DEEP LEARNING WITH MULTIMODAL DATA FOR HEALTHCARE

A Dissertation

by

NUR HAFIEZA BINTI ISMAIL

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Xia Hu
Committee Members,	Ruihong Huang
	Theodora Chaspari
	Joshua Hicks
Head of Department,	Timothy Jacobs

December 2020

Major Subject: Interdisciplinary Engineering

ABSTRACT

Healthcare plays a significant role in promoting and maintaining health, preventing and managing disease, reducing health disability and premature death, and educating a healthy lifestyle. However, healthcare information is well known for its big data that is too vast and complex to manage manually. The healthcare data is heterogeneous, containing different modalities or types of information such as text, audio, images, and multi-type. Over the last few years, the Deep Learning (DL) approach has successfully solved many issues. The primary structure of DL lies in the Artificial Neural Network (ANN). It is also known as representation learning techniques as these approaches can effectively identify hidden patterns of the data without requiring any explicit feature extraction mechanism. In other words, DL architectures also support automatic feature extraction. It is different than machine learning techniques, where there is no need to extract features separately in DL.

In this dissertation, we proposed three DL architectures to handle multiple modalities data in healthcare. We systematically develop prediction models for identifying health conditions in several groups, including Post-Traumatic Stress Disorder (PTSD), Parkinson's Disease (PD), and PD with Dementia (PD-Dementia). First, we designed the DL framework for identifying PTSD among cancer survivors via social media. After that, we apply the DL time series approach to forecast PD patients' future health status. Last, we build DL architecture to identify dementia in diagnosed PD patients. This work is motivated by several medical theories and health informatics perspectives. We have handled multimodal healthcare data information throughout these years, including text, audio features, and multivariate data. We also carefully studied each disease's background, including the symptoms and test assessment run by healthcare. We explored the online social media potential and medical applications capability for disease diagnosis and a health monitoring system to employ the developed models in a real-world scenario.

The DL for healthcare can become very helpful for supporting clinician's decisions and improving patient care. The leading institutions and medical bodies have recognized the benefits it

brings, and the popularity of the solutions are well known. With support from a reliable computational system, it could help healthcare decide particular needs and environments and reduce the stresses that medical professionals may experience daily. Healthcare has high hopes for the role of DL in clinical decision support and predictive analytics for a wide variety of conditions.

DEDICATION

I dedicated this dissertation to my mother and my father, for making me be who I am and supporting me all the way!

ACKNOWLEDGMENTS

It would not have been possible to write this dissertation without the help and support of the kind people around me. There are many people I owe a debt of gratitude and thanks to for my academic journey here in the United States of America.

Foremost, I would like to express my sincere gratitude to my advisor Dr. Xia Hu for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me throughout my research and writing of this dissertation. I could not have imagined having a better advisor and mentor than him.

Besides my advisor, I would like to thank the rest of my committee members: Dr. Ruihong Huang, Dr. Theadora Chaspari, and Dr. Joshua Hicks, for their time, effort, encouragement, and insightful comments.

Also, I wish to express my deep and warm thanks to my labmates at the Data Analytics at Texas A&M (DATA) Lab, in the Department of Computer Science and Engineering, for accompanying and helping me all the time. Their valuable advice for my work has been extremely helpful. You guys have been amazing friends to be surrounded with throughout this journey. Plus, you guys have made it possible for me to enjoy this foreign land and survive my entire graduate school.

Last but not least, I am very grateful to have my husband, my son, my parents, my siblings, my entire family, and friends who have advised and pushed me forward throughout this endeavor. They have been the biggest support system for me during this roller coaster journey. The product of this dissertation would not be possible without all of them.

This work, in part, is supported by the Ministry of Higher Education Malaysia and University Malaysia Pahang.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Xia Hu [advisor], Professor Ruihong Huang, and Professor Theodora Chaspari, of the Department of Computer Science and Engineering and Professor Joshua Hicks of the Department of Psychological & Brain Sciences.

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

Funding Sources

This work, in part, is supported by the Ministry of Higher Education Malaysia and University Malaysia Pahang.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION[1]

1.1 Social Computing for Healthcare

Nowadays, most people have more interaction with people in the virtual world than in the real world. Especially during the current world Coronavirus Disease of 2019 (COVID-19) pandemic, it has affected many of us worldwide. This situation has forces many people to work remotely from home to stop the spread. Many work and services are transferred to a virtual environment to reduce physical interaction and contact with others. This situation has immediately changed our daily work and routine, which now we are spending a lot of time more on screen than before.

Social computing uses technology information to allow its users or netizen to create unlimited online communities and content. A netizen describes an internet user or a person who actively participates in online communities or the internet [2]. The virtual environment is used widely for many activities, including entertainment, economy, education, politics, healthcare, etc. It shows that this technology is more convenient and efficient in many ways for most people. It improves the collaboration, communication, and interaction among participants in communities.

In healthcare, social computing applications started to receive more attention to promote better services to medical professionals, patients, and the public. It has become a landmark in providing health information to society [3]. Many users choose to rely on this platform to track their health status and seek more health information. The healthcare social computing applications can provide the doctor's decision regarding the patient's health condition and allow two-way interaction between doctor and patient to help patients make decisions.

Examples of social computing applications that applicable for healthcare purposes are online social media, mobile health applications, blogs, podcasting for education, and social knowledge sharing [4]. Figure 1.1 presents examples of social computing applications for healthcare. Social media is one of the most popular internet activities with a high number of user engagement rates.

[1]Part of this chapter is reprinted with permission from "Social Media and Psychological Disorder" by Nur Hafieza Ismail, Mengnan Du, and Xia Hu, 2019, In: Bian J., Guo Y., He Z., Hu X. (eds) Social Web and Health Research. Springer, Cham., Pages 171-192, doi.org/10.1007/978-3-030-14714-3_9, Copyright 2019 by Springer, Cham.

By requiring limited internet knowledge from users, social media has offered a better way of connecting with people than conventional approaches [5]. For example, Facebook can create a "page" for any purpose and allow other users to be members. For patients, social networks supply a pleasant environment for information seeking, socializing, and getting support from others who were currently facing similar issues [6]. It also builds online communities for patients affected by diseases, their families, and medical experts.

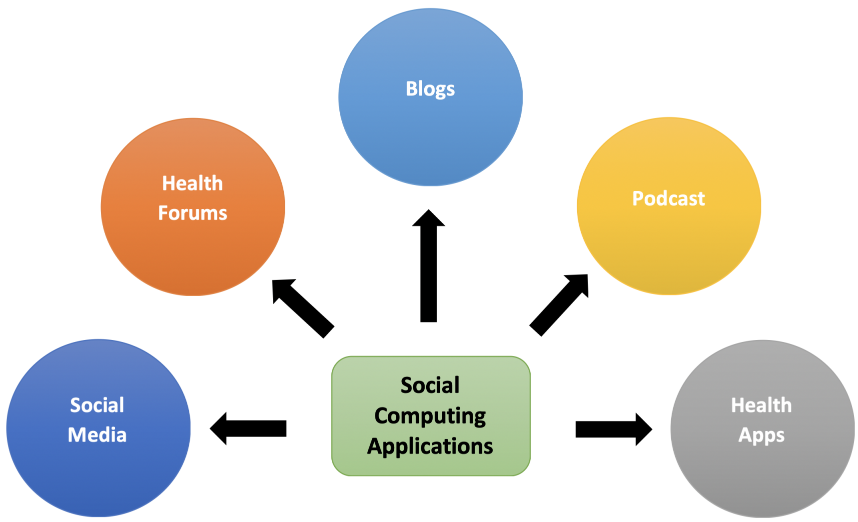

Figure 1.1: The examples of social computing applications for healthcare

Mobile health applications or mHealth apps are mobile communications technology that provides healthcare services. Smartphone devices become popular in healthcare settings, contributing to the rapid growth in the development of medical software applications for these platforms [7]. For example, for tracking the patient's health status, providing diagnostic treatment, clinical decision-making, communications and consulting, and enhancing medical education [8]. Thus, mHealth application usage can help prevent and manage chronic health conditions such as heart disease, diabetes, mental illnesses, neurological diseases, and obesity.

A blog is an online site with postings, date entries, and comments section for two-way interaction. The postings typically include text, audio, video, and link managed by content creators or blog owners. Blogs that run bu hospitals have also become part of the treatments. Patients can gain extra information from a blog to understand their disease and health tips to improve outcomes [9]. In turn, it helps the hospital improving its services from the given comments. A podcast is a digital audio content that can be download to a computer or smartphone. The podcast users gain information by listening to the recorded speeches. The research presented that health information delivered via podcast shows more significant positive effects weight loss journey to the participants [10]. It shows that the podcast's potential growth as a healthcare tool can be beneficial for users.

Social knowledge sharing or wikis is a collection of digital medical encyclopedias regarded as quality information [4]. The ability to offer reliable content and interaction is the main advantage of this approach. The wikis health-related contents are more trustworthy than others because each of the published articles come with references. The writing is also more transparent and systematic, adding several sub-sections containing a detailed explanation of the discussed topic. In conclusion, social computing as healthcare tools shows many positive impacts on target users in understanding diseases and health status. Hopefully, in the future, this technology has enhanced the ability to provides reliable virtual early diagnosis to the patients based on the reported symptoms. This thesis presented three approaches to identifying three different diseases via social media, health applications, and health records.

1.2 Deep Learning for Healthcare

DL is part of machine learning that involves multiple processing layers to learn data representations. It can be categorized as supervised, semi-supervised, and unsupervised learning, in which the models learned to do data classification and identification from the dataset during the experiment phase. For traditional machine learning approaches, the essential features will be identified by an expert to decrease the data complexion and make the extracted features easier for algorithms to learn for prediction. Figure 1.2 visualizes the differences of tasks involved in machine learning and DL. DL methods such as Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) are frequently used in many applications. These methods require a significant amount of data to perform well for classification. DL has shown promising results in Natural Language Processing (NLP) and RNN that can handle sequential data, including

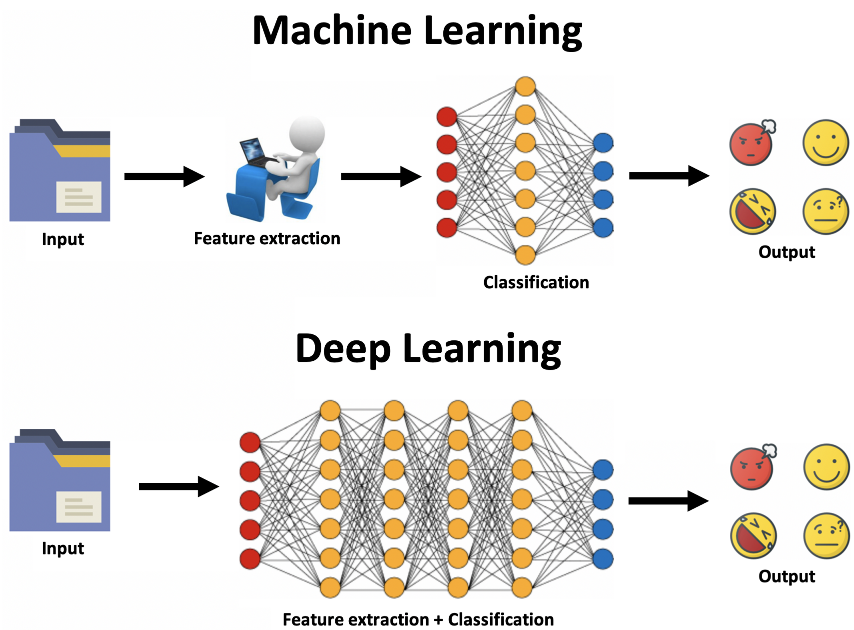

Figure 1.2: The overview of tasks involved in machine learning and DL

text and speech [11].

DL, also known as hierarchical learning, uses a layered algorithmic architecture to analyze data [12]. DL able to solves the problems that were unsolvable with traditional machine learning techniques. Data is filtered through multiple layers during training, with each layer using the output from the previous layer as an input to be processed. DL architecture is based on the way biological neurons connect to process information in the brains. In artificial neural networks, the basis for DL models, each layer may be assigned a specific portion of a transformation task, and input data travel through the layers several times to produce the optimal output. These "hidden" complex layers will perform the mathematical functions that turn raw input data into a meaningful output that can be understood by a human. DL can increase the computational work and provide accurate results in most cases [13]. The models can become more accurate as they process more data by learning from previous results to enhance their ability to make connections and correlations.

DL advanced technologies are revolutionizing various industries such as retail, finance, manufacturing, including healthcare. Medical experts are continually trying to implement new technologies to accurately describe the patient's symptoms, as health is a priority field. In healthcare, DL has increasingly become revolutionary for healthcare, offering the opportunity to interpret data more efficiently and precisely. DL could help medical professionals and researchers to discover hidden opportunities in data and better serve the healthcare industry. It provides doctors the analysis of any disease accurately and helps treat the patients, resulting in better medical decisions. The DL technology analyzes based on the patient's current symptoms and medical history to provides the best treatments. Moreover, this technology is gaining insights from patient symptoms/signs and tests. Healthcare organizations are highly interested in DL's ability to support better patient care while reducing costs and improving efficiencies.

1.3 Multiple Modalities Data in Healthcare

DL collects a large amount of data, including patient records and medical reports, and then uses its neural networks to process it. These health records are stores in multiple data formats, such as text, numbers, audio, and images. Applying DL technology to health data, hidden information, and clinical data patterns can be uncovered to help physicians better treat their patients. Besides being precise, the DL tools are also fast. DL is constantly finding its way into innovative devices that have practical applications in the real-world clinical environment. Figure 1.3 illustrates the various data-type in healthcare.

DL and neural networks are already building many NLP tools that have become popular for dictating documentation and translating speech-to-text [14]. Neural networks for NLP are designed commonly for classification, and they can identify individual linguistic or grammatical elements by "grouping" similar words together and mapping them to one another. DL helps the network understand the complex meaning of semantics. But the task is complicated by the variety of daily conversation and communication. For example, words that always appear next to each other in phrases daily may have meant something very different from when the same words appear in a different context. While the acceptably accurate speech-to-text capability of dictation tools has become relatively common, generating reliable and responsive insights into medical text data analysis is significantly more challenging. Unlike images of defined rows and columns of pixels, the text of clinical notes in Electronic Health Records (EHRs) is notoriously messy, incomplete, and

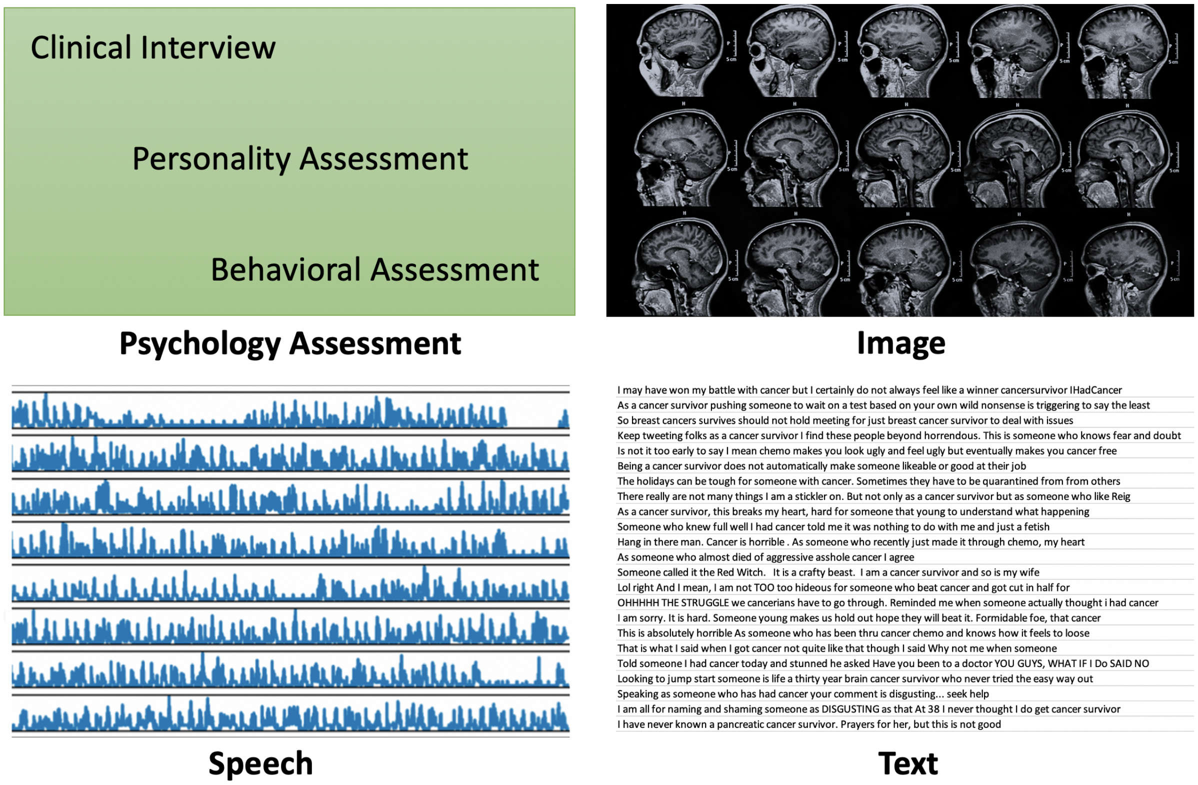

Figure 1.3: Multiple modalities data in healthcare

inconsistent.

Medical imaging, including Magnetic Resonance Imaging (MRI) scans, Computed Tomography (CT) scans, and Electrocardiography (ECG), are used to diagnose chronic diseases such as heart disease, cancer, and stroke. DL is particularly well-suited to analyzing medical images [15]. DL is designed with the assumption that they will be processing images, allowing the networks to operate more efficiently and handle larger images. The result indicates that the DL technique is approaching or even surpassing human diagnosis precision when identifying images' critical features. Thus, it helps doctors to analyze the disease better and provide patients with the best treatment.

Alzheimer's and Parkinson's are progressive diseases, where the symptoms gradually worsen over several years. In its early stages, memory loss is not critical, but individuals will lose the ability to continue a conversation and respond to their surroundings during late-stage diseases. These neurodegenerative diseases are incurable and debilitating conditions that result in progressive degeneration and death of nerve cells [16]. These most common neurodegenerative diseases are the most significant challenges that the medical industry faces because there is no cure for these diseases or a way to stop its progression [17]. The doctors have to keep monitoring the patients' progress in treating the symptoms. DL time-series technique can detect neurodegenerative disease at an early stage by analyzing the patients' records. This technique is also used to understand the condition and help patients living with the disease and their caregivers coping with symptoms and improve quality of life.

Researchers have confirmed that finding patterns among multi-modal data can increase the accuracy of diagnosis, prediction, and overall learning system performance. However, multi-modal learning is challenging due to the heterogeneity of the data [18]. Accessing enough high-quality data to train models accurately is also problematic. Data that is biased or skewed towards particular age groups, ethnicities, or other characteristics could create models that are not equipped to assess a wide variety of real-life patients in a real environment accurately. Still, DL represents the most promising pathway forward into reliable analytics applications. The DL tool improved the accuracy of traditional approaches for identifying unexpected hospital readmissions, predicting length of stay, and forecasting inpatient mortality [19]. This predictive performance was achieved without the manual selection of variables deemed necessary by an expert, similar to other DL applications to EHR data. Thus, the healthcare industry has high hopes for DL's role in clinical decision support and predictive analytics for a wide variety of conditions.

1.4 Dissertation Major Contributions

This dissertation addresses three primary scientific needs toward disease diagnosis. (1) How to employ the DL approach for text classification in identifying the individual with a history of cancer and currently struggles with negative emotions and mental illness from text posting on social media? (2) DL can learn complex structures from multiple inputs and outputs for time series forecasting purposes. How to remotely monitor PD patients' disease progression from speech information using the DL time series approach? (3) DL approach works very well in complex data such as NLP, image classification, and speech recognition but infrequently apply for multi-type

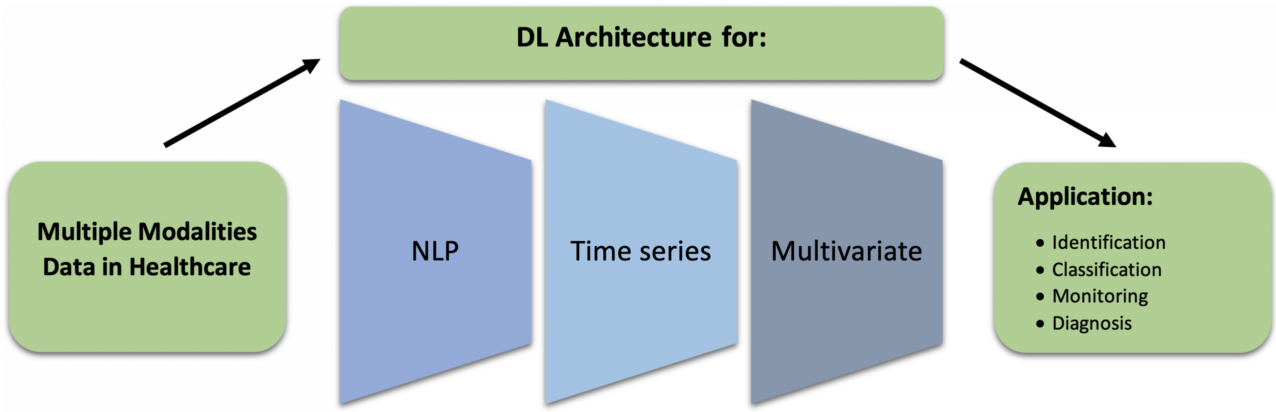

Figure 1.4: Summary of the contributions of this dissertation

data. How to perform effective and efficient DL in identifying dementia in PD patients?

