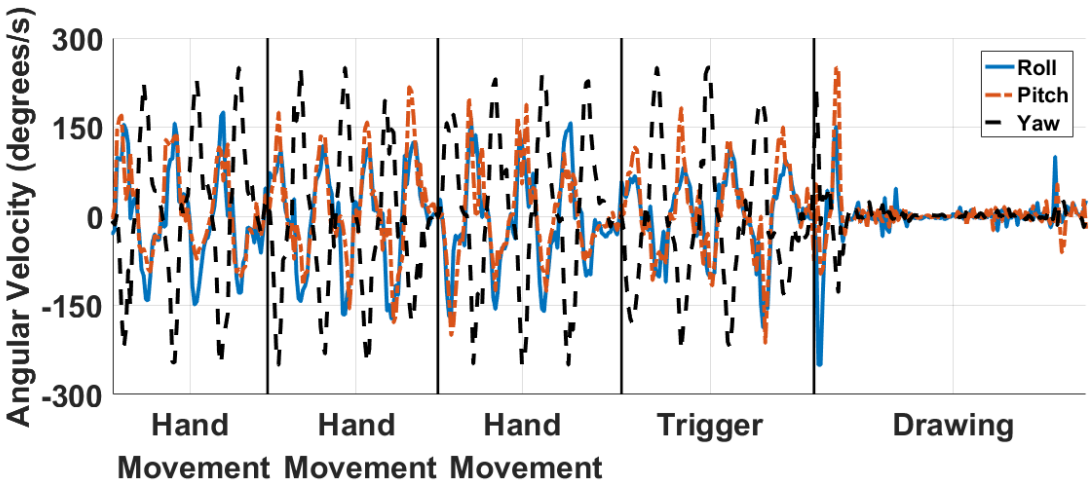


Figure 6.4: Roll, pitch and yaw components of the gyroscope data.

which shows fluctuations in eigen values for successive events.

Parameters used in this work: We used 14 channels EEG ($n_c = 14$) sample at $T_s = \frac{1}{128}s$ intervals. We used three different delay scales ($d = 3$), and minimum delay of $\tau = 1$ sample. Finally, the EEG data is segmented non-uniformly into 5 blocks using the data obtained from the motion processing. The first three blocks correspond to the repetitive task performed by the user, the fourth block contains the trigger event, and the last block covers the cognitive task.

6.4 Experimental Results

6.4.1 Experiment Setup

During our experiments, subjects wore the Emotiv EPOC+ headset and the motion processing prototype. Then, the subject remained seated on a chair in front of a desk, as shown in Figure 6.5. We placed four small boxes on the one side of the table, and asked the subject to move all boxes from one side to another. The task is considered complete when all the boxes are moved to the opposite side. The subject

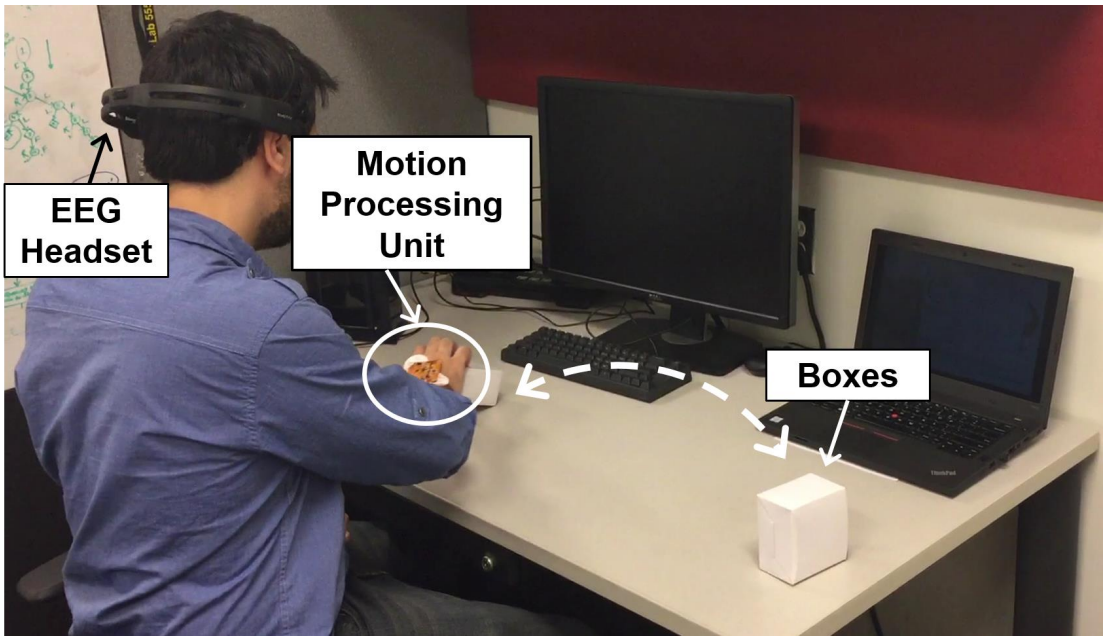


Figure 6.5: The experimental setup.

repeated the whole task until s/he was asked to stop, or draw an arbitrary object on a piece of chapter upon completing the current task. Four neurologically healthy subjects participated in the study. Subjects were not under the influence of any medication that could interfere with EEG. The execution of the repetitive task and drawing lasted approximately $50s$, and each subject repeated the experiment certain number of times. Four subjects participated in the experiments and data from 22 such experiments were used for analysis.

6.4.2 Analysis of the EEG Data

We first used the accelerometer and gyroscope data to identify the start and end times of each repetitive task. Then, this information is used to divide the EEG data into multiple blocks labeled as *hand movement*, *trigger event* and *drawing*. Hand movement corresponds to the repetitive task. If the subject is asked to draw an

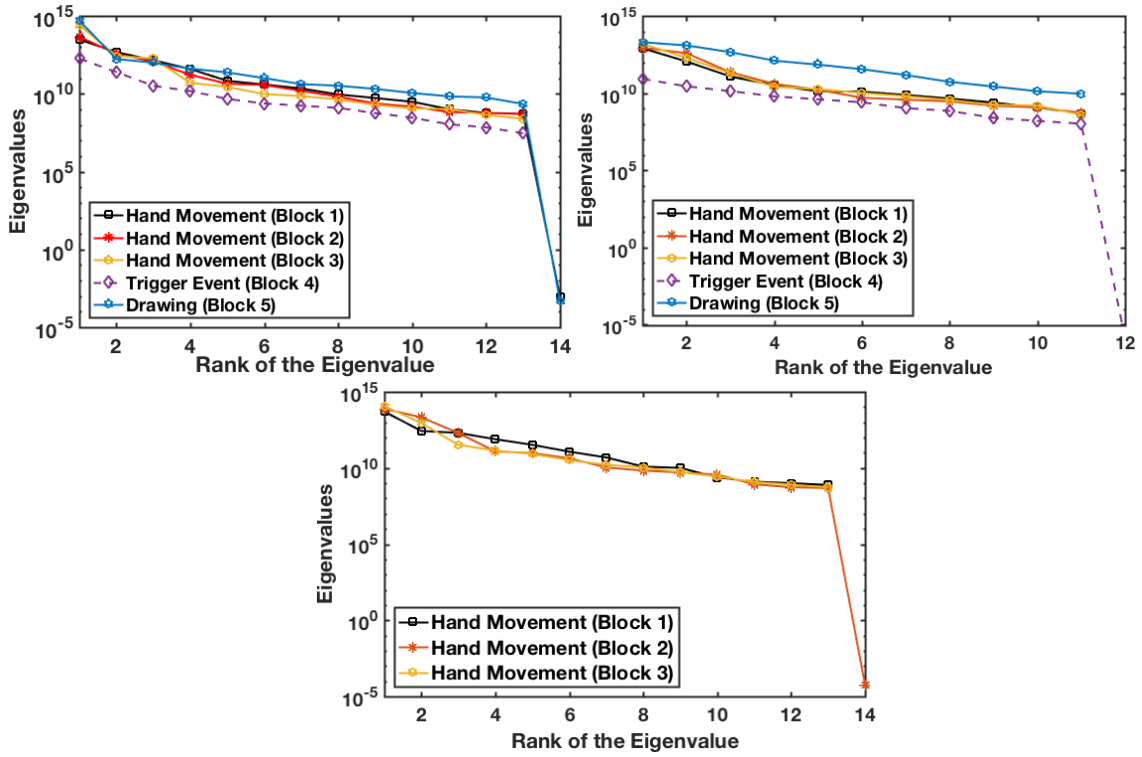


Figure 6.6: Space-delay eigenspectra for three different subjects. (a) Subject 1: Performed both the repetitive and drawing tasks, (b) Subject 2: Performed both the repetitive and drawing tasks.(The EEG channels were removed in this experiment due to sensor misplacement), (c) Subject 3: Performed *only* the repetitive hand movement.

object, we refer to the corresponding block as the trigger event. Finally, if the subject draws an object, the corresponding block is labeled as drawing.

The eigenvalues of the space-delay correlation matrix are shown in Figure 6.6. In particular, Figure 6.6(c) plots the eigenvalues for Subject 3, when he performed only the repetitive task. We observe that the eigenvalues follow a very similar pattern for each of block. Eigenvalues for two other subjects, who were asked to draw an object, are plotted in Figure 6.6(a) and Figure 6.6(b). The eigenvalues show a very clear distinction, when the trigger happens. This means that the subject starts thinking

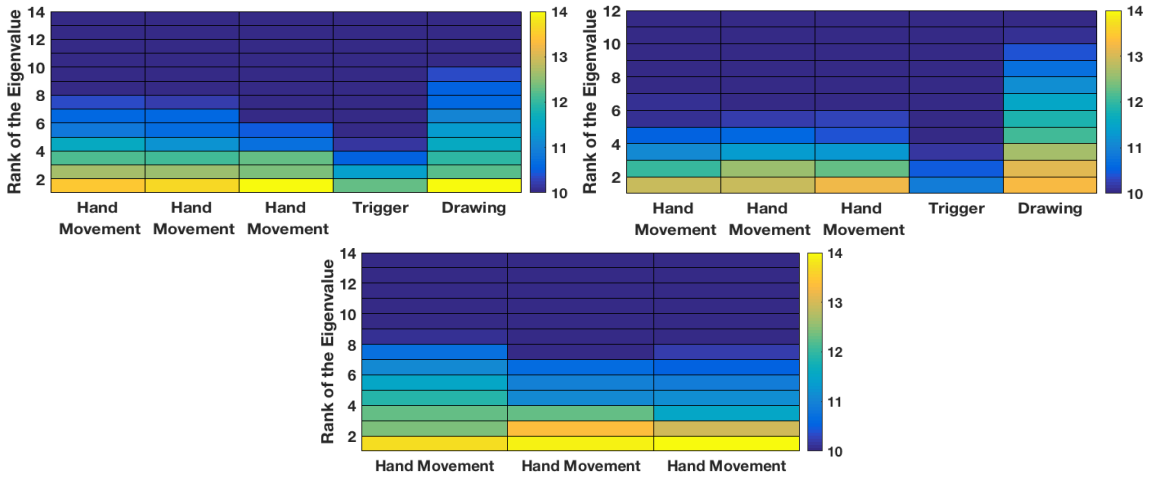


Figure 6.7: Eigenspectrogram of three different subjects. (a) Subject 1, (b) Subject 2, (c) Subject 3.

about the drawing before completing the repetitive task. This is clearly reflected in the eigenvalue distribution of the fourth block. Similarly, the spectra shown in Figure 6.7 shows that the trigger event has a distinctive signature. Each column in these plots correspond to a different block of time, while each row shows the eigenvalues at a given rank. All three blocks in Figure 6.7(c) have a similar behavior as they all correspond to repetitive task as in Figure 6.6(c). However, the spectra that correspond to the trigger event can be easily distinguished from the other blocks, as shown in Figure 6.7(a) and Figure 6.7(b). The second spectrum, that corresponds to Subject 2, used 12 out of 14 EEG channels, since two channels were discarded due to sensor displacement. This shows that the proposed approach can also work effectively with fewer channels. Correct predictions were made in 15 out of 22 cases.

6.5 Discussion

This chapter presented a methodology for cognitive activity detection to predict future human action for human intent decoding. The proposed approach employs a

wearable system that can monitor hand movements and sense EEG signals. Hand movement is used to divide the EEG signals into multiple blocks. For each block, the eigenspecta extracted from space-delay covariance matrix is used as to classify multi-channel EEG data. Our experiments show that the cognitive activity can be used successfully to predict future human activity and the spatio-temporal eigen spectra is an important EEG feature. Thus, from this study we can infer that human cognitive activities can be predicted before its execution using EEG recordings. Further investigation should be done with more human trials and performing different cognitive activities. This study also suggests the importance of multimodal signal analysis in brain machine interface.

Chapter 7

STUDY V: DISCRIMINATION BETWEEN NEUTRON AND GAMMA RAYS EMITTED DURING RADIOACTIVE DECAY

7.1 Overview

Radioactive decay can result in a plurality of emitted radiations, including neutrons and gamma rays. While these components of mixed radiation fields interact with matter via fundamentally different mechanisms, their conversion to light and subsequently to electrical signals via scintillation detectors results in signals that are intrinsically similar and difficult to discriminate by any trivial means. This chapter presents methods to discriminate between neutron and gamma ray pulses collected from using a scintillator. The reduced availability of ^3He is a motivation for developing alternative neutron detectors. ^6Li-enriched CLYC (Cs_2LiYCl_6), a scintillator, is a promising candidate to replace ^3He. The neutron and gamma ray signals from CLYC have different shapes due to the slower decay of neutron pulses. Discriminating neutrons in a mixed field of gamma rays and neutrons is a challenging task especially when the event rate is high. There have been other methods that successfully distinguish neutrons at less than 100 kHz event rates, but separating them at higher event rates has not been satisfactory. In this work, we propose two approaches to discriminate neutron and gamma ray pulses; one approach uses pulse decay time and energy spectral density and the other one uses the normalized cross correlation method. In this chapter, we have described the two novel approaches in details and discusses the experimental results. This work was a collaboration between the Radioactive Monitoring Devices (RMD) Inc. and the School of Electrical, Computer and Energy

Engineering, lead by the author under the guidance of Dr. Keith Holbert.

7.2 Background

For instruments capable of detecting multiple types of radiation, it is often important to distinguish detection pulses based on the type of radiation that induces them. For example, CLYC (Cs2LiYCl6) elpasolite crystals respond to both neutrons and gamma rays. The ^{6}Li-enriched CLYC scintillators have proven to be a promising candidate to replace ^{3}He [168] and for dual gamma-ray and neutron measurements for homeland security and nuclear safeguards [106]. For such applications, neutron-induced pulses may be separated from those produced by gamma rays using signal processing techniques, termed as pulse shape discrimination (PSD) methods. The neutron and gamma-ray pulse shapes depend on the type of scintillator used while the pulse amplitude is proportional to the radiation energy. For the CLYC detector, neutron pulses have a short, blunt peak and a long decay time, whereas gamma-ray pulses have a tall, sharp peak and a quick initial decay period. This PSD becomes increasingly difficult at high event rates, when multiple neutrons and gamma rays are detected within a short period, resulting in overlapping pulses. PSD for CLYC under such conditions is particularly challenging and is the primary focus of this work. Multiple PSD techniques that distinguish neutrons from gamma rays have already been developed. Some of the well-known PSD techniques include the charge comparison method [69], pulse gradient analysis [69], [43], simplified digital charge collection [69], neutron gamma model analysis [69], frequency gradient analysis [110], wavelet transform based method [204], bipolar trapezoid pulse shaping technique [62], zero-crossing method [134] and integral method [92]. All these techniques achieved satisfactory results for low event rates, but pileups were not considered. One of the other works that dealt with pileup events ia [156], in which artificial neural network

techniques were used to discriminate between neutron and gamma-ray pulses. With a liquid scintillation detector, several PSD techniques were compared by an artificial neural network classifier and pileup rejection technique [156].

In this work, we propose two digital signal processing approaches to distinguish neutrons and gammas. The first approach includes calculation of the time taken for a certain percentage decay of the detected pulse. In addition, the energy spectral density (ESD) of these pulses is calculated and the pulse shape discrimination (PSD) is carried out using the decay time and the ESD. This method is limited to static feature extraction which is not adaptive to the event rates, whereas the second approach employs dynamic feature extraction based on the event rate. We have performed both theoretical work and experimental data analysis for different event rates encountered while distinguishing neutrons from gamma rays. The experimental data was collected by a team at the Radiation Monitoring Devices (RMD), lead by Erik B. Johnson, who gave their valuable input to this research study.

7.3 Experimental Setup and Data Acquisition

Scintillation counting is a commonly used radiation detection method. It has many advantages compared to other techniques, including high detection efficiency, fast response time, wide choice of scintillator material and a large detection volume.

