Diyala Journal of Engineering Sciences

Journal homepage: https://djes.info/index.php/djes

ISSN: 1999-8716 (Print); 2616-6909 (Online)

# An EEG Dataset for Alzheimer's Disease Patients in Iraq: Electrophysiological Recordings Across Cognitive Stages

Nigar M. Shafiq Surameery[1,2*], Abdulbasit Alazzawi[1], Aras T. Asaad[3], Abas Nariman Sidiq Tilako[4,6] and Sarwer Jamal Al-Bajalan[5,6]

[1]Department of Computer Science, College of Science, University of Diyala, Diyala, Iraq,
[2]Information Technology Department, College of Computer and Information Technology, University of Garmian, Kalar, Sulaimani, Kurdistan Region-Iraq,
[3]School of Computing, The University of Buckingham, Buckingham MK18 1EG, UK
[4]Department of Neurology and Stroke, Shorsh Military General Teaching Hospital,70th Forces, Sulaimaniyah, Iraq
[5]Department of Clinical Sciences, College of Medicine, University of Sulaimani, Sulaimaniyah, Iraq
[6]Department of Neurology, Shar Teaching Hospital, Sulaimani Directorate of Health, Sulaimaniyah, Iraq

**ARTICLE INFO**

***Article history:***
Received January 6, 2025
Revised March 14, 2025,
Accepted March 17, 2025,
Available online June 1, 2025

***Keywords:***
Alzheimer's Disease
EEG
Machine Learning
Deep Learning
AD Dataset

**ABSTRACT**

Alzheimer's disease (AD) is one of the common neurodegenerative disorders presenting with progressive cognitive decline and synaptic impairment. The need for accurate and accessible diagnostic methods is especially pressing in resource-limited regions. This study presents the first open electroencephalography (EEG) dataset for Alzheimer's Disease (AD) in Iraq, recording electrophysiological activity in four stages of cognition, namely, Healthy, Mild, Moderate, and Severe. The dataset comprised recordings from 53 participants and was recorded using a 40-channel EEG device based on the 10-20 electrode placement standard. Considering this EEG dataset, advanced preprocessing, such as the removal of artifacts, Independent Component Analysis (ICA), and Automatic Subspace Reconstruction (ASR), was performed to ensure high signal quality suitable for all state-of-the-art Machine Learning (ML) and Deep Learning (DL) analyses. Benchmark classification was conducted with a wide variety of ML and DL models, including Random Forest (RF), Gradient Boosting methods, Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and Long Short-Term Memory networks (LSTMs), showing promising results-a maximum of 96.85% by Random Forest within ML techniques and a maximum of 96.05% by DNNs in DL techniques. This dataset fills a critical gap in regional AD research and moves toward developing low-cost, noninvasive diagnostic tools. Future work may be performed on more extended and more diverse datasets with more sophisticated multimodal approaches for improving AD diagnosis and early intervention. The dataset can be downloaded from https://shorturl.at/Z0b8D.

## 1. Introduction

Alzheimer's is a serious and growing global health concern. This common neurodegenerative disorder is characterized by synaptic dysfunction along with loss of learning and cognitive abilities, consequently leading to a progressive decline in mental and cognitive functions. The eventual development of Alzheimer's may result in death [1].

It is befitting to note that early diagnosis of AD has shifted the paradigm in the stages of disease intervention and improved the efficacy of treatment. Several diagnostic tools and

* Corresponding author.
E-mail address: nigar.mahmoud@garmian.edu.krd
DOI: 10.24237/djes.2024.18203

technologies were employed in diagnosing AD, including neuroimaging and cerebrospinal fluid markers, PET, MRI, and EEG [2].

EEG has become essential in the clinical and scientific study of neurological disorders. It is a non-invasive method that transposes electrical activity generated in the cerebral hemispheres into signals that may provide changes in the cortex and even in deeper tissues. Given its portability, cost, and non-invasive advantages, EEG has been employed to analyse different brain diseases, including Alzheimer's disease [3].

The normal resting-state EEG still shows rather significant intrasubject variability because of interference from various sources, as well as the current state of the brain. The major challenges that can be found with the use of EEGs in diagnosing Alzheimer's disease are noise, the non-ergodic nature of the signal, a nonlinear underlying process, and the inter-individual variability regarding gender, age group, stress, and other health conditions of the patients. These factors can complicate the interpretation and analysis of the EEG record and may lead to diagnostic inaccuracies, hence complicating the differentiation of Alzheimer's disease from other neurological disorders [4].

Consequently, research now focuses on devising and enhancing sophisticated methods of processing and analyzing EEG data to achieve higher accuracy and reliability in diagnosing Alzheimer's disease. It, therefore, involves applying signal processing methods, machine learning, and deep learning techniques, among others, with statistical methods to excavate informative features from such complex, time-varying EEG signals. In Iraq, Alzheimer's disease is a rapidly increasing disorder among aged people, with many yet undiagnosed. Neurologists widely use either neuroimaging or morphological tests, which depend either on radiation exposure or are costly. These tests also bear high potential risks or side effects to the patients depending on the test used to detect Alzheimer's [5].