Figure 1.4 illustrates the significant contributions of this dissertation. We propose a series of DL network settings and frameworks to model real-world healthcare information for different diseases. Using DL, we developed a classification system for healthcare and the patients to monitor progression of health status and identify the disease's new symptoms. The developed models would be a guideline for developing other healthcare diagnosis applications. Our key contributions can be summarized as follows:

- From text information, we proposed a framework to identify specific cancer survivors that have PTSD on Twitter. We formally define the problem of data crawling and filtering techniques to obtained tweets that represent the target group. We stream the data using cancer as a keyword to filter the tweets with cancer-free and use PTSD related keywords to reduce the time spent on the annotation task. The proposed CNN learns the input representations to identify cancer survivors with PTSD.
- For time series speech data, we aim to provide a monitoring system for PD patients and clinicians to observe the symptoms' progression remotely. We developed a framework of multivariate multi-step time series approach to monitoring the voice impairment in different future time-frame based on voice features. We develop multivariate multi-step time series forecasting model with observed multiple input variables to forecast multiple future time

steps. We construct a multi-channel CNN network setting for time series to be applied to the PD speech telemonitoring dataset.

- Using multivariate healthcare data, we propose a framework in detecting dementia among untreated PD patients that were diagnosed for two years or less to decrease the PD-dementia risk. We design a DL architecture to identify the dementia symptoms from non-motor information. The Montreal Cognitive Assessment (MoCA) total scores have been used as a scale in classifying the PD-Dementia status. The trained model can be used as an alarm tool for detecting PD-Dementia.
- We evaluate the effectiveness and efficiency of the proposed DL architectures of text classification, time series, and multivariate data carefully.

1.5 Dissertation Organization

This dissertation is organized into five chapters. This dissertation aims to develop a series of DL frameworks in handling multiple modalities data in healthcare to present alternative options for diagnosis tools within different diseases. In Chapter 2, we used text postings from Twitter to identify cancer survivors with PTSD on social media. In Chapter 3, we used PD patients' telemonitoring speech data to forecast future disease progression and applied the DL time series approach. In Chapter 4, we propose work in identifying dementia status in PD patients. In Chapter 5, we conclude the dissertation and present several potential topics as future work.

2. TEXT CLASSIFICATION USING A DEEP LEARNING APPROACH FOR IDENTIFYING CANCER SURVIVORS LIVING WITH POST-TRAUMATIC STRESS DISORDER ON TWITTER[1]

2.1 Introduction

PTSD is a psychological disorder that occurs in some people after witnessing or experiencing traumatic events [20]. People who have suffered from war, a severe accident, a natural disaster, a sexual assault, and medical trauma are potentially at risk of developing PTSD. Almost half of the cancer fighters are diagnosed with a psychiatric disorder, with the majority of them having chronic depression [21]. Cancer diagnosis, treatments (chemotherapy and radiation), post-treatment care, and recovery could affect the patients' psychological condition and cause anxiety or trauma. Unstable mental health among cancer survivors is hazardous because they are at high risk of self-destruction and may also harm others once they lose self-control of their behaviors [22].

The diagnostic procedure for mental illness is different from physical illnesses. Traditional mental illness diagnosis begins with patients' self-reporting about unusual feelings and caregivers' perception of the patients' behavior to the doctor. To diagnose a patient, a doctor will conduct a physical examination, order lab tests, and perform a psychological evaluation that requires a period of observation of the symptoms. The psychological assessment will be conducted by a psychiatrist who has an extensive breadth of knowledge and experience not only in mental health but also in general medicine. This process of making a diagnosis is not easy, and it takes a lot of time and effort to find effective treatments. Thus, in this work, we want to capture the presence of PTSD symptoms in cancer survivors from online social media postings so they can have an early meeting with a doctor and receive immediate treatment to calm the stress.

Information about the user's online activities has been used to identify several health problems

[1]Part of this chapter is reprinted with permission from "Using Deep Neural Network to Identify Cancer Survivors Living with Post-Traumatic Stress Disorder on Social Media" by Nur Hafieza Ismail, Ninghao Liu, Mengnan Du, Zhe He and Xia Hu, The 5th International Workshop on Semantics-Powered Data Mining and Analytics (SEPDA 2019), Paper 12.

in the previous work [23]. The growth of social media sites in recent years has made it a promising information source for investigating issues on mental health. For example, Twitter has a large user base with hundreds of millions of active users [24]. It has simple features that allow users to share their daily thoughts and feelings [25]. The online activities, especially postings on the timeline, may present an insight into emotional crash towards significant incidents that happened in life. Related studies have shown the potential of Twitter for detecting the early symptoms of mental illness [26].

Previous works in mental illness on social media aimed to examine the attitude of self-declared mentally ill patients based on their interactions with others and social aspects from their written comments and postings [27, 28]. The study conducted by De Choudhury et al. [26] used crowd-sourcing to access Twitter users who have been diagnosed with major depressive disorder by a psychiatrist. The Linguistic Inquiry Word Count (LIWC) was used to characterize linguistic styles in tweets. Most of the previous studies only focus on identifying mental illness in social media users generally. In addition, some experimental procedures such as crowd-sourcing, Twitter Fire-house, manual labeling, and LIWC are expensive and time-consuming. Data collection, data pre-processing, and analysis are challenging due to the following reasons. First, there are no available techniques that can verify if a tweet contains elements about both cancer-free and PTSD. Second, the extracted information related to mental health is not fully utilized in developing psychological screening tools for cancer survivors.

To tackle these challenges, we propose a technique to classify cancer survivor and PTSD related tweets. To identify PTSD in cancer survivors, we first crawl the tweets using "cancer" as a keyword. After that, we use a set of cancer survivor and PTSD keywords to filter out irrelevant tweets, which can reduce the time required for manual labeling. To create a ground truth dataset for this work, we make an effort to check the extracted tweets again manually. The primary purpose of the manual checkup is to make sure that the extracted tweets are correctly labeled as to whether the tweet contained a genuine statement of a cancer survivor with PTSD diagnosis. In this work, we used the Deep Neural Network (DNN) approach that learns to extract meaningful representations of texts and identify key features from the input dataset. Specifically, we present a framework that can automatically identify PTSD from cancer survivors based on their tweets. The major contributions of this work can be summarized as follows:

- We formally define the problem of data crawling and extracting techniques for retrieving the tweets that represent the cancer survivors with PTSD.
- We present a framework and training the proposed CNN to identify cancer survivors living with PTSD based on phrases on Twitter.
- We evaluate the model's prediction performance by producing a label with associated probability for new tweets.

2.2 Related Work

Researchers from diverse backgrounds, such as psychology and medical informatics, have proposed early models for detecting mental health issues. They explored different types of dataset, feature extraction approaches, and modeling methods to develop a reliable model. The physiological features, such as facial expression, vocal acoustic, blood flow, and nervous system responses can indicate the presence of a person's current emotions [29]. Various sensor measurements in medical examination results such as ECG, Electroencephalography (EEG), Electromyography (EMG), functional Magnetic Resonance Imaging (fMRI), and respiratory transducer have been used to identify emotional changes in PTSD diagnosis [30]. Besides, some experts also consider speech audio, interview video, and questionnaire, in both formal and informal ways [31, 32]. Nevertheless, collecting this information with these techniques is time-consuming and labor-intensive.

The alternative approach is to crawl the public online postings on social media, which are accessible, expeditious, and provides boundless access to a broader population. Almost 60% of adults use online resources for searching and sharing information about health [33]. Compared to asking doctors or friends, people feel more open to communicate and ask questions on social media. They can also have a conversation with people with a similar background and those who are currently facing the same health concerns on the forums. Previous work has shown that text posts, votes, and comments on Reddit, a popular online discussion board, can reveal early symptoms of mental health conditions [33]. In a previous study on public health, Paul and Dredze [25] showed that Twitter has a capability to show the linguistic style of the users from their tweets. These previous studies have motivated us to use Twitter data to grasp the implicit and explicit information behind the language used by PTSD patients with a cancer history.

Cured cancer patients are often concerned about cancer recurrence, which can be even more stressful and upsetting compared to first time diagnosis [34]. Patients reported that it is harder to decide the treatment, the side effects are more serious, and the fears of pain increase [35]. This psychological impact that may lead to PTSD problems is one of the most significant concerns in clinical oncology [36]. Receiving immediate attention to PTSD can help to improve the quality of life. Nevertheless, the lack of quantifiable data for PTSD is one of the main obstacles for making reliable diagnoses and providing effective treatment [37]. These issues have been our second research motivation to collect data for cancer survivors living with PTSD.

There are several techniques applied to uncover essential features from mental health datasets. Commonly, medical experts who conduct similar research analyze the collected dataset using statistical methods, such as t-test, chi-square tests, correlations, linear regression, and logistic regression [38, 39, 40]. The dataset for mental health is gathered using a questionnaire to collect sociodemographic information, clinical variables, medical comorbidity, and self-reported depression to identify mental illness signs or symptoms. The analyses report the characteristics of each item in the percentage or scale value. From there, they can identify the most correlated factors for mental health diagnosis.

Numerous analytical methods and techniques, including supervised and unsupervised learning algorithms, have been applied for monitoring mental health symptoms. Regression analysis and the Support Vector Machine (SVM) [41, 42, 43], decision tree, and ANN [44] performed well with a high diagnostic accuracy. For the unsupervised models, a linear discriminant analysis model can generate the topics found in engagement content on social media to investigate the engagement implication on mentally ill people [45]. DNN, a deep belief network model, was trained to extract

PTSD features from a speech dataset using a transfer learning approach [46]. The DNN has shown promising results in NLP. In this work, we developed a model using the DNN approach that can learn from different levels of representation of text input. This approach can learn from the input data and has been used widely to make predictions in various areas of automatic speech recognition, image recognition, and NLP. DNN automatically learns the representations from the input data and uses them for classification [47]. In comparison, traditional machine learning requires labor-intensive feature engineering that may result in a biased set of features.

2.3 Methods

In this section, we will introduce the problem statements and the proposed framework, including feature extraction, knowledge transfer, and CNN architecture. Then, in the experiments section, we will explain the data preparation process.

2.4 Problem Statement

We present the problem and goal of our proposed work in detail below:

- Problem: We consider a relation exists in n tweets with m characteristics of cancer survivors living with PTSD. Each relation between a tweet t_i and characteristics p_j is represented as e_{ij} = (t_i, p_j). In particular, in our setting, a relation is composed of textual information related to the tweet t_i with characteristics of cancer survivors living with PTSD p_j. Also, we assume that each tweet is associated with a label $L(t_i)$ = 1 if the tweet belongs to cancer survivor living with PTSD, otherwise $L(t_i)$ = 0. In this work, we will use italic characters x for scalars, bold characters **h** for vectors, and bold capital characters **W** for matrices.

- Goal: We aim to actively explore the cancer survivors living with PTSD on Twitter. In particular, given a tweet containing characteristics of cancer survivors living with PTSD E = $\{e_{ij}=(t_i,\ p_j)\}$, our goal is to produce a prediction $\widehat{L}(t_i) \in [0,\ 1]$ for each tweet and its probability score s_i.

2.5 The Proposed Framework for Classifying Tweets about Cancer Survivors Living with PTSD

Figure 2.1 presents our proposed framework on classifying tweets about cancer survivors living with PTSD using CNN model. It involves two central parts. First, we extract a set of particular lexicons that are frequently mentioned by sufferers from previous studies on depression, which relates to PTSD. Second, the extracted lexicons are then used to capture tweets that contain PTSD symptoms in the cancer survivors dataset. The detailed process of our proposed framework will be explained in three subsections: (1) feature extraction, (2) knowledge transfer, and (3) CNN architecture.

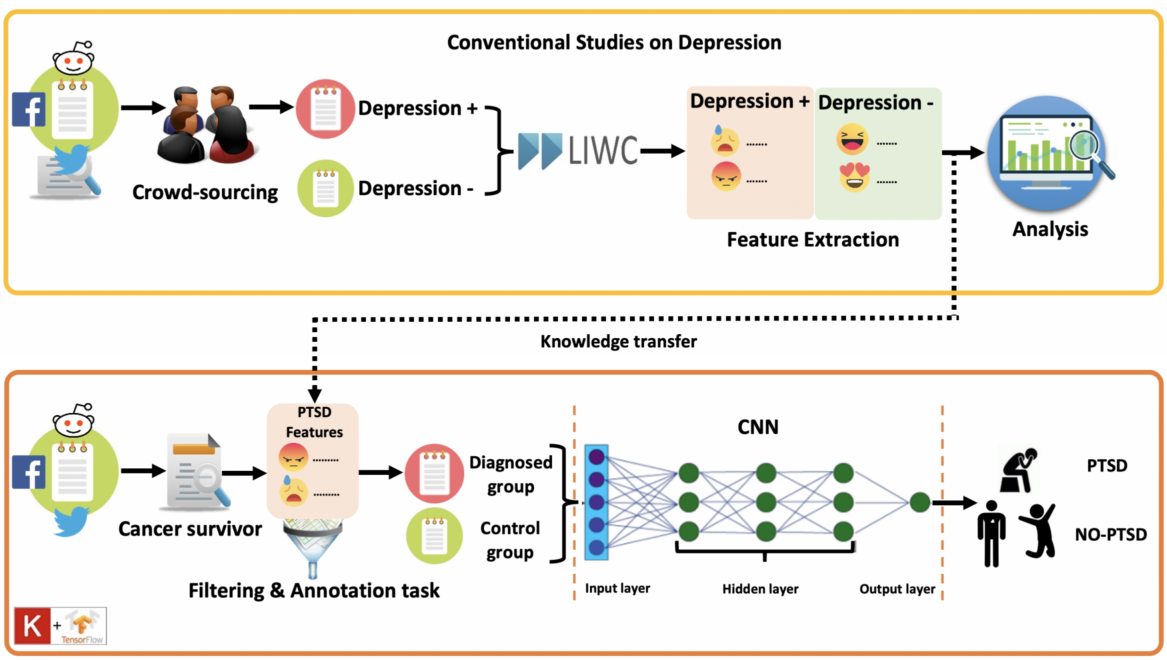

Figure 2.1: The overview of our proposed framework for classifying tweets about cancer survivors living with PTSD using CNN model

2.5.1 Feature Extraction

Feature extraction, also known as variable selection, aims to discover a small amount of valuable information that can best represent the whole large dataset. This process requires specific methods of extraction to be applied to the input data to create an accurate prediction model. The top section of Figure 2.1 shows the overview of previous studies related to predicting depression on social media [26]. The crowd-sourcing approach has been employed to identify written postings, whether they are depression positive or depression negative.

Next, they used LIWC, a text analysis tool, to perceive the characteristics of linguistic style in both groups. It employed to count the psychological expression lexicon in the tweets and assess the proportions of words used in several linguistic categories. The given comprehensive list produced by the tool will automatically present the most frequently used lexicon by depressed people, and statistical methods are applied to visualize the analysis results. Table 2.1 shows the depression lexicon.

2.5.2 Knowledge Transfer

In this part, knowledge transfer can be interpreted as a task that uses depression lexicon in developing our PTSD positive dataset. This approach is similar to the transfer learning method in which the pre-trained models are used to reduce training time and to increase the performance of the model. Depression and PTSD often co-occur. Almost all PTSD patients also have a presence of depression in clinical and epidemiological samples. This co-occurrence reflects overlapping symptoms in both types of mental disorders [48]. The word "cancer" is strongly correlated to negative emotions such as mortality, fear, and stigma [49]. The definition of PTSD in our context is a failure to recover from a traumatic event of cancer. Thus, any expression of negative sentiment related to cancer in a tweet posted by cancer survivor is considered as PTSD. Even though there is no existing PTSD lexicon available, we could use the depression lexicon as a proxy to remove irrelevant tweets. We first used the depression lexicon to labeled our ground truth. To ensure labelling accuracy, we manually reviewed these tweets to make sure they indeed represent cancer survivors with PTSD symptoms. Thus, we opted to utilize the depression lexicon taken from previous work to identify PTSD tweets.

The lower section of Figure 2.1 presents our proposed framework to identify cancer survivors with PTSD. We crawled the raw dataset using "cancer" as a keyword through Twitter's Application

Programming Interface (API) in a period of three months from August 2019 - October 2019. We conducted the extraction process in two steps using two sets of keywords. First, we created the cancer survivor dataset using related hash-tags and terms such as "cancer survivor", "cancer-free", "I had cancer", "post-cancer", "survive from cancer", and "free from cancer". Second, we used the depression features from Table 2.1 to filter out tweets that are unrelated to PTSD signals. Next, in the annotation task, we checked the tweets manually to make sure the extracted tweets are correctly identified. The extraction process helped us save lots of time for the annotation task. The total data has decreased from 900,000 to only 5,000 after we conducted the extraction process and the annotation task. Also, we added the word "PTSD" in the Symptoms category to capture the word PTSD in tweets. Next, the extracted tweets were fed into CNN algorithms in the modeling phase.

Table 2.1: The depression lexicon

Category	**Unigrams**
Symptoms	anxiety, withdrawal, severe, delusions, adhd, weight, insomnia, drowsiness, suicidal,appetite, dizziness, nausea, episodes, attacks, sleep, seizures, addictive, weaned, swings, dysfunction, blurred, irritability, headache, fatigue, imbalance, nervousness, psychosis, drowsy, PTSD
Disclosure	fun, play, helped, god, answer, wants, leave, beautiful, suffer, sorry, tolerance, agree,hate, helpful, haha, enjoy, social, talk, save, win, care, love, like, hold, cope, amazing, discuss
Treatment	medication, side-effects, doctor, doses, effective, prescribed, therapy, inhibitor,stimulant, antidepressant, patients, neurotransmitters, prescriptions, psychotherapy, diagnosis, clinical, pills, chemical, counteract, toxicity, hospitalization, sedative, drugs
Relationship and life	home, woman, she, him, girl, game, men, friends, sexual, boy, someone, movie, favorite, jesus, house, music, religion, her, songs, party, bible, relationship, hell, young, style,church, lord, father, season, heaven, dating

2.5.3 CNN Architecture

The architecture of our proposed CNN model is inspired by [50] for sentiment analysis of the text. We adopted one convolutional layer during network configuration for cancer survivors with PTSD tweets classification, as displayed in Figure 2.2 adapted from [1]. We trained the CNN with the embedding layer. It requires specifying the vocabulary size, the size of the real-valued vector space, and the maximum length of words in input tweets. For convolutional feature maps, we used word embedding with 200-dimension for text representation. Thirty-two filters were applied by referring to the conservative setting for word processing, with a kernel size of 8, and with a rectified linear unit (ReLU) activation function. Followed by a pooling layer, the filters will generate feature maps and reduce the output by half. The last layer uses a sigmoid activation function to output a boolean, i.e., positive and negative, in the tweets based on the concatenation of the previous vectors. Then, the extracted model is saved for later evaluation. The following subsections present the critical elements involved during network configuration.

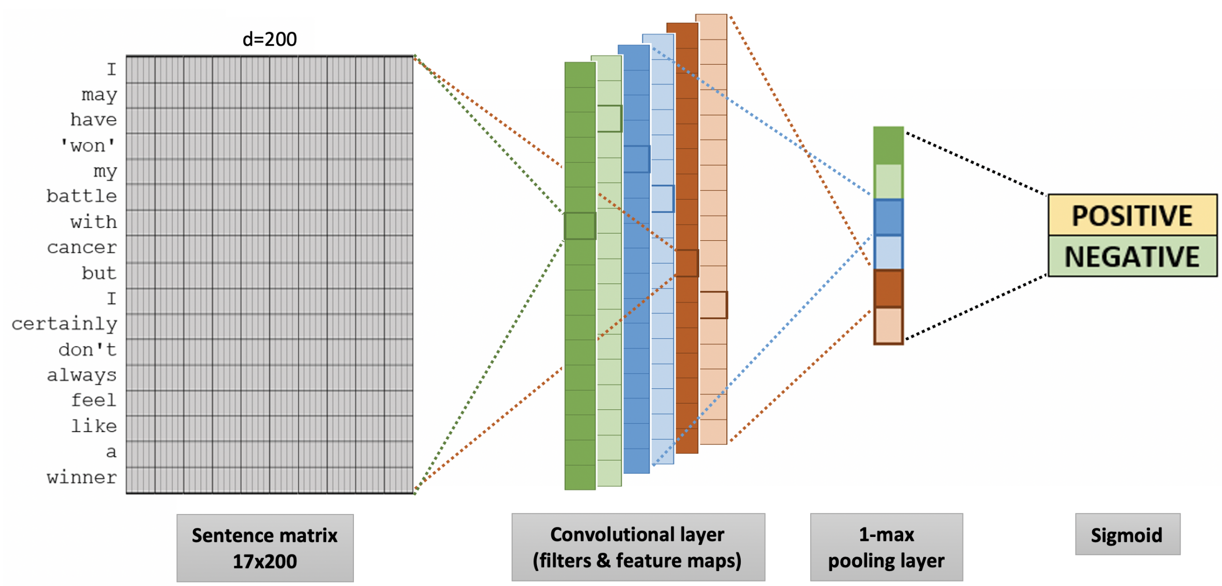

Figure 2.2: The CNN architecture to classify tweets posted by cancer survivors living with PTSD adapted from [1]

2.5.3.1 *Text Representation*

After the data cleaning process, we applied the embedding layer that is initialized with random weights. It learned an embedding for all of the words in the training dataset. The first step embeds the vocabulary file V to check the validity of the tokens in tweets. Each input tweet is presented as a sequence of individual word tokens: $[t_1, \ldots, t_n]$ where n denotes the total number of tokens in the tweet. Tokens are represented by one-hot vectors $\mathbf{t} \in R^{1\times d}$ to look up word embeddings $\mathbf{T} \in R^{d\times|V|}$. For every input tweet s, we built a string of words matrix $\mathbf{S} \in R^{d\times|s|}$, where every single column i represents a word embedding $\mathbf{t}_i$ of position i in a string. The CNN applies multiple configurations to the input string of words matrix $\mathbf{S}$ using convolution, non-linear activation, and pooling operations. It learns how to capture and to reconstruct features of individual tokens in a given tweet from word embeddings into higher-level concepts.