7.3.1 Experimental Setup

The CLYC material along with the cylindrical package is placed on a super-bialkali photomultiplier tube (PMT) and optically coupled with silicone grease (Saint-Gobain BC-630). The signal output of the PMT is connected to an analog-to-digital converter (ADC) board through a differential amplifier. The ADC is a AD9230 12-bit, running at 250 MSPS, whose output is connected to a Xilinx Zynq system on-chip (SOC)

field programmable gate array (FPGA) which is a Virtex 4. The Zynq SOC has programmable logic (PL) and processing system (PS). PL is used to collect the data and the PS, which has an advanced RISC machine (ARM) processor, is used to process the data using PSD algorithms.

7.3.2 Data Acquisition

To acquire actual neutron and gamma ray pulses for analysis, the above detector system was exposed to neutrons from an AmBe source located a few centimeters from the crystal. Gamma ray sources were also placed above the crystal to subject it to additional gamma rays. The number of gamma sources (^{22}Na, ^{137}Cs, and ^{57}Co) and their distance from the crystal was varied to control the event rate detected, while the position of the neutron source was fixed. The fifteen data sets summarized in Table 7.1 were acquired for about 0.3 s to 1.4 s each using a sampling frequency of 250 MHz. Some of the data were acquired using two gamma rays sources and one neutron source, and some of them using three gamma ray sources. Only one gamma ray source was moved to control the event rate. At high event rates, above 600 kHz, neutron and gamma ray pulses are distorted because of the significant presence of pileup.

7.4 Neutron-Gamma Discrimination using Decay Time Algorithm

Before applying the neutron detection algorithm, each event or pulse (n or γ) must be detected and extracted from the data stream. We employed a pulse detection technique that efficiently separates the events from pileups during high event rates.

Table 7.1: Experimental Data

Event Rates (kHz)	Gamma Sources	Neutron Source	Gamma to CLYC distance(inches)	Pulses Generated
6	$^{22}Na, ^{137}Cs$	AmBe	15.625	n and γ
13	$^{22}Na, ^{137}Cs$	AmBe	9	n and γ
92	$^{22}Na, ^{137}Cs$	AmBe	2	n and γ
370	$^{22}Na, ^{137}Cs$	AmBe	-	n and γ
583	$^{22}Na, ^{137}Cs$	AmBe	-	n and γ
745	$^{22}Na, ^{137}Cs$	AmBe	-	n and γ
1100	$^{22}Na, ^{137}Cs$	AmBe	-	n and γ
1390	$^{22}Na, ^{137}Cs$	AmBe	-	n and γ
1660	$^{22}Na, ^{137}Cs$	AmBe	-	n and γ
531	$^{22}Na, ^{137}Cs, ^{57}Co$	-	-	γ
736	$^{22}Na, ^{137}Cs, ^{57}Co$	-	1.5	γ
948	$^{22}Na, ^{137}Cs, ^{57}Co$	-	1.125	γ
1100	$^{22}Na, ^{137}Cs, ^{57}Co$	-	0.875	γ
1390	$^{22}Na, ^{137}Cs, ^{57}Co$	-	0.625	γ
1650	$^{22}Na, ^{137}Cs, ^{57}Co$	-	0.125	γ

7.4.1 Pulse Detection

For each event rate, the collected data have both active and quiet periods. Active zones are those where the pulses (n or γ) are embedded. The inactive zones are the parts without any events and are mostly noise. Obviously, low event rates have longer inactive zones than in higher event rates. The pulse detector moves through the data sample-by-sample. An event is detected when the sample amplitude exceeds a predetermined magnitude (by 30 units in this work) from the signal baseline. The event is then defined from 150 ns before the pulse peak until 650 ns after the peak. After the pulse is detected, it is smoothed using a digital smoothing polynomial filter (the Savitzky-Golay smoothing filter [163]).

7.4.2 Decay Time Determination

A neutron pulse has a significantly longer decay time than that of a gamma-ray pulse within this leading component of the total light pulse. In this part of the algorithm, that knowledge is exploited. After a pulse is detected using the method described above, its decay time is calculated. In this case, the time taken for the pulse to decline from its peak to 64% of its amplitude is defined as the decay time. The reason for choosing 64% is empirical. The algorithm is explained in details in section 2.1.6.

7.4.3 Energy Spectral Density

The energy spectral density (ESD) for a signal is a measure of its energy distribution as a function of frequency. It is a useful concept that allows us to determine the bandwidth associated with each neutron or gamma ray pulse. (see section 2.1.6 for details)

Pulse shape discrimination (PSD) is the technique used to distinguish between neutron and gamma-ray pulses using some of their significant properties. In this case, we use the decay time and the ESD to construct PSD plots with the former as the ordinate and the latter as the abscissa.

7.5 Results

PSD plots are shown for those data sets acquired with the AmBe source present at low event rate: 13 kHz (Figure 7.1 (a)), and high event rate: 745 kHz (Figure 7.1 (b)). In these plots, the events with decay time greater than or equal to 160 ns along with normalized ESD greater than or equal to 0.96 are marked as neutrons, that is, the red and blue dots represent neutrons and gamma rays, respectively. It can be

seen that the neutrons are quite separable from the gamma rays in the PSD plot even at high event rates.

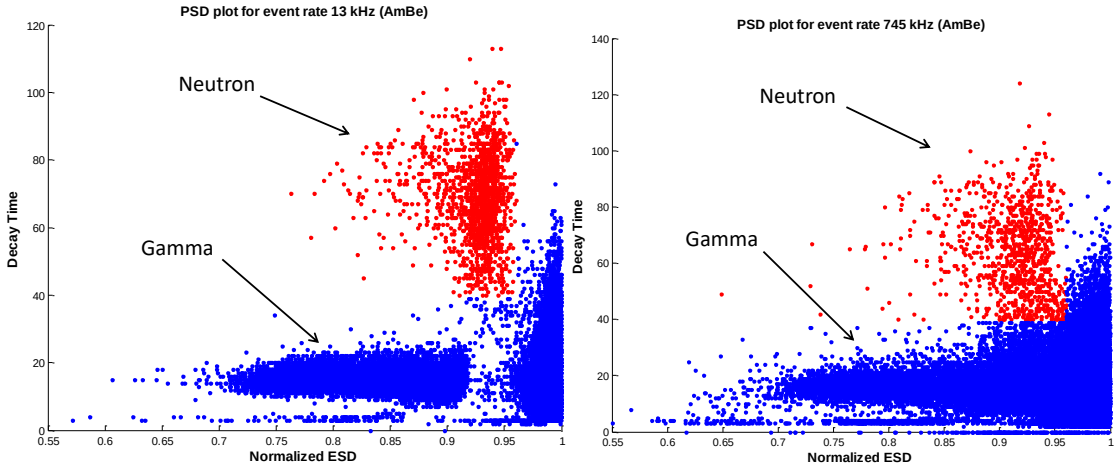


Figure 7.1: PSD plot in presence of AmBe source at (a) 13 kHz and (b) 745 kHz

Next we repeat the same analysis, but this time for the data sets collected without a neutron source. In the case of only gamma sources, PSD plots are shown for a low event rate of 531 kHz (Figure 7.2 (a)) and a high event rate of 1650 kHz (Figure 7.2 (b)).

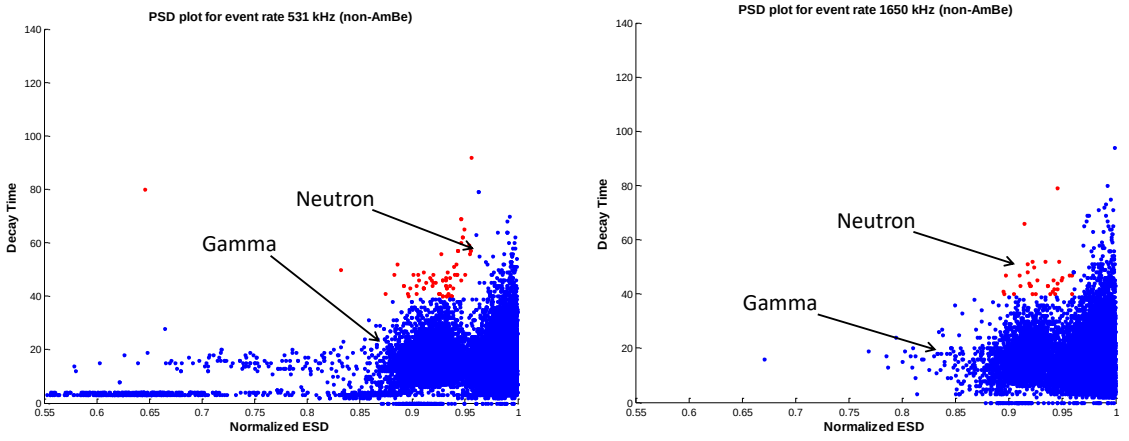


Figure 7.2: PSD plot in absence of AmBe source at (a) 531 kHz and (b) 1650 kHz

The plots show a few neutrons being detected, which can be said to be false detections, but compared to the number of events they are very few and insignificant ranging between 10 to 25 out of approximately 60,000 events.

A standard figure of merit (FOM) (see section 2.4.4) is used to determine the efficacy of the neutron discrimination in the PSD plot. The FOM is calculated for different event rates and it ranges between 0.8 to 1.35.

7.6 Neutron-Gamma Discrimination using Normalized Cross Correlation

To execute this approach, we have utilized an analytical model to represent typical gamma-ray and neutron pulses. One focus of this work is on the different cases of pileup events, showing how the PSD is accomplished for different pileup delays.

7.6.1 Reference Pulse Modeling

Numerical fits were performed on one particular neutron and one specific gamma ray pulse taken from the measured waveforms. These representative pulses were selected from the 13 kHz event rate data set such that pulse pileup was not present. The pulse fitting was carried out in Matlab using a pulse model expanded from that of [101] to include t_0 which is a time reference for the start of the signal,

$$p(t) = \sum A_i \exp[\frac{-(t-t_0)}{\tau_i}] - \exp[\frac{t-t_0}{\tau_0}] \tag{7.1}$$

where each decay time constant τ_i has an associated amplitude A_i, and the last term accounts for the initial pulse rise with τ_0. Recognizing that other investigators have observed differing numbers of decay times in the pulses [19], the number of terms i was varied from 2 to 4. The free fit resulted in 3 terms for the gamma ray and 2 terms for the neutron, which is consistent with the CVL mechanism for gamma rays

needing 1 more term than that of neutrons. Table 7.2 lists the fitted parameters. Figure 7.3 shows a good match between the experimentally obtained gamma-ray and neutron pulses (templates) and the model fits.

Table 7.2: Parameters obtained from fitting representative pulses

Parameters	Gamma Ray		Neutron	
t_0(ns), τ_0(ns)	0.6	18.4	2.2	40.3
A_1, τ_1(ns)	128	2136	165	1538
A_2, τ_2(ns)	187	447	182	257
A_1, τ_1(ns)	981	43.3	-	-

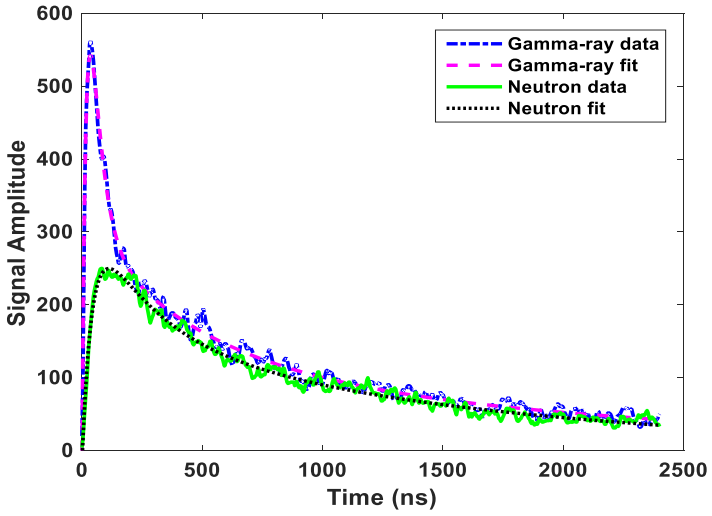


Figure 7.3: Experimentally obtained gamma-ray and neutron template pulses and their reference models.

7.6.2 Normalized Cross Correlation and PCA

In signal processing, the cross correlation is a measure of similarity between two series, expressed as a function of the time lag relative to one another. We systematically analyzed different cases of cross correlation between the reference gamma-ray and neutron pulses ($R_{\gamma\gamma}, R_{\gamma n}, R_{n\gamma}, R_{nn}$) (see section 2.1.6). There is a significant difference in the shapes of the NCC curves for neutron and gamma-ray pulses. Upon examining and comparing each curve , it can be seen that the features that are most discriminative are the magnitude and slope. The NCC curves along with their gradients are used to discriminate between neutron and gamma rays. We have used the PCA to find the points with larger variance. In this approach we have used a standard version and a modified version of PCA. In the modified version of PCA, we first extract 20 different points that have maximum deviation from normality using the KS test (see section 2.2.2). Both the algorithms were first trained using a simulated dataset of pileups, after which they are tested with the experimental data.

7.6.3 Pileup Analysis

We created a series of pileups of neutron and gamma-ray pulses with different A_i values. Pileups are created with 150 unique neutron and gamma-ray pulses, by randomizing the amplitudes using a Gaussian distribution having a standard deviation of 20. Classification is done for different values of lag (20 to 500 ns) between respective pulses.

Both the standard and the modified PCA algorithms are used for the classification. These two algorithms can be further categorized based on the reference pulse for NCC. They are denoted as follows,

- PCA_{gam}: Standard PCA technique with gamma-ray pulse as reference for NCC,

- PCA_{neu}: Standard PCA technique with reference neutron pulse for NCC,
- $PCA-KS_{gam}$: modified PCA technique with gamma-ray pulse as reference for NCC,
- $PCA-KS_{neu}$: modified PCA technique with reference neutron pulse for NCC.

We first created two instances of pileups: (a) gamma followed by neutron and (b) neutron followed by gamma. Figures 7.4 (a) and 7.4 (b) show the PSD with $PCA-KS_{neu}$ and $PCA-KS_{gam}$, respectively, for the two instances of pileups with a lag of 200 ns. In both cases, the neutrons and gamma rays are well separated.

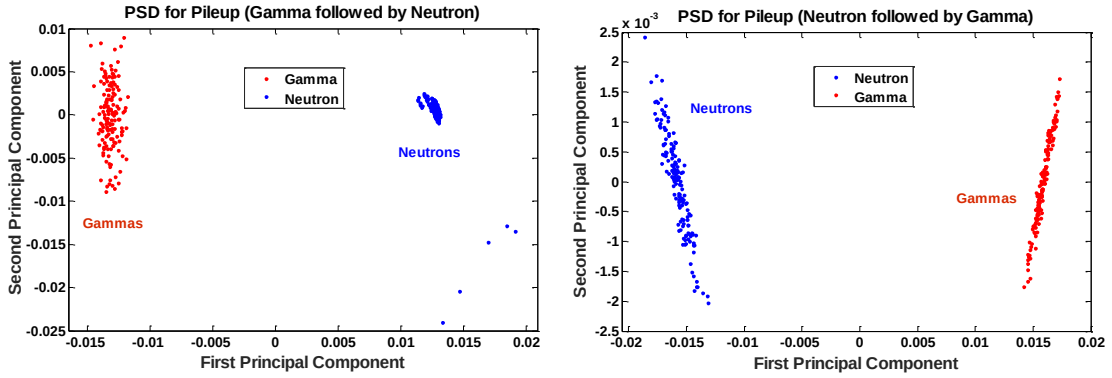


Figure 7.4: PSD for pileup simulation (gamma ray followed by neutron) for lag = 200 ns, using (a) $PCA-KS_{neu}$ and (b) $PCA-KS_{gam}$

The four techniques are compared by calculating the accuracy of detecting neutrons for different pileup delays (20 to 500 ns). It indicated that the PCA_{gam} and $PCA-KS_{gam}$ methods provide exact classification for 100 ns delay and above, for pileups where a gamma pulse is followed by a neutron pulse; whereas the PCA_{neu} and $PCA-KS_{neu}$ require at least a 200 ns delay for 100% accuracy. In case of pileups where a neutron is followed by a gamma ray, all the techniques show similar results.

