Efficient Detection of Mild Cognitive Impairment and Alzheimer's Disease from Brain Signal Data

Ashik Mostafa Alvi

Thesis submitted for the fulfilment of the requirements for the degree of

Doctor of Philosophy

Victoria University, Melbourne, Australia

Institute for Sustainable Industries and Liveable Cities

April, 2023

Abstract

Brain signal data are recordings of the electrical activity of the brain made using the electroencephalography (EEG). EEG is considered the future of neuroscience, as it has emerged as the latest gold standard for detecting neurological disorders such as dementia, mild cognitive impairment (MCI), Alzheimer's disease (AD), Parkinson's disease, schizophrenia, epilepsy, and so on. Due to its cost-effectiveness and portability, EEG is becoming the first choice when it comes to analyses and the detection of neuro-degenerative disorders. In addition, it has been widely accepted that analyzing EEG data is the better method for solving the challenge of learning about the brain's dynamics. EEG measures the electrical activity of the brain in real-time and can provide valuable information about brain function and dysfunction. The ability to analyze EEG data is crucial for improving our knowledge of cognitive processes and aiding in the identification of brain illnesses.

To discover the most recent trends and gaps in the EEG research, a lot of scholarly articles have been studied. However, the following research problems are still existing:

- Poor classification performances due to old and shallow traditional machine learning (TML) algorithms.
- Lack of efficient noise removal algorithms being used.
- Computationally expensive models are being proposed, which consume more power and time to diagnose brain diseases.
- Limit the diagnosis to detect one disorder, even though there may be more than one brain disorder with similar symptoms.
- Failed to accommodate a huge volume of EEG data.

In this dissertation, all five of the aforementioned research challenges have been answered. To address the aforementioned research issues, this dissertation focuses on the detection of AD and MCI among 600 neuro-diseases. The main objective of this dissertation is to develop computer-aided diagnostic methods for effectively detecting both AD and MCI handling a large volume of EEG data with improved performance and cost-effective deep learning (DL) models.

For the purpose of diagnosing AD and MCI from healthy volunteers (HVs), I have developed five DL-based frameworks using EEG data. The following are the innovations achieved at the end of this extensive work:

- ✓ Twenty different LSTM models with varying parameter configurations were analysed in the first investigation, with the best one in terms of performance being recommended.
- ✓ The first GRU examination, in which a newly developed adaptive noise reduction and compression technique has also been employed, outperformed the performance of the LSTM effort.
- ✓ The second suggested GRU approach, which is lightweight and computationally cheap, has undergone further enhancements, and the MCI classification accuracy has climbed to a new level.
- ✓ The Deep Residual Network has been used in our proposed deep residual Alzheimer's disease and MCI detection network architecture. This study is the first multi-class breakthrough in this field.
- ✓ Eventually, the Cognitive Decline Recognition Network framework, a specially developed convolutional neural network, has enabled multi-class AD-MCI detection to reach a new level of accuracy of over 99% while accommodating a huge volume of EEG data.

The outcomes of the experiments have shown how successful the techniques are in detecting MCI and AD. Additionally, technology experts and neuro researchers may find this work useful in creating a new automatic diagnosis system for AD-MCI detection.

Declaration of Authenticity

I, **Ashik Mostafa Alvi**, declare that the PhD thesis entitled "**Efficient Detection of Mild Cognitive Impairment and Alzheimer's Diseases from Brain Signal Data**" is no more than 80,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

I have conducted my research in alignment with the Australian Code for the Responsible Conduct of Research and Victoria University's Higher Degree by Research Policy and Procedures.

Signature

Date: 24/04/2023

ACKNOWLEDGEMENTS

Credit should be given where it is due. The credit for my ability to pursue higher education, which has been a wonderful privilege, must first and foremost go to the All-Powerful Allah. We often say, "Allah is the best of planners, and whatever He does, it is for our best." Surely, I could not have asked for more than what the almighty has given me. May Allah bless me with more success and patience in the future.

I would like to start showing my gratitude to my principal supervisor, Professor Hua Wang, for being a wonderful supportive supervisor who never fails to encourage and motivate us. I remember there were no high-spec computing devices available in our institute, and I was badly in need of a high-spec machine for my work. When I informed my principal supervisor about this need, he immediately requested that the college provide a budget to purchase a supercomputer. I really appreciate your advice on reading at least two excellent articles each week. This has really shaped my writing and thinking capabilities. I also want to thank you for organising those refreshment BBQ parties and outside work hangouts. Keeping our mental health in tip-top form throughout COVID-19 was made possible by your weekly meetings. Nevertheless, I sincerely appreciate you sharing your expertise and experience with me while I finish my thesis. Without your generosity, openness, and assistance, the path may not be as even. You are the one who has greatly influenced me to do excellent work and publish in prestigious D1 publications. I want to express my sincere gratitude for helping me become who I am today.

I owe a lot to Dr. Siuly Siuly. Thank you, madam, for being there for me from the very first days of my Melbourne life. Apart from the professional activities, you have supported me like family. You're so involved in my work and have guided me throughout the whole journey. I appreciate your kindness, patience, support, and graciousness towards me. I highly acknowledge your efforts towards making this thesis a success.

I also like to recognise Dr. Frank Whittaker, our industry partner, Mr. Md Nurul Ahad Tawhid, my VU colleague, and Dr. Kate Wang, who contributed to our effort. This study was made possible by a Linkage Scholarship from the Australian Research Council (ARC). I am grateful to the ARC for such an opportunity and to Victoria University, particularly Associate Professor Randall W. Robinson, Dr. Khandakar Ahmed, Palmina

Fichera, Jo Xuereb, Trish Dwyer, Debra Fitzpatrick, and PariSima Nassirnia, for their support during my VU life.

To my family, I must now express my gratitude. My father, who was always there for my study needs, was my biggest source of academic support and encouragement throughout my education. Whether I needed help with homework, guidance on a challenging subject, or simply someone to talk to about my academic goals and aspirations, you're always there to offer his guidance and support. Never discouraged me to take the next step and lost my trust; you did every possible thing from his end. You are my hero, the reason I am here, and deserves praise for everything I have achieved. I am incredibly grateful for everything he has done for me and for the role he played in shaping my academic success. My mother, who has made significant sacrifices to make sure I get the very best, is next in line to be appreciated. You are my powerhouse and prayer source. I've heard it stated that these two people have given me strength and realism. Last but not least, thank you to my wonderful wife, who has been so encouraging and impressive during the past few years of my PhD path. I really appreciate your contributions, and it is because to your support that this timely submission was made possible. Finally, I'd want to express my gratitude to my sister for looking after my parents while I was abroad pursuing my degree. A special thanks to my high school math teacher, Mr. Nantu Ranjan Biswas, for building my interest in science and developing my basics. Thanks to everyone, and please keep loving and supporting me like always.

DEDICATION

This thesis is dedicated to my parents for their unconditional love and support, and to my wife for always being there for me. Thanks to everyone; this one is for you.

LIST OF ABBREVIATIONS

Abbreviation	Full Form
AD	Alzheimer's disease
ADI	Alzheimer 's Disease International
AI	Artificial Intelligence
AR	Auto-Regressive
BLSTM	Bidirectional Long Short-Term Memory
BM	Boltzmann Machine
BS	Batch Size
CDR-Net	Cognitive Decline Recognition Network
CFV	Candidate Feature Vector
CNN	Convolutional Neural Network
CPT	Continuous Performance Test
CSV	Comma-Separated Value
CT scans	Computed Tomography scans
CWT	Continuous Wavelet Transform
DCssCDBM	Discriminative Convolutional High-Order Boltzmann Machine
DF	Dominant Frequency
DFT	Discrete Furrier Transform
DL	Deep Learning
dNAT	Deep Neuronal Activity Topography
DNN	Deep Neural Network
DRAM-Net	Deep Residual Alzheimer's Disease and MCI Detection Network
DT	Decision Tree
DRN	Deep Residual Network
DWT	Discrete Wavelet Transform
EDF	European Data Format
EEG	Electroencephalography
ELM	Extreme Learning Machine
EPNN	Enhanced Probabilistic Neural Network
ERP	Event-Related Potential

EWT	Empirical Wavelet Transform
FD	Fractal Dimension
FFT	Fast Furrier Transformation
FKNN	Fuzzy K-Nearest Neighbor
FN	False Negative
FP	False Positive
FSVM	Fuzzy Support Vector Machine
FT	Furrier Transformation
GRU	Gated Recurrent Unit
HE	Hurst Exponent
HMM	Hidden Markov Model
HOS	Higher Order Spectra
HT	Hilbert Transform
HV	Healthy Volunteer
ICA	Independent Component Analysis
KNN	K-Nearest Neighbour
LDA	Linear Discriminant Analysis
LOOCV	Leave-One-Out Cross Validations
LR	Logistic Regression
LSTM	Long Short-Term Memory
MCI	Mild Cognitive Impairment
MEG	Magnetoencephalography
ML	Machine Learning
MLP	Multi-Layer Perceptron
MMSE	Mini-Mental State Examination
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
NB	Naïve Bayes
NN	Neural Network
PAA	Piecewise Aggregate Approximation
PCA	Principal Component Analysis
PD	Parkinson’s disease
PDC	Partial Directed Coherence

PE	Permutation Entropy
PET	Positron Emission Tomography
PLV	Phase Locking Value
PNN	Probabilistic Neural Network
PSNR	Peak Signal-to-Noise Ratio
RF	Random Forest
RNN	Recurrent Neural Network
RP	Relative Power
sLORETA	Standardized Low-Resolution Brain Electromagnetic Tomography
SVM	Support Vector Machine
SVMrbf	Support Vector Machine with Radial Basis Function
SWT	Stationary Wavelet Transformation
SZ	Schizophrenia
TML	Traditional Machine Learning
TN	True Negative
TP	True Positive
WHO	World Health Organization
wICA	Wavelet Enhanced Independent Component Analysis
WT	Wavelet Transform

LIST OF PUBLICATIONS

Based on the extensive works, the following papers have been either published or submitted for publication in international journals and conference proceedings.

Journal Articles

1. Alvi, A. M., Siuly, S., & Wang, H. (2022). A long short-term memory based framework for early detection of mild cognitive impairment from EEG signals. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(2), 375-388.
2. Alvi, A. M., Siuly, S., Wang, H., Wang, K., & Whittaker, F. (2022). A deep learning based framework for diagnosis of mild cognitive impairment. *Knowledge-Based Systems*, 248, 108815.
3. Alvi, A. M., Siuly, S., & Wang, H. (2022). Neurological abnormality detection from electroencephalography data: a review. *Artificial Intelligence Review*, 55(3), 2275-2312.
4. Alvi, A. M., Siuly, S., De-Cola, M. C., Wang, H. & Whittaker, F. (2023). CDR-Net: A Computerized Framework to Detect Alzheimer's Diseases and Mild Cognitive Impairment. *IEEE Transactions on Cognitive and Developmental Systems* (Under Review)

International Conference Articles

1. Alvi, A. M., Siuly, S., De Cola, M. C., & Wang, H. (2022, October). DRAM-Net: A Deep Residual Alzheimer's Diseases and Mild Cognitive Impairment Detection Network Using EEG Data. In *Health Information Science: 11th International Conference, HIS 2022, Virtual Event, October 28–30, 2022, Proceedings* (pp. 42-53). Cham: Springer Nature Switzerland.
2. Alvi, A. M., Siuly, S., Wang, H., Sun, L., & Cao, J. (2020). An adaptive image smoothing technique based on localization. In *Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference (FLINS 2020)* (pp. 866-873).
3. Alvi, A. M., Siuly, S., & Wang, H. (2022, August). Challenges in electroencephalography data processing using machine learning approaches. In *Databases Theory and Applications: 33rd Australasian Database Conference, ADC*

2022, Sydney, NSW, Australia, September 2–4, 2022, Proceedings (pp. 177-184). Cham: Springer International Publishing.

4. Alvi, A. M., Siuly, S., & Wang, H. (2022, January). Developing a deep learning based approach for anomalies detection from EEG data. In *Web Information Systems Engineering–WISE 2021: 22nd International Conference on Web Information Systems Engineering, WISE 2021, Melbourne, VIC, Australia, October 26–29, 2021, Proceedings, Part I* (pp. 591-602). Cham: Springer International Publishing.

DETAILS OF INCLUDED PAPERS: THESIS WITH PUBLICATION

Please list details of each scholarly publication and/or manuscript included in the thesis submission. Copies of published scholarly publications and/or manuscripts submitted and/or final draft manuscripts should also be included in the thesis submission.

This table must be incorporated in the thesis before the Table of Contents.

Chapter No.	Publication Title	Publication Status • Published • Accepted for publication • In revised and resubmit stage • Under review • Manuscript ready for submission	Publication Details • Citation, if published • Title, Journal, Date of acceptance letter and Corresponding editor's email address • Title, Journal, Date of submission
2	Neurological abnormality detection from electroencephalography data: a review	Published	Alvi, A. M., Siuly, S., & Wang, H. (2022). Neurological abnormality detection from electroencephalography data: a review. Artificial Intelligence Review, 55(3), 2275-2312.
2	Challenges in electroencephalography data processing using machine learning approaches	Published	Alvi, A. M., Siuly, S., & Wang, H. (2022, August). Challenges in electroencephalography data processing using machine learning approaches. In Databases Theory and Applications: 33rd Australasian Database Conference, ADC 2022, Sydney, NSW
3	A long short-term memory based framework for early detection of mild cognitive impairment from EEG signals	Published	Alvi, A. M., Siuly, S., & Wang, H. (2022). A long short-term memory based framework for early detection of mild cognitive impairment from EEG signals. IEEE Transactions on Emerging Topics in Computational Intelligence
4	An adaptive image smoothing technique based on localization	Published	Alvi, A. M., Siuly, S., Wang, H., Sun, L., & Cao, J. (2020). An adaptive image smoothing technique based on localization. In Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS
4	Developing a deep learning based approach for anomalies detection from EEG data	Published	Alvi, A. M., Siuly, S., & Wang, H. (2022, January). Developing a deep learning based approach for anomalies detection from EEG data. In Web Information Systems Engineering–WISE 2021: 22nd International Conference on Web Information Systems Engineering
5	A deep learning based framework for diagnosis of mild cognitive impairment	Published	Alvi, A. M., Siuly, S., Wang, H., Wang, K., & Whittaker, F. (2022). A deep learning based framework for diagnosis of mild cognitive impairment. Knowledge-Based Systems, 248, 108815.

Declaration by [candidate name]:

ASHIK MOSTAFA ALVI

Signature: Ashik Mostafa Alvi

Digitally signed by Ashik Mostafa Alvi
Date: 2023.04.20 17:48:04 +10'00'

Date: 20/04/2023

DETAILS OF INCLUDED PAPERS: THESIS WITH PUBLICATION

Please list details of each scholarly publication and/or manuscript included in the thesis submission. Copies of published scholarly publications and/or manuscripts submitted and/or final draft manuscripts should also be included in the thesis submission.

This table must be incorporated in the thesis before the Table of Contents.

Chapter No.	Publication Title	Publication Status • Published • Accepted for publication • In revised and resubmit stage • Under review • Manuscript ready for submission	Publication Details • Citation, if published • Title, Journal, Date of acceptance letter and Corresponding editor's email address • Title, Journal, Date of submission
6	DRAM-Net: A Deep Residual Alzheimer's Diseases and Mild Cognitive Impairment Detection Network Using EEG Data	Published	Alvi, A. M., Siuly, S., De Cola, M. C., & Wang, H. (2022, October). DRAM-Net: A Deep Residual Alzheimer's Diseases and Mild Cognitive Impairment Detection Network Using EEG Data. In Health Information Science: 11th International Conference, HIS 2022
7	CDR-Net: A Computerized Framework to Detect Alzheimer's Diseases and Mild Cognitive Impairment	Under Review	CDR-Net: A Computerized Framework to Detect Alzheimer's Diseases and Mild Cognitive Impairment, IEEE Transactions on Cognitive and Developmental Systems, 13th March, 2023

Declaration by [candidate name]: ASHIK MOSTAFA ALVI

Signature: Ashik Mostafa Alvi Digitally signed by Ashik Mostafa Alvi Date: 2023.04.20 18:53:29 +10'00'

Date: 20/04/2023

PRIVACY STATEMENT Victoria University (VU) values your privacy and is committed to handling your personal information in accordance with the Privacy and Data Protection Act 2014 (Vic) and other applicable privacy legislation. The personal information collected on this form will be used primarily for the purposes of assessing and processing this application. VU may also use and disclose your personal information to verify the information provided by you, to comply with government and other reporting requirements and/or to carry out associated activities connected with this application. Your personal information may also be disclosed to Commonwealth and State agencies such as the Department of Education and Training and the Department of Home Affairs in accordance with VU's obligations under the Education Services for Overseas Students Act 2000 (Cth) (ESOS Act), the National Code of Practice for Providers of Education and Training to Overseas Students 2018 (National Code) and other applicable legislation. Your personal information will not otherwise be used or disclosed without your consent, unless permitted by law. By completing and submitting this application, you agree to VU collecting, using and disclosing your personal information as described above and in accordance with VU's Privacy Policy and Student Information Privacy Collection Statement (which provides further detail about the types of personal information VU may collect from you and how it is managed) available on the Privacy page on our website vu.edu.au/privacy You have a right to access your personal information n held by VU. If you have any questions regarding privacy, please refer to the Privacy page on our website, our frequently asked questions at ASKVU or phone us on 9919 6100 or 1300 VIC UNI (or 1300 842 864).

PRIVACY INFORMATION: We collect and protect your personal information in accordance with our Privacy Policy vu.edu.au/privacy.

Victoria University CRICOS Provider No. 00124K (Melbourne) and CRICOS Provider No. 02475D (Sydney). RTO Code: 3113. ABN: 83 776 954 731

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 : INTRODUCTION

1.1 Overview and Motivation

The future of neuroscience lies in electroencephalography (EEG). EEG is the latest gold standard for diagnosing most neurological disorders like dementia, mild cognitive impairment (MCI), Alzheimer's disease (AD), and so on. It is a cheap, portable, and non-invasive option to discover neuro-disorders compared to the remaining expensive and time consuming options like computed tomography (CT) scans, positron emission tomography (PET), mini-mental state examinations, and magnetic resonance imaging (MRI). Though EEG sounds like a promising option, there are some challenges involved in EEG signal processing, from EEG signal recording to disease classification. This chapter has introduced the topic and its background.

The brain is the most perplexing body part of a human, and most of the brain-related diseases remain undiscovered until they reach a dangerous level. Dysfunction and death of the brain cells are the main causes of neurodegenerative diseases and disorders. Brain abnormalities include more than 600 disorders [92], including AD, epilepsy [92], [3], stroke, dementia, traumatic disorders, MCI, schizophrenia (SZ), Parkinson disease (PD), brain tumours, migraine, multiple sclerosis, and autism [80]. The brain, nerve roots, spinal cord, peripheral nerves, neuromuscular junction, cranial nerves, autonomic nervous system, and muscles are influenced by these neuro-disorders. Being impartial about countries, sex, age, income or education, these disorders have effect on people. The continuous development in the medical science and technology sectors has increased the life expectancy of humans, and it is very difficult to stay aside without the touch of technology [8], [37], [10]. More facilitations in the health care sector are essential to confirm the life expectancy of the people. However, increased life anticipation leads to an extensive possibility of being affected by age-related neurodegenerative diseases, which seriously influence our regular lives. Influenced people and their families are being deprived of having a quality life because of these neurological disorders.

According to the World Health Organisation's (WHO), nearly one billion individuals worldwide were afflicted with brain illnesses such as Alzheimer's disease, epilepsy, stroke, and headache [106]. This indicates that nearly 1 in 6 of the world's

population suffers from neurological disorders. As stated by the WHO, 6.8 million people die every year as a result of neurodegenerative diseases, and these disorders affect hundreds of millions of people worldwide: globally, around 47 million have dementia [92]; more than 50 million are affected by epilepsy [97]. Among the brain-related disorders, different types of dementia are the most common. These are very common neuro-diseases for the elderly community who are 65 years old or more, and the severity rate increases exponentially with age [62], [30], [71]. Every year, there are 7.7 million new dementia cases. According to the Australian Bureau of Statistics, the amount of deaths owing to AD and dementia have increased by 32.0% from 2009 to 2013, and these brain diseases are the second-leading cause of death in Australia [1]. There are many neuro-disorders, but this study has picked the top five: MCI, dementia, AD, schizophrenia, and Parkinson's disease (PD), as they are all major public health issues. These disorders are usually diagnosed using electroencephalography data. The recent studies of the aforementioned five neuro-disorders are reported in this chapter.

There are several techniques for the diagnosis of MCI, dementia, AD, SZ, and PD, such as positron emission tomography (PET), magnetoencephalography (MEG), magnetic resonance imaging (MRI), and computed tomography (CT) scan [80]. These techniques are very time-consuming and costly. There is another technique called mini-mental state examination (individual interview method) to measure the severity of the disease. On the other hand, electroencephalography (EEG) is the most popular tool to investigate the presence of neuro-disorder biomarkers [81], [4], and [83]. In recent years, EEG has become a better cost-effective and time-efficient alternative for detecting neuro-diseases like dementia, MCI, SZ, PD, and AD in healthy volunteers (HVs). The EEG records the electrical activities in the cerebral cortex with respect to time. There are electrodes that are placed on the cranium to collect the electric potentials. The International 10-20 system is the most common layout of EEG, having around 21 electrodes [89]. 10-10 and 10-5 are some higher-density layouts with 64 and 128 electrodes, respectively, that are also used sometimes [48].

For the last two decades, EEG has been used as a promising tool for diagnosing neurodegenerative diseases. It is a period impressible biomarker tool that is mostly used in detecting cortical disorders like autism spectrum disorders [95], Huntington's disease [68], epilepsy and seizure [38], [79], [49], [90], [102], [43], [93], Parkinson’s disease [61], [101], cerebral ischemia [60], and dementia [14], [29], [6]. When analysing the EEG

signal, it has five frequency bands: delta (δ) from 0.1 Hz to 4 Hz, theta (θ) from 4 Hz to 8 Hz, alpha (α) from 8 Hz to 12 Hz, beta (β) from 12 Hz to 30 Hz, and gamma (γ) > 30 Hz. A relative power content comparison among HC, MCI, and AD subjects of different frequency bands is shown in **Fig. 1.1**.

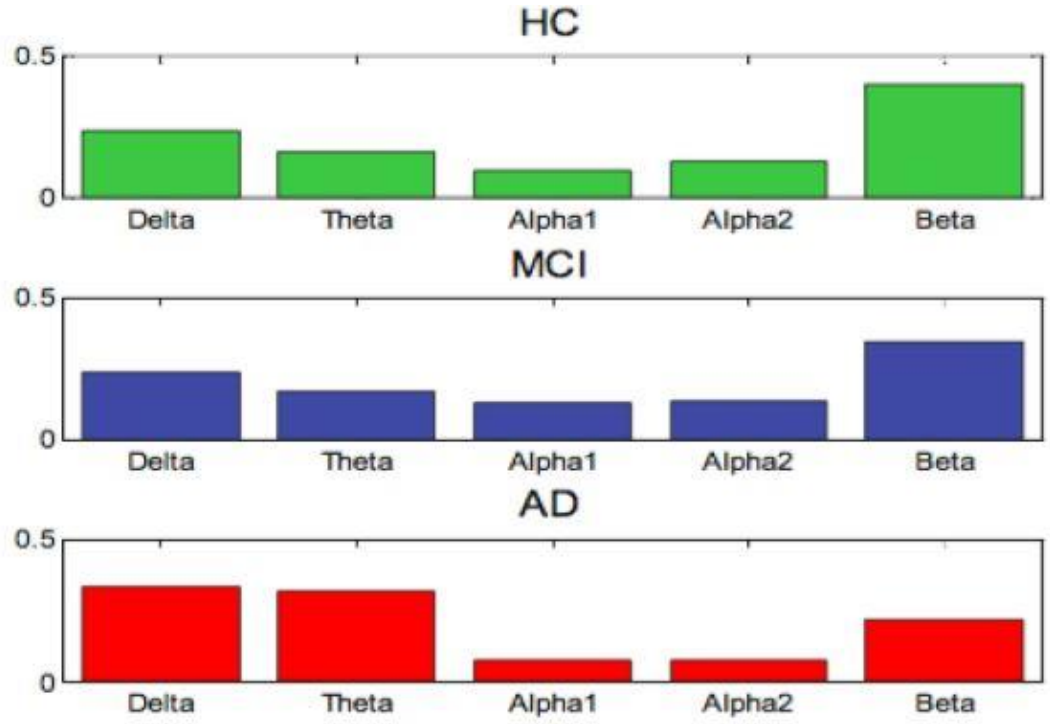

Fig. 1.1: Relative comparison on power content of the frequency bands averaged [55].

Functional alteration is mirrored in EEG signals; hence, it is a very useful clinical tool to detect neuronal degeneration. Because of its promising performance and cost effectiveness, EEG has become more popular compared to other existing medical tools like PET, MRI, and MEG. MEG has a limitation in localising the position of occurrence in the brain, and it is also very expensive. EEG has shown better temporal commitment [92] than other neuroimaging methods [91], [82].

In addition, EEG recording is very sensitive to artifacts such as body movement, electricity grid inference, eye blinks, heartbeats, etc. These drawbacks are one of the main reasons for not detecting neurodegenerative disorders at an early stage. For detecting neuro-disorders from EEG signals, there are few common components of EEG signal pre-processing: removing noises from the EEG data [7], [9], compression of large volume EEG data, segmentation, down/up sampling the sampling frequency. The raw EEG data contains artifacts and other noises. Therefore, the first step begins with removing noise from the recorded EEG data. After separating the signal from the noise, the filtered signal

was then passed through some time, frequency, or time-frequency algorithms to extract the meaningful attributes. Finally, the pulled out attributes are employed as inputs to a classification model to detect the neuro-disorders.

A rigors amount of machine learning (ML) techniques [65], [66] have been used in a number of studies to identify these diseases early on. And it is an ongoing area of research to discover neurological disorders at the quickest possible time with EEG data. But this is a cross-field area of research that evolves both signal processing and ML together. That creates some extra challenges for the researchers. Raw EEG signals have to be processed using time- or frequency-domain methods to get effective and important features out of them. Finally, a promising classifier with adequate parameters to differentiate those disorders from HVs is essential.

1.2 Problem Statements

To determine the most recent trends and gaps in EEG research, an extensive amount of scientific papers are reviewed. The research problems along with the suggested solutions that are uncovered after examining prior EEG efforts are reported below:

Problem 1: Noise removal is an integral part of every research project that evolved from raw data collection. While examining earlier EEG attempts, I have discovered that the studies [16], [17], [28], [42], [45], [75] that did not work well either did not de-noise the EEG data correctly or the de-noising efforts were not good enough to clean the raw EEG data. As a result, I have identified this noise reduction issue as one of the primary shortcomings in earlier EEG studies.

Solution: In this dissertation, I have developed an adaptive de-noiser approach that is discussed in **Chapter 4** and used it to compress the EEG data as well. **Chapter 4** reflects the solution to this **research problem #1**.

Problem 2: In order to improve the performance of the suggested models, prior attempts [13], [32], [84], [100] utilized computationally expensive models, which need more power and longer time to diagnose. Since not everyone has access to a powerful computer, I identified this as a further area in need of investigation.

Solution: Six models are included in this long study, where I have provided four lightweight and computationally affordable models and discussed and compared their

temporal complexity. The answer to this **research problem #2** is presented in **Chapters 4–7**.

Problem 3: Numerous brain illnesses have symptoms that are quite similar. Most of the prior EEG initiatives [13], [32], [70], [86], [157], [172] resolved binary classification issues due to a lack of data, computational assistance, and diagnostic expertise. It was either MCI identification from HVs or AD identification from HVs. Whereas MCI is treated as the preliminary stage of AD, and they share some similar symptoms. Lack of multi-classifiers that will be able to differentiate between AD, MCI, and normal cognitive function simultaneously poses a significant obstacle in achieving precise and comprehensive diagnostic capabilities for cognitive disorders. Another limitation I have realized is this.

Solution: The innovative and novel AD-MCI- HV detection frameworks I have presented in **Chapters 6 and 7** using an EEG dataset of 109 participants established a new standard for multi-class AD-MCI-HV detection performance which has resolved the **research problem #3**.

Problem 4: The effectiveness of earlier EEG attempts [32], [52], [58], [70], [75], [121], [134], [135], [138], [155], [172] has also been hampered by the use of traditional ML methods. The shallow and simple design of such ML techniques, such as support vector machine (SVM), K-nearest neighbor (KNN), decision tree (DT), Naïve Bayes (NB), etc., prevented them from learning characteristics and extracting features from complicated EEG data. It is one of the significant challenges in cutting-edge EEG research that has made it difficult to develop an appropriate model.

Solution: In this dissertation, I have proposed six new deep learning (DL)-based frameworks using EEG data. Since the outset, I have avoided using traditional ML approaches and have shown via comparisons that DL techniques surpass such classical ML techniques. The answer to this **research problem #4** may be found in **Chapters 3 through 7**.

Problem 5: Performance is an important factor in classification problems. Since I am working with medical data, there should be no room for error. Previous suggested strategies [13], [32], [52], [58], [70], [75], [86], [125], [133], [134], [135], [173] for diagnosing brain illnesses like AD or MCI didn't provide enough insight.

Solution: All previous performance records have already been broken and a new benchmark for AD-MCI detection have been established by the six initiatives suggested in this dissertation. Evidence for solutions to **research problem #5** is reported in **Chapters 3–7**.

1.3 Objectives

The research reported in this dissertation concentrates on how various EEG signals from various brain activity might be categorized to assess various brain illnesses. In this dissertation, I have created a total of six techniques for processing and categorizing massive amounts of EEG data into AD, MCI, and HV categories. Each of the research that have been described outperforms previously published efforts when compared to current state-of-the-art techniques. Even I have made attempts to surpass our previous records.

I arrived at the six methodologies in this dissertation as a result of the specified research challenges. Three main goals for the analysis and categorization of EEG signals have been established after looking into the research challenges. The following are our main objectives:

1) Develop a method for noise removal and handle the huge volume of EEG data.
2) Create frameworks for MCI detection and increase the bar in terms of performance.
3) Introduce a lightweight and inexpensive model to compute for identifying neurological disorders from EEG data.
4) Deliver multi-class (AD, MCI, and HV) models for efficient neuro-disorder detection using brain signal data.
5) Study different DL methods, particularly variants of recurrent neural network (RNN) and convolutional neural network (CNN) models, to build frameworks for AD and MCI detection in HVs.

With regard to the aforementioned goals, I am fortunate to have created six innovative techniques. Three frameworks for MCI detection have been established in this dissertation. These MCI recognition frameworks were all built with DL algorithms. Every one of them uses a novel approach in the EEG pre-processing phase. The three platforms that determine MCI are listed below:

- A long short-term memory (LSTM) based framework for early detection of mild cognitive impairment from EEG signals (MCI-LSTM)
- Developing a deep learning based approach for anomalies detection from EEG data (MCI-GRU along with Method 1)
- A deep learning based framework for diagnosis of mild cognitive impairment (MCI-GRU)

Our earlier suggested methods (1-4) were created using an EEG dataset that only included MCI and HVs. As a result, I was unable to examine AD. While this was going on, I worked with the IRCCS Centro Neurolesi "Bonino-Pulejo" and acquired a large EEG dataset that included people with AD, MCI, and normal physiology. I developed two novel frameworks using this multiclass EEG dataset that outperform earlier multiclass attempts. The two multi-class AD-MCI identification frameworks that I have suggested are listed below:

- DRAM-Net: A Deep Residual Alzheimer's Diseases and Mild Cognitive Impairment Detection Network Using EEG Data
- CDR-Net: A Computerized Framework to Detect Alzheimer's Diseases and Mild Cognitive Impairment

Below is a quick outline of their accomplishments:

1.3.1 Develop a method for noise removal and handle the huge volume of EEG data

Method 1(Adaptive De-noiser): In order to deal with artifacts and other noises, I have presented an adaptive noise reduction method that has been tested with digitized and static data such as color images in **Chapter 4**. For our objectives, I transformed the digitized EEG data (amplitude in μV) into image intensity values. Peak signal-to-noise ratio for our suggested adaptive strategy is 15.38 dB, exceeding the approaches currently in use. To down-sample EEG data for an MCI diagnosis system, I have adopted this technique.

1.3.2 Create frameworks for MCI detection and increase the bar in terms of performance

Method 2 (MCI-LSTM): This is our first effort to use EEG data to solve the MCI detection challenge. I investigated RNNs and discovered that LSTM may work well with

EEG data. The 20 best LSTM models out of more than 35 examined have been included in the study reported in **Chapter 3**. I have put forwards the most effective LSTM-based framework with four steps for detecting MCIs in HVs after an extensive examination. The first phase included gathering and describing EEG data; the second involved de-noising using the Butterworth filter, segmenting the data to produce 300 new temporal segments, and down-sampling with the Average filter. After that, the two-layer LSTM model, which includes 1024 and 512 neurons, is provided with the processed EEG data to categorize. The performance of the suggested framework is assessed in the final stage. The best model scored 96.41% accuracy, 96.55% sensitivity, and 95.95% specificity during 5-fold cross-validation.

1.3.3 Introduce a lightweight and inexpensive model to compute for identifying neurological disorders from EEG data

Method 3(MCI-GRU along with Method 1): This time, I have explored the alternative RNN subtype, gated recurrent unit (GRU). While GRU and LSTM both have comparable features, GRU is lighter and comes with a reset gate. Four steps make up our suggested GRU-based MCI detection framework (**Chapter 4**): EEG data collection, pre-processing (de-noising using the Butterworth filter, 6-second segmentation of each subject's EEG recording, and down-sampling using the suggested method 1), classification using the GRU network, and outcome assessment. Two hidden, deep layers as well as a dense layer make up our suggested GRU network. The GRU network's dense layer included a single neuron, whereas the first hidden layer contained 1024 neurons and the second, 512. On a publicly accessible EEG dataset, our suggested model performed with 96.91% accuracy, 97.95% sensitivity, 96.16% specificity, and a 96.39% F1 score.

Method 4(MCI-GRU): This is a thorough and in-depth study that combines LSTM and GRU with other traditional ML techniques like SVM and KNN. Gathering raw EEG data, pre-processing it (de-noising using stationary wavelet transformation, 6 second segmentation, and down-sampling using an average filter), identifying hidden features and differentiating MCI subjects from HVs, and performance testing of the suggested structure make up the proposed framework. The purpose of this MCI research (**Chapter 5**) was to identify MCI participants by using distinguishing characteristics of deep learning. Neither a separate approach for feature extraction nor one for classification is utilised. SVM and KNN are unable to extract features on their own, whereas LSTM

and GRU can. Therefore, in order to lower the computational cost, I have employed the LSTM-extracted features for both SVM and KNN. GRU has distinguished itself and earned the best performance out of the four classifiers that were reported.

1.3.4 Deliver multi-class (AD, MCI, and HV) models for efficient neuro-disorder detection using brain signal data

Method 5(DRAM-Net): Our first response to the multiclass AD-MCI challenge is the Deep Residual Alzheimer's Disease and MCI Detection Network (DRAM-Net). A DL-based classifier known as the deep residual network is used to construct this system. In our suggested DRAM-Net design (**Chapter 6**), there are four residual blocks. This multi-class research includes steps for collecting EEG data, pre-processing (down-sampling to 256 Hz frequency, noise reduction using stationary wavelet transform (SWT), and 5-second temporal segmentation), DRAM-Net architecture to identify AD, MCI, and HVs, and experiment assessment. In this study, I have reported performance by class. In addition to exceeding other multi-class studies, our suggested DRAM-Net architecture has claimed accuracy of 96.66% for the HV class, 98.06% for the MCI class, and 97.79% for the AD class. It has also achieved an overall multiclass accuracy of 96.26%.

1.3.5 Study different DL methods, particularly variants of RNNs and CNN models, to build frameworks for AD and MCI detection in HVs

Method 6(CDR-Net): In order to develop a framework that would include the answers to all the research questions I have raised, I have further examined the multiclass EEG dataset. The Cognitive Decline Recognition Network (CDR-Net) architecture has been constructed and reported in **Chapter 7** to recognize MCIs, ADs, and HVs employing EEG data in order to get around the research challenges. The proposed CDR-Net architecture enables the acquisition of EEG data, data pre-processing (down-sampling from 1024 Hz to 256 Hz, noise cleaning using SWT, 5-second segmentation, and 8-bit digital picture construction), feature extraction and classification using CDR-Net, as well as performance evaluation and cross-validation stages. The ultimate part of the CDR-Net architecture is a custom designed multi-layer CNN model with a *softmax* classifier. Better multiclass accuracy, sensitivity, and specificity were achieved using our recommended CDR-Net design, with values of 99.25%, 99.13%, and 99.32%, respectively. Concerns

about stability, consistency, and data over- and under-fitting are alleviated by performing 10-fold and leave-one-out cross validations.

1.4 Contributions

To accomplish our goals in this dissertation, I endured really hard. In essence, I am successful in contributing the following to the completion of this dissertation:

- The construction of an LSTM-based deep learning system for rapid identification of MCI in this dissertation is the first to employ EEG signal data as input.
- Twenty LSTM models have been evaluated to determine the optimal LSTM version for MCI detection.
- I have released the first report of an EEG classification research using deep learning techniques based on GRU.
- To reduce the computational burden of the proposed frameworks, I developed adaptive filtering as a down-sampling strategy.
- To save computational costs, I employed the LSTM-extracted features for SVM and KNN rather than employing different feature extraction techniques.
- I have presented a deep residual network that is specifically made for AD-MCI identification for the first time.
- To further aid in the processing of EEG data more quickly, I investigated the average filtering approach for down-sampling.
- Created and implemented a state-of-the-art, accurate, dependable, and efficient CDR-Net framework for identifying AD, MCI, and HVs using EEG data.
- Our investigations use 5 seconds temporal segments and 5 seconds of EEG recording is good enough to judge the patient’s case
- Improved classification performance on the same and distinct EEG datasets, both for multi-class and binary classifications, in comparison to older techniques.
- Our recommended frameworks have improved classification accuracy while requiring less computational effort.
- Performed 10-fold and LOOCV cross validations to check the consistency and stability of these suggested frameworks.
- Executed ablation experiments to identify the most appropriate classifier.

1.5 Structure of this Dissertation

This dissertation is divided into 8 chapters, each of which presents significant research results. The remainder of the dissertation is structured as follows:

Chapter 2 discusses contemporary EEG signal categorization techniques and their underlying knowledge. The underlying knowledge and the basics of EEG, and several neuro-disorders are introduced in this chapter. Following that, I described previously used techniques based on several domains. Additionally, a comparative outcome analysis with disease-specific performance comparisons is presented. This chapter concludes with a thorough assessment of prior efforts and a few outstanding research issues for the future.

Chapter 3 introduces our initial approach for MCI detection in this dissertation. It opens with an overview before introducing the context and earlier MCI research that are relevant. Following that, a framework for LSTM-based MCI detection has been described. Results of experiments are presented later. A very thorough comparative discussion and a essence come to an end in this chapter. It contains the answer to **research problems 4 and 5**.

Chapter 4 has integrated two studies together. I conducted two separate investigations, one involving the development of an adaptive de-noiser and the other using a GRU-based MCI detection framework that incorporates the adaptive de-noiser. An overview, an introduction to the subject, and a few associated MCI studies come first in this chapter. Afterwards, the suggested GRU-based MCI detection framework appears. This GRU-based MCI identification framework part discusses the recently created adaptive de-noiser. The results, discussion, and conclusion round up this chapter. **Research problems 1, 2, 4, and 5** have an answer in it.

In **Chapter 5**, a complete investigation employing the same prior EEG dataset and the GRU classifier is reported. Beginning with an overview, an introduction to the subject matter and research challenges, and background investigations. Following that, the suggested framework has been presented. The framework discussion section includes four classifiers and their respective talks. Investigations and their findings are then described, followed by a discussion of comparisons. An essence and some recommendations for succeeding study are provided in this chapter's conclusion. It provides answers to **research problems 2, 4, and 5**.

Chapter 6 has revealed a new era for AD-MCI detection. This research is a multiclass EEG study that includes patients with AD, MCI, and normal cognitive function. A summary of the chapter is presented at the beginning, and then the subject matter, research concerns, and related works are covered. The DRAM-Net framework has subsequently been presented. Then, some fascinating findings and discussions are presented. An essence and recommendations for further investigation are provided in this chapter's conclusion. **Research problems #2, 3, 4, and 5** are addressed inside this chapter.

Chapter 7 incorporates our last proposed method to resolve the MCI-AD identification problem using EEG data. A summary is given at the outset. It then includes a thorough introduction, background research, and a description of the proposed CDR-Net design. The investigations and findings from the study are then included. Following that, there has been a documented comparison. An essence and recommendations for further research come at the end of Chapter 7. Chapter 7 answers **research problems #2, 3, 4, and 5**.

Chapter 8 includes an overview and a contribution summary of this dissertation's work, and this dissertation concludes with a discussion of our shortcomings and suggestions for further study.

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: Neurological abnormality detection from electroencephalography data: a review

Surname: ALVI First name: ASHIK MOSTAFA

Institute: Institute for Sustainable Industries and Liveab Candidate's Con 60

Status:

Accepted and in press: ☐ Date:

Published: ☑ Date: 01/09

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be includ ned in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted] 24/01/2023

Signature **Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

Name(s) of Co-Author(s)	Contribution (%)	Nature of Contribution	Signature	Date
Ashik Mostafa Alvi	60	Conceptualization, Methodology, Data Curation, Investigation, Software,	[redacted]	24/01/2023
Siuly Sluly	20	Formal Analysis, Validation, Writing–Review and Editing, Supervision.		24/01/2023
Hua Wang	20	Resources, Supervision, Writing–Review and Editing, Funding Acquisition, Project		26/01/23

[redacted]

Updated: September 2019

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: Challenges in Electroencephalography Data Processing Using Machine Learning Approaches [redacted]

Surname: ALVI First name: ASHIK MOSTAFA

Institute: Institute for Sustainable Industries and Liveab Candidate's Con [redacted] 60

Status:

Accepted and in press: ☐ Date:

Published: ☑ Date: 27/08

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be includ [redacted] ned in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted] 24/01/2023

Signature **Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

Name(s) of Co-Author(s)	**Contribution (%)**	**Nature of Contribution**	**Signature**	**Date**
Ashik Mostafa Alvi	60	Conceptualization, Methodology, Data Curation, Investigation, Software,	[redacted]	24/01/2023
Siuly Siuly	20	Formal Analysis, Validation, Writing–Review and Editing, Supervision.	[redacted]	24/01/2023
Hua Wang	20	Resources, Supervision, Writing–Review and Editing, Funding Acquisition, Project	[redacted]	26/01/23

Updated: September 2019

CHAPTER 2 : LITERATURE REVIEW

2.1 Overview

The efficient detection of neurological abnormalities (disorders) is very important in clinical diagnosis for modern medical applications. As stated by the World Health Organization (WHO), brain diseases like Alzheimer's disease (AD), epilepsy, stroke, and headache infect almost one billion people globally. Electroencephalography (EEG) is the current reference standard for the diagnosis of most neurological diseases as it is inexpensive, bearable, and non-invasive compared to other tests (e.g., computed tomography, positron emission tomography, mini-mental state examination, and magnetic resonance imaging). In the preliminary stage, many studies are performed using EEG signals to detect neurodegenerative abnormalities. This chapter attempts to provide a comprehensive survey of the recent studies that have used EEG signals to detect the neurological diseases: dementia, mild cognitive impairment (MCI), AD, schizophrenia, and Parkinson's disease. This chapter focuses on the following key research questions: (1) what are the key components of EEG signal processing? (2) What algorithms have been used in this processing? (3) which signal processing techniques have received more attention? This chapter provides a clear description of the mentioned neuro-diseases along with the relevant studies. Moreover, this study presents all the recent efforts of the methods that are obtained at various steps of signal data processing, including the feature extraction and classification phases. Finally, an elaborated comparison of the existing efforts with their drawbacks and performance is reported. This will guide medical field researchers and technology experts to discover more accurate solutions for neuro-diseases and come up with a neurological abnormality detection framework.

The contents of this chapter have been published in the *Artificial Intelligence Review* [106] and also in the *Proceedings of Databases Theory and Applications: 33rd Australasian Database Conference, ADC 2022* [160].

2.2 Introduction

Human life expectancy has grown as a result of constant advancements in medical research and technology, and it is becoming increasingly impossible to live without being touched by technology [8], [10], [37]. To confirm people's life expectancy, more changes in the health-care industry are necessary. However, greater life expectancy increases the risk of developing age-related neurodegenerative illnesses [187], which have a significant impact on our daily lives.

The brain is the most complex part of the human body, and the majority of brain illnesses go unnoticed until they become serious. The primary cause of neurologic illnesses and disorders is the dysfunction and death of brain cells. There are more than 600 neuro-diseases, which include AD [44], MCI [107], [100], [104], [105], dementias, epilepsy [3], [92], [184], stroke, schizophrenia (SZ) [11], [85], Parkinson's disease (PD) [40], migraine, autism, brain tumours, etc. These neurological diseases have an impact on the brain, nerve roots, spinal cord, peripheral nerves, neuromuscular junction, cranial nerves, autonomic nervous system, and muscles [106].

According to the WHO, nearly one billion individuals worldwide were afflicted with brain illnesses such as AD, epilepsy, stroke, and headache [106]. This indicates that nearly 1 in 6 of the world's population suffers from neurological disorders. Regardless of country, gender, age, wealth, or education, these neuro-diseases affect individuals. WHO claims that every year, 6.8 million people die as a result of neurological illnesses, which impact hundreds of millions of people throughout the world: about 47 million people have dementia, and more than 50 million people have epilepsy [97]. These are highly frequent neuro-diseases in the senior population, especially those aged 65 and more, and the severity rate rises exponentially with age [30].

The existing biomarkers to detect these neuro-disorders are a computed tomography (CT) scan, magnetoencephalography (MEG), positron emission tomography (PET), magnetic resonance imaging (MRI), and a mini-mental state examination (a one-to-one interview procedure). These diagnostic tools are either expensive, time-consuming, or require manual methods. EEG has emerged as a non-invasive, portable, easy, inexpensive, and robust biomarker for investigating neurological disorders. The electroencephalogram captures the electrical activity in the cerebral cortex throughout time. To capture the electric potentials, electrodes are implanted on the skull. The most

common EEG configuration, with about 21 electrodes, is the International 10–20 system [89]. It is worthy of mention that I am dealing with more than 600 neuro-diseases in the current world. In our study, I have focused on the five common and deadliest neuro-disorder detection challenges, which are: dementias, MCI, AD, PD, and SZ. The following is a summary of the key contributions to this effort:

- Introduced 5 neuro-diseases and their recent studies
- A through comparative analysis has been done along with the research gaps
- All the narrow spikes entangle in EEG signal processing have been explored.
- I have investigated the obstacles related to classification and reported here.
- Finally, observing the common challenges both in signal processing and machine learning (ML), I have suggested some solutions.

This systematic review is concentrated on bringing EEG to bear as a scrutinizing clinical tool to detect or classify MCI, dementia, AD, SZ, and PD. The remainder of the chapter is divided into four main sections. In section 2.3, all the mentioned neuro-diseases are sub-sectioned and related recent works are discussed. All the methods used in the recent studies, which are mentioned in Section 2.3, are elaborately discussed in Section 2.5. There is a list of public EEG dataset repository reported inside Section 2.4. Section 2.5 is partitioned into two subsections: feature extraction methods and classification methods. A detailed comparison and analysis have been provided in Section 2.6. Section 2.7 has reported all the related challenges involved in EEG data processing. A summarized discussion of the findings is talked over in section 2.8, and finally, this study ends with the essence in section 2.9.

2.3 Neuro Diseases

In this section, I deliver a summarized description of the aforementioned five neuro-diseases with some recent research works and their implemented data structures. I have reported all the work that has been done with EEG data only. Among all five neuro-disorders, I have found more studies have been done on AD detection and prediction. And also, almost all the studies have struggled to have a large dataset for their experiment. This did not allow them to have full or high accuracy in their works. I have placed the works below in sorted order with respect to the diseases.

2.3.1 MCI

MCI is a degenerative neurological disorder caused by cognitive degeneration. It is the early stage of AD, caused by a diminished vocabulary, progressive memory erosion, and a reduced potentiality to perform precise motor motions, all of which impair the performance of day-to-day activities [100]. It is still an ongoing area of study to detect dementia at the MCI stage, which is very challenging in most cases. Annually, almost 25% of patients with MCI are metamorphosed into AD [19]. Therefore, predicting dementia at the MCI stage can support the patients with early medical treatment. By implementing appropriate methods with a quick and timely diagnosis, I can protect and significantly develop subjects' lives. As it is an elderly neuro-disorder, therefore, most of the time the symptoms remain unchecked.

Sharma *et al*. [75] worked with 44 subjects (15 dementia subjects, 16 MCI subjects, and 13 healthy volunteers) who were more than 40 years old and living in India [92]. The EEG data that they used had 21 channels, and one ECG channel was run for 13 minutes, keeping the sampling rate at 256 Hz. A different approach named the Finger Tapping Test (FTT) was used in [75], [76], [77], which calculates the connectivity and motor speed of the cortico-cortical sensorimotor channel. In FTP, MCI patients were asked to squeeze the space bar of the computer at 10-second intervals with their left and right index fingers in accordance with the instructions shown on the computer screen. The Continuous Performance Test (CPT) was also used in [75], [76] to calculate attention and impulsivity. In the CPT test, MCI patients were asked to squeeze the space bar when they see the letter B on the computer monitor with a gap of 1 to 4 seconds. These two tests were done to measure the attention and the motor speed of the subjects to identify dementia, MCI, and healthy volunteers (HVs). A deep learning (DL)-based study [86] with 35 subjects (28 HVs and 7 MCI subjects) was conducted with 16 channel EEG data.

A solo EEG channel-based study to classify MCI had been proposed by Khatun, Morshed, and Bidelman [52]. The authors had 23 subjects, and among them, 15 of them were HVs, and 8 of them had MCI. The subjects were asked to identify the vowel continuum speech sounds as the sounds were playing during the EEG data recording. Using the event-related potential (ERP), 590 features were picked. A dataset having 25 MCI subjects and 11 HVs has been used to predict MCI at an early stage [58]. The authors recorded nineteen to thirty-two channels of EEG data at a rate of 250 Hz, and the band

pass was from 0.016 to 100 Hz. After EEG recording, they used independent component analysis to erase the eye movement artifacts. 86 MCI patients were diagnosed for two years, and 25 patients turned into AD patients in [70]. EEG data used in this experiment was collected using the international 10-20 system in a dimly lit room. The authors kept the sampling frequency at 500 Hz, and subjects kept their eyes closed for 20 minutes.

Characteristic	MCI	Control	*P*
Age (years)	66.4±4.6	65.3±3.9	0.4
Education (years)	10.3±3.8	11.1±3	0.4
GHQ scores	20.5±9.4	17.9±6.6	0.3
BMI (kg/m^2)	25.7±2.2	26.6±3.6	0.3
Fasting glucose (mg/dl)	115.5±24.3	121.8±36.9	0.3
Total cholesterol (mg/dl)	170.6±61.4	169.1±42.6	0.9
Triglycerides (mg/dl)	157.3±100.9	160±80.7	0.9
Creatinine (mg/dl)	1.2±0.2	1.3±0.3	0.1
MMSE scores	27.6±0.9	29.0±0.8	<0.001
NUCOG scores	82.4±3.6	91.1±3	<0.001
Gensini scores	33.3±31.9	20.3±21.7	0.1

GHQ – General health questionnaire; BMI – Body mass index; MMSE – Mini-mental state examination; NUCOG – Neuropsychiatry unit cognitive assessment tool; MCI - Mild cognitive impairment

Fig. 2.1: Subjects' demographics and psychiatric test scores [50].

Kashefpoor, Rabbani, and Barekatain [50] worked with 27 subjects, of whom 11 were MCI patients and 16 were HVs. All the subjects were in a resting, eyes closed, state, and EEG data were recorded in the morning. **Fig. 2.1** is a screenshot illustrating the demographic information of subjects who participated in this study. Using the same data, authors proposed an integrated MCI detection framework [100]. Another MCI detection framework [84] was proposed using this same EEG dataset.

2.3.2 Dementia

Dementia is prescribed clinically as a degeneration of brain activities, with mostly the subject's intellectual and memory degeneration characterized by the quietus of brain cells [23]. In recent years, there have been significant efforts to detect dementia at an early stage. Dementia is peered with neurodegenerative disorder diversity [6]. Efforts through neuroimaging, neurophysiological, and biochemical biomarkers have been made [23], [36]. If dementia has not been detected in its early stages, it can increase the severity of

having AD. The Alzheimer ’s Disease International (ADI) report says 66% of humans are affected by dementia in developing countries. There is someone who develops dementia every 3 seconds [44].

Durongbhan *et al.* [29] reported a classification framework using time-frequency and frequency domains. The authors worked with 40 subjects, of which 20 were HVs and 20 were patients with dementia. The 10-20 international system was modified with a 10/10 overlapping system, and an XLTEK 128 channel was exercised for sampling at 2 KHz. Dominant-frequency analysis was used to detect dementia in its early stages [14]. The authors used publicly available data that was collected with the patients in a relaxed state, with both eyes open and closed. The analogue pass band of frequencies was 0.01–100 Hz, whereas the sampling rate was 256 Hz.

2.3.3 Alzheimer’s Diseases

AD is the most common among the different categories of dementia, and almost 70% of cases fall into the AD category [22]. The most dangerous form of dementia is AD, which causes serious memory loss and cognitive impairment. When dementia reaches AD, there is no permanent cure for it; only alleviative care can stop the growth of AD for the time being. After being medically proven, an AD patient lives for only 5–8 years [21], [39]. Therefore, many studies have been done, and it is still an ongoing area of work to detect AD at an early stage so that the patients can have a medically sound life.

Fiscon *et al.* [32], [31], proposed some supervised algorithms using EEG signals to classify AD subjects. This study was done using 109 subjects, of whom 86 were suffering from AD or MCI and 23 were HVs. A multi-channel EEG signal was collected using 19 electrodes according to the 10–20 system. Kim and Kim [53] worked with 20 subjects, among whom 10 were AD patients and 10 were HVs. The EEG data were collected for 1 minute for each subject having 32 electrodes according to the 10-20 system, and the sampling rate was 500 Hz. All the subjects were in a resting state with their eyes open. In [96], authors collected 144 subjects EEG data from two different sources, and among them, 102 were HVs and 42 were AD patients. This study tried to prove that only three electrodes are enough to detect AD subjects.

Subjects from three different nationalities—England, Italy, and Romania—were used in [5]. A total of 41 British citizens, among them 17 AD patients and 24 HVs, 10 Italian nations, among them 5 AD subjects and 5 HVs, and 11 Romanian citizens, among

them 8 AD patients and 3 HVs, were engaged in the EEG data recordings. A total of 169 patients' (58 MCI subjects, 49 AD subjects, 22 subjective cognitive impairment (SCI) subjects, and 40 other pathology subjects) EEG data that were collected from 2009 to 2013 in Charles-Foix Hospital, France, was used to detect AD in a differential framework [42]. These data were collected in a relaxed state with both eyes closed, and the sampling frequency was 256 Hz. Along with the international 10-20 system, 11 additional electrodes were used.

In [16], transcranial magnetic stimulation (TMS)-based EEG recorded data from 26 right-handed AD patients was used. Among all 26 subjects, 20 patients participated in the Face Name Association Memory Task (FNAT). The EEG data were processed at a sampling rate of 5000 Hz. 37 right-handed AD patients and 37 MCI patients participated in [13]. Though the AD subjects were prescribed cholinesterase inhibitors (ChEis) at 20 mg/day and a dose of antidepressants (citalopram) at 30 mg/day, the MCI subjects were not following any clinical healing. EEG data were collected under the 10-20 system with a sampling frequency of 1024 Hz. A comparison of F8 and O2 amplitudes of a MCI and an AD subject of this study is shown in **Fig. 2.2**. Xiaojun and Haibo [18] studied 12 subjects, of which 4 were HVs, 4 were MCI patients, and 4 were AD patients. The EEG data were collected from 64 channel electrodes with a sampling rate of 500 Hz, and the duration was 60 seconds for each participant.