Therefore, the primary motivation for this study was to collect EEG recordings from elderly patients with Alzheimer's disease across different stages (Mild, Moderate, and Severe) as well as healthy age-matched controls during the resting state. We present the first publicly available EEG dataset for AD diagnosis in Iraq, aiming to enhance the understanding of Alzheimer's disease, particularly in its early stages. This dataset serves as a benchmark for identifying EEG-based biomarkers for diagnostic and prognostic purposes. By providing a high-quality dataset and demonstrating its effectiveness in ML and DL models, this work contributes to the development of noninvasive, cost-effective diagnostic tools for AD, particularly in regions with limited access to advanced neuroimaging techniques.

This study makes many important contributions:

1. The first Alzheimer’s disease EEG dataset in Iraq: The paper outlines the publicly available dataset fills a serious gap in local research. The database consists of 53 recordings from subjects that capture a range of cognitive states.
2. Comprehensive Preprocessing Pipeline: The dataset undergoes extensive preprocessing using Independent Component Analysis (ICA) and Automatic Subspace Reconstruction (ASR) to remove noise and artifacts, ensuring high-quality signals for analysis.
3. Benchmarking with ML/DL Methods: The data is evaluated with a variety of machine learning (ML) and deep learning (DL) models, hence determining benchmark classification performances. In the case of ML models, Random Forest achieves a 96.85% accuracy, with Deep Neural Networks having a 96.05% accuracy in the DL methods.
4. Advancing EEG Based AD Classification**:** The results demonstrate that EEG signals contain distinctive patterns across AD stages, highlighting the potential of EEG-based biomarkers for reliable early diagnosis.

This dataset paves the way for developing innovative methods for achieving a definitive diagnosis of Alzheimer's within the regions where the traditional methods remain costly or not accessible.

## 2. Related work

This section provides an overview of EEG research datasets used for classifying Alzheimer's disease and sets the background in which this work is situated. Most publicly available EEG datasets focus on specific dimensions of research, such as the state of rest, olfactory stimulation, or tests through cognitive tasks. Such datasets are of utmost importance during the training and testing of machine learning models for diagnosing AD, enhancing reproducibility, and supporting comparative research in computational neuroscience.

The dataset of EEG by [5] includes 36 AD patients, 23 FTD patients, and 29 healthy control subjects. In the recording, 19 scalp electrodes were positioned according to the 10-20 system, and the data was digitized at a sampling frequency of 500 Hz. The dataset contains raw and preprocessed EEG signals; the preprocessing includes band-pass filtering from 0.5 to 45 Hz, artifact removal, and ICA. It is unique in providing BIDS-format EEG recordings specific to dementia conditions and hence finds great importance in machine learning and the analysis of brain activity.

The current study by [6] introduces a significant dataset on the electrophysiological processing of the brain to olfactory stimuli in AD and MCI. The dataset comprises recordings from 13 patients diagnosed with AD, 7 patients diagnosed with MCI, and 15 healthy elderly participants. EEG recordings were made with Fp1, Fz, Cz, and Pz channels over 120 trials of olfactory stimulation with rose and lemon odorants. Further filtering was done between 0.5-40Hz, ICA-based eye blink removal, followed by removing epochs containing those artifacts by hand. The dataset provided here is unique in its investigation of smell dysfunction and neurodegenerative disease linkage, giving insights into the potential early biomarkers of AD.

However, Authors in [7] proposed a casual EEG data set to illustrate the capability of QGs as a potential new diagnostic tool for Alzheimer's disease. The abovementioned data included EEG records from the 19 scalp electrode positions of 24 AD patients and 24 healthy elderly controls; 8-second segments sampled at 128 Hz. This dataset is valuable in assessing various network metrics, based on QG, to discriminate EEG activity in healthy individuals from AD patients, providing a valuable benchmark for signal processing and classification studies.

Another EEG dataset presented by[8] includes three different EEG datasets used for phase-based functional connectivity analysis in studying Alzheimer's disease and other cognitive conditions.

First, the data for resting-state EEG with open and closed eyes are recorded of 16 healthy volunteer record-ings-channel EEGs. Then, continuous EEG data will be down-sampled to 200 Hz, referenced to an average reference, and alpha-filtered within the 8-13 Hz range. Clean 5-second segments for each subject—precisely 1000 to 2000 samples—will be extracted since such segments have always proved reliable for studying resting-state brain connectivity.

The other is the VSTM binding task data, which consists of EEG recordings from 19 healthy young volunteers using a 30-channel EEG setup. Continuous EEG signals were sampled at 250 Hz and filtered between 0.01-40 Hz with a band-pass filter, resulting in 1-second epochs that yielded, on average,65.7 ± 9.27 trials per subject. The connectivity analysis focused on the beta band (13–32 Hz), using PLI to estimate functional connectivity during the VSTM task.