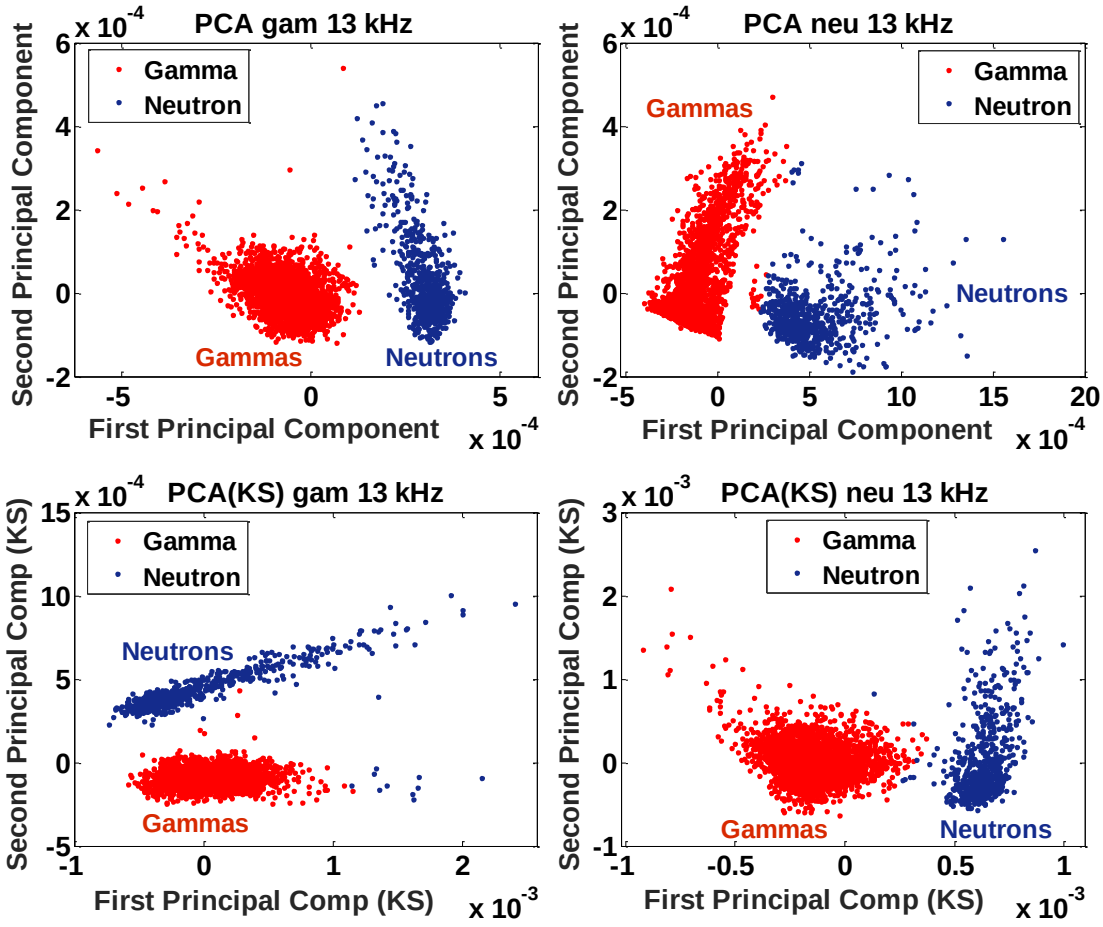


Figure 7.5: PSD at 13 kHz event rate, (a) PCA_{gam}, (b) PCA_{neu}, (c) $PCA-KS_{gam}$ and (d) $PCA-KS_{neu}$

7.6.4 Experimental Data Analysis

The combined PCA-NCC algorithm is evaluated with the experimental data. The PSD is applied to the 9 event rates using the four different PCA techniques presented in the previous section. Using both the reference gamma ray and neutron pulses, the NCCs of the chain of measured pulses are calculated first, after which the NCC peaks are detected. The gradient is then computed for each NCC curve (lag: –100 to +100 ns). Then, the standard and modified PCA techniques are utilized for feature

extraction from each NCC gradient. In both cases, the first and the second principal components are the metrics used for the PSD plot. We implemented the k-means classifier (see section 2.3.1 (I)) to perform the binary $n-\gamma$ discrimination in the PSD plot.

We present the PSD plots for event rates of 13 kHz and 1660 kHz in Figures 7.5 and 7.6 respectively. It can be seen in Figure 7.5 that for the low event rate (13 kHz), the neutrons have a good separation from the gamma rays. As the event rate increases, the neutron and gamma clusters tend to move closer to one another. Considering the fact that there are fewer neutron pulses in the data, it can still be observed that both pulse types have clearly defined regions. To assess the efficiency of the PSD and the effectiveness of the four PCA approaches, the FOM is calculated for each technique and all the event rates. The FOM is fairly constant as the event rate increases, with values ranging from 0.85 to 1.8. Overall, the modified PCA (KS) gamma and neutron exhibit the best FOM of the four techniques, especially at higher event rates.

7.7 Discussion

This chapter presented a study to discriminate between neutron and gamma ray pulses collected from using a CLYC scintillator. It is a challenging task in the field of nuclear radiation study, especially at higher sampling rate, due to pileups. We provided two digital signal processing technique for distinguishing neutrons and gamma rays using pulse decay time and normalized cross correlation. From the first approach, we identified the pulse decay time and the energy spectral density as the primary features, while in the second approach, the NCC gradient vector produced unique features to discriminate neutron and gamma pulses. This first PSD method is a combination of both time-domain and frequency-domain analyses. This algo-

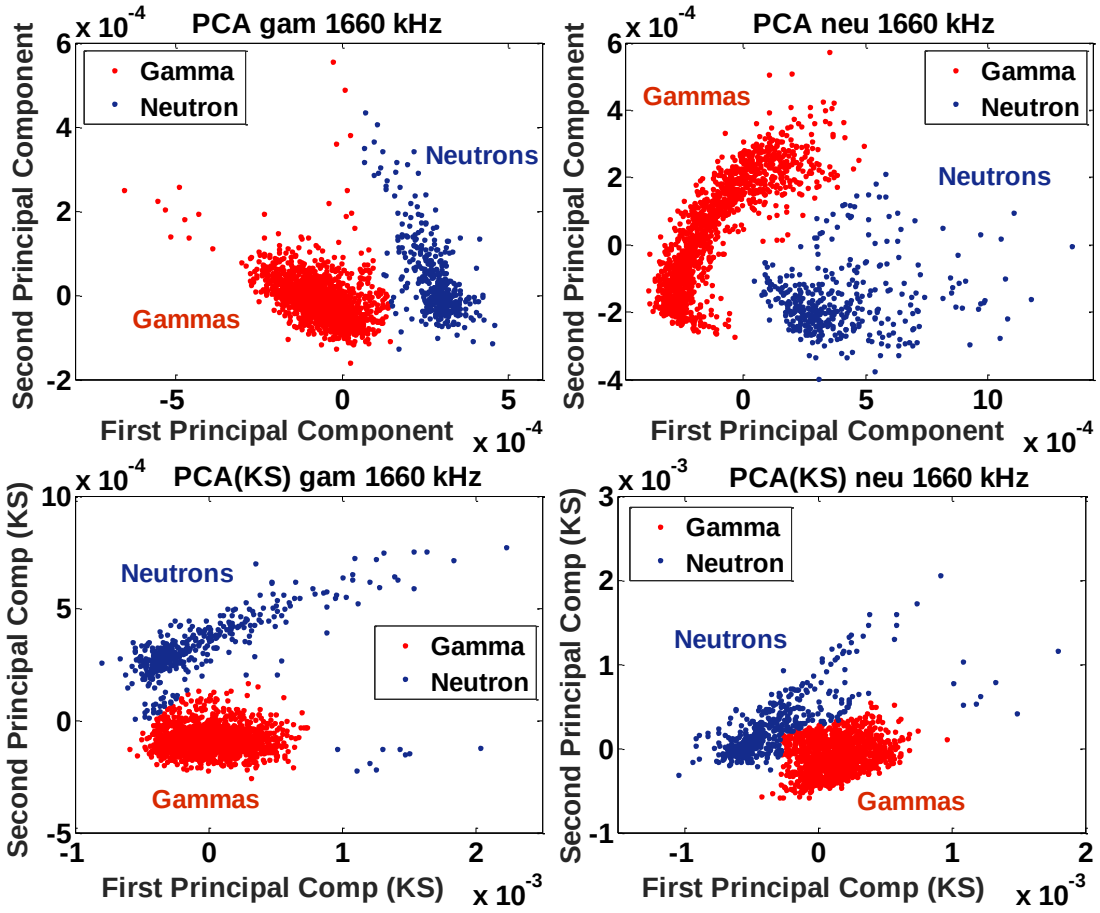


Figure 7.6: PSD at 1660 kHz event rate, (a) PCA_{gam}, (b) PCA_{neu}, (c) $PCA-KS_{gam}$ and (d) $PCA-KS_{neu}$

rithm development can be implemented within a neutron detector coupled to a field-programmable gate array (FPGA). The algorithm uses basic level calculations for decay time and ESD calculation. Overall, it is an algorithm with some fundamental logic, thus making it computationally cheap. In the second PSD method, four novel techniques for $n-\gamma$ pulse shape discrimination are introduced and their performance is evaluated using both simulated and measured pulses. From the real data analysis, it is seen that at high event rates, where pileups constitute the majority of the data, the PSD plot clusters reveal good separation between the gammas and neutrons.

Chapter 8

OTHER STUDIES

In this chapter, we have present 3 more studies, two of which were based on clinical trials and one based on a simulated wireless communication study. We have described these studies briefly in the following sections.

8.1 Identifying Behavioral Tasks using Local Field Potential Signals from Patients with Parkinson's Disease

In this section, we present unique clinical study and a recent area of investigation in improving the effects of Parkinson's disease (PD). The Deep Brain Stimulation has been used to treat intractable pain associated with PD for several decades. It's response to physiologic signals inside or outside the brain is considered the next frontier in brain stimulation research. This study was collaborated with the Colorado Neurological Institute, Denver, where data was collected from 13 patients with PD. We propose adaptive learning methods for identifying different behavioral tasks of these patients using local field potential (LFP) signals collected during DBS implantation surgeries. Using time-frequency signal processing methods, features are first extracted and then clustered in the feature space using two different methods. The first method requires training and uses a hybrid model that combines support vector machines and hidden Markov models. The second method does not require any *a priori* information and uses Dirichlet process Gaussian mixture models. Using the DBS acquired signals, we demonstrate the performance of both methods in clustering different behavioral tasks of PD patients and discuss the advantages of each method under different conditions.

8.1.1 Background

Parkinson's disease (PD) is a degenerative disorder of the central nervous system. It is diagnosed based on clinical history and motor signs of tremor, rigidity, and bradykinesia [81]. PD is caused by the death of dopamine-generating cells in a region of the midbrain, substantia nigra. The most perceptible symptoms of the disease are movement-related; these include shaking, rigidity, and slowness of movement.

In the US, a large number of people suffer from PD and new PD patients are diagnosed each year [84]. When drug therapy is no longer effective for PD patients, Deep Brain Stimulation (DBS) is applied [81, 82]. Current DBS therapy involves an open-loop system that was proven effective for treatment of essential tremor and PD. However, open-loop DBS has various shortcomings [82]. In particular, DBS may cause cognitive, speech, and balance side effects. By customizing DBS therapy to a patient's task, these stimulation side effects may arise only when they are non-detrimental to the patient's current goals. The design and advancement of a closed-loop implantable pulse generator (IPG) to sense and respond physiologic signals inside or outside the brain is considered the next frontier in brain stimulation research. As a consequence, advancements in classifying the behavior of PD patients using LFP signals are important in the development of a next generation closed-loop DBS system. These LFP recordings are often performed with either invasive microelectrodes or DBS leads and reflect oscillatory activity within nuclei of the basal ganglia. LFP differs from normal EEG or electrocorticography (ECOG) and ranges between 5–1,000 μV with frequency less than 200 Hz (see Figure 8.1 (a)).

In this study, we first extract behavioral task informative time-frequency based features from LFP signals, and then we develop adaptive learning based methods to identify the different tasks. The LFP signals used to assess the performance of

our proposed methods were obtained from a study [81, 112] involving twelve patients undergoing DBS implantation for treatment of idiopathic PD. The signals were simultaneous LFP signals recorded during behavioral tasks (see Figure 8.1 (b)). The tasks described four types of behaviors: simple motor task, language, language with motor task, and language without motor task.

For feature extraction, we apply the matching pursuit decomposition (MPD) time-frequency approach [26, 116] to the LFP signals, resulting in four-dimensional feature parameter vectors. Using the MPD, the signal is decomposed into a linear expansion of Gaussian basis functions that are selected from a redundant basic dictionary. The dictionary signals are time-shifted, frequency-shifted and scaled versions of an elementary Gaussian atom; these basis functions are chosen in order to best match the signal structure. The features obtained from the MPD are the time-shift, frequency-shift, scale change and weight coefficient of each selected Gaussian in the signal expansion. The MPD algorithm is described in details in section 2.1.5.

We propose two different adaptive learning methods for task identification. The first method quantizes the feature vectors using K-means clustering and applies the support vector machine (SVM) classifier [35], and the discrete hidden Markov model (DHMM) classifier [153]. The SVM and DHMMs are combined using a top-level and then a hybrid approach. The second method uses Dirichlet process Gaussian mixture models (DP-GMMs) [61] that can be used to represent complex statistical data distributions with multiple modes without training. These learning methods are described in details in section 2.3.1 (V) and 2.3.1 (VI), respectively.

8.1.2 Data Collection

Twelve subjects undergoing DBS as standard of care for treatment of idiopathic PD were enrolled in this study. All subjects provided informed consent for participa-

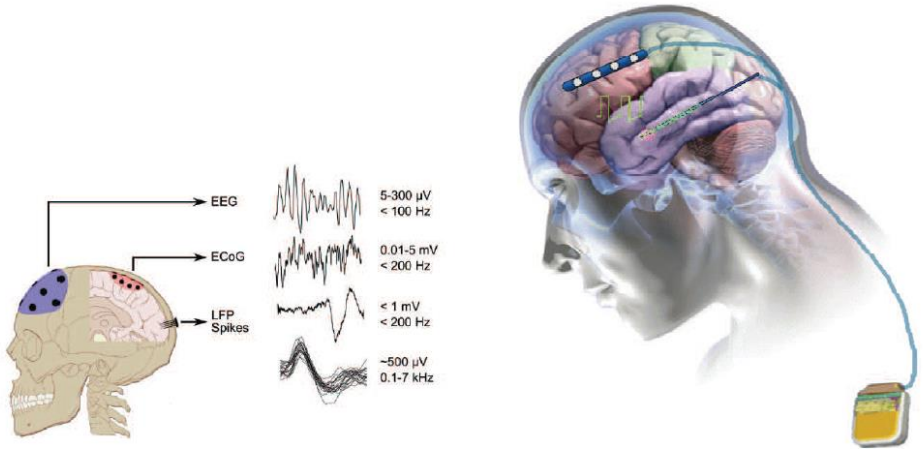


Figure 8.1: (a) Different brain signals: EEG and ECoG (macroscopic) and LFP and action potentials or spikes (microscopic). (b) Electrodes to record LFP signals from the subthalamic nucleus.

tion in this research study, in a manner approved by the internal review board of the University of Washington. Eleven independent recordings/lateralities were measured from the seven participants. Two subjects underwent sequential recordings from each side, another two had simultaneous bilateral recording, and another two subjects were recorded with a second electrode design and amplifier system in the same hemisphere. There were seven left, two right, and two simultaneous bilateral recordings (nine left and four right), for a total of 13 recordings for analysis.

8.1.3 Problem Formulation and Methods

Problem Formulation of Behavior Tasks

Let the LFP signal set from the four behavior tasks of the PD patients from the ith experiment, $i = 1, \ldots, I$, be given by $\mathcal{S} = \{\mathbf{s}_i^1\ \mathbf{s}_i^2\ \mathbf{s}_i^3\ \mathbf{s}_i^4\}$. Here, $\mathbf{s}_i^m = [s_i^m[1]\ \ldots\ s_i^m[K]]^T$ corresponds to the signal vector for the mth task, $m = 1, \ldots, 4$, $s_i^m[k]$ is the discretized LFP signal at time k, $k = 1, \ldots, K$, and T denotes vector transpose. The $m = 1$ task

corresponds to the simple motor task, $m=2$ corresponds to the language task, $m=3$ corresponds to the language with motor task, and $m=4$ corresponds to the language without motor task.

Feature Extraction

We used the Matching Pursuit Decomposition (MPD) time-frequency method (see 2.1.5) to extract observation features as it provided distinct patterns for the different behavior tasks. The feature set from the ith experiment is given by $\mathcal{F}=\{\mathbf{F}_i^1\ \mathbf{F}_i^2\ \mathbf{F}_i^3\ \mathbf{F}_i^4\}$. Here, $\mathbf{F}_i^m=[F_{i,1}^m\ \ \ldots\ \ F_{i,P}^m]^T$ denotes the feature matrix extracted from the LFP signal $\mathbf{s}_i^m$. The pth feature vector, $p=1,\ldots,P$, is given by $F_{i,p}^m=[\alpha_{i,p}^m\ \tau_{i,p}^m\ \nu_{i,p}^m\ \sigma_{i,p}^m]^T$, corresponding to the MPD amplitude, time-shift, frequency-shift and scale change parameters. Using the extracted features, the mth DHMM (or mth DP-GMM) $\mathbf{\Lambda}^m$, $m=1,\ldots,4$ is trained to model the mth task. The MPD feature vectors form 64 nodes; 30 consecutive feature vectors represent an LFP signal attained during the mth behavior task.