2.5.3.2 *Convolutional Feature Maps*

The purpose of convolutional layer is to extract meaningful patterns from the input dataset using a number of filters. During convolutional operation, the input matrix $\mathbf{s} \in R^{1\times|s|}$ and a filter $\mathbf{F} \in R^{d\times m}$ of the same dimensionality d with width m will produce a new vector of $\mathbf{c} \in R^{|s|+m1}$, where each function is computed as follows:

$$\mathbf{c}_i = (\mathbf{S} * \mathbf{F})_i = \sum_{k,j} (\mathbf{S}_{[:i-m+1:i]} \otimes \mathbf{F})_{kj}, \tag{2.1}$$

where $\otimes$ is the element-wise multiplication and $\mathbf{S}_{[:i-m+1:i]}$ is a matrix slice with m size along with the columns. From the Figure 2.3, we can see that the filter overlays across the row vectors in the dimension table of S, producing a vector $\mathbf{c} \in R^{|s|-m+1}$ as the output. Each component c_i is the result of computing an element-wise product between a row slice of $\mathbf{S}$ and a filter matrix $\mathbf{F}$, which is then summed up to obtain a single value. To grab more features and to form richer representation from the dataset, a series of filter $\mathbf{F} \in R^{n\times d\times m}$ overlays the sentence matrix $\mathbf{S}$ and produces a feature map matrix $\mathbf{C} \in R^{n\times|s|-m+1}$.

2.5.3.3 Activation Functions

After the convolution step, we applied ReLU activation defined as $max(0, \mathbf{x})$, which is the simplest non-linear activation function $\alpha()$ on the hidden layers. It has a lot of advantages. For example, it can generate a good result in a short time by reducing the training time for large networks.

2.5.3.4 Pooling

The output from the convolutional layer with ReLU activation function will be passed to the pooling layer. The goal of pooling is to control the overfitting by combining the information and reducing the spatial size of the representation. In our model, we use max pooling to get the maximum value. It operates on columns of the feature map matrix C and returns the largest value: pool($\mathbf{c}_i$): $R^{|s|+mn-1} \rightarrow R$.

The convolutional layer utilizes the activation function, and the pooling layer acts as a non-linear feature extractor. Given that multiple feature maps are used in parallel to process the input, CNN can build rich feature representations of the data. The output of the convolutional and pooling layers are passed to a fully connected sigmoid layer. The main reason for using a sigmoid function is that it pushes the output to be between 0 and 1. Since the likelihood of any class exists only between the range of 0 and 1, sigmoid is appropriate for this setting.

2.6 Experiment

We conducted the experiments to evaluate the proposed framework for classifying cancer survivor with PTSD diagnosis from tweets. First, we briefly describe the experiment setting and the dataset preparation process. Second, we introduce the baselines methods. Third, we report the experimental performances. Finally, we discuss our findings.

2.6.1 Experiment Settings

In these experiments, the dataset with PTSD positive represents the diagnosed group, while PTSD negative represents the control group. For the diagnosed group, we retrieved tweets from users who publicly stated that they survived cancer and had PTSD symptoms. To construct the PTSD negative group, we mixed the tweets posted by cancer survivors with positive sentiment and tweets from the Kaggle dataset. We made use of tweets from the "Twitter User Gender Classification" dataset from the Kaggle website [2].

We used this dataset because we want to make sure that the PTSD negative dataset not only contains about cancer survivors with positive sentiment tweets but also other topics. Both groups have the same total number of five thousand tweets to create balanced datasets. The data preparation phase has three steps: (1) applying 5-fold cross-validation for MLP, CNN, and CNN n-gram algorithms; applying Term Frequency–Inverse Document Frequency (TD-IDF) for naive bayes and SVM algorithms; (2) cleaning the dataset to remove punctuation, stop words, and numbers; (3) defining a vocabulary of preferred words from a training dataset by stepping through words and keeping only tokens with minimum occurrences of five. This setting reduces the vocabulary size because we want to use only frequent tokens that appear in the dataset. We used Keras API running on Tensorflow to train DNN models. All the models were trained with ten epochs through the training data. The efficient Adam implementation of stochastic gradient descent was used. We keep track of performance in addition to loss during training. Table 2.2 shows the details of our CNN network setting.

Table 2.2: CNN network setting

Layer(type)	**Output Shape**	**Param #**
embedding_1&_2 (Embedding)	(None, 20, 200)	84000
conv1d_1 (Conv1D)	(None, 13, 32)	51232
max_pooling1d_1&_2 (MaxPooling1D)	(None, 6, 32)	0
flatten_1&_2 (Flatten)	(None, 192)	0
dense_1 (Dense)	(None, 10)	1930
dense_1 (Dense)	(None, 1)	11

[2] https://www.kaggle.com/crowdflower/twitter-user-gender-classification#gender-classifier-DFE-791531.csv

2.6.2 Baseline Methods

We present baseline methods used for evaluate our proposed algorithm. The input of our dataset was in a text format with positive and negative labels. Therefore, we chose four baselines that are capable of handling text dataset: naive bayes [51], SVM [52], MLP [53], and CNN n-gram [54]. Naive bayes and SVM are considered as traditional machine learning algorithms. While MLP, CNN, and CNN n-gram are the DL algorithms.

2.6.2.1 Naive Bayes

Naive bayes is based on the Bayes Theorem. For text classification, it will predict the membership probabilities for each class label, such as the probability that tweet belongs to a particular class label. The chosen class will have the highest probability value compared to other classes.

2.6.2.2 SVM

SVM is an algorithm that determines the best boundary between vectors that belong to a given group label and vectors that do not belong to the group. This technique can be applied to any vectors that encoded any data. Thus, for SVM text classification, we first must transform the texts into vectors.

2.6.2.3 MLP

The MLP is a feed-forward neural network that is frequently used for prediction models. The MLP used Bag-of-Words (BoW) to represent tweets. This technique can extract features from the text by measuring the occurrence of words within the documents. However, the BoW model suffers from sparse representation, which may have effects the space and time complexity. Moreover, it looses semantics of the input sentences by ignoring the word order and grammar.

2.6.2.4 CNN n-gram

The kernel size in convolutional layer defines the number of tokens that act as a group of the parameters. We set a model with two input channels for processing bi-grams and tri-grams of text in tweets due to the short length of words used in each tweet. This algorithm involves using multiple versions of the standard model with differently sized kernels for tweet classification. This setting allows tweets to be processed at different number of contiguous words sequence, while the model learns how to integrate these interpretations best. The output from both channels was concatenated into a single vector and processed by a dense layer and an output layer.

2.6.3 Experimental Results

We ran the experiments using five different network settings. Our results indicate that CNN can effectively identify cancer survivor with PTSD. Experimental results in Table 2.3 show the 91.29% accuracy for CNN, which is higher than other baselines. We ran the experiments multiple times for MLP, CNN, and CNN n-gram algorithms due to the stochastic nature of DNN to get the reasonably accurate result.

Table 2.3: Experiment results of identifying cancer survivors with PTSD

Data Setting	**Method**	**Accuracy (%)**
TD-IDF	Naive bayes	86.50
	SVM	49.00
	MLP	49.99
5-Fold Cross Validation	CNN n-gram	63.28
	CNN	**91.29**

Figure 2.3 presents the time taken during the DNN training process with MLP and CNN n-gram, which took slightly less time compared to CNN. Figure 2.4 shows the loss values in the training set of all models, where CNN and CNN n-gram display low losses. A model with the lowest loss value is better because loss value indicates errors made for examples during training. To test the CNN performance, we ran the experiment using only depression-lexicon as features. The experiment result is much worse, with 67.03% accuracy compared to our model. The results show that the model performed better with our set of vocabulary compared to a set of depression-lexicon taken from previous work. Even though we used depression-lexicon to help us to filter out unrelated tweets; however, our cancer survivor and PTSD tweets still contained unique characteristics and have different linguistic-style compared to depression users.

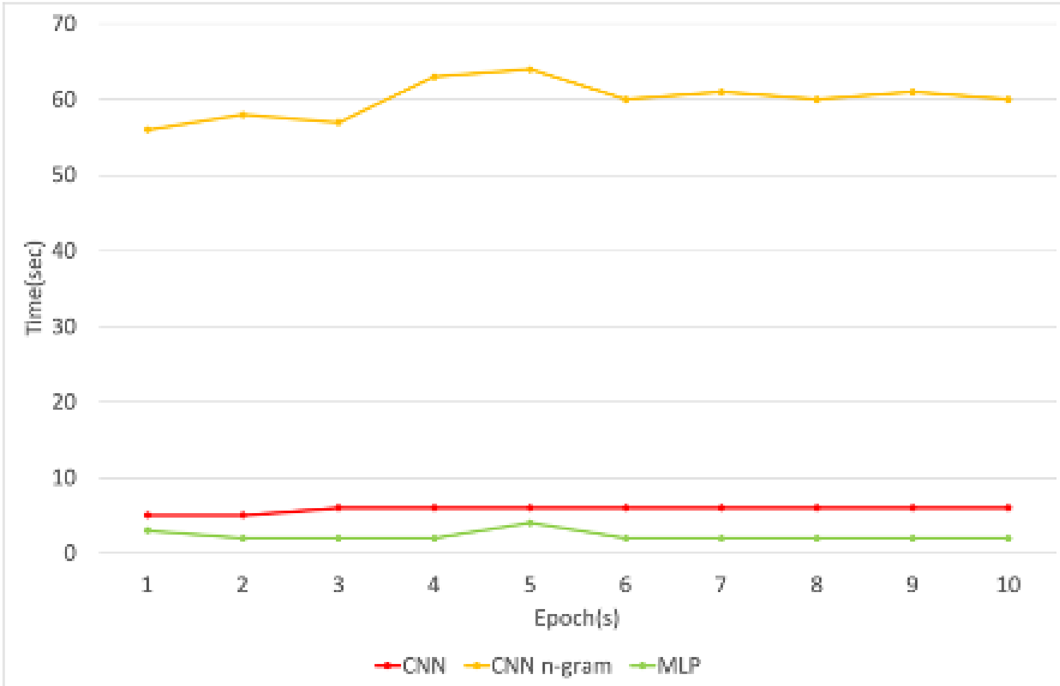

Figure 2.3: The learning time taken

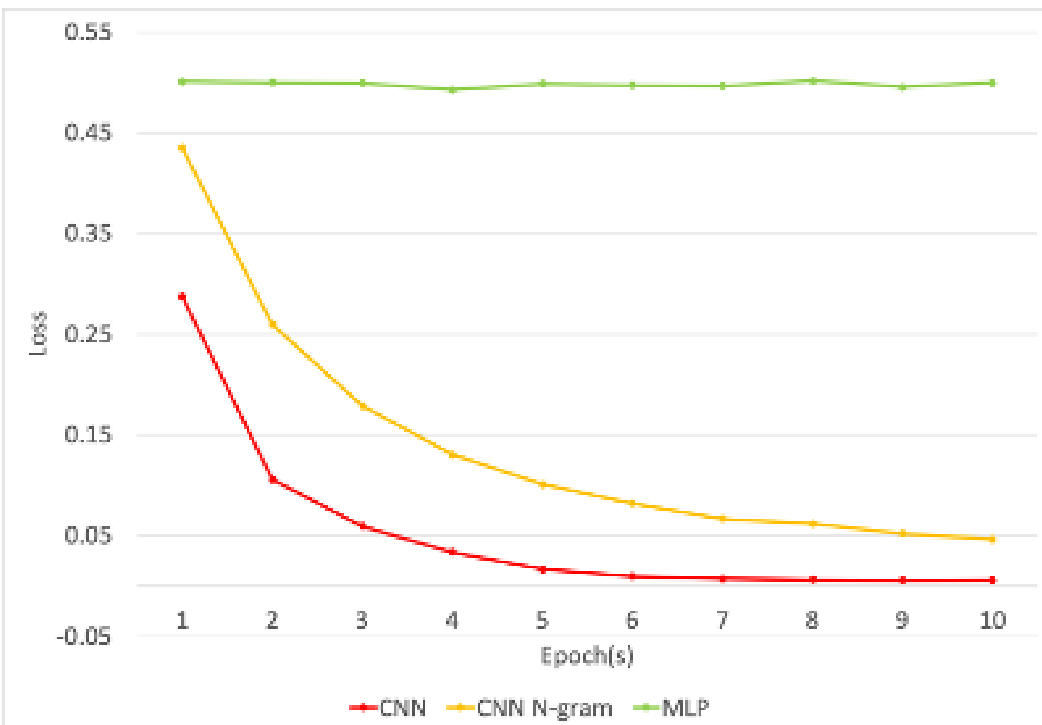

Figure 2.4: The loss values

2.7 Case Study

We constructed a simple prediction system using the CNN model. It will identify cancer survivors as either PTSD positive or PTSD negative together with probability value on new tweets. The tweet samples and the output results are shown in Table 2.4. Surprisingly, the system was able to classify tweets correctly. For example, the second tweet is a statement that consists of negative sentiment but not related to the cancer survivor, and the system classified it as a PTSD negative. The right labeling with high probability value is essential for the diagnosis.

Table 2.4: Predictions for new tweets

Tweet: "I have had more difficulty post cancer than during my active treatment. To me it is a neverending path (hate the word journey)." **POSITIVE (100.000%)**
Tweet: "I hate myself, I don't feel like living anymore." **NEGATIVE (99.831%)**
Tweet: "I got a gold MacBook that only use for music and homework. Still keep it in apple box." **NEGATIVE (99.949%)**

Meanwhile, Table 2.5 shows two examples of misclassified tweets. To test the model reliability, we replaced the word "cancer" with "tumor" and "cyst", which are highly correlated to cancer. Unfortunately, the model failed to detect the presence of cancer-free and PTSD in both tweets. For example, the first tweet contains self-mention about having a bladder tumor and feel depressed, but our system classified it as a PTSD positive, which is wrong. This is because our model should detect the presence of PTSD symptoms in someone who is currently free from cancer. However, the prediction outcome that has a probability rate lower than 90% thus is less convincing and can be ignored for diagnosis.

Misclassified tweets may occur due to several reasons. First, it may be because the words "tumor" and "cyst" occasionally appeared in the dataset. Second, a small number of participants from this group were active in social media. To alleviate this problem, we need a larger dataset for

Table 2.5: Predictions for new tweets (misclassified)

Tweet: "I have bladder tumor. I am totally heartbroken." **POSITIVE (86.548%)**
Tweet: "I am not a superwoman but I survived this pancreatic cyst. Time to enjoy with my family again!" **POSITIVE (86.548%)**

training to leverage cancer-free with PTSD lexicon. Moreover, our model also should contain the information of diverse cancer types so the system will able to recognize them as a part of cancer rather than treating them as unknown words. To get special insights, we should make an effort to gather data from multiple sources. To the best of our knowledge, this is the first work that deployed the extracted model of cancer survivors living with PTSD into a prediction system that is capable of evaluating new tweets. The experimental results showed a high potential of a low-cost text classification technique that can be directly applied to other medical conditions that might affect patients' mental health.

2.8 Discussion and Conclusion

PTSD is one of the severe anxiety disorders that could affect individuals who are exposed to traumatic events, including cancer. Cancer survivors are at risk of short-term or long-term effects on physical and psycho-social well-being. Therefore, the evaluation and treatment of PTSD are essential parts of cancer survivorship care. In this work, we demonstrated that Twitter could be used to identify PTSD among internet users who had cancer. We propose a prediction model that can produce promising results in cancer survivors with PTSD diagnosis. Experimental results demonstrated that CNN is capable of capturing important signals from texts. The social media users with cancer history who suffer from PTSD will benefit from the prediction system. It will act as an alarming system by detecting the PTSD presence based on users' postings.

Essentially, we hope that our proposed data collection approach can facilitate current trauma screening questionnaire-based methods instead of replacing them. With the high rise of social media and a massive number of active users around the world, we hope to encourage more untreated cancer survivors that affected by PTSD to seek medical attention immediately. Moreover, the World Health Organization (WHO) stated that psychological disorder is the second largest of disability in the world population. However, only 10% of them obtained proper treatment.

Furthermore, we identify a cancer survivor who experienced PTSD only with one tweet. In this work, we did not use historical tweets because cancer is so daunting that some of the cancer survivors are even afraid to say “the C word” [55]. Many aspects of cancer events can lead to PTSD, such as various diagnostic testing, stressful waiting periods, the moment of bad news, and the painful treatments. For cancer survivors, PTSD can be triggered by continuous monitoring, follow-up visits, sudden physical pain, death of a public figure due to cancer, and fear of cancer recurrence. The traumatic event of cancer might not be as clear as a life-threatening car crash, but it can completely change someone’s life. They may feel grief for possible lost future opportunities and may impact self-esteem because of disfigurements due to their disease. Because of that, we can spot tweets with negative sentiment related to cancer history when they express saddens, fear, stress, and enraged in their posting. Moreover, from our experience, when we went through their timeline, we noticed that they do not always express how they feel every day. This situation has made it hard for us to identify PTSD after cancer cases using historical tweets.

On the other hand, our model was trained to solely utilize the textual postings. The users’ contextual information, such as gender, ages, etc., is not considered in this work. To better improve our model in the future, additional main keywords that represent “cancer-free” such as “cyst” and “malignant tumor” should be included during data crawling. From the case study, we can conclude that our proposed model cannot provide the right diagnosis when we replaced the word ‘cancer’ with “cyst” and “tumor” in the sentence. It is important because those words are highly correlated with “cancer”. Hence, we also want to identify developing conditions such as suicidal ideation and the side effect of PTSD treatment. Besides, we plan to explore another modality in uncovering PTSD indicators such as audio, image, or combination of both, for better diagnosis.

PTSD can also affect cancer survivors’ caregivers. Witnessing a loved one having cancer and watching the little one in pain are traumatic events that caregivers have to face. The Cancer.Net website reported that almost 20% of families of childhood cancer survivors had a parent who was suffering from PTSD. They also found that this anxiety disorder is common among parents of children receiving cancer treatment to develop PTSD symptoms. Thus, we believe that our work also can be utilized to identify PTSD in cancer survivors' caregivers. However, we must formally define the problem and identify the implicit and explicit characteristics of caregivers because some of them may have a difficult time admitting they are depressed.

3. MULTIVARIATE MULTI-STEP DEEP LEARNING TIME SERIES APPROACH IN FORECASTING PARKINSON'S DISEASE FUTURE SEVERITY PROGRESSION[1]

3.1 Introduction

PD is a long-term, progressive, and incurable neurodegenerative disorder that slowly damaging the nerve cells that produce dopamine located in the brain. Dopamine is a neurotransmitter, which is one of the chemicals responsible for transmitting signals between the nerve cells (neurons) of the brain. Deficiency of dopamine production affects the motor function, which leads to body movements issues. Consequently, patients will begin to experience difficulty in completing simple tasks and conducting other daily routines. PD characteristics may include tremor, bradykinesia (slowed movement), rigid muscles, impaired posture and balance, loss of automatic movements, writing changes, and speech changes. Patients also tend to develop non-motor symptoms such as mood disorder, cognitive changes, behavioral disorders, and dementia. These symptoms will gradually get worse over time, resulting in increased disease severity in patients [56].

The changes in patients' voice and speech patterns are the early and common PD symptoms that can be captured by audio sound recording software or device. The reason is that patients' voice tends to stutter and progressively becomes affected by time. Speech disability can affect several parts of verbalization state such as the production of spoken language (dysprosody), voice production (disphony), and articulation/pronunciation (dysarthria). Each patient recorded speech contains a set of multiple features that could reveal an abnormality in each patient speech elements [57]. The speech impairment characteristics of PD patients are tremor, silent voice, hoarseness, soft and monotonous speech, imprecise pronunciation, and breathiness [58]. Approximately 90% of the early PD patients show speech impairments. Thus, it can be a strong indicator in developing reliable PD diagnosis [59]. Previous work reported that the speech symptoms have a strong association

[1]Part of this chapter is reprinted with permission from "Multivariate Multi-step Deep Learning Time Series Approach in Forecasting Parkinson's Disease Future Severity Progression" by Nur Hafieza Ismail, Mengnan Du, Diego Martinez, and Zhe He, BCB '19: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Pages 383-389. https://doi.org/10.1145/3307339.3342185, Copyright 2019 by Association for Computing Machinery.

with PD severity as the disease progresses. From their observation, speech dysfunctional due to advanced PD became more apparent with the speech severity progression from a low-volume with a monotone voice to a certain extent where the patient's voice slowly faded and disappeared [58].

Usually, diagnosed patients are required to attend routine check-up and need to be present in the clinic. PD monitoring procedure is quite different from other diseases. PD symptoms evaluation mainly rely on human expertise [60]. For example, the patients' reading performance and speech rate will be observed by medical staffs. Moreover, traditional tracking PD symptoms progression often uses the Unified Parkinson's Disease Rating Scale (UPDRS). This procedure is time-consuming because it requires the motor skills examinations assisted by trained medical staff. Thus, symptoms monitoring also quite costly and logistically inconvenient for both sides (patient and clinical staff).

Since PD patients usually present particular characteristics in speech, voice recording data is feasible for diagnosis. Cellphones or smartphones can be a portable recording device which is convenient in monitoring patients' health remotely. Plus, the majority (95%) of US adults population owns a mobile phone and thus the remote PD monitoring can be implemented in real life [61]. The audio waveform of recorded speech can be transformed into several voice parameters such as subtle changes in voice frequencies (jitter), voice cycle-to-cycle magnitude difference (shimmer), volume (amplitude), vocal cord opening pressure, etc. In previous PD speech analysis, patients usually have a short maximum time of phonation, high rate of jitter and shimmer, reduced pitch range and higher phonation threshold pressure [62].