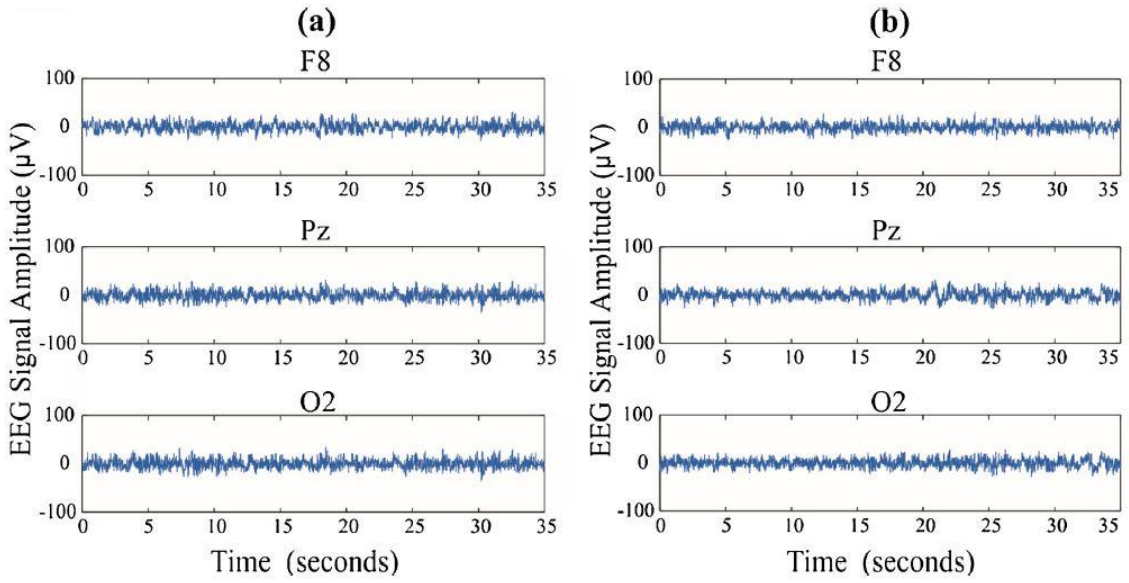

Fig. 2.2: EEG signal's sample at three different positions: frontal (F8), parietal (Pz), and occipital lobes (O2) of (a) an MCI and (b) an AD patient [13].

With a sampling rate of 256 Hz and a total of 38 subjects in [73]. There were eleven HVs, eight MCI subjects, and nineteen AD patients in the dataset. 5 AD patients who are also suffering from epilepsy were chosen to have their EEG data collected for 24 hours with a 30-minute routine EEG epoch [41]. After discarding the other events of the day, the sharing of discards and their relevance to sleep time were measured. 76 subjects (27 HVs and 49 AD subjects) participated in [33], where the authors tried to characterize AD using amplitude modulation analysis. The collected EEG data had twenty channels, and the sampling rate was 200 Hz. Chiang and Pao [24] proposed a fuzzy probability model to detect AD at an early stage. This study was done using the EEG data of 22 HVs and one AD patient collected at a sampling frequency of 512 Hz. 34 subjects (17 HVs and 17 AD subjects) participated in [2], where the EEG data were collected with a 128-channel geodesic sensor for 180–240 seconds. The sample rate of the EEG signal was 500 Hz using a 12-bit Analogue to Digital Converter (ADC).

2.3.4 Schizophrenia

Schizophrenia is a neuropsychiatric disorder that affects one percent of the world's population [11]. According to the WHO [92], there are more than 21 million people who are suffering from schizophrenia [98]. It is one of the four leading causes of childhood disability in developed countries. SZ prefixes influence the subjects by manifesting as auditory hallucinations, paranoid or bizarre delusions, and/or disorganized speech and thinking in the context of significant social and/or occupational dysfunction [46]. A limited number of studies were done to detect schizophrenia (SZ) compared to AD. Jalili and Knyazeva [46] reported a graph-based theoretical measure of neuro-functional networks. A total of 28 subjects (14 SZ subjects and 15 HVs) participated in this study by recording EEG data in the resting state with their eyes closed for 3–4 minutes with the help of the Geodesic Sensor Net. 13 SZ subjects and 18 HVS EEG data were used to classify SZ patients, as proposed in [20]. For each subject, the EEG recording lasted for two minutes while seated upright and keeping eyes open. Jahmunah *et al.* [45] collected a 19-channel EEG dataset from 14 SZ patients and 14 HVs. 19 electrodes, as stated by the international 10-20 system, were used, and the experiment lasted for fifteen minutes with a sampling frequency rate of 250 Hz. A visual disparity between a HV and a SZ subject’s EEG signal after filtering the raw signal from noise appeared in **Fig. 2.3**. Another recent study [67] was done using a public dataset of 45 schizophrenia patients and 39 HVs. With a sampling frequency of 128 Hz and a duration of 60 seconds for each

participant, the EEG data of 16 channels was collected. Li *et al.* [56] reported a resting and P300 task-based EEG dataset of 23 SZ subjects and 25 HVs. EEG data were recorded using a symptoms amplifier and 16 channel electrodes with a sampling frequency rate of 1000 Hz. A total of 20 subjects (11 SZ patients and 9 HV) were included in a recent study [28]. The EEG data in this study is different from other studies because it was recorded while the subjects were freely viewing natural scenes. 32 scalp channels were placed to record the EEG data; another 6 channels were arranged around the eyes to save the electro-oculogram (EOG), and the sampling frequency rate was 2048 Hz. In another study with 60 schizophrenia patients and 76 HVs, authors tried to find the difference in the gamma activity among the subjects [17]. A 256-channel Biosemi Active Two method was implemented to store the EEG data at a frequency rate of 512 Hz. 14 paranoid SZ patients and 14 control subjects participated in a multivariate iterative filtering (MIF)-based SZ detection study [25]. EEG was recorded for 15 minutes on 19 channels, and the electrodes were placed according to the 10-20 international EEG electrode placement standard. Siuly *et al.* [85] used the empirical mode decomposition (EMD) method for SZ identification from EEG signals. The data for this study was gathered from 81 people, including 49 SZ sufferers and 32 HVs. Collected EEG signals were decomposed into intrinsic mode functions (IMFs) by EMD. An empirical wavelet transformation-based SZ identification framework was proposed in [51]. This study contained 22 HVs and 26 SZ subjects.

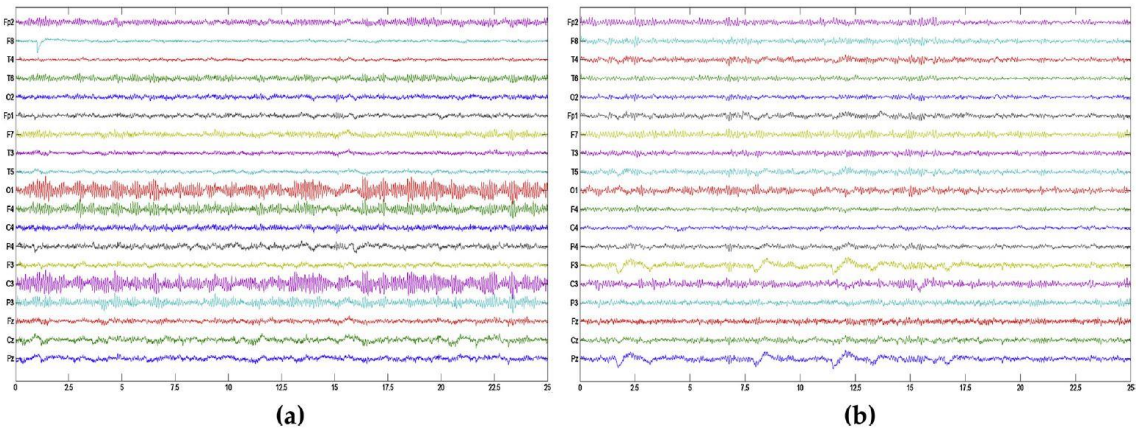

Fig. 2.3: EEG signals of (a) HV and (b) schizophrenia after pre-processing [45].

2.3.5 Parkinson

Parkinson's disease is a neurodegenerative disorder characterised by the death of the neurones. According to the WHO, 10 million people have died because of PD [97]. At the time of birth, the human brain contains the maximum number of brain cells, which are also known as neurones [87]. Apart from other cells of the human body, neurones cannot get fixed automatically [78]. These neurones produce dopamine, a chemical substance necessary for controlling body movement. Therefore, the death of nerve cells stops them from producing dopamine, and that is the main cause leading to PD. The principal prefixes of PD are damaged fine motor skills, unsteady posture, tremor, slow movements, loss of balance, and stiffness [40]. Very few studies were done to detect Parkinson's disease using EEG signals.

A convolutional neural network (CNN) model was used with 20 PD subjects and 20 HVs EEG data in [61]. For each data point, the EEG recording lasted for 5 minutes in the resting state with the eyes closed. The sampling frequency was 128 Hz, and an emotive EPOC neuroheadset having fourteen channels was utilized. Raw EEG data of a HV and a PD patient are provided for comparison in **Fig. 2.4**. Using the same dataset, Yuvaraj, Acharaya, and Hagiwara [101] reported different classifier performances. Genetic Algorithm (GA) and Binary Particle Swarm Optimization (BPSO) were used to choose highly valued features in a ML-based PD study [64]. This study collected data from 252 subjects, including 188 PD patients and 64 HVs. Three repetitions of the sustained phonation of the vowel ‘a' were used to target each of the characteristics of PD throughout the data gathering process. The frequency of the microphone used to gather the data was 44.1 kHz.

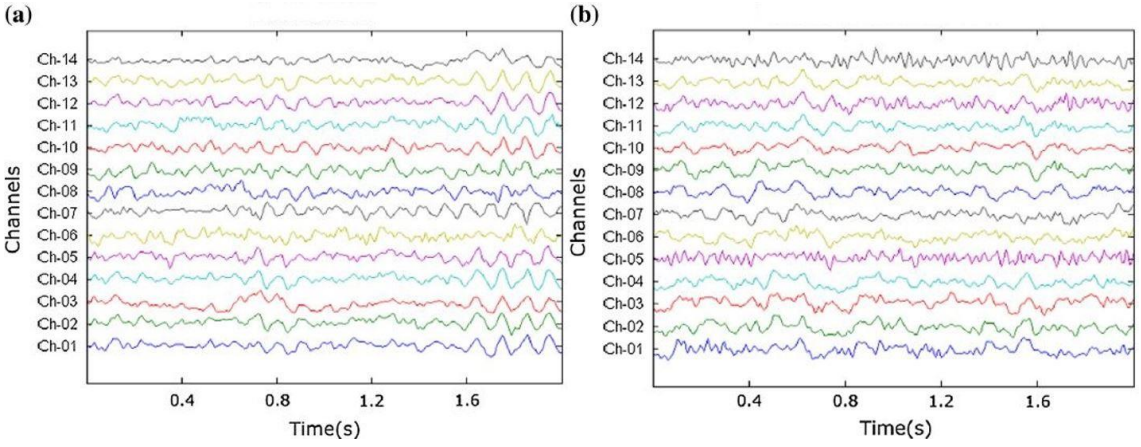

Fig. 2.4: A Sample EEG signals of (a) HV and (b) PD subject [61].

2.4 Public EEG Dataset Repositories

Table 2.1, gives some publicly available EEG dataset web link of these neuro diseases.

Table 2.1: Publicly available EEG dataset sources.

Neuro diseases	**Neuro diseases**
MCI	EEG Signals from Normal and MCI (Mild Cognitive Impairment) Cases. [Online]. Available: http://ww25.biosigdata.com/?download=eegsignals-from-normal-and-mci-cases#comment-2909
AD	https://doi.org/10.17862/cranfield.rd.7673702
Schizophrenia	https://repod.icm.edu.pl/dataset.xhtml?persistentId=doi:10.18150/repod.0107441
Schizophrenia	https://www.kaggle.com/broach/button-tone-sz
Parkinson	https://www.frontiersin.org/articles/10.3389/fninf.2017.00067/full

2.5 Previous Efforts

All the feature extraction methods that were used in the existing research are described here in this section. Data source, collection techniques, and preconditions have been discussed in the earlier section. Almost every study that I have analyzed has two parts to its methodology. The first part talks about EEG data processing and feature extraction, and the second part discusses the classification technique. I have put all the reference methods into the two subsections and described them accordingly.

2.5.1 EEG Data Processing and Feature Extraction

After collecting the EEG data, it is necessary to check for noise and outliers. The raw EEG data often contains artifacts like eye blinking, sinusoidal noises produced by alternating current (AC) movement, electro-oculograms, electrocardiograms, etc. Therefore, it is a must to preprocess the raw EEG data before feature extraction. A Butterworth filter is commonly used to discard higher-frequency signals. It is a low-pass

filter used in [29], [42], [45], [28], and [17]. A 6th-order Butterworth filter was implemented to filter frequencies from 1 to 49 [61], [101].

Independent Component Analysis (ICA) is a statistical algorithm that finds out the secret attributes and factors from a group of arbitrary signals or variables. It is mostly implemented in signal processing to reveal the important factors of signals. Artefacts like eye movements, electromyograms, electro-oculograms, and electrocardiograms were discarded using ICA [75], [58], [16], and [17]. Poil *et al.* [121] used the JADE ICA algorithm to break the EEG signals into 23 components. CleanLine is an EEGLAB plugin that is often employed to clean the signal. Notch (Butterworth) is a bandstop filter used for removing sinusoidal signals. A Wiener-type filtering technique named the SOUND algorithm is used in [16] to discard other noises from the EEG signal after applying ICA.

2.5.1.1 Time Domain Algorithms

To measure the correlation of brain capabilities at each electrode position, the authors of [96] calculated the Triple Correlation Value, which is measured by multiplying the three signals EVA(t), EVB(t), and EVC(t), using time shifts of τ_1 and τ_2 with regard to the signals, and finally integrating over time. The calculation of the triple correlation value is given in (1), and the three electrodes are shown in **Fig. 2.5**.

$$S_i(\tau_1, \tau_2) = \frac{1}{N}\int_i^{i+1} |EVA(t) * EVB(t-\tau_1) * EVC(t-\tau_2)|dt \quad (1)$$

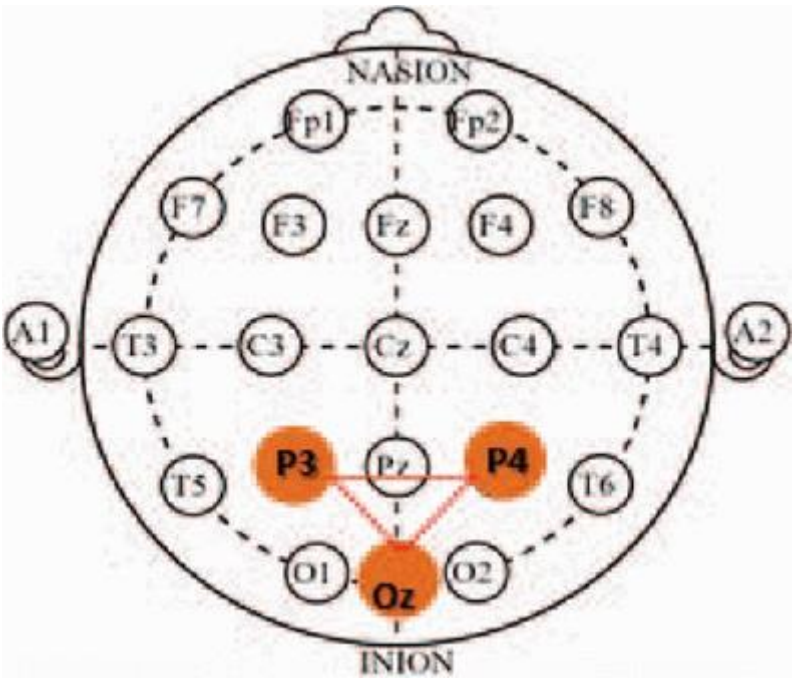

Fig. 2.5: Electrode sets located at P3, P4, and Oz in the International 10-20 EEG System [96].

The triple correlation value is only measured when all three signals have the same sign. Therefore, it can limit the rotation plane. The total number of times when all three electrodes have the same value is represented by N. This method was used to differentiate AD patients from HVs when the triple correlation value of all the subjects' distribution was plotted with respect to the time axis.

To find the similarity between two signals, the cross correlation value is calculated and used in [5]. It returns a sequence of vectors with a value range of -1 to +1. The cross-relation value is equal to 1 if both signals are identical to each other; otherwise, it returns 0.

To measure the complexity for early detecting AD, epoch-based entropy is measured [42]. Using the Hidden Markov Model (HMM) [72], the epoch-based entropy is calculated on piecewise stationary epochs of the EEG signal. It estimates the local density at epoch level.

Jalili and Knyazeva [46] worked with the connectivity matrices of the neuron networks of the brain. They considered the partial and impartial correlations to find out the functional connectivity through both indirect and direct links. The authors measured the Pearson correlation coefficient for all feasible couples.

Source localization is a non-parametric algorithm to present a linear equation between EEG calculations and brain networks, which was used in a study [2]. It is a time-domain calculation to measure the similarity between electrodes at any given time. For N number of electrodes, P dipole, and T discrete time samples sources at any given time, source localization matrices are measured as follows:

$$b = \begin{bmatrix} b(r_1) \\ \vdots \\ b(r_N) \end{bmatrix} = A \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = Ax \tag{2}$$

$$b(t) = \begin{bmatrix} b(r_1, 1) & \cdots & b(r_1, T) \\ \vdots & \ddots & \vdots \\ b(r_N, 1) & \cdots & b(r_N, T) \end{bmatrix} = A \begin{bmatrix} b(x_1, 1) & \cdots & b(x_1, T) \\ \vdots & \ddots & \vdots \\ b(x_p, 1) & \cdots & b(x_p, T) \end{bmatrix} \tag{3}$$

In (2), b represents the EEG measurements vector (N1),... exhibits the location of each electrode, and x shows the cortical dipole moments vector (P1). The current flow for a given electrode through each dipole position is indicated by each row of the transfer

matrix A (NP). In (3), b(t) (NT) is the EEG measurement matrix, and x(t) (PT) is the dipole moments matrix at different time frames.

Boostani, Sadatnezhad, and Sabeti [20] reported auto-regressive (AR) coefficients, which is a time-domain algorithm used for signal processing. Each EEG sample is treated as an estimation of earlier weighted samples in the AR model. Siuly *et al.* [84] used the AR model along with permutation entropy (PE) for feature extractions. The authors calculated the AR coefficients by using the Burg method [88]. It applies an nth-order AR model to the input signal by reducing the backwards and forwards estimation errors. Therefore, setting the order of the model is an important decision here. The Finite Sample Criteria (FSC) method was performed to find out the best order for AR. A vector autoregressive (VAR) model was used to generate a CNN in [67]. To measure the connectivity between different EEG channels, this VAR model was used. VAR coefficients estimators can be calculated using the least-squares (LS) method.

From the filtered EEG data, only 14 optimal features were chosen among 157 non-linear features using a Student's t-test [103] in a study [45]. Student's t-test is a statistical method to verify a hypothesis from the mean of a small sample calculated from a normally distributed dataset. It is used when the standard deviation of the dataset is unknown. The most discriminant 30 features for a SZ study [25] are selected for classification, and they are ranked using the student t-test.

Phase Locking Value (PLV) and coherence are useful tools to build brain networks. PLV can be used to measure changes in the large-range synchronization of brain activity from EEG signals. Li *et al.* [56] worked with PLV and P300, two time-domain techniques, and measured the phase synchronization among each pair of electrodes. The authors defined the PLV value range as 0 to 1. A high PLV value means strong phase synchronization. On the other hand, P300 is an endogenous potential and part of the event-related potential (ERP), which helps in finalizing decisions. While recording the EEG signals, P300 surfaces as a positive drift in the supplied voltage with an evanescence of 250 ms to 500 ms. After filtering the raw EEG data, all the P300 trails for each subject were averaged in the same study. From the trial-averaged value, C3, C4, P3, and P4's P300 amplitudes were set as their utmost peaks from 300 ms to 500 ms. Then, SZ and HVs were distinguished by the differences in P300 amplitudes using an independent t-test.

ERP computes the response of the brain when a sensory, motor, or cognitive event occurs in front of the subject. After applying the Butterworth filter and ICA to the recorded EEG data, the authors measured the ERP with all possible trials for each electrode [52], [28]. For every image group, the authors [28] averaged across frontal electrodes (Fz, F3, F4, AF3, and AF4), central electrodes (Cz, C3, and C4), parietal electrodes (Pz, P3, P4, CP1, and CP2), and occipital electrodes (Oz, O1, O2, PO3, and PO4). From each of these four average ERPs, the average between 400 ms and 600 ms was used as a feature for classification. In this same study, authors used the Wilcoxon Rank Sum (WRS), which is a non-parametric method to check the mean ERP results. WRS is often used because it allows two independent groups to have two different sample sizes. WRS was also used in another study [2] for statistical evolution.

2.5.1.2 Frequency Domain Algorithms

Khatun and Morshed [52] extracted 590 features using ERP and used the top 25 features for classification. While extracting the features, ERP and Candidate Feature Vector (CFV) were used. CFV is a nonlinear algorithm to separate features from EEG data and has been described in the nonlinear method section. The authors used relative power (RP) along with CFV in the EEG bands for detecting AD, MCI, and HVs. RP measures the relationship between two passing frequency bands. In [52], [20] all the four bands [delta (0-4Hz), theta (4-7Hz), alpha (8-12Hz), and beta (12-30Hz)] were used, and in [53], only three bands [theta (4-7Hz), alpha (8-12Hz), and beta (12-30Hz)] were used to extract features. The *power.m* function of MATLAB was used, and 64 features were extracted in the study [52]. By normalizing each band power by the total power of 4–30 Hz, three RP values were computed, and 96 relative power features were used for each EEG trial in the study [53].

The cross-frequency connectivity technique was reported in [58], [33]. All the participating subjects had to interact with the questions and descriptions asked by the authors [58], and the alterations in the scalp functionality of the MCI and HV participants were captured in the EEG recordings. Cross-frequency coupling between θ and α / β were noticed between groups and within groups.

The Hilbert Transform (HT) converts real-valued signals to analytic signals and is reported in [70], [5], [33], and [56]. 13 biomarkers were extracted using HT in [70]. The temporal envelope of each of the 5 bands of the EEG signals is calculated using HT

[33]. Using the HT, which transforms analytic signals, phase synchronisation was detected in [5].

Linear coherence of signals can be measured using magnitude squared coherence and reported in [5]. The square of the modulus of the average cross-power spectral density (PSD) normalized to the product of the average auto-PSDs is known as the magnitude squared coherence. PSD computes the average power allotted as a frequency function and is also used in [24].

Linear coherence between two signals x and y is measured using (4), where $C_{xy}(f)$ is defined as the coherence between two channels x and y at a given time, $C_{xy}(f)$ is the cross PSD, $P_{xx}(f)$ and $P_{xy}(f)$ are the two PSD estimations of x and y channels, respectively.

$$C_{xy}(f) = \frac{|P_{xy}(f)|}{P_{xx}(f)P_{xy}(f)} \tag{4}$$

Phang *et al.* [67] reported Partial Directed Coherence (PDC), which measures the direct dependencies between frequency channels in a brain network. It also uses the Furrier Transformation (FT) and VAR coefficients described in the time domain algorithm. The squared PDC has a normalized value between 0 and 1, which determines the ratio between the channels outflow of data.

After applying fast Furrier transformation (FFT), Dominant Frequency (DF) was measured for each epoch [14]. DF in α range was found available in 30% of the epoch, DF variability above 3.5 Hz in 68% of the epochs and DF in θ is above 40% of the epochs.

2.5.1.3 Time – Frequency Domain Algorithms

Wavelet Transform (WT) has been used in [75], [100], and [42] after cleaning the raw EEG data from artifacts/noises. Some of the noises have various time-frequency domain properties. Therefore, WT is a powerful time-frequency tool to discard those multi-domain noises. WT along with ICA wavelet noise removal was done in [75], and using the stationary wavelet transform (SWT), recorded EEG signals were decomposed into coefficients with various frequency ranges, and frequencies that were bigger or smaller compared to the range were discarded in [100]. Power line interference and baseline drift were removed using SWT and segmented in [84]. Then, segmented data were compressed

using Piecewise Aggregate Approximation (PAA). Morlet wavelets were used to calculate wavelet time-frequency in [42].

The MUSIC-Empirical Wavelet Transformation (EWT) algorithm was proposed by Amezquita-Sanchez and Adeli [12] and used in a study [13] to detect AD. It is an adaptive WT that performs well in buggy signals with non-stationary and non-linear characteristics like the EEG signals. In the mentioned study, initially the MUSIC algorithm is implemented to compute the frequencies visible in the investigated EEG signals, and then the fitting limitations are set to the visible frequencies that have already been analyzed. This leaves the WT filters blank. The extremely non-stationary EEG signals are decomposed into modes in a Fourier spectrum using EWT [51].

Rodrigues *et al.* [73] used the discrete wavelet transform (DWT) to have a multi-resolutional analysis from successive filtering and shorter signal resolution versions. Using a scale function and a WT function, the DWT works. The authors achieved signal disintegration into multiple frequency bands by continuous high-pass and low-pass filtering in the time domain, followed by sub-trialling by a factor of two until the highest elevation of disintegration is achieved. In another study [32], DWT was applied to each EEG signal for 180 seconds. There were 48 coefficients and 19 electrodes, which produced 912 features.

The mean of all continuous wavelet transform (CWT) coefficients is known as the CWT and was reported in the study [29]. Availing a mother wavelet in that fixed frequency band over the complete signal duration, CWT is measured.

The Fourier Transform (FT) is a commonly used algorithm to transform time-domain EEG signals into frequency-domain signals through non-periodic functionalities. A fuzzy logic-based study [24] to detect AD had applied FR to the filtered EEG data before assigning fuzzy membership values to the bands. In another study [32], the FFT was used for 180 seconds for each channel with 16 Furrier Coefficients, and FFT is dependent on the Discrete Furrier Transform (DFT). 304 characteristics were separated from the data after implementing FFT in the same work. As the authors of [18] found that the most important features lie in the frequency domain, they applied FFT to measure the power spectrum of the data. They used FFT for 0.5 s for each channel only in the theta, alpha, and beta bands because the oscillatory cortical activity mostly found these bands. Aghajani *et al.* [2] used the dipole approximation algorithm of FFT for frequency

analysis. On the other hand, Welch's algorithm with a 2 second, 75% overlapping window of FFT was applied in [17]. γ band was computed in 31–48 Hz by the addition of absolute power on frequency bins. After that, authors calculated the Log10 transform with summed frequency. FFT was applied on 12 epochs per channel in another study [14]. In [29], the features of FFT were set as the mean magnitude of the FFT coefficients in a particular frequency band.

Mazaheri *et al.* [120] measured the time-frequency representation (TFR) of power using moving Hanning tapers for each trial and frequency while maintaining a variable time window of 3 rounds. This power oscillatory analysis was done using this TFR.

The sparse bump model is a time-frequency map that was used in [42]. It has time-frequency representatives that go on for approximately four time periods and are concentrated at a particular frequency. These patterns are the ambassadors of transient local synchronization in the neuronal network.

2.3.1.4 Other Neuromarkers

ANOVA is a statistical measure to verify the significance of any experiment. It had been used in [58] to identify the event-related shifts in the θ and α / β between constraints, in [75] to find out the difference between independent groups (MCI, dementia, and HV), in [13] to discover the most unique features for differentiating AD subjects from MCI, in [41] along with Bonferroni posthoc analysis to analyses the sensitivity of the EEG signal's epoch, and in [33] to differentiate normally distributed features across the three groups (MCI, AD, and HV) at different locations on the scalp.

Reference Electrode Standardization Technique (REST), an offline band pass filter, was used in [56]. It has a duration of 1 s of EEG data segmenting [-200 ms, 0 ms] for baseline emendation and other artifacts.

Kashefpoor, Barekatain, and Rabbani [50] reported a correlation-based process for feature extraction. They tested channels and features individually, zone correlation-based process for feature extraction. They tested channels and features individually, zone-group-mean, and zone-grouped to select the best discriminatory feature.

The Candidate Feature Vector (CFV) in [52] included features that were distinct by ERP. The highest amplitudes and their corresponding latencies, as well as the average amplitudes of the interval holding the distinguished points, were also added to the CFV.

CFV performs well in isolating groups in the tests, implying evoked responses. 80 feature points were selected from the CFV.

For analyzing the nonlinear properties of the EEG signal, higher order spectra (HOS) were used in [101]. HOS is a spectral portrayal of higher-order statistics. Because of the difference in gaussianity and levels of nonlinearity in the time domain, HOS can hold the information. A third-order spectrum named bispectrum was applied in this work. Bispectrum is the easiest to calculate and most reachable of HOS.

For measuring the orderliness and self-similarity of any signal, the Fractal Dimension (FD) was used in [13], [20]. FD works with the entropy of a signal, and this entropy holds the information of that signal. In both studies, three algorithms—the Higuchi method, Katz method, and Petrosian method—were used to calculate the FD.

The Hurst Exponent (HE) is another approach to measuring the similarities available in a time-domain signal like the FD, mentioning the order of persistence between the signal series. HE has a value between 0 and 1, which means it works on scaling-law statistics. When the HE value is 0.5, it means it is a random signal. If the HE value is greater than [92] 0.5 or less than [92] 0.5, it means the measured signal has low or high self-similarities, respectively. In [13], HE was used to separate the MCI patients from the AD patients.

2.5.2 Classification Methods to Detect Neuro Diseases

Classification is a way to identify the category or class of a new observation based on a training set of data that contains the characteristics. Classification methods are being applied to almost every research problem. ML algorithms are an application of artificial intelligence that enable the model to learn and detect automatically. ML algorithms are often chosen and perform well when it comes to supervised learning.

K-nearest neighbor (KNN) is a well-known classical ML algorithm to classify unknown subjects and is used in [65], [29], [66], [100], [13], and [45]. The default 10-fold cross-validation procedure was implemented, and training was done continuously with the same data for 50 iterations [29]. To increase the performance and reduce the fluctuation, they evaluated and averaged the 50 trained models. The collected EEG data were separated into N folders, and among them, N-1 folders were used for training the models. The rest of the data was used for testing, and this process continued for N times.

$$Loss = \frac{FP+FN}{TP+TN+FP+FN} \quad (5)$$

$$Accuracy = 1 - L \quad (6)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (7)$$

$$Specificity = \frac{TN}{TN+FP} \quad (8)$$

Loss, accuracy, sensitivity, and specificity are calculated using (5), (6), (7), and (8), respectively, where TP means true positive, FP means false positive, TN means true negative, and FN means false negative. This is how the authors measured the performance of the classification model in [29], [100]. KNN was implemented with other classification algorithms just to compare the performance in [100], [45]. Fuzzy logic was used along with KNN in [101], [50]. The Takagi-Sugeno neuro-fuzzy (NF) inference system was applied, which has the IF-THEN rules inside the model for universal approximation in [50]. Interpretability and accuracy are the two contradictory requirements in fuzzy modelling. Linguistic fuzzy modelling looks after the interpretability, and the Takagi-Sugeno model takes care of the accuracy. Features were used as the input of the NF system, and their corresponding values were the output, where MCI = 1 and HV = 2. In the study, two-thirds of the data were used to train the NF, and the outputs were used to model the KNN classifier [50]. Fuzzy K-Nearest Neighbor (FKNN) was implemented in [101]. Fuzzy membership values were computed based on the Euclidean distance. The fuzzy strength parameter was also calculated to find out the contribution of each neighbor based on the distance. The authors found the best result when the strength parameter was set to 1.24 and k = 3.

Decision Tree (DT) is a decision-aided classifier that has tree-type models and their feasible effects. DT is a flowchart-type structure. I have found DT has been implemented in [101], [100], [32], [13], [24], and [45]. Extracted features were used to build the tree that had the rules to detect the subjects into two labels, and the performance was measured by the Gini index in [101]. The C4.5 DT algorithm was used as it has the capability to manage noisy datasets [32], [15]. The quality of the model was measured using the leave-one-out cross-validation method. DT was utilised for comparing with other classifiers in [100], [13], [24], and [45]. Siuly *et al.* [85] introduced the ensemble bagged tree (EBT) for a SZ study, where it outperformed other comparing classifiers with an IMF value of 2.

Extreme Learning Machine (ELM) is a relatively recent ML approach that can achieve the global optimal solution while providing higher generalization performance at a significantly quicker learning pace. ELM was first designed for single-hidden-layer feed-forwards networks (SLFNs), then expanded to include "*generalized*" SLFNs. Siuly *et al.* [84] used 50 neurons in the hidden layers of their ELM architecture. '*hardlim*' was set as the activation function.

Support Vector Machine (SVM) is a supervised learning model widely used for regression and classification. It is often used for detecting neurodegenerative disorders and is suitable for use when the data size is small or medium. To detect the MCI at an early stage, this binary classifier was chosen in [75], [52], [100], [58], and [24]. It was picked to compare with a fuzzy logic-based method in [24]. As there were fewer overlapping features between the HVs, dementia, and MCI groups, a hierarchy-based SVM classifier was chosen using the Gaussian kernel [75]. The model consisted of three binary class labels: control vs. dementia, control vs. MCI, and MCI vs. dementia. Khare *et al.* [51] used a similar Gaussian kernel-based SVM classifier for their SZ study. The top 25 features chosen by the random forest algorithm were implemented to develop the SVM classifier [52]. Polynomial and radial-based kernels were used with the cost, C = {10-2, 10-1, 1, 101, 102} in the same study. Yin *et al.* [100] developed a 3-D evaluation method to pick features, and selected features were used to design the SVM classifier. To achieve the capacity to classify in the presence of a polynomial kernel function having an order of 2, the SVM was used. Another Gaussian kernel-based study where radial basis functions (SVMrbf) were used to train the model to differentiate MCI from AD convertors from the two classes of MCI stable and HVs [58] This was performed on each related variable in their dataset and also on amalgamations of those variables. Liner inseparability was the main concern of the authors (SVMrbf). Using the polynomial kernel functions of orders 2 and 3, the nonlinear signals, which were difficult to separate, were transformed into a higher-dimensional feature space in a PD detection study [101]. SVM performed better [92] than other used classifiers in this same study when the smoothing parameter (σ) had a value of 0.284. The radial basis function and linear kernels were also part of this experiment. Houmani *et al.* [42] worked with multi-class probabilistic SVM classifiers for AD detection, having four different groups of subjects: SCI subjects, AD subjects, MCI subjects, and subjects with other pathologies. To overcome the unbalanced dataset issue and also differentiate between each pair of classes, they used the polynomial SVM

classifier. Using Platt's estimation method [69], results coming from the SVM classifier were mapped to posterior probabilities. Another AD detection study using a supervised linear SVM classifier was used to build an optimal hyperplane, making a decision surface that increased the margin of differentiation between the nearest data points associated with varied classes [2]. Any possible bias was erased by using the leave-one-out cross-validation (LOOCV) method to have an unfailing estimate of the performance of the algorithm. A SZ study used SVM and fuzzy support vector machine (FSVM) classifiers to compare with the proposed method [20]. When the input vectors go through the kernel functions (SVMrbf), their dimensions are enlarged, and a hyperplane is coached to detect the classes. The instances that are positioned on the margin are considered by the SVM when adjusting its hyperplane. It is one of the drawbacks of SVM reported in this study. Even the FSVM could not solve the margin solidity problem as the amount of noisy instances in the EEG signal features was so large. An SVM classifier with radial basis function (SVMrbf) was picked and applied to detect SZs and HVs in [45]. The SVM classifier, which has multiple polynomial kernels, including 1st-order (SVM1), 2nd-order (SVM2), and 3rd-order (SVM3), was also used to perform the classification. Again, SVM was trained to compare with the proposed DL method in another SZ study [67]. Time-domain Vector Autoregressive (VAR) features performed better [92] than the frequency-domain Partial Directed Coherence (PDC) features in achieving high accuracy. With the radial basis kernel, SVM was coached using the training features in another SZ work [56]. While training the SVM, a grid search technique was used to find the optimised parameters. Finally, based on the extracted EEG features (i.e., the spatial pattern of the network (SPN), network properties, and P300 amplitudes), the trained SVM classifier was used to distinguish the SZs and HVs. Das and Pachori [25] proposed a MIF-based SZ study where SVM with a cubic kernel was used.

Logistic regression (LR) is a statistical algorithm often used in classification problems that uses a logistic function to model a binary dependent variable. Khatun *et al.* [52] used the "*liblinear*" package of Weka 3.7 to perform LR for MCI detection. A LR grid search was implemented where the cost was 10-2, 10-1, 1, 101, and 102. Another MCI study used LR for binary classification [70]. The binary classes, either AD transformers (i) or MCI durables (ii), are regressed with linear integration of biomarkers. The probability was set to 50% as the classification threshold; if the number of subjects was greater than 92, they belonged to the AD converter class; otherwise, they belonged

to the MCI stable class. A SZ study used the random regression hierarchical linear model (HLM), where absolute power was the dependent variable [17]. Age, education level, and gender were utilized as covariates to adjust for potential confounding. A channel-wise comparison between SZ subjects and HVs was performed. Additional statistical analysis was done with gamma power changes in the subjects to find out the connection between psychopathological and clinical variables. Finally, to find out which frequencies are dominant in changes in SZ observed in the gamma range, another subsidiary analysis was done.

Linear discriminant analysis (LDA) is an algorithm commonly used in pattern recognition or ML. It perceives a linear combination of the inputted features that differentiates two or more classes. It is one of the basic classifiers that creates a single hyperplane in the data space. LDA was trained to classify MCI to AD converters from the two classes of MCI stables and HVs in an MCI study [58]. It was applied to each relevant variable and also to the combinations of those variables. Apart from the standard LDA, a boosted version of direct LDA (BDLDA) was proposed in a SZ study to improve the accuracy [20]. BDLDA is composed of the Fisher linear discriminant analysis (FLDA) and the Adaboost. To get a direction in the feature space along which the distance of the relative to the in class scatter prolongs the highest and, as a result, maximizes the class differentiability, authors used the FLDA. Then, by using the Adaboost, they sequentially employed a base classifier on a weighted version of the training sample set. To increase the within-class differentiability, a linear mapping was required and a pairwise class discriminant distribution [57] was used. To identify a matching class according to the set of findings, LDA was used in another SZ detection [45]. It did not perform well as it was tested with only two discriminative features. But the computation time was very fast for the LDA. Li *et al.* [56] used the extracted features from the different brain states (rest or work) to train the LDA. Then the model was tested based on P300 amplitudes, SPN features, and network properties to separate the SZ subjects from the HVs. In another study [28], three different LDAs were used to detect SZ: LDA, Rule-based LDA, and a combination of the posterior probability of two LDAs. Using the ERP occipital imparity as an instrument to separate the SZs from the HVs, the LDA was trained with the mean power from each patient as the input feature. While testing the same classifier based on the mean amplitude in each group, they called it a rule-based classifier. They also utilized the mean power from the occipital electrodes of each patient as the

input feature of the classifier. By adding three new rules and combining the two classifiers, the third LDA was created. To interpret the subject's label, the posterior probabilities were compared.

Naïve Bayes (NB) is a probabilistic ML algorithm that works on the Bayes theorem and has the belief that the attributes are independent and random. Ranked attributes were inputted to the NB classifier to perform an automated PD detection [101]. But it was outperformed by SVM. NB was also used for AD detection to compare its performance with other used classifiers [32], [13], and [24]. But it did not perform well compared with other commonly used classifiers.

DL-based methods [185], a subfield of ML, have been used in various investigations. DL is inspired by artificial neural networks, which are a type of human brain function and structure. A neural network (NN) was implemented to compare with the proposed classifier in an AD detection study [24]. CNN is a widely used DL algorithm often used for analyzing visual imagery. It is a fully connected network, and it learns more as the network gets deeper. Oh *et al.* [61] proposed a thirteen-layer CNN model to detect the PD. Adam optimization [54] with a learning rate of 0.0001 was utilized. The architecture used the *Relu* activation function for all layers and *Softmax* for the last layer. The dropout value was set to 0.5 for the dropout layer. Using the brute force algorithm, the kernel size and amount of filters were obtained. For an automated human mood detection study [27] using EEG oscillations, five deep CNN models are considered: AlexNet, VGG16, ResNet50, SqueezeNet, and MobilNetv2. The AlexNet features with Alpha rhythm give greater accuracy scores of 91.07 percent in the Oz channel, according to the trial findings. A SZ study was done using a novel multi-domain connectome CNN (MDC-CNN) [67]. The proposed model worked using brain network having two parts: connectivity feature extraction and the classification based on the CNN. VAR coefficients of time-domain, PDC of frequency-domain, and complex network's (CN) topological measures are the various measures of the directed brain network computed in the first stage. Second stage started with the extracted connectivity features from the different domains using as input to construct the deep CNN classifier. There were two 2D-CNNs to store the spatial architecture in the 2-D PDC and VAR connectivity matrices and a 1D-CNN for the CN computes of the entire brain network. A *softmax* activation function was used in the output layer.

Xiaojun and Haibo [18] demonstrated an AD detection algorithm with a multitask learning strategy using a discriminative convolutional high-order Boltzmann machine (DCssCDBM) with hybrid feature maps. There were two hidden layers: the energy function and the *softmax* activation function. Probabilistic maximum pooling was done using dual variables. Based on the Siamese DCssCDBM model with distributed weights, multitask learning was done.

Probabilistic Neural Network (PNN) is one of the feed-forwards architectures that has multiple layers and works with an exponential activation function. PNN was used to compare with the proposed model in a PD detection study [101]. In a SZ study, PNN was implemented but did not perform well [45]. This SZ study used the hidden layer for calculating the probability density value and the summing layer for accumulating the results. An AD study modified the PNN and used the Enhanced Probabilistic Neural Network (EPNN) architecture [13]. After being evaluated by the ANOVA test, most discriminative attributes are inputted into the EPNN classifier to distinguish the AD and MCI subjects.

Kim and Kim [53] mentioned a Deep Neural Network (DNN) technique to detect AD subjects. There were four hidden layers, and the trial-and-error method was utilized to determine the number of nodes in each hidden layer. As the number of nodes in each hidden layer affects the learning, they changed the hidden nodes from 5 to 15 for each hidden layer.

A bidirectional long short-term memory (BLSTM) network is proposed for MCI detection in [86]. With a total of 164 hidden nodes, the BLSTM layer is built with one forwards and one backwards layer of a long short-term memory (LSTM) network. A Rectifying Linear Unit (*ReLu*) Layer is an activation layer that outputs the original input if the input is positive and 0 if the input is negative. The following layer is a completely connected layer, in which all nodes are linked to each other. The *softmax* layer restricts the function's output to a 0–1 range. The three main component characteristics generated from the principal component analysis (PCA) processing of the 16-channel EEG and EMG data are fed into the network. An EMG signal-based hand movement recognition framework has been proposed in [59]. Logarithmic spectrogram-based graph signal (LSGS), AdaBoost k-means (AB-k-means), and an ensemble of feature selection (FS) approaches make up the suggested system.

Using only three electrode potentials of the EEG data, the authors [96] proposed a new method named Deep Neuronal Activity Topography (dNAT) for AD detection. They assumed a triangular-shaped pyramid region in the brain network collected by three electrodes named P3, Oz, and P4. By exploring the time series of these three electrodes, they classified the AD subjects. **Fig. 2.5** displays those three electrodes that were used in this study.

Al-Jumeily *et al.* [5] designed two different approaches to detect AD. One is taking the average of the synchrony measures, and the other is principal component analysis (PCA). For all frequency bands, they used the synchrony computation technique on the time series of each pair of electrodes in two separate regions. And by using the PCA, they eliminated redundant features. After that, they used the synchrony measures on two pairs of regions. Then the results were compared with the neural synchronization of the AD and HVs by the Mann-Whitney U-test.

Standardized low-resolution brain electromagnetic tomography (sLORETA) along with the TMS protocol were introduced in [16] for AD classification. sLORETA [63] is a distributed-source imaging algorithm. Standardized neuronal present source density allotment compatible with the scalp topography in the cortical grey matter and the hippocampus of the template of the Montreal Neurological Institute 152 (MNI-152) that comprises 6239 grey matter voxels at 5-mm spatial resolution is calculated by sLORETA [34]. sLORETA calculated the underlying sources under the hypothesis that the neighboring voxels should have maximally alike electrical activity; the outcome was a blurred, widespread result in which the voxel with the maximum present density is identified as the middle of the source of the signal.

Rodrigues *et al.* [73] designed a hybrid algorithm using the cepstral analysis of EEG DWT multiple band decomposition for early AD detection. After performing the DWT decomposition with the EEG data, a cepstral analysis was done using the Cepstrum, which is a tool to check the time lags between the frequency bands. The results of the cepstral vectors were analyzed with Leven's test, and normality was checked with the Kolmogorov-Smirnov test. The difference between classes was measured by the Kruskal-Wallis test ($p<0.05$) as the data distribution was not close to the parametric tests' assumption. IMFs were used to calculate twenty-two statistical features, and using the Kruskal-Wallis test, five characteristics were chosen as important characteristics in a SZ

study [85]. The Kruskal–Wallis test is used to choose highly discriminating features in another SZ study [51].

An Epileptiform discharge analysis was demonstrated to detect AD in the sleeping state in [41]. The IBM SPSS 20 software was used to perform the statistical analysis. Any discharge was reviewed as Epileptiform discharge (ED) (localized spikes and waves; spike-and-wave complexes; ≥ 3s rhythmic synchronization) when both raters analyzed it so. The daily distribution of the EDs was shown as the index of detection probability and defined as the amount of EDs at a certain time of the day divided by the entire number of EDs saved during a complete day of recording.

Fraga *et al.* [33] applied the amplitude modulation analysis technique (AMAT) constructed on the spectra of temporal signal processing for AD classification. After performing the HT, amplitude envelopes were measured from the five subbands. A secondary frequency decomposition was conducted to pick the temporal dynamics of the subbands using the second-order band-pass modulation filters. Then cross-frequency modulation interaction was explored by the percentage of modulation energy found at a given frequency. There were 5 scenarios, which represent the modulation bands relative to the energy. The cross-frequency modulation parameters were measured for each of the 19 EEG channels. Using the Jarque-Bera test [47], the normality was measured, and the critical values were calculated by Monte-Carlo simulation [26]. One-way ANOVA was performed for normally distributed features. When the normal distribution was not clear for some of the features, the Kruskal-Wallis test was utilized. For correcting the multiple comparisons, the Dunn-Sidak post hock test was performed. Pearson correlations were performed between two classes (HV and AD).

A fuzzy and associative Petri net (APN) technique was proposed to design an algorithm for AD classification in [24]. The entropy of all the extracted features was measured by the fuzzy membership function named minimize entropy principle approach (MEPA) [74]. Different associative production rules (ARPs) are associated to create the APN model. Each input state's true preposition was set by the fuzzy membership function, and using the certainty function, the true extent of the consequence preposition was found.

Jalili and Knyazeva [46] implemented the graph-theoretic measure to classify SZ subjects. The brain's functional networks were constructed first using the connectivity

matrices. There were direct and indirect links to represent the functional connectivity. The Pearson correlation coefficient was calculated for all possible pairs. Then the network was measured using the binary adjacency matrices. Using the small worldness index, which measured functional segregation and integration, the modularity index, the centrality measure, resiliency measures, and synchronizability measures, the brain graph network was built to classify. The statistical difference between the SZ and HVs was measured by the Wilcoxon rank sum test.

In [56], using the extracted inherent spatial pattern network (SPN) features of the resting state and P300 task state, it was constructed to differentiate SZ from HVs. EEG data analysis started with noise removal, and then P300 trails were trail averaged for each patient. Brain networks were built using the PLV. Brain network properties such as characteristic path length, global efficiency, and local efficiency clustering coefficient were measured. The SPN approach was used to pick the discriminative spatial patterns of the brain network. A 16 × 16 adjacency matrix was constructed to compute SPN filters. Then the SPN filters were used for training to separate SZ from HV.

2.6 Comparative Result and Analysis

EEG is a promising tool to work with neurodegenerative disorders. One can say that it is an electrophysiological brain monitoring system [99]. It is not only cheaper [92] than other modalities like MEG, PET, MRI, etc. but also easy to deploy. And it has been using it for almost 100 years now. Hans Berger [35] recorded the first human EEG recording in 1924. Since then, EEG has become a trendy tool for biomedical researchers. A summarized comparative result and analysis based on recent literature in relation to the MCI detection are narrated below:

Table 2.2 presents some of the recent studies done using the EEG data to detect MCI. It has been noticed that not all the studies manage to have a healthy sample size to work with. And that is the main reason for choosing ML algorithms for the classification task. Traditional ML algorithms (e.g., SVM, KNN, LR, DT, LDA, NB, etc.) perform well, but the sample size is not large. Recently proposed ELM architecture has outperformed all the methods with 98.78% accuracy. A DL-based BLSTM network achieved high accuracy, though. And from the summary, it is observed that SVM has been used in most of the studies, and it performed excellently while classifying MCI subjects. DL-based approach BLSTM performed well with a small dataset, and ELM exhibits the highest

performance in terms of accuracy. Other [92] ELM, KNN, and LR performed well. As FFT is not suitable for non-stationary signals like EEG, it had an impact on the accuracy.

Very few recent studies were conducted to detect dementia. **Table 2.3** reports those studies where KNN with CWT showed promising results. But failed to produce high accuracy when the subjects were more than [92] 70 years old. For the below-70-year-old group, the same models accuracy decreased, but again, with CWT, it had better accuracy. The proposed models were not compared with any of the existing solutions. One of the studies did not mention the sample size.

Most of the state-of-the-art literature was on AD detection using EEG data, which is portrayed in **Table 2.4**. Some studies managed to have a good sample size, but some struggled. ML algorithms were mostly used for the classification task, and among them SVM and DT performed well. There were some DL approaches that ended up having good accuracy as well. New and modified approaches like dNAT, EPNN, sLORETA, AMAT, etc. have shown promising results too. But some of the studies did not show any performance comparison with their proposed method.

Table 2.5 provides a summary of some recent works that were conducted to detect schizophrenia from the EEG data. I have found that most of these studies struggled to have a good sample size. Graph theory has been used for the first time to detect SZ and has shown promising accuracy. Again, ML algorithms like SVM, LDA, and DT were used and performed well. Altered approaches like Ensemble Bagged Tree (EBT) proved their ability to classify SZ subjects quite accurately. A cubic kernel-based SVM achieved the highest accuracy of 98.9% with 28 subjects (14 SZ subjects and 14 HVs). These algorithms are not suitable when working with a larger dataset or if the dataset contains some artifacts.

Only a few recent studies have used EEG data to automatically classify Parkinson patients. **Table 2.6** illustrates those studies and their methods and classifiers, along with their performances. CNN and SVMrbf did well in classifying PD subjects. In one of these studies, SVMrbf achieved 99.62% accuracy, but the sample size was not large.

Table 2.2: Overall summary of EEG based MCI studies.

Authors	**Number of subjects**	**Methods**	**Classifiers**	**Performance/result**	**Challenges**
Sharma *et al.* [75]	44 subjects (15 dementia subjects, 16 MCI subjects and 13 HVs)	ICA and WT, ANOVA	SVM combined with FTT and CPT	-MCI versus Dementia: FTT 84%, CPT 88% -MCI versus HV: FTT 89.8%, CPT 73.4%, 84.1% when eyes open state	Using different EEG features can improve the accuracy and there was lack of boundaries between HV, MCI, and dementia
Kashefpoor *et al.* [50]	27 subjects (11 MCI patients and 16 HVs)	Correlation based pursuit	Sugeno neuro-fuzzy (NF) inference system with KNN	88.89% accuracy, 100% sensitivity, and 83.33% specificity	Small sample size
Khatun *et al.* [52]	23 subjects (15 HVs and 8 MCIs)	ERP, CFV and RP	SVM and LR	87.9% accuracy, 84.8% sensitivity, and 95% specificity (using SVM)	SVM has a shallow architecture and the sample size is too small to work with
Yin *et al.* [100]	22 subjects (11 MCI patients and 11 HVs)	SWT, 3-D evaluation algorithm	SVM, KNN, and DT	-SVM: 96.94% accuracy, 96.89% sensitivity, and 96.99% specificity -KNN: 96.89% accuracy, 97.25% sensitivity, and 96.54% specificity -DT: 95.47% accuracy, 95.38% sensitivity, and 95.55% specificity	Small sample size
Mazaheri *et al*. [120]	36 subjects (25 MCIs and 11 HVs)	TFR, Cross-frequency representation, ANOVA	LDA and SVM	-SVM: d-prime of 2.51 - LDA: d-prime of 1.75	Anomalies present in the MCI dataset and smaller sample size
Poil *et al.* [121]	86 subjects (61 MCI patients and 25 AD patients)	ICA and HT	LR	88% sensitivity, 82% specificity	No other classifiers were used to compare

Table 2.2 (continued)

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
Sharma *et al.* [75]	35 subjects (7 MCI patients and 28 AD patients)	PCA	BLSTM	91.93% accuracy (with 40–60 age group), 65.73% accuracy (with >60 age group)	No other classifiers were used for comparison and the performence starts to decline with different age groups
Siuly *et al.* [84]	27 subjects (11 MCI patients and 16 HVs)	SWT, PAA, AR and PE	ELM	98.78% accuracy, 98.32% sensitivity, and 99.66% specificity	Computationally expensive model to build

Table 2.3: Overall summary of EEG based Dementia studies.

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
Anuradha *et al.* [14]	Not mentioned	FFT	Dominant frequency analysis	DF in α range was present in 30%, DFV more than 3.5 Hz in 68% of the epochs and DF in θ range more than 40% of the epochs	No comparison with any well-known classifiers and weak methodology
Durongbhan *et al.* [29]	40 subjects (20 HVs and 20 dementia cases)	Butterworth Filter, FFT, CWT	KNN	-With FFT > 78% accuracy and with CWT > 80% accuracy (for all group) -With FFT > 89% sensitivity and with CWT > 90% sensitivity (for below 70 groups) -With FFT > 81% specificity and with CWT > 83% specificity (for below 70 groups)	The model does not seem to be more accurate and no comparison with other classifiers

Table 2.3 (continued)

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
				-With FFT > 64% sensitivity and with CWT > 67% sensitivity (for above 70 groups) -With FFT > 84% specificity and with CWT > 85% specificity (for above 70 groups)	

Table 2.4: Overall summary of EEG based AD studies.

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
Fiscon *et al.* [32]	109 subjects (86 AD and MCI and 23 HVs)	FFT and DWT	DT (C4.5), NB	DT with DWT HV VS AD: 83.3% accuracy HV VS MCI: 91.7% accuracy MCI VS AD: 79.1% accuracy DT with FFT HV VS AD: 72.2% accuracy HV VS MCI: 71.7% accuracy MCI VS AD: 80.2% accuracy	FFT is not suitable for nonstationary signal like EEG
Kim and Kim [53]	20 subjects (10 AD patients and 10 were HVs)	RP	DNN	Correct classification rate > 50%	Very weak method and accuracy. No comparison with any other classifiers.
Watanabe *et al.* [96]	144 subjects (102 HVs and 42 AD patients)	Triple correlation value, index value	dNAT	70% accuracy	Very weak method and accuracy. No comparison with any other classifiers

Table 2.4 (continued)

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
Al-Jumeily *et al.* [5]	62 subjects (32 HVs and 30 AD patients)	HT, Cross-correlation value, Magnitude squared coherence	PCA	No performance was measured	Very weak method and accuracy. No comparison with any other classifiers
Houmani *et al.* [42]	169 subjects (58 MCI subjects, 49 AD subjects, 22 SCI subjects and 40 other pathology subjects)	Butterworth filter, Epoch-based entropy measure, WT, and Sparse Bump model	SVM	-SCI versus AD: 91.6% accuracy, 100% specificity, and 87.8% sensitivity -SCI versus AD versus Other: 81.8% ~ 88.8% accuracy	More features were needed for classifying AD from Other pathologies and the model failed to discriminate all these pathologies from the MCI stage
Amezquita-Sanchez *et al.* [13]	37 AD patients and 37 MCI patients	MUSIC-EWT, FD, HE, ANOVA	EPNN, DT, NB, KNN	90.3% accuracy (EPNN)	No HV participated in the study
Bagattini *et al.* [16]	26 AD patients	SOUND, ICA	sLORETA	Specificity index = 0.63 and sensitivity index = 0.92	Only one group of subjects and no comparison with other classifiers
Bi and Wang [18]	12 subjects (4 HVs, 4 MCIs, and 4 AD patients)	FFT	DCssCDBM and SVM	-DCssCDBM: 95.04% accuracy -SVMrbf: 78.3% accuracy	Very small sample size
Rodrigues *et al.* [73]	38 subjects (11 HVs, 8 MCI patients, and 19 AD subjects)	DWT	Cepstral analysis	Statistically significant parameters variation p value = 0.003 < 0.05	No comparison with any other classifiers

Table 2.4 (continued)

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
Horvath *et al.* [41]	5 AD patients	ANOVA	Epileptiform discharge analysis	-Correlation between the sensitivity and the length (r = 0.972, p = 0.005) -Average sensitivity of an 8-h EEG-epoch was $\geq\geq$ 0.8 -During 82% Epileptiform discharges occurred	Very small sample and only one group of subjects
Fraga *et al.* [33]	76 subjects (27 HVs and 49 AD subjects)	HT, Cross-frequency connectivity, ANOVA	AMAT	-Delta modulation of the beta frequency band disappeared with an increase in disease severity (from mild to moderate AD) -Delta modulation of the theta band appeared with an increase in severity -Delta modulation of the beta frequency band showed to be a reliable discriminant feature between HVs and mild-AD patients	No comparison with any existing works or classifiers
Chiang and Pao [24]	22 HVs and 1 AD patient	PSD, FT	APN, DT, NN, SVM, Bayes Net (BN)	-APN: 96.25% accuracy -SVM: 94.81% accuracy -NN: 89.34% accuracy -DT: 95.39% accuracy -BN: 94.81% accuracy	Very small sample size with only one AD subject
Aghajani *et al.* [2]	34 subjects (17 HVs and 17 AD subjects)	Source localization, FFT, WRS	SVM	SVM: 84.4% accuracy, 75% sensitivity, and 93.7% specificity	No comparison shown with other classifiers

Table 2.5: Overall summary of EEG based Schizophrenia studies.