Adaptive Learning Clustering Algorithms

In order to classify the behavior tasks, several methods have been employed. First, we trained multiple DHMM classifiers for each of the four cognitive tasks separately, and then a top-level SVM fused all the low-level DHMMs. The performance of the task clustering and identification showed an improvement of 6% when compared to our approach in [91]. Second, we trained a hybrid SVM-DHMM model, which is an implementation of structural SVMs for sequence tagging based on integrating SVMs with HMMs (see 2.3.1(V)). The hybrid approach is more advantageous than the top-level approach when there are dependencies between neighboring labels and there exist overlapping features between clusters. Finally, the extracted feature vectors were

directly provided to the DP-GMM for unsupervised clustering (see 2.3.1(VI)). The approach allows for an unlimited number of mixture components, and the appropriate number of clusters does not need to be known *a priori* and is learned adaptively from data. The Gibbs sampler Markov chain Monte Carlo (MCMC) technique [73] was used to iteratively estimate the DP-GMM parameters [88].

Task m	3	4
3	**0.93**	0.07
4	0.09	**0.91**

(a) language with motor $m=3$ vs language without motor $m=4$

Task m	1	2
1	**0.94**	0.06
2	0.10	**0.90**

(b) simple motor $m=1$ vs language $m=2$

Task m	1	4
1	**0.94**	0.06
4	0.09	**0.91**

(c) simple motor $m=1$ vs language without motor $m=4$

Task m	1	3
1	**0.89**	0.11
3	0.10	**0.90**

(d) simple motor $m=1$ vs language with motor $m=3$

Table 8.1: DHMMs by top-level SVM confusion matrix.

8.1.4 Results

We use the LFP signals collected from twelve PD patients to demonstrate our clustering methods. The signal segments associated with different behavioral tasks were labeled by physicians during data collection. The behavioral tasks are: simple motor task ($m=1$), language with motor task ($m=3$), and language without motor task ($m=4$). The language tasks ($m=2$) combines tasks 3 and 4. The signal sampling rate is 4 kHz, and for different behavioral tasks, the number of data segments varied from 80 to 109. We used k-fold cross validation technique to separate the training

and test data and four hidden states for the DHMMs. Linear SVM provided the best performance for both the top-level and hybrid approaches. As a preprocessing step, we quantified the feature vectors and mapped them into new feature values using the K-means classifier. The feature vectors are clustered into 64 cluster nodes. Each consecutive 30 feature vectors represent deep brain signal attained during a PD patient's behavior. Tables 8.1, 8.2, and 8.3 provide the confusion matrices that summarize the identification results.

Task m	3	4
3	**0.92**	0.08
4	0.08	**0.92**

(a) language with motor $m=3$, language without motor $m=4$

Task m	1	2
1	**0.93**	0.07
2	0.08	**0.92**

(b) simple motor $m=1$ vs language $m=2$

Task m	1	4
1	**0.94**	0.06
4	0.08	**0.92**

(c) simple motor $m=1$ vs language without motor $m=4$

Task m	1	3
1	**0.91**	0.09
3	0.10	**0.90**

(d) simple motor $m=1$ vs language with motor $m=3$

Table 8.2: Hybrid HMM-SVM confusion matrix.

For DP-GMM clustering, first MPD features were extracted from signals from each behavioral class. The MPD was run for 15 iterations for each signal set which gave a 300×4 feature matrix. The best clustering results were obtained using the amplitude and time-shift MPD parameters as the feature set. The parameters used in the DP-GMM were set as follows; innovation parameter $\alpha = 0.6$, 0.01 truncation error, 2,000 burn-in and 1,000 sampling iterations for the Gibbs sampler. The confusion matrices are provided in Table 8.3.

Task m	3	4
3	**0.96**	0.04
4	0.28	**0.72**

(a) language with motor $m=3$, language without motor $m=4$

Task m	1	3
1	**0.92**	0.08
3	0.22	**0.78**

(b) simple motor $m=1$, language with motor $m=3$

Task m	1	4
1	**0.84**	0.16
4	0.10	**0.9**

(c) simple motor $m=1$, language without motor $m=4$

Task m	1	3	4
1	**0.78**	0.11	0.11
3	0.035	**0.90**	0.035
4	0.095	0.095	**0.90**

(d) simple motor $m=1$, language with motor $m=3$ and without motor $m=4$

Table 8.3: DP-GMM confusion matrix.

8.1.5 Discussion

Thus, we provided a unique study to identify behavioral task performed by patients with PD and using DBS. The LFP recordings from DBS is used to find the markers of the behavioral tasks. We provided a description of the data collection experiments and a complete mathematical formulation for the overall proposed behavioral task identification system. We applied the MPD feature extraction method, and employed two proposed adaptive learning methods in order to classify different behavioral tasks. Our simulations using real data demonstrate that the classification rate has improved significantly when compared to previous efforts. Moreover, the advantages of each approach under different conditions is discussed. This study furthers the investigation on use of DBS and a way forward to construct a feedback DBS system to ameliorate the effects of PD.

8.2 EEG Processing to Predict Postoperative Delirium

This section introduces another unique clinical study, where electroencephalogram (EEG) recordings were collected from patients undergoing open heart surgery, to predict post operative delirium. Cardiac surgeries involving deep hypothermia circulatory arrest present a risk of cognitive impairment. This study attempts to uncover EEG biomarkers predictive of postoperative delirium, which is associated with long term health complications. We predict postoperative delirium diagnoses by examining changes in ensemble neural activity during surgeries through spatiotemporal eigenspectra extracted from patient EEG data. Artifact detection and feature normalization schemes are developed to facilitate this. At most 14 of 16 cases were correctly predicted with a p-value of 0.0015. An area under the receiver operating characteristics (ROC) curve's convex hull of 0.9091 was achieved. This work was a collaboration between Mayo Clinic and the engineering team at ASU, lead by Owen Ma from BLISS Lab.

8.2.1 Background

Many cardiac surgeries, including coronary artery bypass surgery and operations involving the ascending, arch, or proximal descending aorta require induction of circulatory arrest to facilitate repairs. Doing so, however, neccessitates neural protection. Reducing body temperature down to (deep) hypothermic levels lowers the brain's metabolic demand and curbs neuronal death [208]. Deep hypothermia circulatory arrest (DHCA) is typically effective [71], but a risk of cognitive impairment remains [60, 77, 155]. Antegrade cerebral perfusion (ACP) and retrograde cerebral perfusion (RCP) have been utilized to supplement DHCA, but mixed opinions have been raised regarding their efficacy [78, 143]. Detecting delirium during recovery in the intensive

care unit is associated with long term health complications [18]. It would be beneficial to monitor biological signals intraoperatively for cues of cognitive impairment in real-time, allowing for interventions to improve the success rate of these surgeries. This study seeks to accomplish that task by analyzing electroencephalogram (EEG) signals.

Consistent EEG changes with anoxia were described in 1960 based on EEG findings observed during intraoperative cardioplegia, and descriptions of EEG findings during hypothermia emerged following the inclusion of hypothermia and cardiopulmonary bypass (CPB) in cardiac surgery [147]. Warren J. Levy investigated the relationship between temperature changes during DHCA and both spectrograms and estimated entropy of EEG recordings [107, 108]. He identified burst-suppression [4] activity, a possible indicator of hypoxia, within spectrograms. Most recently Westover et al. studied the burst-suppression phenomenon through spectrograms in further depth with additional statistical measures, arguing that this activity is a fundamental indicator of poor metabolic support [15]. Past EEG based DHCA research has mainly focused on signal characteristics relating to hypoxia (or anoxia), one of the primary causes of neuronal death.

The primary focus of this study is to formulate an EEG based feature predictive of postoperative delirium. Furthermore, the data collected in this study measures neural activity while the human body transitions into and out of conditions that would be lethal under non-surgical circumstances. It presents an opportunity to learn the nature of neurological functions and responses during such stress. This data captures the resiliency of the human body–despite halting all measurable neural and cardiovascular function, remarkably many patients recover without issue.

8.2.2 Surgical Procedure and Data Collection

The Institutional Review Board (IRB) approved this study, a collaboration between Arizona State University and the Mayo Clinic. Over 100 participants comprise the study population, but a subset consisting of 18 patients were examined. Two patients are excluded from this report because they did not undergo DHCA. These patients underwent surgery at a Mayo Clinic facility primarily to repair their aortic arch, their aortic root, or both. A general outline of an arch replacement surgical procedure has been documented by Sundt et al., Mayo Clinic surgeons [178]. The set of procedures pertaining to this study utilized RCP in conjunction with DHCA. Selective ACP was implemented additionally at the surgeon's discretion. Patients are cooled until they exhibit electrocerebral silence (ECS), a peripheral indicator that metabolic demand has been minimized [15, 89], and is defined here as no detectable cerebral activity while reading at a sensitivity of 2 uV/mm.

EEG data was recorded at a sampling frequency of 500 Hz using a Xltek EMU128FS from the induction of anesthesia until unassisted circulation is resumed. Electrodes were placed at the following locations: C3, C4, Cz, F3, F4, F7, T6-P8, FPZ, Fp1, Fp2, Fz, O1, O2, Oz, P3, P4, Pz, T3-T7, T4-T8, T5-P7, T6-P8, TP11, and TP12 [144]. Additionally, there were numerous reference nodes spaced over the head and an electrocardiogram channel. Core body temperature was recorded using an esophageal probe. Only occasional time stamps include a temperature reading, but the start of cooling and warming is well documented. Sufficient samples are provided such that a general sense of the temperature's rate of change can be determined. As patients recover in the intensive care unit, nurses engage them in delirium assessment protocol using the Richmond Agitation-Sedation Scale (RASS) and the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) [57]. The latter produces a binary

decision about whether a patient is delirious or not. This study attempts to determine a relationship between EEG features and the CAM-ICU score assigned to patients. Figure 8.2 shows a single channel EEG signal and its spectrogram collected from a patient during the surgical procedure.

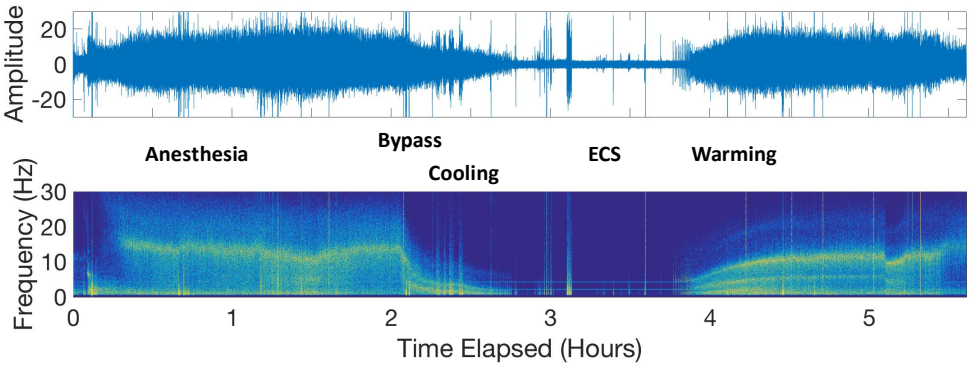


Figure 8.2: Single channel EEG signal and its spectrogram, showing all the stages of the surgical procedure

8.2.3 EEG Signal Processing and Decision Methodology

EEG Signal Processing

The data was first filtered before analysis took place. The data is first low-pass filtered at 40 Hz and downsampled to a new sampling rate of 100 Hz. The downsampled signals were high-pass filtered with the cut off frequency at 1 Hz to center the data and remove effects of "DC drift".

Spatiotemporal correlation is used to study ensemble interaction is through. EEG signals are not statistically stationary, so the data are analyzed in windows of reasonable lengths under which they can be approximated as a stationary process. Any changes in latent probabilistic models can still be observed through windowing. The Spatiotemporal eigenspectra is used to capture this, which show fluctuations in cor-

relation within the ensemble over time [201].

Decision Methodology and Performance Metrics

The general approach in predicting the delirium assessment is to investigate how EEG signal dynamics differ before and after DHCA. Predictions presented in this study are made during the warming procedure and prior to allowing autonomous blood circulation. It is suspected that patients will test negative for delirium, if characteristics before DHCA resemble those afterward. In terms of the eigenspectra, we hypothesize that patients will test positive for delirium, if the eigenvalue distributions deviate far from the reference and remain as such throughout the warming procedure.

$$\mu[l] = \frac{1}{M} \sum_{m=1}^{M} s_m[l] \tag{8.1}$$

An example metric, $\mu[l]$, that captures this characteristic–likely suboptimally- is the mean z-score for each window. This is expressed in 8.1, where $s_m[l]$ is the z-score of the m^{th} eigenvalue corresponding to the l^{th} window. Predictions are made based on two thresholds, a minimum required sustain period, T_s and an amplitude threshold, T_z. To be predicted delirium-negative, $\mu[l]$ during warming must exceed T_z for at least T_s sequential windows. Whenever an artifact occurs, the decision if $\mu[l]$ has crossed the amplitude threshold is assigned the last known valid decision. The choice of this particular time based requirement may not be precise, but the intention of including it is to ensure a steady state status has been established. Each patient is labeled either 1 for a positive delirium detection or 0 otherwise. The true labels are the CAM-ICU labels provided. The entire process chain is shown in Figure 8.3.

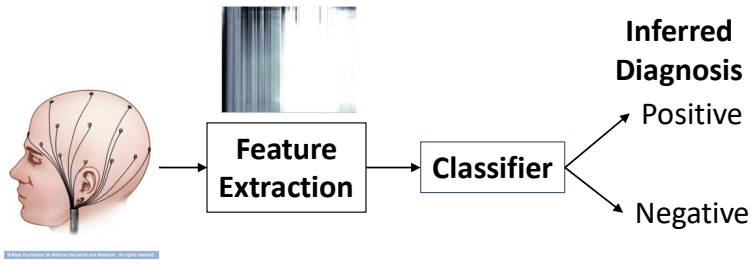


Figure 8.3: Process chain to predict postoperative delirium

8.2.4 Results

A receiver operating characteristics (ROC) curve (see section 2.4.3) is produced by setting a particular T_s that maximizes the area under the curve (AUC) while varying T_z. We evaluate AUCs achieved for a range of T_s and observe in Fig. 4.1 that the optimal set of T_s is between 26.25 and 38.33 minutes. We select one of these optimal choices of T_s and produce its corresponding ROC curve. An additional AUC was computed from the convex hull of the points forming the same ROC curve. The highest achieved AUC was 0.8364, and the AUC for the ROC convex hull was 0.9091.

A signi

cance test was also administered to determine if the optimal classi

er's performance performs better than randomly guessing. For each patient, 1000 rounds of prediction labels are randomly generated. The positive predictive value (PPV), negative predictive value (NPV), and accuracy were computed for each round. Means and standard deviations of those three metrics were estimated. P-values were generated for each metric based on the simulated statistical parameters. Prediction accuracy was statistically signi

cant with a p-value of 0.0015. Both PPV and NPV were also statistically signi

cant with a p-value for the former was 0.0015 and 0.0067 for the latter. The

confusion matrix that achieved the highest accuracy is shown in Table 8.4.

Table 8.4: Highest Accuracy Achieved Confusion Matrix

	Truth Positive	Truth Negative
Predicted Positive	5	2
Predicted Negative	0	9

8.2.5 Discussion

In this study, a procedure to predict postoperative delirium assessments based on intraoperative EEG data is developed. It is hypothesized that electrophysiological dynamics after DHCA should begin to resemble those beforehand in non-delirium cases. This is measured through normalized spatiotemporal eigenspectra to monitor the correlation between EEG signals. Patients were classified based on how far eigenvalue structure has deviated during warming from the average distribution prior to incision. If the eigenvalue distributions during warming remain dissimilar in terms of standard deviations throughout warming, then that patient is predicted to test positive for delirium. Using this rule, 13 of 16 patients were correctly classified without false negatives. From these results, it appears there is potential to detect potential complications as early as the warming phase, using the spatio-temporal eigen spectra.

8.3 Detection of Weak Signals in a SISO Wireless Communication Scenario

The last section of this chapter presents a simulation study to detect signals at low signal-to-noise ratio (SNR). In this work we consider the problem of quantifying the detection latency of weak signals—a signal whose energy is much lower than the noise floor. Detection of weak signals—signals whose energy is much smaller compared to the noise—arises in a variety of areas such as wireless communications, radar, and radio astronomy. We define detection latency as the total number of samples needed to correctly detect the weak signal. We characterize the latency by considering a single-input, single-output (SISO) wireless communications scenario where the receiver is tasked with detecting the transition from a channel probing waveform to a 500 kHz signaling tone. We consider constant modulus linear frequency modulated (LFM) chirp and the GPS L2C (M) pseudorandom noise waveforms for channel probing. Thus, this is a detection problem, where each sample of a gives sequence is used as a feature to make a decision. We also introduce the notion of *hypothesis tracking* built on the sequential detection framework to reduce detection latency at the receiver. We compare the performance between two hypothesis tracking methods.