To tackle these challenges, we propose a monitoring framework that is able to forecasts PD progression of 16 speech features and UPDRS scores in the week/s period. We develop multivariate multi-step time series forecasting using DNN methods to observed multiple input variables and forecast multiple future time steps. We design a multichannel CNN network setting for time series to be applied to the PD speech telemonitoring dataset. The developed model's performance is compared with baseline models. We summarize our contributions as follows:

- We propose a framework of multivariate multi-step time series approach to monitoring the voice impairment in several different future time-frame based on previous and current knowledge of speech features.

- We design the DNN architecture to train the time series forecasting model using 16 features of speech signals and UPDRS.
- Our experiment results show that our model can provide forewarning to PD patients to take immediate action in delaying the symptom progression. Plus, it is also conducive for PD patients' health to be monitored remotely by a clinician without having to go to the doctor for a check up.

3.2 Related Work

Several studies have been conducted in PD to improve and maintain the patients' quality of life. In the medical perspective, to precise identification of PD symptoms and signs from the early stage has become a pressing issue and has rise interdisciplinary researchers' interest to widely explore this topic. Even though PD is incurable, with the right medication and proper treatment, the disease progression can be delayed [59]. To monitor disease progression, routine check-ups are essential for patients. The main challenge is PD effects on patients often overlapping with other diseases, making it laborious in the diagnosis procedure, especially at the early stage. Conventionally, the medical diagnosis procedure requires a medical history of the patient, caregiver's feedback, and several physical tests to inspect the motor fluctuations and dyskinesia development. Moreover, there is uncertainty regarding laboratory tests conducted on PD patients suspect, making it harder to diagnose, especially at the beginning level of PD. Since PD commonly affects the patients' motor skills performance, the dysfunction on physical movement and speech can be captured and recorded using electronic devices.

In previous work, different types of motor skills of body parts have been collected, such as typing pattern, gait disturbances in multiple walking patterns, hand movements, doing other routine activities, and speech patterns [63, 64, 65]. For fluctuation of movements on legs and arms tremor among PD patients can be differentiated with healthy non-PD people relaxed muscle movements [66]. However, except speech data, other symptoms need to be recorded using specific acceleration sensors wearable devices attached on body. While for speech, it can be recorded using a noninvasive tool such as smartphones. Thus, the convenience of using smartphones has motivated us to explore the characteristics of vocal features for this work. This simple procedure is suitable to attain the necessity in PD symptoms progression constant monitoring by clinicians so that they can immediately take action in dosing of medication, the side effect of drugs, and request of further check-up for allowing patients to perform at their best.

Various statistical and machine learning techniques have been applied to different types of PD dataset. In previous work, patients' body movements in doing daily activities were recorded using wearable devices. To identify the severity of motor skills, three types of dynamic learning structure algorithms were applied. Dynamic neural networks, dynamic SVM, and hidden markov models techniques were compared using global error rate and local error rate. The experimental results show that the dynamic neural network algorithm achieved the best results compared to dynamic SVM and hidden markov models [67]. The sensor system was developed to record the kinetic information during Deep Brain Stimulation Therapy to observe PD motor symptoms such as rigidity, bradykinesia, and tremor. This system was built using three machine learning models: a simple decision tree, linear SVM, and fine K-Nearest Neighbors (KNN). The predicted UPDRS scores are nearly correct using the fine KNN model [68]. The recorded PD patients' speech data are also vastly explored, because of abnormality in dysphonia features. A PD prediction tool was designed using parallel feed-forward neural network due to its ability in reducing the prediction error [69].

Furthermore, the speech symptom is also vastly explored by researchers. The recorded PD patients' voice signals contain several dysphonia features that are useful for classification. The authors proposed a new hybrid intelligent system using a combination of pre-processing techniques such as model-based clustering (gaussian mixture model) and multiple approaches of feature reduction/selection (principal component analysis, linear discriminant analysis, sequential forward selection, and sequential backward selection). It also used three supervised classifiers, such as the least-square SVM, probabilistic neural network, and general regression neural network [70]. A PD prediction tool is designed using parallel feedforward neural network due to its ability to reduce the prediction error. The output result is then compared against a rule-based tool in handling an imbalanced dataset for making the end decision [69]. Using a similar speech dataset, forty speech features from PD patients were extracted and the developed model (regression and DNN) was able to classify four severity groups (Healthy, Early, Intermediate & Advance) [71]. DNN was also employed to PD patients' audio voice dataset to calculate the UPDRS scores and categorize them into either "severe" or "not severe" classes [56]. The literature shows that DNN always presents a better performance compared to other machine learning classifier in PD diagnosis [72]. DNN has the capability in classifying unstructured data including audio and speech signals [56]. It consists of multiple layers of neurons that stacked together to generate reliable models for prediction or classification.

Time series forecasting is a mathematical estimation of particular values in the future that involves temporal measurements from previously observed information. The model was built based on specific assumptions about dynamic behaviors of the underlying system using statistical and mathematical approaches. There are numbers of available methods for forecasting, each of which was built based on different algorithms in different environments and has distinct assumptions on domain systems in temporal structure [73, 74]. Time series forecasting has been explored broadly in various fields due to its ability to tackle many issues that arise in real-world situations [74]. This approach has been successfully demonstrated in many fields such as business activities [75], financial management [76], meteorology for hurricanes and global warning [77], and healthcare applications [78]. In medical applications, time series forecasting model is used to predict the disease progression, estimate the mortality rate, and assess the possible risk over time. For examples, it has been used in monitoring cardiovascular diseases [74] and chronic kidney disease [79].

However, the model for time series is always challenging to handle multiple input variables and next to predict multiple output variables in multiple time steps. It can be divided into two types of forecasting which are short-term, intermediate-term, and long-term. Short-term is used to forecast future objective scenario within a minute to 24 hours and intermediate-term is within several days to many months, while long-term is commonly predicting scenario in years from the available dataset using intensive analysis and calculations [80]. These characteristics are suitable in many clinical situations. For example, to forecast patient severity of illness in the Intensive Care Unit (ICU), a multivariate time series approach with the multi-task Gaussian process were used to evaluate the clinical data that are sparse and heterogeneous. The short-term technique was applied in this situation where the medical doctor can monitor the patient's next progression based on the current state and allowed the doctor to immediately taking actions before the patient condition's worsen [81].

In this work, we applied multivariate multi-step time series forecasting using DNN methods on speech features. DNN is able to automatically learn arbitrary complex mappings from inputs to outputs and support multiple inputs and outputs. These are powerful features that offer a lot of promise for time series forecasting, particularly on problems with complex-nonlinear dependencies, multivalent inputs, and multi-step forecasting. DNN methods do provide a lot of promising capabilities for time series forecasting, specifically the automatic learning of temporal dependence and the automatic handling of temporary structures like trends and seasonality. Thus, Parkinson's telemonitoring dataset that contains weekly report is fit to apply in multichannel CNN model.

3.3 Methods

In this section, we will briefly introduce the problem statements, the proposed framework including data preparation procedure and DNN architecture.

3.3.1 Problem Statement

We denote a multivariate time series with D variables of length T as X = (x_1, x_2, ..., x_T)$\in R^{T\times D}$, where for each $t \in$ 1, 2, ..., T, $x_t \in R^D$ represents the observations (a.k.a., measurements) of all variables and x_t^d denotes the measurement of d-th variable of x_t. Let $s_t \in R$ denote the time-stamp when the tth observation is obtained and we assume that the first observation is made at time-stamp 0 (i.e., s_1 = 0). In this paper, we are interested in time series forecasting problem, where we compute the overall Root Mean Squared Error (RMSE) over ten different setting weeks and per-week RMSE for each timestep.

3.3.2 The Proposed Framework for Forecasting Parkinson's Disease Future Progression

Accurate decision support systems could help healthcare professionals in monitoring PD progression based on patient data. A new recorded audio voice through PD application installed in the electronic device will be saved in the PD data online storage. The current and past audio data are converted into speech features together with UPDRS scores. This multivariate input is then fed into the CNN time series model to forecast PD progression. The experimental output is evaluated to identify the PD voice symptoms status for incoming week/s. If the PD symptoms are getting worse, patients will be scheduled for a further checkup in the clinic.

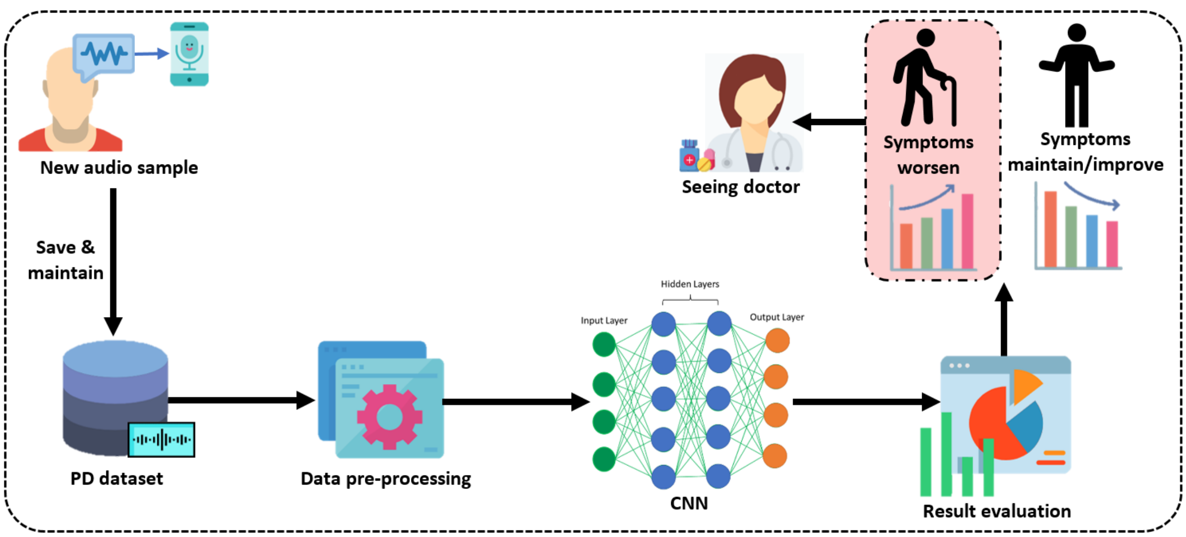

Figure 3.1: The overview of our proposed framework for forecasting PD future progression

This monitoring system can be developed by DNN methods. They can utilize DNN to learn from the past and additional current data and recognize the patterns. We trained the CNN multichannel model to learn the multivariate multi-step time series to forecasts the PD future progression in incoming week/s. The general framework of the proposed model is shown in Figure 3.1. Our proposed framework contains four main parts, which are PD dataset, data pre-processing, CNN multichannel model, and result evaluation.

3.3.3 Parkinsons Tele-monitoring Dataset

Table 3.1: Description of the features and UPDRS scores of the PD telemonitoring dataset

Description	Label	Feature label
Several measures of variation in fundamental frequency	F01	MDVP:Jitter (%)
	F02	MDVP:Jitter (Abs)
	F03	MMDVP:Jitter:RAP
	F04	MDVP:Jitter:PPQ5
	F05	Jitter:DDP
Several measures of variation in amplitude	F06	MDVP:Shimmer
	F07	MDVP:Shimmer(dB)
	F08	Shimmer:APQ03
	F09	Shimmer:APQ05
	F10	Shimmer:APQ11
	F11	Shimmer:DDA
Two measures of ratio of noise to tonal components in the voice	F12	NHR
	F13	HNR
A nonlinear dynamical complexity measure	F14	RPDE
Signal fractal scaling exponent	F15	DFA
A nonlinear measure of fundamental frequency variation	F16	PPE
Clinician's motor UPDRS score, linearly interpolated	O01	Motor-UPDRS
Clinician's total UPDRS score, linearly interpolated	O02	Total-UPDRS

The dataset used in this study contains a total of 5875 recordings from 42 subjects, which include 14 women and 28 men. Each patient has about 200 voice recordings. Every recorded voice has 16 vocal attributes based on traditional measurements (NHR, HNR, shimmer, Jitter) and nonlinear dynamical systems theory (RPDE, DFA, PPE). Each subjects' voice was recorded with phonations of the sustained vowel/a/. The dataset also contains Total-UPDRS and Motor-UPDRS scores. UPDRS is a standard scale used by clinicians during PD diagnosis and monitoring PD progression. This instrument is well-establish and has been widely used by a medical professional specializing in PD during patient's meetings. The ranges of Total-UPDRS and Motor-UPDRS are 0-176 (0 indicating healthy and 176 indicating total disability) and 0-108 (with 0 indicating healthy state and 108 indicating severe motor impairment), respectively. The dataset is available in

UCI machine learning repository [82]. Table 3.1 presents the 16 features of the dataset along with UPDRS scores. While Figure 3.2 shows the overall data distribution based on Total-UPDRS and Shimmer. Lastly, Table 3.3 displays the data frequency based on Total-UPDRS score.

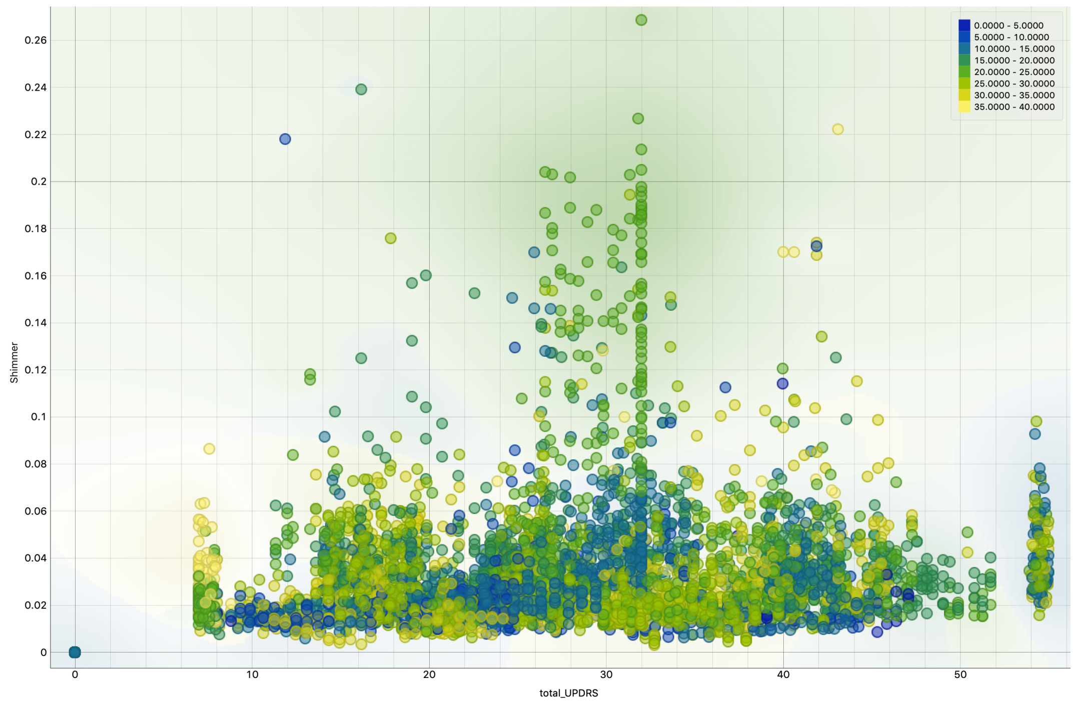

Figure 3.2: The data distribution

3.3.4 Data Setting for Multivariate Multi-step Time Series

Multivariate time series data is data where there is more than one observation for each time step. The Parkinson's telemonitoring dataset is considered as multivariate time series dataset because it describes the voice symptoms progression of early-stage of PD patients over six months. The data were collected using the telemonitoring device and speech audio was recorded every week in the patients' homes. To fit the dataset for time series forecasting, we prepare ten different data arrangement for ten different experiments. In total, we used 5042 recordings which is equal to 86% of the dataset. We could not use the rest of the dataset because some of the subjects have recorded their voices only up to 18 weeks. Table 3.2 presents the data setting for ten different experiments.

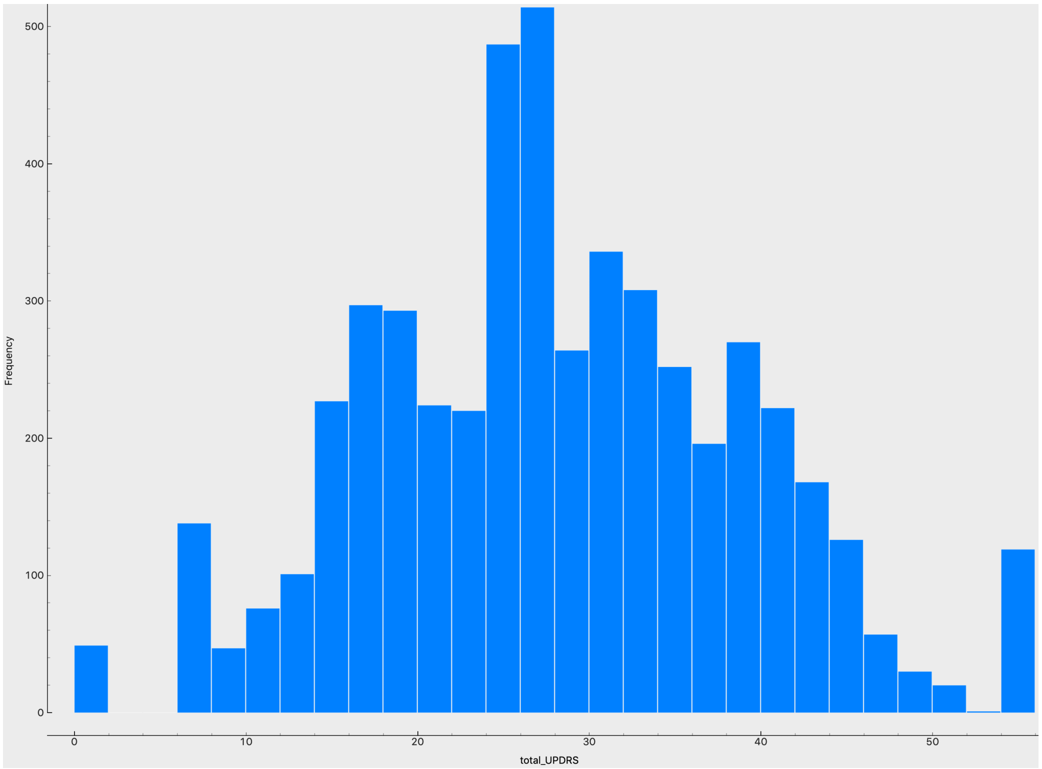

Figure 3.3: The data proportion in Total-UPDRS score

Table 3.2: Data setting for ten different experiments

Experiment	Input	Forecast Output
01	Week01	Week02
02	Week01-02	Week03-04
03	Week01-03	Week04-06
04	Week01-04	Week05-08
05	Week01-05	Week06-10
06	Week01-06	Week07-12
07	Week01-07	Week08-14
08	Week01-08	Week09-16
09	Week01-09	Week10-18
10	Week01-10	Week11-20

3.3.5 CNN Multichannel Architecture

CNN model is feasible for time series forecasting. Many types of CNN models can be used for each specific type of time series forecasting problem. We provide every individual one-dimensional time series to the model as an input of every individual channel. The model employs a different kernel and reads each input sequence onto a different set of filter maps, substantially learning features from time series input variables. This setting is suitable for problems that the output sequence is some function of the observations at the steps of time ahead from several different features, including the forecasted feature. The increase in data amount needs a more prominent and advanced model which require more time for training. The architecture of our proposed multichannel CNN model is inspired by [83]. Figure 3.4 shows the illustration of multichannel CNN for forecasting PD future progression.

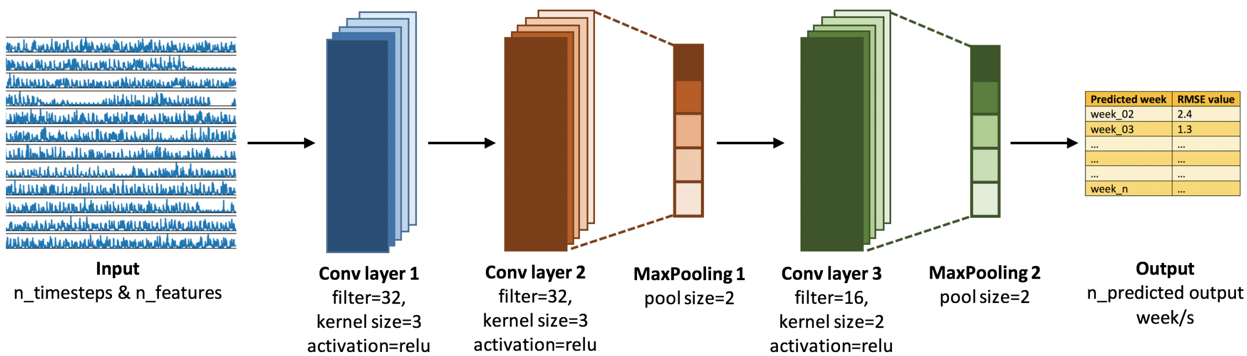

Figure 3.4: The architecture of multichannel CNN network for forecasting PD future progression

For the experiment, a three-stage CNN model was applied as the core building block in the network. The total variables and timesteps (week/s) from PD dataset used as an input network. Each feature learning consists of convolution (filter), activation, and pooling operators. After input stage, we construct two convolutional layers with a filter size of 32, kernel size of 3, ReLU as an activation function, followed by max-pooling with a pool size of two. Here we applied ReLU because it could train the neural networks faster [84]. Next, we add the third convolutional layer with a filter size of 16, kernel size of 2, ReLU, and max-pooling. The fully connected layer then describes the features up to 100 nodes. We train the model with 70 epochs and 16 samples for batch size. We also use the Adam optimization algorithm to estimate and update the network weights continuously based on training dataset. This network setting performs well with our multivariate variables. Pooling (max-pooling) reduces the resolution of input and make it robust to small variations for previously learned features. At the end of the three-stage feature extraction, the feature maps are flattened and fed into a fully connected layer for forecasting.