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
Jalili and Knyazeva [46]	28 subjects (14 SZ subjects and 15 HVs)	Partial-unpartial correlation, WRS	Graph theory	-Unpartial correlation values were differentiated between SZ and HV groups and the percentage of sensor pairs correlation values (p<0.05) different from 8.73% in β band and 4.04% in δ band -Partial correlation values were differentiated between SZ and HV groups and the percentage of sensor pairs correlation values (p<0.05) different from 7.72% in β band and 5.03% in δ band	No comparisons were shown between previous works or classifiers
Boostani *et al.* [20]	13 SZ subjects and 18 HVs	AR coefficients, Band power, FD	BDLDA, LDA, SVM, Adaboost, FSVM	-BDLDA: 87.51% accuracy -LDA: 85.36% accuracy -Adaboost: 85.41% accuracy -SVM and FSVM:<50% accuracy -LDA: 85.36% accuracy -Adaboost: 85.41% accuracy -SVM and FSVM:<50% accuracy	The accuracy was affected due to outliers present in the dataset
Jahmunah *et al.* [45]	14 SZ patients and 14 HVs	2nd order Butterworth filter, Student's *t*-test	DT, KNN, LDA, PNN, SVMrbf	-SVMrbf: 92.91% accuracy -KNN: 90.3% accuracy -DT: 88.3% accuracy -PNN: 86.33% accuracy -LDA: 78.2% accuracy	Small sample size and SVM has a shallow architecture
Phang *et al.* [67]	45 Schizophrenia patients and 39 HVs	VAR coefficient, PDC	SVM, MDC-CNN	-MDC-CNN: 93.06% accuracy -SVM: 78.60% accuracy	Computationally expensive model

Table 2.5 (continued)

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
Li *et al.* [56]	23 SZ subjects and 25 HVs	HT, REST, PLV, P300 amplitude	SPN, SVM, LDA	SPN: 90.48% accuracy, 89.47% sensitivity, and 89.47% specificity	Small sample size
Devia *et al.* [28]	20 subjects (11 SZ patients and 9 HVs)	Butterworth filter, ERP, WRS	LDA, Rule based LDA	Combination of LDA and rule based LDA: 71% accuracy	Small sample size and no comparison with other classifiers
Baradits *et al.* [17]	60 SZ patients and 76 HVs	Butterworth filter, ICA, FFT	HLM	False discovery rate corrected, $p < 0.05$	No comparison with other classifiers or previous studies
Das and Pachori [25]	28 subjects (14 SZ patients and 14 HVs)	Student's *t*-test	SVM (Cubic)	Accuracy: 98.9%, Sensitivity: 99.00%, Specificity: 98.8%	Small sample size and a shallow architecture
Siuly *et al.* [85]	49 SZ patients and 32 HVs	EMD and Kruskal Wallis test	Ensemble Bagged Tree (EBT)	Accuracy: 89.59%, Sensitivity: 89.76%, Specificity: 89.32%	Complex model to build
Khare *et al.* [51]	26 SZ patients and 22 HVs	EWT and Kruskal Wallis test	SVM	Accuracy: 88.7%, Sensitivity: 91.13%, Specificity: 89.29%	Shallow architecture

Table 2.6: Overall summary of EEG based Parkinson studies.

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
Oh *et al.* [61]	20 PD subjects and 20 HVs	A threshold-based technique	CNN	88.25% accuracy, 84.71% sensitivity and 91.77% specificity	No comparison with other classifiers and computationally expensive model

Table 2.6 (continued)

Authors	Number of subjects	Methods	Classifiers	Performance/result	Challenges
Yuvaraj *et al.* [101]	20 PD subjects and 20 HVs	A threshold-based technique, HOS	DT, FKNN, KNN, NB, PNN, SVM	-SVMrbf: 99.62% accuracy, 100% sensitivity, and 99.25% specificity -PNN: 93.98% accuracy, 96.21% sensitivity, and 90.61% specificity -NB: 88.72% accuracy, 94.17% sensitivity, and 80.83% specificity -KNN: 90.83% accuracy, 96.67% sensitivity, and 82.50% specificity -FKNN: 93.01% accuracy, 95% sensitivity, and 90.21% specificity -DT: 90.65% accuracy, 94.17% sensitivity, and 85.63% specificity	SVM has a shallow architecture
Pasha and Latha [64]	188 PD patients and 64 HVs	GA and BPSO	LR, Gaussian Process Classifier (GPC), Multilayer Perceptron (MLP)	GA-LR Accuracy: 86.8% GA-GPC Accuracy: 86.8% GA-MLP Accuracy: 85.1% BPSO-LR Accuracy: 85.5% BPSO-GPC Accuracy: 86.8% BPSO-MLP Accuracy: 89%	Computationally expensive model

2.7 Challenges in Identifying Brain Disorders

EEG signals are not periodic and stationary like other regular signals. There are certain features and attributes in EEG signals. It has five frequency bands: delta (δ) from 0.1 Hz to 4 Hz, theta (θ) from 4 Hz to 8 Hz, alpha (α) from 8 Hz to 12 Hz, beta (β) from 12 Hz to 30 Hz, and gamma (γ) > 30 Hz [106]. Possible challenges starting from EEG recording to EEG classification are reported below:

2.7.1 Challenges in EEG Data Recording

Previous efforts to detect those five neuro-diseases have suffered from poor performance, instability, and cost effectiveness in terms of time. It is because I am not working with the raw EEG signals properly and tuning the classifier ineffectively. In this effort, I have also reported those challenges and tuning points that need to be taken care of attentively to ensure an accurate and efficient neuro-disorder detection method.

Managing medical data for research is quite hard. So, the initial challenge is to find a good number of case subjects and HVs. Here, case subjects indicate patients with any neuro-diseases that I want to automatically differentiate from HVs. And HVs mean humans of any gender without brain diseases. Researchers need to find a sufficient number of subjects to conduct the EEG recording sessions. While selecting the subjects, there are some exclusion criteria that need to be followed:

- Head trauma
- A yesteryear of major psychiatric disorders
- Substance misuse
- Any medication that affects cognition
- Other special requirements based on the case
- Other significant medical condition

The Petersen criteria means a neuropsychiatric interview named the Mini-mental State Examination (MMSE) has to be performed for each of the subjects. Usually, the MMSE score of the HVs stays higher than 26, and less than 26 is the score of the case subjects. After all these criteria have been met, ethical approval has to be obtained from each of the subjects and the hospital or managing authority. After doing so, an EEG cap can be placed on the subject's scalp. Just before that, it has to be determined how many electrodes will be used to record the electrical activity of the brain. There are a couple of options,

like the international 10-20, 10-10, 10-5, etc. systems. They have 21, 64, and 128 electrodes set up, respectively. To find the perfect set up according to the work is a big challenge in EEG recording. **Fig. 2.6** illustrates the positioning of the 21 electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, A1, T3, C3, Cz, C4, T4, A2, T5, P3, Pz, P4, T6, O1, O2) according to the international 10-20 system, which is the most widely accepted and used set up to conduct EEG recordings.

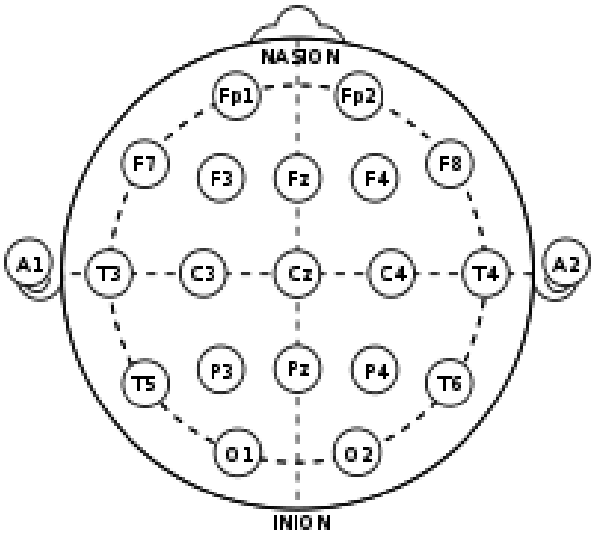

Fig. 2.6: Electrodes position according to international 10-20 system.

The next thing to work on before recording is setting the electrodes skin impedance. It is preferable to set fewer than 5 kΩ, otherwise, there might be some white noise included in the signal. Then, choosing the right sampling frequency is another decision to make. The standard sampling frequency for the aforementioned five neuro-disease studies is 128–256 Hz. Finally, and lastly, the most challenging part of EEG recording is keeping the subject conscious, as it may last for 10–30 minutes.

2.7.2 Challenges in EEG Signal Pre-processing

This stage is to process the raw data before feeding it into the classifier for classification. Raw EEG recordings are frequently contaminated by artifacts or "unwanted signals". Some of the most typical reasons for contaminating an EEG recording are outlier readings, baseline drift, electrode-pops, power supply fluxes and interference (50 Hz), breathing, eye blinking, or muscle electrical activity, among others. If I want to create an accurate detection system, I must first remove the noise from the recording. It is reported as **research problem #1** in this dissertation. Now the initial struggle begins with the noise reduction algorithm. There are several methods for removing artifacts that have been

proven to work. Buttherworth filtering, Wavelet Transform (WT), Furrier Transform (FT), Independent Component Analysis (ICA), Wavelet Enhanced Independent Component Analysis (wICA), etc. are some of the common options for noise and dimensionality reduction. **Fig. 2.7** is a showcase of an MCI subject's (a) raw EEG signal and (b) de-noised signal.

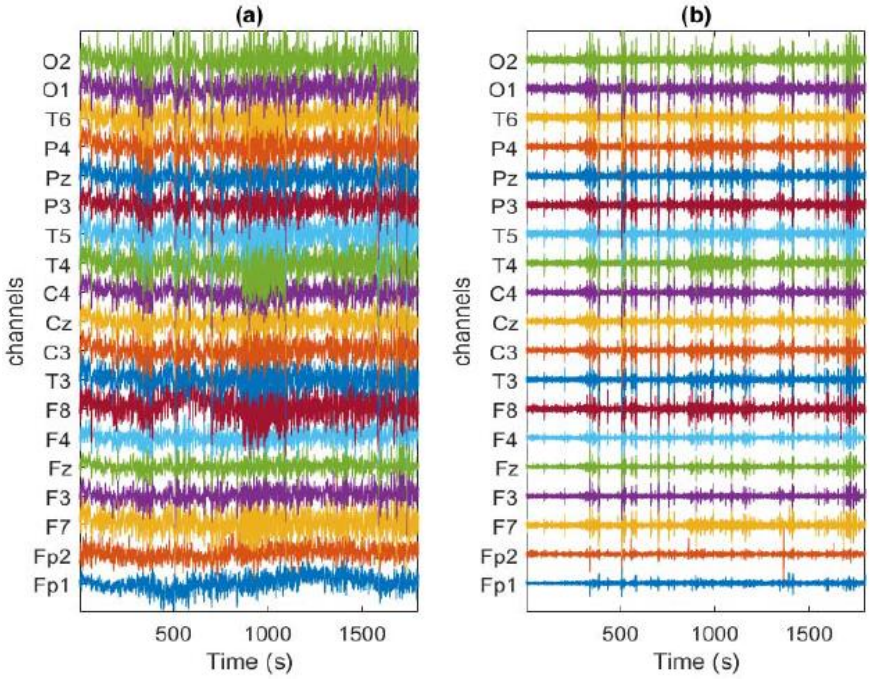

Fig. 2.7: EEG recording of an MCI subject (a) raw EEG signal (b) de-noised EEG signal of the same MCI subject [105].

Based on the classifier and computational resources, the next steps vary. If a traditional machine learning (TML) classifier is picked, then it requires features from the data to be manually extracted using time, frequency, or time-frequency domain methods. This additional feature extraction and selection add computational burden and it is reported as **research problem #2**. Some of the time-domain options for feature extraction include: Phase Locking Value (PLV), P300, Event Related Potential (ERP), Auto regressive (AR) coefficients, etc. are some of the time-domain options for feature extraction. Candidate Feature Vector (CFV), relative power (RP), Hilbert transform (HT), fast Furrier transformation (FFT), etc. are widely used frequency domain methods. WT, Music-Empirical wavelet transformation (EWT), CWT, DWT, etc. are some time-frequency domain analysis methods.

Now, considering the computational power, some extra pre-processing is required, like segmentation, compression, or down-sampling. The motivation for these

steps is to break down the massive EEG dataset without losing important features. There is no well-established method for segmentation. Segmentation is mostly done based on epoch size. E.g., from 30 minutes of EEG recording of a subject, I will consider each epoch of 5 seconds in duration, which will be a temporal segment. For compression Piecewise Aggregate Approximation (PAA) is a widely used method for EEG data. Well-known average [7] or median filtering [9] can be used for down-sampling. But it is not recommended to down-sample the data, as it compromises important features if the computational power can handle the data size. After completing this pre-processing stage, all the filtered data or features are fed into the classifier.

2.7.3 Challenges in EEG Classification

Establishing and improving ML algorithms has taken a lot of time and effort. DL, which is the latest branch of ML, has enabled us to work with huge datasets like EEG. Though there has been a lot of work done using classical ML classifiers like SVM, KNN, LR, NB, DT, LDA, etc. Due to their shallow architecture and linear pattern, these classical ML algorithms struggle to deal with huge amounts of EEG data. The shallow TML classifiers are regarded as **research problem #4** in this dissertation. Additionally, when it comes to multi-class classification, TML classifiers failed miserably. Therefore, lack of accurate multi-classifier has been noted as **research problem #3.** And from there, DL comes to enhance its support. So far, DL methods have not been explored like classical ML methods. Few studies have been done using CNN, recurrent neural network (RNN), the Boltzmann machine (BM), and the extreme learning machine (ELM).

Choosing the right classifier is the very first challenge here. Though it is like a "black box" problem. It is hard to predict the performance of the chosen classifier at the beginning. And there are many parameters associated with the classifier itself. Some of the parameters vary from classifier to classifier. The performance often got affected due to inappropriate classifier and its parameter selection. Hence, it is selected as **research problem #5** in this dissertation. This study reports the common struggles in EEG classification. The least complicated task is to divide the dataset for training, testing, and validation purposes. 10-fold, 5-fold, and leave-one-out cross-validation (LOOCV) are some of the preferred and accepted validation methods to prove the stability of the classifiers.

2.8 Discussion

Fine-tuning the parameters of the classifier can increase its performance. But it requires a lot of time. To reduce the training time, normalization can be done with the dataset. It will reduce the size of the data and provide redundancy over the main memory. Fine tuning involves changing the kernel, order, batch size, number of epochs, activation function, window size, etc. Another challenge lies in the coding. MATLAB and Python-based editors like Pytorch, Spider, Pycharm, and Jupyter Notebook are some of the commonly used tools for EEG classification. As Python has better support for DL, most recent studies have been done in a Python environment.

EEG analysis has drawn the attention of many researchers for neuro-disease detection. Multiple EEG classification frameworks have been demonstrated in [104], [105], and [107]. All the steps discussed in this study have taken place inside [104], [105], and [107]. But there are certain things that need to be taken care of carefully to ensure a quality detection mechanism. It all starts with the EEG data recording. Arranging a sufficient number of subjects for the study is one of the big challenges here. Certain exclusion criteria are there to eliminate illegitimate subjects from the study. As the EEG recording goes on for a long time, keeping the selected subjects stable and quiet for a long time is another challenge.

After EEG signal recording, the artifacts have to be removed from the signal for further processing. Removing these noises with a proper noise filter is a hard task. Further processing involves feature extraction, segmentation, compression, etc. If a DL-based classifier is chosen, then just after de-noising, the data is ready to be fed into the classifier. Otherwise, features have to be manually extracted using some time- or frequency-domain algorithms. Selecting the right feature extraction algorithm is challenging, and the quality of the extracted features mostly depends on the quality of the data after de-noising. Segmentation and compression are done to reduce the computational overhead.

This effort also explores recent state-of-the-art literature that was done using EEG signal data to detect neurodegenerative disorders. The elementary aim of this work is to review the recent studies that were designed for predicting or detecting MCI, dementia, AD, SZ, and PD. It is found that most of the studies tussled to have a good sample size for their work, as I know it is not easy to acquire medical data for research. As the availability of the internet has increased, people have become even more dependent on it

[12]. The trend is now towards EEG-based online automation. The existing tools like PET, MRI, and MEG are expensive and time-consuming as well. Therefore, scholars are now mostly interested in finding any pattern using the inexpensive EEG data for neurological diseases.

Among the literature, ML-based algorithms are very illustrious and have shown promising accuracy in classifying. More specifically, SVM, DT, and KNN had impressive performance over classifying neuro-biomarkers. This is because of the sample size. SVM, DT, and KNN algorithms perform well when the dataset is not large. Deep learning approaches were used in very few works and demonstrated competitive performance. But DL algorithms are computationally expensive, and they perform better when the sample size is large. On the other hand, PNN, BN, NB, and NN did not perform well in any of these five neuro-disorders classification.

Succeeding studies should focus on working collaboratively with hospitals and other researchers in this field, which can help in having a diverse and large dataset. One MCI study [70] was found with a decent data size. There were few AD studies [32], [96], [5], [42], [13], and [33] that were conducted with a reasonable sample size. Three SZ works [85], [67], and [17] had passable data sizes. On the other hand, [50], [52], [100], [58], [32], [16], [18], [41], [24], [45], [56], and [28] had a small sample size. Having a diverse and large dataset enables the model to be more accurate while classifying the subjects. And also, ML algorithms need the extracted features to be fed into the system as input, which is another reason not to have a good, accurate system. I have seen some DL approaches such as neural networks (NN) used by [24], CNN proposed by [61], [27], [67], DCssCDBM utilized by [18], probabilistic neural networks (PNN) introduced by [101], [45], enhanced probabilistic neural networks (EPNN) implemented by [13], deep neural networks (DNN) employed by Kim and Kim [53], bidirectional long short-term memory (BLSTM) applied by [86], etc. The raw data consist of artifacts and other noises. And these noises are not mixed with the raw signal in a unified manner. Therefore, first detecting the noise and then removing it from the signal should be the goal. These types of algorithms are known as adaptive noise-removing algorithms. Having a good quality adaptive noise removing algorithm plays an important role in achieving high accuracy. Moreover, DL algorithms are good when the sample size is large, and they do not need the extracted features to be fed. DL algorithms extract features for themselves according to the requirements. But these algorithms are computationally expensive.

2.9 Essence

The classification stage does not have the huge challenges of the previous stages if the EEG data is processed properly. The classifier may struggle to perform well. But proper fine tuning can increase the performance, which is the challenging part. There are lots of parameters when I talk about fine-tuning a classifier. Parameters include kernel, order, batch size, number of epochs, activation function, window size, etc. These have an impact on the learning rate of the classifier as well as the training time.

EEG has opened a new door towards brain disease research. There has been an extensive amount of work going on to detect different brain diseases at an early stage. Some neuro-patients only live for 5–8 years after being medically confirmed. EEG is a very easy biomarker to investigate. This study is a pathway for upcoming EEG research. Challenges and common struggles are discussed, along with possible solutions. New studies will find it interesting and easy to work with EEG data.

It is an important area of medical science that needs to be taken seriously and with high priority. And the properties of the brain signal data, which are also known as EEG data, have been well studied and better understood. Having an automated and well-accurate system for the elderly is a must now. After analyzing efforts to classify neuro-diseases like AD, MCI, dementia, SZ, and PD, I have seen hand-engineered ML algorithms dominate the ladder. They are preferable if the sample size is not too large. But they cannot extract the features by themselves. This is where the DL methods rule out the TML methods. DL methods have the ability to extract the features by themselves and feed them into the system. But they really perform well when I have a large dataset, ensuring there is enough data left for the training. Few of the reported studies used DL algorithms, and they performed well because the authors segmented the dataset. Segmentation helped those DL methods increase the number of samples. From the reviewed articles, I have implied five open research questions (ORQs): (1) How can I manage to have a large dataset for a universal classifier? (2) How can I reduce the noises and

artifacts using a suitable method from the EEG data? (3) What are the modifications required in the DL models to save time in terms of identifying cases? (4) How can I accumulate the huge volume of EEG data in the classifier more efficiently? Which algorithm suits well with the EEG data so that I get good accuracy? Therefore, succeeding studies must pay attention to having an automated system that will have good accuracy in detecting neurodegenerative disorders. ML algorithms with a shallow architecture have performed well with a small sample size. Almost every study struggles to have a large dataset. Some studies also grappled with the noise of the EEG data. Therefore, succeeding work should place more emphasis on having a large dataset and an adaptive noise-removing filter for EEG data.

From the next chapter onwards, I have reported all the investigations and frameworks that I have developed. Each of those methodologies has been compared and justified with proper validations and performance matrices. Furthermore, with the exception of Chapter 7, all of them have been published in top journals and at previous conferences.

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: A Long Short-Term Memory Based Framework for Early Detection of Mild Cognitive Impairment From EEG Signals

Surname: ALVI First name: ASHIK MOSTAFA

Institute: Institute for Sustainable Industries and Liveab Candidate's Con 60

Status:

Accepted and in press: ☐ Date:

Published: ☑ Date: 08/07

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be includ ned in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted] 24/01/2023

Signature **Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

Name(s) of Co-Author(s)	**Contribution (%)**	**Nature of Contribution**	**Signature**	**Date**
Ashik Mostafa Alvi	60	Conceptualization, Methodology, Data Curation, Investigation, Software,	[redacted]	24/01/2023
Siuly Siuly	20	Formal Analysis, Validation, Writing–Review and Editing, Supervision.		24/01/2023
Hua Wang	20	Resources, Supervision, Writing–Review and Editing, Funding Acquisition, Project		26/01/23

Updated: September 2019

CHAPTER 3 : A LONG SHORT-TERM MEMORY BASED FRAMEWORK FOR EARLY DETECTION OF MILD COGNITIVE IMPAIRMENT FROM EEG SIGNALS

3.1 Overview

Mild cognitive impairment (MCI) is an irreversible progressive neuro-degenerative disorder that seems to be a precursor to Alzheimer's disease (AD) and may lead to dementia in elderly people. It is a major public health challenge for healthcare in the 21st century. Because there is no cure or therapy to halt or reverse the course of MCI, early identification is critical for successful treatment programmes to enhance patients' quality of life. Currently, electroencephalography (EEG) has emerged as an efficient tool to investigate MCI. Traditional methods for finding MCI from EEG data use shallow machine learning (ML)-based architectures that cannot find important biomarkers in deep, hidden layers of the data and also have trouble dealing with a large amount of EEG data. To reduce this issue, this research will use EEG data to provide a deep learning (DL)-based framework using the long short-term memory (LSTM) model for effective identification of MCI individuals from healthy volunteers (HVs). The suggested framework consists of four phases: de-noising, segmentation, down-sampling, uncovering deep hidden features using the LSTM model, and identifying MCI patients with the sigmoid classifier. This study has designed 20 different LSTM models and investigated them in a publicly available MCI database to find the best one. After performing 5-fold cross validation, the best model achieved 96.41% accuracy, 96.55% sensitivity, and 95.95% specificity. The proposed LSTM-based DL model provides a robust biomarker and will guide technologists in creating a new automatic diagnosis system for MCI detection. It also upholds the answer to **research problems 4 and 5**.

The contents of this chapter have been published in the *IEEE Transactions on Emerging Topics in Computational Intelligence* [105].

3.2 Introduction

Mild cognitive impairment, sometimes referred to as MCI, is a gradual neurological illness that is characterized by cognitive degradation. Memory loss is more severe in those with MCI than in others in a comparable age group. Continuous memory degradation, decreased vocabulary, and a lessened potentiality to fulfil accurate motor movements, which all damage carrying out day-to-day affairs, are the root of what causes MCI [100]. Neuronal cell death and malfunction are the main causes of this neurological illness. According to findings from recent research, individuals with MCI have an increased risk of developing dementia and, in particular, AD [108], [109], [110]. It usually happens to people over 65, and the severity rate goes up by a factor of 10 with age [62], [30], [104], and [111]. Dementia is one of the top reasons for impairment and reliance among the elderly across the globe. MCI and dementia are growing worldwide and forecasted to escalate proportionately more in developing countries [112], [107]. It is Australia's second-leading root of mortality [113] and ranks as the 7th biggest cause of death globally [114]. By 2021, it is anticipated that 472,000 Australians would be suffering from dementia [38]. This number is expected to increase to 590,000 by 2028 and 1,076,000 by 2058 if they do not get any medical innovation. As per a survey published by Alzheimer's Disease International (ADI), there were more than 50 million dementia patients globally in 2019, with that number expected to rise to 152 million by 2050 [115]. Until now, both dementia and MCI have no known treatment options; only early diagnosis can temporarily slow the worsening of symptoms, intending to enhance the patients' and carers' wellbeing [116]. Thus, it is very important to identify MCI at an early stage so that they can get prompt medical treatment, and this is an ongoing area of research now. A patient’s life can be saved and improved by performing suitable methods with a rapid and timely diagnosis.

To diagnose MCI, there are several tools in current medical systems, such as the Mini-Mental State Examination (MMSE), Magnetic Resonance Imaging (MRI), Computed Tomography (CT) scan, blood tests, Positron Emission Tomography (PET), spinal fluid, and EEG [107]. Among these, because of its low cost, noninvasive nature, and accessibility, the EEG is commonly used to study MCI, whereas PET, MRI, and CT scan are expensive options, and the MMSE test is a manual question-answer option. EEG recordings preserve the electrical movements in the cerebral cortex relative to time, which are the primary motivators for evaluating neurological conditions. In light of this, I have

addressed the use of EEG as a valuable technique for detecting MCI at a preliminary phase.

Numerous investigations have been undertaken over the last several decades to diagnose MCI at an early stage so that it does not evolve into AD or other cognitive illnesses. Kashefpoor *et al.* [50] worked with the spectral features of EEG data, fed them to the Takagi-Sugeno neuro-fuzzy (NF) inference method with the K-nearest neighbor (KNN) classifier, and reached 88.89% accuracy in detecting MCI. The same research team had proposed a dictionary-based method [117] with a larger MCI dataset of 61 subjects, where they claimed 88.9% accuracy. Hadiyoso *et al.* [118] designed a KNN method to identify MCI and wrapped up with 81.5% accuracy. A continuous multi-stream hidden Markov model (MS-CHMM) was developed in [116]. But this study was done with only 14 subjects and achieved 95.9% accuracy. A DL-based approach was proposed in [119], where synchronization-based measurements were used to design a brain network. The connectivity matrix organization was done before supplying the data to the convolutional neural network (CNN). A total of 107 people took part in this investigation, which concluded with 92.06% accuracy. Khatun *et al.* [52] measured the event-related potentials and candidate feature vector (CFV) and extracted 590 features for their proposed support vector machine (SVM) with a radial-basis kernel. Their study obtained 87.9% accuracy. Mazaheri *et al.* [120] investigated oscillatory analyses and connections between theta and alpha/beta bands at different frequencies for pre-processing the EEG data and designed linear discriminant analysis (LDA) and SVM with a Gaussian kernel for classification. This study reported 80% sensitivity and 95% specificity. Poil *et al.* [121] performed de-noising of the EEG data using the JADE ICA algorithm and developed a supervised logistic regression (LR) model. This LR-based classifier achieved 88% sensitivity and 82% specificity. Seventy-four patients participated in an enhanced probabilistic neural network (EPNN)-based study [13], where the empirical wavelet transform (MUSIC-EWT) was used for feature extraction. This method ended up having 90.3% accuracy. Yang *et al.* [122] developed a four-layer CNN model to classify MCI patients and achieved 90.37% accuracy.

I have reviewed related literature inside this section, and a comparative discussion has been reported in Section 3.5. Most of the studies struggle to balance satisfactory time efficiency and accuracy for real-time applications. Moreover, previous approaches are based on manual feature extraction techniques and ML algorithms with shallow

architectures, which cannot draw out dense hidden characteristics of EEG signals from several ledges. Additional processing components, difficult-to-model complicated notions, and multi-level metaphors are all required by these approaches. So far, after analyzing previous studies, I have found that previous proposed models either have good accuracy by costing time or poor accuracy by minimizing time complexity. Our study has developed a new method to balance accuracy with minimum cost and time.

To resolve these issues, the paradigm proposed in this effort is based on LSTM for automatic detection of MCI from EEG signals. The main motivation for considering LSTM in this framework is because LSTM is best suited to sequential data like EEG. A deep LSTM network is composed of hidden LSTM layers that have the potential to selectively memorise essential data for a prolonged period of time and are often implemented for sequential prediction. EEG data are sequential data by nature. In our proposed framework, EEG data are de-noised using the Butterworth filtering technique. In the passing bands, it also allows for additional linear stage retort and an overall smooth report. To ensure our lodged model is quick enough at detecting MCI, the average filter has been employed to partition and down-sample the filtered EEG data. When there are no or few oddities in the dataset, the average filter outperforms all other filters. And the EEG dataset is de-noised before applying the average filter. I have designed our study so that it is a balanced algorithm by being an accurate model with less time complexity. Below is a synopsis of the accomplishments of this experiment:

- This effort is the first to use EEG signal data to construct an LSTM-based DL system for rapid identification of MCI.
- I explored average filtering technique for down-sampling to further help for faster fulfillment of the suggested framework.
- The optimal LSTM prototype for MCI detection has been obtained by evaluating twenty LSTM models.
- Our model is able to enhance classification accuracy with less computation time.

The following is the order in which the remainder of the chapter is organised: The suggested MCI identification architecture is described in detail in Section 3.3. Section 3.4 provides comprehensive results and findings. Section 3.5 offers a discussion of alternatives. This effort ends with a mention of the limitations, succeeding work plan, and essence in Section 3.6.

3.3 Proposed Framework

This study presents a LSTM-based MCI detection framework using EEG data. **Fig. 3.1** displays the architecture of the suggested MCI detection framework. The framework consists of four major parts: (1) EEG data acquisition; (2) data pre-processing (collected EEG data is then de-noised using Butterworth filtering, segmented, and down-sampled using the Average filter); (3) LSTM-based architecture development for feature extraction and detection of MCI subjects; and (4) performance evaluation. A detailed description of this entire framework's major parts is reported below:

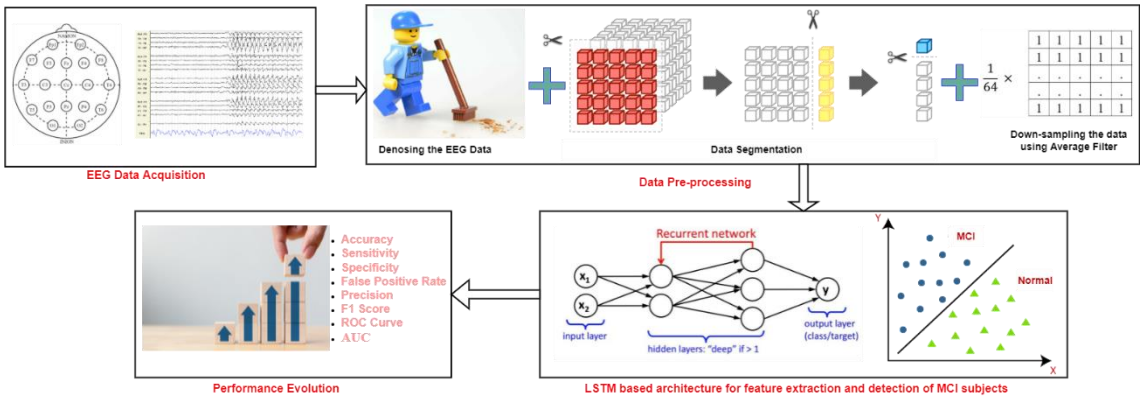

Fig. 3.1: An illustration of the MCI diagnosis system based on LSTM.

3.3.1 EEG Data Acquisition

This study employed an EEG dataset of 27 participants from the cardiac catheterization units of Sina and Nour Hospitals, Isfahan, Iran [50], [123]. This EEG dataset includes 16 HVs and 11 MCI subjects, aged between 60 and 77 years. This data collection was ethically approved by the deputy of research and technology at the Isfahan University of Medical Sciences, Isfahan, Iran, and all the subjects were well informed and gave their consent. Petersen's criteria, meaning a neuropsychiatric interview for the MCI disease had been performed for each subject. To authenticate the participants, the MMSE value was employed. MCI participants had an MMSE score of 21 to 26, whereas HVs had a score greater than 26. Drug abuse, dementia, a serious medical condition, a history of significant mental problems, or head trauma were all considered excluding factors. **Table 3.1** summarizes the gathered dataset's demographic characteristics.

The recordings were conducted in the morning in a silent room for 30 minutes while all the subjects kept their eyes closed and were in a relaxed state. The 19 electrodes (Fp1, Fp2,F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2) were placed

Table 3.1: Details Related to the Dataset's Demography.

Features	HV	MCI
Age (in years) (*mean* ± *SD*)	65.3 ± 3.9	66.4 ± 4.6
Education (years) (*mean* ± *SD*)	11.1 ± 3.0	10.3 ± 3.8
MMSE score (*mean* ± *SD*)	29.0 ± 0.8	27.6 ± 0.9
NUCOG score	91.1 ± 3.0	82.4 ± 3.6
GHQ Scores	17.9 ± 6.6	20.5 ± 9.4
BMI (kg/m^2)	26.6 ± 3.6	25.7 ± 2.2
Fasting glucose (mg/dl)	121.8 ± 36.9	115.5 ± 24.3
Total cholesterol (mg/dl)	169.1 ± 42.6	170.6 ± 61.4
Triglycerides (mg/dl)	160 ± 80.7	157.3 ± 100.9
Creatinine (mg/dl)	1.3 ± 0.3	1.2 ± 0.2
Gensini scores	20.3 ± 21.7	33.3 ± 31.9

GHQ – General health questionnaire; BMI – Body mass index; MMSE – Mini-mental state examination; NUCOG –Neuropsychiatry unit cognitive assessment tool

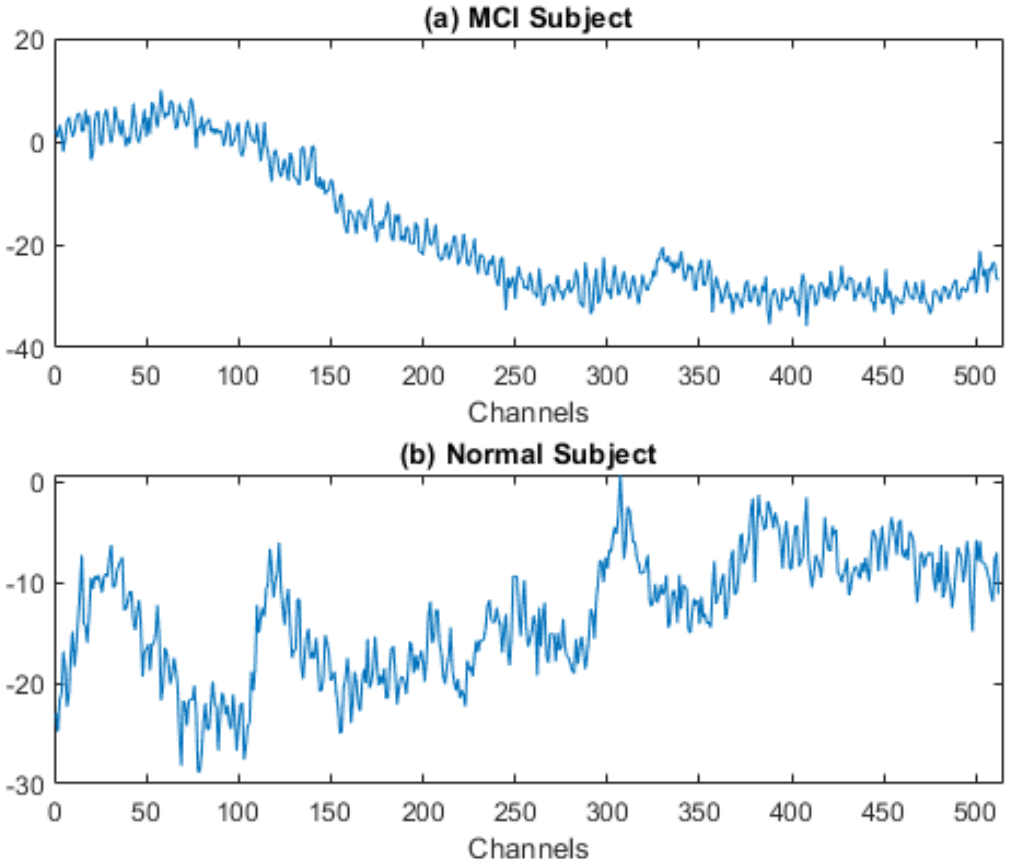

Fig. 3.2: Fresh EEG recordings of (a) an MCI person and (b) a HV for 2 seconds at the Fp2 electrode location.

over the head following the International 10-20 system, and EEG data were processed using a 32-channel EEG instrument with 256 Hz sampling frequency (Galileo NT, EBneuro, Italy). **Fig. 3.2** shows two seconds of EEG data from an MCI and a HV at the Fp2 electrode site. The electrodes' skin had an impedance that was lower than 5 kΩ. The medical expert ensured all the subjects were conscious while EEG recordings were being made to avoid drowsiness.

3.3.2 Data Pre-processing

To prepare EEG data for analysis, there are three stages to follow: (i) de-noising, (ii) segmentation, and (iii) down-sampling. All these three steps are elaborated below:

3.3.2.1 De-nosing the EEG data

EEG recordings are often corrupted by artifacts or so-called unwanted signals [9]. Electrode-pops, eye blinking, outlier readings, power supply fluxes and interference (50 Hz), baseline drift, breathing, or muscle electrical activity, among other things, are some of the most typical causes of EEG recording contamination. Therefore, it is a must to deduct the noise from the recording if I want to have a good classifier. This study has employed the Butterworth filter to deal with the noise. The Butterworth filter is used in this work to exclude undesired high-frequency signals caused by eye blinks, body movement, heartbeats, power grid inference, and other sources. And as I also know, the Butterworth filter helps to produce additional linear stage retort while maximizing the smooth reaction.

In this study, I received the raw EEG recordings of the 27 subjects in EDF. The MATLAB EEGLAB toolbox [124] is used to view the recordings and clean the noise. Butterworth is a low-pass filter mostly used to remove higher-frequency signals and is applied in [125], [126]. To eliminate artifacts, the recorded EEG data are broadband-refined between 0.5 and 50 Hz using a third-order Butterworth filter. Then, all 27 recordings are digitized with a length of 60Seconds×30Minutes×256Hz each and stored as .mat files.

The initial EEG recordings contain some unnecessary signals. After de-noising all the recordings, the EEG data become more valid and usable. **Fig. 2.7** illustrates a visual differentiation of untouched EEG recordings from an MCI participant with de-noised

recordings from the same participant. After cleaning the signals, I have just segmented the whole dataset and down-sampled it to 4 Hz.

3.3.2.2 Data segmentation

The signals produced by an EEG are not periodic, they are not steady, and they have a very large magnitude. As a result, even if the recording is segmented, representative information can be extracted from each data segment. In this study, I have digitized the signal for each subject; there are 460800 rows, and all 19 electrodes are in the columns. For processing this huge dataset of 27 subjects, I need very high computational support, which is also time-consuming. Therefore, I planned to segment each of the 30 minutes of recording and take 6 seconds of data for each segment. But the sampling frequency has remained the same, which is 256 Hz. Following the segmentation of each participant, I constructed 300 new temporal segments from each individual's recording, all of which are clearly identified with the matching participant's tag (MCI/Normal). Then, I have 8,100 segments of size 19 (channels) × 1,536 (samples), where 1,536 equals 6 seconds × 256 Hz.

3.3.2.3 Average filter to down-sample temporal segments

Our main goal is to have an accurate, DL-based MCI detection framework that can quickly generate results. For this purpose, I have reduced the sampling frequency. This study employs average filtering for down-sampling. The average filter is effective if such a dataset is devoid of distortion and oddities. Therefore, I have found it suitable for our dataset and employed it.

In this study, before feeding the filtered data to the classifier, I further down-sample it. I set the new sampling frequency to 4 Hz, and for this reason, I used the well-known Average filter. For each cycle, the sliding window approach has been employed, with a window size of 64 × 1. **Fig. 3.3** explains this down-sampling method. The algorithm has run 24 times for each channel, and the number of channels has been kept constant, which is 19.

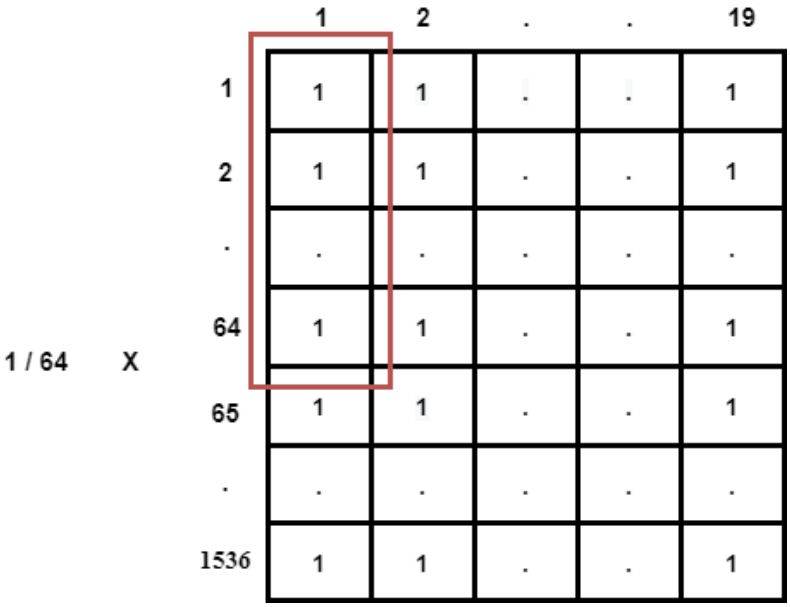

Fig. 3.3: 64 × 1 sliding window of the Average filter.

3.3.3 LSTM-based architecture for feature extraction and detection of MCI subjects

The primary goal of this effort is to differentiate MCI individuals from HVs with satisfactory accuracy using non-stationary data like EEG. I explore the LSTM network model to work with EEG data for two main reasons. (1) EEG is a very huge volume of data, and (2) the characteristics of EEG data are non-steady and chaotic. Deep learning-based classifiers perform well when fed with huge volumes of data, and they can handle non-stationary signal data like EEG well, whereas ML-based algorithms need separate feature extractions and high processing times. Because of the nature of the EEG data, I considered recurrent neural network (RNN) for its memory hosting option. Particularly, an RNN version with LSTM has been chosen.

3.3.3.1 Design of LSTM

Memorizing the history of sequential data is the unique strength of LSTM, and it helps to have more accurate prediction. LSTM is designed to resolve the declining slopes when fault derivatives are back-transmitted across several layers over time in recurrent networks [127]. The intimate state of each LSTM unit illustrates its memory. Based on the past internal states' history and the most recent input, the cells become skilled enough to publish, clean their storage, or replace it. The LSTM regulates the flow of data into and

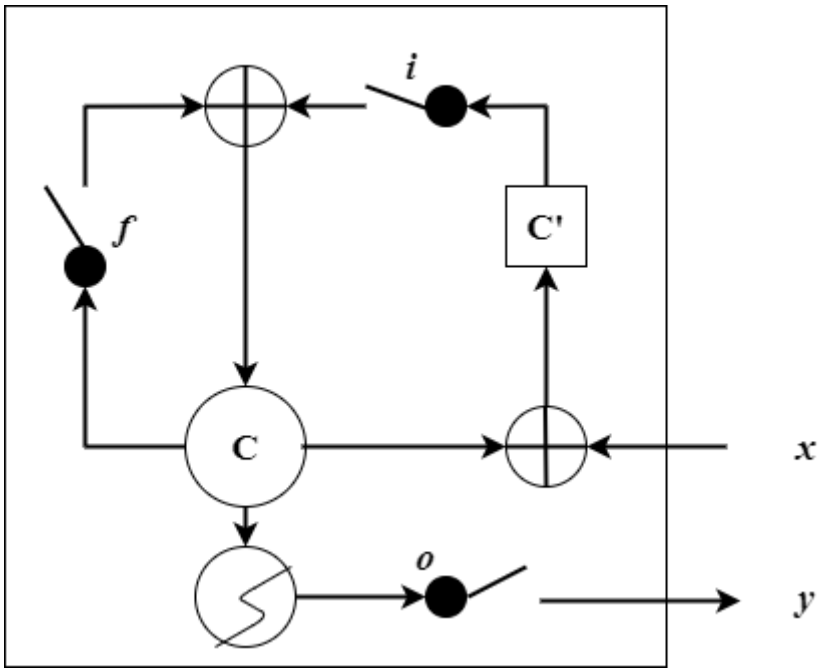

Fig. 3.4: Basic LSTM architecture.

out of its memories using three gates. The input signals at time t are represented by x, the output signals by y, the entry gate by I, the forget gate by f, and the output gate by o in **Fig. 3.4**.

Python has been chosen to build up the LSTM architecture because it provides excellent DL assistance with richer packages such as Scikit-Learn, Keras [128], Pandas, SciPy, and NumPy.

After finishing all pre-processing procedures, processed data are stored in MATLAB's CSV format. And these CSV files are taken as input to the system with the help of the Pandas library. The NumPy module is then used to transform all of the data frames into double-dimensional matrices. I have also ensured that all the labels for each of the subjects are kept in another NumPy array. After that, I reduced the dimension of the input data with the reshape function of the NumPy library. Next, the dataset is split randomly into the training and testing sets using the Scikit-Learn library. I have taken 25% of the data in our testing set. Finally, an LSTM model is developed using the Keras library with a TensorFlow backend [129]. There are many parameters available for adding layers. Twenty different models are designed with different parameters in each of those. There are 1–5 layers presented in those 20 models, with different numbers of neurones present in every layer with the same "*tanh*" activation function for all the input layers. Prior to the scheduled thick layer, there is a flattening layer. "*Sigmoid*" is employed as the

Table 3.2: Performances of different LSTM models.

Model Number: Number of Nodes in LSTM Layers	Accuracy	Sensitivity	Specificity	False Positive Rate	Precision	F1 Score	AUC
1 (128)	0.9437	0.9579	0.9332	0.0667	0.9132	0.9350	0.9456
2 (256)	0.9481	0.9684	0.9332	0.0667	0.9140	0.9405	0.9508
3 (1024)	0.9530	0.9579	0.9452	0.05479	0.9276	0.9425	0.9515
4 (128, 256)	0.92345	0.9124	0.9315	0.0684	0.9071	0.9098	0.9219
5 (256, 128)	0.9437	0.9498	0.9357	0.0642	0.9156	0.9324	0.9428
6 (128, 512)	0.9293	0.9288	0.9297	0.0702	0.9066	0.9175	0.9293
7 (512, 128)	0.9476	0.9486	0.9469	0.0530	0.9291	0.9387	0.9477
8 (256, 512)	0.9358	0.9439	0.9297	0.0702	0.9079	0.9256	0.9368
9 (512, 256)	0.9516	0.9603	0.9452	0.0547	0.9278	0.9438	0.9527
10 (256, 1024)	0.9456	0.9486	0.9392	0.0607	0.9196	0.9339	0.9439
11 (1024, 256)	0.9511	0.9509	0.9443	0.0556	0.9261	0.9383	0.9476
12 (512, 1024)	0.9540	0.9521	0.9554	0.0445	0.9400	0.9460	0.9538
13 (1024, 512)	**0.9641**	**0.9655**	**0.9595**	**0.0404**	**0.9429**	**0.9539**	**0.9625**
14 (2048, 512)	0.9486	0.9638	0.9340	0.0659	0.9147	0.9386	0.9489
15 (1024, 1024, 512)	0.9481	0.9428	0.9520	0.0479	0.9351	0.9389	0.9474
16 (256, 128, 64)	0.9303	0.9276	0.9323	0.0676	0.9096	0.9185	0.9300
17 (300, 200, 100)	0.8943	0.8728	0.9066	0.0933	0.8728	0.8728	0.8897
18 (256, 128, 64, 32)	0.9032	0.8879	0.9178	0.0821	0.8879	0.8879	0.9028
19 (128, 64, 32, 64)	0.8869	0.8646	0.9006	0.0993	0.8646	0.8646	0.8826
20 (32, 64, 128, 64)	0.8355	0.8378	0.8270	0.1729	0.7804	0.8081	0.8324

activation function for the dense layer, which is just one neurone in size. The loss function remains the same for all twenty models, which is "binary_crossentropy". "*Adam*" optimizer and accuracy metrics are used for all these models. Across the model compilation, a check pointer is used to indicate when the model should be cut short and ensure a time-efficient model has been established. Early stopping is done by monitoring the validation loss value, and for that reason, the *min_delta* parameter has been set at 0.001 and the tolerance quantity to 10. The best model is also saved in HDF5 format for further validation. The batch size and number of epochs remain constant throughout all these 20 models, which are 300 and 100, respectively.

3.3.3.2 Evaluation of Different LSTM models

I have actually designed and run more than 30 different LSTM models with different parameters. But only the top twenty models are reported in this study due to poor performance **Table 3.2** displays those twenty models with the different combinations of nodes along with their performances. In addition, **Fig. 3.5** portrays Model 13 which has stood top inside **Table 3.2**. All the layers (input, output, flatten, dense) along with their shape have been visualized in **Fig. 3.5**.

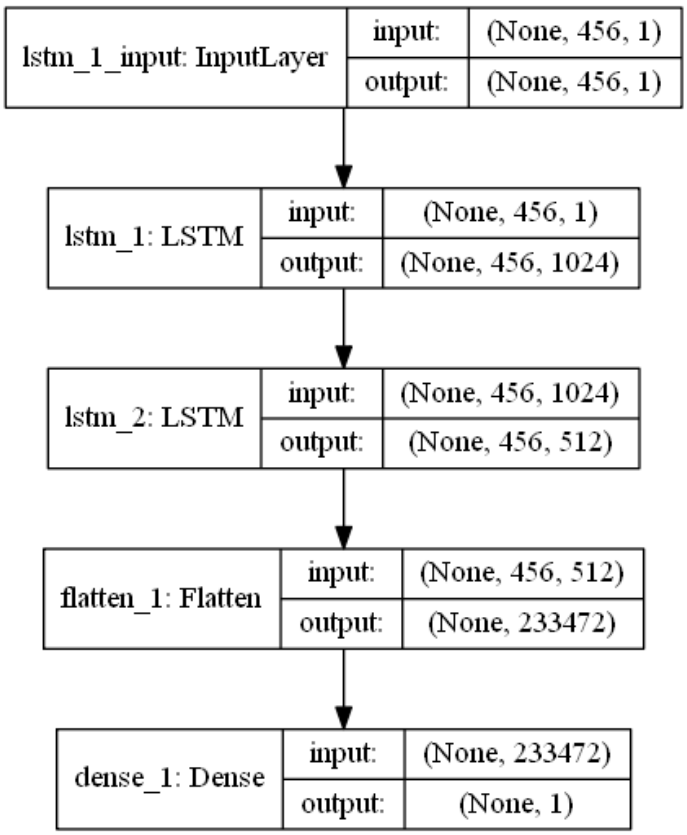

Fig. 3.5: Proposed LSTM architecture's input-output layers along with their shapes.

3.3.4 Performance Evaluation

In this study, various performance measurements are used to justify the performance of the models I have created. I have measured the performance of each of the models by (6), (7), (8), (9), (10), (11), (12), (twelve), the receiver operating characteristic (ROC) graph, and the area under the ROC curve (AUC). The plots of sensitivity (true positive rate) vs. false positive rates are illustrated on the ROC curve. [84].

$$False\ Positive\ Rate = \frac{FP}{FP+TN} \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$F1\ Score = 2\ \times \frac{Precision \times Sensitivity}{Precision+Sensiivity} \tag{11}$$

$$AUC = 1 - ROC \tag{12}$$

Hither,

TP: True Positive (number of MCI individuals recognized accurately as MCI)

TN: True Negative (number of HV recognized as HV)

FP: False Positive (number of HV recognized inaccurately as MCI)

FN: False Negative (number of MCI individuals recognized inaccurately as HV)

3.4 Experiments and Results

The EEG was recorded with the use of 19 separate electrodes that were positioned according to the 10-20 international standard [89], and 27 people participated. All the EEG recordings are transformed and stored in the European Data Format (EDF). This study has introduced the Butterworth filter for signal de-noising and converted the data into MATLAB format. After segmenting all of the artifact-free data, the average filter is applied to each segment for down-sampling. Each of the filtered temporal segments has been stored as a comma-separated value (CSV) file using MATLAB. Finally, down-sampled temporal partitions are used to categorize MCI patients using a deep LSTM model. All of the tests are executed on a Windows computer with an Intel Core i5-8400T CPU running at 1.70 GHz and 8GB of RAM.

The whole experiment set up is evaluated multiple times to check for consistency. After getting the EEG recordings of 27 subjects, I pre-processed the EEG data to ensure the data quality and features remained good enough to produce an accurate classification result. Finally, I have used that filtered data to feed into the LSTM models and checked the performance of the classifiers.

I have designed 20 different LSTM models and checked their performance. **Table 3.2** summarises the effectiveness of all 20 prototypes. It is seen from **Table 3.2** that Model 13 achieved the highest accuracy of 96.41%, the highest specificity of 95.95%, the highest precision of 94.29%, the highest F1 score of 95.39%, the highest AUC value of 96.25%, and the lowest false positive rate (4.04%). In model 13, each epoch took only 1242 seconds on average for compilation and only 280 seconds to evaluate. Model 13 LSTM network composed of an input layer, two hidden LSTM layers having "*tanh*" activation function and 1024 and 512 nodes, respectively, one flattening layer, and a dense layer having "*sigmoid*" activation function. Model 2 yields the highest sensitivity of 96.84% and has an input layer, one hidden LSTM layer with 256 neurones and "*tanh*" activation function, one flattening layer, and a dense layer with "sigmoid" activation function. Among all the twenty models, I found model/experiment 20 (having five hidden LSTM layers with 32, 64, 128, 64, and 32 nodes, respectively) did not perform well compared to the other 19 models.

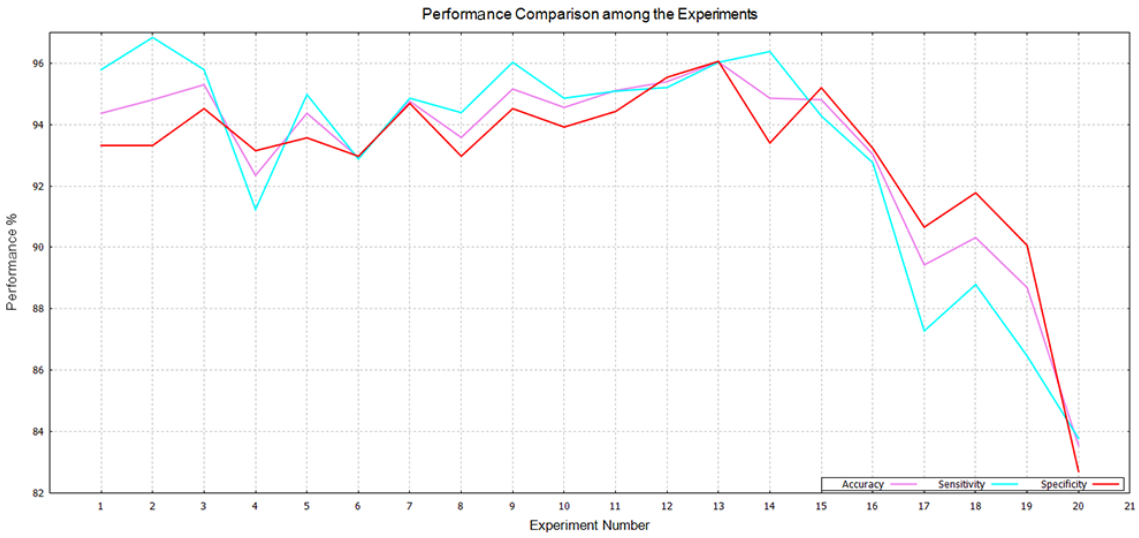

Fig. 3.6: Visual comparison of accuracy, sensitivity and specificity of 20 different LSTM models.

An important pattern is discovered in **Fig. 3.6**. The performance matrices (accuracy, sensitivity, and specificity) of the LSTM models remain above 90% when I have 2–3 hidden LSTM layers, and performance starts decreasing when I increase the number of hidden layers and neurons. Experiment/Model 13 has overall stability with good performance. And that is why I have considered and proposed this model for MCI classification.