8.3.1 Background

Methods for detecting weak signals most often rely on accumulating large number of observations of the signal of interest to achieve a suitable integrated SNR for detection. However, in time-sensitive applications such as unmanned aerial system (UAS) control, remote surgery, or vehicle to vehicle communications, these methods are not suitable. In these sorts of applications, sequential analysis often employed to detect the signal in the least number of samples as possible.

A broad overview of non-sequential detection methods employed for the weak

signal detection problem are discussed in [192, 207]. These methods include matched filter, lock-in amplifiers, wavelet transforms, and empirical mode decomposition, to name a few. The first sequential detection method, the sequential probability ratio test (SPRT), was introduced by Wald in his seminal work [189]. References [125] and [22] represent, possibly, the earliest theoretical studies of applying sequential analysis to the weak signal detection problem. In these works, the authors derive expressions for error probabilities for the tests, and compare against original approximations in Reference [189]. To address shortcomings within the SPRT framework—primarily, the absence of an upper bound on stopping time and model mismatch can often cause the number of samples for which a decision is made to be larger than the sample size—a class of truncated tests were proposed in References [7] and [5].

While sequential analysis has been applied to various problems, not much literature exists for characterizing the detection latency in a low instantaneous SNR regime ($<$ 0 dB). In this work we utilize the sequential analysis framework to quantify the effect of SNR on detection latency for some commonly used waveforms in the wireless communications and radar areas. The contribution of this work is as follows:

- Develop two algorithms that utilize sequential analysis to quantify the effect of SNR on latency within a wireless communications scenario
- Characterize the latency and probability of false alarm performance of the algorithms utilizing a constant modulus pseudorandom noise waveform and linear frequency modulated chirp waveform

8.3.2 Problem Formulation

We consider a single-input single-output (SISO) wireless communications scenario under flat fading channel conditions. Further, we assume the system operates in an

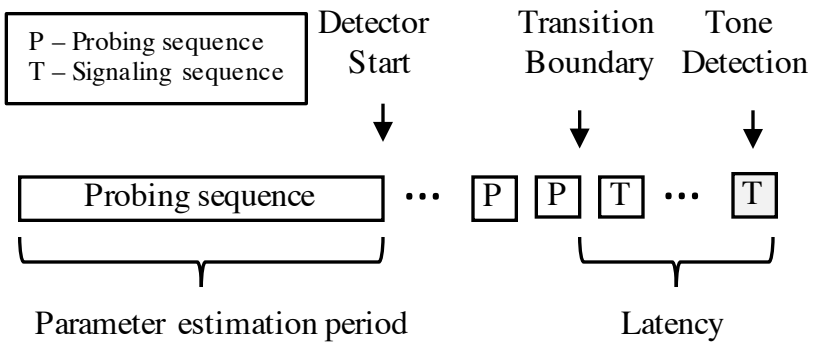


Figure 8.4: Illustration of the detection scenario considered in this work. We wish to characterize the latency of detecting a switch in transmit waveforms. We consider a flat fading channel and low SNR regime. We define latency as the minimum number of samples needed to make a correct decision once the transition boundary has occurred, excluding propagation delay.

instantaneous SNR regime of less than 0 dB. For some amount of time the transmitter emits a channel probing waveform. The receiver utilizes this time period for synchronization, and to estimate the channel parameters necessary to perform detection. The transmitter switches to a signaling waveform once triggered by some event. We define the point at which this switch occurs as the transition boundary.

Our goal is to characterize the latency associated with detecting various signaling waveforms at the receiver. We define latency as the minimum number of samples needed to make a correct decision, excluding propagation delay, once the transition boundary has occurred. The exact occurrence of the transition boundary is unknown. However, it is assumed the receiver has some knowledge of the approximate time frame for which the switch will occur. We also assume that the receiver has a priori knowledge of both waveforms. This scenario is illustrated in 8.4.

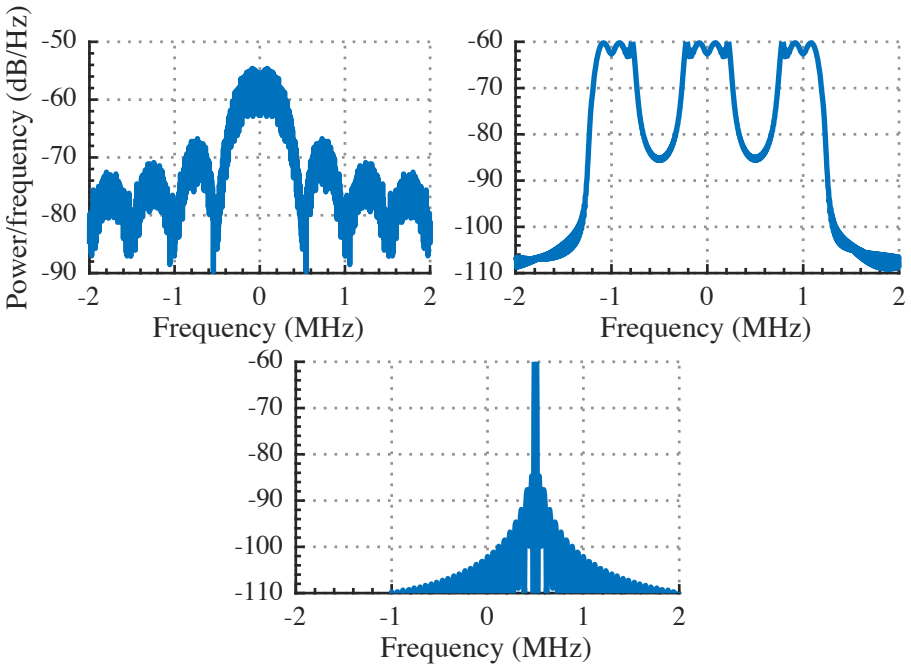


Figure 8.5: Spectrum plots of the channel probing and signaling waveforms considered in this work. (a) A GPS L2C (M) channel probing pseudorandom waveform. (b) A linear frequency modulated (LFM) chirp channel probing waveform with nulls around 500 kHz. (c) A 500 kHz tone was used as the signaling waveform.

8.3.3 Waveforms

Comb Chirp Waveform

The first channel probing waveform we consider is a linear frequency modulated (LFM) chirp of the form

$$\{\mathbf{s}_0\}_n = \exp\left\{\frac{j\,\pi\,B\,n^2}{T}\right\}, \tag{8.2}$$

where B is the relative chirp bandwidth and $T = \frac{N}{f_s}$ is the chirp duration. We construct a waveform using multiple LFM chirps concatenated in time. This allows nulls to be inserted at specified frequencies while keeping the waveform constant amplitude. We show the spectrum of the comb chirp waveform used in the simulations

in 8.5 (b).

GPS L2C (M) Waveform

The second channel probing waveform we consider is the Global Positioning System (GPS) L2C (M) waveform [87]. The waveform is generated in two stages. First, a 25 bits/second pseudorandom binary phase-shift (BPSK) data stream is encoded using a rate 1/2 convolutional encoder. Finally, the encoded bit sequence is multiplexed with a 10,230 chip code generator to produce the transmission waveform. We show the spectrum of the GPS L2C (M) waveform used in the simulations in 8.5 (a).

Signaling Waveform

We utiilze a complex tone for the signaling waveform

$$\{\mathbf{s}_1\}_n = \exp\left\{\frac{j\,2\,\pi\,f_0\,n}{f_s}\right\}, \tag{8.3}$$

where f_0 denotes the fundamental frequency, and f_s denotes the digital to analog converter sampling rate. The complex tone is utilized because it is resolvable in the least amount of samples. However, any arbitrary constant amplitude waveform can be utilized for signaling. We show the spectrum of the tone waveform used in the simulations in 8.5 (c). For channel probing we consider constant modulus waveforms to limit low-noise amplifier saturation at the receiver.

8.3.4 Detection Algorithms

A larger distance between the upper threshold value (γ_1) and lower threshold value (γ_0) produces large latency values, but decreases the probability of false alarm. In most cases, the latency operates as a function of this distance. To overcome this we track local changes of the test statistic to reduce the latency required in making a

positive detection. We refer to this process as *hypothesis tracking* and present two methods of doing this. The detection problem along with the two hypothesis tracking methods are described in details in section 2.3.3.

The two hypothesis tracking algorithms are,

- *Serial Hypothesis Tracking*: In this algorithm, the cumulative test statistic in 2.18 is calculated and compared against the thresholds for each received sample.
- *Parallel Hypothesis Tracking*: Here, we assume that each new sample is potentially the start of the signaling sequence and compare multiple cumulative test statistics at the same time.

8.3.5 Results

ROC Analysis

We generate the ROC curves for both algorithms by computing probability of detection (P_{D}) values when only the signaling waveform is present, and computing probability of false alarm (P_{FA}) values when only the channel probing waveform is present. In this analysis, we consider four SNR regimes: -20 dB, -10 dB, -3 dB, and 5 dB. The ROC curves are generated using 50 logarithmically spaced threshold points over the intervals $\gamma_1 = \{10^{-2}, \ldots, 10^{2}\}$ and $\gamma_0 = \{-10^{-2}, \ldots, -10^{2}\}$ using 1000 Monte Carlo simulations.

We show the ROC curves for both algorithms in 8.6. For both channel probing waveforms, serial tracking outperforms parallel tracking for SNR values of less than 0 dB. For SNR values of greater than 0 dB both algorithms achieve similar performance. We also observe that serial tracking performance was roughly the same for both channel probing waveforms. However, parallel tracking performed better when the GPS L2C (M) channel probing waveform was used.

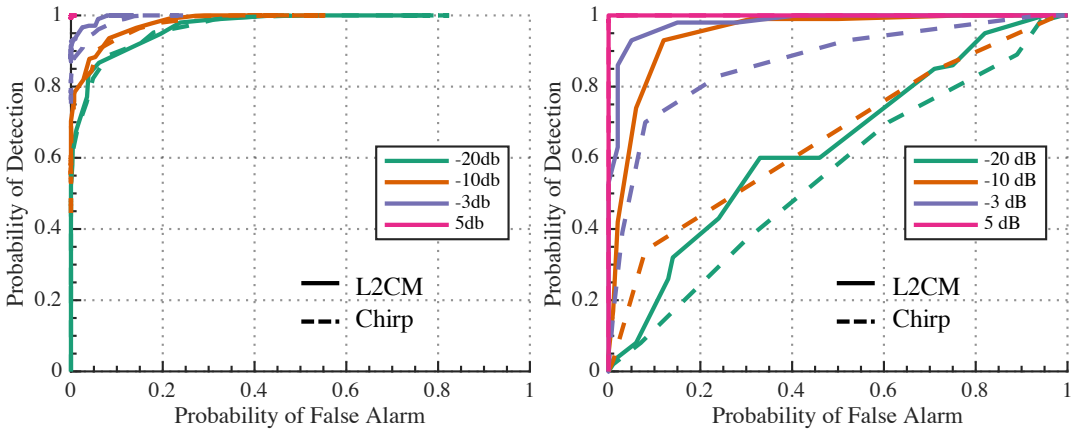


Figure 8.6: Receiver operating characteristics (ROC) for (a) Serial Hypothesis Tracking and (b) Parallel Hypothesis Tracking. For both channel probing waveforms, serial tracking outperforms parallel tracking for SNR values of less than 0 dB. For SNR values of greater than 0 dB both algorithms achieve similar performance. Serial tracking performance was roughly the same for both channel probing waveforms. However, parallel tracking performed better when the GPS L2C (M) channel probing waveform was used.

Latency Analysis

We characterize the latency of each receiver detection algorithm by considering two scenarios:

1. *Scenario 1*: detector observes only the 500 kHz signaling tone.

2. *Scenario 2*: detector observes 768 μs of channel probing sequence followed by 768 μs of 500 kHz signaling tone.

We identified the approximate optimal thresholds for each algorithm using the ROC curves shown in 8.6.

For serial tracking, a threshold setting of $\gamma_1 = 1.1$ and $\gamma_0 = -0.9$ was used across all SNR regimes. For parallel tracking, a threshold setting of $\gamma_1 = 10.5$ and $\gamma_0 = -9.5$ for -10 dB, -3 dB and 5 dB, whereas $\gamma_1 = 3.35$ and $\gamma_0 = -3.1$ were used for -20 dB SNR. We define latency as the minimum number of samples needed for the detector to make a correct decision once the transition boundary has occurred—by this definition the transition boundary corresponds to 0 μs latency. We computed the latency over 100 Monte Carlo trials and analyzed the cumulative distribution functions (CDFs) for each scenario.

We show plots of the CDFs for the measured latencies of the two detectors in 8.7. In each scenario, serial tracking outperforms parallel tracking for SNR values of less than 0 dB. This result makes sense because of inherent delays (varies depending on when the transition boundary occurs within the buffer) that occur as a result of the test statistic buffer used in parallel tracking. Similar latency values among each algorithm was observed for SNR values greater than 0 dB.

In each scenario, the measured latency remained approximately the same for each waveform across all SNR values. Further, no significant difference between latencies was observed in the algorithms between scenarios. This suggest that the algorithms are relatively robust to the location of the transition boundary.

8.3.6 Discussion

Thus, we presented a wireless communication simulation study to detect weak signals whose energy is much smaller compared to noise. We considered the problem of quantifying the latency for detection of weak signals. We considered a single-input, single-output (SISO) wireless communications scenario under flat fading channel conditions. We also introduced the notion of *hypothesis tracking* and presented two methods for doing this: Serial and Parallel Hypothesis Tracking. In the scenario, the

receiver was tasked with detecting a 500 kHz signaling tone as fast as possible. We designed two detectors based on these hypothesis tracking methods and compared their performance at SNRs of -20 dB, -10 dB, -3 dB, and 5 dB. Our results showed that serial hypothesis tracking outperformed parallel hypothesis tracking in terms of latency and with smaller probability of false alarm.

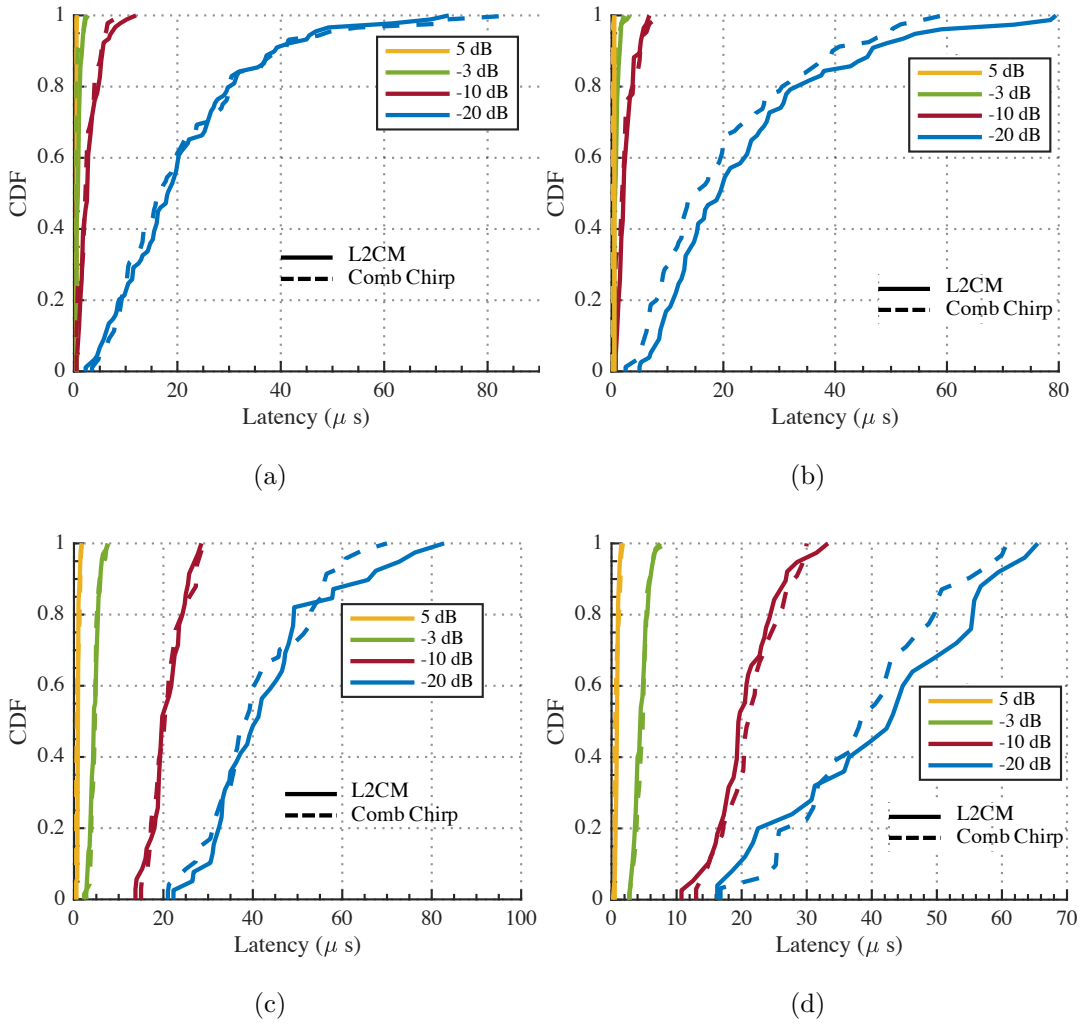


Figure 8.7: Cumulative distribution functions as a function of latency for the two scenarios considered in this work. In Scenario I the detector only observes the signaling waveform, while in Scenario II the detector observes 768 μs of the channel probing waveform followed by the signaling waveform. (a) The latency distributions of the Serial Tracking Method, Scenario 1 are consistent between the L2CM and comb chirp waveforms. (b) The comb chirp waveform performs marginally better than the L2CM waveform with serial tracking in Scenario 2. (c) The latency distributions of the (c) Parallel Tracking Method, Scenario 1 and (d) Algorithm 2, Scenario 2 are similar between the L2CM and comb chirp waveforms. In each scenario, serial tracking outperforms parallel tracking for SNR values of less than 0 dB. Further, in each scenario the measured latency remained approximately the same for each waveform across all SNR values.