3.4 Experiment

We conducted the experiments to evaluate the proposed framework for forecasting PD future progression from weekly dataset containing standard temporal structure. We used Keras API running on Tensorflow to train DNN models.

3.4.1 Baseline Methods

We chose two baselines that are capable of handling multivariate variables for time series: CNN multiheaded and encoder-decoder Long Short-Term Memory (LSTM).

3.4.1.1 Multiheaded CNN

We designed the extended CNN that contain individual sub-model for all input variables, which also called as multiheaded CNN model. For multiheaded CNN model, we established an individual CNN model for 17 and 18 input variables for this network. The hyperparameters and total layers are modified to fit for this model. This model can utilize the available API functions to be more flexible. Each variable can be looped over and generate a sub-model that holds a one-dimensional week/s sequence from data and extracts a flat vector as outputs. These outputs contain a summary of the learned features from the input sequence. All of the extracted vectors were integrated by concatenation to produce one lengthy vector, and then interpreted by fully connected layers to be ready for a prediction. As we constructed the submodels, we define the inputs for the model and utilize the stack of flattening layers in the merged layer. To fit our PD data, it required about 17 and 18 arrays as input for the network which one for each submodel. It is essential for model training, model evaluation, and doing the forecasting with a developed model. To do that, we generate the 3D arrays that contain [samples, timesteps, 1], with one feature.

3.4.1.2 Encoder-Decoder LSTM

LSTM is a recurrent neural network that has the ability to learn and forecast long sequences. A benefit of LSTM in addition to learning long temporal sequences is that they can learn to make a one-shot multi-step forecast which may be useful for time series forecasting. However, LSTMs disadvantages are it can be complicated in the configuration process and requires a lot of preparation to get the dataset in a suitable format for learning. LSTM add the explicit handling of order between observations when learning a mapping function from inputs to outputs, which are not offered by MLP or CNN. This neural network includes native support for input dataset comprised of observations' sequences. We designed the encoder-decoder LSTM to use about 17 or 18 time series variables to forecast the next incoming week/s of PD voice symptoms progression. All individual one-dimensional time series were allocated as a different input sequence to the model. The LSTM will create an internal representation of each input sequence that will together be interpreted by the decoder. We used training epochs of 50 given the 8-fold increase in the amount of input data.

3.4.2 Result Evaluation Metric

The forecast will be comprised of ten different time frame settings, for week/s ahead, as shown in Table 3.2. It is common for multiple steps forecasting problems to evaluate each forecasted temporal step individually. The units of the speech features are numerical, and it would be functional to apply an error metric that employed the same units. From the literature, the RMSE and Mean Absolute Error (MAE) are suitable format for time series. However, RMSE was frequently used in previous work and we decided to adopt it for the experiment. Different from MAE, RMSE allocates a high weight to forecast errors. For this work, RMSE will be used to measure the performance for each lead time from week02 to week20. It is useful for summarizing the model's performance using an RMSE score to aid in model selection. The model's performance is based on multiple week/s forecasts.

3.4.2.1 RMSE

Accuracy is often regarded as the dominant criterion for selecting a forecasting method [85]. The accuracy of a forecasting method is determined by analyzing the forecast error, which is defined as the actual value minus the forecast (or fitted) value of the variable for time period t; namely: e_t = A_t - F_t, where e_t is the forecast error at time t; At the actual speech features' value at time t; F_t the forecast speech features' value at time t + 1. For instance, forecast optimization typically chooses a model that minimizes RMSE, which is calculated as RMSE = $\sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$.

3.4.3 Experimental Results and Discussion

We ran the experiments on three different DNN network settings which are CNN multichannel, CNN multiheaded, and encoder-decoder LTSM. We ran the experiments approximately 1000 times due to the stochastic nature of DNN to get a reasonably accurate result. We also conducted about 90 experiments with ten different combinations of the dataset. We applied three different variables combinations, which are: (1) 16 speech features and motor-UPDRS for Table 3.3, (2) 16 speech features and total UPDRS for Table 3.4, (3) 16 speech features, motor-UPDRS, and total-UPDRS for Table 3.5. Table 3.3 displays the RMSE values of using 17 variables (16 features and motor-UPDRS). The performance of three DNN methods were compared. Each method has ten different data temporal setting. For example, in Table 3.3, the first column (Multichannel CNN) represents the ten different data settings (1a-10a). The grey cells in the table present the input data in week/s used for the experiment, while the white cells contain the RMSE values for predicted week/s. The total number of forecasting output week/s will be the same with the total number of input week/s. For example, the column experiment 2a used two weeks (week01-week02) as an input to forecast the PD progression of two weeks ahead (week03-week04). From these three tables, we can see that experiments results are better using 16 speech features and motor-UPDRS as an input. The RMSE values are much lower in Table 3.3 compared to the other two tables. While Table 3.7 that used 16 speech features and both UPDRS scores presents the highest RMSE values.

Figure 3.3-3.5 present the average of RMSE values for every experiment in graph format for

	16 Features & Motor UPDRS																													
Method	Multichannel CNN										Multiheaded CNN										Encoder-decoder LSTM									
Experiment	1a	2a	3a	4a	5a	6a	7a	8a	9a	10a	1b	2b	3b	4b	5b	6b	7b	8b	9b	10b	1c	2c	3c	4c	5c	6c	7c	8c	9c	10c
Week01																														
Week02	2.5										4.2										2.9									
Week03		2.3										4.2										3.1								
Week04		2.4	2.4									4.2	4.3									2.9	3.1							
Week05			2.3	2.6									4.3	4.6									2.9	3.7						
Week06			2.6	2.4	2.7								4.3	4.6	4.9								3.3	3.2	3.5					
Week07				2.6	2.5	2.8								4.4	4.7	4.8								3.4	2.8	4.6				
Week08				2.9	2.7	2.6	3.3							4.5	4.7	4.7	5.2							3.4	2.8	4.4	4.9			
Week09					2.8	2.7	3.1	3.5							4.8	4.7	5.1	5.2							2.9	4.4	4.8	5.2		
Week10					3.1	2.7	2.9	3.1	4.0						4.7	4.8	4.9	5.1	5.6						3.0	4.3	4.8	5.0	5.3	
Week11						2.8	2.7	3.1	3.4	1.8						4.9	4.8	5.0	5.3	2.6						4.3	4.7	4.9	4.9	1.3
Week12						3.2	2.8	3.0	3.2	2.5						5.1	4.8	5.0	5.3	4.9						4.3	4.6	4.8	4.8	2.4
Week13							2.9	3.0	3.2	3.7							5.0	5.1	5.3	5.2							4.6	4.8	4.8	4.8
Week14							3.1	3.0	3.1	3.3							4.9	5.1	5.4	5.3							4.6	4.8	4.9	4.8
Week15								3.1	3.3	3.3								5.0	5.3	5.3								4.8	4.9	4.8
Week16								3.4	3.3	3.2								5.0	5.3	5.2								4.8	4.8	4.8
Week17									3.6	3.3									5.2	5.2									4.6	4.8
Week18									3.8	3.3									5.8	5.2									4.7	4.7
Week19										3.4										5.2										4.6
Week20										3.6										5.7										4.7

Table 3.3: The output of RMSE values in forecasting 16 speech features and motor-UPDRS

	16 Features & Total UPDRS																													
Method	Multichannel CNN										Multiheaded CNN										Encoder-decoder LSTM									
Experiment	1d	2d	3d	4d	5d	6d	7d	8d	9d	10d	1e	2e	3e	4e	5e	6e	7e	8e	9e	10e	1f	2f	3f	4f	5f	6f	7f	8f	9f	10f
Week01																														
Week02	3.5										5.6										3.4									
Week03		3.0										5.8										3.9								
Week04		3.3	2.8									6.0	5.9									4.2	4.4							
Week05			3.0	3.4									5.9	6.0									4.1	4.9						
Week06			3.2	3.1	3.2								5.7	5.9	6.7								4.2	4.1	5.2					
Week07				3.1	3.0	3.1								5.8	6.7	6.3								4.0	4.6	5.6				
Week08				5.9	3.1	3.2	3.7							5.9	6.1	6.3	6.4							4.0	4.5	5.3	5.9			
Week09					3.4	3.4	3.9	4.4							6.1	6.2	6.2	6.4							4.6	5.4	5.7	6.4		
Week10					3.9	3.5	4.1	3.9	4.6						6.1	6.1	6.2	6.3	6.8						4.7	5.7	5.4	6.2	6.5	
Week11						3.6	3.9	3.8	4.5	2.1						6.0	6.1	6.2	6.6	4.1						5.9	5.2	6.0	6.3	1.6
Week12						4.4	4.0	3.9	4.4	3.0						6.0	6.0	6.2	6.6	7.6						6.3	5.1	5.8	6.2	2.6
Week13							4.1	3.9	4.3	4.6							6.0	6.2	6.5	7.0							5.2	5.8	6.1	6.4
Week14							4.4	4.0	4.2	4.3							6.0	6.3	6.6	6.7							5.2	5.7	6.0	6.2
Week15								4.1	4.1	4.3								6.3	6.5	6.8								5.7	6.0	6.2
Week16								4.4	4.2	4.3								6.5	6.5	6.8								5.7	6.0	6.1
Week17									4.4	4.3									6.5	6.6									6.0	6.0
Week18									5.2	4.4									7.4	6.5									6.2	6.0
Week19										4.3										6.5										5.9
Week20										4.5										7.4										6.0

Table 3.4: The output of RMSE values in forecasting 16 speech features and total-UPDRS

	16 Features, Motor UPDRS & Total UPDRS																													
Method	Multichannel CNN										Multiheaded CNN										Encoder-decoder LSTM									
Experiment	1g	2g	3g	4g	5g	6g	7g	8g	9g	20g	1h	2h	3h	4h	5h	6h	7h	8h	9h	10h	1i	2i	3i	4i	5i	6i	7i	8i	9i	10i
Week01																														
Week02	3.3										5.6										3.2									
Week03		2.9										5.7										4.1								
Week04		3.3	2.8									5.7	5.8									4.0	4.2							
Week05			2.6	3.0									5.8	6.0									3.9	4.7						
Week06			3.0	2.8	3.0								5.9	5.9	6.2								4.3	3.9	5.8					
Week07				2.9	2.8	3.3								5.8	6.1	6.4								3.8	5.6	6.2				
Week08				3.5	2.9	2.6	3.2							5.9	6.4	6.4	6.3							4.1	5.6	5.9	6.3			
Week09					3.1	2.9	3.2	4.6							6.6	6.2	6.2	7.3							5.7	5.6	6.2	6.6		
Week10					3.7	2.9	3.4	4.0	4.1						6.6	6.2	6.2	7.3	6.8						5.8	5.5	6.1	6.3	6.6	
Week11						2.8	3.3	3.9	3.8	2.8						6.1	6.1	7.4	6.7	4.3						5.3	6.0	6.1	6.3	1.2
Week12						3.2	3.5	4.0	3.8	3.6						6.2	6.1	7.6	6.7	6.5						5.3	6.0	6.1	6.2	2.4
Week13							3.7	4.0	3.8	4.3							6.1	7.7	6.7	7.1							6.0	6.1	6.2	6.5
Week14							4.0	4.0	3.9	3.9							6.1	6.9	6.6	6.9							6.0	6.1	6.1	6.3
Week15								4.1	4.0	3.9								7.0	6.7	7.0								6.0	6.1	6.2
Week16								4.2	4.2	3.7								6.9	6.6	7.1								6.0	6.0	6.2
Week17									4.1	3.7									6.6	6.9									6.0	6.1
Week18									4.6	3.7									7.5	7.0									6.1	6.1
Week19										3.9										7.1										6.0
Week20										4.2										7.7										6.1

Table 3.5: The output of RMSE values in forecasting 16 speech features, motor-UPDRS, and total-UPDRS

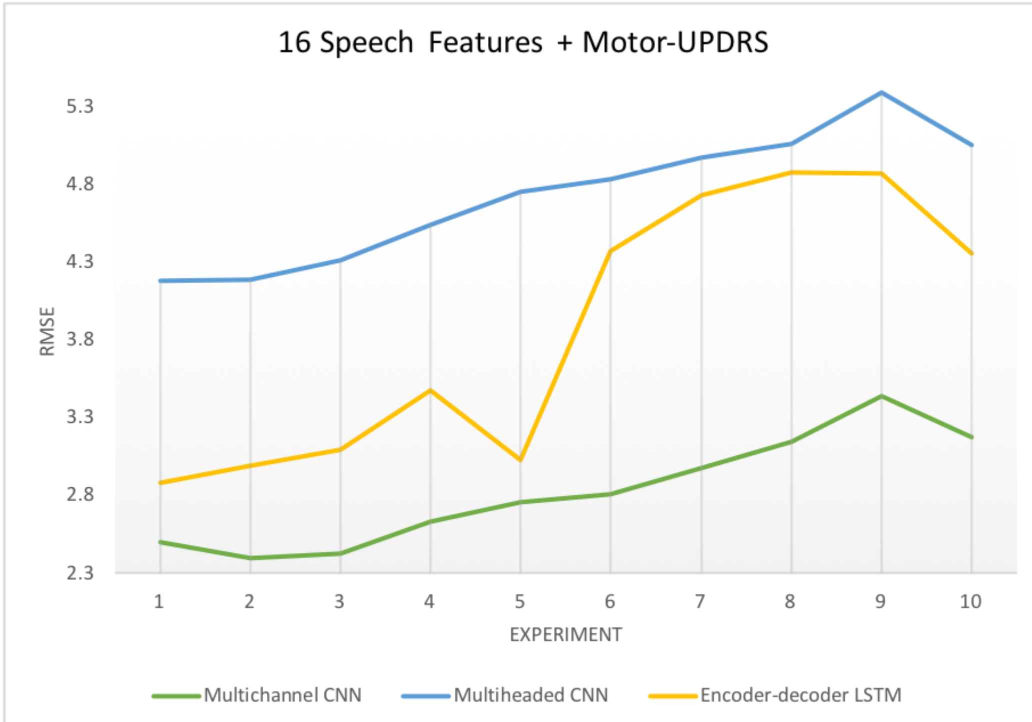

Figure 3.5: The average RMSE values for 16 speech features and motor-UPDRS

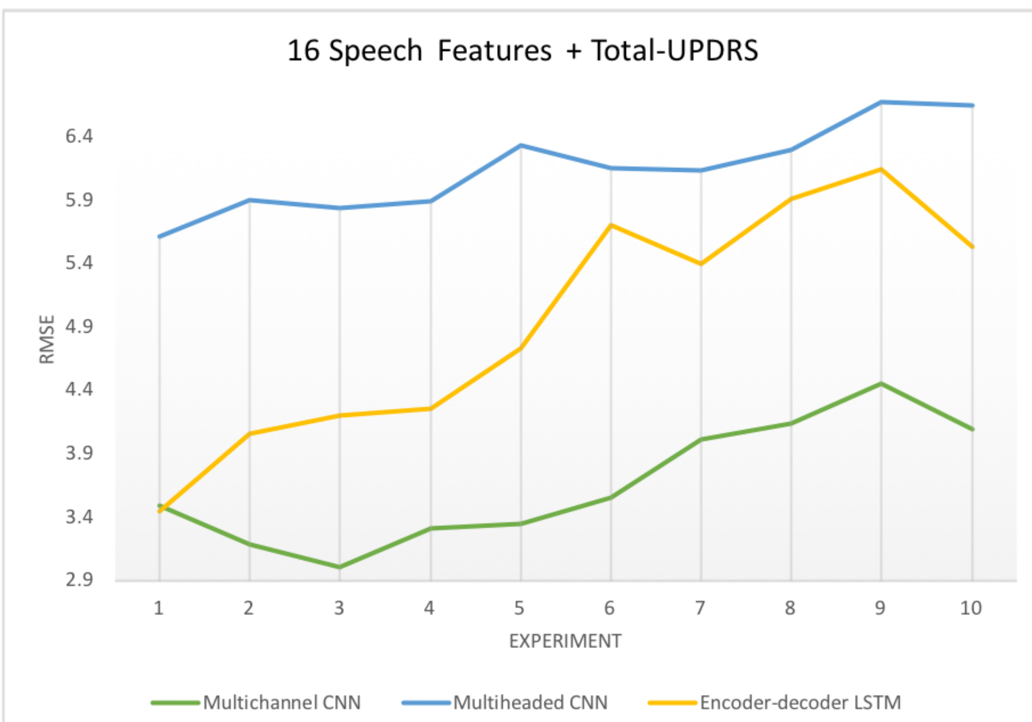

Figure 3.6: The average RMSE values for 16 speech features and total-UPDRS

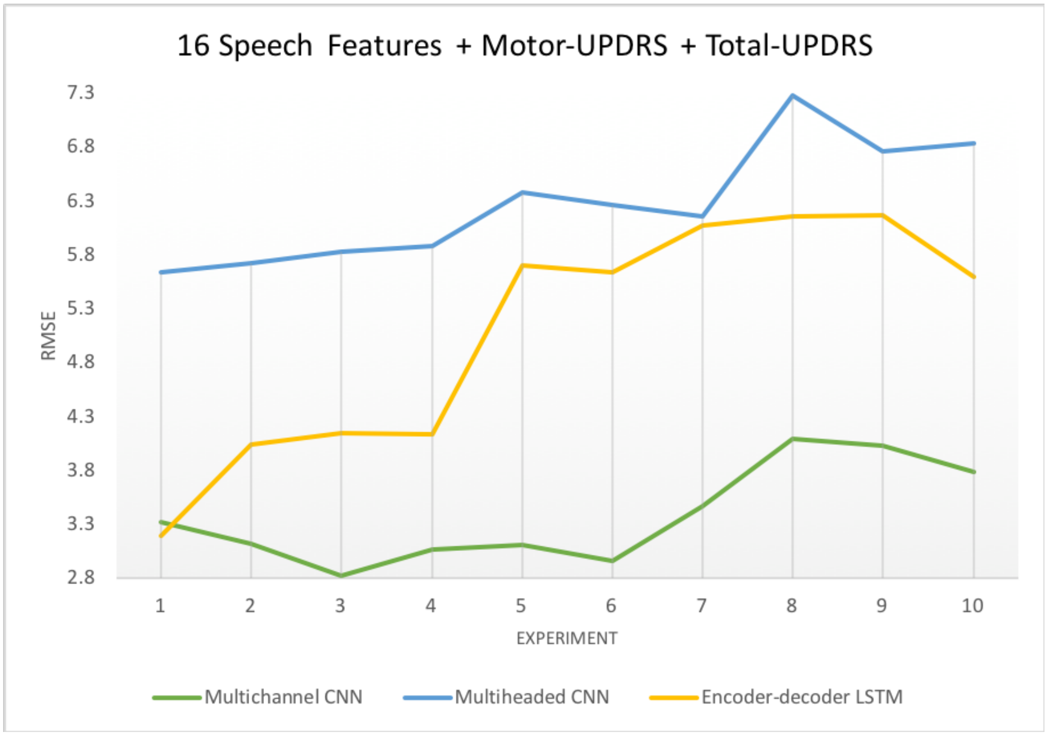

Figure 3.7: The average RMSE values for 16 speech features, motor-UPDRS, and total-UPDRS

better representation. The detail setting for each experiment can be referred in Table 3.2. Figure 3.5 shows the RMSE scores for all three models using 16 speech features with motor-UPDRS. Figure 3.6 shows the RMSE scores for all three models using 16 speech features with total-UPDRS. Figure 3.7 shows the RMSE scores for all three models using 16 speech features, motor-UPDRS, and total-UPDRS. Multichannel CNN model is the best compared to others by producing the least RMSE values in all experiments settings. Results indicate that multichannel CNN can effectively forecast the future progression of PD patients. Instead of mapping inputs to outputs alone, the network is capable of learning a mapping function for the inputs over time to output.

3.5 Conclusions

PD is a neurodegenerative disorder that affects the dopamine neurons production in specific part in the brain. PD is also recognized as the second most common degenerative nerve disorder in the United States after Alzheimer's disease. About 1% of the world population which estimated 7 to 10 million people after the age of 60, with an average age of 62 are PD sufferers. Every year, approximately 60,000 Americans are diagnosed with PD, and the researchers believe this number will continue to grow. By providing a computational prognosis tool for PD using patients' dataset that contains clinical PD rating scale based on speech features could alleviating the speech changes symptom. This can help huge amount of people who want to know the progression of unusual symptoms that they are currently facing based on previous and current recorded speech.

Remote tracking of UPDRS using voice measurements is an effective screening step before an appointment with a clinician. Developing computational tools using DNN techniques can assist the medical expert in forecasting PD progression for the patient faster and recognize the subjects at an early stage. This can be a useful guide for clinical staff, following the progression of clinical PD symptoms regularly. PD is often tricky in diagnosis and also time-consuming to monitor the progress of the symptoms, but even at early stages, small vocal differences may be machine-detectable. Using this information, it becomes possible to monitor PD using voice recordings from potential patients.

In this work, we develop a model to forecast the motor symptoms severity progression through voice speech features. We proposes a multi-step time series approach to forecasting the PD symptoms progression using DNN methods. Three different setting of DNN methods such as multi-channel CNN, multi-headed CNN, and encoder-decoder LSTM are employed and their performance are compared. The experimental results on available public PD dataset show that the proposed multichannel CNN model remarkably helping in forecasting of PD progression in several incoming weeks. Our proposed model can assist medical practitioners in healthcare practice for monitoring PD symptoms severity growing. It also can be implemented as an efficient clinical decision support system for PD treatments as it demonstrated that real PD data could efficiently forecast PD progression.