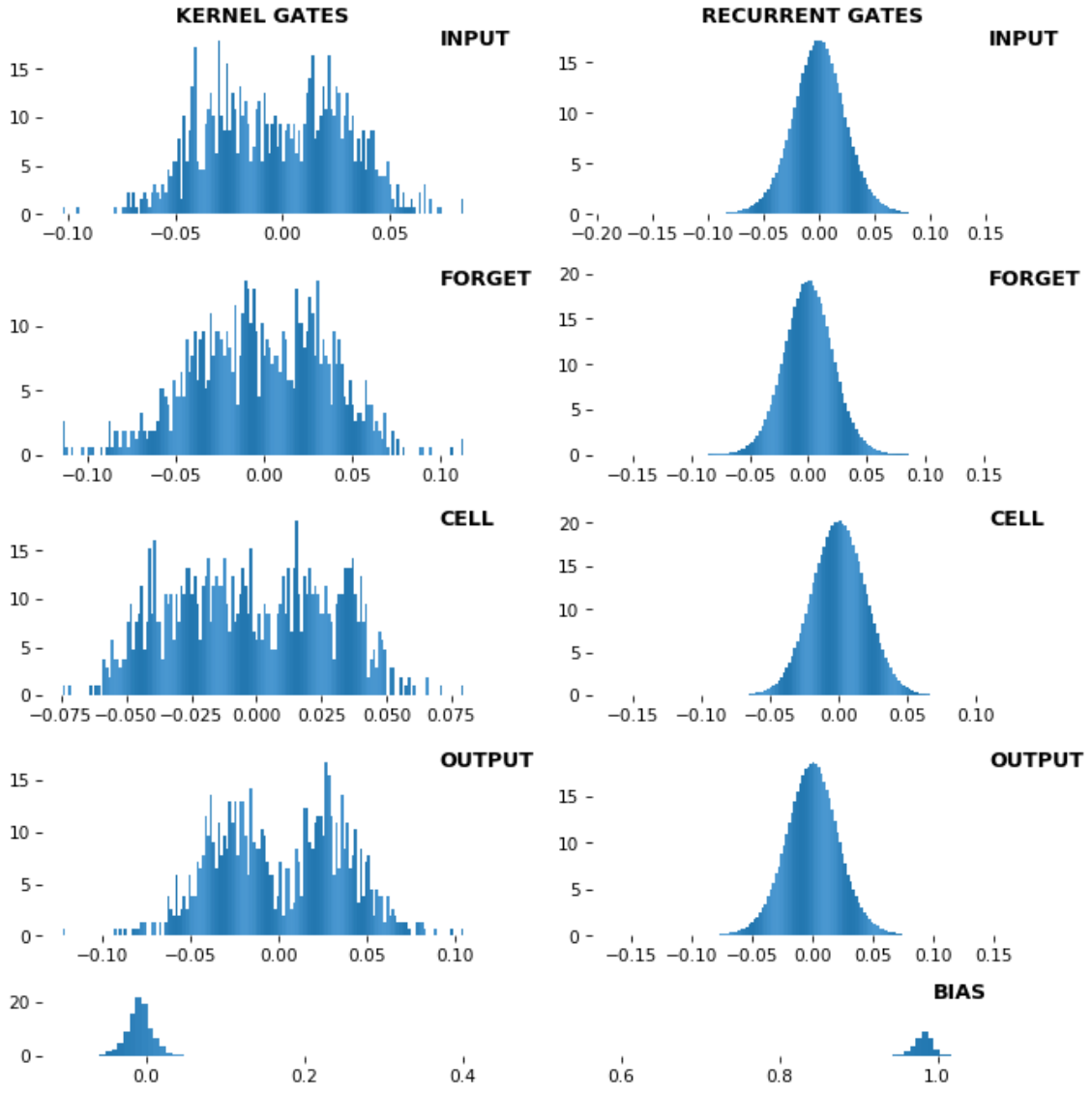

Fig. 3.7: Histogram of weight distributions as per kernel and gate.

Now talking about the architecture of the proposed LSTM model, I had to try so many different parameters and values to find out the optimal architecture for this MCI detection problem. To be able to further assess the learning speed of the propounded framework, I visualized the weight distributions per gate, kernel, and direction. **Fig. 3.7** portrays the weight distribution per gate, kernel, and direction. The load variations per kernel and

inside each kernel per gate are shown in a histogram subplot grid. From this histogram, I can see how well the weight is distributed, like a bell curve. The weight distribution of the kernel and recurrent gates remain within the (-0.1, 0.1) range. A very low bias range for both kernel and recurrent gates indicates that learning occurred quickly. **Fig. 3.8** illustrates a heatmap of the identical loads, complete with gate displacements indicated by vertical lines and bias weights. From the heatmap of the weight distribution, it is also clear that the LSTM architecture is very efficient in terms of low bias and learning rate, as color variation in the heatmap is almost constant, remaining between light blue (-0.1) and light red (0.1).

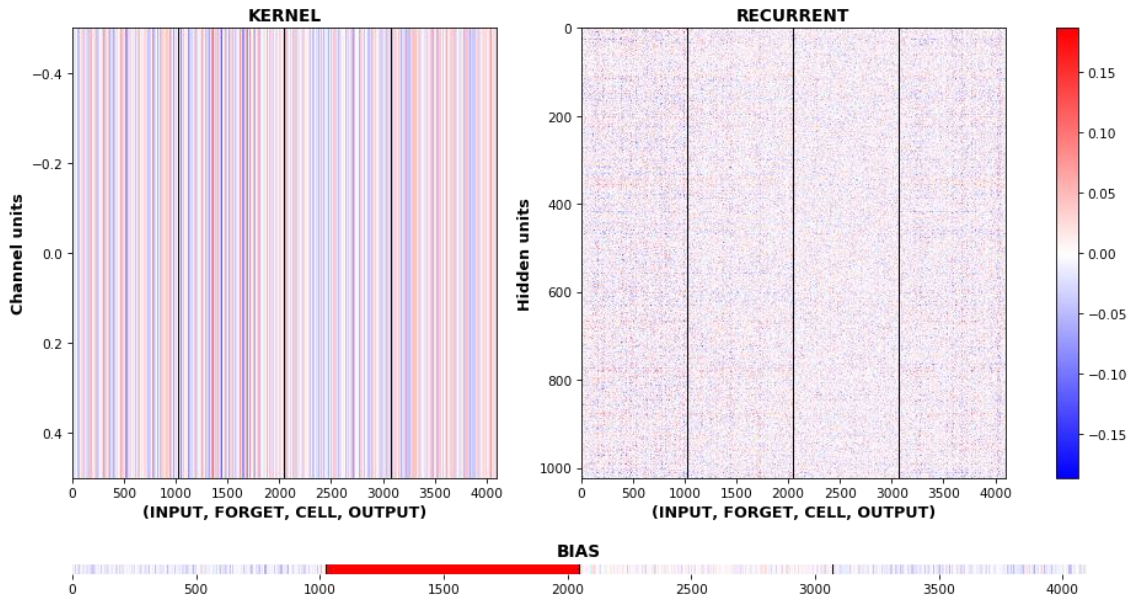

Fig. 3.8: Heatmap of the same weights distribution as per kernel and gate.

To validate the proposed LSTM architecture, I have carried out 5-fold cross-validation. Each of the five rounds of the 5-fold cross validation yielded comprehensive data for the LSTM model (**Table 3.3**), allowing me to assess its precision, adequacy, sensitivity, specificity, and false-positive rate in more detail. According to **Table 3.3**, there are no significant differences between the five obtained outcome levels. Except for fold 5, which offers 97.28 percent accuracy, all of the 5-fold alternatives have classification accuracy over 96 percent. The false positive rate is well under 4% in most of the folds, which confirms the stability and reliability of the proposed LSTM framework.

Fig. 3.9 illustrates the confusion matrix of our proposed study. It took only 31 epochs to reach such accuracy. 1126 (96.41%) of the 1168 healthy testing examples are accurately identified as HVs, whereas 42 (3.59%) are incorrectly labelled as MCI.

Furthermore, 827 (96.55%) of 857 MCI testing samples are correctly classified as MCI, while 30 (3.50%) are misclassified as HVs. According to the stated performance, the erroneously labelling rate is quite low, and it begins to decline as the number of testing examples grows.

Table 3.3: 5-Fold Cross Validation Procedure of the Proposed LSTM-Based Classifier.

Folds	Accuracy (%)	Sensitivity (%)	Specificity (%)	False Positive Rate (%)	Precision (%)	F1 Score (%)	AUC (%)
1	96.42	94.47	97.67	2.33	96.30	95.37	96.07
2	96.17	97.30	95.38	4.61	93.65	95.44	96.34
3	96.11	96.80	95.64	4.36	93.81	95.28	96.22
4	96.05	96.63	94.83	5.17	92.66	94.60	95.73
5	97.28	97.54	96.24	3.76	95.05	96.28	96.89
Average	96.41	96.55	95.95	4.04	94.29	95.39	96.25

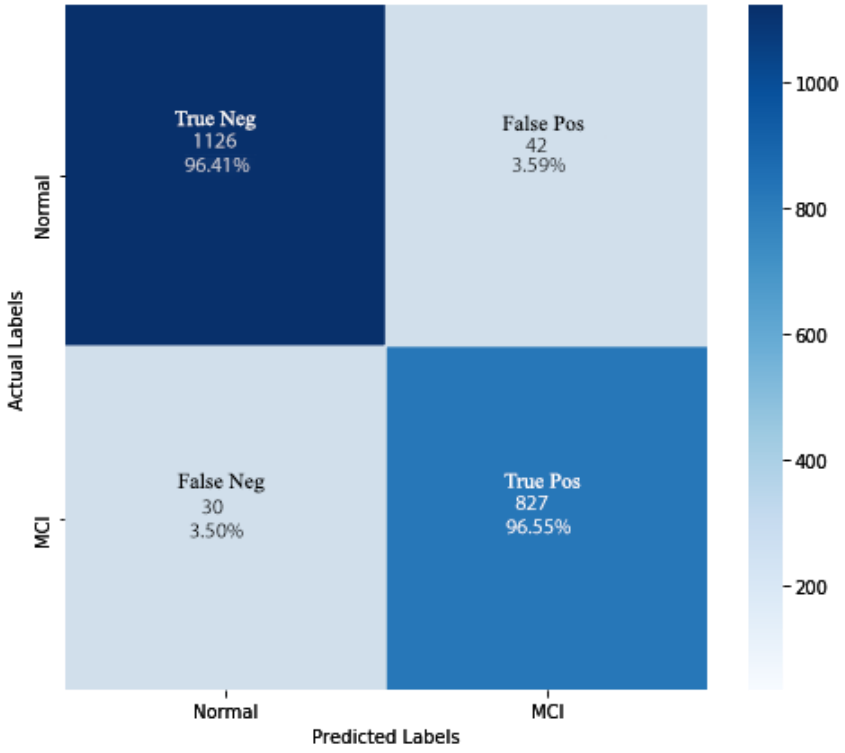

Fig. 3.9: The recommended LSTM model's confusion matrix.

The ROC curve represents how well the classification operates. It illustrates how sensitivity and specificity are related. The formula for the ROC curve is as follows: 1—specificity = FP / (FP+TN) and sensitivity = TP / (TP+FN) on the y-axis. In **Fig. 3.10**, the ROC curve is shown in blue. The value of the area that lies under the ROC curve is referred to as the AUC value. The AUC value ranges from 0 to 1. The effectiveness of

the suggested scheme is better when the AUC value is closer to 1 and the standard deviation value is lower than 0 [84]. The AUC is shown in **Fig. 3.10** as the region beneath the blue ROC curve. And model 13 achieved a 96.05% AUC value.

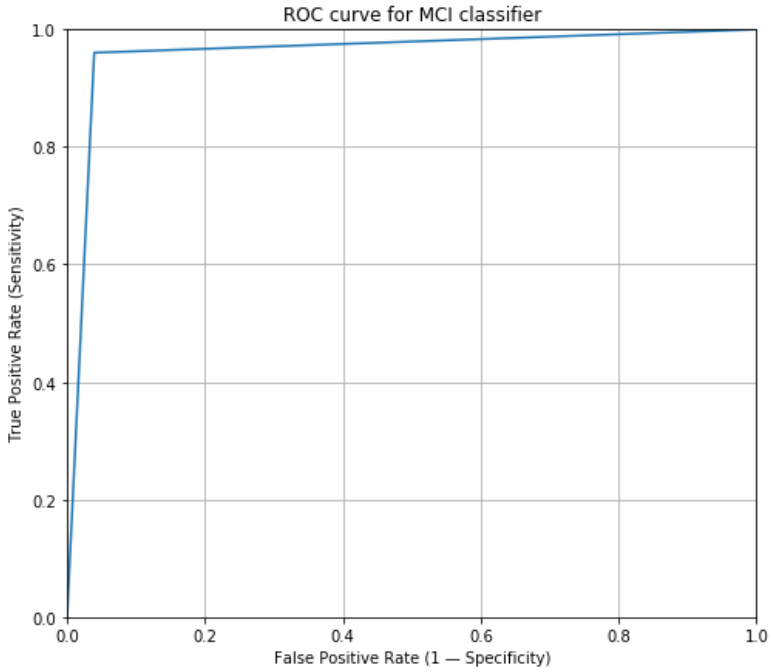

Fig. 3.10: The presented LSTM model's ROC curve.

3.5 Discussion

This thorough study has been done to identify MCI at an early stage. Many studies have been conducted with the same public MCI dataset that I used in this study and also with other private MCI datasets. Next, I discuss some of the significant recent works.

3.5.1 Previous Attempts

3.5.1.1 With Same Dataset

I have reported eight recent studies conducted with this same MCI dataset along with our previous works in **Table 3.4**. A few of these studies have already been described in Section 3.2. Yin *et al.* [100] ameliorated the categorization accuracy of 96.94% compared with the first work done using this dataset in [50] by using the 3D evolution method for feature picking and SVM as a classifier. But this study had a reduced number of subjects (11 MCI and 11 HVs), and it was a computationally expensive model because of its' preprocessing steps. Hadiyoso *et al.* [130], [126] put forth multiple efforts with this same dataset to detect MCI subjects and denoise the raw EEG data. They used the Hjorth

descriptor and K-NN for classification and achieved 80% accuracy. While denoising the raw EEG data, the study finished with a mean Root Mean Squared Error (RMSE) value of 0.0295. Our recent work [84] outperformed all the previous works and achieved 98.78% accuracy using the Extreme Learning Machine (ELM) as a classifier. Stationary wavelet transform (SWT) for noise removal, Piece-wise Aggregate Approximation (PAA) for compression, Permutation Entropy (PE), and an auto-regressive (AR) model are all part of the pre-processing phases. This study achieved high accuracy but was computationally expensive.

It can be deduced from these previous literatures using this same EEG dataset that most of the studies struggle to have satisfactory results. Our previous work outperformed all the studies, but in that study, I sacrificed time. In this proposed work, I came up with a balanced framework that not only performs better but is also time-efficient. More details about the literature using the same dataset are described in **Table 3.4**.

Table 3.4: List of previous research works for the same EEG dataset of MCI patients (that is used in this study).

Studies	**Subjects**	**Methods**	**Performance**
M. Kashefpoor, H. Rabbani, and M. Barekatain [50]	27 subjects (11 MCI and 16 HVs)	- A correlation based pursuit - NF inference system with KNN	Accuracy: 88.89% Sensitivity: 100% Specificity: 83.33%
S. Siuly *et al.* [84]	27 subjects (16 cognitively HVs and 11 MCI patients)	- Noise removing using SWT - Data compression using PAA - Permutation entropy (PE) and auto-regressive (AR) model - ELM	Accuracy: 98.78% Sensitivity: 98.32% Specificity: 99.66%
J. Yin *et al.* [100]	22 subjects (11 MCI and 11 HVs)	- Noise removing using SWT - Spectral-temporal feature extraction - Feature selection via 3-D evaluation - SVM classifier	Accuracy: 96.94% Sensitivity: 96.89% Specificity: 96.99%

S. Hadiyoso *et al.* [118]	27 subjects (16 HVs and 11 MCI)	- Spectral Analysis - K-Nearest Neighbor (K-NN) as the classifier	Accuracy: 81.5% Sensitivity: 81.82% Specificity: 81.25%
S. Hadiyoso, and T. Latifah [130]	5 MCI participants and 5 HVs	- Normalized to get a signal in the range -1 to +1. BPF filters (0.5 - 40 Hz) are applied to reduce noise artifact, DC and AC frequencies - Hjorth Descriptor - K-Nearest Neighbor (K-NN) as the classifier	Accuracy:80%
F. Jamalooa, M. Mikaeilia, and M. Noroozian [116]	EEG data of fourteen subjects (7 MCI and 7 HVs)	- Blinking artifacts were repaired using independent component analysis - Multi-stream continuous Hidden Markov Model (MS-CHMM) - Multi-metric functional connectivity analysis based on MS-CHMM - Phase Locking Value(PLV)	Accuracy: 95.9 ± 0.4 alpha % 97.2 ± 0.5 gamma Sensitivity:97.2% Specificity: 97.2%
S. Hadiyoso, and I. Wijayanto [126]	16 HVs and 11 MCI patients	- Empirical Mode Decomposition - Finite Impulse Response (FIR)	Mean RMSE 0.0295 Standard deviations RMSE 0.0665
M. Kashefpoora, H. Rabbania, and M. Barekatain [117]	32 HVs and 29 MCI subjects	- Dictionary learning(DiL) for Pre-processing: K-SVD stands for K-means and singular value decomposition (SVD) - Label Consistent KSVD (LC-KSVD) is one of the most recent DiL-based classification - Correlation-based LC-KSVD (CLC-KSVD) another classifier	Accuracy: 88.9% Sensitivity: 83.3% Specificity: 100%

3.5.1.2 With Different Dataset

It has been an ongoing effort to detect MCI, and there has been extensive work going on. In this subsection, I have reported fifteen recent efforts conducted utilizing the EEG data for MCI detection in **Table 3.5**. Using 408 subjects and the penalized logistic regression algorithm as a classifier, Farina *et al.* [125] developed a multiclass classifier and achieved 65% sensitivity and a 0.77 AUC value. No other performance matrices were reported, and the performance was not good enough. Another multiclass model was proposed in [131] using 189 subjects (63 AD participants, 63 MCI patients, and 63 HVs), where Multi-Layer Perceptron (MLP) was used as the classifier. The model earned 96.24% accuracy, 95.31 ± 0.8% precision, and a 95.58 ± 0.6% F1 score. This MLP-based study had a shallow architecture, and the MCI patients data used here had similar patterns closer to the AD category. Khatun *et al.* [52], [132] proposed two event-related potential (ERP) studies with SVM classifiers achieving 87.9% accuracy in both studies having 23 subjects (15 HV and 8 MCIs) and 17 subjects (10 HV and 7 MCIs), respectively. Sharma *et al.* [75] performed the continuous performance test (CPT) and finger tapping test (FTT) while recording the EEG signal of 44 subjects (15 dementias, 16 MCIs, and 13 HVs) and processed the data with power spectral density, skewness, kurtosis, spectral skewness, spectral kurtosis, spectral crest factor, spectral entropy (SE), and fractal dimension (FD) methods. Finally, they implemented the multiclass SVM classifier and achieved 89.8% accuracy, 84% sensitivity, and 94% specificity. Using the fuzzy neighborhood preserving analysis with QR decomposition (FNPAQR) dimensionality reduction technique, the patterns of 35 subjects (five vascular dementia (VaD) patients, fifteen stroke-related MCI patients, and fifteen control subjects) EEG recordings were preserved. SVM and KNN were both employed, but SVM outperformed them with an accuracy and sensitivity of 91.48%. S. J. Ruiz-Gómez *et al.* [133] reported a LDA, quadratic discriminant analysis (QDA), and multi-layer perceptron artificial neural network (MLP)-based study with 111 subjects (37 AD patients, 37 MCI patients, and 37 elderly HV patients) to develop a multiclass classifier. The accuracy, sensitivity, and specificity were 78.43%, 82.35%, and 70.59%, respectively. Another LDA-based study [134] was reported with 40 participants (27 HV and 13 MCIs) and scored 86.5% accuracy, 66.7% sensitivity, and 96% specificity.

From these related works with different MCI datasets, it is clear that our proposed study has outperformed all these other works. Most of these studies did not even end with

satisfactory results and struggled with the sample size as well. More details about the literature using the different datasets are described in **Table 3.5**.

Table 3.5: List of previous research works for the different EEG datasets of MCI patients.

Studies	**Subjects**	**Methods**	**Performance**
F. R. Farina *et al.* [125]	408 volunteers (102 AD subjects, 121 MCI patients, and 185 HV)	1. Band-pass filtered between 0.1 and 70 Hz, notch filtered at 50 Hz and average referenced across all scalp electrodes 2. Spectral analysis 3. EEG weighted phase lag index calculation 4. Penalized logistic regression algorithm with a well- established regularization technique known as the Elastic Net	Sensitivity: 65% AUC: 0.77
H. Chen, Y. Song, and X. Li [119]	107 subjects (50 chil- dren with ADHD, 57 HV)	1. EEG segmentation 2. Synchronization measurement 3. Brain network 4. Connectivity matrix organization 5. CNN for classification	Accuracy: 92.06 ± 1.50% AUC: 0.977 ± 0.0064
C. Ieracitano *et al.* [131]	189 subjects (63 suf- fering from AD, 63 MCI patients, and 63 HV)	1. Artifact removing and segmentation 2. Continuous Wavelet Transform (CWT) 3. Bispectrum analysis (HOS analysis) 4. Multimodal (CWT+BiS) features) is used as input, the MLP classifier	Accuracy: 96.24% Precision: 95.31 ± 0.8% F1 Score: 95.58 ± 0.6%
S. Khatun, B. I. Morshed, and G. M. Bidelman [52]	23 subjects (fifteen were HVs and eight participants were found to have MCI)	1. Event-Related Potential Processing 2. Candidate feature vector (CFV) 3. SVM method with radial basis kernel (sigma = 10/cost = 102)	Accuracy: 87.9% Sensitivity: 84.8% Specificity: 95%
S. Khatun, B. I. Morshed, and G. M. Bidelman [132]	A total of seven- teen subjects (ten persons were con- trol and seven per- sons had MCI)	1. Event Related Potential (ERP) 2. Neural Feature Extraction 3. Behavioral Feature Extraction 4. SVM was applied as a classifier	Accuracy: 87.9% Sensitivity: 85% Specificity: 90% F1 Score: 94%
N. Sharma *et al.* [75]	A total of 44 (15 de- mentia, 16 MCI, and 13 HVs)	1. Power spectral density, skewness, kurtosis, spectral skew- ness, spectral kurtosis, spectral crest factor, spectral entropy (SE), fractal	Accuracy: 89.8% Sensitivity: 84% Specificity: 94%

		dimension (FD) 2. Finger tapping test (FTT) and continuous performance test (CPT) 3. SVM was applied as a classifier	
A. Mazaheri *et al.* [120]	36 patients (with 25 amnestic MCI and 11 normal elderly con-trols) Within three years, 15 of the 25 individ- uals with MCI had progressed to AD	1. Oscillatory analyses: Time-frequency representations (TFRs) 2. Cross-frequency coupling between theta and alpha/beta: one-sample t-test of the Fisher r-to-z transformed correlations 3. ANOVA test 4. Theta power increase related to lexical processing was attenuated in MCI converters 5. Facilitatory effects of lexical processing due to repeated word presentation 6. Alpha suppression attenuates with each word repetition, but only for HVs and not for either MCI patient group 7. LDA and SVM with Gaussian kernels, i.e. radial-basis functions (SVMrbf)	Sensitivity: 80% Specificity: 95%
S. Poil *et al.* [121]	86 patients (25 patients had converted to AD, 39 subjects remained MCI, 9 subjects were diagnosed with subjective complaints and 13 patients with other disorders)	1. JADE ICA algorithm 2. Hjorth's activity, mobility and complexity parameters 3. Wackermann's global field strength, global frequency, and spatial complexity 4. Time domain Parameters 5. Barlow's amplitude, frequency and spectral purity 6. Absolute, relative power, and power ratios, furthermore, the central frequency, power in central frequency, band width and spectral edge 7. Student's t-test 8. Logistic regression	Sensitivity: 88% Specificity: 82%

J. P. Amezquita-Sanchez *et al.* [13]	Seventy-four patients (37 with AD and 37 with amnestic MCI)	1. Empirical wavelet transform (MUSIC-EWT) 2. Fractality dimension (FD) from the chaos theory 3. Enhanced probabilistic neural network (EPNN)	Accuracy: 90.3% Sensitivity: 92.1% Specificity: 87.9%
N. K. Al-Qazzaz *et al.* [135]	35 participants (5 VaD subjects, 15 stroke-related MCI participants, and 15 HV)	1. ICA-WT denoising 2. Fractal dimension 3. Two-way ANOVA was conducted 4. SVM and K-NN	Accuracy: 91.48% Sensitivity: 91.48%
S. J. Ruiz-Gómez *et al.* [133]	111 subjects: (37 AD patients, 37 MCI pa- tients, and 37 elderly HV)	1. Relative power in the conventional frequency bands, median frequency, individual alpha frequency, spectral entropy, Lem- pel–Ziv complexity, central tendency measure, sample entropy, fuzzy entropy, and auto-mutual information 2. Relevance and redundancy analyses were also conducted through the fast correlation-based filter (FCBF) to derive an optimal set of them 3. LDA, quadratic discriminant analysis (QDA) and multi-layer perceptron artificial neural network (MLP)	Accuracy: 78.43% Sensitivity: 82.35% Specificity: 70.59%
R. Pozˇar, B. Giordani, and V. Kavcic [134]	40 participants (27 of them being HVs and 13 with MCI)	1. A combination of functional connectivity, topological and cognition measurements is powerful for prediction of MCI 2. Linear discriminant analysis	Accuracy: 86.5% Sensitivity: 66.7% Specificity: 96%
D. Yang *et al.* [122]	24 participants (15 MCI patients and 9 HV)	1. Three mental tasks: N-back, Stroop, and verbal fluency (VF) tasks 2. Examining the oxygenated hemoglobin changes (ΔHbO) in the region of interest 3. During the tasks and seven temporal feature maps (i.e., two types of mean, three types of slope, kurtosis, and skewness) in the prefrontal cortex were	Accuracy: 90.37% Precision: 82.19% F1 Score: 84.04%

		investigated 4. A four-layer convolutional neural network (CNN) was applied to identify the subjects into either MCI or HV	
C. J. Huggins *et al.* [136]	52 AD participants, 37 MCI patients, and 52 HV	1. ICA for denoising 2. CWT to create time-frequency graphs with a wavelet coefficient scale range of 0–600 3. AlexNet, a deep CNN model of five hidden convolutional layers used for three class classification	Accuracy: 98.9 ± 0.4%
J. Poza *et al.* [137]	37 AD subjects, 19 MCI participants, and 29 HV	1. CWT to model non-stationary EEG data 2. Relative Power (RP) to analyses spectral distribution of the EEG oscillatory components 3. Spectral Flux (SF) to measure changes in time-frequency spectra and within- and between-electrode spatio-temporal fluctuations 4. Grand-mean RP and SF values over all channels were analysed using one-way MANOVAs	Accuracy: 77.3% (HV vs AD: combination of RP and within-electrode SF at the beta band) Accuracy: 79.2% (HV vs MCI: within-electrode SF at beta and gamma bands)
L. T Timothy *et al.* [138]	18 MCI patients and 18 HV	1. Recurrence quantification analysis (RQA) to extracts and Characterizes various aspects of a nonlinear system. The recurrence rate (RR) is an effective indication of the regularity or predictability of a particular time series provided by RQA. 2. Cross recurrence quantification analysis (CRQA) to deter- mine the similarity of states between dynamical systems 3. ANOVA was applied to compare RQA RR and CRQA RR for different pairs of regions 4. The experiments were set out on EEG from two states (i) resting eyes closed (EC)	Efficiency: 80.6% (EC: RQA RR) Efficiency: 72.2% (EC: CRQA RR) Efficiency: 72.2% (EC: combination of RQA and CRQA) Efficiency: 72.2% (STM: RQA RR) Efficiency: 86.1% (STM: CRQA RR) Efficiency: 91.7% (STM: combination of RQA and

		and (ii) short term memory task (STM) 5. LDA as classifier	CRQA)

3.5.2 Summary of the Proposed Work

Our main goal is to differentiate MCI patients from HVs with high accuracy and quickly. All the experiments have been conducted in a very ordinary computational setting with no GPU. Our model can run on any normal or regular PC. No external GPU or high RAM is required. And with such a computational setup, the best model was still proposed, compiled, and evaluated very quickly. It is not like I want to save money by compromising an algorithm's performance. I wanted to assure both time efficiency and accuracy. On average, our proposed model took 1242 seconds to compile for each epoch and only 280 seconds to evaluate the entire model. And still, it achieved 96.41% accuracy, 96.55% sensitivity, 95.95% specificity, 4.04% false positive rate, 94.29% precision, 95.39% F1 score, and 96.25% AUC value in 5-fold cross validation. To justify our proposed framework, I built 20 different LSTM models to identify the best parameters, and the performance of those 20 models is also reported in this study. The proposed LSTM architecture is well visualized in **Fig. 3.7** and **Fig. 3.8**. To prove the stability and reliability, I have carried out 5-fold cross validation, as reported in **Table 3.3**. There are no huge variations in the cross-validation results of the performance matrices.

The presented LSTM model has one input layer, two concealed LSTM layers with a total of 1024 and 512 neurons, one flattening layer, and a dense layer as part of its overall design. The tanh activation function has been used inside the hidden layers and the "*sigmoid*" activation function inside the dense layer.

I have reported previous studies done with the same publicly available MCI data and also with other different MCI datasets. Our proposed algorithm performed better compared to all the previous work done with different MCI datasets. This proposed LSTM-based MCI detection classifier also beats all the previous literature done with the same dataset except our ML-based previous works [84, 100], where the focus was only on accuracy. They required an extra step, which is feature extraction, and those models are very computationally costly. [100], [130], and [116] used fewer subjects, and [117] worked with more subjects, including those 27 subjects for their experiments. This MCI

dataset has not yet been studied using the LSTM approach, to our knowledge. Our proposed LSTM model has shown promising results in an efficient time.

3.6 Essence

The presented MCI detection framework has the ability to identify MCI patients very quickly. In this study, I focused on two things together: time and accuracy. It is a big challenge to work with DL-based classifiers and expect them to perform well. I pre-processed the EEG dataset quite well, and that ensured a time-efficient MCI classifier. The model has been compared with previous literature, and it is still well ahead if I consider both time and accuracy.

There have been extensive amounts of experiments done to find out the pattern and an accurate model for identifying MCI subjects from HVs using the EEG data. Around 35 experiments and models have been built, and in this study, I have reported the top 20 of them. All the experiments have been conducted on an ordinary computer without a graphics card. Another fruitful discovery I have made from the experiments is that the performance matrices (accuracy, sensitivity, and specificity) of the LSTM models remain above 90% when I have 2-3 hidden LSTM layers, and performance starts decreasing when I increase the number of hidden layers and neurons. Model 13, with two concealed LSTM layers with 1024 and 512 nodes, fared well in comparison to the other 19 models, according to our findings. It took only 1242 seconds on average for each epoch while building up the model, and only 280 seconds to test the whole model. The performance matrices are satisfactory as well, with 96.41% accuracy, 96.55% sensitivity, 95.95% specificity, 4.04% false positive rate, 94.29% precision, 95.39% F1 score, and 96.25% AUC value in five-fold cross-inspection. There is no significant variation found in the five-fold cross-inspection. It would have been better if I could manage to have a big EEG dataset, and that would have been even better for our proposed LSTM model as they tend to perform better with huge data. Another lacking or succeeding work plan of ours is to run this model with other publicly available MCI datasets or work collaboratively with the same field researchers. I also plan to extend our work to other neuro-diseases like AD, Parkinson's disease, schizophrenia, etc. I plan to work on other neuro-diseases and finally build a web-based, automated neuro-disorder system.

Since LSTM is an expensive variant of RNN, I want to explore further RNN variants like gated recurrent unit (GRU). GRUs are lightweight and consume less power

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: An adaptive image smoothing technique based on localization

Surname: ALVI First name: ASHIK MOSTAFA

Institute: Institute for Sustainable Industries and Liveab Candidate's Con 60

Status:

Accepted and in press: ☐ Date:

Published: ☑ Date: 13/03

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be includ ned in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted] 24/01/2023

Signature **Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

Name(s) of Co-Author(s)	Contribution (%)	Nature of Contribution	Signature	Date
Ashik Mostafa Alvi	60	Methodology, Data Curation, Investigation, Software, Validation, Formal Analysis,	[redacted]	24/01/2023
Siuly Siuly	10	Validation, Writing–Review and Editing, Supervision		24/01/2023
Hua Wang	10	Supervision, Funding Acquisition, Project Administration		26/01/23
Lili Sun	10	Resources, Writing–Review and Editing, Funding Acquisition		26/01/23
Jinli Cao	10	Validation, Writing–Review and Editing, Funding Acquisition		26/01/23

Updated: September 2019

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: Developing a Deep Learning Based Approach for Anomalies Detection from EEG Data

Surname: ALVI First name: ASHIK MOSTAFA

Institute: Institute for Sustainable Industries and Liveab Candidate's Con 60

Status:

Accepted and in press: ☐ Date:

Published: ☑ Date: 01/01

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be includ ned in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted] 24/01/2023

Signature **Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

Name(s) of Co-Author(s)	**Contribution (%)**	**Nature of Contribution**	**Signature**	**Date**
Ashik Mostafa Alvi	60	Conceptualization, Methodology, Data Curation, Investigation, Software,	[redacted]	24/01/2023
Siuly Siuly	20	Formal Analysis, Validation, Writing–Review and Editing, Supervision.		24/01/2023
Hua Wang	20	Resources, Supervision, Writing–Review and Editing, Funding Acquisition, Project		26/01/23

Updated: September 2019

than LSTMs. In the next chapter, I have documented the next MCI detection framework using a GRU model and the same EEG data to see the improvement in performance and in computational expense.

CHAPTER 4 : DEVELOPING A DEEP LEARNING BASED APPROACH FOR ANOMALIES DETECTION FROM EEG DATA

4.1 Overview

Electroencephalography (EEG) plays a leading role in brain studies, mental and brain disease diagnosis, and treatment. Traditional machine learning (TML) approaches are employed in most of the recent efforts in identifying anomalies from EEG data. But their scalloped architecture is one of the reasons why they fail to detect correctly and efficiently. Furthermore, these systems need to be fed the discriminant features manually. To overcome these issues, this study aims to develop an EEG data analysis system involving a multi-layer gated recurrent unit (GRU) for mild cognitive impairment (MCI) anomaly detection. There are four steps to the suggested framework: (1) collecting raw EEG data; (2) data pre-processing (de-noising, segmenting, and down-sampling); (3) discovering hidden significant characteristics of EEG data and classification using a GRU-based scheme; and (4) evaluating the model's performance. Our proposed model was tested on a publicly available EEG dataset and achieved 96.91% accuracy, 97.95% sensitivity, 96.16% specificity, and 96.39% F1 score. This study will guide future bio-medical researchers and technology experts towards a deep learning (DL)-based automated anomaly detection system for EEG data. **Research problems 1, 2, 4, and 5** have their solutions in this chapter.

The contents of this chapter have been published in the *Proceedings of Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference (FLINS 2020)* [9] and also in the *Proceedings of Web Information Systems Engineering–WISE 2021: 22nd International Conference on Web Information Systems Engineering, WISE 2021* [104].

4.2 Introduction

Artificial intelligence (AI) has received a lot of attention from researchers in recent years. AI, machine learning (ML), and DL appear in a slew of technology-focused articles [140]. DL is allowing change and innovation in many aspects of our modern lives. It is at the

heart of the majority of AI advancements reported in recent tech news [143]. Different AI fields, for example, genomics [149], graph theory [148], computer vision [145], natural language processing [94], cloud computing [182], sentiment analysis [147], automation [141], big data [139], and so on, are filled with DL applications. It has placed its foot not only in the tech world but also in the agricultural [142], health care [81], and business [144] sectors. Many AI, ML, and DL algorithms have been developed over the years. Talking about ML algorithms, they perform well with supervised learning with decent data sizes. To expand the area of AI and overcome some of the limitations of ML algorithms, DL algorithms were introduced. As I report, the computational support has increased as time goes on, and that makes DL algorithms even more efficient.

Electroencephalography is a tool to record electrical activity in the brain. Cerebral electrical potentials are measured by electrodes placed on the scalp. EEG is a non-invasive, portable, non-stationary, easy to use and interpret instrument mostly suitable for brain-related abnormality detection [106], [188].

Most of the EEG studies have at least three basic steps: deducting noises from the collected raw EEG signal without losing much data, extracting important features out of the filtered data, and EEG classification [186] with the extracted features. Due to the complex nature of EEG data, TML algorithms find it hard to extract the features properly. EEG holds multiple channels' data, and that makes the behaviour complex. TML cannot extract deep characteristics out of the data, and separate feature extraction and selection methods become mandatory for this job. On the other hand, DL algorithms can handle feature extraction and selection by themselves. By nature, DL algorithms have multiple hidden deep layers, which can easily tackle complex data like EEG. In recent years, as DL has advanced, a growing number of DL-based EEG investigation algorithms have been developed [183]. The size of the recorded data is enormous, and discovering the pattern of the data requires enough computational assistance. TML algorithms, having a shallow architecture, fail to work with huge data. And I know how well the DL algorithms can perform if I have a lot of data.

While recording the EEG data, artifacts and different types of noise get mixed with the signal. Therefore, a noise-removing stage is always there to ensure the data only contains Bain's electrical potentials. The feature extraction step looks after the bold features, which are important for the classifiers' training. There are many time-,

frequency-, and time-frequency-based algorithms to help the researchers in feature extraction. Finally, separated features are set as an input to the planned classifier for EEG classification. Various TML algorithms have been employed for EEG classification, such as K-nearest neighbour (KNN), support vector machine (SVM), linear regression (LR), random forest (RF), linear discriminant analysis (LDA), and so on.

To handle the noise, stationary wavelet transformation (SWT) was performed in [84]. For compressing filtered EEG data, the authors used piecewise aggregate approximation (PAA). Permutation entropy (PE) and the auto-regressive (AR) model were selected for feature extraction, and the extreme learning machine (ELM) was introduced for EEG classification. Independent component analysis (ICA) was used for discarding eye blinking artifacts in [116]. The same study installed a multi-stream continuous hidden Markov model (MSCHMM) and multi-metric functional connectivity analysis based on MSCHMM for identifying the discriminant features.

Kashefpoor *et al.* [50] conducted an EEG study with a correlated-based pursuit for feature extraction and employed the Takagi-Sugeno neurofuzzy (NF) inference system with KNN This study achieved 100% sensitivity, but the accuracy and specificity fell short at 88.89% and 83.33%, respectively. Yin *et al.* [100], [84] removed the artifacts using SWT and introduced a spectral-temporal way of feature extraction. Extracted features were evaluated and selected based on a 3-D evaluation algorithm. Finally, a SVM classifier was displayed and achieved 96.94% accuracy. Another effort [118] using spectral analysis and KNN was reported and attained 81.5% accuracy. EEG classification was done using the phase locking value (PLV) and achieved 95.9% accuracy. Dictionary learning had been introduced in [117] for pre-processing the EEG data. K-means and singular value decomposition (K-SVD), label consistent KSVD (LC-KSVD), and correlation-based LC-KSVD (CLC-KSVD) were three classifiers tried in that study, and CLC-KSVD achieved 88.9% accuracy among those three.

There were a couple of DL-based studies that tried to work with EEG classification. Amezquita-Sanchez *et al.* [13] used the empirical wavelet transform, also known as MUSIC-EWT, for noise reduction and the fractality dimension (FD) from the chaos theory for data compression. Finally, processed EEG data were fed into an enhanced probabilistic neural network (EPNN) and achieved 90.3% accuracy. A four-layer convolutional neural network (CNN) was designed for EEG classification along

with a biological experimental set-up to measure the oxygenated haemoglobin changes [122]. The performance of this CNN model was 90.37% accurate at differentiating abnormalities in the EEG signals. Chen *et al.* [119] reported a graph theory-based EEG classification where they performed segmentation and created a brain network, a connectivity matrix. After completing the pre-processing steps, CNN was brought into play for classifying EEG subjects. 92.06 ± 1.5% of the time, the CNN classifier was correct. LeCun's general LeNet CNN architecture was selected with an EEG dataset that was filtered with Fast Fourier Transformation (FFT) to classify EEG healthy volunteers (HVs) in [146] and performed very poorly. The accuracy of this work was just 69.23%, but the specificity was high at 88.89%.

It can be concluded from the reviewed literature that the TML algorithms performed consistently and better compared to a few of the DL-based efforts. But TML algorithms are not a time-effective option when it comes to feature extraction. An extra step is always required when it comes to TML algorithm-based studies. And having the complex nature of EEG data, it requires more attention for identifying discriminant features. The shallow architecture of TML methods makes it more time consuming and difficult to handle large amounts of data such as EEG.

To overcome the reported research gaps, I have introduced a GRU-based EEG classification framework. GRU is a DL algorithm and a variant of the recurrent neural network (RNN). To our best knowledge, GRU has never employed any EEG study containing MCIs in it. I know RNN is best suited to sequential data like EEG. And GRU is a memory-efficient DL model, as it does not hold much memory like long short-term memory (LSTM), which is another variant of RNN. A deep GRU network is designed with hidden GRU layers that may selectively memorise important input for a certain amount of time decided by the reset gate and are commonly used in sequential prediction. Our proposed framework starts with reducing the noise using the Butterworth filter, which has the potential to achieve a more linear phase reaction and a comprehensive flat response. Then, I have segmented the filtered data and down-sampled it using a newly developed adaptive filter to reduce the computational overhead. The proposed adaptive filter works fine with any dataset when there are no or a minimum number of outliers. Finally, a two-layer GRU network has been designed to classify EEG data. The main efforts of this study can be summarised as follows:

- For the first time, I have reported GRU based DL study for EEG classification.
- I have developed an adaptive filtering as a down-sampling approach to aid the suggested model's computational overhead.
- Our proposed model is computationally inexpensive to design and achieve efficient and competitive retrieval resultswhen compared to the existing DL models.

The rest of the chapter is sorted as follows: Section 4.3 provides a description of the proposed model framework. A detailed result and analysis are given in Section 4.4. A comparative discussion is provided in Section 4.5. Finally, this study finishes by adhering to the essence in Section 4.6.

4.3 Proposed Framework

The DL-based anomaly detection framework from EEG data is designed with 4 sub-modules: (1) collecting raw data; (2) data pre-processing (de-noising, segmenting, and down-sampling); (3) discovering hidden significant characteristics of data and classification using a GRU-based scheme; and (4) evaluating the model's performance. **Fig. 4.1** illustrates the overall framework in a nutshell. I have collected a publicly available raw EEG dataset from 27 subjects. Data pre-processing has three inner steps: (1) de-noising the raw EEG data, (2) EEG data segmentation, and (3) down-sampling the segmented data using an adaptive filter [7], [9]. Processed data is fed into the GRU classifier for feature extraction and anomaly detection. There are eight standard

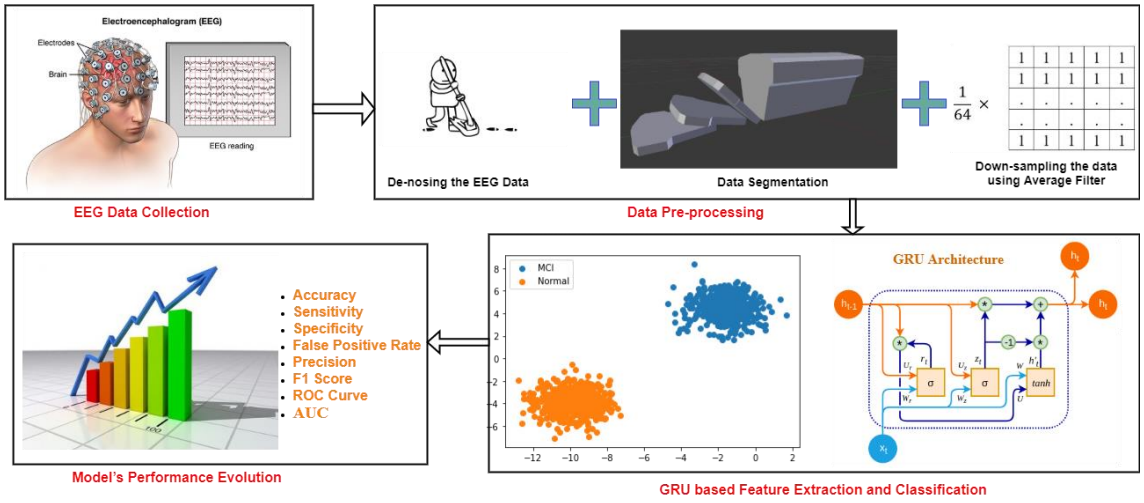

Fig. 4.1: Overview of the GRU based anomaly detection framework from EEG data.

evaluation matrices used to validate the performance of our proposed model. All the sub-modules are described in detail below:

4.3.1 Collecting Raw EEG Data

This proposed anomaly detection framework is tested with a publicly available EEG dataset of 27 subjects, where there are 16 HVs and 11 MCI subjects ageing between 60 and 77 years. These EEG data were collected in the Sina and Nour Hospitals, Isfahan, Iran [50], [123]. All the participants gave their consent, and the deputy of research and technology at the Isfahan University of Medical Sciences, Isfahan, Iran, ethically approved this EEG data collection. Each of the subjects underwent a neuropsychiatric interview for MCI illness according to Pererson's criteria. The participants were validated using the Mini-mental State Examination (MMSE) score. Subjects with an MMSE score of 21 to 26 were treated as MCI, while those with a score of more than 26 were considered HV. Head trauma, dementia, and a history of significant mental problems, serious medical conditions, or drug abuse were all considered exclusion criteria.

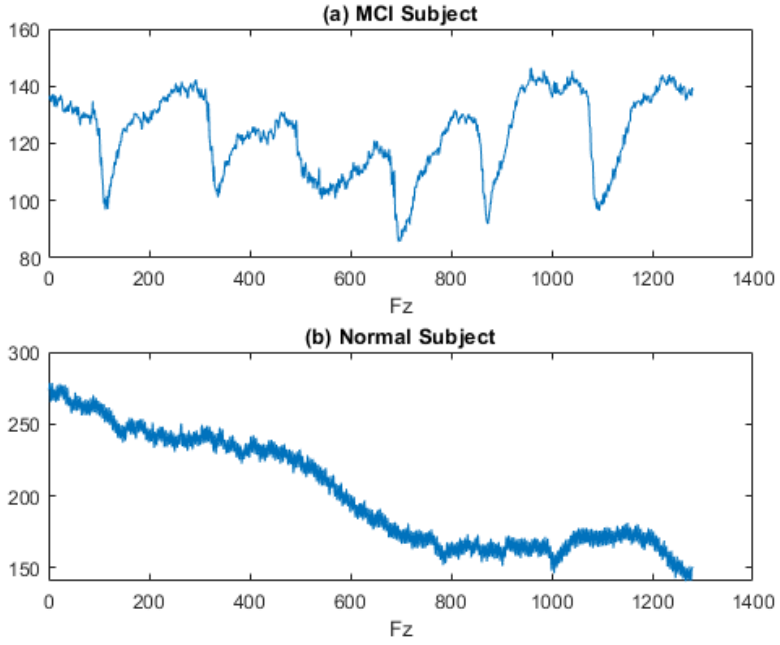

Fig. 4.2: Raw EEG recordings of (a) an MCI subject and (b) a HV for 5 seconds at the Fz electrode location.

The EEG recordings took place in the morning in a quiet room for 30 minutes, with all of the subjects in a relaxed state. The 19 electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2) were placed over the scalp using the International

10-20 system, and EEG data were digitised using 32-channel EEG equipment with a sampling frequency of 256 Hz. **Fig. 4.2** shows 5-second EEG recordings of an MCI and a HV at the Fz electrode location. The electrodes' skin impedance was less than 5 kΩ. To avoid sleepiness, the medical professional made sure that all of the participants were awake during the EEG recording.

4.3.2 Data Pre-processing (De-noising, Segmenting, and Down-sampling)

The raw EEG data often gets mixed with unwanted signals and other outliers. And for an individual subject, the recording lasted for 30 minutes. So, I can understand the size of the data. These huge EEG datasets need to be segmented for smooth processing. Our proposed pre-processing sub-module has three inner steps: (a) de-noising using the Butterworth filter, (b) segmentation, and (c) down-sampling using an adaptive filter to ensure smooth feature extraction and classification.

4.3.2.1 De-noising the EEG Data

Artifacts or so-called undesirable signals, frequently contaminate EEG recordings. Outlier values, electrode pops, breathing, power supply fluxes and interference (50 Hz), baseline drift, eye blinking, or muscle electrical activity, among other things, are some of the most prevalent causes of EEG recording contamination. As a result, if I want a decent classifier, I must first remove the noise from the recording.

Our proposed study engaged the Butterworth filter to de-noise the unwanted signals. The Butterworth filter used in this work excludes undesirable high-frequency signals such as those caused by body movement, electricity grid inference, eye blinks, and heartbeats. It aids in the production of more linear phase reactions while also optimizing the flat response.

The recordings were viewed, and noise was removed using the MATLAB EEGLAB [124] package. To remove artifacts, a low-pass, third-order Butterworth filter with a cut-off frequency of 50 Hz was employed. Each having a duration of 60 seconds × 30 minutes × 256 Hz. I digitized each subject's signal, which has 460800 rows and all 19 electrodes representing the columns. The recordings were saved as .mat files.

4.3.2.2 Segmentation

By character, EEG signals are non-stationary, non-periodic, and massive in size. We'll need a lot of computer power to analyze this big dataset of 27 people, and it will take a long time to build a model. To deal with this type of data, segmentation is a good option. It also helps to increase the sample size by not losing any data, as each segment still contains important features.

For our work, I intended to split each of the 30 minutes of recording and capture 6 seconds for each segment by keeping the sampling frequency the same, which is 256 Hz. Following the segmentation of each subject, I produced 300 additional segments from each subject's recording, all of which are clearly labelled with the appropriate subject's label (MCI/Normal). Finally, I had a total of 27 × 300 = 8,100 individuals, each with 1,536 rows and 19 columns. For our work, each of these temporal segments is treated as a digital image for further processing.

4.3.2.3 Down-sampling using adaptive filter

Image processing is still being challenged by noise. Noise causes the intensity manipulation of the image. So, removing or reducing the noise from the image is a must before working with it. It is an active area of research because none of the established or proposed noise-reducing methods can bring back the original image. And also, there are different types of noises. Different proposed algorithms work fine with different types of noise and also up to a certain level of noise. In this part of the work, an adaptive noise removal algorithm is proposed that works fine with impulse noises and does not blur the edges of the inputted image. While removing the noises, the algorithm uses an adaptive mask, which is n × n squared or crossed mu, where n is usually an odd number. Our proposed algorithm has achieved a peak signal-to-noise ratio of 15.38 dB, outperforming the existing filters.

Noise is a bug that can be added while capturing the image. Image noise is the variation of color and brightness information in an image [91], [168]. The latest cameras have the noise-removing algorithms installed within them. But still, the capturing sensors and lighting factors, i.e., sunlight, nighttime mood, etc., can help in acquiring the noises. The real intensity values of the image are being replaced by the noise. When a sensor captures an image, it filters the image with different smoothing algorithms.

$$f(x,y) = reflectance(x,y) * illumination(x,y) \quad (13)$$

An image is created in Eq. (13), where f(x, y) represents the intensity level at position (x, y), the reflectance data range is [0, 1], and the illumination's range is [0, ∞]. This is how an image is created. Digital image-capture devices have better support for noise reduction. Nowadays, those devices have a good image-capturing sensor. Image noise can be produced by the circuitry of a scanner, digital camera, or sensor [161].

Applications of digital imaging are increasing day by day. Starting from a magnetic computed tomography scan, aerospace, satellite television, resonance imaging, astronomy, a vision system, a vision-based robotic system, etc. The most sensitive and vast use of digital image processing is in the biomedical sector. Identifying a disease takes more time because of long testing examinations. At present, several diseases are identified; also, proper treatments are prescribed later than a sequence of long examinations [162]. Taking blood samples and using different scans and diagnoses to detect a disease is the regular or manual procedure. Biomedical imaging is playing a role here to reduce the time for detecting the diseases. Hair cracks, eye cataracts, and tissue damage detections are some of the biomedical applications of digital imaging. If these images are corrupted with noise, that can cause a lot of harm to a patient. So, removing the noise from the image is a must for all these important applications of digital imaging.

In the related works section, it is used to briefly discuss the related previous works of noise removal algorithms. Our proposed method has been elaborately described in the proposed method section. In the experiments and findings section, our proposed algorithm's performance is measured and comparisons are shown with respect to the existing previous works. This sub-study ends with an essence.

Analogue and digital are the two types of signals I mostly work with. Both of the signals are corrupted with noise while capturing. So, removing noise is a prerequisite for any type of signal processing. Noise removal is an active area of research in Signal Processing. There have been several works done regarding noise removal from images. And still, it is an ongoing research area as the noise and image acquisition devices vary.

Wang *et al.* [163] had used the non-local means (NLM) and the bilateral filters to reduce the noise in color images. They introduced a weight function as texture information with the NLM and bilateral filters. The mean squared deviation (MSD) was

used to distinguish the edge areas from the smooth areas. This algorithm showed better performance in reducing Gaussian multiplicative noise from color images.

Among all the well-established noise removal filters, the median filter performs better at restoring the fine details of the image. There have been lots of works done with the median filters [167, 168]. Finally, when the mask is fixed, the algorithm calculates the average value of the selected mask, visits each pixel, and compares it with the average value. If any of the pixels values were greater than the average value, then the algorithm replaced them with the median value of the selected mask. This algorithm showed better performance in terms of time complexity.

In a recent work [164], salt and pepper noise was targeted to reduce grayscale images. Recursive adaptive switching median filter (RASMF) was proposed to do the task, which has two parts: noise detection and noise cancellation. A pixel is corrupted if it has the maximum or minimum intensity level, and the noise detection function assigns candidate value 1 to a noisy pixel otherwise 0. The noise cancellation part replaced a noisy pixel with the median value of eight non-noisy pixels within the mask. The filtering window was switching until it had eight noise-free pixels inside it.

Depending on the probability density function of the impulse noise, noisy pixels had been replaced with the difference between the median value of the floating window, which has four members: (i-1, j), (i, j-1), (i, j+1), (i+1, j), and the level of rejection in [166]. A distortion function is used to identify noise, and 1 value is assigned if the pixel is noisy and 0 otherwise. From the level of rejection, which was calculated by the variance of the pixels, each noisy pixel's replacement value was calculated.

Very few works have been done with the mean filter to reduce the noise, as this filter is very sensitive to salt and pepper noise [8], [167]. In [165], an adaptive decision-based mean filter was proposed for only salt and pepper noise. It was a decision-based noise filtering model that replaced the noisy pixel with the mean of the four noise-free pixels in the north, south, east, and west directions.

Fuzzy logic has been used for image noise removal as well. The Fuzzy Inference System (FIS) was used to determine the noise level of the pixel of the current mask in [7]. There were 97 fuzzy if-then rules inside the FIS to decide the noise level. Maximum, minimum, median, and mean intensity values were passed as an input to the FIS.

Prabu, Balamurugan, and Vengatesan [162] had also used the FIS to filter out the noise from medical images. It was actually a histogram-based algorithm that had been equalized and input into the FIS. Based on the FIS decision, noisy pixels get replaced by the median value of the current window.

Proposed Method. In this chapter, I have proposed an adaptive noise removal filter that has two basic parts. a) Noise Detection Unit; b) Noise Removal Unit Based on the noise detection, different filters are applied to the centered pixel.

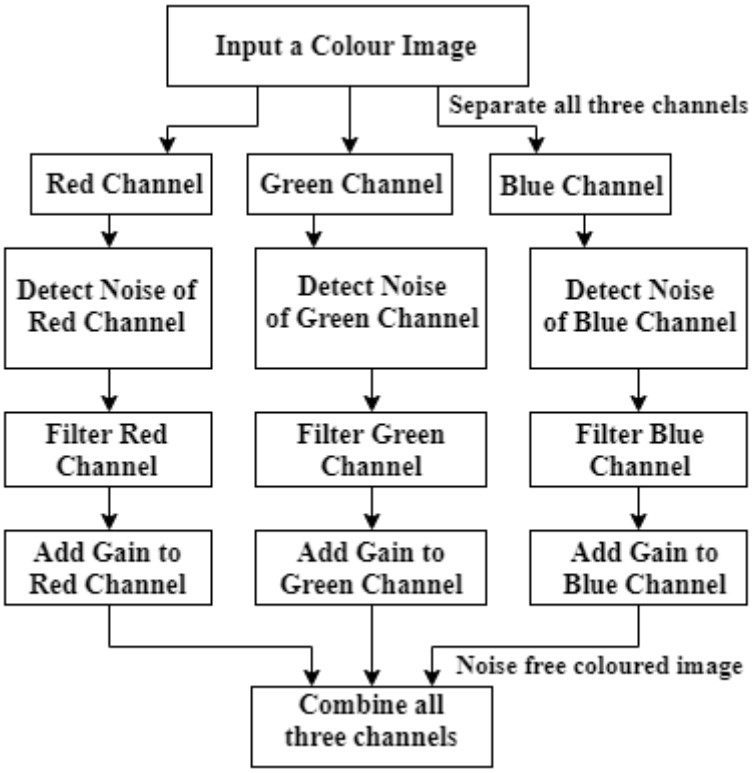

Fig. 4.3: Proposed algorithm's flowchart.