Chapter 9

SUMMARY

In this dissertation, we have presented some unique studies from which we were able to make significant hypotheses about various real-world complex systems. All these studies involved processing real-world data to solve problems of classification, regression, detection or modeling. With the use of various SP and ML algorithms, we were able to investigate these problems and draw valuable inference from them. In each study, we have presented a set of problems related a particular field of interest, the processes used to tackle those problems and the results we achieved.

In the first study (Chapter 3) we proposed a method to estimate free-living physical activity from recorded wrist-worn accelerometer data. Our approach first classified and modeled laboratory based physical activities (ambulatory and non-ambulatory) performed by a diverse group of participants, then used these models to estimate physical activity intensity of 20 participants in free-living settings. High classification accuracy was achieved in both settings, suggesting that the models were accurate. This study has important implications for physical activity researchers. First, because these data were collected using raw waveform accelerometry on the wrist, these computations can be replicated across a large range of wearable sensors. Second, while ample detail are provided here for data scientists to replicate this approach, the methods can be implemented real-time and more user-friendly tools are currently being prepared to make this approach available for physical activity researchers.

In the second study (Chapter 4), we presented a study analyzing the relationship between the human gut-microbiome and bone health using data collected from 26 female adolescent participants. The study showed that there exists a high correla-

tion (> 0.8) between the gut-microbiome and bone density, which suggests that the gut-microbes can be strongly related to bone health. The study also identified the microbes that were highly correlated with different measures of bone density. Some of the microbes that were identified in this study are known to affect various biological processes which are responsible for bone regulation. Thus, this study provides valuable contribution to the gut-biome research community which is encouraging for further investigation.

The third study (Chapter 5) presents a unique social study, where we analyzed the relationship between communication and productivity in an organization setting. Speech recordings were collected from the employees, for a period of almost 3 years, which were analyzed in this study to extract communication features. The results suggested that there exists a moderate relationship between communication and productivity. We also inferred that productivity can be more dependent on the structure and topography of the communication network, rather than individual characteristics. In this study we provided unique tools to analyze communication from speech and use machine learning models to predict measures of productivity, which can be very useful for social science researchers.

In the forth study (Chapter 6), we presented a unique study to decode human intent using multi-modal signal analysis. We designed experiment to predict a cognitive task using a motion processing unit and a brain machine interface unit. We found markers of cognitive activity in the eigenspectrogram of the EEG. Thus, this study suggests that cognitive activities can be detected prior its execution using EEG readings. Chapter 7 presents a nuclear radiation study, where the primary objective is to discriminate neutron and gamma ray radiation pulses, detected using a CLYC scintillator. This is a challenging in this field, as it becomes very difficult to detect neutrons from gamma-ray pulses at high sampling frequency. In this study we

presented two algorithms to find the pulse shape discrimination at various sampling rates, achieving high accuracy. The developed algorithms can be implemented within a neutron detector coupled to a field-programmable gate array (FPGA).

In chapter 8, we presented three other studies, two of which were clinical studies and one was a wireless communication simulation study. In the first clinical study, where behavioral tasks in Parkinson's disease patients were classified using local field potential readings from brain collected from deep brain stimulation. The second study provided method to predict post-operative delirium from patients undergoing open heart surgery from their EEG readings. We also showed the unique EEG characteristics during deep hypothermia. The third study presents two algorithms to detect a weak signal in a high SNR wireless communication channel.

Thus, we investigated eight studies, addressing both fundamental and complex problems using SP and ML algorithms. Not only are the results important to the field of interest, it also provides an ensemble of tools that can be used by researchers in the field. To conclude, this dissertation contributes to the utilization of SP and ML in various interdepartmental studies exploring previously unanswerable questions in real-world applications and opens the prospect for future studies.

[1] Ainsworth, B. E., W. L. Haskell, S. D. Herrmann, N. Meckes, D. R. Bassett, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover and A. S. Leon, "2011 compendium of physical activities: A second update of codes and MET values", Medicine and Science in Sports and Exercise **43**, 8, 1575–1581 (2011).

[2] Al-Rahayfeh, A. and M. Faezipour, "Eye Tracking and Head Movement Detection: A State-of-Art Survey", Translational Eng. in Health and Med. **1**, 2100212–2100212 (2013).

[3] Altun, Y., I. Tsochantaridis and T. Hofmann, "Hidden markov support vector machines", in "Int. Conference on Machine Learning", (2003).

[4] Amzica, F., "Basic physiology of burst-suppression", Epilepsia **50**, 12, 38–39 (2009).

[5] Anderson, T. W., "A modification of the sequential probability ratio test to reduce the sample size", Ann. Math. Statist. **31**, 1, 165–197, URL `https://doi.org/10.1214/aoms/1177705996` (1960).

[6] Arcelli Fontana, F. and M. Zanoni, "A tool for design pattern detection and software architecture reconstruction", Information Sciences (2011).

[7] Armitage, P., "Restricted sequential procedures", Biometrika **44**, 1/2, 9–26 (1957).

[8] Atkeson, C., A. Moore and S. Schaal, "Locally weighted learning", AI Review (1996).

[9] Barford, L. A., R. S. Fazzio and D. R. Smith, "An introduction to wavelets", Hewlett-Packard Labs, Bristol, UK, Tech. Rep. HPL-92-124 **2**, 1–29 (1992).

[10] Belkaid, Y. and T. W. Hand, "Role of the microbiota in immunity and inflammation", (2014).

[11] Ben-Hur, A., "Support vector clustering", Scholarpedia (2008).

[12] Bi, L., X.-A. Fan and Y. Liu, "EEG-based Brain-Controlled Mobile Robots: A Survey", Human-Machine Sys. **43**, 2, 161–176 (2013).

[13] Bishop, C. M., *Pattern Recognition and Machine Learning* (2006).

[14] Bliss, D. W. and P. A. Parker, "Temporal synchronization of mimo wireless communication in the presence of interference", IEEE Transactions on Signal Processing **58**, 3, 1794–1806 (2010).

[15] Brandon Westover, M., S. Ching, V. M. Kumaraswamy, O. Akeju, E. Pierce, S. S. Cash, R. Kilbride, E. N. Brown and P. L. Purdon, "The human burst suppresion electroencephalogram of deep hypothermia", Clinical Neurophysiology **126**, 10, 1901–1914 (2015).

[16] Breiman, L., "Bagging predictors", Machine Learning **24**, 2, 123–140 (1996).

[17] Breiman, L., "Random forests", Machine Learning **45**, 1, 5–32 (2001).

[18] Brown IV, C. H., "Delirium in the cardiac surgical intensive care unit", Current Opinion in Anaesthesiology **27**, 2, 117–122 (2014).

[19] Budden, B. S., L. C. Stonehill, J. R. Terry, A. V. Klimenko and J. O. Perry, "Characterization and Investigation of the Thermal Dependence of Cs2LiYCl6:Ce3+ (CLYC) Waveforms", IEEE Transactions on Nuclear Science (2012).

[20] Buehler, P., M. Everingham, D. P. Huttenlocher and A. Zisserman, "Upper body detection and tracking in extended signing sequences", International Journal of Computer Vision (2011).

[21] Bulling, A., "Human Visual Behaviour for Collaborative Human-Machine Interaction", in "Proc. of the 2015 ACM Int. Joint Conf. on Pervasive and Ubiquitous comp. and Proc. of the 2015 ACM Int. Symposium on Wearable comp.", pp. 901–905 (2015).

[22] Bussgang, J. and D. Middleton, "Optimum sequential detection of signals in noise", IRE Transactions on Information Theory **1**, 3, 5–18 (1955).

[23] Buzsáki, G., C. A. Anastassiou and C. Koch, "The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes", (2012).

[24] Caporaso, J. G., K. Bittinger, F. D. Bushman, T. Z. Desantis, G. L. Andersen and R. Knight, "PyNAST: A flexible tool for aligning sequences to a template alignment", Bioinformatics (2010).

[25] Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pẽa, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld and R. Knight, "QIIME allows analysis of high-throughput community sequencing data", (2010).

[26] Chakraborty, D., N. Kovvali, J. Wei, A. Papandreou-Suppappola, D. Cochran and A. Chattopadhyay, "Damage classification structural health monitoring in bolted structures using time-frequency techniques", Journal of Intelligent Material Systems and Structures **20**, 1289–1305 (2009).

[27] Charles, J., T. Pfister, M. Everingham and A. Zisserman, "Automatic and efficient human pose estimation for sign language videos", International Journal of Computer Vision (2014).

[28] Chen, Y. C., J. Greenbaum, H. Shen and H. W. Deng, "Association Between Gut Microbiota and Bone Health: Potential Mechanisms and Prospective", (2017).

[29] Cheng, S. Y. and M. M. Trivedi, "Turn-Intent Analysis using Body Pose for Intelligent Driver Assistance", Pervasive comp. **5**, 4, 28–37 (2006).

[30] Choo, J., O. U. Elci, K. Yang, M. W. Turk, M. A. Styn, S. M. Sereika, E. Music and L. E. Burke, "Longitudinal relationship between physical activity and cardiometabolic factors in overweight and obese adults", European Journal of Applied Physiology **108**, 2, 329–336 (2010).

[31] Chowdhury, A. K., D. Tjondronegoro, V. Chandran and S. G. Trost, *Ensemble Methods for Classification of Physical Activities from Wrist Accelerometry*, no. March (2017).

[32] Clifton, L., D. A. Clifton, M. A. Pimentel, P. J. Watkinson and L. Tarassenko, "Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors", IEEE Journal of Biomedical and Health Informatics (2014).

[33] Costello, E. K., K. Stagaman, L. Dethlefsen, B. J. M. Bohannan and D. A. Relman, "The application of ecological theory toward an understanding of the human microbiome.", Science (New York, N.Y.) (2012).

[34] Craig, S. B., L. G. Bandini, A. H. Lichtenstein, E. J. Schaefer and W. H. Dietz, "The impact of physical activity on lipids, lipoproteins, and blood pressure in preadolescent girls.", Pediatrics **98**, 3 Pt 1, 389–95 (1996).

[35] Cristianini, N. and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods* (Cambridge University Press, 2000).

[36] Cruz, J. A. and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis", Cancer Informatics **2**, 117693510600200030 (2006).

[37] Cuayáhuitl, H., "Robot Learning from Verbal Interaction: A Brief Survey", New Frontiers in Human-Robot Interaction p. 62 (2015).

[38] de la Iglesia, D., M. García-Remesal, A. Anguita, M. Muñoz-Mármol, C. Kulikowski and V. Maojo, "A machine learning approach to identify clinical trials involving nanodrugs and nanodevices from ClinicalTrials.gov.", PloS one (2014).

[39] De Vries, S. I., M. Engels and F. G. Garre, "Identification of children's activity type with accelerometer-based neural networks", Medicine and Science in Sports and Exercise **43**, 10, 1994–1999 (2011).

[40] Del Rosario, M. B., K. Wang, J. Wang, Y. Liu, M. Brodie, K. Delbaere, N. H. Lovell, S. R. Lord and S. J. Redmond, "A comparison of activity classification in younger and older cohorts using a smartphone", Physiological Measurement **35**, 11, 2269–2286 (2014).

[41] Di Stefano, M., G. Veneto, S. Malservisi and G. R. Corazza, "Small intestine bacterial overgrowth and metabolic bone disease.", Digestive diseases and sciences (2001).

[42] Dinger, M. K. and T. K. Behrens, "Accelerometer-determined physical activity of free-living college students", Medicine and Science in Sports and Exercise **38**, 4, 774–779 (2006).

[43] D'Mellow, B., M. D. Aspinall, R. O. Mackin, M. J. Joyce and A. J. Peyton, "Digital discrimination of neutrons and γ-rays in liquid scintillators using pulse gradient analysis", Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment (2007).

[44] Dong, B., A. H. K. Montoye, K. A. Pfeiffer and S. Biswas, "Energy-aware activity classification using wearable sensor networks", Proceedings of SPIE – The International Society for Optical Engineering pp. 8723–87230Y (2013).

[45] Doshi, A. and M. Trivedi, "A Comparative Exploration of Eye Gaze and Head Motion Cues for Lane Change Intent Prediction", in "Intelligent Vehicles Symposium", pp. 49–54 (2008).

[46] Duda, R. O., P. E. Hart and D. G. Stork, "Pattern Classification", (2001).

[47] Dutta, A., P. Chandhran, K. E. Holbert and E. B. Johnson, "Using decay time to discriminate neutron and gamma ray pulses from a clyc detector", in "2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)", pp. 1–7 (2015).

[48] Dutta, A. and K. E. Holbert, "Discrimination of neutron-gamma ray pulses with pileup using normalized cross correlation and principal component analysis", IEEE Transactions on Nuclear Science **63**, 6, 2764–2771 (2016).

[49] Dutta, A., O. Ma, M. P. Buman and D. W. Bliss, "Learning approach for classification of GENEActiv accelerometer data for unique activity identification", BSN 2016 - 13th Annual Body Sensor Networks Conference pp. 359–364 (2016).

[50] Dutta, A., O. Ma, M. Toledo, M. P. Buman and D. W. Bliss, "Comparing Gaussian mixture model and hidden Markov model to classify unique physical activities from accelerometer sensor data", Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016 pp. 339–344 (2017).

[51] Dutta, A., O. Ma, M. Toledo, A. F. Pregonero, B. E. Ainsworth, M. P. Buman and D. W. Bliss, "Improving methods for using supervised lab-based physical activity classification to estimate free-living physical activities", IEEE Transactions on Biomedical Engineering In preparation (2018).

[52] Dutta, A., O. Ma, C. M. Whisner and D. W. Bliss, "Analyzing relationship between gut-microbiome and bone health", IEEE Transactions on Biomedical Engineering In preparation (2018).

[53] Dutta, A., D. Pressler, E. Steiner, A. Yahja, V. Berisha, D. W. Bliss, M. S. Poole and S. R. Corman, "Analyzing the relationship between productivity and human communication in an organizational setting", Communication Methods and Measures In preparation (2018).

[54] Edgar, R. C., "Search and clustering orders of magnitude faster than BLAST", Bioinformatics (2010).

[55] Ekelund, U., P. W. Franks, S. Sharp, S. Brage and N. J. Wareham, "Increase in physical activity energy expenditure is associated with reduced metabolic risk independent of change in fatness and fitness", Diabetes Care **30**, 8, 2101–2106 (2007).

[56] Ellis, K., J. Kerr, S. Godbole, J. Staudenmayer and G. Lanckriet, "Hip and wrist accelerometer algorithms for free-living behavior classification", Medicine and Science in Sports and Exercise **48**, 5, 933–940 (2016).

[57] Ely, E. W., "Delirium protocol", (2005).

[58] Emotiv, "EEG Neuroheadset", `https://www.emotiv.com/support/` (accessed December 3, 2016) (2016).

[59] Erden, M. S. and T. Tomiyama, "Human-Intent Detection and Physically Interactive Control of a Robot without Force Sensors", Robotics **26**, 2, 370–382 (2010).

[60] Ergin, M. A., S. Uysal, D. L. Reich, A. Apaydin, S. L. Lansman, J. N. McCullough and R. B. Griepp, "Temporary neurological dysfunction after deep hypothermic circulatory arrest: A clinical marker of long-term function deficit", The Annals of Thoracic Surgery **67**, 6, 1887–1890 (1999).