4. MULTIVARIATE DATA, THE NEUROPSYCHOLOGY ASSESSMENT FOR IDENTIFYING DEMENTIA IN PARKINSON'S DISEASE PATIENTS

4.1 Introduction

Dementia is not a disease. It is a general term used to describe abnormal changes in the brain. The characteristics of dementia are memory difficulties, problem-solving, language, and thinking abilities, which can interfere with daily life and independent function. It eventually will affect the emotions, feelings, behavior, and relationships with others. Dementia is more synonymous with Alzheimer's disease caused by brain cell damage and destruction. The damage cells will interfere with brain activity and neuron connection. However, in most cases, PD patients will eventually experience dementia as their disease progresses at least a year after the diagnosis. In the early years, the symptoms start slowly and gradually get worse in the following years. The previous studies reported that about 50-80% of PD patients experience dementia [86]. PD is a progressive neurodegenerative disease that affects 1–2% of people older than 60. Although PD has long been considered a motor disorder predominantly, its frequent association with dementia has recently gained increasing recognition. Patients with PD have an almost sixfold increased risk of developing dementia compared with age-matched individuals without PD. In a 12-year population study of patients with PD, the cumulative incidence of dementia increased steadily with age and disease duration reaching 80–90% by age 90. Dementia contributes significantly to the morbidity and mortality of PD. The factors commonly associated with PD-Dementia are declining years, severe stage of parkinsonism (rigidity, postural disturbance, and gait disorder), and psychiatric symptoms (depression, anxiety). It also includes trouble interpreting visual information, memory loss, sleep disorder, low concentration, and judgment, and Mild Cognitive Impairment (MCI) [87].

Until today, there are still no available treatments to cure or stop the damage of brain cells caused by PD-Dementia. However, the drug and non-drug treatments could help ease and improve the symptoms suffered by PD-Dementia patients. Thus, if the PD patients start to experiencing poor memory, thinking trouble, or other dementia symptoms, they are advised to seek the doctor treatments as soon as possible to determine the cause. The diagnosis could help patients in receiving the optimum benefits of early available therapies. Dementia cannot be determined with only one test, but a combination of tests and guidelines are required for the diagnosis. The medical procedures needed in detecting dementia are neuropsychological and cognitive tests, psychiatric evaluation, brain scans, laboratory tests, and genetic tests.

Thus, to prolong and reduce the dementia risk, we propose a framework in detecting dementia among PD patients using neuropsychological assessment. We classify the samples using the MoCA scores as a guideline during the preprocessing stage. We classify them into three different categories, which are No dementia, PD-MCI, and PD-Dementia. In this study, we used the Parkinson's Progression Markers Initiative (PPMI) dataset, which available upon request from the Image Data Archive (IDA) website that was maintained by the University of Southern California (USC). We design a DNN architecture specific for analyzing electronic health records for PD-Dementia detection. We summarize our contributions as follows:

- We propose a framework of monitoring the neuropsychological progression symptoms among PD patients.
- We design the DNN architecture to train the dementia detection model using non-motor information of PD patients.
- Our experiment results show that the trained model can be used as an alarm tool in detecting PD-Dementia. Plus, the model is reliable in monitoring the patients' health status remotely.

4.2 Related Work

Many dementia studies have been conducted in diagnosis, management, and treatment of this condition. Continuous research exploration on this topic is essential to identify the symptoms and prevent it from getting worse. Early detection will help patients face any physical and emotional changes that distract them from everyday lives. Various diseases correlated to dementia such as Alzheimer's Disease (AD), Frontotemporal, Lewy Body (LB), Vascular, Huntington, and PD. The overlap symptoms of multiple dementia have made it harder to received an accurate diagnosis. Plus, there are only a few dementia works done in PD. Even though the PD-Dementia cases are small compared to other related brain diseases, but PD patients are at high risk of getting dementia. Plus, the PD patients with dementia symptoms are at increased risk of morbidity and mortality. Thus, we take the initiative of conducting this study so that PD-Dementia patients will appropriately be treated.

Usually, healthy people will be experiencing brain cells losses as they age, but for dementia patients, the losses are far more significant that affect a person's functioning. During the severe phase, patients with dementia will entirely rely on other family members to conduct necessary daily activities. Some older people with dementia cannot control their feelings and emotions. Thus, it will affect their usual behavior and personalities. Neuropsychological studies reported that cognitive capacities deficit is associated with dementia. The neuropathological characteristics were prominent among PD-Dementia patients [88]. PD-Dementia has its own medical profile and neuropathology characteristics, differentiated from other diseases with dementia.

In the previous work, non-motor symptoms such as Rapid Eye Movement (REM) sleep behavior disorder and Korean MoCA were used as the leading indicator to distinguish PD-Dementia and AD with Dementia (AD-Dementia). 110 PD-Dementia and 118 AD patients with age at least 60 years involved in this work [89]. The patients' profile was analyzed to develop the PD-Dementia detection model. In the other work done by [90], the visual measures and retinal thinning have shown vital risk factors of dementia development in PD. These features were used in the experiment because the human eye structure linked with dopaminergic layers in which dopamine deficiency is one of the PD-Dementia symptoms. About 146 participants contributed to the study, with 112 PD patients and others being healthy.

Additionally, the PD patients' historical data, including demographic, age, behavior, current health status, and dopaminergic treatment, were considered in the research conducted by [91]. 85 PD-Dementia and 444 PD patients data used for the experiments. The results proved that the psychiatric and cognitive features contribute the most in identifying dementia in PD patients. In other similar work done by [92], about 140 PD patients were involved in the study to examine the neuropathological substrates of cognitive deficits, which contributes to PD-Dementia. All the participants are either having a stable cognitive function or dementia symptoms for two or more years. The results show that 92 of PD patients developed dementia while the 48 remained cognitively intact. 542 PD patients were involved in the with 46 had PD-Dementia, 64 had LB with Dementia (LB-Dementia) [93]. The experiment analysis shows that the overall incidence rate of PD-Dementia is higher than LB-Dementia. The incidence rate also increased progressively with age.

Earlier, much statistical analysis and data mining techniques have been used widely by researchers in developing accurate dementia detection models. The examples of methods used for the experiments are latent class analysis, logistic regression, kruskal-wallis, mann-whitney, decision tree, and random forest [94, 91, 95]. In the last five years, DL has shown phenomenal growth in disease diagnosis [96]. DL applies representation learning techniques that beneficially the hidden new knowledge from the input data without requiring manual features extraction from that data. It is because the network will automatically learn to extract features during the training process [63]. Thus, in this work, we explored the neuropsychological assessment and current age to detect the dementia signs in PD patients. We construct a DNN model that effectively identify these essential groups, PD-MCI and PD-Dementia.

4.3 Methods

This section will introduce the proposed framework, including PPMI neuropsychological assessment, data preprocessing, and MLP architecture.

4.3.1 The Proposed Framework for Identifying Dementia in Parkinson's Disease Patients from Neuropsychological Assessment

Figure 4.1 shows our proposed framework in identifying the dementia status in PD patients. This proposed framework contains four main parts, which are EHR database, data preprocessing, DL for neuropsychological assessment inputs, and result evaluation. The PD patient's information and their progression of health status saved in the database. We extracted the patient's neuropsychological information from the record and labeled them into three categories: No Dementia, PD-MCI, and PD-Dementia. In this preprocessing phase, the labeling process has conducted the dataset using MoCA scores information that we retrieved from the same database. After that, the preprocessed data then fed as an input into the DL algorithm for training. To check the model reliability in classifying the PD patient's current status, we evaluate the model using several establish evaluation metrics. From the model, if the level shows as PD-MCI or PD-Dementia, the patient will see a doctor for further checkups. The doctor may suggest changing the medication or provide a new treatment to ease the symptoms.

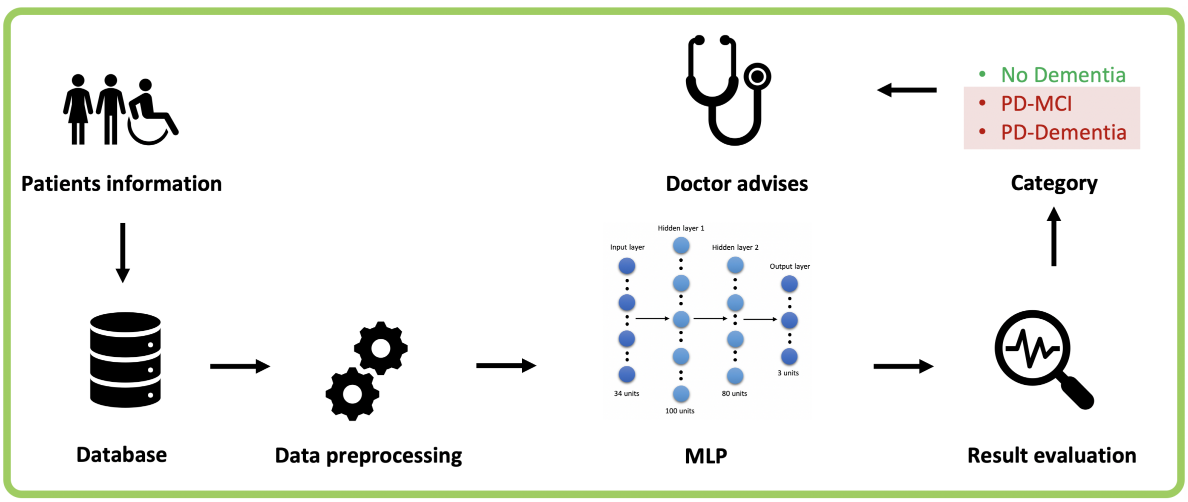

Figure 4.1: The overview of our proposed framework for identifying dementia in PD patients using DL model

4.3.2 PPMI Neuropsychological Assessment

PPMI is a repository for observational clinical study containing various types of medical imaging, biospecimen information, clinical observation, and behavioral evaluation to identify PD progression biomarkers to get a better understanding of the disease [97]. Data used in the preparation of this work were obtained from the PPMI database (www.ppmiinfo.org/data). For up-to-date information on the study, visit www.ppmiinfo.org. The data was downloaded on March 16th, 2020.

The data stored in PPMI was contributed by 21 clinical sites that followed the publicly available standardized data protocols through the PPMI website. The institutional review board approved the study conducted by PPMI at each clinical site, and the participants provided written informed consent. The supplement information of the PPMI study has published a few years back [97].

In this study, we used PD patients' information about the non-motor assessment of neuropsychological tests. The neuropsychological tests that we used for the experiments are semantic fluency and MoCA. We used the non-motor assessment to identify PD-Dementia because it is a commonly reported symptom among PD patients. Plus, it is often unremarkable in medical practice. Conversely, PD diagnosis is usually determined based on motor symptoms [98]. However, patients with PD show common non-motor conditions when they started developing dementia-related disabilities in daily life [99].

The example of non-motor assessment is cognitive and neuropsychological tests. These tests will measure language skills, memory, visual ability, and other brain functioning abilities. Neuropsychological testing usually is time-consuming and needs a proper introduction to the patients. This session requires special monitoring from the neuropsychologist. Nevertheless, if the patient reported having cognitive impairment, their doctor can conduct a quick test during checkups. The suspect dementia patient will undergo a MoCA test to assess their cognitive status. The MoCA screening test investigates orientation, memory, and attention and the ability to recognize items, follow verbal and written instructions, and reproduce a complex shape. Plus, MoCA has better sensitivity and specificity than a famous Mini-Mental State Examination (MMSE) [100]. Another neuropsychological test used in this work is semantic fluency. The goals were to examine the quantitative for word generation and qualitative for grouping and switching categories in verbal fluency.

The participants were 403 PD patients diagnosed with PD for two years or less who are not taking PD medications and 64 PD patients who do not show evidence of a dopaminergic deficit. Each patient took the neuropsychological tests multiple times with a least approximately one year gap for every test. The total data used for this study is 2873. Table 4.1 presents 34 features of

Table 4.1: Description of the features of neuropsychological assessment from PPMI dataset

Description	**Type**	**Label**	**Feature label**
Semantic Fluency	Numerical	VLTANIM	Total number of animal
	Numerical	VLTVEG	Total number of vegetable
	Numerical	VLTFRUIT	Total number of fruits
	Numerical	DVS SFTANIM	Derived-Sem. Fluency - Animal scaled score
	Numerical	DVT SFTANIM	Derived-Sem. Fluency - Animal T-score
MoCA	Numerical	MCATOT	MoCA total score
	Numerical	MCASER7	Attention - Serial 7s
	Numerical	MCAVFNUM	Verbal Fluency - Number of words
	Numerical	MCASNTNC	Sentence repetition
	Numerical	MCAABSTR	Abstraction
	Categorical	MCAALTTM	Alternating trail making
	Categorical	MCACUBE	Visuoconstructional skills (cube)
	Categorical	MCACLCKC	Visuoconstructional skills (clock cont)
	Categorical	MCACLCKN	Visuoconstructional skills (clock num)
	Categorical	MCACLCKH	Visuoconstructional skills (clock hands)
	Categorical	MCALION	Naming - Lion
	Categorical	MCARHINO	Naming - Rhino
	Categorical	MCACAMEL	Naming - Camel
	Categorical	MCAFDS	Attention - Forward digit span
	Categorical	MCABDS	Attention - Backward digit span
	Categorical	MCAVIGIL	Attention - Vigilance
	Categorical	MCAVF	Verbal Fluency
	Categorical	MCAREC1	Delayed Recall - Face
	Categorical	MCAREC2	Delayed Recall - Velvet
	Categorical	MCAREC3	Delayed Recall - Church
	Categorical	MCAREC4	Delayed Recall - Daisy
	Categorical	MCAREC5	Delayed Recall - Red
	Categorical	MCADATE	Orientation - Date
	Categorical	MCAMONTH	Orientation - Month
	Categorical	MCAYR	Orientation - Year
	Categorical	MCADAY	Orientation - Day
	Categorical	MCAPLACE	Orientation - Place
	Categorical	MCACITY	Orientation - City
Age	Numerical	AGE ASSESS	Age at assessment

semantic fluency and MoCA tests together with the patients' age. Eleven features are numerical type, and the remaining 23 features are categorical.

Figure 4.2 shows the data proportion for each group. The total instances for each group are; No dementia is 1716, PD-MCI is 1108, and PD-Dementia is 52. Figure 4.3 presents the overall

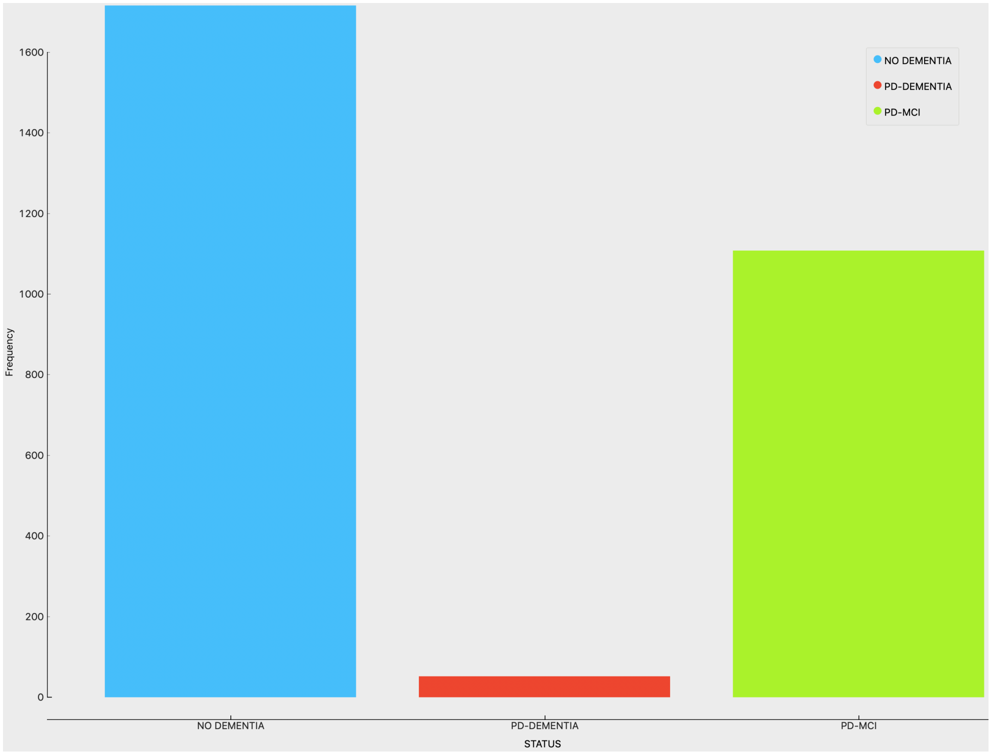

Figure 4.2: The data proportion in three classes

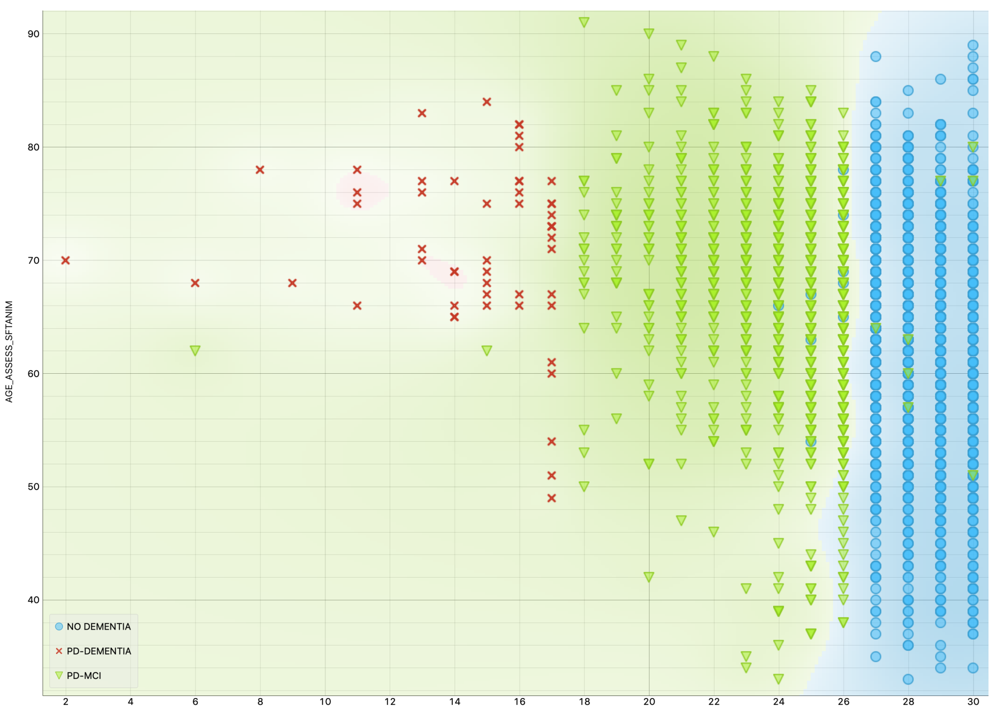

Figure 4.3: The proportion of the data

data distribution based on the MoCA and age of the participants.

4.3.3 Data Preprocessing

We classified each of 2873 data samples into three different labels: No dementia, PD-MCI, and PD-Dementia. The label is based on MoCA scores defines by The Health Navigator New Zealand website, a non-profit community. The website goals are to provide reliable, trustworthy health information and self-care resources to the public. Table 4.2 presents the MoCA scores for three different labels.

Table 4.2: MoCA scores range interpretation

Dementia Stage	MoCA Score	Cognitive Status
No Dementia	26 - 30	No cognitive impairment
PD-MCI	18 - 25	A mild but noticeable decline in cognition
PD-Dementia	0 - 17	Definite cognitive decline & impairment

4.3.4 MLP Architecture for Multivariate Data

This study used the classical type of neural network, MLP algorithm with backpropagation, which will automatically learn to map from inputs to outputs. It's network settings are flexible and has shown excellent performance in a broad range of problems. MLP works very well with tabular datasets for classification prediction tasks and regression prediction problems. Our MLP contains simple architecture with four layers with a different number of nodes. Figure 4.4 shows the overview of our MLP. The input layer includes 34 neurons, represents the same total of 34 input features. The two hidden layers have 100 neurons and 80 neurons with ReLu as an activation function. Lastly, the output layer contains three neurons with three labels that we want to predict: No dementia, PD-MCI, and PD-Dementia. MLP able to supports multi-class classification by adopting a softmax activation function in the output layer. MLP model supports multi-label classification in which a sample can belong to more than one class. For each class, the raw output passes through the logistic function. Values larger or equal to 0.5 are rounded to 1, otherwise to

0. For a predicted outcome of a sample, the indices where the value is 1 represent the sample's assigned classes. MLP trains using some form of gradient descent, and the gradients calculated using backpropagation. It will minimize the cross-entropy loss function and produce a vector of probability value that will estimate the probability of the sample's class for classification purposes. We also used the Adam optimization algorithm to continuously evaluate and update the network weights based on the training dataset. We have tried several different network settings to produce the best and reliable MLP architecture for our problem.

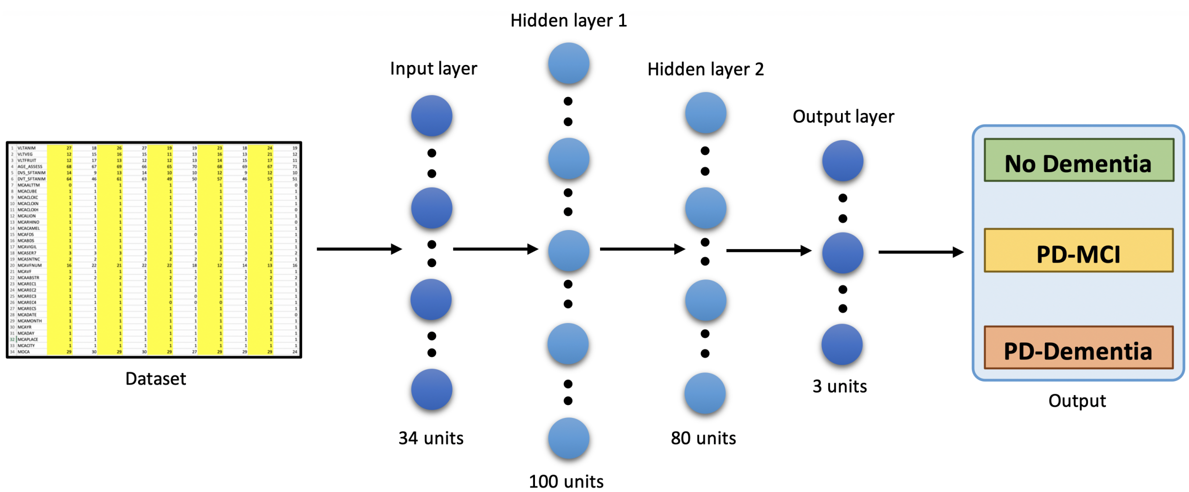

Figure 4.4: The MLP architecture

4.4 Experiment

We conducted experiments to evaluate the proposed network settings for identifying dementia in PD patients. We applied a 10-fold cross-validation setting to train and test all the models. We briefly introduce the baselines methods. After that, we report the experimental performances, and we discuss our findings.