Dataset. All the experiments are done with a Kodak color image dataset.

Algorithm. The system takes as input a color image. After reading a color image, it separates the three channels (red, green, and blue) and then filters them separately. The filtered channels are finally combined to build the colored image. The algorithm plan is shown in **Fig. 4.3**. The algorithm is divided into two units: (a) noise detection and (b) noise removal. The proposed adaptive filtering requires the following steps:

Algorithm 1 The pseudo-code of the noise removal method

Step 1: input a noisy image (im_noisy)
Step 2: compute [H W] ← size (im_ noisy),
max ← 0, min ← 0 // two variables to store the maximum and minimum intensity value within the 3 × 3 mux window.
Step 3: for each i € H-1 and j € W-1, store the 9 pixel's intensity in P matrix and also compute the max and min intensity value.
Step 4: compute the difference between max and min, compute maximum occurred intensity from the P matrix.
Step 5: Judgement: if max_occurence_intensity >= mode threshold value, then apply mode filter;
if max_occurence_ intensity < mode threshold & & difference between max and min > average threshold && difference between max and min <= median threshold, then apply average filter;
if max_occurence_intensity < mode threshold && difference between max and min > median threshold, then apply median filter

Experiments and Findings. The above adaptive algorithm works fine with all types of additive noise, depending on the threshold value given. The algorithm beats all the existing smoothing algorithms when the image has multiplicative noise. It also gives satisfactory output for salt and pepper noise. The peak signal-to-noise ratio (PSNR) value is measured for comparing the performance of this algorithm.

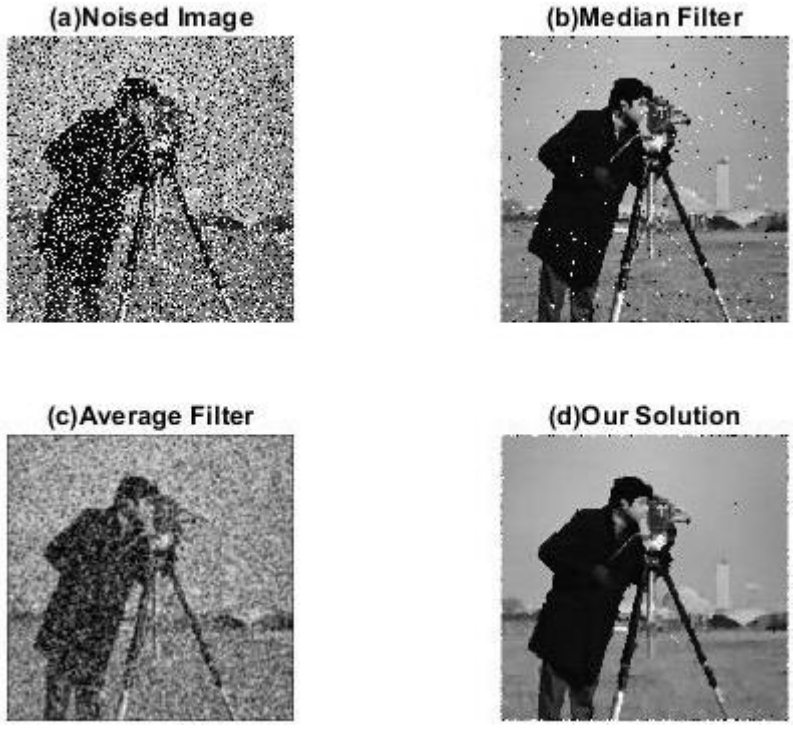

Fig. 4.4: Visual comparison on a grayscale image.

Fig. 4.4 illustrates a compassion among the mean, median, and our proposed filter applied to a grayscale image. Another color image comparison is shown in **Fig. 4.5**. **Table 4.1**

represents the mathematical comparison where our proposed algorithm beats the existing filters.

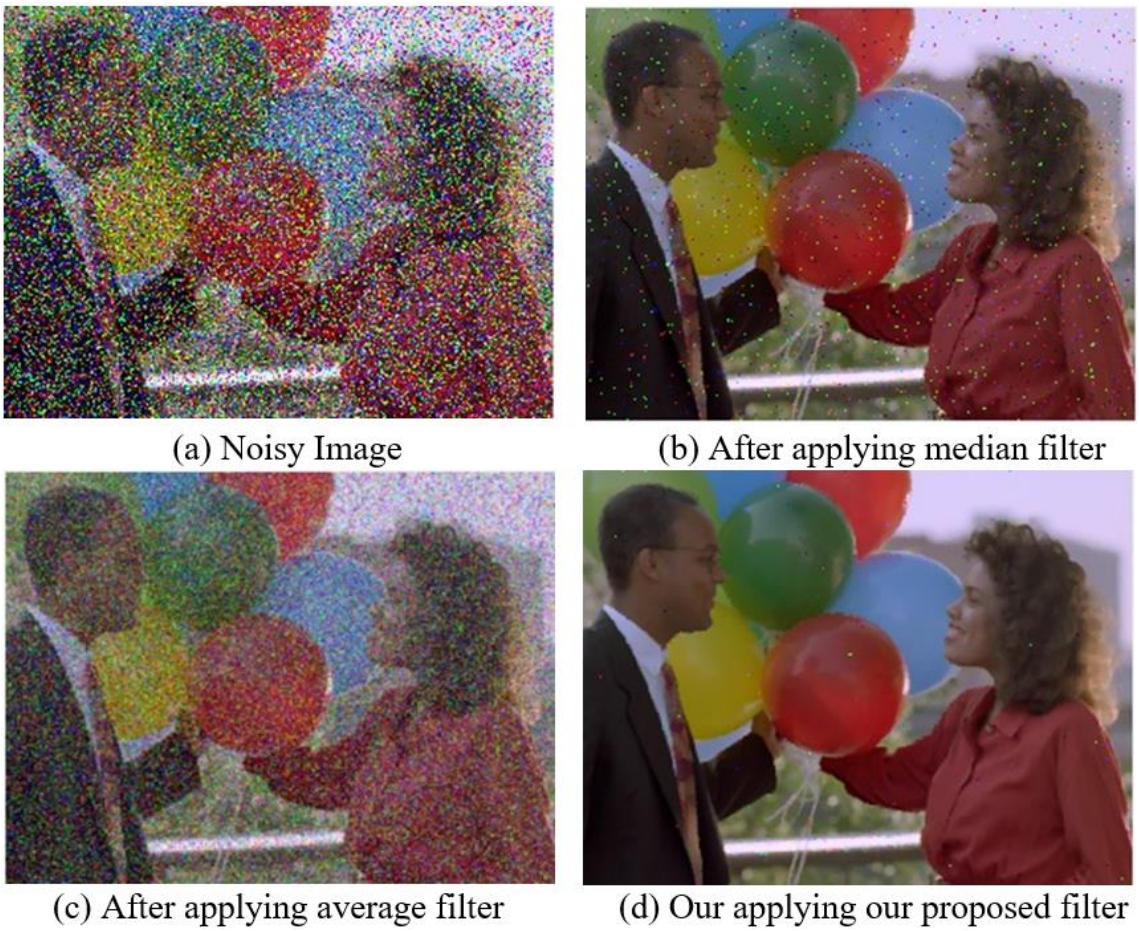

Fig. 4.5: Visual comparison on a colour image.

Table 4.1: PSNR and MSE comparison.

Filters	MSE (Mean Square Error)	PSNR in dB
Median Filter	1.929738601684570e+03	15.275818764893970
Average Filter	2.031163055419922e+03	14.789564798109934
Our Algorithm	1.884438812255859e+03	15.378983203988462

Summary. In this work, I have proposed an adaptive algorithm that is more powerful at reducing the noise, particularly the multiplicative noise. I have shown the comparison of two different images and their peak signal-to-noise ratio (PSNR) and mean square error (MSE) values, where it is proved to work better than existing algorithms. This adaptive algorithm works fine with all types of images and all types of noise, but only at certain thresholds that I have mentioned in the algorithm section.

Applying This Filter. To ensure an efficient DL model, I need to have a large dataset. But I had only 27 subjects. From these 27 EEG recordings, I created 8,100 subjects by segmenting the dataset. Now, I want to resolve the computational power issue because I want a cost-effective model that can be trained on a regular machine.

To reduce the computational overhead, I down-sampled the data using the proposed adaptive filter. The adaptive filter works fine when the data is outlier- and noise-free. So, this filter fits perfectly with our dataset and helped me reduce the data dimension. Initially, the sampling frequency was 256 Hz, and by using the adaptive filter, I down-sampled it to 4 Hz. The sliding window method was employed, with a window size of 64 × 1 for each iteration. **Fig. 3.3** is a showcase of the sliding window. While keeping the number of channels constant, I run the window column-wise. For each of the channels, the window ran for 24 times, giving me a new row number for each of the subjects. I saved the data in CSV format after performing all of the pre-processing processes in MATLAB.

4.3.3 Discover Hidden Significant Characteristics of Data and Classification using GRU based Scheme

The main goal of this study is to develop a simple, computationally cheap, DL-based model that performs efficiently. To investigate further, I introduced a GRU-based classifier in this study to identify the hidden features and differentiate HV EEG signals from abnormal ones. To our best knowledge, this dataset has never been subjected to GRU analysis. I find GRU suitable for this EEG study due to the nature of EEG data and the characteristics of GRU. As I know, EEG is a huge, non-stationary, chaotic dataset, and it suits well with DL-based methods. It is suitable for any recurrent network-based method. As LSTM, which is a variant of recurrent networks, has a tendency to hold the memory for a long time, it is not suitable if I want a cost-effective model. When it comes to GRU, in particular, it becomes a perfect match as it has a forget gate inside its architecture, which helps to set free the memory.

To build up our proposed GRU network, I used Jupyter notebook as an IDE and Python as the programming language, as it has good support for DL with libraries like Keras [128], SciPy, Pandas, NumPy, and Scikit Learn.

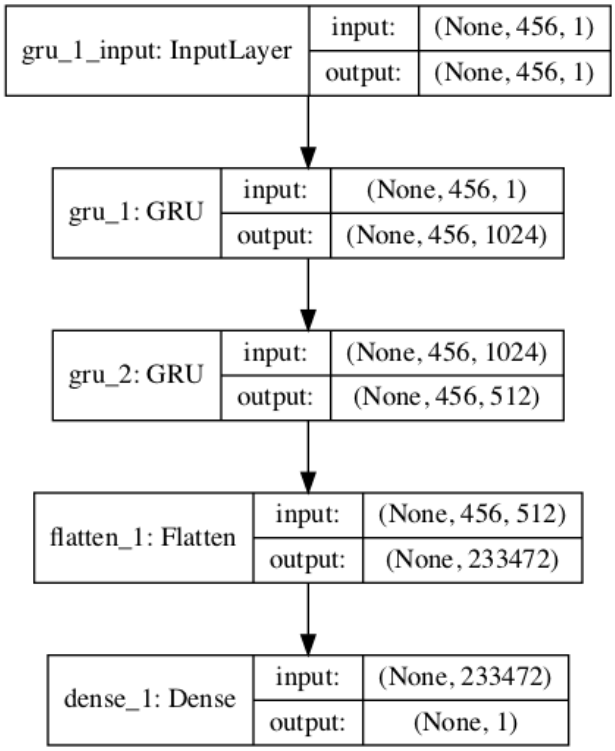

Fig. 4.6: Proposed GRU Architecture.

With the aid of the Pandas library, all pre-processed CSV files were read into the system. The NumPy library was then used to transform all of these data frames into two-dimensional arrays. At the same time, each subject's category label was stored in a separate NumPy array. Then, using the NumPy library's reshape function, I decreased the input data's dimension. With the Scikit-Learn package, the dataset was then randomly split into the training and validation sets. In our validation set, I included 20% of the data. Finally, the Keras library with a Tensor Flow backend [129] was used to construct the proposed GRU network.

Our proposed GRU network had two deep, hidden layers and a dense layer. The dense layer had a single neuron, and the first hidden layer of the GRU network had 1024 neurons, and the second one had 512 neurons in it. I had used "*tanh*" as the activation function in both of the deep hidden layers and ‘sigmoid’ in the dense layer. There was a flattening layer employed just before the dense layer. **Fig. 4.6** illustrates the architecture of the GRU network. "*Binary_crossentropy*" was used as the loss function as it is a binary classification problem that I was trying to solve. I had picked the "*Adam*" optimizer for our work.

To ensure our proposed GRU network did not overfit, I deposited early and kept a checkpoint throughout the training process. The check pointer was set on the validation loss and validation accuracy, with the *min_delta* set to 0.001 and patience to 10. For fitting the data, I set the batch size to 300 and the epoch to 100. After every check point, it sorted the best model in "*hdf5*" format, which was later used for further model evaluation.

4.3.4 Model's Performance Evaluation

Our proposed GRU-based model has been tested with eight standard evaluation matrices. Accuracy, sensitivity, specificity, false-positive rate, precision, the F1 score, the receiver operating characteristic (ROC) curve, and the area under the ROC curve (AUC) are the performance measure matrices used to justify our proposed study. The plots of sensitivity (true positive rate) vs. false positive rates are shown on the ROC curve in **Fig. 4.9**.

4.4 Result and Analysis

All the raw EEG data was stored in European Data Format (EDF). For the pre-processing steps, I have used MATLAB as a tool. Using the Butterworth filter, I have de-noised the raw EEG data and saved it as MATLAB-formatted data (.mat). After that, all of the artifact-free data was segmented, and each segment was down-sampled using the proposed adaptive filter. Comma-separated value (CSV) files were created for each of the filtered segments. Finally, these segments are input into the GRU network for feature extraction and classification. All the experiments are performed on a MacBook Pro with a 2.6GHz Intel Core i7 CPU, 16GB RAM, and a 4GB graphics card.

There were 100 epochs with a batch size of 300 while training the model. To ensure our model does not overfit, I have set an early stop on validation loss and accuracy. The *min_delta* value was set to 0.001 and patience to 10. Finally, our model achieved 96.91% accuracy within only 36 epochs. **Fig. 4.7** depicts both the training and validation accuracy and loss over epochs. In the beginning of the training process, the training loss was very high, but it dropped drastically just after the 1st epoch. The validation loss decreased from 42.82% to 12.59% over the 36 epochs. The validation accuracy improved from 66.05% to 96.91% over the training process. The overall performance of our proposed model is reported in **Table 4.2**. The model has achieved 96.91% accuracy, 97.95% sensitivity, 96.16% specificity, 3.84% false positive rate, 94.89% precision, 96.39% F1 score, and 97.05% AUC value.

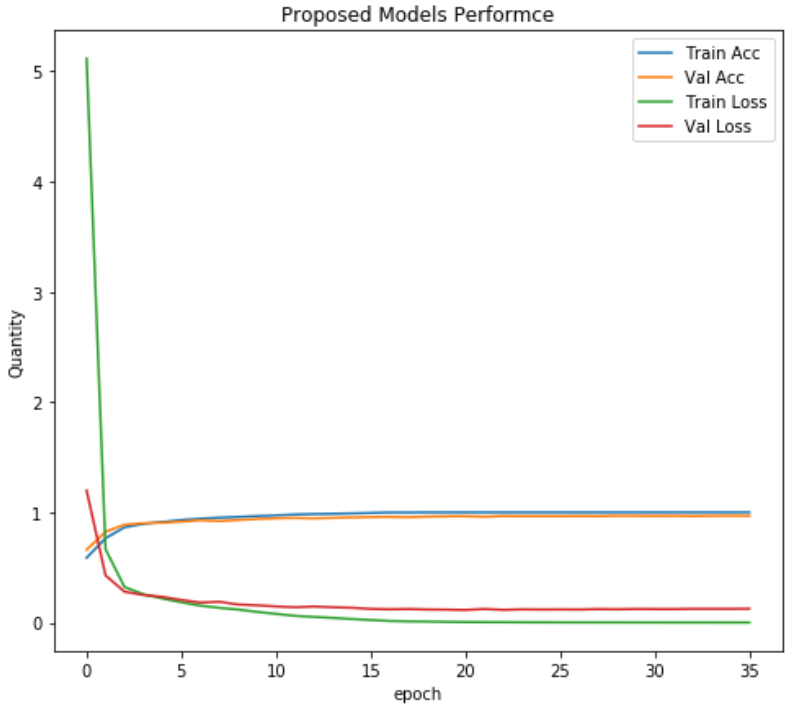

Fig. 4.7: Proposed Model's Accuracy and Loss over epochs.

Table 4.2: Overall performance of the proposed GRU model.

Evolution Matrixes	Predicted Outcome
Accuracy	96.91%
Sensitivity	97.95%
Specificity	96.16%
False Positive Rate	3.84%
Precision	94.89%
F1 Score	96.39%
AUC	97.05%

Fig. 4.8 shows the confusion matrix of our GRU-based classifier. There were 938 HVs and 682 MCI testing samples to measure the performance. Among the HV testing samples, 902 (96.16%) samples were correctly identified as HVs, and 36 (3.83%) samples were misclassified as MCI. Again, among the MCI testing samples, 668 (97.95%) samples were correctly recognized as MCI, and 14 (2.5%) samples were misclassified as HV. It can be said that the misclassification rate is very low compared to the properly identified rate. And the false alarm rate tends to decrease if I increase the number of testing samples.

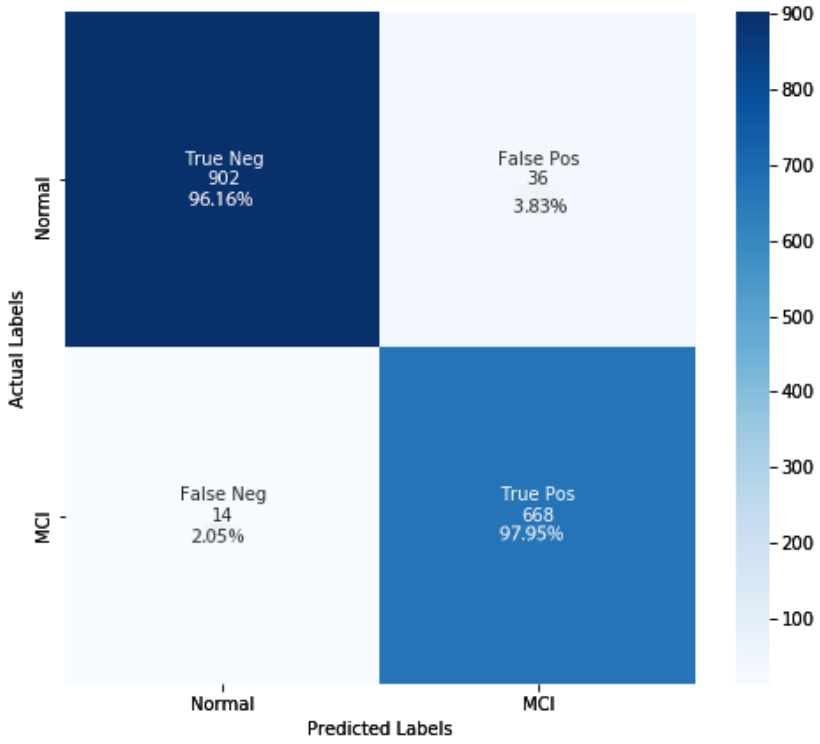

Fig. 4.8: Confusion matrix of the GRU based model.

The receiver operating characteristic (ROC) curve is a measure to recognize a classifier's performance. It demonstrates the relationship between sensitivity and specificity. The ROC curve is shown by the x-axis, which depicts 1 – specificity = FP / (FP+TN), and the y-axis, which depicts sensitivity = TP / (TP+FN). The blue curve in **Fig. 4.9** illustrates the ROC curve, and the area under the ROC curve is known as the AUC value. The bigger the value, the better the model is. The AUC value always stays between 0 and 1. In our case, after multiplying by 100, the AUC value I have is 97.05 percent, which proves the efficiency of our proposed model.

4.5 Discussion

This study is committed to differentiating HV EEG signals from irregular EEG signals. For that purpose, I have chosen 27 EEG recordings (16 HVs and 11 MCIs). The framework I have designed performed really well in differentiating HV EEG samples from the abnormal ones, with an accuracy of 96.91%. It is observed that the false alarm rate is 9.84%, which is very low. The reason behind it could be a smaller dataset and artifacts affecting the EEG data recording. To our best knowledge, the GRU model has never been set up with this dataset, and it has shown promising performance. Previous attempts to classify EEG data with TML methods had shown promising performance, but

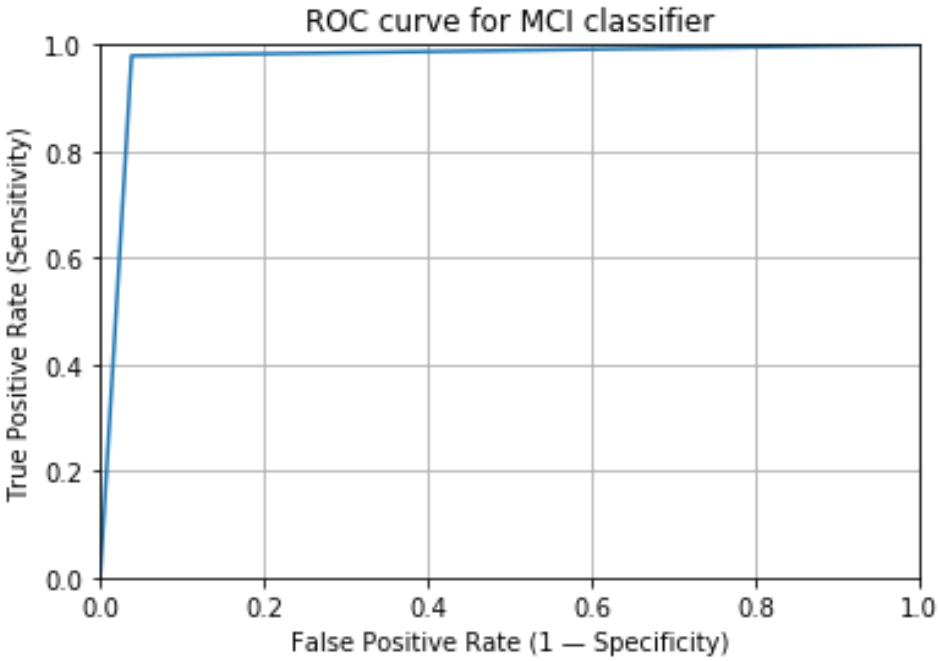

Fig. 4.9: ROC curve showing the efficiency of the proposed model.

the computational expense to build the shallow architectures and feature extraction methods made it costly. An overall comparison with previous attempts including both TML and DL methods alongside our proposed method is reported in **Table 4.3**.

Previous studies with a shallow architecture failed to have a cost-effective and efficient model. From the literature, it can be said that very few studies have used DL-based models to separate the regular EEG signals from the abnormal ones. Most of the previous attempts evolved with TML algorithms like SVM, KNN, and LR, etc., which do not have the power to extract the features for themselves. They need to be fed manually extracted features by different feature extraction methods. It is one of the main reasons for missing important features, which has an impact on the classifier's performance. And due to this extra step of feature extraction, the cost of the model increases, whereas the DL-based model can do this feature extraction by itself. But it often leads to a non-efficient model if the sample size is too small.

Our proposed GRU model with two deep hidden layers and a dense layer is too simple to design. It has 1024, 512, and 1 neurons respectively. The hidden layers have tanh," whereas the dense layer has "*sigmoid*" as activation functions. With such a simple design, it achieves very good accuracy. It took 36 epochs to reach such efficiency, and on average, each epoch lasts for only 1120 seconds.

Table 4.3: Comparison with previous efforts.

Studies	Accuracy	Sensitivity	Specificity
Kashefpoor *et al.* [50]	88.89%	100%	83.33%
Yin *et al.* [100]	96.94%	96.89%	96.99%
Hadiyoso *et al.* [118]	81.5%	81.82%	81.25%
Kashefpoor *et al.* [117]	88.9%	83.3%	100%
Chen *et al.* [119]	92.06%	Not reported	Not reported
Jamalooa *et al.* [116]	95.9%	97.2%	97.2%
G. Vrbancic and V. Podgorelec [146]	69.23%	25%	88.89%
Amezquita-Sanchez *et al.* [13]	90.3%	92.1%	87.9%
Yang *et al.* [122]	90.37%	Not reported	Not reported
Proposed method	**96.91%**	**97.95%**	**96.16%**

4.6 Synopsis

Our proposed GRU-based anomaly identification framework has proven its capability in differentiating HV EEG data. GRU has never used this publicly available EEG dataset to separate the anomalies. It is a big challenge to work with a smaller sample size, particularly when I use a DL-based model as a classifier. So, I had to pre-process the raw data with some de-noising, segmentation, and down-sampling. De-noising helped me to remove the unwanted signals from the raw data, and segmentation helped me to increase the sample size. As I wanted to have a simpler model that could be trained on a regular machine with a simple configuration, I down-sampled the segmented data. This down-sampling helped me reduce the data dimension, and while doing this, I had to sacrifice some of the data. That is one of the main reasons I do not have a perfect model. But while down-sampling, I tried our best not to lose much of the data that may cause me to have a less efficient model.

In the end, I can say that our proposed GRU-based anomaly detection framework is a balanced model with a simpler architecture, not much cost to design, and high performance in identifying HV EEG signals from the irregular ones. The succeeding study should focus on having a larger dataset and investigating this GRU architecture with different amounts of hidden layers and activation functions. I believe that this chapter will provide future EEG researchers with a pathway towards a perfect HV EEG identification model.

At the end of this framework, I still had some further fine-tuning in mind with this same EEG data and GRU classifier. Though this proposed GRU-based framework has outperformed all previous MCI efforts with the same EEG data as well as our own LSTM-based MCI detection effort, I wanted to give it further tries to improve this current GRU framework. There are a few new EEG pre-processing methods along with three other classifiers, and I have proposed an improved GRU-based MCI detection framework in the next chapter using the same EEG data.

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: A deep learning based framework for diagnosis of mild cognitive impairment

Surname: ALVI First name: ASHIK MOSTAF

Institute: Institute for Sustainable Industries and Liveab Candidate's Con 60

Status:

Accepted and in press: ☐ Date:

Published: ☑ Date: 19/07

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be includ ned in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted] 24/01/2023

Signature **Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

Name(s) of Co-Author(s)	**Contribution (%)**	**Nature of Contribution**	**Signature**	**Date**
Ashik Mostafa Alvi	60	Conceptualization, Methodology, Data Curation, Investigation, Software,	[redacted]	24/01/2023
Siuly Siuly	10	Formal Analysis, Validation, Writing–Review and Editing, Supervision.		24/01/2023
Hua Wang	10	Resources, Supervision, Writing–Review and Editing, Funding Acquisition, Project		26/01/23
Kate Wang	10	Formal Analysis, Validation, Writing–Review and Editing.		26/01/23
Frank Whittaker	10	Writing–Review and Editing, Funding Acquisition, Project Administration.		26/01/23

Updated: September 2019

CHAPTER 5 : A DEEP LEARNING BASED FRAMEWORK FOR DIAGNOSIS OF MILD COGNITIVE IMPAIRMENT

5.1 Overview

Detecting mild cognitive impairment (MCI) from electroencephalography (EEG) data is a challenging problem as existing methods rely on machine learning (ML)-based shallow architectures that are unable to successfully uncover relevant biomarkers from deep hidden layers of data. This study will design a deep learning (DL)-based framework, including a gated recurrent unit (GRU) model, for effective detection of MCI participants from healthy volunteers (HVs) utilising EEG data. MCI is a gradual, irreversible neurodegenerative illness that is frequently a precursor to Alzheimer's disease (AD) and can result in dementia in the elderly. There is no cure or treatment to stop or reverse the course of MCI; early identification is critical for the successful application of treatment strategies to enhance the quality of life of patients. The proposed framework consists of four steps: gathering raw EEG data; pre-processing of the raw data (de-noising, segmentation, and down sampling); uncovering hidden features and classification of MCI subjects from HVs; and performance evaluation of the proposed model. The proposed GRU model has been compared with long-short-term memory (LSTM), support vector machine (SVM), and K-nearest neighbour (KNN) classifiers. The stability of the framework is evaluated through 5-fold cross-validation. The suggested GRU-based DL model serves as a reliable biomarker and aids technicians in developing a new automatic MCI detection method. This chapter also shares answers to **research questions 2, 4, and 5**.

The contents of this chapter have been published in the *Knowledge-Based Systems* [107].

5.2 Introduction

MCI is a neurological disorder engendered by cognitive decline. Memory loss is far more severe in those with MCI than it would be in someone of a similar age group. MCI is caused by memory loss, a reduction in vocabulary, and a diminished ability to make precise motor motions, all of which have an impact on day-to-day activities [106]. The

primary causes of this neurologic illness are neuronal cell death and malfunction [66], [65]. According to recent research, MCI patients are more likely to acquire dementia, particularly AD [108], [109], [110]. Dementia is one of the leading sources of impairment and reliance among the elderly across the world. MCI and dementia are on the rise globally, with emerging nations expected to have a greater increase [112], [10]. It is Australia's second most common cause of death [113] and the world's seventh most common cause of death [114]. If no medical breakthroughs are made, the anticipated 472,000 Australians living with dementia in 2021 would rise to 590,000 by 2028 and 1,076,000 by 2058 [135]. The Alzheimer's Disease International (ADI) report estimates that there were more than 50 million dementia sufferers globally in 2019, with that number expected to rise to 152 million by 2050 [115]. There is currently no permanent cure for MCI or dementia; however, early detection can temporarily halt the progression of symptoms, therefore improving patient and caregiver quality of life [139]. Therefore, it is critical to diagnose MCI early on so that patients may receive quick medical attention, and it is a study topic that is still being explored. The life of an MCI subject can be saved and enhanced by using the right procedures and getting a quick and accurate diagnosis.

Mini-Mental State Examinations (MMSE), Magnetic Resonance Imaging (MRI), Computed Tomography (CT) scans, blood tests, Positron Emission Tomography (PET), spinal fluid, and EEG are some of the instruments used in today's medical systems to identify MCI. PET, MRI, and CT scanning are expensive alternatives, whereas the MMSE exam is a manual question-and-answer test. EEG, on the other hand, is widely acknowledged as a standard instrument for examining MCI since it is a cost-effective, non-invasive, and portable instrument. EEG recordings reflect the electrical activity in the cerebral cortex over time. It is one of the most useful tools for assessing cognitive issues. Therefore, this study looks at the use of EEG as a technique for early detection of MCI.

Two traditional machine learning (TML) and two DL-based approaches to detect MCI have been presented to address the mentioned research gaps. This is an extension of our previous GRU-based MCI detection framework [104]. This proposed study starts with discarding unwanted signals and artifacts from the acquired public EEG data using the stationary wavelet transform (SWT). Filtered signals have been segmented and down-sampled from 256 Hz to 4 Hz using an average filter. Finally, LSTM, GRU, SVM, and KNN classifiers have been designed to differentiate MCI subjects from HVs. SVM and

KNN have used the LSTM-extracted features. This extensive work has introduced a different EEG signal de-noising technique and three new classifiers compared to our previous effort to identify MCI at the preliminary stage. The following are the major contributions of this study:

- For the first time, an LSTM and GRU-based DL study for MCI classification has been accomplished.
- I have utilized the LSTM-extracted features for SVM and KNN instead of using separate feature extraction methods to save computational cost.
- To enhance the suggested model's quicker performance, I have examined the average filtering strategy for down sampling.
- With reduced calculation time, our approach improves classification accuracy, and the consistency of the performance has been checked by 5-fold cross validation scheme.

The rest of the chapter is assembled as follows: Background reviews are presented in Section 5.3. The suggested model framework is described in detail in Section 5.4. Section 5.5 contains the full outcomes and analyses, as well as the results of the investigations. Section 5.6 has a qualitative discussion. Finally, in Section 5.7, this study concludes with a summary of findings and recommendations for further investigations.

5.3 Background Studies

Many studies have been undertaken over the last few decades to diagnose MCI at an early stage so that it does not progress to AD or other forms of dementia. Siuly *et al.* [84] used the piecewise aggregate approximation method to compress the EEG data, permutation entropy and auto-regressive to extract the features, and an extreme learning machine (ELM) to classify MCI subjects. Recently, a multiclass study to diagnose MCI, AD, and HVs using discrete wavelet transform (DWT), power spectral density (PSD), and coherence for feature extraction has been proposed [152]. The bagged tree was chosen as the classifier for this study. Amezquita-Sanchez *et al.* [135] performed the empirical wavelet transform, also known as MUSIC-EWT, for noise reduction and fractality dimension (FD) from the chaos theory for data compression. Processed data was fed into an enhanced probabilistic neural network (EPNN) and achieved 90.3% accuracy. A single channel-based MCI study [52] was conducted with 23 subjects, and event-related

potential (ERP) was measured to capture the features. SVM attained 87.9% accuracy, 84.8% sensitivity, 95% specificity, and an 85% F score in leave-one-out cross validation.

Al-Qazzaz *et al.* [135] studied 5 vascular dementia (VaD) subjects, 15 patients with MCI, and 15 HVs to develop a multiclass MCI-dementia detection model. Independent component analysis (ICA) and wavelet transform (WT) were used for EEG analysis. Linear features were extracted using relative powers (RP) and power ratios of ANOVA analysis, and non-linear features were uncovered using permutation entropy and FD. KNN and SVM were applied, and SVM outperformed KNN and achieved 91.48% accuracy. Another SVM-based study [75] was conducted where finger tapping test (FTT), continuous performance test (CPT), power spectral density (PSD), skewness, kurtosis, spectral skewness, spectral kurtosis, spectral crest factor, spectral entropy (SE), and FD were measured for EEG analysis of 44 subjects. This study achieved 89.8% accuracy in FTT mode.

A DL-based approach [86] along with wavelet analysis was recently proposed to classify MCI patients. A bidirectional long short-term memory (BLSTM) network had been developed and gained 69.53% accuracy. Meghdadi *et al.* [153] measured PSD at each channel, spectral coherence between pairs of channels, principal component analysis (PCA), and linear discriminant analysis (LDA) for classification. This model obtained an AUC value of 0.85 for cross-validation. Using the statistical pattern recognition (SPR) method, another MCI prediction [154] was accomplished where logistic regression (LR) was used for differentiating MCI subjects and received a 0.78 AUC value, 71% sensitivity, and 69% specificity.

Kashefpoor *et al.* [50] used a Takagi-Sugeno neurofuzzy (NF) inference system with KNN to perform an EEG investigation using a correlation-based pursuit for feature extraction. Although the study's sensitivity was 100 percent, its accuracy and specificity were only 88.89 percent and 83.33 percent, respectively. The EEG data was preprocessed using dictionary learning [117]. The study used three classifiers: K-means and singular value decomposition (K-SVD), label-consistent KSVD (LC-KSVD), and correlation-based LC-KSVD (CLC-KSVD), with CLC-KSVD achieving the highest accuracy of the three. Chen *et al.* [119] published a graph theory-based EEG classification that included segmentation and the creation of a brain network connectivity matrix. A Convolutional neural network (CNN) was used to categorize EEG individuals once the preprocessing

stages were completed. The CNN classifier was right 92.06 percent of the time. For EEG categorization, a four-layer CNN was created, combined with a medical experiment to assess variations in oxygenated hemoglobin [122]. This CNN model was 90.37 percent accurate in identifying MCI subjects from HVs.

According to the examined literature, TML algorithms outperformed a few DL-based attempts more consistently and effectively. But when it comes to feature extraction, TML algorithms are not a time-saving alternative. An extra step is always necessary when it comes to studying the TML algorithm. Furthermore, due to the complexity of EEG data, detecting discriminant characteristics requires greater attention. The shallow design of TML approaches makes it more time-consuming and difficult to manage large amounts of data, such as EEG.

5.4 Proffered Framework

This extended study has submitted a comprehensive MCI identification framework using EEG data. The proffered framework consists of four steps, and **Fig. 5.1** exhibits them in a nutshell. (i) Accumulating EEG Data; (ii) Pre-processing Raw Data; (iii) Uncovering Hidden Features and Classification of MCI and Normal Subjects; and (iv) Performance Evaluation are the four steps towards the entire framework described in the following sub-sections.

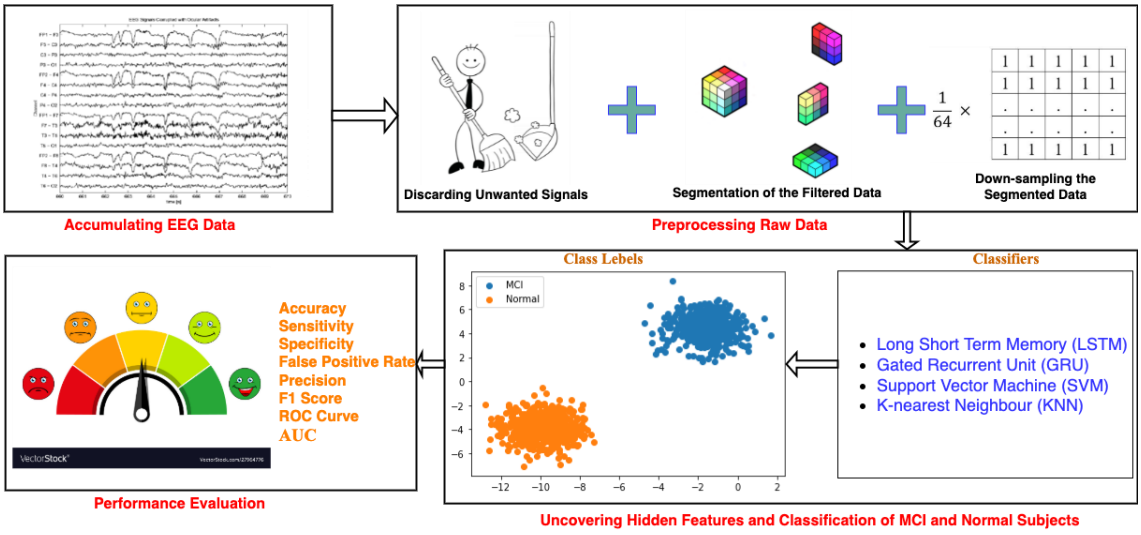

Fig. 5.1: A comprehensive MCI identification framework.

5.4.1 Accumulating EEG Data

This MCI study has exercised a publicly available EEG dataset of 27 subjects from the Sina and Nour Hospital, Isfahan, Iran [50], [123]. 16 HVs and 11 MCI subjects were part of this EEG data collection. All of these subjects were aged between 60 and 77 and had given their consent for the study. The deputy of research and technology, Isfahan University of Medical Sciences, Isfahan, Iran, had ethically approved the data collection. A neuropsychiatric interview, also known as a mini-mental state examination (MMSE), was performed for each subject following Petersen's criteria. The MMSE scores of MCI and HVs were between 21 and 26, respectively. None of these 27 subjects had head trauma, dementia, or a history of major psychiatric disorders, serious medical disease, or substance misuse. **Table 5.1** provides the demographic and psychiatric information of the participating subjects.

Table 5.1: Demographic and Psychiatric Information of the Dataset.

Characteristic	MCI	Normal
Age (years)	66.4 ± 4.6	65.3 ± 3.9
Education (years)	10.3 ± 3.8	11.1 ± 3
GHQ Scores	20.5 ± 9.4	17.9 ± 6.6
BMI (kg/m^2)	25.7 ± 2.2	26.6 ± 3.6
Fasting glucose (mg/dl)	115.5 ± 24.3	121.8 ± 36.9
Total cholesterol (mg/dl)	170.6 ± 61.4	169.1 ± 42.6
Triglycerides (mg/dl)	157.3 ± 100.9	160 ± 80.7
Creatinine (mg/dl)	1.2 ± 0.2	1.3 ± 0.3
MMSE scores	27.6 ± 0.9	29 ± 0.8
NUCOG scores	82.4 ± 3.6	91.1 ± 3
Gensini scores	33.3 ± 31.9	20.3 ± 21.7

GHQ – General health questionnaire; BMI – Body mass index; MMSE – Mini-mental state examination; NUCOG – Neuropsychiatry unit cognitive assessment tool

Following the international EEG recording standard also known as international 10-20 system, 19 electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4,

T6, O1, O2) were used to capture the electron movement across the scalp. All subjects were instructed to close their eyes and feel relaxed and conscious to avoid drowsiness during the EEG recordings. For each subject, recording lasted for 30 minutes, and signals were digitised using a 32-channel EEG device with a sampling frequency of 256 Hz, keeping the impedance of the electrodes skin less than 5 kΩ. **Fig. 5.2** displays 2 seconds of EEG recordings at the Fp2 electrode position of an MCI and a HV.

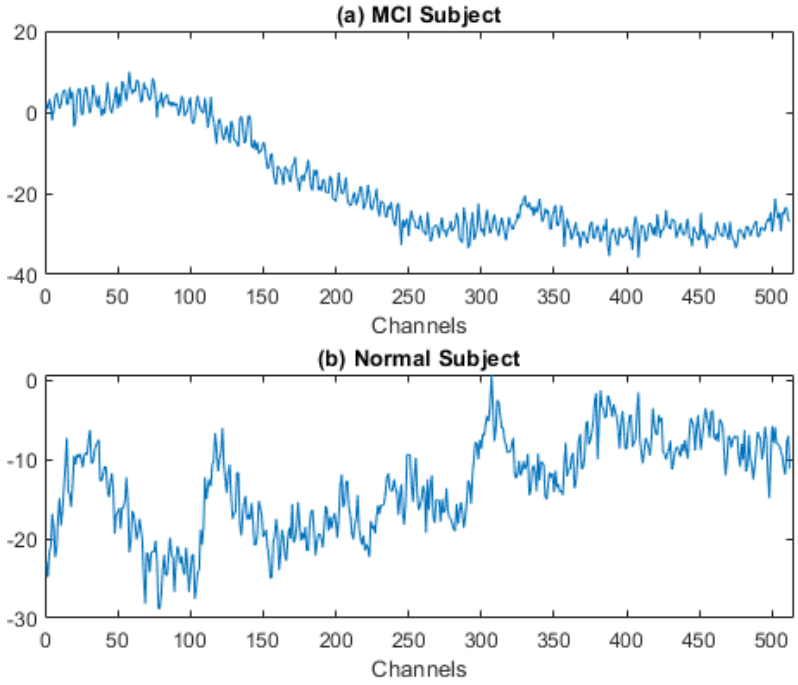

Fig. 5.2: 2 seconds EEG signal at Fp2 electrode position of (a) an MCI, (b) HV.

5.4.2 Preprocessing Raw Data

To ensure the quality of the acquired EEG dataset, recorded signals have been processed with different mechanisms. (i) Discarding unwanted signals, (ii) Segmentation of the Filtered Data, and (iii) Down-sampling the Segmented Data are the processing steps that are followed and discussed below.

5.4.2.1 Discarding Unwanted Signals

It is a very common scenario when unwanted signals, so-called noises, get mixed with signals while recording [9], [7]. These noises corrupt the original signals and make it mandatory to de-noise the signal before further processing. Some common artifacts and noises that contaminate EEG signals are electrode pops, outlier values, baseline drift, power supply fluxes and interference (50 Hz), breathing, eye blinking, muscular electrical

activities, etc. To ensure a high quality of EEG signal processing, it is a must to employ a noise remover to discard unwanted signals and artifacts.

This MCI study has performed SWT to remove unwanted artifacts and signals. SWT is capable of dealing with both high and low-frequency noises, and it can maximise the flat response of the given signal. To eradicate DC components (0–0.5 Hz) and base line drift in each of the channels, the symmetrized wavelet "*sym9*" with an 8th order has been deployed. Finally, high-frequency noises (32–128 Hz) and decomposition of the stationary wavelets have been carried out with second-order approximations for each of the channels.

After de-noising, 27 filtered signals are digitized with a length of 60 seconds x 30 minutes x 256 Hz each and saved as .mat files. Each of these 27 .mat files contains 460800 rows representing voltage amplitude and 19 columns representing channels. All these operations have been brought to pass in MATLAB. There is a visual comparison of (a) the raw EEG signal of an MCI subject and (b) the de-noised signal of that same MCI subject shown in **Fig. 2.7**.

5.4.2.2 Segmentation of the Filtered Data

By nature, EEG signals are non-stationary, non-periodic, and huge in size. Therefore, for smooth processing of EEG data, it is a very common practice in EEG studies to segment the data. It reduces the computational overhead a bit and increases the sample size while not losing any features.

In this study, I have segmented the whole 30-minute recordings of each subject, giving me 300 temporal segments out of each subject's recording. The new duration of each temporal segment is 6 s, which gives me 1536 rows and 19 rows for new temporal subjects. After segmenting 27 subjects, I now have 8,100 temporal subjects.

5.4.2.3 Down-sampling the Segmented Data

I was able to expand the sample size thanks to segmentation, but I still have a massive quantity of data to cope with. And these massive datasets have impacts on training and testing time. It requires a lot of computational power to process the data. It is a compromise between losing data in bulk and preserving critical characteristics at the same time. To overcome this situation, I exert the average filtering method.

The segmented dataset has been down-sampled using the average filter, which reduces computational cost. The average filter works well when the dataset is outlier- and noise-free. Therefore, this filter suits our dataset and accomplishes the job. An average filter with a sliding window of size 64 x 1 has been used to reduce the sampling frequency. I execute the sliding window column-wise while maintaining the number of channels constant. Up until the last segmentation step, all the temporal segments held a sampling frequency of 256 Hz. Now, the sample frequency has been changed to 4 Hz, giving me 24 rows (6 seconds x 4 Hz) and 19 columns (channels). Each of the down-sampled temporal segments has been saved in comma-separated values (CSV) format in MATLAB.

5.4.3 Uncovering Hidden Features and Classification of MCI and Normal Subjects

Hidden features are event-relevant potentials that convert acquired EEG signals into a compact number of related values. This MCI study aims to discover important features of DL and identify MCI subjects based on those discriminative features. For feature extraction, no separate method is used, and for classification, four different ML models are tested with the processed EEG dataset. LSTM, GRU, SVM, and KNN are the four ML classifiers chosen for this study. Pre-processed EEG data from the previous down-sampling step is used as input to the classifiers. For all four classifiers, training, testing, and validation sets remain constant at 70%, 20%, and 10%, respectively. To ensure no overfitting occurs during the training process, early stopping has been set to monitor validation loss, and *min_delta* was set to 0.001 while keeping the patience value of 10. The batch size and number of epochs have been set to 300 and 250, respectively, for all the classifiers.

5.4.3.1 Long Short-Term Memory (LSTM)

LSTM is a variant of the recurrent neural network (RNN) and works well with sequential data like EEG. RNN's distinctive strength is its ability to remember the past of successive data, which aids in more accurate prediction. Each LSTM unit's deep state demonstrates its memory. Cells develop the ability to output, clean their memory, or replace it depending on the history of their internal states and the most recent input. In **Fig. 3.4**, x denotes the input signals at time t, y denotes the output signals, and I, f, and o denote the

input, forget, and output gates, respectively. The stream of data into or out of LSTM's memory is managed using those three gates.

Working with a sequence s = (s_1, s_2, s_3,..., s_t), the RNN modifies its recurrent covert state c_t at each time step t depending on the current input vector s_t and the previous covert state c_t-1, as shown in (14) where m denotes a nonlinear function. RNNs have hidden layers whose activation is based on the preceding time. Because part of the network's linkages form a directed cycle, this is a recurrent design in which the current time step t considers the network's state in time step $t - 1$. The layers share properties based on different time steps, making them suitable for use with sequential data.

$$c_t = \begin{cases} 0 & if\ t = 0 \\ m(c_{t-1}, s_t) & otherwise \end{cases} \quad (14)$$

RNNs have hidden layers whose activation is based on the preceding time. Because part of the network's linkages form a directed cycle, this is a recurrent design in which the current time step t considers the network's state in time step t – 1. The layers share properties based on different time steps, making them suitable for use with sequential data. In LSTM, the recurrent covert unit of (14) is altered as given in (15).

$$c_t = f\ (W1s_t + W2c_{t-1} + x) \quad (15)$$

W1 and *W2* are two weight matrices, and *x* is a bias vector, with *f* serving as a pointwise nonlinear activation function. Equation (15) allows an LSTM to operate with orders of any length, and the gradients of *f* may grow exponentially as the network is being trained. When the error derivatives are back-propagated across multiple layers, LSTM cells are modelled to capture the influence of diminishing gradients.

In this study, multiple experiments have been run with different LSTM networks set up. Among them, an LSTM network with two hidden layers having 1024 and 512 nodes, respectively, has performed well. It also has a flat, dense layer with a "*sigmoid*" activation function and a single node. Hidden layers are set up with the "*tanh*" activation function, "*binary_crossentropy*" as the loss function, and "*Adam*" as the optimizer.

5.4.3.2 Gated Recurrent Unit (GRU)

GRU is another RNN variant and a cost-effective option in terms of data memorization. LSTM and GRU are comparable; however, GRU contains fewer parameters. Similarly, both the LSTM and GRU have gates. The *update gate* u_t is one, while the *reset gate* r_t is

the other. u_t specifies which of the candidate's previous states will be replaced with a new value, and r_t specifies how much of the candidate's previous state should be ignored while determining the candidate's hidden state. A sample GRU architecture in **Fig. 5.3** is characterized by two internal variables that keep track of the prior *h* and present *h* inner states, respectively. GRU maintains the stream of data without the need for a memory unit to guard against the diminishing gradient issue.

The current value of the covert state c_t is derived using linear interpolation between the immediate candidate covert state $\tilde{c}t$ and the old value of covert state c_{t-1}. From (15), $\tilde{c}_t$ is calculated.

$$c_t = (1 - u_t) \odot c_{t-1} + u_t \odot \tilde{c_t} \tag{16}$$

$$\tilde{c_t} = g(W1_c s_t + W2_c(r_t \odot c_{t-1}) + x_c) \tag{17}$$

$$u_t = \sigma\,(W1_u s_t + W2_u c_{t-1} + x_u) \tag{18}$$

$$r_t = \sigma\,(W1_r s_t + W2_r c_{t-1} + x_r) \tag{19}$$

Here, g and σ are nonlinear GRU activation functions. The $\odot$ operator represents the multiplication of individual elements.

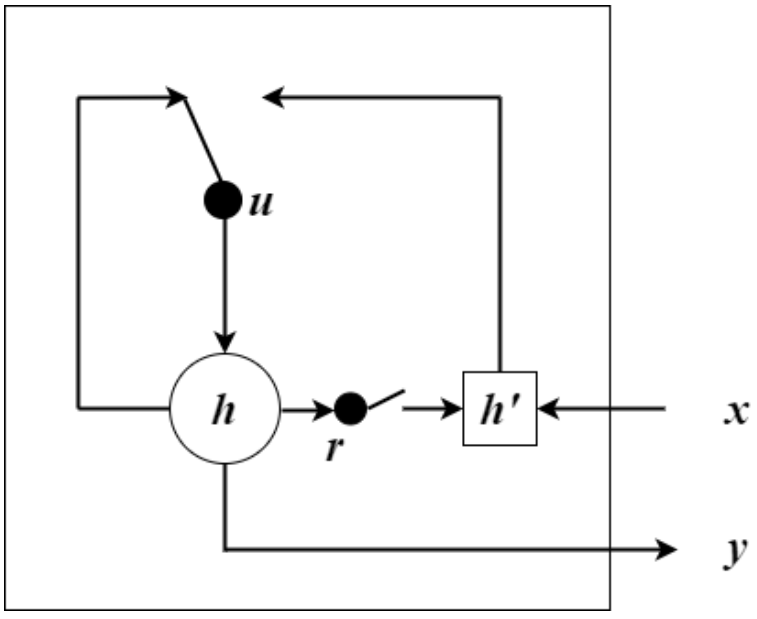

Fig. 5.3: Architecture of a simple GRU.

This study works with EEG data, and it requires extra attention for its sequential and complex data pattern. GRU is another updated version of the classical RRN that is famous for working with sequential and complex data patterns like EEG. There are hidden layers

in the recurrent network that are very good at capturing the complex EEG data's hidden features, like how the brain waves change over time. These recurrent networks also help keep the information from the previous layer safe and sound. But unlike LSTM, GRU does not hold redundant information in its memory. By using the update and reset gates, it efficiently controls the data flow in the memory. The GRU network that has been designed for this study has the exact same configuration as the LSTM network. But this GRU network is light in terms of memory holding and training processes. The GRU algorithm used in this study has been reported below:

Algorithm 2 The pseudo-code of **GRU**

Input: Down-sampled CSV files from the pre-processing step

Output: Trained model in hdf5 format

1. **Repeat**
2. store each CSV file in pandas dataframe
3. **Until** any temporal segments are left
4. **repeat**
5. initialize ***Q***, convert all the dataframes to numpy arrays and store in ***Q***
6. **until** any temporal segments are left
7. **Store** the labels in a numpy array named ***L***
8. **Reshape** ***Q*** and make it one dimensional
9. **Split** the dataset into train, test, and validation sets
10. **Build** the model

    Add a sequential layer

    Add a hidden layer with 1024 neuron and ‘tanh’ activation function

    Add a hidden layer with 512 neuron and ‘tanh’ activation function

    Add a flatten layer

    Add a dense layer with ‘sigmoid’ activation function
11. **Initialize**: The early stopping with min_delta=0.001, patience=10, monitor = ‘val_loss’
12. **Fit** the model with the training dataset, epoch 250, batch size 300

5.4.3.3 Support Vector Machine (SVM)

SVM is very popular among EEG researchers. From the background study, it can be concluded that many MCI studies were conducted using an SVM classifier. The common and traditional ML practise for EEG studies is to extract the hidden features for the SVM classifiers using different time- or frequency-domain methods. But in this MCI study, I have kept it simple and computationally inexpensive by extracting the hidden features using an LSTM network. Using the predefined SVC function in the Scikit-Learn library, an SVM classifier has been executed.

The feature extraction and classification layers have been separated and stored in the "*getFeature*" vector. Then, *x_train* and *x_test* sets have been reorganised from the "*getFeature*" vector. Polynomial has been set as the kernel function, regularisation parameter *C* has been holding a value of 2, kernel scale is set to auto, and degree of the polynomial kernel is set to 2.

5.4.3.4 K-nearest Neighbor (KNN)

KNN is another TML algorithm often used for EEG studies. Using the Scikit-Learn library, the KNN algorithm has been implemented. The same extracted features from the LSTM network have been fitted to this KNN classifier. Different values for the number of neighbours have been tried out, and the best result is when the number of neighbours is 10. The algorithm to compute the nearest neighbour has been set to *auto*, the "*Manhattan*" distance metric has been chosen, and weights have been selected to "distance". The rest of the parameters have been holding their default values.

5.4.4 Performance Evaluation

Numerous experiments have been conducted with the mentioned classifiers and their different configurations to achieve the best possible result. To check the stability and consistency of those results, 5-fold cross validations have been done for each of those four classifiers. These four models have been examined with eight standard evaluation matrices. (6), (7), (8), (9), (10), (11), the receiver operating characteristic (ROC) curve, and (12) the area under the ROC curve (AUC) are the performance evaluation matrices picked to find the best model. The ROC curve depicts the graph of sensitivity (true positive rate) versus false positive rates.

5.5 Investigations and Outcomes

A publicly available EEG dataset of 16 HVs and 11 MCI subjects has been used to test our study. EEG recoding of those 27 subjects took place in accordance with the international 10-20 system, with 19 electrodes placed across the scalp and stored in European Data Format (EDF) format.

The proposed pre-processing steps are done using the MATLAB 2019b version, and after the down-sampling stage, EEG data are saved as CSV files for easy access to the feature extraction and classification part. The code for the hidden feature extraction and classification parts has been written in Python, and Jupyter Notebook has been used as the IDE. All the investigations were accomplished on a MacBook Pro machine with a 2.6GHz Intel Core i7 CPU, 16GB of RAM, and a 4GB graphics card.

Table 5.2: Overall Performance of Various ML Classifiers.

Classifiers	Accuracy %	Sensitivity %	Specificity %	F1 Score %	AUC %
LSTM	95.47 ± 1.17	96.31 ± 0.91	95.18 ± 1.29	94.53 ± 1.13	95.42 ± 1.03
GRU	**95.51 ± 3.11**	**97.52 ± 0.96**	**96.50 ± 0.97**	**95.69 ± 2.26**	**96.48 ± 1.85**
SVM	92.31 ± 2.04	95.06 ± 0.30	90.53 ± 0.91	91.25 ± 2.25	92.33 ± 2.02
KNN	80.85 ± 1.78	87.58 ± 0.84	80.71 ± 1.46	82.53 ± 1.12	81.65 ± 3.62

Table 5.2 reports the overall performance of the four classifiers mentioned. To validate the performance, a 5-fold cross-validation scheme has been measured for each of these classifiers. Performance metrics (accuracy, sensitivity, specificity, F1 score, and AUC) with a 5-fold standard deviation have been revealed in **Table 5.2**. It can be seen from **Table 5.2** that GRU has outperformed all other classifiers with the highest accuracy of 95.51 ± 3.11%, sensitivity of 97.52 ± 0.96%, specificity of 96.50 ± 0.97%, F1 score of 95.69 ± 2.26%, and an AUC value of 96.48 ± 1.85%. Two TML algorithms, SVM and KNN, have failed to perform well in terms of performance metrics. KNN has achieved the lowest accuracy of 80.85 ± 1.78%, a sensitivity of 87.58 ± 0.84%, a specificity of 80.71 ± 1.46%, an F1 score of 82.53 ± 1.12%, and an AUC value of 81.65 ± 3.62%. LSTM was close to the lead with GRU, but SVM has fallen far behind.