[61] Escobar, M. D. and M. West, "Bayesian density estimation and inference using mixtures", Journal of the American Statistical Association **90**, 430, 577–588 (1995).

[62] Esmaeili-Sani, V., A. Moussavi-Zarandi, N. Akbar-Ashrafi, B. Boghrati and H. Afarideh, "Neutron-gamma discrimination based on bipolar trapezoidal pulse shaping using FPGAs in NE213", Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment (2012).

[63] Feichtenhofer, C., A. Pinz and R. P. Wildes, "Spatiotemporal multiplier networks for video action recognition", in "Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017", (2017).

[64] Florez Pregonero, A. A., *Monitors-Based Measurement of Sedentary Behaviors and Light Physical Activity in Adults*, Ph.D. thesis (2017).

[65] Forney, J., G.D., "The viterbi algorithm", Proceedings of the IEEE **61**, 3, 302–309 (1973).

[66] Foster, K. R., R. Koprowski and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research - commentary", BioMedical Engineering OnLine **13**, 1, 94, URL `https://doi.org/10.1186/1475-925X-13-94` (2014).

[67] Freedson, P. S., E. Melanson and J. Sirard, "Calibration of the Computer Science and Applications, Inc. accelerometer", Medicine and Science in Sports and Exercise **30**, 5, 777–781 (1998).

[68] Frigo, M. and S. G. Johnson, "FFTW: An adaptive software architecture for the FFT", in "ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings", (1998).

[69] Gamage, K. A., M. J. Joyce and N. P. Hawkes, "A comparison of four different digital algorithms for pulse-shape discrimination in fast scintillators", Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment (2011).

[70] Ganchev, T., T. Ganchev, N. Fakotakis, N. Fakotakis, G. Kokkinakis and G. Kokkinakis, "Comparative evaluation of various MFCC implementations on the speaker verification task", in Proc. of the SPECOM-2005 (2005).

[71] Gega, A., J. A. Rizzo, M. H. Johnson, M. Tranquilli, E. A. Farkas and J. A. Elefteriades, "Straight deep hypothermic arrest: Experience in 394 patients supports its effectiveness as a sole means of brain preservation", The Annals of Thoracic Surgery **84**, 3, 759–767 (2007).

[72] Geyik, C. S., A. Dutta, U. Y. Ogras and D. W. Bliss, "Decoding human intent using a wearable system and multi-modal sensor data", in "2016 50th Asilomar Conference on Signals, Systems and Computers", pp. 846–850 (2016).

[73] Gilks, W. R., S. Richardson and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall/CRC, 1996).

[74] Glodo, J., R. Hawrami, E. Van Loef, U. Shirwadkar and K. S. Shah, "Pulse shape discrimination with selected elpasolite crystals", IEEE Transactions on Nuclear Science (2012).

[75] Gupta, U., J. Park, H. Joshi and U. Y. Ogras, "Flexibility Aware Systems on Polymer: Concept to Prototype", IEEE Trans. on Multi-Scale Computing Systems (*To appear.*) (2017).

[76] Gutierrez, R. M., A. Dutta and D. W. Bliss, "Latency analysis for matched filter based detection of weak signals", in "MILCOM", (2018).

[77] Hagl, C., M. A. Ergin, J. D. Galla, S. Lansman, D. McCullough, J. Spielvogel, P. Sfeir, C. A. Bodian and R. B. Griepp, "Neurologic outcome after ascending aorta–aortic arch operations: Effect of brain protection technique in high-risk patients", The Journal of Thoracic and Cardiovascular Surgery **121**, 6, 1107–1121 (2001).

[78] Hagl, C., N. Khaladj, M. Karck, K. Kallenbach, R. Leyh, M. Winterhalter and A. Haverich, "Hypothermic circulatory arrest during ascending and aortic arch surgery: the theoretical impact of different cerebral perfusion techniques and other methods of cerebral protection", European Journal of Cardio-Thoracic Surgery **24**, 3, 371–378 (2003).

[79] Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA data mining software: an update", SIGKDD Explorations **11**, 1, 10–18 (2009).

[80] HANSEN, B. H., E. KOLLE, S. M. DYRSTAD, I. HOLME and S. A. ANDERSSEN, "Accelerometer-Determined Physical Activity in Adults and Older People", Medicine & Science in Sports & Exercise **44**, 2, 266–272 (2012).

[81] Hebb, A. O., F. Darvas and K. J. Miller, "Transient and state modulation of beta power in human subthalamic nucleus during speech production and finger movement", Neuroscience **202** (2012).

[82] Hebb, A. O., J. J. Zhang, M. H. Mahoor, C. Tsiokos, C. Matlack, H. J. Chizeck and N. Pouratian, "Creating the feedback loop: Closed-loop neurostimulation", Neurosurgery Clinics of North America **25**, 1, 187–204 (2014).

[83] Hikihara, Y., C. Tanaka, Y. Oshima, K. Ohkawara, K. Ishikawa-Takata and S. Tanaka, "Prediction models discriminating between nonlocomotive and locomotive activities in children using a triaxial accelerometer with a gravity-removal physical activity classification algorithm", PLoS ONE **9**, 4 (2014).

[84] Hirtz, D., D. J. Thurman, K. Gwinn-Hardy, M. Mohamed, A. R. Chaudhuri and R. Zalutsky, "How common are the "common" neurologic disorders?", Neurology **6**, 326–327 (2007).

[85] Hutkins, R. W., J. A. Krumbeck, L. B. Bindels, P. D. Cani, G. Fahey, Y. J. Goh, B. Hamaker, E. C. Martens, D. A. Mills, R. A. Rastal, E. Vaughan and M. E. Sanders, "Prebiotics: Why definitions matter", (2016).

[86] Ibáñez, L., M. Rouleau, A. Wakkach and C. Blin-Wakkach, "Gut microbiome and bone", (2018).

[87] IRN-IS-200H-003, *Interface Specification: NAVSTAR GPS Space Segment/-Navigation User Segment Interface*, Global Positioning Systems Directorate, irn-is-200h-003 edn. (2015).

[88] Ishwaran, H. and L. F. James, "Gibbs sampling methods for stick-breaking priors", Journal of the American Statistical Association **96**, 453, 161–173 (2001).

[89] James, M. L., N. D. Andersen, M. Swaminathan, B. Phillips-Bute, J. M. Hanna, G. R. Smigla, M. E. Barfield, S. D. Bhattacharya, J. B. Williams, J. G. Gaca, A. M. Husain and G. C. Hughes, "Predictors of electrocerebral inactivity with deep hypothermia", The Journal of Thoracic and Cardiovascular Surgery **147**, 3, 1002 – 1007 (2014).

[90] Jetté, M., K. Sidney and G. Blümchen, "Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity", Clinical Cardiology **13**, 8, 555–565 (1990).

[91] Jiang, H., J. J. Zhang, A. Hebb and M. H. Mahoor, "Time-frequency analysis of brain electrical signals for behavior recognition in patients with parkinson's disease", in "Asilomar Conference on Signals, Systems and Computers", (2013).

[92] Johnson, E. B., C. Whitney, S. Vogel, J. F. Christian, K. Holbert and P. Chandhran, "High event rate, pulse shape discrimination algorithm for CLYC", in "2015 IEEE International Symposium on Technologies for Homeland Security, HST 2015", (2015).

[93] Jolliffe, I. T., *Principal Component Analysis, Second Edition* (2002).

[94] Jones, R. M., J. G. Mulle and R. Pacifici, "Osteomicrobiology: The influence of gut microbiota on bone in health and disease", (2016).

[95] Kanungo, T., D. Mount, N. Netanyahu, C. Piatko, R. Silverman and A. Wu, "An efficient k-means clustering algorithm: analysis and implementation", IEEE Transactions on Pattern Analysis and Machine Intelligence (2002).

[96] Karray, F., M. Alemzadeh, J. A. Saleh and M. N. Arab, "Human-Computer Interaction: Overview on State of the Art", International Journal on Smart Sensing and Intelligent Systems **1**, 1, 137–159 (2008).

[97] Kay, S., *Fundamentals of Statistical Signal Processing: Detection theory*, vol. 2 of *Prentice Hall Signal Processing Series* (Prentice-Hall PTR, 1998).

[98] Kennedy, P. J., J. F. Cryan, T. G. Dinan and G. Clarke, "Kynurenine pathway metabolism and the microbiota-gut-brain axis", (2017).

[99] Killworth, P. D. and H. R. Bernard, "Informant Accuracy in Social Network Data", Human Organization **35**, 269–286, URL `http://sfaa.metapress.com/index/10215J2M359266N2.pdf` (1976).

[100] Kleinsmith, A. and N. Bianchi-Berthouze, "Affective Body Expression Perception and Recognition: A Survey", Affective comp. **4**, 1, 15–33 (2013).

[101] Knoll, G., "Radiation detection and measurement", Wiley (2005).

[102] Kourou, K., T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction", (2015).

[103] Kretschmer, T., "Information and Communication Technologies and Productivity Growth", OECD Digital Economy Papers , 195, 1–27 (2012).

[104] Kudo, M. and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers", Pattern Recognition (2000).

[105] Kulic, D. and E. A. Croft, "Affective State Estimation for Human–Robot Interaction", Robotics **23**, 5, 991–1000 (2007).

[106] Lee, D. W., L. C. Stonehill, A. Klimenko, J. R. Terry and S. R. Tornga, "Pulse-shape analysis of Cs2LiYCl6:Ce scintillator for neutron and gamma-ray discrimination", Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment (2012).

[107] Levy, W. J., "Quantitative analysis of eeg changes during hypothermia", Anesthesiology **60**, 4, 291–297 (1984).

[108] Levy, W. J., E. Pantin, S. Mehta and M. McGarvey, "Hypothermia and the approximate entropy of the electroencephalogram", Anesthesiology **98**, 1, 53–57 (2003).

[109] Li, L., A. J. Brockmeier, J. S. Choi, J. T. Francis, J. C. Sanchez and J. C. Príncipe, "A Tensor-Product-Kernel Framework for Multiscale Neural Activity Decoding and Control", Computational Intelligence and Neuroscience **2014**, 2 (2014).

[110] Liu, G., M. J. Joyce, X. Ma and M. D. Aspinall, "A digital method for the discrimination of neutrons and γ rays with organic scintillation detectors using frequency gradient analysis", IEEE Transactions on Nuclear Science (2010).

[111] Liu, Z., T. Z. Desantis, G. L. Andersen and R. Knight, "Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers", Nucleic Acids Research (2008).

[112] Lopez-Azcarate, J., M. Tainta, M. C. Rodriguez-Oroz, M. Valencia, R. Gonzalez, J. Guridi, J. Iriarte, J. A. Obeso, J. Artieda and M. Aleg, "Coupling between beta and high-frequency activity in the human subthalamic nucleus may be a pathophysiological mechanism in Parkinson's disease", Neurobiology of Disease **30**, 6667–6677 (2010).

[113] Ma, O., A. Dutta, D. W. Bliss and A. Z. Crepeau, "Predicting postoperative delirium in patients undergoing deep hypothermia circulatory arrest", in "2017 51st Asilomar Conference on Signals, Systems, and Computers", pp. 1313–1317 (2017).

[114] Mackay, D. J. C., "Introduction to Gaussian processes", Neural Networks and Machine Learning **168**, 1996, 133–165 (1998).

[115] Mahmoudi, B. and J. C. Sanchez, "A Symbiotic Brain-Machine Interface through Value-based Decision Making", PloS one **6**, 3, e14760 (2011).

[116] Mallat, S. G. and Z. Zhang, "Matching pursuits with time-frequency dictionaries", IEEE Transactions on Signal Processing **41**, 3397–3415 (1993).

[117] Mannini, A., S. S. Intille, M. Rosenberger and A. M. Sabatini, "Wrist or Ankle", **45**, 11, 2193–2203 (2014).

[118] Mannini, A., M. Rosenberger, W. L. Haskell, A. M. Sabatini and S. S. Intille, "Activity recognition in youth using single accelerometer placed at wrist or ankle", Medicine and Science in Sports and Exercise **49**, 4, 801–812 (2017).

[119] MATLAB, *version 9.4.0 (R2018a)* (The MathWorks Inc., Natick, Massachusetts, 2018).

[120] McCabe, L., R. A. Britton and N. Parameswaran, "Prebiotic and Probiotic Regulation of Bone Health: Role of the Intestine and its Microbiome", (2015).

[121] McDonald, D., E. Hyde, J. W. Debelius, J. T. Morton, A. Gonzalez, G. Ackermann, A. A. Aksenov, B. Behsaz, C. Brennan, Y. Chen, L. DeRight Goldasich, P. C. Dorrestein, R. R. Dunn, A. K. Fahimipour, J. Gaffney, J. A. Gilbert, G. Gogul, J. L. Green, P. Hugenholtz, G. Humphrey, C. Huttenhower, M. A. Jackson, S. Janssen, D. V. Jeste, L. Jiang, S. T. Kelley, D. Knights, T. Kosciolek, J. Ladau, J. Leach, C. Marotz, D. Meleshko, A. V. Melnik, J. L. Metcalf, H. Mohimani, E. Montassier, J. Navas-Molina, T. T. Nguyen, S. Peddada, P. Pevzner, K. S. Pollard, G. Rahnavard, A. Robbins-Pianka, N. Sangwan, J. Shorenstein, L. Smarr, S. J. Song, T. Spector, A. D. Swafford, V. G. Thackray, L. R. Thompson, A. Tripathi, Y. Vázquez-Baeza, A. Vrbanac, P. Wischmeyer, E. Wolfe, Q. Zhu, and R. Knight, "American gut: an open platform for citizen science microbiome research", mSystems **3**, 3, URL `https://msystems.asm.org/content/3/3/e00031-18` (2018).

[122] McDonald, D., M. N. Price, J. Goodrich, E. P. Nawrocki, T. Z. Desantis, A. Probst, G. L. Andersen, R. Knight and P. Hugenholtz, "An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea", ISME Journal (2012).

[123] McLachlan, R. D., *Pressure as a Non-dominant Hand Input Modality for Bimanual Interaction Techniques on Touchscreen Tablets*, Ph.D. thesis, University of Glasgow (2015).

[124] Mesmer-Magnus, J. R. and L. DeChurch, "Information sharing and team performance: a meta-analysis.", The Journal of applied psychology **94**, 2, 535–46, URL `http://www.ncbi.nlm.nih.gov/pubmed/19271807` (2009).

[125] Middleton, D., "Statistical criteria for the detection of pulsed carriers in noise.", Journal of Applied Physics **24**, 4, 371–391 (1953).

[126] Miranda, R. A., W. D. Casebeer, A. M. Hein, J. W. Judy, E. P. Krotkov, T. L. Laabs, J. E. Manzo, K. G. Pankratz, G. A. Pratt, J. C. Sanchez *et al.*, "DARPA-Funded Efforts in the Development of Novel Brain-Computer Interface Technologies", Journal of Neuroscience Methods **244**, 52–67 (2015).

[127] Misra, H., S. Ikbal, S. Sivadas and H. Bourlard, "Multi-resolution spectral entropy feature for robust ASR", in "ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings", (2005).

[128] Montoye, A. H., L. M. Mudd, S. Biswas and K. A. Pfeiffer, "Energy Expenditure Prediction Using Raw Accelerometer Data in Simulated Free Living", Med Sci Sports Exerc **47**, 8, 1735–1746 (2015).

[129] Moon, T., "The expectation-maximization algorithm", IEEE Signal Processing Magazine **13**, 6, 47–60 (1996).

[130] Murphy, P. J., "Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis.", The Journal of the Acoustical Society of America (1999).

[131] Murphy-Chutorian, E. and M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey", Pattern Analysis and Machine Intelligence **31**, 4, 607–626 (2009).

[132] Muztoba, M., U. Gupta, M. Tanvir and U. Y. Ogras, "Robust Communication with IoT Devices using Wearable Brain Machine Interfaces", in "Proc. of Int. Conf. on Computer-Aided Design", (2015).

[133] Nakagawa, S., L. Wang and S. Ohtsuka, "Speaker identification and verification by combining MFCC and phase information", IEEE Transactions on Audio, Speech and Language Processing (2012).

[134] Nakhostin, M., "Recursive algorithms for digital implementation of neutron/gamma discrimination in liquid scintillation detectors", Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment (2012).

[135] Nef, T., P. Urwyler, M. Büchler, I. Tarnanas, R. Stucki, D. Cazzoli, R. Müri and U. Mosimann, "Evaluation of three state-of-the-art classifiers for recognition of activities of daily living from smart home ambient data", Sensors (Switzerland) (2015).

[136] Nehaniv, C. L., K. Dautenhahn, J. Kubacki, M. Haegele, C. Parlitz and R. Alami, "A Methodological Approach Relating the Classification of Gesture to Identification of Human Intent in the Context of Human-robot Interaction", in "Robot and Human Interactive Comm.", pp. 371–377 (2005).