4.4.1 Baseline Methods

We present the baselines to evaluate our proposed MLP architecture. The input of our dataset was in a Comma-Separated Values (CSV) format that uses a comma to separates each value. Row in the file represents data record in which each record consists of several fields that are also separated by commas. Therefore, we chose five baselines frequently used for machine learning, such as decision tree, random forest, naive bayes, KNN, and SVM.

4.4.1.1 Decision Tree

Decision tree is a simple supervised learning algorithm that is frequently used for classification and regression purposes. It learns from data to approximate a sine curve with a set of if-then-else decision rules. The deeper the tree, the more complex the decision rules, and the fitter the model. The decision tree creates a model in a tree structure form. It will break down a dataset into smaller and smaller subsets, and at the same time, an associated decision tree is incrementally developed. The final result is a tree with decision nodes that could have two or more branches. The leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called the root node. Decision trees can handle both categorical and numerical data.

4.4.1.2 Random Forest

Random forest is an ensemble learning method generally used for classification and regression tasks. This algorithm contains a large combination of multiple decision trees, which tree is constructed during the training process. When developing individual trees, an arbitrary subset of attributes is drawn (hence the term "Random"), from which the best feature for the split is selected. The final model is based on the majority vote from individually developed trees in the forest.

4.4.1.3 Naive Bayes

A fast and straightforward probabilistic classifier based on bayes' theorem with the assumption of feature independence. This supervised algorithm has been used in a wide variety of classification tasks. Naive bayes is a classification algorithm suitable for binary (0 and 1) classification and multi-class classification problems.

4.4.1.4 K-Nearest Neighbor

The KNN is a simple non-parametric, and supervised machine learning algorithm that records all instances during the training process and will classify new samples based on similarity measures such as distance functions. A sample is classified by a majority vote of its neighbors, with the sample being assigned to the class most common amongst its K nearest neighbors measured by a distance function. It searches for "k" closest training examples in feature space and uses their average as the prediction.

4.4.1.5 SVM

SVM is a supervised machine learning algorithm that can be employed for classification and regression tasks. It separates the attribute space with a hyper-plane, maximizing the margin between the instances of different classes or class values. The technique often yields supreme predictive performance results.

4.4.2 Experimental Result and Discussion

Table 4.3: Evaluation results

Model	**AUC**	**Accuracy**	**F1**	**Precision**	**Recall**
Decision Tree	0.498	0.597	0.446	0.356	0.597
SVM	0.708	0.698	0.636	0.773	0.698
KNN	0.917	0.841	0.837	0.842	0.841
Naive Bayes	0.976	0.902	0.904	0.908	0.902
Random Forest	0.989	0.969	0.968	0.968	0.969
MLP	**0.995**	**0.975**	**0.975**	**0.975**	**0.975**

We ran the experiments using six different methods. Our results indicate that our proposed MLP effectively can identify the current health status of PD patients. We compared each method's experiment results with five different evaluation metrics such as Area Under the ROC Curve (AUC), accuracy classification score, F1-measure, precision, and recall. It is crucial to better understand the trade-off in performance for different threshold values when interpreting probabilistic predictions. Experimental results in Table 4.3 shows our MLP has the highest accuracy value of 97.5%. It also has the highest AUC, F1, precision, and recall values compared to other models. These results indicate that our MLP performed better and the most reliable model to identify dementia in PD patients.

Table 4.4: The proportion of predicted instances using our MLP

		Predicted			
		PD-Dementia	**PD-MCI**	**No Dementia**	Total
Actual	**PD-Dementia**	**85.4%**	1.5%	0	52
	PD-MCI	14.6%	**96.2%**	1.3%	1108
	No Dementia	0	2.3%	**98.7%**	1716
	Total	41	1123	1712	2876

Additionally, Table 4.4 shows the proportion of the corretly predicted instances in percentage. For the PD-Dementia, about 85.4% or 35 instances were correctly classified by our MLP. While, for the PD-MCI, about 96.2% or 1080 instances were predicted correctly. Lastly, about 98.7% or 1690 instances were labeled correctly for No Dementia.

Table 4.5: Description of the added brain features to the dataset

Description	**Type**	**Label**	**Feature label**
DATScan	Decimal	CAUDATE_R	Right caudate
	Decimal	CAUDATE_L	Left caudate
	Decimal	PUTAMEN_R	Right putamen
	Decimal	PUTAMEN_L	Left putamen

We extend the experiment by using dataset with added brain features. The description of four Dopamine Transporter Scan (DATScan) information are show in Table 4.5. DATScan is a scanning tool used to determine the diagnosis of PD. This dataset contained 1480 samples. Table 4.6 shows the proportion of correctly predicted instances with brain features. We using the same network setting to analyzed the new instances. The table shows that the percentages values are dropped

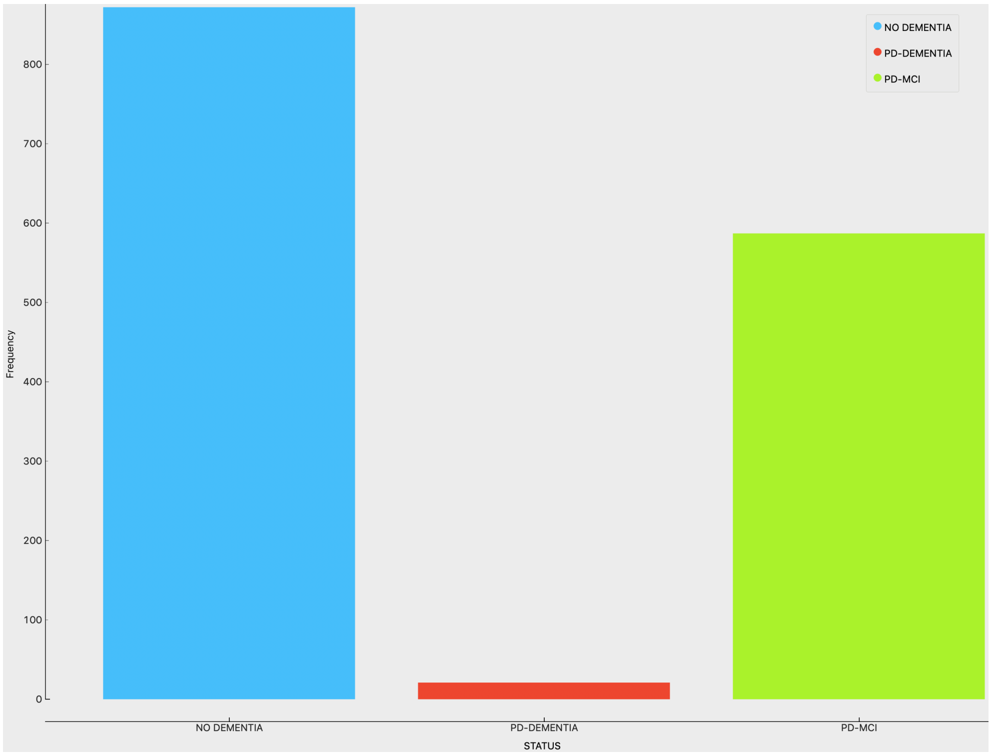

Figure 4.5: The data proportion with added brain features in three classes

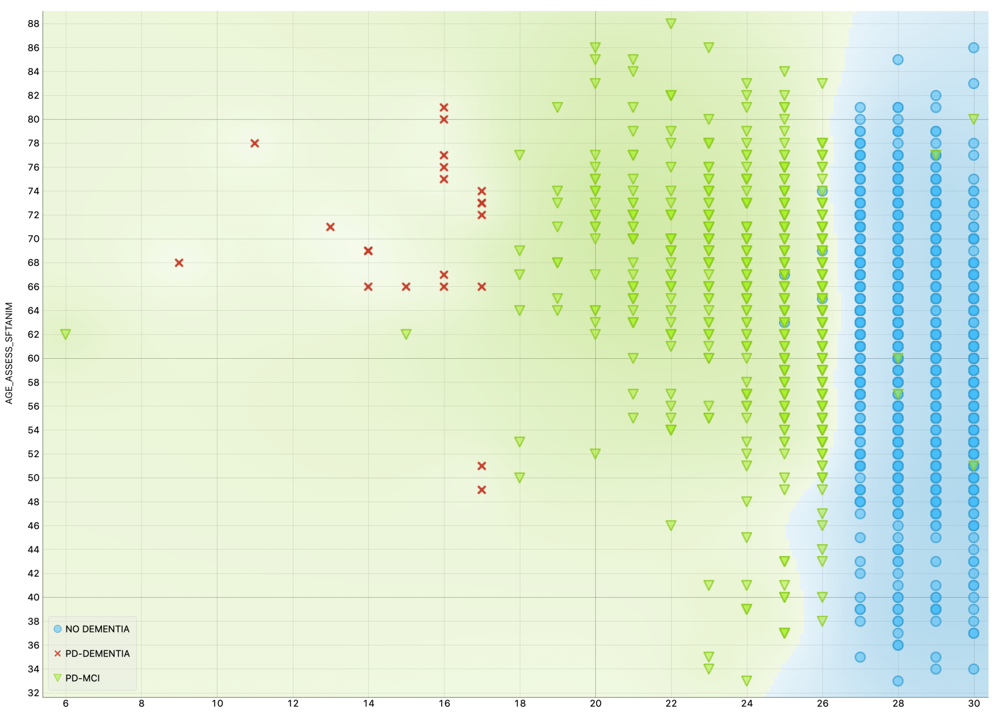

Figure 4.6: The distribution of the data with added brain features

for all classes especially for the PD-Dementia. It is because the limited instances of this label. However, we believe that the prediction value would improved if we could include more instances in this experiments. Figure 4.5 shows the total data with added brain features in three different classes. While Figure 4.6 presents the data distribution with added brain features based on the MoCA and age of the participants.

Table 4.6: The proportion of predicted instances using our MLP with added brain features

		Predicted			
		PD-Dementia	**PD-MCI**	**No Dementia**	Total
Actual	**PD-Dementia**	**78.6%**	1.7%	0	21
	PD-MCI	21.4%	**95.3%**	2.5%	562
	No Dementia	0	3.1%	**97.5%**	872
	Total	14	590	876	1480

From this presented proportion, we can conclude that our proposed MLP is reliable and can be trusted to monitor the PD patients' health status for the dementia task. Surprisingly, for the PD-Dementia group, the MLP has no miss-classified instances for No Dementia. It is a good sign that the model will not miss-diagnose the PD patient with dementia as No Dementia. Error in diagnosis is an unwanted situation that we have to avoid. This situation is dangerous because the patients will be ignored for treatment and will continue to suffer from progressive dementia. If this situation continues, we afraid that it will be too late to treat the symptoms. The model also presents high correctly classify instances proportion for PD-MCI. PD-MCI phase is a transitional conditional between mild PD state and dementia state. PD-MCI patients could get suitable treatment to reduce the risk of getting dementia.

4.5 Conclusions

PD is a progressive neurodegenerative disease that affects 1–2% of people older than 60. Although PD has long been considered a motor disorder predominantly, its frequent association with dementia has recently gained increasing recognition. Patients with PD have an almost sixfold increased risk of developing dementia compared with age-matched individuals without PD. In a 12-

year population study of patients with PD, the cumulative incidence of dementia increased steadily with age and disease duration reaching 80–90% by age 90. Dementia contributes significantly to the morbidity and mortality of PD. Key risk factors or correlates consistently associated with PD-Dementia are older age, more severe parkinsonism (incredibly rigidity, postural instability, and gait disturbance), specific psychiatric symptoms (depression, psychosis), and MCI [87].

PD causes physical symptoms at first, including tremor on some of the body parts. Over time, PD patients might start to experience cognitive function impairment, such as concentration problems and forgetfulness. As the disease gets worse with time, many PD patients develop dementia. PD-Dementia can cause permanent memory loss and affect social relationships with others. PD-Dementia reduced the self-care ability in patients. The medical experts are still unable to uncover the concrete answers to why dementia often occurs in PD patients. Until today, there is no available treatment to cure PD and PD-Dementia. The current treatment by healthcare only helps patients to prolong or delay the symptoms. Thus, scheduled monitoring and early diagnosis are essential to detect dementia in PD patients. The sufferers could get the proper treatment to help them ease the symptoms.

In this work, we proposed an MLP architecture to identify dementia symptoms in PD patients. We used non-motor features such as semantic fluency and MoCA information to train our model. We also run an experiment using a dataset with additional features of brain information, which is DATScan. Both results are compared, and we conclude that the model performed very well in identifying dementia even though we trained the model without DATScan. This result shows that we can virtually conduct an early diagnosis, which will help patients save cost and time. For healthcare, they can improve operations activities because they can reduce time on unnecessary appointments.

In the future, we believe that by including the other biomarkers such as brain images and motor features as input, it will be a more useful indicator in developing PD-Dementia prediction models in the future [89]. We are unable to use brain images due to the limited number of a dataset. However, we believe in the future more brain information will be added to the PPMI database.

5. CONCLUSION AND FUTURE WORK[1]

5.1 Conclusion

DL, a subfield of machine learning, has seen a dramatic resurgence in the past few years, primarily driven by increases in computational power and the availability of massive new datasets. The field has witnessed striking advances in machines' ability to understand and manipulate data, including images, language, speech, and multivariate input. Healthcare and medicine stand to benefit immensely from DL because of the high volume of data being generated and the increasing proliferation of electronic medical devices and digital record systems.

The future of healthcare has never been more exciting. Not only do Artificial Intelligence (AI) and present an opportunity to develop solutions that cater to particular needs within the industry, but DL in healthcare can become incredibly powerful for supporting clinicians and improving patient care. While DL in healthcare is still in the early stages of its potential, it has already seen significant results. The leading institutions and medical bodies have recognized the benefits it brings, and the popularity of the solutions are well known. The future of healthcare still lies in the hands of medical professionals. However, with support from a reliable computational system, it could help healthcare decide particular needs and environments and reduce the stresses that they experience daily. Healthcare has high hopes for the role of DL in clinical decision support and predictive analytics for a wide variety of conditions.

In this dissertation, we proposed a series of DL algorithms to handle multiple modalities data in healthcare. We systematically develop three prediction models for identifying diseases, including PTSD, PD, and PD-Dementia. They are motivated by different theories in medical and computational perspectives. We have handled multimodal healthcare data information throughout these years, including text, speech information, and multi-type data. We also carefully studied each disease's background, including the symptoms and test assessment run by healthcare. Thus, with the

[1]Part of this chapter is reprinted with permission from "Social Media and Psychological Disorder" by Nur Hafieza Ismail, Mengnan Du, and Xia Hu, 2019, In: Bian J., Guo Y., He Z., Hu X. (eds) Social Web and Health Research. Springer, Cham., Pages 171-192, doi.org/10.1007/978-3-030-14714-3_9, Copyright 2019 by Springer, Cham.

underlying special properties of each disease that we discovered, our developed models could be employed in a real-world scenario. Besides, we have explored the online social media and medical applications capability and potential in disease diagnosis and a health monitoring system.

For example, online social media and forum discussion sites contain a significant amount of information. The popularity of social media, advanced NLP systems, and excellent text classification methods have opened up a vast opportunity for researchers to explore more about linguistic style in virtual life. The hidden knowledge behind the written text is very precious in several cases, including mental health problems and threat cases (bomb threats, school threats, etc.). For mental health issues, an early depression identification with an alarming system will inform the social media users to seek the doctor immediately for diagnosis and treatments. Based on our work and the previous report, we can conclude that the DL algorithms have shown promising results as innovative tools with high-value applications in the real-world medical environment. This research will broadly impact fields such as information retrieval, social computing, and health informatics.

5.2 Future Research Direction

While performing research on DL architecture for healthcare for multiple modalities data, I have not only accumulated several rewarding but also challenging research questions at the same time. They could be categorized into two directions, i.e., interpretable DL models for healthcare and drug discovery precision medicine.

5.2.1 Interpretable DL Models for Healthcare

Although the field of DL technologies for healthcare is advancing rapidly, there are still lots of future work needed to promote this field's progress further. (1) The detection algorithms are often regarded as black-boxes and criticized by their lacking of interpretability [101]. The necessity for interpretable computational systems is strong as DL becomes more common in healthcare. More interpretable models are on-demand to increase the acceptance of users for detection models. (2) Accurate prediction requires an analysis of a massive amount of factors, including demographic information, medical history, lifestyle, and other information that could contribute to diseases. (3)

Using Automated Machine Learning (AutoML) [102] in the medical field is possible and suitable because it provides techniques and methods for non-machine learning experts without having to code. AutoML can also process a significant amount of data, identify the critical features, and extract patterns for modeling that are more accurate and work faster than classic models. These characteristics are essential in disease diagnosis, which requires precise prediction in a real-time manner on online social media and healthcare applications.

The need for interpretable and reliable computational healthcare systems for patient's monitoring grows along with the popularity of DL applications used in many other fields. Explainable DL systems aim to provide self-explanation about the meaningful reasoning behind system decisions and predictions. Precisely, explainable algorithms can control and monitor adverse or unwanted effects, such as bias in decision-making or treatment. DL explanations could help healthcare in many ways, such as improving patient's care and confidence in diagnosis while relying on DL decisions.

The model explanation can be designed using several different of output formats for different user groups [103]. Visual explanations apply visual elements to describe the reasoning behind the develop models. Attention maps and visual saliency in the form of saliency heatmaps [104] are examples of visual explanations that are widely used in previous work. Verbal explanations describe the model or output analysis with natural language. This kind of explanations are popular for question-answering applications and recommendation systems.

5.2.2 Drug Discovery and Precision Medicine

Machine learning was applied in drug discovery about 30 years ago, and several tools have been developed for this purpose. However, unlike traditional machine learning, DL has unique criteria that make it very popular and influential, flexible in its architecture. This exciting characteristic has made it possible to custom-tailored the DL network for a particular problem. Thus, drug development and precision medicine are also necessary fields that need to be explored using the DL approach. These tasks require computational processing of heterogeneous information of genomic, clinical, and population-level data. The goal is to identify previously unknown relations between genes, pharmaceutical products, lifestyle, and living environments. DL is an ideal approach for interdisciplinary researchers and pharmaceutical organizations to reveal new patterns or key features in these datasets. It is because many precision medicine researchers are still ambiguous about what they should be looking for.

Genetic medicine is one of the medicine branches involving integration and application technologies for data collection in the population-level study to understand diseases and genetics affect drug response. It is a new field to be explored, so the new unforeseen discoveries can be exciting for researchers and the medical community because it has a high potential to improve patients care and reduce the drug side effects. Moreover, the combination of predictive analytics and molecular modeling will hopefully uncover new insights into how and why certain cancers or diseases form in individual patients. By leveraging the dataset combination such as EHR data, clinical guidelines, and real-time monitoring data set using DL, it may be possible to reduce the number of unnecessary biopsies that are performed due to suspicious findings in the mammograms. The advanced DL technologies can accelerate the process of analyzing data and shrinking the processing time.

REFERENCES

[1] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[2] A. Mxoli, N. Mostert-Phipps, and M. Gerber, "Risks and benefits of social computing as a healthcare tool," *Society and Information Technologies: ICSIT*, 2016.

[3] J. Sarasohn-Kahn, "The wisdom of patients: Health care meets online social media," *California HealthCare Foundation Oakland, CA*, 2008.

[4] Y. Punie, W. Lusoli, C. Centeno, G. Misuraca, and D. Broster, "The impact of social computing on the eu information society and economy," *JRC Scientific and Technical Report, JRC IPTS, EUR*, vol. 24063, 2009.

[5] N. H. Ismail, M. Du, and X. Hu, "Social media and psychological disorder," in *Social Web and Health Research*, pp. 171–192, Springer, 2019.

[6] P. E. Kummervold, D. Gammon, S. Bergvik, J.-A. K. Johnsen, T. Hasvold, and J. H. Rosenvinge, "Social support in a wired world: use of online mental health forums in norway," *Nordic Journal of Psychiatry*, vol. 56, no. 1, pp. 59–65, 2002.

[7] S. Wallace, M. Clark, and J. White, "'it's on my iphone': attitudes to the use of mobile computing devices in medical education, a mixed-methods study," *BMJ Open*, vol. 2, no. 4, 2012.

[8] C. L. Ventola, "Mobile devices and apps for health care professionals: uses and benefits," *Pharmacy and Therapeutics*, vol. 39, no. 5, p. 356, 2014.

[9] R. Atkinson, T. Hare, M. Merriman, and A. Vogel, "Therapeutic benefits of expressive writing in an electronic format," *Nursing Administration Quarterly*, vol. 33, no. 3, pp. 212–215, 2009.

[10] G. Turner-McGrievy, S. Kalyanaraman, and M. K. Campbell, “Delivering health information via podcast or web: Media effects on psychosocial and physiological responses,” *Health Communication*, vol. 28, no. 2, pp. 101–109, 2013.

[11] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[12] X. Du, Y. Cai, S. Wang, and L. Zhang, “Overview of deep learning,” in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 159–164, IEEE, 2016.