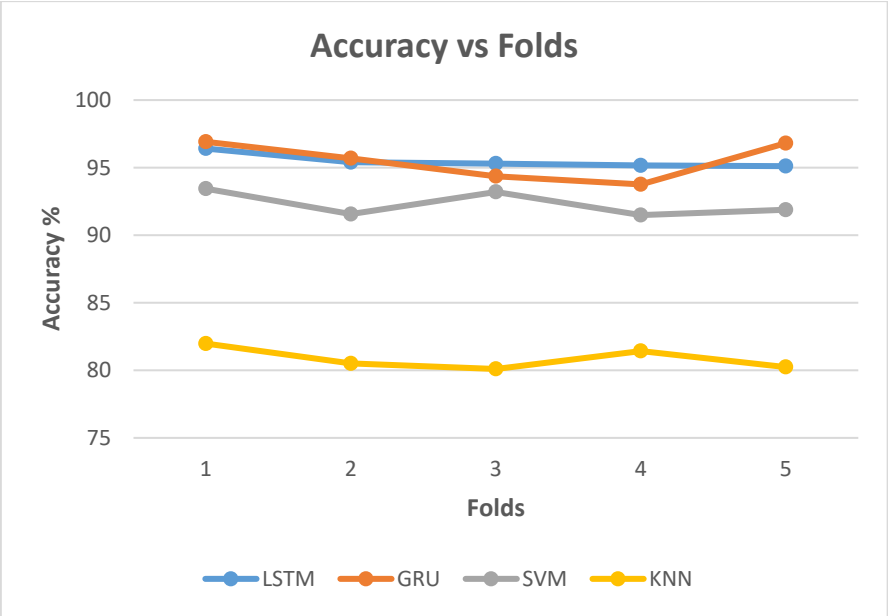

Fig. 5.4: Overall accuracies over folds of LSTM, GRU, SVM, and KNN classifiers.

The accuracy versus folds graph has been plotted in **Fig. 5.4**. GRU has remained in the top position with an average accuracy of 95.51% for most of the folds except 3 and 4. LSTM has gained the highest accuracy in the 3rd and 4th folds. Another observation is that, apart from KNN, the accuracy of the remaining classifiers gradually decreased in the fourth fold. In **Fig. 5.5**, sensitivity versus folds is shown. GRU has maintained the top position among the other classifiers.

There are enough distances among the plotted sensitivity lines. Fold-wise specificity has been illustrated in **Fig. 5.6**. In the first fold, the LSTM and GRU specificity lines almost have the same value. **Fig. 5.7** shows the fold-wise false positive rates of the mentioned classifiers. GRU has the lowest average false positive rate of 4.11%. In the 1st, 3rd, and 4th folds, LSTM and GRU almost have the same false-positive rates.

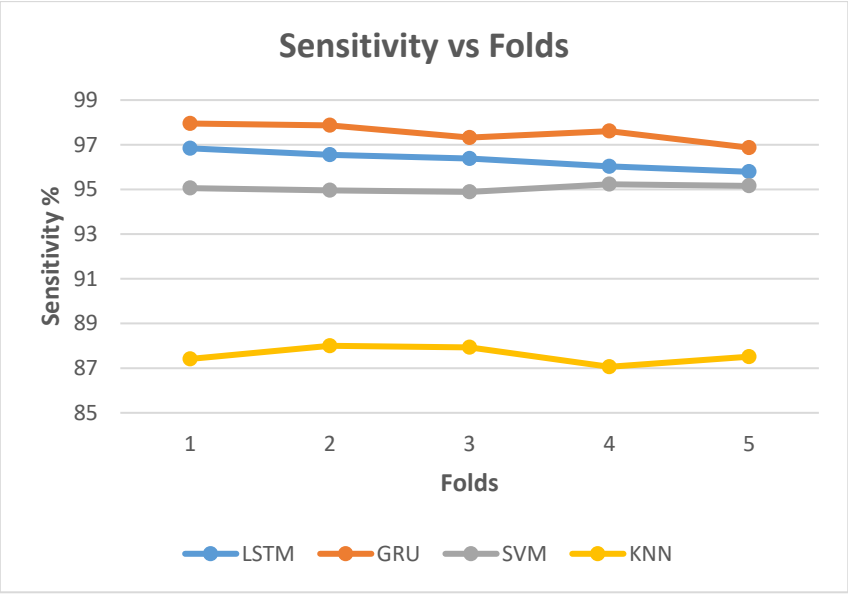

Fig. 5.5: Overall sensitivities over folds of LSTM, GRU, SVM, and KNN classifier.

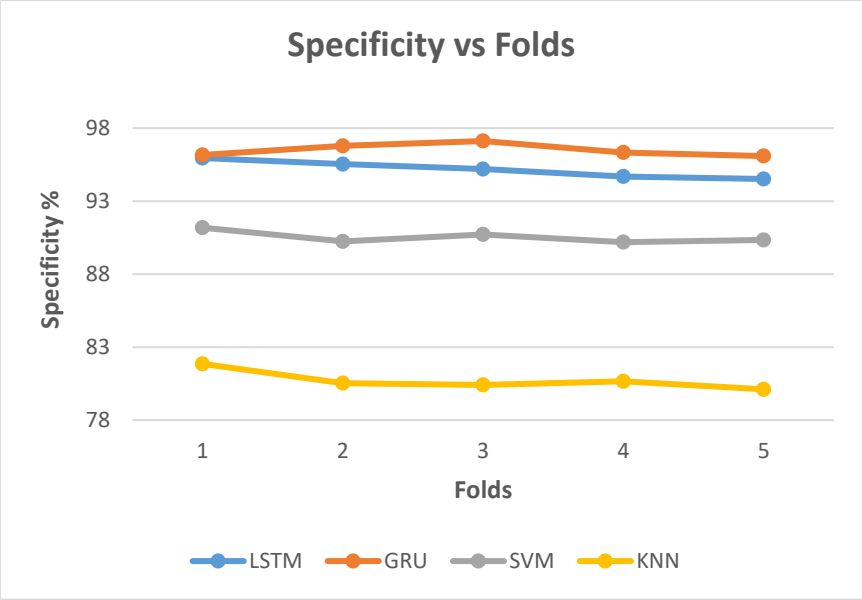

Fig. 5.6: Overall specificities over folds of LSTM, GRU, SVM, and KNN classifiers.

In **Fig. 5.8**, ROC curves show the quality of the designed classifiers. Curves of LSTM and GRU are almost overlapping with each other. The AUC value is a reference for the classifier's quality, meaning a higher area under the ROC curve is better in terms of performance. AUC values for different classifiers are presented in **Table 5.2**.

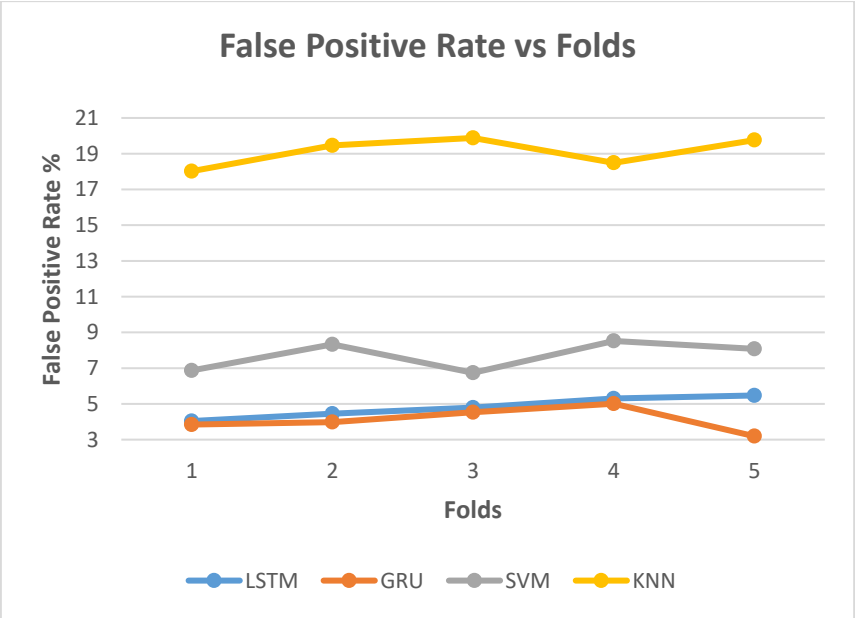

Fig. 5.7: Overall false positive rates over folds of LSTM, GRU, SVM, and KNN classifiers.

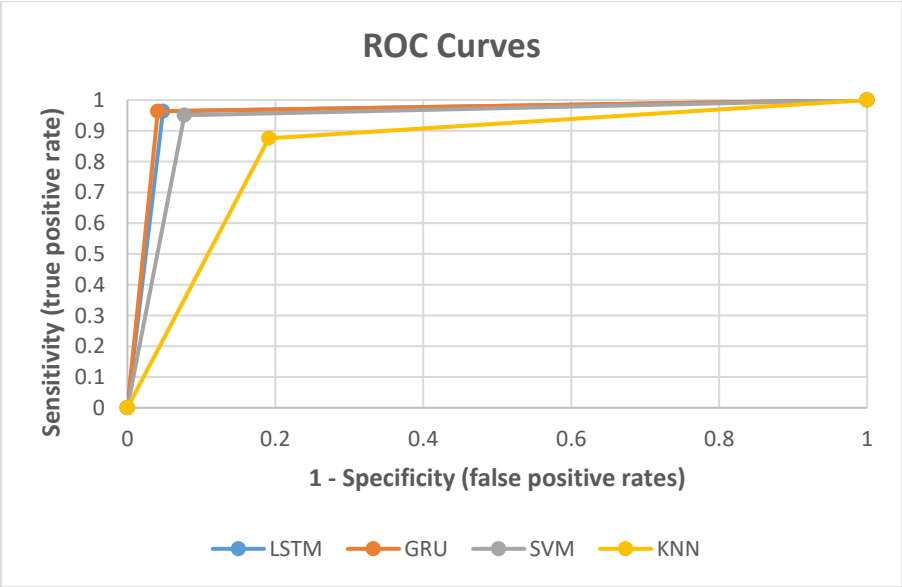

Fig. 5.8: ROC curves of LSTM, GRU, SVM, and KNN classifiers.

5.6 Discussion

This study aims to identify MCI subjects from HVs using a portable, cost-effective tool like the EEG. 27 subjects (16 HVs and 11 MCIs) participated in this study. Recorded EEG signals from those 27 subjects are preprocessed and fed into four different classifiers named LSTM, GRU, SVM, and KNN. GRU has achieved the highest accuracy of 96.91%, sensitivity of 97.95%, and specificity of 96.16%. It can be deduced from **Fig. 5.7** that GRU has the lowest false alarm rate as well, which is 4.11%. The false alarm rate

is still in a considerable range, and it is because of proper pre-processing of the recorded EEG data. This is an extensive extension of our previous GRU-based MCI detection framework [104]. To our knowledge, the LSTM and GRU models have never been used with this dataset but have shown encouraging results. Previous attempts to categorize EEG data using TML approaches have yielded promising results, but the computational cost of creating shallow structures and feature extraction algorithms renders it too expensive. Therefore, no feature extraction method has been applied, and the designed SVM and KNN reused the extracted features from the introduced LSTM network.

Table 5.3: Comparison with Previous MCI Efforts.

Studies	Accuracy %	Sensitivity %	Specificity %
Kashefpoor *et al.* [50]	88.89	**100**	83.33
Amezquita-Sanchez *et al.* [135]	90.3	92.1	87.9
Chen *et al.* [119]	92.06	Not reported	Not reported
Yang *et al.* [122]	90.37	Not reported	Not reported
Kashefpoor *et al.* [117]	88.9	83.3	**100**
Khatun *et al.* [52]	87.9	84.8	95
Al-Qazzaz *et al.* [135]	91.48	91.48	Not reported
Sharma *et al.* [75]	89.8	84	94
Sridhar and Manian [86]	91.93	Not reported	Not reported
Engedal *et al.* [154]	69	71	69
Proposed GRU Model	**96.91**	97.95	96.16

An overall comparison with previous attempts, along with our proposed method, is reported in **Table 5.3**. GRU has not only outperformed the rest of the designed classifiers but also previous MCI efforts mentioned in the table. For cross validation and consistency checking, 5-fold cross validation has been followed for each of the designated classifiers. Previous experiments that used a shallow design failed to provide a model that was both cost-effective and efficient. According to the literature, there have been relatively few

studies that have used DL-based models to classify MCIs from HVs. Most earlier attempts used typical ML methods such as SVM, KNN, and LR, which lack the ability to extract features on their own. Different feature extraction methods must be used to feed them with manually extracted features. It's one of the most common causes of missing critical characteristics that affect the classifier's performance. The cost of the model rises as a result of this extra feature extraction phase, whereas a DL-based model can accomplish it on its own. If the sample size is too small, it will almost always result in a non-efficient model.

The GRU model I propose contains two deep, hidden layers and a dense layer that is far too easy to create. It has a total of 1024, 512, and 1 neuron. The hidden layers' activation functions are "*tanh*," whereas the dense layer's activation function is "sigmoid". It maintains excellent accuracy despite its modest construction. It took 36 epochs to achieve such efficiency, with each epoch lasting just 1120 seconds on average.

5.7 Synopsis and Hereafter Work

Our proposed GRU-based MCI detection system has demonstrated its capacity to distinguish between HV and MCI subjects. Working with a reduced sample size is difficult, especially when using a DL-based model as a classifier. As a result, I had to de-noise, segment, and down-sample the raw data before I could use it. I was able to eliminate undesirable signals from the raw data using SWT, and I was able to expand the sample size using segmentation. Because I needed a simpler model that could be trained on a standard computer with a simple setup, I down sampled the segmented data. I was able to minimize the data size by using this down-sampling, but I had to sacrifice some of the data in the process. One of the key reasons for the lack of a flawless model is this. However, in down-sampling, I made every effort to avoid losing too much data, which may have resulted in an inefficient model.

Finally, I can state that our suggested GRU-based MCI detection framework is a well-balanced model with a simple architecture, low construction cost, and great performance in distinguishing HV EEG signals from MCI ones. Succeeding research should concentrate on obtaining a larger dataset and examining this GRU architecture with other hidden layer counts and activation parameters. I hope that this chapter will pave the way for oncoming EEG researchers to develop a flawless normal EEG identification model.

This was the last MCI detection study with the EEG data I have received from [50], [123]. In the next chapter, I have incorporated a multiclass EEG dataset containing 109 subjects (49 ADs, 39 MCIs, and 23 HVs) and developed a novel multiclass AD-MCI detection framework named the deep residual Alzheimer's disease and MCI detection network (DRAM-Net). It is the very first breakthrough of its kind, breaking multiclass accuracy by a long margin.

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book:	DRAM-Net: A Deep Residual Alzheimer's Diseases and Mild Cognitive Impairment Detection Network Using EEG Dat[redacted]
Surname:	ALVI
First name:	ASHIK MOSTAFA
Institute:	Institute for Sustainable Industries and Liveab
Candidate's Con[redacted]	55

Status:

Accepted and in press:	☐	Date:	
Published:	☑	Date:	25/10[redacted]

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be includ[redacted]ned in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted]	24/01/2023
Signature	**Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

Name(s) of Co-Author(s)	**Contribution (%)**	**Nature of Contribution**	**Signature**	**Date**
Ashik Mostafa Alvi	55	Conceptualization, Methodology, Data Curation, Investigation, Software,	[redacted]	24/01/2023
Siuly Siuly	15	Formal Analysis, Validation, Writing–Review and Editing, Supervision.		24/01/2023
Maria Cristina De Cola	15	Data Curation, Validation, Writing–Review and Editing.		24/01/2023
Hua Wang	15	Resources, Supervision, Writing–Review and Editing, Funding Acquisition, Project		26/01/23

[redacted]

Updated: September 2019

CHAPTER 6 : DRAM-NET: A DEEP RESIDUAL ALZHEIMER'S DISEASES AND MILD COGNITIVE IMPAIRMENT DETECTION NETWORK

6.1 Overview

Mild cognitive impairment (MCI) and Alzheimer's disease (AD) are two common neurodegenerative disorders that belong to the dementia family and are mostly found in elders. There is evidence that MCI may lead to AD. Since there is no treatment for AD after it has been diagnosed, it is a significant public health problem in the twenty-first century. Existing CML methods fail to detect AD and MCI more efficiently and accurately because of their shallow and limited architecture. Electroencephalography (EEG) is emerging as a portable, non-invasive, and cheap diagnostic tool to analyse MCI and AD, whereas other diagnostic tools like computed tomography, positron emission tomography, mini-mental state examination, and magnetic resonance imaging are expensive and time-consuming. To address these obstacles, a deep residual Alzheimer's disease and MCI detection network (DRAM-Net)-based framework has been introduced to detect MCI and AD using EEG data. This multi-class study contains EEG data collection, pre-processing (down-sampling, de-noising, and temporal segmentation), DRAM-Net architecture to classify AD, MCI, and healthy volunteers (HVs), and experiment evaluation stages. Our proposed DRAM-Net framework has obtained 96.26% overall multiclass accuracy, outperforming existing multi-class studies, and has also claimed accuracy of 96.66% for the HV class, 98.06% for the MCI class, and 97.79% for the AD class. This study will create a new pathway for succeeding neuro-disease researchers and technology experts. **Research questions #2, 3, 4, and 5** are solved inside this chapter 6.

The contents of this chapter have been published in the *Proceedings of Health Information Science: 11th International Conference, HIS 2022* [171].

6.2 Introduction

MCI and AD are linked to dementias induced by neurodegeneration. These neurological conditions are mostly caused by the loss and dysfunction of neurons in the brain's cells. Symptoms of MCI and AD include memory loss, decreased vocabulary, and a decreased ability to perform accurate motor movements [107]. Approximately 50 million individuals worldwide are expected to suffer from dementia, and 60 percent of those instances are linked to AD [156]. MCI is treated as the preliminary stage of AD [105]. People with AD or MCI have a wide variety of brain processes impaired, including memory and learning, as well as the ability to do complicated tasks like executive and motor skills and the ability to pay attention to others [155]. With age, the likelihood of experiencing it increases by a factor of 10 and often affects those over the age of 65 [62]. AD and MCI are currently untreatable, and once diagnosed with AD, patients live for 5–8 years [37], [39]. However, early detection may delay the progression of the disease and improve the quality of life for patients and their caregivers.

Existing tools including computed tomography, positron emission tomography, mini-mental state examination, and magnetic resonance imaging to detect MCI and AD are expensive, invasive, and time-consuming [106], [160]. Whereas, EEG is a newly emerging portable, low-cost, easy to understand and access, and quick tool to identify neuro-disorders like AD and MCI [104]. EEG recordings retain the electrical movements in the cerebral cortex relative to time, which are the fundamental drivers for assessing neurological disorders. The procedure for capturing EEG data means placing electrodes on the scalp according to a specific design, with the international 10–20 system being the most popular setup [155]. In account of this, I have addressed the use of EEG as a useful approach for identifying MCI and AD at an earlier stage.

In order to identify MCI early and prevent it from progressing into AD or other cognitive diseases, several studies have been conducted in the last few decades. Morabito *et al.* [157] had performed a binary (AD vs. MCI) epoch-based classification using a convolutional neural network (CNN) with 11 MCI and 4 AD subjects. Recorded EEG data were preprocessed by Power Spectral Density (PSD), and then a multi-dimensional CNN with a *softmax* classifier model was employed to complete the binary classification. This effort gained an accuracy rate of up to 98.97% (95% confidence range: 98.68%–99.26%). A recent deep learning (DL)-based approach [159] proposed two DL models:

modified CNN and convolutional auto encoder (Conv-AE) neural networks (NNs) to differentiate 61 HVs, 56 MCI, and 63 AD participants. Time–frequency representation (TFR) with continuous wavelet transform (CWT) has been used for processing the EEG data prior to feeding to the NNs. The CNN and Conv-AE NN models achieved 92% and 89% average accuracy, respectively. Ieracitano *et al.* [131] carried out an AD-MCI study with multiple machine learning (ML) methods with 63 ADs, 63 MCIs, and 63 HVs. CWT and higher order statistics (HOS) from the bispectrum (BiS) features were extracted and fed into multi-layer perceptrons (MLP), auto encoders (AE), logistic regression (LR), and support vector machines (SVM). MLP outperformed the rest of the ML [132] and [8] classifiers used in this study with an accuracy of 89.22%. Another traditional machine learning (TML)-based effort [139], [132] with 109 participants (49 AD, 37 MCI, and 23 HVs) had been introduced, where Fast Fourier Transform (FFT) and Discrete Wavelet Transform (DWT) were performed to obtain the spectrum features and de-noise the signal. A Decision Tree (DT) with the C4.5 algorithm was applied to perform the classification task, and it received 83%, 92%, and 79% accuracy while differentiating HV vs. AD, HV vs. MCI, and MCI vs. AD, respectively. Pirrone *et al.* [155] proposed a DT, SVM, and K-nearest neighbor (KNN)-based AD-MCI study with 48 AD, 37 MCI, and 20 HVs, and KNN remained at the top with 97%, 95%, 83%, and 75% accuracy (HV vs. AD, HV vs. MCI, MCI vs. AD, and HV vs. AD vs. MCI).

The studies I have reviewed have used the same multi-class EEG dataset that I have used. While analyzing these studies, I have concluded that most of the studies were conducted using TML algorithms [65], [66] and most importantly, the multi-class performance is poor (below 90%). TML algorithms have the tendency and limitation to overlook some important features of the complex EEG data because their architecture does not allow them to capture those. To resolve these problems, I have come up with a DL-based effort to not only enhance the performance but also extract those extra hidden complex features of EEG data that have significant involvement in the classifier's learning rate. I have proposed the DRAM-Net framework consisting of four stages: EEG data collection, pre-processing, DRAM-Net architecture to identify MCI, AD, and HVs, and experiment evaluation. Below is the key contribution of this proposed DRAM-Net framework:

- For the first time, I have introduced a deep residual network (DRN), customly designed for AD-MCI detection

- Our experiment uses 5 seconds temporal segments and 5 seconds of EEG data is good enough to decide the patient's condition
- This proposed DRAM-Net framework has outperformed all the existing multi-class AD-MCI studies with this EEG dataset

The rest of the chapter is organized as follows: Section 6.3 introduces the proposed DRAM-Net framework. Results and discussion are elaborated in Section 6.4. Finally, this study finishes with a oncoming study and essence in Section 6.5.

6.3 Proposed DRAM-Net Framework

This study represents the DRAM-Net framework using EEG data. **Fig. 6.1** outlines the whole framework in a nutshell. This proposed DRAM-Net framework is made up of four steps. In the EEG data collection step, I gathered the raw EEG data of 109 subjects, Pre-processing includes down-sampling, noise removal, and temporal segmentation to ensure that the EEG data is clean and ready to feed to the network, DRAM-Net architecture, where the pre-processed data is fed and classifications are performed, and finally, the quality of the proposed model is checked in the experiment evaluation step. A detailed talk about DRAM-Net is reported below.

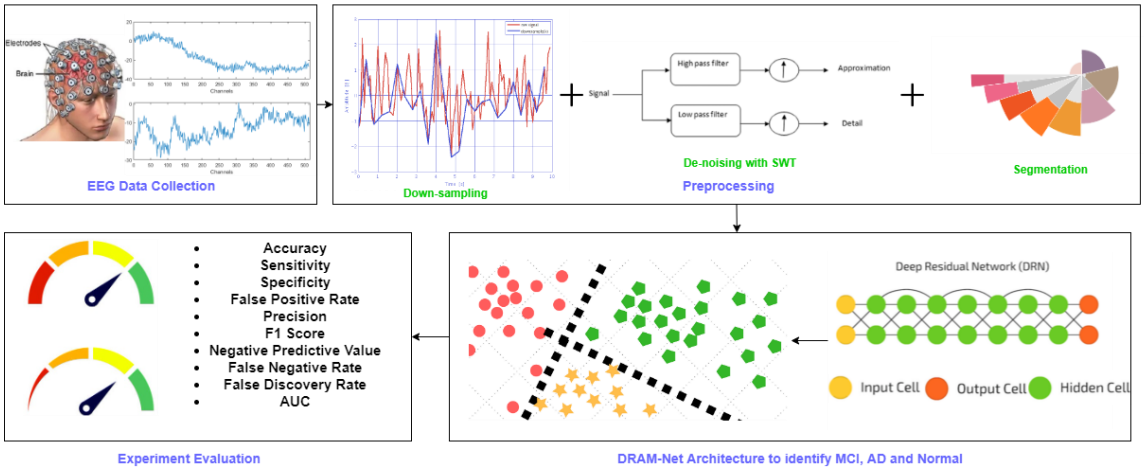

Fig. 6.1: DRAM-Net Framework for AD, MCI, and HV's detection.

6.3.1 EEG Data Collection

This multi-class study uses an EEG dataset containing 109 subjects enrolled in the IRCCS Centro Neurolesi "Bonino-Pulejo" [155], [132]. Among these 109 subjects, 49 AD, 39 MCI, and 23 HV were present. The average age of patients with AD and AD is 78.4 (6.4%) and 74.1 (9.4%), respectively, whereas the average age of HVs is 65.6 (7.4%).

The inclusion criteria were the diagnosis of AD or MCI, while the exclusion criteria were the existence of neurological or psychiatric illnesses that may cause cognitive imbalance, complicated systemic disorders, and the presence of epileptiform patterns in the EEGs, hydrocephalus, stroke, traumatic brain injuries, or other neurological abnormalities. The objective of these EEG recordings was explained to all the participants, and everyone signed the consent form. Following the World Health Organization's standard, the recording tool was placed in place and approved by the local Ethics Committee of the RCCS Centro Neurolesi "Bonino-Pulejo". All the participants were in a resting condition, keeping their eyes closed. 19 electrodes (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz) were placed across the scalp of each subject following the international 10-20 standard. The recording took place for about 300 seconds for each subject, with a sampling frequency of 256 or 1024 Hz.

6.3.2 Preprocessing

6.3.2.1 Down-sampling

The collected raw EEG data were not uniform in terms of sampling frequency. Therefore, I have made it uniform by down-sampling the sampling frequency to 256 Hz for all subjects using MATLAB.

6.3.2.2 De-noising with SWT

EEG recordings may be contaminated by a variety of factors, including electrode pops, eye blinks, outlier readings, power supply fluxes and interference (50 Hz), baseline drift, breathing patterns, and other electrical activities [9], [7], and [37]. To remove these artifacts, Stationary Wavelet Transformation (SWT) has been applied to the down-sampled EEG recordings. A detailed description of this de-nosing process can be found in [107].

6.3.2.3 Temporal Segmentation

This study starts with an EEG dataset of 109 subjects. However, just to increase the number of training and testing samples and also to create a quick and efficient model, I have segmented the dataset. As mentioned in the EEG data collection step, each recording tool records for about 300 seconds. For this work, each subject's 300 seconds of de-noised recordings were segmented into 5-second chunks. After this segmentation process, 4525

AD, 3789 MCI, and 1663 HV temporal segments have been generated. All the newly created segments are stored as comma-separated values (CSV) files using MATLAB.

6.3.3 DRAM-Net Architecture to Identify AD, MCI, and HVs

This proposed DRAM-Net framework aims to discover AD, MCI, and HVs more accurately and efficiently. To ensure our objectives, the raw collected data have been properly pre-processed. I have introduced a DRN for the first time with any EEG study and designed it specifically for AD, MCI, and HVs detection.

This classification process starts with importing all the temporal segments stored as CSV files. With the help of Python's Pandas library, these CSV files and corresponding labels are inputted. All the dataframes were converted to two-dimensional matrices using the NumPy library. Then, data are randomly divided into testing and training sets using the Scikit-Learn library. 85% of the data has been used for training and 15% for testing. Lastly, the DRAM-Net architecture has been created with the Keras module having TensorFlow as the backend.

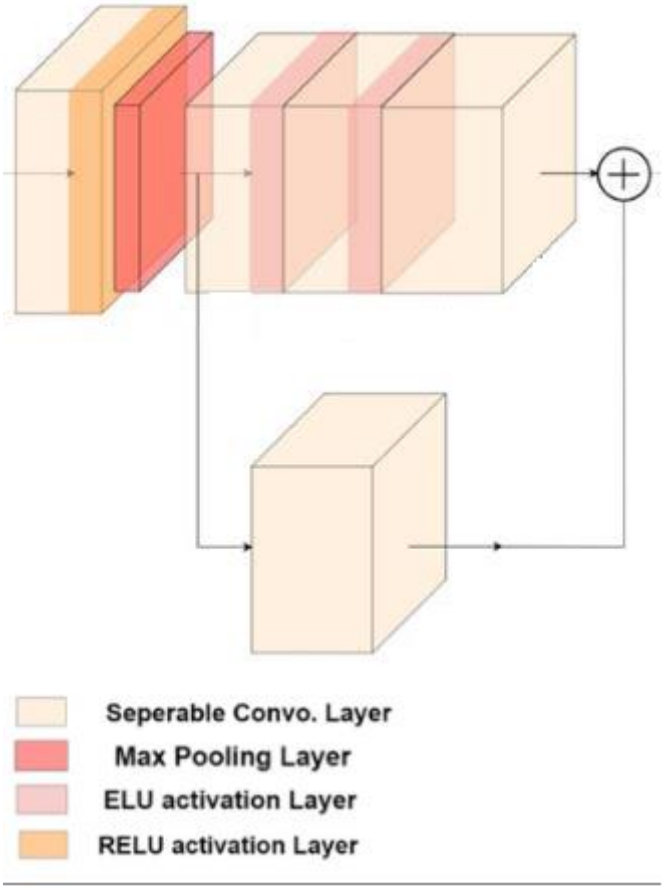

Fig. 6.2: Residual block of the proposed DRAM-Net.

DRAM-Net starts with a SeparableConv2D layer having 16 neurons. I have kept the kernel size and strides fixed for all the layers, which are (4, 4) and (1, 1), respectively. A batch normalization has been performed using the *BatchNormalization()* function in the Keras library. A *ReLU* activation function has been added after that. There are four residual blocks present in our proposed DRAM-Net architecture. **Fig. 6.2** illustrates a single residual block. I have included a *MaxPooling2D* layer before and after each of those residual blocks. Each of these four residual blocks is composed of 3 *SeparableConv2D* layers having 16, 32, 64, and 96 neurons respectively. Inside the residual block, *SeparableConv2D* layers are activated with the *ELU* function. The final convolutional layer of each residual block is combined with the shortcut connection, and then a *RELU* activation layer is applied [158]. Lastly, a global average pooling function is carried out to create a 1D vector, which is subsequently fed into a fully connected layer.

To stop over fitting, an early stopping set up has made by monitoring validation loss. *The min_delta* and *patience* parameters have been holding 0.001 and 10 values, respectively. The best model has been stored in HDF5 format. The batch size and number of epochs have been set to 32 and 150, respectively.

The training and testing process has been run on a Windows PC having 256 GB of RAM, an AMD Ryzen Threadripper PRO 3995WX 64-Core 2.70 GHz processor, and an NVIDIA RTX A6000 graphics card.

6.4 Results and Discussions

This study aims to establish a DL-based method for efficient detection of AD, MCI, and HVs using EEG data. The pre-processing steps of this proposed study have produced 9977 temporal segments (4525 AD, 3789 MCI, and 1663 HVs). This has enabled me to have an accurate and efficient model as I have enough samples to extract the deep hidden features and use them while training the DRAM-Net.

$$Negative\ Predictive\ Value = \frac{TN}{TN+FN} \times 100 \quad (20)$$

$$False\ Negative\ Rate = \frac{FN}{TP+FN} \times 100 \quad (21)$$

$$False\ Discovery\ Rate = \frac{FP}{TP+FP} \times 100 \quad (22)$$

Following (6), (7), (8), (9), (10), (11), (12), (20), (21) and (22) are the performance matrices used in experiment evaluation, where TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively.

Our proposed DRAM-Net framework took only 32 epochs to beat the existing efforts accuracy with the same EEG data set. **Fig. 6.3** shows the overall accuracy of DRAM-Net over 32 epochs. Initially, the accuracy was below 60%. As the learning rate increases, the accuracy grows. It is important to mention that the pre-processing steps have helped the DRAM-Net architecture to learn very quickly by removing the artifacts and extracting hidden but important features.

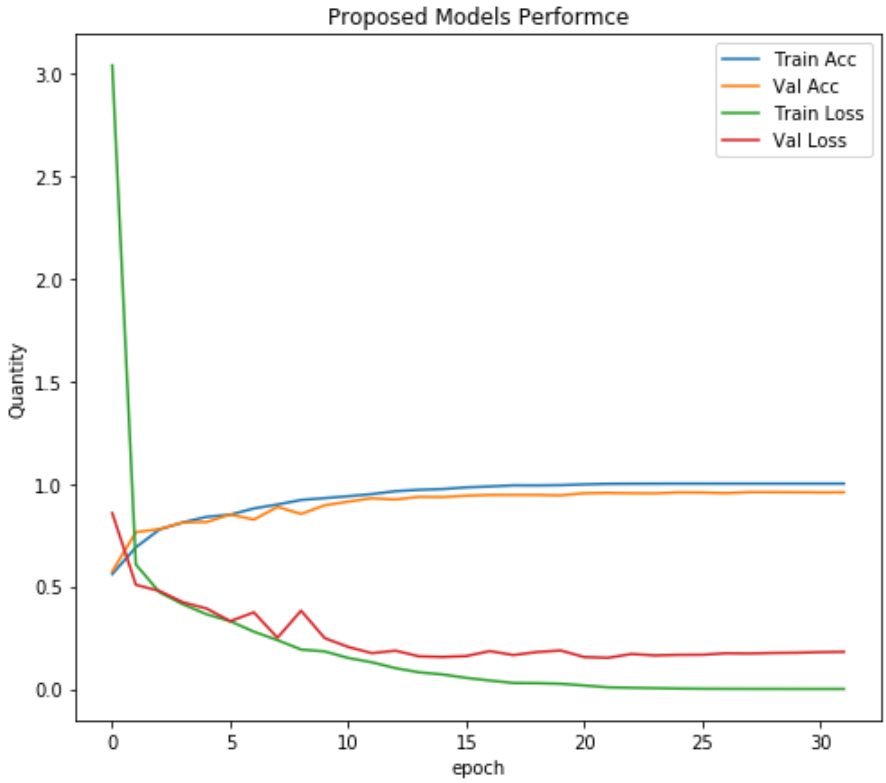

Fig. 6.3: Epoch vs Overall Accuracy of the proposed DRAM-Net.

After the splitting, there were 1497 samples in the testing set. The overall confusion matrix of the proposed DRAM-Net is portrayed in **Fig. 6.4**. Samples from 240, 537, and 664 have been correctly identified as HVs, MCIs, and ADs, respectively. And the misclassification rate is very low. A class-wise performance report has been noted in **Tables 6.1–6.3**.

While classifying the HVs, the accuracy has dropped a bit compared to the other two classes. **Table 6.1** represents the performance report for HV class. DRAM-Net

achieved 96.66% accuracy, 89.55% sensitivity, 98.21% specificity, 1.79% false positive rate (FPR), 91.60% precision, 90.57% F1 score, 97.73% negative predictive value (NPV), 10.45% false negative rate (FNR), 8.40% false discovery rate (FDR), and 93.88% area under the ROC (AUC) curve value.

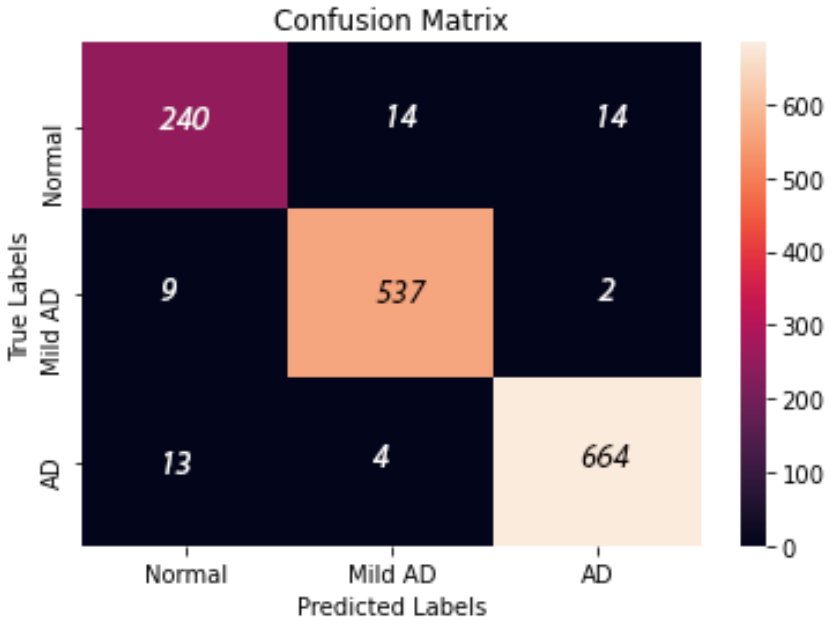

Fig. 6.4: Confusion Matrix of the proposed DRAM-Net.

Table 6.1: Classification report of HV class.

Confusion Matrix		**Performance Matrixes**	
True Positive	240	**Accuracy**	96.66%
True Negative	1207	**Sensitivity**	89.55%
False Positive	22	**Specificity**	98.21%
False Negative	28	**False Positive Rate**	1.79%
		Precision	91.60%
		F1 Score	90.57%
		Negative Predictive Value	97.73%
		False Negative Rate	10.45%
		False Discovery Rate	8.40%
		AUC	93.88%

DRAM-Net has performed so well while identifying MCI participants. The overall performance report of the MCI class is reported in **Table 6.2**. DRAM-Net gained 98.06% accuracy, 97.99% sensitivity, 98.10% specificity, 1.90% FPR, 96.76% precision, 97.37% F1 score, 98.83% NPV, 2.01% FNR, 3.24% FDR, and 98.05% AUC.

Table 6.2: Classification report of MCI class.

Confusion Matrix		**Performance Matrixes**	
True Positive	537	**Accuracy**	98.06%
True Negative	11	**Sensitivity**	97.99%
False Positive	18	**Specificity**	98.10%
False Negative	931	**False Positive Rate**	1.90%
		Precision	96.76%
		F1 Score	97.37%
		Negative Predictive Value	98.83%
		False Negative Rate	2.01%
		False Discovery Rate	3.24%
		AUC	98.05%

Table 6.3: Classification report of AD class.

Confusion Matrix		**Performance Matrixes**	
True Positive	664	**Accuracy**	97.80%
True Negative	17	**Sensitivity**	97.50%
False Positive	16	**Specificity**	98.04%
False Negative	800	**False Positive Rate**	1.96%
		Precision	97.65%
		F1 Score	97.57%
		Negative Predictive Value	97.92%
		False Negative Rate	2.50%
		False Discovery Rate	2.35%
		AUC	97.77%

Table 6.3 presents the overall classification report for the AD class. While testing AD samples after model creation, DRAM-Net has received 97.80% accuracy, 97.50% sensitivity, 98.04% specificity, 1.96% FPR, 97.65% precision, 97.57% F1 score, 97.92% NPV, 2.50% FNR, 2.35% FDR, and 97.77% AUC. The confusion matrix for the AD class is also part of **Table 6.3**, where 664 samples were correctly identified as AD.

It is possible to visualize the effectiveness of a classification system by looking at the ROC curve. **Fig. 6.5** displays all three ROC curves generated by each of the three classes. The green, blue, and yellow ROC curves represent the HV class, MCI class, and AD class, respectively.

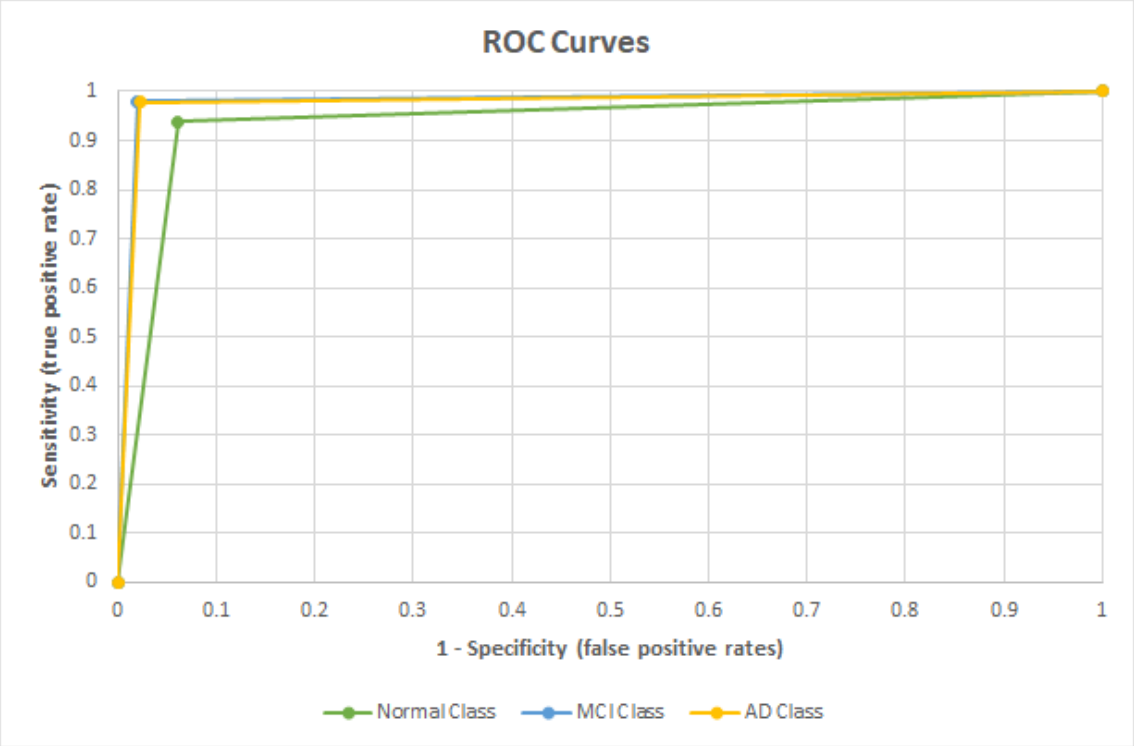

Fig. 6.5: ROC curves of HV, MCI, and AD classes.

This study has been conducted with 109 participants, and our proposed DRAM-Net has achieved satisfactory performance. From the literature, it can be seen that most of the studies performed binary classification tasks using this multi-class dataset. When it comes to the multi-class classification, all the previous efforts failed to perform well. Moreover, the binary classification performances were not satisfactory. Previous efforts used TML methods like SVM, KNN, LR, and DT to perform this complex task. But the nature and pattern of EEG are too complex, and they require extra attention. In addition, TML efforts

require extra steps to trim down the features for the classifier. And while doing this, important features got mixed up and lost.

Table 6.4: Comparison with previous AD and MCI studies with the same EEG dataset.

Efforts	Classes	Accuracy %
Morabito *et al.* [157]	AD vs MCI	98.97% (95% confidence range: 98.68%–99.26%)
Fouladi *et al.* [159]	HV vs MCI vs AD	92%
Ieracitano *et al.* [131]	HV vs MCI vs AD	89.22%
Fiscon *et al.* [132]	HV vs AD, HV vs MCI, and MCI vs AD	83%, 92%, and 79% respectively
Pirrone *et al.* [155]	HV vs AD, HV vs MCI, MCI vs AD, HV vs MCI vs AD	97%, 95%, 83%, and 75% respectively
Our Proposed DRAM-Net Framework	**HV vs MCI vs AD**	**96.26%**

DL-based methods are so advanced that they can take care of the feature extraction step by step by themselves. And connected hidden layers of DL methods allow for the extraction of hidden complex features from the EEG data. Our proposed DRAM-Net, consisting of four hidden residual blocks that also have multiple *SeparableConv2D* layers, is capable of trimming down important features from the non-stationary EEG data. A compilation of existing efforts using the same EEG dataset as our proposed DRAM-Net is reported in **Table 6.4**. Morabito *et al.* [157] performed binary classification, which performed well but did not solve the multi-class problem. Fouladi *et al.* [159], Ieracitano *et al.* [131], and Pirrone *et al.* [155] performed multi-classification and achieved 92%, 89.22%, and 75%, respectively. On the other hand, our proposed DRAM-Net has outperformed, achieving 96.66% accuracy for the HV class, 98.06% accuracy for the MCI class, 97.79% accuracy for the AD class, and 96.26% overall accuracy.

6.5 Succeeding Study and Essence

Our proposed DRAM-Net framework focuses on efficient detection of MCI, AD, and HVs using EEG data. Previous efforts have struggled to perform well while solving multi-class problems. And also, pre-processing steps are causing extra time and effort for

existing TML-based efforts. From these issues, I was motivated and developed a DL-based effort named DRAM-Net. DRAM-Net achieved 98.06%, 97.79%, and 96.66% while classifying MCI, AD, and HVs. Overall accuracy of our proposed framework is 96.26%.Oncoming studies should focus on increasing the number of subjects and the pre-processing steps to enhance the data quality and the model's learning rate. Choosing a DL method over TML algorithms can help to extract hidden complex EEG data features. This effort will guide the technology and medical experts to continue EEG research at a new level and develop new ideas and methods for neurological disorders.

Further investigations have been done with the same multi-class EEG dataset to improve the multi-class performance and computational speed. Therefore, I have worked with a custom-designed CNN model to construct the final AD-MCI identification framework. The final AD-MCI detection framework that I created is in the following chapter. Using the current multi-class EEG dataset, I have developed the ultimate solution to the AD-MCI identification problem. I have named it the Cognitive Decline Recognition Network (CDR-Net). It has reached over 99% accuracy in both 10-fold and leave-one-out cross validation.

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: CDR-Net: A Computerized Framework to Detect Alzheimer's Diseases and Mild Cognitive Impairment

Surname: ALVI First name: ASHIK MOSTAFA

Institute: Institute for Sustainable Industries and Liveab Candidate's Con 55

Status:

Accepted and in press: ☐ Date:

Published: ☐ Date:

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be includ ned in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted] 24/01/2023

Signature **Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

Name(s) of Co-Author(s)	Contribution (%)	Nature of Contribution	Signature	Date
Ashik Mostafa Alvi	55	Conceptualization, Methodology, Data Curation, Investigation, Software,	[redacted]	24/01/2023
Siuly Siuly	15	Formal Analysis, Validation, Writing–Review and Editing, Supervision.		24/01/2023
Maria Cristina De-Cola	15	Data Curation, Validation, Writing–Review and Editing.		24/01/2023
Hua Wang	15	Resources, Supervision, Writing–Review and Editing, Funding Acquisition, Project		26/01/23

Updated: September 2019

CNNCHAPTER 7 : CDR-NET: A COMPUTERIZED FRAMEWORK TO DETECT ALZHEIMER'S DISEASES AND MILD COGNITIVE IMPAIRMENT

7.1 Overview

Alzheimer's disease (AD) and mild cognitive impairment (MCI) are two dementia-related brain illnesses that are prevalent among elders in the twenty-first century. MCI is treated as the initial stage of AD, and once the illness reaches the AD stage, there is no escape from certain death. The accuracy and efficacy of current multiclass computer-based approaches to diagnose AD and MCI are constrained by traditional machine learning (TML) classifiers due to their shallow architecture. This makes it challenging to make a prompt and accurate diagnosis of AD and MCI. This research proposes a framework employing electroencephalography (EEG) to diagnose MCI, AD, and healthy volunteers (HVs) to boost multiclass performance and efficacy. EEG is a portable, non-invasive, and affordable means to identify brain problems as compared to expensive and time-consuming techniques like computed tomography (CT) scans, positron emission tomography (PET), magnetic resonance imaging (MRI), and the mini-mental state examination (MMSE). To circumvent these issues, the Cognitive Decline Recognition Network (CDR-Net) architecture has been developed to identify MCI, AD, and HVs using EEG data. The proposed architecture allows for the acquisition of EEG data, data pre-processing (down-sampling, noise cleaning, segmentation, and digital picture construction), feature extraction and classification using CDR-Net, as well as performance assessment and cross-validation stages. Our suggested CDR-Net architecture produced better multiclass accuracy, sensitivity, and specificity of 99.25%, 99.13%, and 99.32%, respectively. By using 10 folds and leave-one-out cross validations, stability, consistency, and data overfitting and underfitting concerns are addressed. This chapter will pave the way for oncoming modules that recognise several brain disorders. **Chapter 7** also resolves **research questions #2, 3, 4, and 5**.

The contents in this chapter have been submitted for publication to the *IEEE Transactions on Cognitive and Developmental Systems*.

7.2 Introduction

MCI and AD are both linked to dementias caused by neuronal damage. The death or malfunction of brain cells, called neurons, is a major contributor to these neurological disorders. Loss of memory, a reduced vernacular, and a diminished capacity for precise motor motions are all signs of MCI and AD. Recent discoveries from studies indicate that patients with MCI have a significantly increased risk of developing dementia, in particular AD [108], [109], [110]. MCI is regarded as the antecedent phase of AD [105]. With age, the likelihood of experiencing it increases by a factor of 10 and often affects those over the age of 65 [62], [30], [71]. The prevalence of MCI and AD is rising globally, with emerging nations expected to have a proportionally greater increase [112], [107]. It ranks as Australia's second-most common cause of death [113] and the seventh top death factor worldwide [114]. According to the 2018 World Alzheimer Report, 60 percent of dementia cases are attributable to AD. More than 50 million people worldwide are suffering from dementia. It is projected that there would be 152 million new cases by the year 2050, representing a more than threefold increase [156]. A patient with AD has a life expectancy of 5–8 years after being diagnosed, and MCI is also presently incurable [21], [39]. Nonetheless, the disease's course may be slowed and the quality of life for patients can be enhanced by caretakers if the diagnosis is made early.

To diagnose MCI and AD, there are currently costly, invasive, and time-consuming methods available, including magnetic resonance imaging (MRI), computed tomography (CT)), positron emission tomography (PET), and the mini mental state examination (MMSE) [106], [160]. In contrast, EEG is a recently developed, portable, affordable, simple to use and comprehend, and rapid technology to diagnose neurological illnesses, including AD and MCI. The relative timing of brain waves in the cerebral cortex is preserved by EEG recordings, which is the key to determining the severity of neurological illnesses. The technique for collecting EEG data entails putting electrodes on the scalp in a specified pattern, with the international 10-20 approach being the most widely used configuration [99]. In light of this, electroencephalography's potential as a tool for early diagnosis of MCI and AD has been emphasised.

Numerous investigations were conducted over the past few decades in an effort to diagnose MCI at an early stage and prevent its progression to AD or another cognitive disorder. With 11 MCI and 4 AD participants, Morabito *et al.* [157] chose a convolutional

neural network (CNN) to accomplish a binary (AD vs. MCI) epoch-based classification. Power Spectral Density (PSD) was utilised to pre-process the acquired EEG data, and a multi-dimensional CNN with a softmax classifier model was then employed to conclude the binary classification. This attempt had an accuracy rate of up to 98.97% (95% confidence interval: 98.68%–99.26%). Two deep learning (DL) models—modified CNN and convolutional auto encoder (Conv-AE) neural networks (NNs)—were utilised in a recent study [159] to differentiate between 61 HVs, 56 persons with MCI, and 63 people with AD. The EEG data were processed using time-frequency representation (TFR) and continuous wavelet transform (CWT) before being submitted to the NNs. Both the CNN and Conv-AE NN models had average accuracy of 92% and 89%, respectively.

Ieracitano *et al.* [131] undertook an AD-MCI investigation using several machine learning (ML) techniques using 63 ADs, 63 MCIs, and 63 HVs. CWT and higher order statistics (HOS) from the bispectrum (BiS) features were retrieved and loaded into multi-layer perceptrons (MLP), auto encoders (AE), logistic regression (LR), and support vector machines (SVM) classifiers. MLP fared better than the other ML classifiers in this investigation, surpassing them with an accuracy of 89.22%. Fast Fourier Transform (FFT) and Discrete Wavelet Transform (DWT) were implemented to acquire the spectrum characteristics and de-noise the data in another TML-based attempt [32] with 109 participants (49 ADs, 37 MCIs, and 23 HVs). The categorisation challenge was handled by a decision tree (DT) using the C4.5 algorithm, which distinguished between HV and AD, HV and MCI, and MCI and AD with 83%, 92%, and 79% accuracy, respectively. Another recent TML effort has been reported with 48 ADs, 37 MCIs, and 20 HVs, where investigators used PSD, finite impulse response (FIR), and 2nd order Butterworth filters for EEG data pre-processing. The classification task was carried out using the classifiers K-nearest neighbor (KNN), DT, and SVM. KNN outperformed the rest of the two classifiers by gaining 97%, 95%, 83%, and 75% while performing HV vs. AD, HV vs. MCI, MCI vs. AD, and HV vs. MCI vs. AD classifications.

To provide a binary identification system utilising a cubic-SVM, Puri *et al.* [172] laboured with 12 ADs and 11 HVs and reached a classification accuracy of 98.5%. The unprocessed EEG signals were divided into sub-bands by applying low-complexity orthogonal wavelet filter banks with vanishing moments. The Kruskal-Wallis test was adopted to extract and examine the two characteristics, Higuchi's fractal dimension (HFD) and Katz's fractal dimension (KFD), from EEG sub-bands. A new study [173] employing

MLP with 6 MCIs, 11 ADS, and 9 HVs was suggested and achieved an 88% F1 score. Using autoreject and independent component analysis (ICA) techniques, the investigators de-noised the EEG data. The last step was to feed MLP with the extracted power, entropy, and complexity attributes. Poil *et al.* [70] presented a study with 86 subjects (25 ADs and 61 MCIs) to forecast the development of AD at the MCI stage. The authors extracted the features and passed them to the LR classifier for classification using the ICA and Hilbert transform (HT). This binary categorisation study achieved 88% sensitivity and 82% specificity.

Another binary classification study [86] based on DL with 28 ADs and 7 MCI participants was reported. A bidirectional long short-term memory (BLSTM) classifier was used after principal component analysis (PCA) was used to identify the features. With those aged 40 to 60, it had increased by 91.93%, and with those beyond 60, by 65.73% accuracy. Amezquita-Sanchez *et al.* [13] proposed an Enhanced Probabilistic Neural Network after studying 37 MCI and 37 AD participants (EPNN). ANOVA, Hurst Exponent (HE), Fractal Dimension (FD), and the Music-Empirical Wavelet Transformation (EWT) were all investigated as ways to extract characteristics. To compare the classification outcomes, the DT, Naive Bayes (NB), and KNN were also applied. The EPNN outperformed the other used classifiers, with an accuracy of 90.3%. A limited sample size of four ADs, four HVs, and four MCIs was adopted by Bi and Wang [18]. The discriminative convolutional high-order Boltzmann machine (DCssCDBM) classifier received the features after being liberated by the FFT. 95.04 percent accuracy was achieved in this investigation.

From the reviewed articles, it is clear that none of the studies are able to break through multiclass AD-MCI detection with high accuracy and efficiency. Most of the reported efforts struggled to have a satisfactory sample size. Another significant finding is that effective research either identified standard ML approaches as a poor solution for multiclass issues or employed DL methods to address binary classification problems. And here comes one of the research problems that I intend to resolve: the studies that performed better and used TML methods utilised feature extraction and selection methods for those ML classifiers. This extra step is always required when it comes to TML methods like SVM, LR, DT, KNN, and so on. Moreover, TML classifiers are limited and old-fashioned due to their shallow architecture and often fail to work with complex structured data like EEG. As a result, they were unable to extract and link deep data layers'

crucial hidden properties. One of our objectives is to reduce this computational expense by picking a suitable DL classifier that will not ask for any feature extraction or selection method. DL methods have passed the test of working with EEG data with flying colours [104], [107], [105], and [171]. Though the performance of previous multiclass AD-MCI studies is really poor, Therefore, our goal is to set a high-performance standard for AD-MCI detection using EEG data.

In the past, I have explored long- and short-term memory (LSTM) [105], gated recurrent unit (GRU) [107], and deep residual networks (DRN) [171] using EEG data to investigate for these sorts of brain illnesses. This study incorporates multiple well-known CNN models like AlexNet [178], InceptionNet [179], ResNet50 [180], RSISC-16 Net [189], and VGG16 [181]. After evaluating the performance and being inspired by the computational disadvantage, I am motivated to design a custom CNN model.