[137] Neish, A. S., "Microbes in Gastrointestinal Health and Disease", (2009).

[138] Nielsen, H. B., M. Almeida, A. S. Juncker, S. Rasmussen, J. Li, S. Sunagawa, D. R. Plichta, L. Gautier, A. G. Pedersen, E. Le Chatelier, E. Pelletier, I. Bonde, T. Nielsen, C. Manichanh, M. Arumugam, J. M. Batto, M. B. Quintanilha Dos Santos, N. Blom, N. Borruel, K. S. Burgdorf, F. Boumezbeur, F. Casellas, J. Doré, P. Dworzynski, F. Guarner, T. Hansen, F. Hildebrand, R. S. Kaas, S. Kennedy, K. Kristiansen, J. R. Kultima, P. Léonard, F. Levenez, O. Lund, B. Moumen, D. Le Paslier, N. Pons, O. Pedersen, E. Prifti, J. Qin, J. Raes, S. Sørensen, J. Tap, S. Tims, D. W. Ussery, T. Yamada, P. Renault, T. Sicheritz-Ponten, P. Bork, J. Wang, S. Brunak and S. D. Ehrlich, "Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes", Nature Biotechnology (2014).

[139] Nossa, C. W., W. E. Oberdorf, L. Yang, J. A. Aas, B. J. Paster, T. Z. de Santis, E. L. Brodie, D. Malamud, M. A. Poles and Z. Pei, "Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome", World Journal of Gastroenterology (2010).

[140] Obinikpo, A. and B. Kantarci, "Big Sensed Data Meets Deep Learning for Smarter Health Care in Smart Cities", Journal of Sensor and Actuator Networks (2017).

[141] Ohlsson, C. and K. Sjögren, "Effects of the gut microbiota on bone mass", (2015).

[142] Oki, K., M. Toyama, T. Banno, O. Chonan, Y. Benno and K. Watanabe, "Comprehensive analysis of the fecal microbiota of healthy Japanese adults reveals a new bacterial lineage associated with a phenotype characterized by a high frequency of bowel movements and a lean body type", BMC Microbiology (2016).

[143] Okita, Y., K. Minatoya, O. Tagusari, M. Ando, K. Nagatsuka and S. Kitamura, "Prospective comparative study of brain protection in total aortic arch replacement: Deep hypothermic circulatory arrest with retrograde cerebral perfusion or selective antegrade cerebral perfusion", The Annals of Thoracic Surgery **72**, 1, 72–79 (2001).

[144] Oostenveld, R. and P. Praamstra, "The five percent electrode system for high-resolution eeg and erp measurements", Clinical Neurophysiology **112**, 4, 713–719 (2001).

[145] Ortega, J. P., M. del Rocio Boone Rojas and M. J. Somodevilla Garcia, "Research issues on K-means Algorithm : An Experimental Trial Using Matlab", Proceedings of the 2nd Workshop on Semantic Web and New Technologies pp. 83–96 (2009).

[146] Oscos, G. C., T. M. Khoshgoftaar and R. Wald, "Rotation Invariant Face Recognition Survey", in "Info. Reuse and Integration (IRI)", pp. 835–840 (2014).

[147] Pampiglione, G., "Neurophysiological aspects of cerebral ischæmia", Proceedings of the Royal Society of Medicine **53**, 5, 329–332 (1960).

[148] Panagiotakis, C. and G. Tziritas, "A speech/music discriminator based on RMS and zero-crossings", IEEE Transactions on Multimedia (2005).

[149] Pikrakis, A., T. Giannakopoulos and S. Theodoridis, "Gunshot detection in audio streams from movies by means of dynamic programming and Bayesian Networks", in "ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings", (2008).

[150] PINCUS, J. D., "Communication satisfaction, job satisfaction, and job performance", Human Communication Research **12**, 3, 395–419 (1986).

[151] Rabiner, L. and B. Juang, "An introduction to hidden Markov models", IEEE ASSP Magazine (1986).

[152] Rabiner, L. R., "On the Use of Autocorrelation Analysis for Pitch Detection", IEEE Transactions on Acoustics, Speech, and Signal Processing (1977).

[153] Rabiner, L. R., "A tutorial on hidden Markov models and selected applications in speech recognition", Proceedings of the IEEE **77**, 2, 257–286 (1989).

[154] Rautaray, S. S. and A. Agrawal, "Vision based Hand Gesture Recognition for Human Computer Interaction: A Survey", Artificial Intelligence Review **43**, 1, 1–54 (2015).

[155] Reich, D. L., S. Uysal, M. Sliwinksi, M. A. Ergin, R. A. Kahn, S. N. Konstadt, J. McCullough, M. R. Hibbard, W. A. Gordon and R. B. Griepp, "Neuropsychologic outcome after deep hypothermic circulatory arrest in adults", The Journal of Thoracic and Cardiovascular Surgery **117**, 1, 156–163 (1999).

[156] Ronchi, E., P. A. Söderström, J. Nyberg, E. Andersson Sundén, S. Conroy, G. Ericsson, C. Hellesen, M. Gatu Johnson and M. Weiszflog, "An artificial neural network based neutron-gamma discrimination and pile-up rejection framework for the BC-501 liquid scintillation detector", Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment (2009).

[157] Rosenberg, D., S. Godbole, K. Ellis, C. Di, A. Lacroix, L. Natarajan and J. Kerr, "Classifiers for Accelerometer-Measured Behaviors in Older Women", Medicine and Science in Sports and Exercise **49**, 3, 610–616 (2017).

[158] S.A. Hellweg, S. P., "Communication and productivity in organizations", Public Productivity Review , 6(4), 276–288 (1982).

[159] Sadjadi, S. O. and J. H. Hansen, "Unsupervised speech activity detection using voicing measures and perceptual spectral flux", IEEE Signal Processing Letters (2013).

[160] Saltzman, J. R. and R. M. Russell, "The Aging Gut: Nutritional issues", (1998).

[161] Samuel, A. L., "Some Studies in Machine Learning Using the Game of Checkers", IBM Journal of Research and Development (1959).

[162] Sasaki, J. E., A. M. Hickey, J. W. Staudenmayer, D. John, J. A. Kent and P. S. Freedson, "Performance of activity classification algorithms in free-living older adults", Medicine and Science in Sports and Exercise **48**, 5, 941–949 (2016).

[163] Savitzky, A. and M. J. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures", Analytical Chemistry (1964).

[164] Schwarzer, M., K. Makki, G. Storelli, I. Machuca-Gayet, D. Srutkova, P. Hermanova, M. E. Martino, S. Balmand, T. Hudcovic, A. Heddi, J. Rieusset, H. Kozakova, H. Vidal and F. Leulier, "Lactobacillus plantarum strain maintains growth of infant mice during chronic undernutrition", Science (2016).

[165] Sharma, R., V. I. Pavlovic and T. S. Huang, "Toward Multimodal Human-computer Interface", Proc. of the IEEE **86**, 5, 853–869 (1998).

[166] Sharon, G., T. R. Sampson, D. H. Geschwind and S. K. Mazmanian, "The Central Nervous System and the Gut Microbiome", (2016).

[167] Shevade, S. K., S. S. Keerthi, C. Bhattacharyya and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression", IEEE Transactions on Neural Networks **11**, 5, 1188–1193 (2000).

[168] Simpson, A. P., S. Jones, M. J. Clapham and S. A. McElhaney, "A Review of Neutron Detectino Technology Alternatives to Helium-3 for Safeguards Applications", INMM 52 Annual Meeting (2011).

[169] Sirichana, W., B. A. Dolezal, E. V. Neufeld, X. Wang and C. B. Cooper, "Wrist-worn triaxial accelerometry predicts the energy expenditure of non-vigorous daily physical activities", Journal of Science and Medicine in Sport **20**, 8, 761–765 (2017).

[170] Sjögren, K., C. Engdahl, P. Henning, U. H. Lerner, V. Tremaroli, M. K. Lagerquist, F. Bäckhed and C. Ohlsson, "The gut microbiota regulates bone mass in mice", Journal of Bone and Mineral Research (2012).

[171] Skotte, J., M. Korshøj, J. Kristiansen, C. Hanisch and A. Holtermann, "Detection of Physical Activity Types Using Triaxial Accelerometers", Journal of Physical Activity and Health **11**, 1, 76–84 (2014).

[172] Smirnov, N., "Table for Estimating the Goodness of Fit of Empirical Distributions", The Annals of Mathematical Statistics (1948).

[173] Smola, A. J. and B. Schölkopf, "A tutorial on support vector regression", (2004).

[174] Sommer, F. and F. Bäckhed, "The gut microbiota-masters of host development and physiology", (2013).

[175] Spor, A., O. Koren and R. Ley, "Unravelling the effects of the environment and host genotype on the gut microbiome", Nature Reviews Microbiology (2011).

[176] Stotzer, P. O., C. Johansson, D. Mellstrom, G. Lindstedt and A. F. Kilander, "Bone mineral density in patients with small intestinal bacterial overgrowth", Hepatogastroenterology (2003).

[177] Sudhakar, P., A. Farooq and S. Patnaik, "Measuring productivity of software development teams", Serbian Journal of Management **7**, 1, 65–75, URL `http://scindeks.ceon.rs/Article.aspx?artid=1452-48641201065S` (2012).

[178] Sundt III, T. M., T. A. Orszulak, D. J. Cook and H. V. Schaff, "Improving results of open arch replacement", The Annals of Thoracic Surgery **86**, 3, 787–796 (2008).

[179] Tapia, E. M., S. S. Intille, W. Haskell, K. W. J. Larson, A. King and R. Friedman, "Real-Time Recognition of Physical Activities and their Intensitiies Using Wireless Accelerometers and a Heart Monitor", International Symposium on Wearable Computers pp. 37–40 (2007).

[180] Tedesco, S., J. Barton and B. O'Flynn, "A review of activity trackers for senior citizens: Research perspectives, commercial landscape and the role of the insurance industry", Sensors (Switzerland) **17**, 6 (2017).

[181] Texas Instruments Inc., "CC-2650 Microcontroller", `http://www.ti.com/product/CC2650` (2016).

[182] Tremaroli, V. and F. Bäckhed, "Functional interactions between the gut microbiota and host metabolism", (2012).

[183] TROST, S. G., W.-K. WONG, K. A. PFEIFFER and Y. ZHENG, "Artificial Neural Networks to Predict Activity Type and Energy Expenditure in Youth", Medicine & Science in Sports & Exercise **44**, 9, 1801–1809 (2012).

[184] Turnbaugh, P. J., R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis and J. I. Gordon, "An obesity-associated gut microbiome with increased capacity for energy harvest", Nature (2006).

[185] U.S. Department of Health and Human Services, "2008 Physical activity guidelines for Americans", President's Council on Physical Fitness & Sports Research Digest **9**, 4, 1–8 (2008).

[186] Vapnik, V., S. Golowich and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing", in "Advances in Neural Information Processing Systems 9", pp. 281–287 (MIT Press, 1996).

[187] Vaseghi, S., "Spectral Substraction", Advanced Digital Signal Processing and Noise Reduction (2000).

[188] Wagner, S. and M. Ruhe, "A Systematic Review of Productivity Factors in Software Development", Software Productivity Analysis and Cost Estimation (SPACE 2008) pp. 1–6 (2008).

[189] Wald, A., "Sequential tests of statistical hypotheses", Ann. Math. Statist. **16**, 2, 117–186, URL `http://dx.doi.org/10.1214/aoms/1177731118` (1945).

[190] Walter, J. and R. Ley, "The Human Gut Microbiome: Ecology and Recent Evolutionary Changes", Annual Review of Microbiology (2011).

[191] Wang, J., Y. Wang, W. Gao, B. Wang, H. Zhao, Y. Zeng, Y. Ji and D. Hao, "Diversity analysis of gut microbiota in osteoporosis and osteopenia patients", PeerJ (2017).

[192] Wang, J., L. Yang, L. Gao and Q. Miao, "Current progress on weak signal detection", in "2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)", pp. 1812–1818 (2013).

[193] Wang, Q., S. Lohit, M. J. Toledo, M. P. Buman and P. Turaga, "A statistical estimation framework for energy expenditure of physical activities from a wrist-worn accelerometer", Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS **2016-Octob**, 2631–2635 (2016).

[194] Weaver, C. M., "Diet, Gut Microbiome, and Bone Health", (2015).

[195] Welch, W. A., D. R. Bassett, D. L. Thompson, P. S. Freedson, J. W. Staudenmayer, D. John, J. A. Steeves, S. A. Conger, T. Ceaser, C. A. Howe, J. E. Sasaki and E. C. Fitzhugh, "Classification accuracy of the wrist-worn gravity estimator of normal everyday activity accelerometer", Medicine and Science in Sports and Exercise **45**, 10, 2012–2019 (2013).

[196] Wellman, B. and S. D. Berkowitz, *Social Structures: A Network Approach*, vol. 2 (1988).

[197] Whisner, C. M., B. R. Martin, C. H. Nakatsu, G. P. McCabe, L. D. McCabe, M. Peacock and C. M. Weaver, "Soluble maize fibre affects short-term calcium absorption in adolescent boys and girls: A randomised controlled trial using dual stable isotopic tracers", British Journal of Nutrition (2014).

[198] Whisner, C. M., B. R. Martin, C. H. Nakatsu, J. A. Story, C. J. MacDonald-Clarke, L. D. McCabe, G. P. McCabe and C. M. Weaver, "Soluble Corn Fiber Increases Calcium Absorption Associated with Shifts in the Gut Microbiome: A Randomized Dose-Response Trial in Free-Living Pubertal Females", Journal of Nutrition (2016).

[199] Whisner, C. M., B. R. Martin, M. H. Schoterman, C. H. Nakatsu, L. D. McCabe, G. P. McCabe, M. E. Wastney, E. G. Van Den Heuvel and C. M. Weaver, "Galacto-oligosaccharides increase calcium absorption and gut bifidobacteria in young girls: A double-blind cross-over trial", British Journal of Nutrition (2013).

[200] Wiens, J. and B. C. Wallace, "Editorial: special issue on machine learning for health and medicine", Machine Learning **102**, 3, 305–307, URL `https://doi.org/10.1007/s10994-015-5533-9` (2016).

[201] Williamson, J. R., D. W. Bliss, D. W. Browne and J. T. Narayanan, "Seizure prediction using eeg spatiotemporal correlation structure", Epilepsy & Behavior **25**, 2, 230–238 (2012).

[202] Williamson, J. R., D. W. Bliss, D. W. Browne and J. T. Narayanan, "Seizure prediction using eeg spatiotemporal correlation structure", Epilepsy & Behavior **25**, 2, 230–238 (2012).

[203] Young, T., D. Hazarika, S. Poria and E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing", (2017).

[204] Yousefi, S., L. Lucchese and M. D. Aspinall, "Digital discrimination of neutrons and gamma-rays in liquid scintillators using wavelets", Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment (2009).

[205] Zaker, N., A. Dutta, A. Maurer, J. J. Zhang, S. Hanrahan, A. O. Hebb, N. Kovvali and A. Papandreou-Suppappola, "Adaptive learning of behavioral tasks for patients with parkinson's disease using signals from deep brain stimulation", in "2014 48th Asilomar Conference on Signals, Systems and Computers", pp. 208–212 (2014).

[206] Zhang, S., A. V. Rowlands, P. Murray and T. L. Hurst, "Physical activity classification using the GENEA wrist-worn accelerometer", Medicine and Science in Sports and Exercise **44**, 4, 742–748 (2012).

[207] Zhu, S., "Study on weak signal detection methods", (2013).

[208] Ziganshin, B. A. and J. A. Elefteriades, "Deep hypothermic circulatory arrest", Annals of Cardiovascular Surgery **2**, 3, 303–315 (2013).

APPENDIX A

LIST OF ACRONYMS

BMI Body Mass Index

BMC Bone Mineral Content

BMD Bone Mineral Density

CAE Classifier Attribute Evaluation

CC Correlation Coefficient

DBS Deep Brain Simulation

DHMM Discrete Hidden Markov Model

DPGMM Dirichlet Process Gaussian Mixture Model

EEG Electroencephalogram

ESD Energy Spectral Density

FE Feature Extraction

FS Feature Selection

GM Gut Microbiome

GMM Gaussian Mixture Model

GPR Gaussian Process Regression

HMM Hidden Markov Model

LFP Local Field Potential

LR Linear Regression

LWL Locally Weighted Learning

MAE Maximum Absolute Error

ML Machine Learning

MPD Matching Pursuit Decomposition

NCC Normalized Cross Correlation

PCA Principal Component Analysis

PD Parkinson's Disease

RC Random Committee

RF Random Forest

ROC Receiver Operating Characteristic

SISO Single Input Single Output

SP Signal Processing

SNR Signal to Noise Ratio

SVM Support Vector Machine

SVR Support Vector Regression

BLISS Bliss Laboratory of Information, Signals, and Systems

ASU Arizona State University