[13] M. Morchid, “Neural networks for natural language processing,” *Ph.D. Thesis of Avignon Université*, 2019.

[14] T. Young, D. Hazarika, S. Poria, and E. Cambria, “Recent trends in deep learning based natural language processing,” *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[15] K. Suzuki, “Overview of deep learning in medical imaging,” *Radiological Physics and Technology*, vol. 10, no. 3, pp. 257–273, 2017.

[16] H.-M. Gao and J.-S. Hong, “Why neurodegenerative diseases are progressive: uncontrolled inflammation drives disease progression,” *Trends in Immunology*, vol. 29, no. 8, pp. 357–365, 2008.

[17] E. Esposito and S. Cuzzocrea, “New therapeutic strategy for parkinson’s and alzheimer’s disease,” *Current Medicinal Chemistry*, vol. 17, no. 25, pp. 2764–2774, 2010.

[18] S.-F. Zhang, J.-H. Zhai, B.-J. Xie, Y. Zhan, and X. Wang, “Multimodal representation learning: Advances, trends and challenges,” in *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*, pp. 1–6, IEEE, 2019.

[19] N. Flaks-Manov, M. Topaz, M. Hoshen, R. D. Balicer, and E. Shadmi, “Identifying patients at highest-risk: the best timing to apply a readmission predictive model,” *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 118, 2019.

[20] A. Alkan, Z. G. Guc, F. C. Senler, T. Yavuzsen, H. Onur, M. Dogan, E. Karci, A. Yasar, E. B. Koksoy, O. Tanriverdi, *et al.*, "Breast cancer survivors suffer from persistent post-mastectomy pain syndrome and posttraumatic stress disorder (orthus study): a study of the palliative care working committee of the turkish oncology group (tog)," *Supportive Care in Cancer*, vol. 24, no. 9, pp. 3747–3755, 2016.

[21] T. Akechi, T. Okuyama, Y. Sugawara, T. Nakano, Y. Shima, and Y. Uchitomi, "Major depression, adjustment disorders, and post-traumatic stress disorder in terminally ill cancer patients: associated and predictive factors," *Journal of Clinical Oncology*, vol. 22, no. 10, pp. 1957–1965, 2004.

[22] N. C. C. for Mental Health (UK and Others, *Post-traumatic stress disorder: The management of PTSD in adults and children in primary and secondary care*. Gaskell, 2005.

[23] M. De Choudhury, S. Counts, and E. Horvitz, "Social media as a measurement tool of depression in populations," in *Proceedings of the 5th Annual ACM Web Science Conference*, pp. 47–56, ACM, 2013.

[24] M. De Choudhury, S. Counts, and E. Horvitz, "Major life changes and behavioral markers in social media: case of childbirth," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pp. 1431–1442, ACM, 2013.

[25] M. J. Paul and M. Dredze, "You are what you tweet: Analyzing twitter for public health," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[26] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Seventh International AAAI Conference on Weblogs and Social Media*, 2013.

[27] G. Coppersmith, C. Harman, and M. Dredze, "Measuring post traumatic stress disorder in twitter," in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[28] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in twitter," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 51–60, 2014.

[29] S. R. Krothapalli and S. G. Koolagudi, "Characterization and recognition of emotions from speech using excitation source information," *International Journal of Speech Technology*, vol. 16, no. 2, pp. 181–201, 2013.

[30] S. Balters and M. Steinert, "Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices," *Journal of Intelligent Manufacturing*, vol. 28, no. 7, pp. 1585–1607, 2017.

[31] S. Latif, R. Rana, S. Younis, J. Qadir, and J. Epps, "Cross corpus speech emotion classification-an effective transfer learning technique," *arXiv preprint arXiv:1801.06353*, 2018.

[32] J. Gideon, S. Khorram, Z. Aldeneh, D. Dimitriadis, and E. M. Provost, "Progressive neural networks for transfer learning in emotion recognition," *arXiv preprint arXiv:1706.03256*, 2017.

[33] M. De Choudhury, M. R. Morris, and R. W. White, "Seeking and sharing health information online: comparing search engines and social media," in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1365–1376, ACM, 2014.

[34] U. Fischer, O. Zachariae, F. Baum, D. von Heyden, M. Funke, and T. Liersch, "The influence of preoperative mri of the breasts on recurrence rate in patients with breast cancer," *European Radiology*, vol. 14, no. 10, pp. 1725–1731, 2004.

[35] F. M. Lewis and L. W. Deal, "Balancing our lives: A study of the married couple's experience with breast cancer recurrence," *Oncology Nursing Forum-Oncology Nursing Society*, vol. 22-6, pp. 943–956, 1995.

[36] M. L. Stuber, A. E. Kazak, K. Meeske, L. Barakat, D. Guthrie, H. Garnier, R. Pynoos, and A. Meadows, "Predictors of posttraumatic stress symptoms in childhood cancer survivors," *Pediatrics*, vol. 100, no. 6, pp. 958–964, 1997.

[37] J. M. Salsman, S. C. Segerstrom, E. H. Brechting, C. R. Carlson, and M. A. Andrykowski, "Posttraumatic growth and ptsd symptomatology among colorectal cancer survivors: a 3-

month longitudinal examination of cognitive processing," *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer*, vol. 18, no. 1, pp. 30–41, 2009.

[38] C. M. H. Chan, C. G. Ng, N. A. Taib, L. H. Wee, E. Krupat, and F. Meyer, "Course and predictors of post-traumatic stress disorder in a cohort of psychologically distressed patients with cancer: A 4-year follow-up study," *Cancer*, vol. 124, no. 2, pp. 406–416, 2018.

[39] E. Moschopoulou, I. Hutchison, K. Bhui, and A. Korszun, "Post-traumatic stress in head and neck cancer survivors and their partners," *Supportive Care in Cancer*, vol. 26, no. 9, pp. 3003–3011, 2018.

[40] J. James, Y. T. Harris, I. M. Kronish, J. P. Wisnivesky, and J. J. Lin, "Exploratory study of impact of cancer-related posttraumatic stress symptoms on diabetes self-management among cancer survivors," *Psycho-Oncology*, vol. 27, no. 2, pp. 648–653, 2018.

[41] S. M. Brown, A. Webb, R. Mangoubi, and J. Dy, "A sparse combined regression-classification formulation for learning a physiological alternative to clinical post-traumatic stress disorder scores," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[42] E. Sharma and M. De Choudhury, "Mental health support and its relationship to linguistic accommodation in online communities," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 641, ACM, 2018.

[43] S. Dutta, J. Ma, and M. De Choudhury, "Measuring the impact of anxiety on online social interactions," in *Twelfth International AAAI Conference on Web and Social Media*, 2018.

[44] D. Vergyri, B. Knoth, E. Shriberg, V. Mitra, M. McLaren, L. Ferrer, P. Garcia, and C. Marmar, "Speech-based assessment of ptsd in a military population using diverse feature classes," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[45] S. K. Ernala, T. Labetoulle, F. Bane, M. L. Birnbaum, A. F. Rizvi, J. M. Kane, and M. De Choudhury, "Characterizing audience engagement and assessing its impact on so-

cial media disclosures of mental illnesses," in *Twelfth International AAAI Conference on Web and Social Media*, 2018.

[46] D. Banerjee, K. Islam, K. Xue, G. Mei, L. Xiao, G. Zhang, R. Xu, C. Lei, S. Ji, and J. Li, "A deep transfer learning approach for improved post-traumatic stress disorder diagnosis," *Knowledge and Information Systems*, vol. 60, no. 3, pp. 1693–1724, 2019.

[47] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *Journal of the American Medical Informatics Association*, vol. 24, no. 2, pp. 361–370, 2016.

[48] J. D. Flory and R. Yehuda, "Comorbidity between post-traumatic stress disorder and major depressive disorder: alternative explanations and treatment considerations," *Dialogues in Clinical Neuroscience*, vol. 17, no. 2, p. 141, 2015.

[49] A. Leano, M. B. Korman, L. Goldberg, and J. Ellis, "Are we missing ptsd in our patients with cancer? part i," *Canadian Oncology Nursing Journal*, vol. 29, no. 2, p. 141, 2019.

[50] B. Cui, Y. Li, Y. Zhang, and Z. Zhang, "Text coherence analysis based on deep neural network," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 2027–2030, ACM, 2017.

[51] S. Raschka, "Naive bayes and text classification," *Retrieved from sebastianraschka. com/Articles/2014_naive_bayes_1. html*, 2014.

[52] M. Goudjil, M. Koudil, M. Bedda, and N. Ghoggali, "A novel active learning method using svm for text classification," *International Journal of Automation and Computing*, vol. 15, no. 3, pp. 290–298, 2018.

[53] A. Padia, A. Roy, T. W. Satyapanich, F. Ferraro, S. Pan, Y. Park, A. Joshi, T. Finin, *et al.*, "Umbc at semeval-2018 task 8: Understanding text about malware," in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018.

[54] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, “Deep learning-based document modeling for personality detection from text,” *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74–79, 2017.

[55] M. Wess, “Bringing hope and healing to grieving patients with cancer.,” *The Journal of the American Osteopathic Association*, vol. 107, no. 12 Suppl 7, p. ES41, 2007.

[56] S. Grover, S. Bhartia, A. Yadav, K. Seeja, *et al.*, “Predicting severity of parkinson’s disease using deep learning,” *Procedia Computer Science*, vol. 132, pp. 1788–1794, 2018.

[57] T. Khan, J. Westin, and M. Dougherty, “Classification of speech intelligibility in parkinson’s disease,” *Biocybernetics and Biomedical Engineering*, vol. 34, no. 1, pp. 35–45, 2014.

[58] R. J. Holmes, J. M. Oates, D. J. Phyland, and A. J. Hughes, “Voice characteristics in the progression of parkinson’s disease,” *International Journal of Language & Communication Disorders*, vol. 35, no. 3, pp. 407–418, 2000.

[59] M. Nilashi, O. Ibrahim, and A. Ahani, “Accuracy improvement for predicting parkinson’s disease progression,” *Scientific Reports*, vol. 6, p. 34181, 2016.

[60] C.-W. Cho, W.-H. Chao, S.-H. Lin, and Y.-Y. Chen, “A vision-based analysis system for gait recognition in patients with parkinson’s disease,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 7033–7039, 2009.

[61] “Mobile fact sheet,” *https://www.pewinternet.org/fact-sheet/mobile/*, April 1, 2019.

[62] K. Chenausky, J. MacAuslan, and R. Goldhor, “Acoustic analysis of pd speech,” *Parkinson’s Disease*, 2011.

[63] T. Arroyo-Gallego, M. J. Ledesma-Carbayo, Á. Sánchez-Ferro, I. Butterworth, C. S. Mendoza, M. Matarazzo, P. Montero, R. López-Blanco, V. Puertas-Martin, R. Trincado, *et al.*, “Detection of motor impairment in parkinson’s disease via mobile touchscreen typing,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 1994–2002, 2017.

[64] W. Nanhoe-Mahabier, A. Snijders, A. Delval, V. Weerdesteyn, J. Duysens, S. Overeem, and B. Bloem, “Walking patterns in parkinson’s disease with and without freezing of gait,” *Neuroscience*, vol. 182, pp. 217–224, 2011.

[65] S. Sapir, L. Ramig, and C. Fox, “Speech and swallowing disorders in parkinson disease,” *Current Opinion in Otolaryngology & Head and Neck Surgery*, vol. 16, no. 3, pp. 205–210, 2008.

[66] R. LeMoyne and T. Mastroianni, “Use of smartphones and portable media devices for quantifying human movement characteristics of gait, tendon reflex response, and parkinson’s disease hand tremor,” in *Mobile Health Technologies*, pp. 335–358, Springer, 2015.

[67] B. T. Cole, S. H. Roy, C. J. De Luca, and S. H. Nawab, “Dynamical learning and tracking of tremor and dyskinesia from wearable sensors,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 5, pp. 982–991, 2014.

[68] P. Angeles, Y. Tai, N. Pavese, S. Wilson, and R. Vaidyanathan, “Automated assessment of symptom severity changes during deep brain stimulation (dbs) therapy for parkinson’s disease,” in *2017 International Conference on Rehabilitation Robotics (ICORR)*, pp. 1512–1517, IEEE, 2017.

[69] F. Åström and R. Koker, “A parallel neural network approach to prediction of parkinson’s disease,” *Expert Systems with Applications*, vol. 38, no. 10, pp. 12470–12474, 2011.

[70] M. Hariharan, K. Polat, and R. Sindhu, “A new hybrid intelligent system for accurate detection of parkinson’s disease,” *Computer Methods and Programs in Biomedicine*, vol. 113, no. 3, pp. 904–913, 2014.

[71] E. Benmalek, J. Elmhamdi, and A. Jilbab, “Updrs tracking using linear regression and neural network for parkinson’s disease prediction,” *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 4, no. 6, pp. 189–193, 2015.

[72] R. Das, “A comparison of multiple classification methods for diagnosis of parkinson disease,” *Expert Systems with Applications*, vol. 37, no. 2, pp. 1568–1572, 2010.

[73] C. Cheng, A. Sa-Ngasoongsong, O. Beyca, T. Le, H. Yang, Z. Kong, and S. T. Bukkapatnam, "Time series forecasting for nonlinear and non-stationary processes: a review and comparative study," *IIE Transactions*, vol. 47, no. 10, pp. 1053–1071, 2015.

[74] C. Bui, N. Pham, A. Vo, A. Tran, A. Nguyen, and T. Le, "Time series forecasting for healthcare diagnosis and prognostics with the focus on cardiovascular diseases," in *International Conference on the Development of Biomedical Engineering in Vietnam*, pp. 809–818, Springer, 2017.

[75] R. G. Hoptroff, "The principles and practice of time series forecasting and business modelling using neural nets," *Neural Computing and Applications*, vol. 1, no. 1, pp. 59–66, 1993.

[76] K. K. Jae, "Financial time series forecasting using support vector machines," *Neurocomputing*, pp. 307–319, 2003.

[77] C. W. Landsea, "Meteorology: hurricanes and global warming," *Nature*, vol. 438, no. 7071, p. E11, 2005.

[78] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 5, no. 1, pp. 82–87, 1995.

[79] A. Perotte, R. Ranganath, J. Hirsch, D. Blei, and N. Elhadad, "Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis," *AMIA Annual Symposium Proceedings*, vol. 22, no. 4, pp. 872–880, 2015.

[80] H. Al-Hamadi and S. Soliman, "Long-term/mid-term electric load forecasting based on short-term correlation and annual growth," *Electric Power Systems Research*, vol. 74, no. 3, pp. 353–361, 2005.

[81] M. Ghassemi, M. A. Pimentel, T. Naumann, T. Brennan, D. A. Clifton, P. Szolovits, and M. Feng, "A multivariate timeseries modeling approach to severity of illness assessment

and forecasting in icu with sparse, heterogeneous clinical data," *In Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, February.

[82] K. Bache and M. Lichman, "Uci machine learning repository," *University of California, School of Information and Computer Science, Irvine, CA*, 2013.

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

[84] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[85] C. Lim and M. McAleer, "Time series forecasts of international travel demand for australia," *Tourism Management*, vol. 23, no. 4, pp. 389–396, 2002.

[86] J. B. Leverenz, J. F. Quinn, C. Zabetian, J. Zhang, K. S. Montine, and T. J. Montine, "Cognitive impairment and dementia in patients with parkinson disease," *Current Topics in Medicinal Chemistry*, vol. 9, no. 10, pp. 903–912, 2009.

[87] W. Poewe, S. Gauthier, D. Aarsland, J. Leverenz, P. Barone, D. Weintraub, E. Tolosa, and B. Dubois, "Diagnosis and management of parkinson's disease dementia," *International Journal of Clinical Practice*, vol. 62, no. 10, pp. 1581–1587, 2008.

[88] F. Boller, T. Mizutani, U. Roessmann, and P. Gambetti, "Parkinson disease, dementia, and alzheimer disease: clinicopathological correlations," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 7, no. 4, pp. 329–335, 1980.

[89] H. Byeon, "Application of machine learning technique to distinguish parkinson's disease dementia and alzheimer's dementia: Predictive power of parkinson's disease-related non-motor symptoms and neuropsychological profile," *Journal of Personalized Medicine*, vol. 10, no. 2, p. 31, 2020.

[90] L.-A. Leyland, F. D. Bremner, R. Mahmood, S. Hewitt, M. Durteste, M. R. Cartlidge, M. M.-M. Lai, L. E. Miller, A. P. Saygin, P. A. Keane, *et al.*, "Visual tests predict dementia risk in parkinson disease," *Neurology: Clinical Practice*, vol. 10, no. 1, pp. 29–39, 2020.

[91] P. Martinez-Martin, Y. M. Wan, K. Ray Chaudhuri, A. E. Schrag, and D. Weintraub, "Impulse control and related behaviors in parkinson's disease with dementia," *European Journal of Neurology*, vol. 27, no. 6, pp. 944–950, 2020.

[92] D. J. Irwin, M. T. White, J. B. Toledo, S. X. Xie, J. L. Robinson, V. Van Deerlin, V. M.-Y. Lee, J. B. Leverenz, T. J. Montine, J. E. Duda, *et al.*, "Neuropathologic substrates of parkinson disease dementia," *Annals of Neurology*, vol. 72, no. 4, pp. 587–598, 2012.

[93] R. Savica, B. R. Grossardt, J. H. Bower, B. F. Boeve, J. E. Ahlskog, and W. A. Rocca, "Incidence of dementia with lewy bodies and parkinson disease dementia," *JAMA Neurology*, vol. 70, no. 11, pp. 1396–1402, 2013.

[94] M. C. Campbell, P. S. Myers, A. J. Weigand, E. R. Foster, N. J. Cairns, J. J. Jackson, C. N. Lessov-Schlaggar, and J. S. Perlmutter, "Parkinson disease clinical subtypes: key features & clinical milestones," *Annals of Clinical and Translational Neurology*, vol. 7, no. 8, pp. 1272–1283, 2020.

[95] H. Byeon, "Is the random forest algorithm suitable for predicting parkinson's disease with mild cognitive impairment out of parkinson's disease with normal cognition?," *International Journal of Environmental Research and Public Health*, vol. 17, no. 7, p. 2594, 2020.

[96] R. Gautam and M. Sharma, "Prevalence and diagnosis of neurological disorders using different deep learning techniques: A meta-analysis," *Journal of Medical Systems*, vol. 44, no. 2, p. 49, 2020.

[97] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury, *et al.*, "The parkinson progression marker initiative (ppmi)," *Progress in Neurobiology*, vol. 95, no. 4, pp. 629–635, 2011.

[98] M. Salari, A. Chitsaz, M. Etemadifar, M. R. Najafi, O. Mirmosayyeb, M. Bemanalizadeh, F. Panahi, and H. Mirzajani, "Evaluation of non-motor symptoms and their impact on quality of life in patients with parkinson's disease, isfahan, iran," *Iranian Journal of Neurology*, vol. 16, no. 3, p. 118, 2017.

[99] J.-H. Chen, C.-T. Hong, D. Wu, W.-C. Chi, C.-F. Yen, H.-F. Liao, L. Chan, and T.-H. Liou, "Dementia-related functional disability in moderate to advanced parkinson's disease: assessment using the world health organization disability assessment schedule 2.0," *International Journal of Environmental Research and Public Health*, vol. 16, no. 12, p. 2230, 2019.

[100] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005.

[101] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Communications of the ACM*, vol. 63, no. 1, pp. 68–77, 2019.

[102] H. Jin, Q. Song, and X. Hu, "Auto-keras: An efficient neural architecture search system," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1946–1956, 2019.

[103] R. Yu and L. Shi, "A user-based taxonomy for deep learning visualization," *Visual Informatics*, vol. 2, no. 3, pp. 147–154, 2018.

[104] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*, pp. 818–833, Springer, 2014.

APPENDIX A

PUBLICATIONS

- Nur Hafieza Ismail, Mengnan Du, Ninghao Liu, Xia Hu, and Zhe He. A Deep Learning Approach for Identifying Cancer Survivors Living with Post-Traumatic Stress Disorder on Twitter. BMC Medical Informatics & Decision Making. In press.
- Ismail, N. H., Du, M., Martinez, D., & He, Z. (2019, September). Multivariate Multi-step Deep Learning Time Series Approach in Forecasting Parkinson's Disease Future Severity Progression. In Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (pp. 383-389).
- Ismail, N. H., Liu, N., Du, M., He, Z., & Hu, X. (2019). Identification of Cancer Survivors Living with PTSD on Social Media. Studies in health technology and informatics, 264, 1468.
- Ismail, N. H., Liu, N., Du, M., He, Z., & Hu, X. (2019). Using Deep Neural Network to Identify Cancer Survivors Living with Post-Traumatic Stress Disorder on Social Media.
- Ismail, N. H., Du, M., & Hu, X. (2019). Social Media and Psychological Disorder. In Social Web and Health Research (pp. 171-192). Springer, Cham.
- Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. Applied Mathematical Sciences, 9(129), 6415-6426.
- Aziz, A. A., Ismail, N. H., Ahmad, F., & Hassan, H. (2015). A FRAMEWORK FOR STUDENTS'ACADEMIC PERFORMANCE ANALYSIS USING NAÏVE BAYES CLASSIFIER. Jurnal Teknologi, 75(3).
- Aziz, A. A., & Ahmad, N. H. I. F. (2014, September). First semester computer science students' academic performances analysis by using data mining classification algorithms. In

Proceeding of the International Conference on Artificial Intelligence and Computer Science (AICS 2014) (pp. 15-16).

- Aziz, A. A., Ismail, N. H., & Ahmad, F. (2013). MINING STUDENTS' ACADEMIC PERFORMANCE. Journal of Theoretical & Applied Information Technology, 53(3).
- Ismail, N. H., Ahmad, F., & Aziz, A. A. (2013). Implementing WEKA as a Data Mining Tool to Analyze Students' Academic Performances Using Naïve Bayes Classifier. In UniSZA Postgraduate Research Conference.