To address these problems, I offer the CDR-Net framework, which uses a specially created CNN model to conduct AD, MCI, and HV identification using EEG data. Both AD and MCI share the same symptoms of cognitive decline. Inspired by this, I named our framework the Cognitive Decline Recognition Network. It is well known that the CNN model works better with images, so the EEG data I collect is processed and turned into images before being sent to the classifier. CNN [190] is a compact classifier that uses less memory while processing pictures than other DL techniques. Moreover, by itself, it is capable of extracting and choosing significant and in-depth characteristics. These considerations led to CNN being selected as the classifier for our CDR-Net framework. With the pre-processing steps, CNN parameters, and layers I have proposed, this is a whole new, inexpensive framework in terms of our nobility.

Our recommended CDR-Net structure contains four phases, the first of which is the primary EEG data collection phase. I was successful in gathering EEG data from 109 participants (23 HVs, 37 MCIs, and 49 ADs). Then, because there weren't many EEG signals collected at a sampling frequency of 1024 Hz, the raw signals that were collected were down-sampled to 256 Hz to keep things consistent. To eradicate artifacts and noises from the raw signals, the stationary wavelet transform (SWT) has been performed. SWT is widely renowned for its ability to handle both high- and low-frequency noises. Afterwards, cleansed signals are segmented into 5-second frames in order to increase the sample size and acquire key information quickly. Before sending the segmented frames

to the classifier, they are finally converted to 8-bit colour pictures. A specially created multi-layer CNN model with a *softmax* classifier completes the CDR-Net architecture. Different performance measures have been used to evaluate this proposed CDR-Net system, which can identify multiple classes of cognitive abnormalities. To verify the consistency and stability of our presented CDR-Net architecture, the 10-fold and leave-one-out cross validations (LOOCV) were also carried out. An overview of this chapter's achievements can be seen below:

- Developed and underpinned a modern, precise, reliable, and effective CDR-Net framework for detecting AD, MCI, and HV using EEG data.
- Increased the accuracy of multi-class classification compared to earlier methods on same and different EEG datasets.
- Examined the consistency and stability of this proposed CDR-Net by performing 10-fold and LOOCV cross validations.
- Conducted ablation tests to discover the best-suited CNN classifier, which is the heart of the CDR-Net framework.

The remaining sections of this study are arranged as follows: Section 7.3 talks about the proposed CDR-Net framework in detail. Afterwards, Section 7.4 reports on the experiments and findings. Following that, the discussion of this research is included in Section 7.5. Finally, Section 7.6 of this chapter provides an essence and a strategy for the hereafter.

7.3 Proposed CDR-Net Framework

This proposed CDR-Net framework has been built for the detection of multiple brain diseases using EEG data. **Fig. 7.1** showcases the proposed CDR-Net framework. It has four interconnected stages towards EEG data classification. The CDR-Net framework begins with EEG data acquisition. Then, collected EEG signals are pre-processed before feeding into the proposed CDR-Net model. CDR-Net consists of multiple convolutional and max pooling layers for feature extraction. There is a fully connected layer for the classification purpose from the selected features. Finally, the performance of the proposed method has been evaluated by multiple ranking matrices and cross-validation methods like LOOCV and 10-fold cross-validation. A comprehensive description of the four stages along with their sub-stages is reported below:

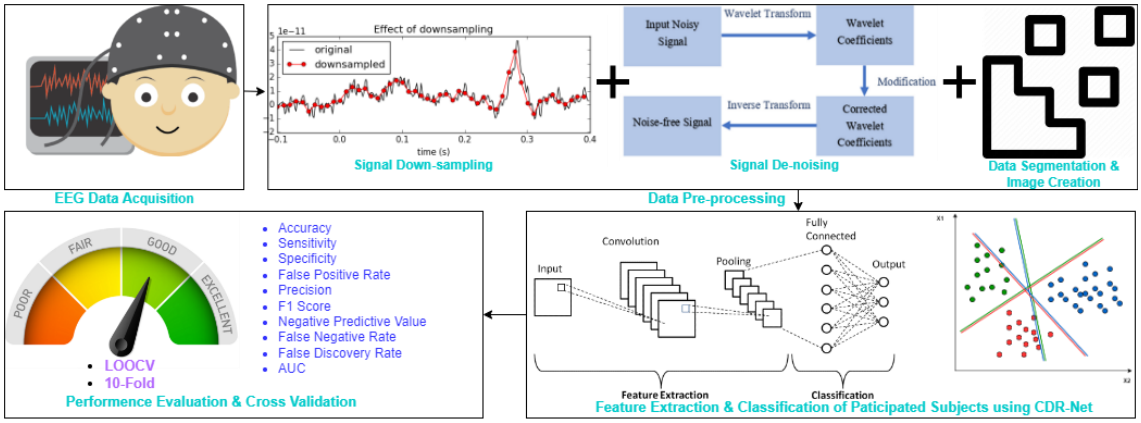

Fig. 7.1: CDR-Net Framework.

7.3.1 EEG Data Acquisition

The CDR-Net framework has been tested and trained with a multiclass EEG dataset holding 109 subjects' EEG recordings. Out of 109 subjects, 23 HVs, 39 MCIs, and 49 AD subjects participated in this data collection at the IRCCS Centro Neurolesi "Bonino-Pulejo" [155], [32]. The ethical committee of IRCCS Centro Neurolesi "Bonino-Pulejo" authorised this data collection (reference number: 40/2013), and all the participants had given their consent by signing the consent form.

This multiclass EEG data acquisition was performed using 19 electrodes (Fp1, F3, C3, P3, O1, F7, T3, T5, Fz, Cz, Pz, Fp2, F4, C4, P4, O2, F8, T4, T6) placed across the scalp following the international 10-20 system's guidelines. All the subjects were advised to keep their eyes closed and remain rested until the data collection is done. Collected brain signals were evaluated with reference to electrical potentials (μV). Each recording lasted for 5 minutes (300 seconds), keeping the sampling frequency at either 256 Hz or 1024 Hz. This uneven sampling frequency has been fixed in the pre-processing step.

Following the World Health Organization's standard, all the participating subjects are classified as AD, MCI, or HV. Any subjects with a history of head trauma, drug abuse, serious medical conditions, or other forms of dementia were excluded from this study. The average ages of patients diagnosed with AD and MCI are 78.4 ± 6.4 and 74.1 ± 9.4 years, respectively, while the average age of HVs is 65.6 ± 7.9 years. **Table 7.1** represents the demographic data of the participants.

Table 7.1: Demographic Data of Participated Subjects.

Classes	Number of Subjects (%)			Mean age (standard deviation) in years		
	Male	Female	Total	Male	Female	Total
MCI	17(46%)	20(54%)	37	75.7±9.7	72.7±9.1	74.1 ± 9.4
AD	20(41%)	29(59%)	49	78.6±4.1	78.2±7.6	78.4 ± 6.4
HV	13(56%)	10(44%)	23	68.1±6.9	62.3±8.3	65.6 ± 7.9
Total	50(46%)	59(54%)	109	74.9±8.2	73.6±9.9	74.2 ± 9.1

7.3.2 Data Preprocessing

Collected raw signals were digitised and converted to MATLAB files, which were then prepared for further processing. There are three pre-processing techniques applied. Initially, the recorded raw signals are down-sampled to 256 Hz. Down-sampled signals were de-noised using the SWT method. Cleaned signals are segmented into 5-second chunks for better learning and time savings. Lastly, temporal segments are converted to Joint Photographic Experts Group (JPG) images. An elaborate description of these pre-processing steps is reported below:

7.3.2.1 Signal Down-sampling

EEG data collection took place over multiple days, and this is one of the big EEG primary data sets with AD, MCI, and HVs. A total of 109 subjects participated in this EEG study, and 256 Hz and 1024 Hz sampling frequencies were used for collecting their brain potentials.

As there were different sensors used, the sampling frequency was not consistent. To bring consistency, I have down-sampled the 1024 Hz sampling frequencies to 256 Hz using MATLAB. From the literature, it can be said that 256 Hz is the standard sampling frequency [32], [171], [159], [131]. Now, all 109 subjects' EEG recordings have the same sampling frequency, which is 256 Hz.

7.3.2.2 Signal De-noising

EEG data collection is often affected by some external or internal factors known as artifacts or noises, which may corrupt the actual data. Eye blinks, muscle movement, electrode pops, breathing patterns, power supply fluxes, interference (50 Hz), outlier readings, and other electrical activities are some of the internal and external sources of noise. Actual signal gets contaminated with these and often misleads to a different essence if they are not removed from the actual signal [9]. Therefore, it is a must to discard the noise before further processing.

Our proposed CDR-Nett employs CDR-Net to dispel artifacts and other undesirable signals. SWT is suitable for trading with both low- and high-frequency unwanted signals, keeping the flat response at the highest position of the passing signal. Previous efforts have used SWT, DWT, CWT, etc. frequency domain noise-removing methods to clean the input signal. Apart from that, the most valuable frequency bands that carry important features are from 0.5 Hz to 32 Hz [170].

An 8th-order Symlet wavelet "*sym9*" filter has been chosen for exterminating DC components of the input signal (0–0.5 Hz) and the baseline electric drift in each of the electrode channels. On the other hand, high-frequency artifacts (32-128 Hz) along with the decomposition of stationary wavelets have been performed with second-order estimation for each of the 19 channels.

After the de-noising phase, 109 cleaned signals are digitised, representing electrical potentials (μV). Each signal has a length of 60 seconds × 5 minutes × 256 Hz and is stored as a MATLAB (.mat) file. That gives me approximately 76800 rows representing voltage amplitude (μV) and 19 columns referring to 19 electrode channels for each subject. This phase of the proposed CDR-Net has been performed in MATLAB, and a visual comparison of the P4 electrode position of (a) the raw EEG signal of an AD subject (3841 to 4240 seconds), (b) the de-noised signal of the same AD subject (3841 to 4240 seconds), (c) the raw EEG signal of an MCI subject (21761 to 22160 seconds), (d) de-noised signal of the same MCI subject (21761 to 22160 seconds), (e) the raw EEG signal of a HV (75521 to 75920 seconds), (f) de-noised signal of the same HV (75521 to 75920 seconds) are shown in **Fig. 7.2**. It is a 400-second window, which has remained constant in all 6 sub-plots in **Fig. 7.2**.

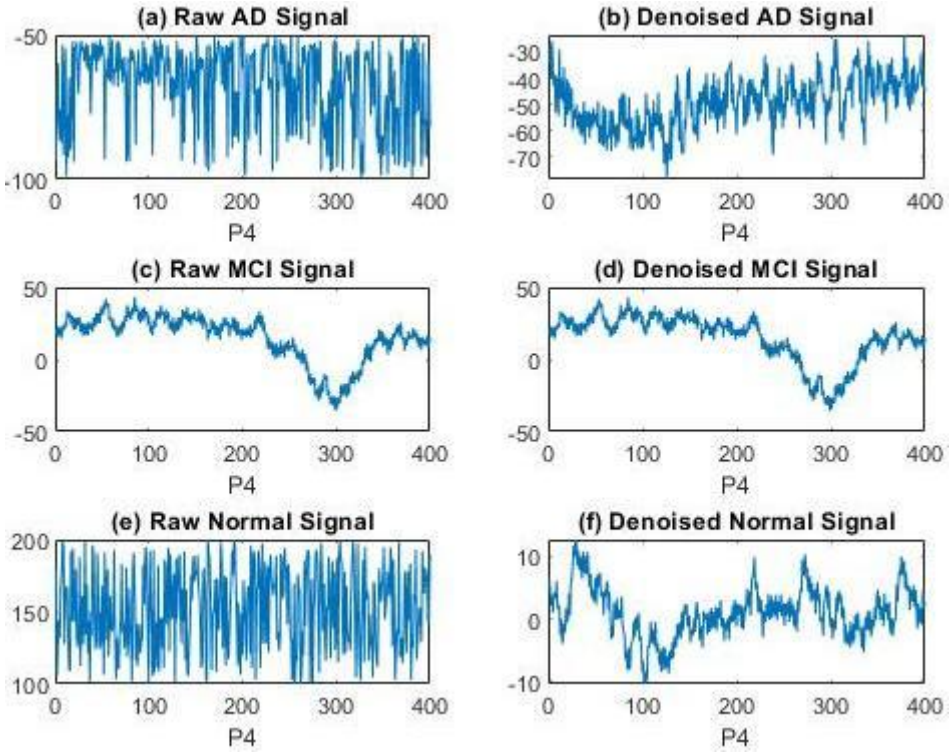

Fig. 7.2: Visual comparison of P4 electrode position of (a) raw EEG signal of an AD subject, (b) denoised signal of the same AD subject, (c) raw EEG signal of an MCI subject, (d) denoised signal of the same MCI subject, (e) raw EEG signal of a HV, (f) denoised signal of the same HV.

7.3.2.3 Data Segmentation and Image Creation

Segmentation is a technique that increases the sample size and reduces the computation expenses without losing any important features or information. As shown in our earlier works, executing segmentation has a positive impact on the entire computing cost, which includes feature extraction, learning, model construction, classification, and testing [105], [171]. It is well known that EEG data are a bit different than electrocardiogram (ECG) or magnetoencephalography (MEG) data. EEGs are complex, non-periodic, non-stationary, and huge in terms of size compared to ECGs and MEGs. Moreover, managing enough medical data is a huge task. To address such a challenge, multiple studies [104], [107], [105], and [171] used data segmentation to increase sample size and deal with such complex EEG data.

In this proposed CDR-Net, EEG recordings are segmented into short time windows and tagged with the same label as the native signal. As noted in the section on EEG data acquisition, each recording lasted around 300 seconds. Investigations on 10, 5, and 3 second segments show that 5 second segments are the best choice for keeping

important features intact and easing with time. In the context of this study, the 300-second de-noised recordings of each participant are cut up into 5-second pieces.

The final tally of segments produced by this technique is 4525 for AD, 3789 for MCI, and 1663 for HV. Each newly created temporal segment includes 1280 rows (5 seconds × 256 Hz) and 19 columns (19 channels). Using MATLAB's *imwrite* function, all the temporal segments are converted to 8-bit color images (.JPG) for better feature picking by the CNN classifier. In summary, the sample size for this study has increased from 109 to 9977. Furthermore, these small temporal segment images help the classifier select and extract important features easily.

7.3.3 Feature Extraction and Classification of Participated Subjects using CDR-Net

The fundamental objective of this multiclass work is to discriminate between persons with MCI and AD and HVs with an accuracy and efficiency that are sufficient when using non-stationary data such as EEG. Having said that, I have investigated multiple CNN models for this multiclass EEG dataset. CNN has grown its effective range from regular color or grayscale imaging to medical imaging like computed tomography images, x-ray images, spectrogram images, and magnetic resonance images. EEG data to identify brain diseases like MCI and AD is yet to be unraveled.

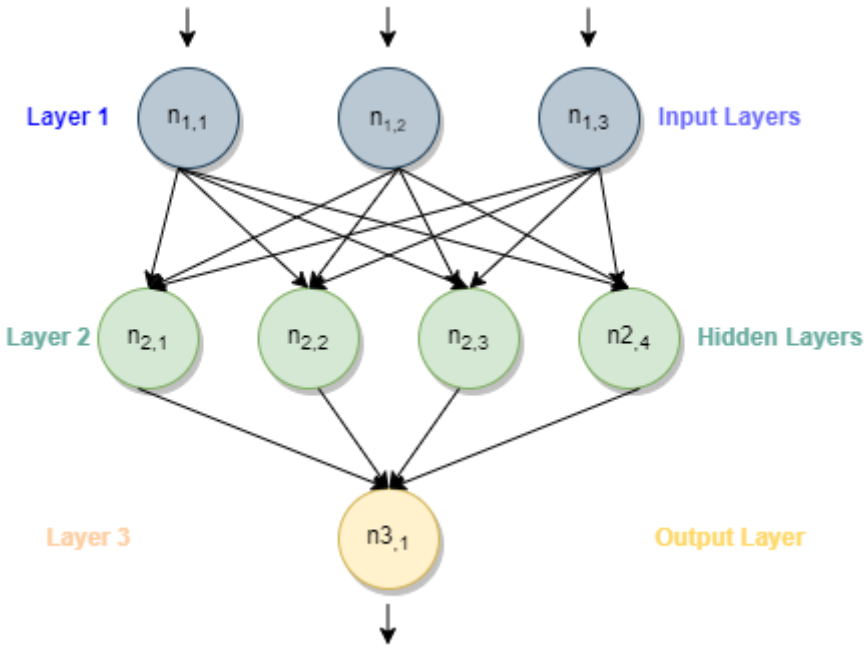

Fig. 7.3: The structure of a classical ANN.

CNN is a subtype of DL, which is an enhanced form of artificial neural networks (ANNs). A classical ANN looks like **Fig. 7.3**, where it consists of 3 layers (input, hidden, and output layers). Inspired by the human brain's neurons, ANN is a collection of connected neurons, which I call nodes in DL. The scaled total provided by the neurons in the preceding layer is applied to the output of the neurons. The weights and biases of the prior layers in the network structure are utilized to determine the final outcome choice of the CNN model, much as the ANN [174]. Therefore, considering (23) and (24), the biases and weights are adjusted for every layer where W, B, u, L, s, C, m, λ, and t stands for weight, bias, upgrading step, learning rate, strata number, cost function, momentum, regularization parameter, and training samples in total, respectively.

$$\Delta W_s(u+1) = -\frac{L\lambda}{r}W_s - \frac{L}{t}\frac{\partial C}{\partial W_s} + m\Delta W_s(u) \tag{23}$$

$$\Delta B_s(u+1) = -\frac{L}{t}\frac{\partial C}{\partial B_s} + m\Delta B_s(u) \tag{24}$$

CNN is made up of three different types of layers: convolutional layers, pooling layers, and completely linked layers [143]. The feature extraction process makes use of convolutional and max-pooling layers. Max-pooling layers are intended for feature selection, while convolution layers are aimed at feature recognition. When an image doesn't need all the high-resolution information or when a reduced output with CNN-extracted areas is required after a down-sampling procedure on the input data, max-pooling layers are used [175]. The fully connected layers receive the outputs from the convolution and pooling layers and utilize them to categorize the input.

Our proposed CDR-Net employs a custom CNN model that consists of four feature extraction blocks. A convolution layer, a *relu* activation function, and a max-pooling layer make up each feature extraction block. There are 32 filers in each convolution layer, and the kernel size is 3 × 3, while the pool size of the max-pooling layer is 2 × 1. There are two 25% dropout layers present after the 2nd and 4th feature extraction blocks. Prior to sending the extracted features to the fully linked layer, there is a 50% dropout layer. Our custom-built CNN model incorporates the *softmax* classifier, the *adam* optimizer, and *categorical cross_entropy* as the loss function. **Fig. 7.4** depicts this custom-built CNN in a visual manner, along with the output shapes of each layer.

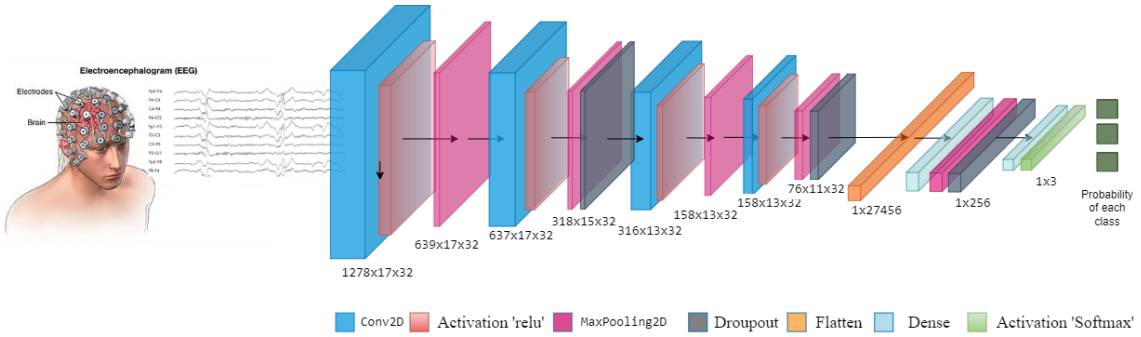

Fig. 7.4: The proposed configuration of CNN classifier.

To make sure the classifier does not overfit and has the best model, I have brought early stopping into play. It has kept an eye on the validation loss and accuracy. The *min_delta* has been set to 0.001 and the patience to 15. **Fig. 7.5** represents CDR-Net framework's hidden layers (conv2d, max_pooling, input, output, activation, dropout, flatten, dense, etc.) along with their shapes.

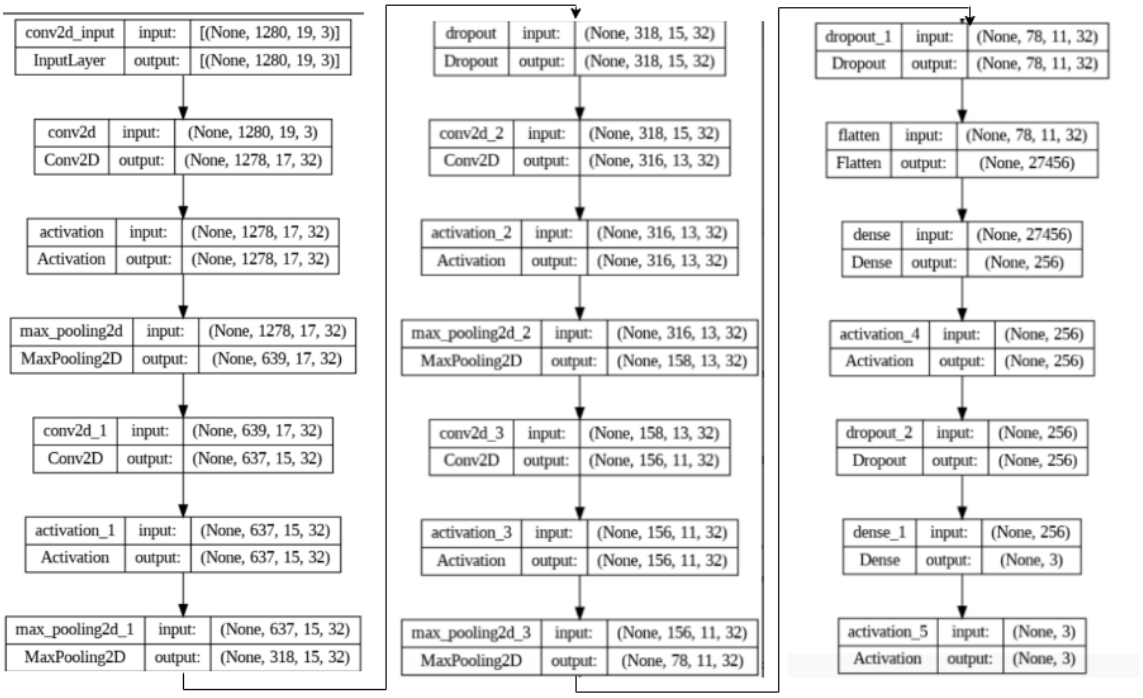

Fig. 7.5: Proposed CDR-Net framework's hidden layers along with their shapes.

7.3.4 Performance Evaluation and Cross Validation

The fundamental objective of this multiclass work is to discriminate between persons with MCI and AD and HVs with an accuracy that is sufficient by using non-stationary data such as EEG. In light of this, I looked at more than twenty distinct CNN setups and reported the top five configurations and their performances in **Table 7.5**.

To make sure the best model gets reported, multiple performance matrices are in place. All of the explored setups are verified through (6), (7), (8), (9), and the receiver operating characteristic (ROC) graph. The ROC curve represents the relationship between the true positive rate (sensitivity) and the false positive rate (1 - specificity) on a graph. Often, the effectiveness of the classifier is shown by the ROC curve. The area under the ROC curve (AUC) is a statistical measure of how well the curve predicts the true outcome. The AUC value runs between 0 and 1. An AUC value around 1 and a standard deviation below 0 indicate that the proposed method is performing optimally [84].

TP, TN, FP, and FN are the four parameters that have been generated for each of the three classes. A sample confusion matrix of how the TP, TN, FP, FN are calculated for class MCI is shown in **Table 7.2**. The green colored cell is representing TP (number of MCI samples correctly classified as MCI), the purple colored cells are standing for TN (number of non-MCI samples classified as either AD or HV), the light blue colored cells are referring to the FP value (number of AD or HV samples incorrectly classified as MCI), and the yellow colored cells are offering the FN value (number of MCI samples incorrectly classified as either AD or HV) for MCI class. The confusion matrices have multiple cells for TN, FP, and FN. Therefore, all the matching cells need to be summed together in order to calculate TN, FP, and FN. Similarly, for the AD and HV classes, the same calculations can be applied to obtain TP, TN, FP, and FN.

Table 7.2: Sample Confusion Matrices for Three Classes of Classification While Classifying MCI.

		Predicted Class			
		HV	**MCI**	**AD**	
Actual Class	**HV**	TP_{HV}	$I_{HV, MCI}$	$I_{HV, AD}$	
	MCI	$I_{MCI, HV}$	TP_{MCI}	$I_{MCI, AD}$	**FN**
	AD	$I_{AD, HV}$	$I_{AD, MCI}$	TP_{AD}	**TN**
	*I = Incorrect Classification		**FP**	**TP**	

Additionally, 10-fold and LOOCV are in place to test the performance consistency and stability of the proposed CDR-Net framework. Both approaches guarantee that the classifier is impartial and does not become too fitted to the data during the training process. While 10-fold cross validation is in progress, the entire preprocessed data are split into 10 chunks. During each fold, 9 split slices are utilized for training the model,

and one slice is kept for testing the model. Each of the data segments is examined once by repeating this method 10 times. The statistics of this 10-fold cross-validation confirm that the model is unbiased and consistent in terms of categorizing AD, MCI, and HV.

Lastly, LOOCV has been carried out to further guarantee the stability and absence of overfitting of the planned CDR-Net framework. As stated in the sub-sub-section 7.3.2.3, each of the subject's EEG recordings is segmented into 5-second chunks and converted to 8-bit color images. Now, all the segments (color images) from a subject are left out from training the model and used for testing the trained model to predict the left-out subject [176]. All of the dataset's individuals go through this procedure once. Despite being a computationally costly approach, it validates the stability, constancy, and impartiality of the suggested CDR-Net structure.

7.4 Experiments and Outcomes

The goal of this study is to develop an accurate model for detecting cognitive abnormalities from EEG data. For this process, CNN has been chosen because it is lightweight when it comes to images. This makes it possible for me to do the study in a quick manner. The setup and findings of the investigations are covered in this section.

7.4.1 Experimental Setup and Tools

The CDR-Net structure is constructed and tested primarily using only two tools. Utilizing MATLAB R2021B, all pre-processing procedures are accomplished, including signal down-sampling, de-noising, segmentation, and the production of 8-bit colour images. The Python environment is used by Jupyter Notebook to carry out the feature extraction, classification, performance assessment, and cross validation phases. A Windows computer with 256 GB of RAM, an AMD Ryzen Threadripper PRO 3995WX 64-Core 2.70 GHz CPU, and an NVIDIA RTX A6000 graphics card were employed for all the investigations.

7.4.2 Findings

Our proposed CDR-Net framework has proven its capability of distinguishing MCI, AD, and HV quite accurately and efficiently. In this part, I have reported class, fold, and batch size (BS)-wise performance for each of the three classes. To examine the performance pattern and choose the optimal BS, several BSs were employed. For each of the three classes, the outcomes of four BSs—32, 64, 125, and 256—are presented and contrasted.

In **Fig. 7.6**, the classification report for the HV class has been visualized. This figure demonstrates that, when data overfitting and underfitting are taken into consideration, a BS of 128 is the optimal choice. The proposed CDR-Net has received the highest average accuracy of 99.59%, sensitivity of 99.49%, specificity of 99.67%, and FPR of 0.33% when the batch is set to 128. The second fold of BS 128 has performed better compared to the rest of the folds. The average accuracy, sensitivity, specificity, and FPR of 10 folds with a 32-BS are 95.43%, 95.38%, 95.47%, and 4.53%, respectively. When compared to the other folds of this kind, the 6th fold of BS 32 has had the worst performance. On the other hand, the 10-fold average accuracy, sensitivity, specificity, and FPR values of BS 64 are 97.18%, 97.39%, 97.02%, and 2.98%, respectively, and those of BS 256 are 97.87%, 97.14%, 98.47%, and 1.53%, respectively. The BSs of 64 and 256 almost achieved similar performance, but 256 performed better than 64 in terms of accuracy, specificity, and FPR. It can also be seen that BS 32 did not perform well compared to the other 3 BSs, and it is due to data underfitting. The sample size has been raised from 109 to 9977 since the preprocessing phase. And this action has enabled the classifier to capture features from a larger BS. However, due to the overfitting issue, the performance of the proposed CDR-Net starts to decrease when the BS goes beyond 128.

The classification report for the MCI class has been illustrated in **Fig. 7.7**. The CDR-Net has been successful in discriminating MCI from other samples 98.92% of the time on average when the BS is set to 128. The average performance of CDR-Net while classifying MCI is: 98.92% accuracy, 97.58% sensitivity, 99.18% specificity, and 0.82% FPR, keeping the BS at 128. The maximum accuracy of this setup, over 99%, has been obtained by the 5th and 8th folds. The 10-fold average performance has decreased when the BS is kept at 64, giving an accuracy of 96.30%, a sensitivity of 90.28%, a specificity of 97.50%, and a FPR of 2.50%. The proposed framework has improved the 10-fold average performance when the BS is increased to 256, giving accuracy of 97.24%, sensitivity of 92.42%, specificity of 98.21%, and FPR of 1.79%. The poorest 10-fold average performance has been reported when the BS is set to 32, returning accuracy of 94.49%, sensitivity of 84.45%, specificity of 96.46%, and FPR of 3.54%, with the 6th, 7th, and 10th folds reporting the lowest. Again, BS 128 has won the race, outperforming the rest of the BSs by 10 folds on average.

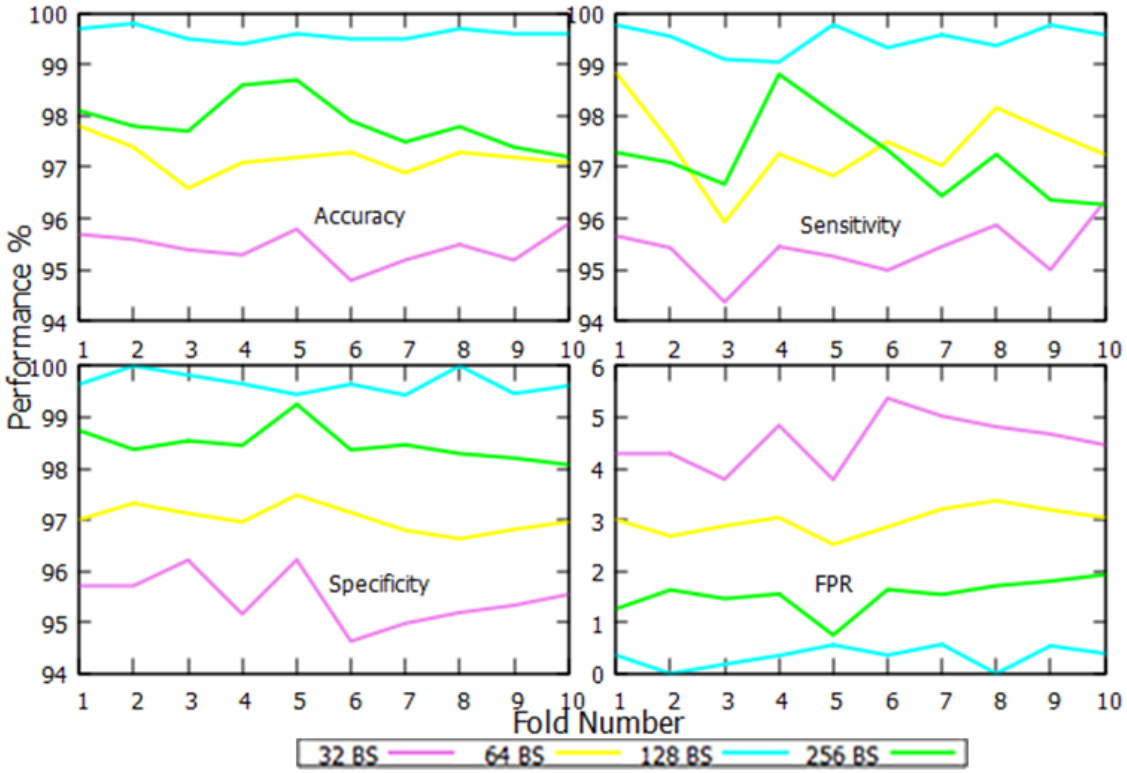

Fig. 7.6: Fold-wise accuracy, sensitivity, specificity, and FPR visualization of HV class.

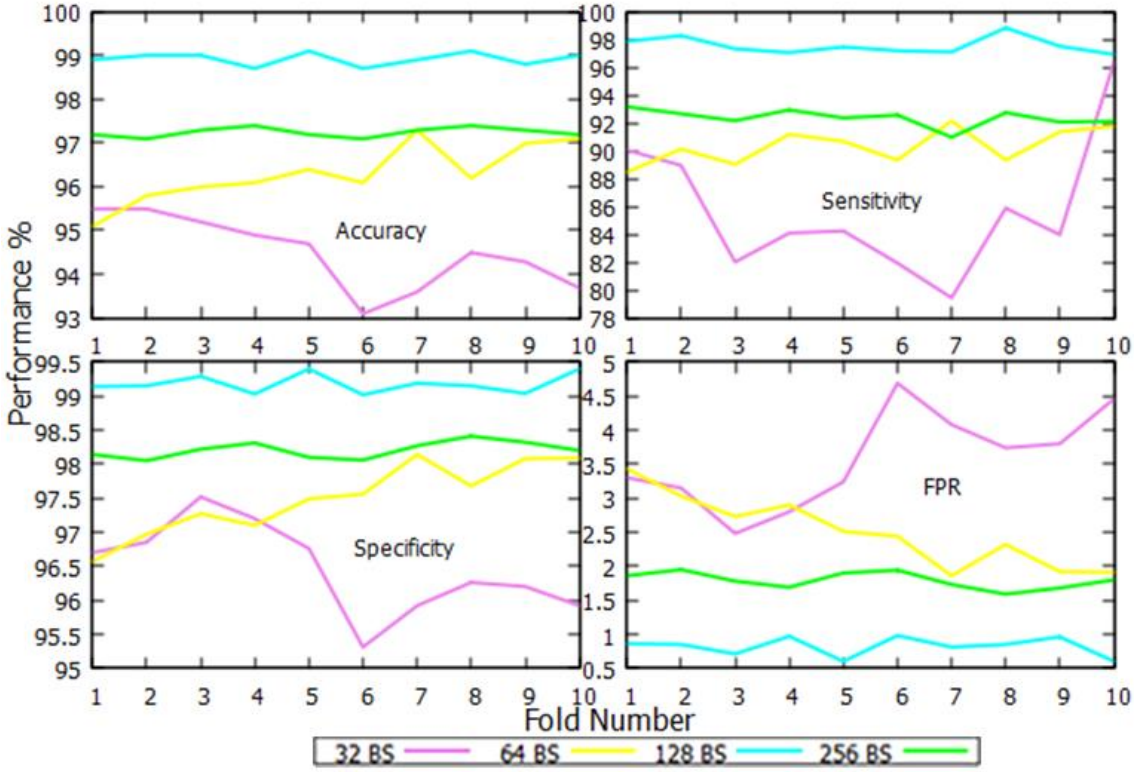

Fig. 7.7: Fold-wise accuracy, sensitivity, specificity, and FPR visualization of MCI class.

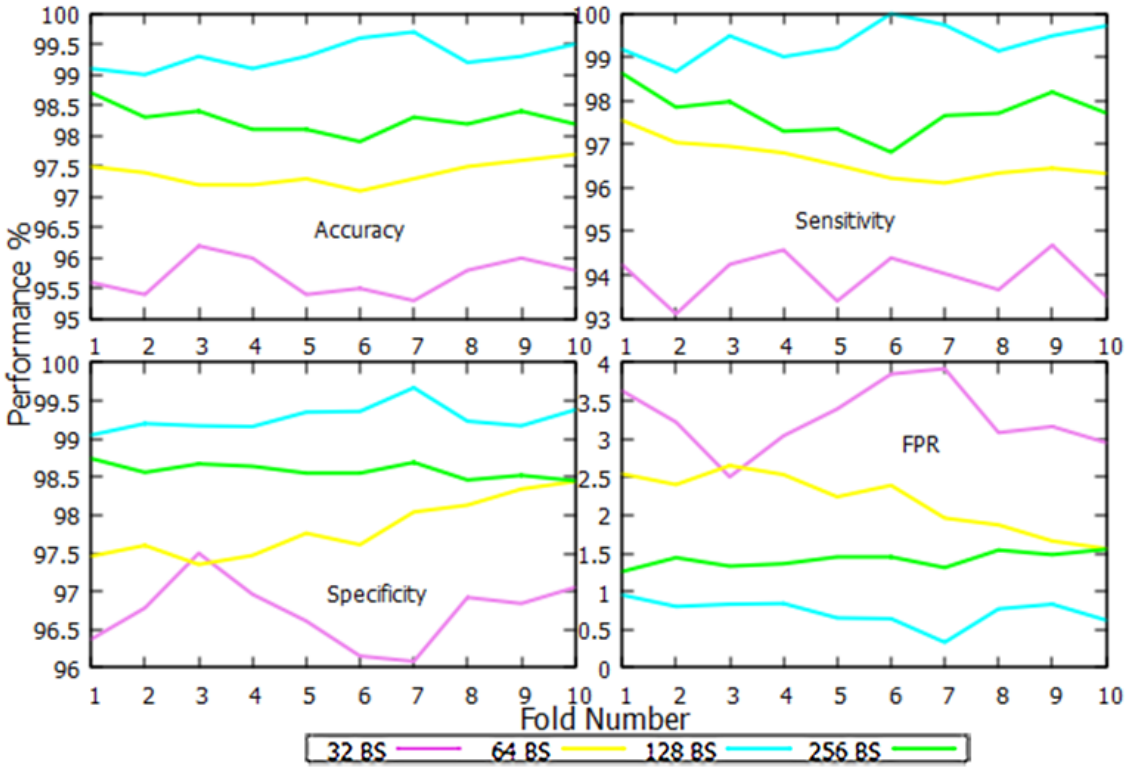

Fig. 7.8: Fold-wise accuracy, sensitivity, specificity, and FPR visualization of AD class.

The classification report for the AD class is no different, delineating that BS 128 is the optimal one shown in **Fig. 7.8**. The 10-fold average performance has again crossed over 99% when the BS is kept at 128. Such a BS setup produced an accuracy of 99.31%, a sensitivity of 99.36%, a specificity of 99.27%, and a FPR of 0.73%. All the folds in this setup have reported over 99% accuracy, but the sixth has the highest. For this contest, the BS of 256 is likewise quite close, with reported averages of 98.26% accuracy after 10 iterations, 97.72% sensitivity, 98.58% specificity, and 1.42% FPR. The BS of 64 has come in third place, obtaining an average accuracy of 97.37%, a sensitivity of 96.63%, a specificity of 97.82%, and a FPR of 2.18% across a 10-fold range. A 32-BS has an average 10-fold accuracy, sensitivity, specificity, and FPR of 95.69%, 93.99%, 96.72%, and 3.28%, respectively, with the most accuracy being contributed by the sixth fold and the lowest by the seventh. Again, it is a clear win for CDR-Net when the BS is set to 128.

Table 7.3: Overall Multiclass Classification Report.

BS	32			64			128			256		
Fold No	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE
1	95.09	94.97	95.19	97.39	98.37	96.66	99.20	99.18	99.21	98.00	97.29	98.56
2	94.89	94.75	95.00	96.49	96.11	96.79	99.20	98.94	99.36	98.10	97.52	98.56
3	95.19	93.92	96.21	96.79	96.15	97.31	99.50	99.49	99.50	97.70	96.67	98.54

4	95.29	95.25	95.32	97.29	97.48	97.15	99.30	99.26	99.33	98.00	97.85	98.10
5	95.59	95.24	95.87	97.19	96.83	97.48	99.20	98.94	99.36	97.70	97.19	98.13
6	94.39	94.33	94.43	96.99	96.82	97.13	99.30	99.20	99.36	98.00	97.33	98.54
7	94.89	94.58	95.14	97.09	97.25	96.97	99.80	99.74	99.84	97.80	96.65	98.85
8	95.09	94.98	95.17	97.19	97.93	96.62	99.10	98.57	99.38	98.09	97.46	98.66
9	95.59	95.45	95.69	96.99	97.25	96.79	99.00	98.99	99.00	97.79	96.61	98.74
10	96.09	96.36	95.88	97.19	97.48	96.97	98.90	98.87	98.91	97.49	96.88	98.06
AVG	**95.21**	**94.98**	**95.39**	**97.06**	**97.16**	**96.99**	**99.25**	**99.13**	**99.32**	**97.87**	**97.14**	**98.47**

The overall multiclass classification debriefing reported in **Table 7.3** tells me about the consistency and high classification performance of our proposed CDR-Net framework. The excellent detection accuracy illustrates how well CDR-Net can cope with a variety of brain illnesses. In addition, **Table 7.3** showcases fold-wise multiclass performance. For BSs of 32, 64, 128, and 256, the average multiclass accuracy is 95.21 percent, 97.0 percent, 99.25 percent, and 97.87 percent, respectively. The CDR-Net framework's average multiclass sensitivity for BSs of 32, 64, 128, and 128 is 94.98%, 97.16%, 99.13%, and 97.17%, correspondingly. Additionally, the average multiclass specificities for BSs of 32, 64, 128, and 256 are 95.39%, 96.99%, 99.32%, and 98.47%, accordingly. Once again, the 128-BS has edged out the competitors. BS 256 outperforms BS 64 in terms of performance.

Table 7.4: Fold and BS wise Time Complexity Report.

BS	32		64		128		256	
Fold No	#Epochs	Avg time per epoch (seconds)	#Epochs	Avg time per epoch (seconds)	#Epochs	Avg time per epoch (seconds)	#Epochs	Avg time per epoch (seconds)
1	27	3.22	30	3.28	33	2.84	22	2.98
2	30	3.14	28	3.18	36	1.97	29	2.66
3	28	3.14	40	3.18	24	2.47	18	2.73
4	20	3.16	25	3.18	25	2.50	41	2.68
5	19	3.16	31	3.18	29	2.37	23	2.69
6	33	3.14	33	3.18	36	2.43	31	2.72
7	32	3.14	29	3.22	26	2.37	25	2.81
8	17	3.17	31	3.21	27	2.53	18	2.61
9	21	3.16	25	3.18	20	2.59	38	2.76
10	22	3.15	24	3.22	18	2.57	20	2.96
AVG	**24.9**	**3.158**	**29.6**	**3.201**	**27.4**	**2.464**	**26.5**	**2.76**

In this study, I have not only considered the accuracy of our proposed CDR-Net framework but also its efficiency. **Table 7.4** showcases the efficiency of each fold and different BSs. In comparison to the other BSs, the 32 BS has needed the fewest average training epochs (24.9), on average. On the contrary, the BS of 64 spent the most time training the model, averaging 29.6 epochs. The BSs of 128 and 265 have consumed 27.4

and 26.5 epochs on average, respectively. Again, according to **Table 7.4**, the shortest average time per epoch for a BS of 128 is 2.46 seconds. The BSs of 32, 64, and 256 have recorded averages of 3.16, 3.2, and 2.76 seconds per epoch, respectively. These evaluations show that training with a BS of 128 has taken less time overall for each fold. Despite the fact that this configuration has taken more than 27 epochs on average in each fold, which is the third lowest of the four BSs, it is well ahead of its league.

To validate the stability and absence of overfitting, LOOCV has been performed. All segments of a subject have been kept aside from training and used for testing to predict the subject's status. This strategy continues until all the subjects are checked. With the use of LOOCV, I have conducted subject-wise detection on the main EEG data that were obtained from 109 participants in this study. Along with the earlier 32, 64, 128, and 256 BSs, I have also documented the performance of BSs 16 and 512 in this LOOCV phase. The average accuracy for batches of 16, 32, 64, 128, 256, and 512 is 88.61 percent, 94.1 percent, 97.0 percent, 99.10 percent, 98.0 percent, and 95.0 percent, respectively. Different BSs and subject-wise accuracy are visualized in **Fig. 7.9**, where the X axis represents the subject numbers, and the Y axis represents the corresponding accuracy. The 16, 32, 64, 128, 256, and 512 subject-wise accuracy of BSs are shown in **Fig. 7.9** by the orange, yellow, green, black, brown, and blue lines respectively. Among different BSs, 128 and 256 have performed better when LOOCV is in action. The worst performance has been reported for BS 16 due to data underfitting.

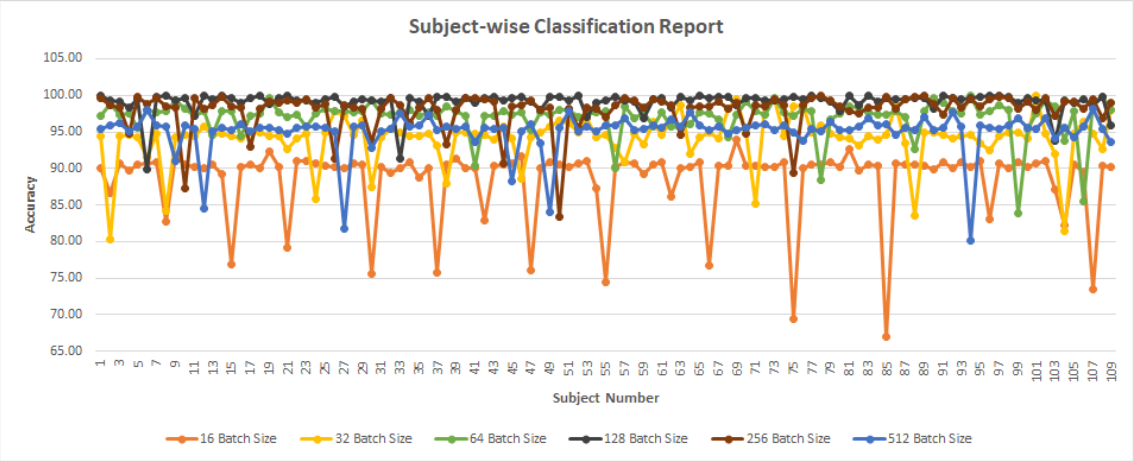

Fig. 7.9: Subject-wise Classification Report.

Table 7.5: Ablation Study.

Configurations	AVG Accuracy_10_Folds %	AVG Subject-wise Accuracy / LOOCV %
Proposed baseline setup	**99.25**	**99.09**
Added a feature extraction block	96.87	98.49
Removed a featured extraction block	94.60	93.03
Halved the filters in the convolution layers	98.19	97.54
Doubled the filters in the convolution layers	98.78	98.10
Changed the kernel size to 5 x 5	98.53	97.47
AlexNet [178]	74.38	69.44
InceptionNet [179]	81.92	76.63
ResNet50 [180]	94.35	89.86
VGG16 [181]	89.47	87.30

In order to strengthen the suggested CDR-Net framework, several performance matrices, including 10-fold and LOOCV, are being used. It is abundantly evident from both the 10-fold and LOOCV results that a BS of 128 is the best option for this dataset. Apart from that, over twenty CNN configurations have been investigated, and the top five setups are listed in **Table 7.5** before I recommend the best CNN configuration for our CDR-Net architecture. In the context of ML, and particularly sophisticated deep neural networks, a technique where certain components of the network are deleted or added in order to better understand the behavior of the network has been referred to as "ablation studies". **Table 7.5** compiles the results of the ablation investigation performance while maintaining a BS of 128. Here are the top five configurations along with well-established and popular CNN models depicted in **Table 7.5**, where a feature extraction block has been inserted and eliminated, the filters in the convolution layers are cut in half and doubled, and then the kernel window size has been raised to 5×5. All these configurations' 10-fold and LOOCV performances indicate that the proposed baseline CNN configuration is the optimal one. The classifier's performance suffered significantly when a feature extraction block was added or removed, but less so when the filter or kernel size was altered. Moreover, when trained on our EEG multiclass dataset, the popular CNN models AlexNet [178],

InceptionNet [179], ResNet50 [180], and VGG16 [181] failed to provide satisfactory results. Additionally, since they are complicated, layered CNN models, extra time was spent on training and testing. Only ResNet50 has achieved an average 10-fold and LOOCV multiclass accuracy of 94.35% and 89.86%, respectively, with the other CNN models lagging behind at 90%. Hence, the suggested design has been justified by the ablation work.

Since the feature extraction process is so intricate and hard to comprehend, DL is sometimes referred to as a "black box". T-distributed stochastic neighbor embedding (t-SNE) [177] is a widely used dimension reduction method which allows to visualize high dimensional like EEG by mapping it to a 2-D space. DL investigators often use t-SNE to visualize the categorization process. In **Fig. 7.10**, the layer-by-layer visualization of the second fold of the proposed CNN configuration's testing phase with a BS of 128 model is illustrated. This image shows that all of the samples are initially clustered together in the input layer for this testing, and that as it advances through the feature extraction blocks, the testing samples become dispersed. The proposed four feature extraction blocks make sure all the AD, MCI, and HV samples are clustered away from each other to have an accurate model. The testing samples are all clearly segregated from one another by the time they reach the dense layer.

Moreover, the class-wise ROC curves are showcased in **Fig. 7.11**. The ROC curves in yellow, blue, orange, and grey represent BSs of 32, 64, 128, and 256. The orange ROC curve, which reflects the 128-BS, has a substantially larger area underneath it in all 3 classes. This indicates a higher AUC reading. The ROC curve and the AUC value both complement one another and elevate the classifier.

7.5 Discussion

This CDR-Net system has been developed to diagnose AD, MCI, and HV using EEG data swiftly and correctly. To do so, I have accumulated the EEG data of 109 people, 49 of whom have AD, 39 have MCI, and 23 have HV. In order to clean the data and increase the sample size, the raw EEG data was down-sampled, de-noised by SWT, segmented, and transformed into an 8-bit color image before being sent to the CNN model. The CNN model has been tested in several setups, and **Table 7.5** reports the results. Our proposed CDR-Net structure has a 99.25%, 99.13%, and 99.32% average multiclass classification

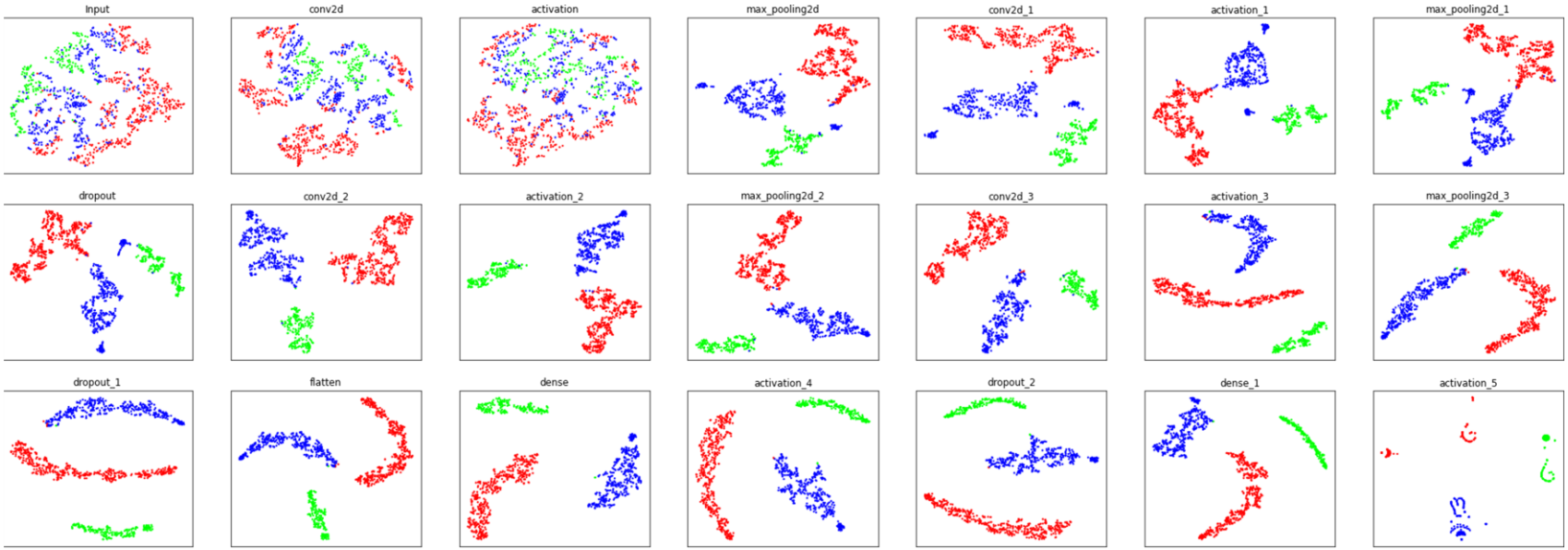

Fig. 7.10: T-SNE Visualization of the Categorization Process.

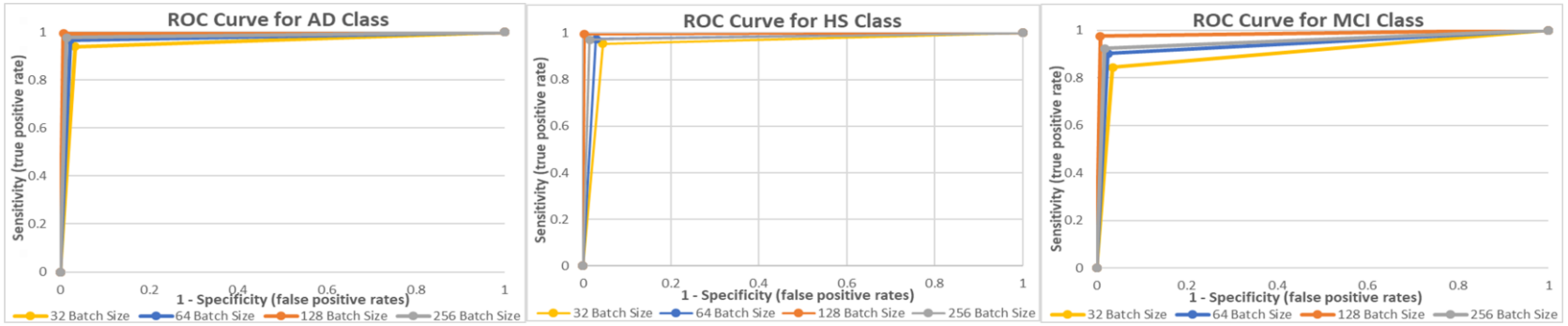

Fig. 7.11: ROC Curve of (a) AD Class, (b) HV Class, and (c) MCI Class.

accuracy, sensitivity, and specificity. 10-folds, LOOCV, and other performance matrices have been employed to demonstrate the performance.

Performance issues and computational complexity impeded earlier efforts to diagnose MCI or AD. The majority of the studied literature used TML techniques to successfully conduct binary classification. However, the performance drastically deteriorates when trying to identify various brain illnesses. It is caused by data noise, a limited sample size, the drawbacks of TML techniques, and, last but not least, the absence of crucial features. MCI and AD share some similar symptoms. The majority of earlier multiclass initiatives muddled up MCI with AD since they share several symptoms. Moreover, TML classifiers also need distinct techniques for choosing features and identifying them. Such two additional processes often muck things up or leave out key features, and they have some computing expenses associated with them. Since they are unable to extract features on their own, the TML classifiers rely heavily on these feature extraction techniques.

Table 7.6 provides a comparative summary between our proposed CDR-Net architecture, earlier investigations, and other well established CNN models. The majority of earlier work used the TML classifiers SVM, DT, KNN, MLP, and LR [155], [32], [131], [172], [173], [70], [13]. For the preprocessing stage, the bulk of ML-based efforts used CWT, FFT, PSD, ICA, and PCA algorithms. With DL classifiers like CNN, BLSTM, DCssCDBM, and EPNN, four investigations were included [171], [157], [159], [86], [13], [18]. In addition, the majority of the investigations that are reported on employed binary classification and have an average accuracy rate of above 90%. Furthermore, only half of the twelve recently published relevant studies employed multiclass categorization. Our prior work [171] among them had the greatest multiclass classification accuracy, with a score of 96.26%, utilizing identical EEG dataset. The second-best multiclass model was constructed by Bi and Wang [18] for predicting MCI, AD, and HV. With a multiclass accuracy of 75% and the KNN as the classifier, Pirrone *et al.*'s [155] model has been found to be the least accurate for multiclass performance. This is due to the classifiers' usage having a simplistic design, which resulted in improper feature extraction. Another important finding is the difficulty [172], [173], and [18] have had in obtaining a sufficient sample size.

After doing an analysis of the literature, I have come to the essence that, owing to their restricted design, TML approaches cannot effectively handle complicated data, such as EEG. It is challenging for such techniques to penetrate deeply into intricate layers and retrieve the vital characteristics. However, since these approaches require additional processing for feature extraction, they have high computational costs. These are the driving forces behind our study's decision to use a DL classifier that does not need additional feature extraction techniques and reduces processing expenses. Additionally, the methodologies used by NNs make it possible to extract important elements from very complicated and deep layers of data. Specifically, CNNs are often employed in digital image processing research and save a significant amount of time during training and testing. Since the deep layers of the NNs are all linked, feature extraction does not need human processing. Our classification performance makes it clear that the proper decisions I have taken have allowed me to achieve less than 1% FPR and over 99% accuracy. Consequently, this high performance is an indication of a good feature extraction and classification process. With an average of 27.4 epochs and 2.46 seconds per epoch, our suggested CDR-Net trained and tested the model for each fold in 67.404 seconds. It is a pointer to the effectiveness of our proposed framework. Comparatively to existing methods, CDR-Net is a simple and highly accurate framework for classifying AD, MCI, and HV using EEG data.

Table 7.6: Comparison with Earlier Efforts.

Efforts	Dataset	Method	Classifier	Classes	Performance
Fouladi *et al.* [159]	61 HV, 56 MCI, and 63 AD	TFR, CWT	CNN, Conv-AE neural networks	HV vs MCI vs AD	CNN 92% Conv-AE 89%
Ieracitano *et al.* [131]	63 AD, 63 HV, and 63 MCI	CWT, HOS, BiS	MLP, AE, LR, SVM	HV vs MCI vs AD	89.22%
Fiscon *et al.* [32]	23 HV, 49 AD, and 37 MCI	Discrete Fourier Transforms, DWT	DT	HV vs AD, HV vs MCI, and MCI vs AD	83%, 92%, and 79% respectively
Pirrone *et al.* [155]	48 AD, 37 MCI, 20 HV	PSD, FIR & Butterworth filter	KNN, DT, SVM	HV vs AD, HV vs MCI, MCI vs AD, HV vs MCI vs AD	97%, 95%, 83%, and 75% respectively for KNN
Perez-Valero *et al.* [173]	6 MCIs, 11 ADS, and 9 HVs	Autoreject, ICA	MLP	HV vs MCI vs AD	88% F1 score.

Puri *et al.* [172]	12 ADs and 11 HVs	HFD, KFD	SVM	AD vs HVs	accuracy of 98.5%
Sridhar and Manian [86]	28 ADs and 7 MCI participants	PCA	BLSTM	AD vs MCI	With those aged 40 to 60, it had increased by 91.93%, and with those beyond 60, by 65.73% accuracy
Amezquita-Sanchez *et al.* [13]	37 MCI and 37 AD	ANOVA, HE, FD, and MUSIC-EWT	EPNN, DT, NB, and KNN	AD vs MCI	accuracy of 90.3% by EPNN
Bi and Wang [18]	four ADs, four HVs, and four MCIs	FFT	DCssCDBM	HV vs MCI vs AD	95.04 percent accuracy
Poil *et al.* [70]	25 ADs and 61 MCIs	ICA, HT	LR	AD vs MCI	88% sensitivity and 82% specificity
DRAM-Net [171]	Our Dataset	SWT	deep residual network	HV vs MCI vs AD	Accuracy 96.26%
Our Proposed CDR-Net Framework	Our Dataset	**SWT**	**CNN**	**HV vs MCI vs AD**	**Accuracy 99.25%**

7.6 Synopsis

The designed CDR-Net system presented here is particularly efficient and accurate in recognizing patients with AD and MCI from EEG data. The emphasis in this study is on decreasing diagnosis time, improving performance over existing efforts, increasing trust in EEG data, and selecting the best classifier to reduce the number of false positives and negatives. Since the initial EEG data I have gathered includes artifacts and is inconsistent, the preprocessing step is given more attention. The raw signals are cleaned using SWT, which has taken care of both high- and low-frequency noises. By down-sampling the data to 256 Hz, the unevenness has been addressed. For more effective feature extraction, the data were divided into 5-second chunks and converted to 8-bit colored pictures at the completion of the preprocessing stage. Before settling on the CNN design with the maximum accuracy, I investigated over twenty different CNN setups. The suggested CNN model has just taken over a minute to complete the training and testing process for each round. In order to verify our suggested framework, the performance of the CDR-Net

architecture has been evaluated using 10-folds, subject-wise detection, and other considerable performance matrices. The final results show that the overall multiclass accuracy, sensitivity, and specificity for CDR-Net are 99.25%, 99.13%, and 99.32%. Succeeding investigations should concentrate on expanding the sample size and multiclass performance. A smaller sample size has an impact on the classifier's performance. The more data I supply to the classifier, the more it can learn and predict correctly. Moreover, it will be more ideal if more brain disorders like autism, schizophrenia, Parkinson's, etc. can be detected using a single model. I intend to develop a web-based technology that can identify many brain abnormalities using EEG data.

The essence of this dissertation is coming up in the next chapter. It contains a summary of each of the methodologies and frameworks proposed by us. The next chapter ends with the future research direction.

CHAPTER 8 : SUMMARY AND FUTURE RESEARCH DIRECTION

8.1 Overview

Enhanced support for the healthcare industry is required as the average lifespan of humans rises thanks to scientific and technological advancements. This is followed by the emergence of age-related illnesses. Electroencephalography (EEG) is the sole instrument that shows promise and is effective, affordable, portable, and rapid to help the enormous population that suffers from mental and neurological diseases. The categorisation of EEG data is crucial in biomedical research for identifying and evaluating brain activity. Determining and analysing the various kinds of EEG signals is a challenging endeavour that requires the examination of vast amounts of EEG data. Extreme caution is required when extracting information and characteristics from the EEG's deep, hidden, complicated layer. In this dissertation, I developed EEG data processing and categorisation algorithms by analysing two large EEG datasets of mild cognitive impairment (MCI), Alzheimer's disease (AD), and healthy volunteers (HVs) with the following objectives:

- Enhance performance from previous initiatives.
- Create a noise reduction method that will guarantee the EEG characteristics are preserved and available for extraction for further processing.
- Offer methods and strategies for handling the enormous amount of EEG data.
- Develop rapid, computationally inexpensive diagnostic techniques for mental illnesses like MCI and AD from HVs.
- Construct and contrast many deep learning (DL)-based approaches for detecting different types of brain disorders.

8.2 Contribution Synopsis

Chasing our objectives, I have built six methods in this dissertation. The remainder are EEG classification algorithms, with one approach being for noise reduction. Every time, I made an effort to do better and beat our previous mark. Finally, I have shown the best

DL-based multiclass classifier that is light and precise enough to identify AD, MCI, and HVs.

Chapter 3: A framework for long short-term memory (LSTM)-based MCI detection is what we've proposed as our initial attempt. EEG data are de-noised using the Butterworth filtering method in our suggested framework. The average filter has been used to partition and down-sample the filtered EEG data in order to guarantee that our lodged model is rapid enough to diagnose MCI. I believe that our method strikes a good middle ground between speed and accuracy in its modelling. I have reported on the leading 20 of the approximately 35 tests and models that have been developed during the course of this investigation. The studies have all been carried out on a regular PC that does not have a graphics card. The tests have also shown me that keeping the number of hidden LSTM layers between two and three keeps the performance matrices (accuracy, sensitivity, and specificity) of the LSTM models at 90%, while increasing the number of hidden layers and neurons cause performance to decline. In comparison to the other 19 models, our data indicate that Model 13, which consists of two hidden LSTM layers and a total of 1024 and 512 nodes, performed very well. During the process of developing the model, it took an average of 280 seconds to complete each epoch, and it took a total of 1242 seconds to test the whole model. The performance matrices also meet expectations, scoring 96.41% accurate, 96.55% sensitive, 94.95% specific, 4.04% false positive rate, 94.29% precise, 95.39% F1 score, and 96.25% AUC value in five-fold cross-inspection. The results of the five-fold cross-inspection did not reveal any variations that were statistically significant. It also upholds the answer to **research problems 4 and 5**. An overview of this chapter's successes may be seen below:

- This work is the first of its kind to make use of EEG signal data in the construction of an LSTM-based DL system for the quick detection of MCI.
- I investigated the use of a method called "average filtering" for down-sampling in order to further assist with the speedier completion of the recommended framework.
- Twenty LSTM models were evaluated in order to find the best LSTM prototype for MCI detection.
- Our approach has the ability to improve classification accuracy while simultaneously reducing the amount of time required for calculation.

Chapter 4: I have continued our work on the same EEG dataset, and this time I have created an adaptive denoiser method in addition to a second novel DL-based framework. A gated recurrent unit (GRU)-based framework for the categorisation of EEG data. The GRU algorithm is a variation of the recurrent neural network (RNN) and belongs to the domain of DL. According to our most reliable sources, the GRU has never conducted an EEG research with participants who had modest cognitive impairments. It is well knowledge that RNN performs better with sequential data, such as EEG. And the GRU is a memory-efficient DL model due to the fact that it does not carry as much memory as the LSTM model, which is another member of the RNN family. It is usual practise to utilise a deep GRU network for sequential prediction, and the architecture of such a network has hidden GRU layers that may selectively recall significant information for a period of time determined by the reset gate. The Butterworth filter, which is used to reduce noise and has the ability to generate a more linear phase reaction as well as a thorough flat response, is the first step in the suggested architecture that I have developed. After that, in order to cut down on the amount of work that needed to be done computationally, I down-sampled the filtered data and segmented it using a newly designed adaptive filter. When there are no or a minimal amount of outliers, the suggested adaptive filter performs well with any dataset. Last but not least, a two-layer GRU network has been designed to categorise EEG data. The peak signal-to-noise ratio of the adaptive noise remover algorithm that I have presented has reached 15.38 dB, exceeding the filters that are already in use, while the overall GRU-based framework has achieved 96.91% accuracy, 97.95% sensitivity, 96.16% specificity, and 96.39% F1 score. **Research problems 1, 2, 4, and 5** have a solution in it. The following is a concise summary of the primary efforts that have been made throughout this chapter:

- A GRU-based DL investigation for EEG categorisation is something that I have reported on for the very first time.
- To reduce the computational burden of the proposed model, I have devised our very own designed adaptive filtering as a down-sampling strategy.
- When compared to the other DL models, the retrieval results achieved by our proposed model are both efficient and competitive, while the construction of our model requires just a minimal amount of computing resources.

Chapter 5: Expanding on our prior research, I conducted an in-depth study that improved performance using the same GRU-based architecture. As opposed to our prior attempt to

diagnose MCI at the early stage, this substantial study has included a different EEG signal de-noising approach and three new classifiers. The aforementioned research goals have been addressed by the presentation of two techniques based on traditional machine learning (TML) and two approaches based on DL, respectively, for the detection of MCI. This is an enhancement of our prior approach for the identification of MCI that was based on the GRU [104]. This study's first steps include removing noise and distortions from the EEG data using a technique known as the stationary wavelet transform (SWT). After using an average filter to do the down-sampling from 256 Hz to 4 Hz, the filtered signals were then segmented. Finally, classifiers such as LSTM, GRU, support vector machine (SVM), and K-nearest neighbor (KNN) have been developed to discriminate MCI patients from HVs. Both SVM and KNN have utilized the characteristics that were retrieved using LSTM. GRU surpassed all other classifiers in this investigation, achieving the greatest levels of accuracy 95.51 ± 3.11%, sensitivity 97.52 ± 0.96%, and specificity 96.50 ± 0.97%, as well as an F1 score of 95.69 ± 2.26% and an AUC value of 96.48 ± 1.85%. SVM and KNN, two of the most used machine learning (ML) algorithms, have both been shown to perform poorly in terms of performance measures. KNN has obtained the lowest accuracy score possible, which is 80.85 ± 1.78%, a sensitivity score of 87.58 ± 0.84%, a specificity score of 80.71 ± 1.46%, an F1 score of 82.53 ± 1.12%, and an AUC value of 81.65 ± 3.62%. LSTM and GRU were neck and neck for the lead, but SVM has dropped a significant distance behind. It also shares answers to **research questions 2, 4, and 5**. The following is a list of the most significant contributions that this chapter has made:

- An LSTM and GRU-based DL research for MCI classification has been successfully completed for the very first time.
- Instead of employing two distinct feature extraction techniques, I have used the LSTM-extracted features for SVM and KNN to save computational overhead.
- In order to improve the proposed model's fast performance, I have looked at the average filtering approach for down-sampling.
- In addition to enhancing classification accuracy while decreasing computation time, our method has been thoroughly tested for consistency of performance using a 5-fold cross validation procedure.

Chapter 6: A big leap has been taken after the previous study. New multi-class EEG data has been a focus as I attempt to develop a multi-class AD, MCI, and HVs detection

system. In the process of pursuing our goals, I have devised a DL-based attempt to not only improve the performance, but also to extract those additional hidden complicated characteristics of EEG data that have a substantial participation in the classifier's learning rate. I propose a deep residual Alzheimer's disease and MCI detection network (DRAM-Net) framework consisting of four steps: acquiring EEG data, processing that data, using the DRAM-Net architecture to distinguish between people with MCI, AD, and HVs, and assessing the results of the experiment. In the step of EEG data collection, I collected the raw EEG data of 109 subjects. In the pre-processing step, I did things like down-sampling, noise removal, and temporal segmentation to make sure the EEG data was clean and ready to feed to the network. I did this by utilising the DRAM-Nett architecture, which is the place where the pre-processed data are fed and the classifications are done. Previous attempts had a difficult time doing adequately when trying to solve issues involving several classes. Additionally, preprocessing processes are leading current TML-based attempts to take more time and effort than they should. In response to these problems, I created DRAM-Net, a DL-based initiative. While classifying MCI, AD, and HVs, DRAM-Net achieved success rates of 97.79%, 98.06%, and 96.66% respectively. Our suggested framework has a 96.26% overall accuracy rate. **Research questions #2, 3, 4, and 5** are responded inside this chapter. The primary contribution that this DRAM-Net framework enhancement offers is outlined below:

- To identify AD-MCI, I present a novel deep residual network developed specifically for this purpose.
- Our investigation uses temporal segments that are five seconds in length, and analyzing the patient's status with only five seconds' worth of EEG data is sufficient.
- Using this particular EEG dataset, the DRAM-Net architecture that was just introduced has outperformed every other known multi-class AD-MCI study.

Chapter 7: The Cognitive Decline Recognition Network (CDR-Net) architecture is our last and most comprehensive answer to the difficulties of EEG data processing as well as the detection of AD and MCI. This design makes use of a convolutional neural network (CNN) model that was developed specifically for the purpose of identifying AD, MCI, and HV using EEG data. The CDR-Nett structure that I suggest consists of four stages, the first of which is the primary data collecting phase for EEG readings. I was able to collect EEG data from 109 subjects (23 HVs, 37 MCIs, and 49 ADs). After then, the raw

EEG signals that were recorded were down-sampled to 256 Hz in order to maintain consistency. This was done since there weren't many EEG data acquired at a sampling frequency of 1024 Hz throughout the study. The raw data have been processed using the stationary wavelet transform, often known as SWT, in order to eliminate artifacts and sounds. The versatility of SWT in handling both high- and low-frequency disturbances is well known. After the signals have been cleaned, they are segmented into 5-second frames and then reassembled so that the sample size may be increased and important information can be located more rapidly. Converting the segmented frames to 8-bit colour images is the last step before submitting them to the classifier. The CDR-Nett architecture is finished off with a custom-made multi-layer CNN model that employs a *softmax* classifier. This suggested CDR-Net system, which has been shown to be capable of detecting many kinds of cognitive impairments, has been evaluated using various performance metrics. The 10-fold and leave-one-out cross validations (LOOCV) were also carried out as part of our efforts to validate the consistency and stability of the CDR-Nett architecture that I have provided here. Over twenty various CNN configurations were looked at before I decided on the CNN design with the highest accuracy. The proposed CNN model has just taken over a minute to finish the training and testing procedure for each cycle. According to the final findings, CDR-Net has an overall multiclass accuracy, sensitivity, and specificity of 99.25%, 99.13%, and 99.32%. **Chapter 7** also solves **research questions #2, 3, 4, and 5**. The accomplishments of this chapter are summarised below:

- Developed and upheld a cutting-edge, accurate, dependable, and efficient CDR-Net system for identifying AD, MCI, and HV using EEG data.
- Improved the accuracy of multi-class categorisation in comparison to prior approaches on both the same EEG dataset and distinct EEG datasets.
- Performed both 10-fold and LOOCV cross validations in order to investigate the consistency and stability of the suggested CDR-Net.
- Undertook ablative experiments in order to find the CNN classifier that worked best for the CDR-Net system, which serves as its core component.

Table 8.1 illustrates the contribution of this dissertation at a glance. Chapter, method, detected disease, and performance are reported inside **Table 8.1**. It can be concluded from **Table 8.1** that I have gradually improved the performance of the proposed method to identify MCI and AD.

Table 8.1: Contribution at a Glance.

Chapter Number	Method	Disease Detected	Performance
3	LSTM	MCI + HV	Accuracy: 96.41% Sensitivity: 96.55% Specificity: 95.95%
4	Adaptive denoiser + GRU	MCI + HV	Accuracy: 96.91% Sensitivity: 97.95% Specificity: 96.16%
5	GRU	MCI + HV	Accuracy: 95.51 ± 3.11 % Sensitivity: 97.52 ± 0.96 % Specificity: 96.50 ± 0.97 %
6	DRN	MCI + AD	Accuracy: 96.26% Sensitivity: 95.01% Specificity: 98.11%
7	CNN	MCI + AD	Accuracy: 99.25% Sensitivity: 99.13% Specificity: 99.32%

8.3 Future Work

In the realm of EEG data categorization, I believe that the method and frameworks that are outlined in this dissertation will prove to be useful in producing positive outcomes. The classification of EEG signals will be the subject of more study in the near future, specifically focusing on the viability of employing the aforementioned methods. In order to assist and make use of these proposed techniques, I have brought to future researchers' attention a few key points, which will now be explained.

Datasets play a vital role in any model's training and decision-making processes. It would have been better if I had been able to collect a dataset that was substantial in size, rich in variety, comprised of many classes, and included EEG data. This would have been especially advantageous for the frameworks that I had suggested, given that such frameworks often function more effectively when presented with large amounts of data. The performance of the classifier is affected when there is a reduced number of observations in the sample. The more information I provide the classifier, the more it is

able to learn and the more accurate its predictions will be. Therefore, the data sample size should be given priority in succeeding studies.

Integration of other neurological illnesses is a further important step that may be taken towards improvement. Running this model on EEG datasets from individuals with different brain disorders or cooperating with experts in the same field is another area where I need to improve or hereafter work plans I have. In addition, I want to broaden the scope of our research to include a variety of other neurological conditions, such as epilepsy, Parkinson disease, schizophrenia, autism, seizure disorders, and so on. Also, it would be beneficial for researchers in this area to combine their efforts and work together in order to collect EEG data from persons who have autism, schizophrenia, Parkinson's disease, and other conditions. In the future, I want to establish a concept that I call "one framework for all neuro-disorders." This indicates that a single model may be used to diagnose many types of brain illnesses. Last but not least, multiclass performance need also to be one of the key focus of any succeeding study.

Occasionally, the preprocessing of EEG data is unnoticed. In further research, the preprocessing processes need to likewise be the primary emphasis, with the goals of improving both the data quality and the pace at which the model is able to learn. The use of a DL approach rather than traditional ML methods may be beneficial for the purpose of extracting deeper complicated EEG data elements. In the future, studies of our proposed LSTM, GRU, and CNN architectures with a variety of additional hidden layer counts and activation parameters will be able to be carried out to see whether or not there are any benefits.

This endeavor will direct those working in technology and medicine to continue EEG research at a higher level and create new ideas and approaches for treating neurological illnesses. With the help of succeeding EEG researchers, I expect that they will be able to create a perfect EEG identification model. I intend to create a web-based application that detects several brain issues using EEG data.

REFERENCES

[1] Australian Bureau of Statistics. (2014). Causes of death, Australia 2014 (No. 3303.0).

[2] Aghajani, H., Zahedi, E., Jalili, M., Keikhosravi, A., & Vahdat, B. V. (2013). Diagnosis of early Alzheimer's disease based on EEG source localization and a standardized realistic head model. *IEEE journal of biomedical and health informatics*, *17*(6), 1039-1045.

[3] Akut, R. (2019). Wavelet based deep learning approach for epilepsy detection. *Health information science and systems*, *7*(1), 8.

[4] Al Ghayab, H. R., Li, Y., Siuly, S., & Abdulla, S. (2019). A feature extraction technique based on tunable Q-factor wavelet transform for brain signal classification. *Journal of neuroscience methods*, *312*, 43-52.

[5] Al-Jumeily, D., Iram, S., Hussain, A. J., Francois-Benois, V., & Fergus, P. (2014). Early detection method of Alzheimer's disease using EEG signals. In *Intelligent Computing in Bioinformatics: 10th International Conference, ICIC 2014, Taiyuan, China, August 3-6, 2014. Proceedings 10* (pp. 25-33). Springer International Publishing.

[6] Al-Qazzaz, N. K., Ali, S. H. B., Ahmad, S. A., Chellappan, K., Islam, M., & Escudero, J. (2014). Role of EEG as biomarker in the early detection and classification of dementia. *The Scientific World Journal*, *2014*.

[7] Alvi, A. M., Basher, S. F., Himel, A. H., Sikder, T., Islam, M., & Rahman, R. M. (2017, July). An adaptive grayscale image de-noising technique by fuzzy inference system. In *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp. 1301-1308). IEEE.

[8] Alvi, A. M., Tasneem, N., Hasan, M. A., & Akther, S. B. (2019). A study to find the impacts of strikes on students and local shopkeepers in Bangladesh. In *World Congress on Sustainable Technologies (WCST-2019)*.

[9] Alvi, A. M., Siuly, S., Wang, H., Sun, L., & Cao, J. (2020). An adaptive image smoothing technique based on localization. In *Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference (FLINS 2020)* (pp. 866-873).

[10] Alvi, A., Tasneem, N., Hasan, A., & Akther, S. (2020). Impacts of blockades and strikes in Dhaka: a survey. *Int J Innov Bus Strat*, *6*(1), 369-377.

[11] American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders-DMS III*. American Psychiatric Association.

[12] Amezquita-Sanchez, J. P., & Adeli, H. (2015). A new music-empirical wavelet transform methodology for time–frequency analysis of noisy nonlinear and non-stationary signals. *Digital Signal Processing*, *45*, 55-68.

[13] Amezquita-Sanchez, J. P., Mammone, N., Morabito, F. C., Marino, S., & Adeli, H. (2019). A novel methodology for automated differential diagnosis of mild cognitive impairment and the Alzheimer's disease using EEG signals. *Journal of neuroscience methods*, *322*, 88-95.

[14] Anuradha, G., Jamal, N., & Rafiammal, S. (2017, September). Detection of dementia in EEG signal using dominant frequency analysis. In *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)* (pp. 710-714). IEEE.

[15] Arisi, I., Bertolazzi, P., Cappelli, E., Conte, F., Cumbo, F., Fiscon, G., ... & Taglino, F. (2018, December). An ontology-based approach to improve data querying and organization of Alzheimer's Disease data. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2732-2734). IEEE.

[16] Bagattini, C., Mutanen, T. P., Fracassi, C., Manenti, R., Cotelli, M., Ilmoniemi, R. J., ... & Bortoletto, M. (2019). Predicting Alzheimer's disease severity by means of TMS–EEG coregistration. *Neurobiology of aging*, *80*, 38-45.

[17] Baradits, M., Kakuszi, B., Bálint, S., Fullajtár, M., Mód, L., Bitter, I., & Czobor, P. (2019). Alterations in resting-state gamma activity in patients with schizophrenia: a high-density EEG study. *European archives of psychiatry and clinical neuroscience*, *269*, 429-437.

[18] Bi, X., & Wang, H. (2019). Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning. *Neural Networks*, *114*, 119-135.

[19] Bibina, V. C., Chakraborty, U., Lourde, M., & Kumar, A. (2018). Signal processing methods of diagnosing alzheimer's disease using eeg a technical review. *Int J Biol Biomed Eng*, *12*(2018), 100-113.

[20] Boostani, R., Sadatnezhad, K., & Sabeti, M. (2009). An efficient classifier to diagnose of schizophrenia based on the EEG signals. *Expert Systems with Applications*, *36*(3), 6492-6499.

[21] Bracco, L., Gallato, R., Grigoletto, F., Lippi, A., Lepore, V., Bino, G., ... & Amaducci, L. (1994). Factors affecting course and survival in Alzheimer's disease: a 9-year longitudinal study. *Archives of Neurology*, *51*(12), 1213-1219.

[22] Cassani, R., Estarellas, M., San-Martin, R., Fraga, F. J., & Falk, T. H. (2018). Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment. *Disease markers*, *2018*.

[23] Cedazo-Minguez, A., & Winblad, B. (2010). Biomarkers for Alzheimer's disease and other forms of dementia: clinical needs, limitations and future aspects. *Experimental gerontology*, *45*(1), 5-14.

[24] Chiang, H. S., & Pao, S. C. (2016). An EEG-based fuzzy probability model for early diagnosis of Alzheimer's disease. *Journal of medical systems*, *40*, 1-9.

[25] Das, K., & Pachori, R. B. (2021). Schizophrenia detection technique using multivariate iterative filtering and multichannel EEG signals. *Biomedical Signal Processing and Control*, *67*, 102525.

[26] Deb, P., & Sefton, M. (1996). The distribution of a Lagrange multiplier test of normality. *Economics Letters*, *51*(2), 123-130.

[27] Demir, F., Sobahi, N., Siuly, S., & Sengur, A. (2021). Exploring deep learning features for automatic classification of human emotion using EEG rhythms. *IEEE Sensors Journal*, *21*(13), 14923-14930.

[28] Devia, C., Mayol-Troncoso, R., Parrini, J., Orellana, G., Ruiz, A., Maldonado, P. E., & Egaña, J. I. (2019). EEG classification during scene free-viewing for schizophrenia detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(6), 1193-1199.

[29] Durongbhan, P., Zhao, Y., Chen, L., Zis, P., De Marco, M., Unwin, Z. C., ... & Sarrigiannis, P. G. (2019). A dementia classification framework using frequency and time-frequency features based on EEG signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(5), 826-835.

[30] Duthey, B. (2013). Background paper 6.11: Alzheimer disease and other dementias. *A public health approach to innovation*, *6*, 1-74.

[31] Fiscon, G., Weitschek, E., Felici, G., Bertolazzi, P., De Salvo, S., Bramanti, P., & De Cola, M. C. (2014, December). Alzheimer's disease patients classification through EEG signals processing. In *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)* (pp. 105-112). IEEE.

[32] Fiscon, G., Weitschek, E., Cialini, A., Felici, G., Bertolazzi, P., De Salvo, S., ... & De Cola, M. C. (2018). Combining EEG signal processing with supervised methods for Alzheimer's patients classification. *BMC medical informatics and decision making*, *18*(1), 1-10.

[33] Fraga, F. J., Falk, T. H., Kanda, P. A., & Anghinah, R. (2013). Characterizing Alzheimer's disease severity via resting-awake EEG amplitude modulation analysis. *PloS one*, *8*(8), e72240.

[34] Fuchs, M., Kastner, J., Wagner, M., Hawes, S., & Ebersole, J. S. (2002). A standardized boundary element method volume conductor model. *Clinical neurophysiology*, *113*(5), 702-712.

[35] Haas, L. F. (2003). Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(1), 9-9.

[36] Hampel, H., Frank, R., Broich, K., Teipel, S. J., Katz, R. G., Hardy, J., ... & Blennow, K. (2010). Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. *Nature reviews Drug discovery*, *9*(7), 560-574.

[37] Hasan, M. A., Tasneem, N., Akther, S. B., Das, K., & Alvi, A. M. (2019). An analysis on recent mobile application trend in Bangladesh. In *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019) 33* (pp. 195-204). Springer International Publishing.

[38] Hassan, A. R., Siuly, S., & Zhang, Y. (2016). Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating. *Computer methods and programs in biomedicine*, *137*, 247-259.

[39] Helzner, E. P., Scarmeas, N., Cosentino, S., Tang, M. X., Schupf, N., & Stern, Y. (2008). Survival in Alzheimer disease: a multiethnic, population-based study of incident cases. *Neurology*, *71*(19), 1489-1495.

[40] Heyn, S. N., & Davis, C. P. (2020). Parkinson's disease symptoms, signs, causes, stages, and treatment.

[41] Horváth, A., Szűcs, A., Barcs, G., & Kamondi, A. (2017). Sleep EEG detects epileptiform activity in Alzheimer's disease with high sensitivity. *Journal of Alzheimer's Disease*, *56*(3), 1175-1183.

[42] Houmani, N., Vialatte, F., Gallego-Jutglà, E., Dreyfus, G., Nguyen-Michel, V. H., Mariani, J., & Kinugawa, K. (2018). Diagnosis of Alzheimer's disease with Electroencephalography in a differential framework. *PloS one*, *13*(3), e0193607.

[43] Hussain, W., Sadiq, M. T., Siuly, S., & Rehman, A. U. (2021). Epileptic seizure detection using 1 D-convolutional long short-term memory neural networks. *Applied Acoustics*, *177*, 107941.

[44] Alzheimer's Disease International. (2020). The global voice on dementia: Dementia statistics.

[45] Jahmunah, V., Oh, S. L., Rajinikanth, V., Ciaccio, E. J., Cheong, K. H., Arunkumar, N., & Acharya, U. R. (2019). Automated detection of schizophrenia using nonlinear signal processing methods. *Artificial intelligence in medicine*, *100*, 101698.

[46] Jalili, M., & Knyazeva, M. G. (2011). EEG-based functional networks in schizophrenia. *Computers in Biology and Medicine*, *41*(12), 1178-1186.

[47] Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163-172.

[48] Jurcak, V., Tsuzuki, D., & Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, *34*(4), 1600-1611.

[49] Kabir, E., & Zhang, Y. (2016). Epileptic seizure detection from EEG signals using logistic model trees. *Brain informatics*, *3*(2), 93-100.

[50] Kashefpoor, M., Rabbani, H., & Barekatain, M. (2016). Automatic diagnosis of mild cognitive impairment using electroencephalogram spectral features. *Journal of medical signals and sensors*, *6*(1), 25.

[51] Khare, S. K., Bajaj, V., Siuly, S., & Sinha, G. R. (2020). Classification of schizophrenia patients through empirical wavelet transformation using electroencephalogram signals. In *Modelling and Analysis of Active Biopotential Signals in Healthcare, Volume 1*. IOP Publishing.

[52] Khatun, S., Morshed, B. I., & Bidelman, G. M. (2019). A single-channel EEG-based approach to detect mild cognitive impairment via speech-evoked brain responses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(5), 1063-1070.

[53] Kim, D., & Kim, K. (2018, July). Detection of early stage Alzheimer's disease using EEG relative power with deep neural network. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 352-355). IEEE.

[54] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[55] Lekshmi, G. S., & Chacko, B. (2016). Robust Alzheimer's disease severity classification in compressed EEG signal. In *National Conference on Emerging Trends in Engineering and Technology (NCETET'16)* (pp. 22-28).

[56] Li, F., Wang, J., Liao, Y., Yi, C., Jiang, Y., Si, Y., ... & Xu, P. (2019). Differentiation of schizophrenia by combining the spatial EEG brain network patterns of rest and task P300. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(4), 594-602.

[57] Lu, J., Plataniotis, K. N., Venetsanopoulos, A. N., & Li, S. Z. (2006). Ensemble-based discriminant learning with boosting for face recognition. *IEEE transactions on neural networks*, *17*(1), 166-178.

[58] Mazaheri, A., Segaert, K., Olichney, J., Yang, J. C., Niu, Y. Q., Shapiro, K., & Bowman, H. (2018). EEG oscillations during word processing predict MCI conversion to Alzheimer's disease. *NeuroImage: Clinical*, *17*, 188-197.

[59] Miften, F. S., Diykh, M., Abdulla, S., Siuly, S., Green, J. H., & Deo, R. C. (2021). A new framework for classification of multi-category hand grasps using EMG signals. *Artificial Intelligence in Medicine*, *112*, 102005.

[60] Muniz, C. F., Shenoy, A. V., O'Connor, K. L., Bechek, S. C., Boyle, E. J., Guanci, M. M., ... & Rosenthal, E. S. (2016). Clinical development and implementation of an institutional guideline for prospective EEG monitoring and reporting of delayed cerebral ischemia. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, *33*(3), 217.

[61] Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2020). A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, *32*, 10927-10933.

[62] World Health Organization. (2012). *Dementia: a public health priority*. World Health Organization.

[63] Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol*, *24*(Suppl D), 5-12.

[64] Pasha, A., & Latha, P. H. (2020). Bio-inspired dimensionality reduction for Parkinson's disease (PD) classification. *Health information science and systems*, *8*, 1-22.

[65] Paul, S., Alvi, A. M., Nirjhor, M. A., Rahman, S., Orcho, A. K., & Rahman, R. M. (2017). Analyzing accident prone regions by clustering. *Advanced Topics in Intelligent Information and Database Systems 9*, 3-13.

[66] Paul, S., Alvi, A. M., & Rahman, R. M. (2021). An analysis of the most accident prone regions within the Dhaka Metropolitan Region using clustering. *International Journal of Advanced Intelligence Paradigms*, *18*(3), 294-315.

[67] Phang, C. R., Noman, F., Hussain, H., Ting, C. M., & Ombao, H. (2019). A multi-domain connectome convolutional neural network for identifying schizophrenia from EEG connectivity patterns. *IEEE journal of biomedical and health informatics*, *24*(5), 1333-1343.

[68] Piano, C., Mazzucchi, E., Bentivoglio, A. R., Losurdo, A., Calandra Buonaura, G., Imperatori, C., ... & Della Marca, G. (2017). Wake and sleep EEG in patients with Huntington disease: an eLORETA study and review of the literature. *Clinical EEG and neuroscience*, *48*(1), 60-71.

[69] Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, *10*(3), 61-74.

[70] Poil, S. S., De Haan, W., van der Flier, W. M., Mansvelder, H. D., Scheltens, P., & Linkenkaer-Hansen, K. (2013). Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage. *Frontiers in aging neuroscience*, *5*, 58.

[71] Prince, M. J., Wimo, A., Guerchet, M. M., Ali, G. C., Wu, Y. T., & Prina, M. (2015). World Alzheimer Report 2015-The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends.

[72] Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. *ieee assp magazine*, *3*(1), 4-16.

[73] Rodrigues, P. M., Freitas, D., Teixeira, J. P., Bispo, B., Alves, D., & Garrett, C. (2018). Electroencephalogram hybrid method for alzheimer early detection. *Procedia computer science*, *138*, 209-214.

[74] Ross, T. J. (2009). *Fuzzy logic with engineering applications*. John Wiley & Sons.

[75] Sharma, N., Kolekar, M. H., Jha, K., & Kumar, Y. (2019). EEG and cognitive biomarkers based mild cognitive impairment diagnosis. *Irbm*, *40*(2), 113-121.

[76] Sharma, N., Kolekar, M. H., & Jha, K. (2020). Iterative filtering decomposition based early dementia diagnosis using EEG with cognitive tests. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(9), 1890-1898.

[77] Sharma, N., Kolekar, M. H., & Jha, K. (2021). EEG based dementia diagnosis using multi-class support vector machine with motor speed cognitive test. *Biomedical Signal Processing and Control*, *63*, 102102.

[78] Silver, J., Schwab, M. E., & Popovich, P. G. (2015). Central nervous system regenerative failure: role of oligodendrocytes, astrocytes, and microglia. *Cold Spring Harbor perspectives in biology*, *7*(3), a020602.

[79] Siuly, S., & Li, Y. (2015). Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification. *Computer methods and programs in biomedicine*, *119*(1), 29-42.

[80] Siuly, S., & Zhang, Y. (2016). Medical big data: neurological diseases diagnosis through medical data analysis. *Data Science and Engineering*, *1*, 54-64.

[81] Siuly, S., Li, Y., Zhang, Y., Siuly, S., Li, Y., & Zhang, Y. (2016). Significance of eeg signals in medical and health research. *EEG Signal Analysis and Classification: Techniques and Applications*, 23-41.

[82] Siuly, S., Alcin, O. F., Bajaj, V., Sengur, A., & Zhang, Y. (2019). Exploring Hermite transformation in brain signal analysis for the detection of epileptic seizure. *IET Science, Measurement & Technology*, *13*(1), 35-41.

[83] Siuly, S., Bajaj, V., Sengur, A., & Zhang, Y. (2019). An advanced analysis system for identifying alcoholic brain state through EEG signals. *International Journal of Automation and Computing*, *16*, 737-747.

[84] Siuly, S., Alçin, Ö. F., Kabir, E., Şengür, A., Wang, H., Zhang, Y., & Whittaker, F. (2020). A new framework for automatic detection of patients with mild cognitive impairment using resting-state EEG signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(9), 1966-1976.

[85] Siuly, S., Khare, S. K., Bajaj, V., Wang, H., & Zhang, Y. (2020). A computerized method for automatic detection of schizophrenia using EEG signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(11), 2390-2400.

[86] Sridhar, S., & Manian, V. (2020). Eeg and deep learning based brain cognitive function classification. *Computers*, *9*(4), 104.

[87] Stiles, J., & Jernigan, T. L. (2010). The basics of brain development. *Neuropsychology review*, *20*(4), 327-348.

[88] Stoica, P., & Moses, R. L. (1997). *Introduction to spectral analysis*. Pearson Education.

[89] Supriya, S., Siuly, S., Wang, H., Cao, J., & Zhang, Y. (2016). Weighted visibility graph with complex network features in the detection of epilepsy. *IEEE access*, *4*, 6554-6566.

[90] Supriya, S., Siuly, S., & Zhang, Y. (2016). Automatic epilepsy detection from EEG introducing a new edge weight method in the complex network. *Electronics Letters*, *52*(17), 1430-1432.

[91] Supriya, S., Siuly, S., Wang, H., & Zhang, Y. (2018). EEG sleep stages analysis and classification based on weighed complex network features. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *5*(2), 236-246.

[92] Supriya, S., Siuly, S., Wang, H., & Zhang, Y. (2021). Epilepsy detection from EEG using complex network techniques: A review. *IEEE Reviews in Biomedical Engineering*.

[93] Supriya, S., Siuly, S., Wang, H., & Zhang, Y. (2021). New feature extraction for automated detection of epileptic seizure using complex network framework. *Applied Acoustics*, *180*, 108098.

[94] Deng, L., & Liu, Y. (2018). *Deep learning in natural language processing*. by the registered company Springer Nature Singapore Pte Ltd..

[95] Wang, J., Barstein, J., Ethridge, L. E., Mosconi, M. W., Takarae, Y., & Sweeney, J. A. (2013). Resting state EEG abnormalities in autism spectrum disorders. *Journal of neurodevelopmental disorders*, *5*, 1-14.

[96] Watanabe, Y., Kobayashi, Y., Tanaka, M., Asada, T., Ishii, K., & Yagi, T. (2017, August). Analysis for Alzheimer's disease using cross-correlation of EEG data. In *2017 10th Biomedical Engineering International Conference (BMEiCON)* (pp. 1-5). IEEE.

[97] World Health Organization. (2006). *Neurological disorders: public health challenges*. World Health Organization.

[98] World Health Organization. (2019). *Schizophrenia*. World Health Organization.

[99] Yang, S., Bornot, J. M. S., Wong-Lin, K., & Prasad, G. (2019). M/EEG-based bio-markers to predict the MCI and Alzheimer's disease: a review from the ML perspective. *IEEE Transactions on Biomedical Engineering*, *66*(10), 2924-2935.

[100] Yin, J., Cao, J., Siuly, S., & Wang, H. (2019). An integrated MCI detection framework based on spectral-temporal analysis. *International Journal of Automation and Computing*, *16*, 786-799.

[101] Yuvaraj, R., Rajendra Acharya, U., & Hagiwara, Y. (2018). A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals. *Neural Computing and Applications*, *30*, 1225-1235.

[102] Zarei, R., He, J., Siuly, S., Huang, G., & Zhang, Y. (2019). Exploring Douglas-Peucker algorithm in the detection of epileptic seizure from multicategory EEG signals. *BioMed research international*, *2019*.

[103] Zhou, N., & Wang, L. (2007). A modified T-test feature selection method and its application on the HapMap genotype data. *Genomics, proteomics & bioinformatics*, *5*(3-4), 242-249.

[104] Alvi, A. M., Siuly, S., & Wang, H. (2022, January). Developing a deep learning based approach for anomalies detection from EEG data. In *Web Information Systems Engineering–WISE 2021: 22nd International Conference on Web Information Systems Engineering, WISE 2021, Melbourne, VIC, Australia, October 26–29, 2021, Proceedings, Part I* (pp. 591-602). Cham: Springer International Publishing.

[105] A. M. Alvi, S. Siuly and H. Wang, "A Long Short-Term Memory Based Framework for Early Detection of Mild Cognitive Impairment From EEG Signals," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 2, pp. 375-388, April 2023, doi: 10.1109/TETCI.2022.3186180.

[106] Alvi, A. M., Siuly, S., & Wang, H. (2022). Neurological abnormality detection from electroencephalography data: a review. *Artificial Intelligence Review*, *55*(3), 2275-2312.

[107] Alvi, A. M., Siuly, S., Wang, H., Wang, K., & Whittaker, F. (2022). A deep learning based framework for diagnosis of mild cognitive impairment. *Knowledge-Based Systems*, *248*, 108815.

[108] D. Australia, "Mild cognitive impairment", Jan. 2021, [online] Available: https://www.dementia.org.au/about-dementia-and-memory-loss/about-dementia/memory-loss/mild-cognitive-impairment.

[109] You, Z., Zeng, R., Lan, X., Ren, H., You, Z., Shi, X., ... & Hu, X. (2020). Alzheimer's disease classification with a cascade neural network. *Frontiers in Public Health*, *8*, 584387.

[110] Rodrigues, P. M., Bispo, B. C., Garrett, C., Alves, D., Teixeira, J. P., & Freitas, D. (2021). Lacsogram: A new EEG tool to diagnose Alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, *25*(9), 3384-3395.

[111] Prince, M. J., Wimo, A., Guerchet, M. M., Ali, G. C., Wu, Y. T., & Prina, M. (2015). World Alzheimer Report 2015-The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends.

[112] Mavrodaris, A., Powell, J., & Thorogood, M. (2013). Prevalences of dementia and cognitive impairment among older people in sub-Saharan Africa: a systematic review. *Bulletin of the World Health Organization*, *91*, 773-783.

[113] Australian Bureau of Statistics. (2018). Causes of death, Australia.

[114] W. H. Organization et al., "The top 10 causes of death: World health organization", vol. 2020, Jan. 2021, [online] Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

[115] Alzheimer's Disease International. (2019). World Alzheimer report 2019: attitudes to dementia. *Alzheimer's Disease International: London*.

[116] Jamaloo, F., Mikaeili, M., & Noroozian, M. (2020). Multi metric functional connectivity analysis based on continuous hidden Markov model with application in early diagnosis of Alzheimer's disease. *Biomedical Signal Processing and Control*, *61*, 102056.

[117] Kashefpoor, M., Rabbani, H., & Barekatain, M. (2019). Supervised dictionary learning of EEG signals for mild cognitive impairment diagnosis. *Biomedical Signal Processing and Control*, *53*, 101559.

[118] Hadiyoso, S., Cynthia, C. L. F. A. R., ER, M. T. L., & Zakaria, H. (2019, September). Early detection of mild cognitive impairment using quantitative analysis of EEG signals. In *2019 2nd International Conference on Bioinformatics, Biotechnology and Biomedical Engineering (BioMIC)-Bioinformatics and Biomedical Engineering* (Vol. 1, pp. 1-5). IEEE.

[119] Chen, H., Song, Y., & Li, X. (2019). A deep learning framework for identifying children with ADHD using an EEG-based brain network. *Neurocomputing*, *356*, 83-96.

[120] Mazaheri, A., Segaert, K., Olichney, J., Yang, J. C., Niu, Y. Q., Shapiro, K., & Bowman, H. (2018). EEG oscillations during word processing predict MCI conversion to Alzheimer's disease. *NeuroImage: Clinical*, *17*, 188-197.

[121] Poil, S. S., De Haan, W., van der Flier, W. M., Mansvelder, H. D., Scheltens, P., & Linkenkaer-Hansen, K. (2013). Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage. *Frontiers in aging neuroscience*, *5*, 58.

[122] Yang, D., Huang, R., Yoo, S. H., Shin, M. J., Yoon, J. A., Shin, Y. I., & Hong, K. S. (2020). Detection of mild cognitive impairment using convolutional neural network: temporal-feature maps of functional near-infrared spectroscopy. *Frontiers in Aging Neuroscience*, *12*, 141.

[123] M. Kashefpoor, H. Rabbani and M. Barekatain, "EEG signals from normal and MCI (mild cognitive impairment) cases", Nov. 2016, [online] Available: http://ww25.biosigdata.com/?download=eegsignals-from-normal-and-mci-cases.

[124] Delorme, A., Mullen, T., Kothe, C., Acar, Z. A., Bigdely-Shamlo, N., Vankov, A., & Makeig, S. (2011). EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Computational intelligence and neuroscience*, *2011*, 10-10.

[125] Farina, F. R., Emek-Savaş, D. D., Rueda-Delgado, L., Boyle, R., Kiiski, H., Yener, G., & Whelan, R. (2020). A comparison of resting state EEG and structural MRI for classifying Alzheimer's disease and mild cognitive impairment. *Neuroimage*, *215*, 116795.

[126] Hadiyoso, S., & Wijayanto, I. (2019, July). Noise removal in mild cognitive impairment eeg recording using empirical mode decomposition. In *2019 IEEE International Conference on Signals and Systems (ICSigSys)* (pp. 74-78). IEEE.

[127] Arifoglu, D., & Bouchachia, A. (2017). Activity recognition and abnormal behaviour detection with recurrent neural networks. *Procedia Computer Science*, *110*, 86-93.

[128] C. F. Ois, "Keras", Jan. 2015, [online] Available: https://github.com/fchollet/keras.

[129] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016, November). Tensorflow: a system for large-scale machine learning. In *Osdi* (Vol. 16, No. 2016, pp. 265-283).

[130] Hadiyoso, S., & Tati, L. E. (2018, December). Mild cognitive impairment classification using hjorth descriptor based on eeg signal. In *2018 International*

Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC) (pp. 231-234). IEEE.

[131] Ieracitano, C., Mammone, N., Hussain, A., & Morabito, F. C. (2020). A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks*, *123*, 176-190.

[132] Khatun, S., Morshed, B. I., & Bidelman, G. M. (2017, May). Single channel EEG time-frequency features to detect mild cognitive impairment. In *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* (pp. 437-442). IEEE.

[133] Ruiz-Gómez, S. J., Gómez, C., Poza, J., Gutiérrez-Tobal, G. C., Tola-Arribas, M. A., Cano, M., & Hornero, R. (2018). Automated multiclass classification of spontaneous EEG activity in Alzheimer's disease and mild cognitive impairment. *Entropy*, *20*(1), 35.

[134] Požar, R., Giordani, B., & Kavcic, V. (2020). Effective differentiation of mild cognitive impairment by functional brain graph analysis and computerized testing. *PLoS One*, *15*(3), e0230099.

[135] Al-Qazzaz, N. K., Ali, S. H. B. M., Ahmad, S. A., Islam, M. S., & Escudero, J. (2018). Discrimination of stroke-related mild cognitive impairment and vascular dementia using EEG signal analysis. *Medical & biological engineering & computing*, *56*, 137-157.

[136] Huggins, C. J., Escudero, J., Parra, M. A., Scally, B., Anghinah, R., De Araújo, A. V. L., ... & Abasolo, D. (2021). Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing. *Journal of Neural Engineering*, *18*(4), 046087.

[137] Poza, J., Gomez, C., Garcia, M., A Tola-Arribas, M., Carreres, A., Cano, M., & Hornero, R. (2017). Spatio-temporal fluctuations of neural dynamics in mild cognitive impairment and Alzheimer's disease. *Current Alzheimer Research*, *14*(9), 924-936.

[138] Timothy, L. T., Krishna, B. M., & Nair, U. (2017). Classification of mild cognitive impairment EEG using combined recurrence and cross recurrence quantification analysis. *International Journal of Psychophysiology*, *120*, 86-95.

[139] Alvi, A. M., Shaon, M. F. I., Das, P. R., Mustafa, M., & Bari, M. R. (2017, December). Automated course management system. In *2017 12th International*

Conference for Internet Technology and Secured Transactions (ICITST) (pp. 161-166). IEEE.

[140] Chollet, F. (2021). *Deep learning with Python*. Simon and Schuster.

[141] Du, J., Michalska, S., Subramani, S., Wang, H., & Zhang, Y. (2019). Neural attention with character embeddings for hay fever detection from twitter. *Health information science and systems*, *7*, 1-7.

[142] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, *147*, 70-90.

[143] Kelleher, J. D. (2019). *Deep learning*. MIT press.

[144] Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*, *281*(3), 628-641.

[145] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, *2018*.

[146] Vrbancic, G., & Podgorelec, V. (2018). Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks. *Elektronika ir Elektrotechnika*, *24*(4), 3-7.

[147] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253.

[148] Zhang, Z., Cui, P., & Zhu, W. (2020). Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *34*(1), 249-270.

[149] Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2019). A primer on deep learning in genomics. *Nature genetics*, *51*(1), 12-18.

[150] You, Z., Zeng, R., Lan, X., Ren, H., You, Z., Shi, X., ... & Hu, X. (2020). Alzheimer's disease classification with a cascade neural network. *Frontiers in Public Health*, *8*, 584387.

[151] Rodrigues, P. M., Bispo, B. C., Garrett, C., Alves, D., Teixeira, J. P., & Freitas, D. (2021). Lacsogram: A new EEG tool to diagnose Alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, *25*(9), 3384-3395.

[152] Oltu, B., Akşahin, M. F., & Kibaroğlu, S. (2021). A novel electroencephalography based approach for Alzheimer’s disease and mild cognitive impairment detection. *Biomedical Signal Processing and Control*, *63*, 102223.

[153] Meghdadi, A. H., Stevanović Karić, M., McConnell, M., Rupp, G., Richard, C., Hamilton, J., ... & Berka, C. (2021). Resting state EEG biomarkers of cognitive decline associated with Alzheimer's disease and mild cognitive impairment. *PloS one*, *16*(2), e0244180.

[154] Engedal, K., Barca, M. L., Høgh, P., Andersen, B. B., Dombernowsky, N. W., Naik, M., ... & Snaedal, J. (2020). The power of EEG to predict conversion from mild cognitive impairment and subjective cognitive decline to dementia. *Dementia and geriatric cognitive disorders*, *49*(1), 38-47.

[155] Pirrone, D., Weitschek, E., Di Paolo, P., De Salvo, S., & De Cola, M. C. (2022). EEG signal processing and supervised machine learning to early diagnose Alzheimer's disease. *Applied Sciences*, *12*(11), 5413.

[156] Patterson, C. (2018). World alzheimer report 2018.

[157] Morabito, F. C., Ieracitano, C., & Mammone, N. (2023). An explainable Artificial Intelligence approach to study MCI to AD conversion via HD-EEG processing. *Clinical EEG and Neuroscience*, *54*(1), 51-60.

[158] Chatterjee, C. C., & Krishna, G. (2019, September). A novel method for IDC prediction in breast cancer histopathology images using deep residual neural networks. In *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)* (pp. 95-100). IEEE.

[159] Fouladi, S., Safaei, A. A., Mammone, N., Ghaderi, F., & Ebadi, M. J. (2022). Efficient deep neural networks for classification of alzheimer's disease and mild cognitive impairment from scalp EEG recordings. *Cognitive Computation*, *14*(4), 1247-1268.

[160] Alvi, A. M., Siuly, S., & Wang, H. (2022, August). Challenges in electroencephalography data processing using machine learning approaches. In *Databases Theory and Applications: 33rd Australasian Database Conference, ADC 2022, Sydney, NSW, Australia, September 2–4, 2022, Proceedings* (pp. 177-184). Cham: Springer International Publishing.

[161] Goyal, P., & Chaurasia, V. (2017, April). Application of median filter in removal of random valued impulse noise from natural images. In *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)* (Vol. 1, pp. 125-128). IEEE.

[162] Prabu, S., Balamurugan, V., & Vengatesan, K. (2019). Design of cognitive image filters for suppression of noise level in medical images. *Measurement*, *141*, 296-301.

[163] Wang, G., Liu, Y., Xiong, W., & Li, Y. (2018). An improved non-local means filter for color image denoising. *Optik*, *173*, 157-173.

[164] Md. Taha, A. Q., & Ibrahim, H. (2020). Reduction of Salt-and-Pepper Noise from Digital Grayscale Image by Using Recursive Switching Adaptive Median Filter. In *Intelligent Manufacturing and Mechatronics: Proceedings of the 2nd Symposium on Intelligent Manufacturing and Mechatronics–SympoSIMM 2019, 8 July 2019, Melaka, Malaysia* (pp. 32-47). Springer Singapore.

[165] Halder, A., Sengupta, S., Bhattacharya, P., & Sarkar, A. (2020). Fast adaptive decision-based mean filter for removing salt-and-pepper noise in images. In *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019* (pp. 783-793). Springer Singapore.

[166] Lyakhov, P. A., Orazaev, A. R., Chervyakov, N. I., & Kaplun, D. I. (2019, January). A new method for adaptive median filtering of images. In *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)* (pp. 1197-1201). IEEE.

[167] Zhu, Y., & Huang, C. (2012). An improved median filtering algorithm for image noise reduction. *Physics Procedia*, *25*, 609-616.

[168] Yang, R., Yin, L., Gabbouj, M., Astola, J., & Neuvo, Y. (1995). Optimal weighted median filtering under structural constraints. *IEEE transactions on signal processing*, *43*(3), 591-604.

[169] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

[170] Morant, J. V., & Tataryn, L. (2017). Neurotherapies and Alzheimer: a protocol oriented review. *NeuroRegulation*, *4*(2), 79-79.

[171] Alvi, A. M., Siuly, S., De Cola, M. C., & Wang, H. (2022, October). DRAM-Net: A Deep Residual Alzheimer's Diseases and Mild Cognitive Impairment Detection Network Using EEG Data. In *Health Information Science: 11th International Conference, HIS 2022, Virtual Event, October 28–30, 2022, Proceedings* (pp. 42-53). Cham: Springer Nature Switzerland.

[172] Puri, D. V., Nalbalwar, S. L., Nandgaonkar, A. B., Gawande, J. P., & Wagh, A. (2023). Automatic detection of Alzheimer's disease from EEG signals using low-complexity orthogonal wavelet filter banks. *Biomedical Signal Processing and Control*, *81*, 104439.

[173] Perez-Valero, E., Lopez-Gordo, M. Á., Gutiérrez, C. M., Carrera-Muñoz, I., & Vílchez-Carrillo, R. M. (2022). A self-driven approach for multi-class discrimination

in Alzheimer's disease based on wearable EEG. *Computer Methods and Programs in Biomedicine*, *220*, 106841.

[174] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in biology and medicine*, *100*, 270-278.

[175] Kumar, A. (2021, September 15). *CNN Basic Architecture for Classification & Segmentation*. Data Analytics. Retrieved January 2, 2023, from https://vitalflux.com/cnn-basic-architecture-for-classification-segmentation/

[176] Sammut, C., & Webb, G. I. (2010). Leave-one-out cross-validation. *Encyclopedia of machine learning*, 600-601.

[177] Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*(11).

[178] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90.

[179] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

[180] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[181] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[182] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2020). Deep learning for 3d point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *43*(12), 4338-4364.

[183] Yang, L., Song, Y., Ma, K., & Xie, L. (2021). Motor imagery EEG decoding method based on a discriminative feature learning strategy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 368-379.

[184] Supriya, S., Siuly, S., Wang, H., & Zhang, Y. (2020). Automated epilepsy detection techniques from electroencephalogram signals: a review study. *Health Information Science and Systems*, 8, 1-15.

[185] Singh, R., Zhang, Y., Wang, H., Miao, Y., & Ahmed, K. (2020). Antisocial behaviour analyses using deep learning. In *Health Information Science: 9th*

International Conference, HIS 2020, Amsterdam, The Netherlands, October 20–23, 2020, Proceedings 9 (pp. 133-145). Springer International Publishing.

[186] Farsi, L., Siuly, S., Kabir, E., & Wang, H. (2020). Classification of alcoholic EEG signals using a deep learning method. *IEEE Sensors Journal*, 21(3), 3552-3560.

[187] Tawhid, M. N. A., Siuly, S., Wang, K., & Wang, H. (2023). Automatic and Efficient Framework for Identifying Multiple Neurological Disorders from EEG Signals. *IEEE Transactions on Technology and Society*.

[188] Tawhid, M. N. A., Siuly, S., Wang, K., & Wang, H. (2022, August). Brain Data Mining Framework Involving Entropy Topography and Deep Learning. In *Databases Theory and Applications: 33rd Australasian Database Conference, ADC 2022, Sydney, NSW, Australia, September 2–4, 2022, Proceedings* (pp. 161-168). Cham: Springer International Publishing.

[189] Deepan, P., Sudha, L. R., Kalaivani, K., & Ganesh, J. (2022). Scene Classification of Remotely Sensed Images using Optimized RSISC-16 Net Deep Convolutional Neural Network Model. *EAI Endorsed Transactions on Scalable Information Systems*, 9(5).

[190] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Wang, K. (2022). Convolutional neural network for multi-class classification of diabetic eye disease. *EAI Endorsed Transactions on Scalable Information Systems*, 9(4), e5-e5.