The third dataset deals with Alzheimer's disease. In that, EEG records were taken from 12 Alzheimer's disease patients aged 72.8 ± 8.0 years, along with 11 healthy control subjects. Recordings were obtained with a 16-channel EEG device accordinRecordings0-20 system, and the data acquisition was performed at the University Hospital of Valladolid in Spain. Artifact removal was done by visual inspection, and clean 5-second epochs were extracted-28.8 ± 15.5 per subject for further processing. The EEG signals were filtered separately for the alpha and beta bands: 8-13 Hz and 13-32 Hz, respectively. The EEG PLI adjacency matrices comprising MATLAB functions and scripts concerning network analysis are included in this dataset, with the raw EEG data, thus allowing

further studies relating to Alzheimer's disease and connectivity network analysis.

Another EEG data is contributed by [9] that involves EEG activity recorded for 59 patients suffering from AD (MMSE score = 10–19), 7 patients suffering from MCI, and 102 healthy control participants. For the recording of EEG activity, a 21-channel EEG system was used: Walter EEG PL-231, Germany - 256 Hz and another 21-channel configuration: "TruScan" 32, Alien Technik Ltd., Czech Republic - 128 Hz. Data recorded at 256 Hz were down-sampled to 128 Hz. All electrodes were placed according to the 10-20 international system. EEG pre-processing consisted of detrending, notch filtering at 50 Hz, removal of all myogenic potentials, eye movements, ECG artifacts, and electrode artifacts by an experimenter, and segmentation into nonoverlapping 7.8125-s segments (1000-time samples). It has been used to test the performance of a novelty detection approach for diagnosing AD and MCI using features extracted by adaptive filtration with a linear neural unit. This dataset could support multi-device analyses and robustness against de-vice-specific variability.

Another EEG dataset was also presented in the research by[10] for recordings conducted on 160 probable AD patients and 24 healthy participants. The whole dataset involves four groups of subjects: (A) 12 healthy elderly with eyes open, (B) 12 healthy elderly with eyes closed, (C) 80 probable AD patients with eyes open, and (D) 80 probable AD patients with eyes closed. EEG was recorded using 19 scalp electrodes according to the international 10-20 system, namely, Fp1, Fp2, F3, F4, F7, F8, Fz, C3, C4, Cz, P3, P4, Pz, T3, T4, T5, T6, O1, O2. Each segment of ongoing EEG was 8 seconds and was sampled at 128 Hz frequency. A signal was filtered within a frequency range of 0.5-30 Hz. Those signals resulting from movements were removed. This dataset was generously provided free of cost by Dr. Dennis Duke, and it finds extensive use in the computational analysis of EEG signals related to the classification of Alzheimer's disease using various time-series analysis techniques such as wavelet coherence, fractal dimension, quadratic entropy, wavelet energy, quantile graphs, and visibility graphs.

To the best of our knowledge, no EEG dataset on AD classification exists in Iraq. This research will, therefore, propose a pioneering effort to establish the first publicly available EEG dataset for AD classification in Iraq, contribute to the global research community, and, by doing so, develop diagnostic tools for this region.

## 3. Experimental design

The following figure depicts the general structure that the experimental design has taken in the present work. The basic steps comprising the creation and analysis of the EEG dataset on Alzheimer's Disease are depicted: data collection, preprocessing, feature extraction, model selection, and evaluation. Each of these individual steps has a significance of its own in order to insure the integrity, accuracy, and effectiveness of the classification models elaborated in the current work.

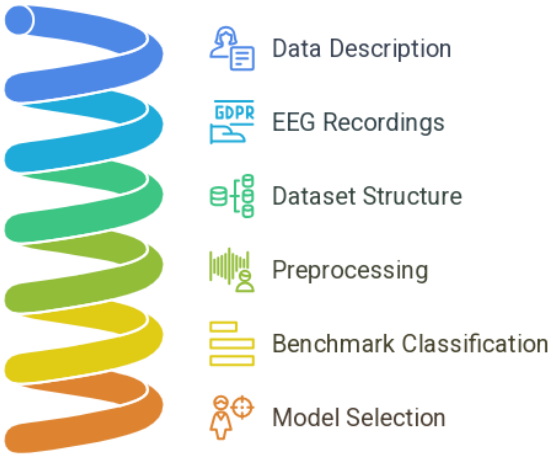


**Figure 1.** EEG data Collection and analysis for AD

### *3.1 Data description*

This dataset contains the EEG resting state-open and closed eyes recordings from 53 subjects in total. A total of 8 of them were diagnosed with Alzheimer's disease (Mild group), 12 were diagnosed with (Moderate group), 10 were diagnosed with (Severe group), and 23 were Normal cases.

### 3.2 Participants

Participants were recruited from individuals who sought treatment for memory complaints at the memory clinic at “Shar Hospital,” “Shorsh Hospital,” “Anwar Shekha Private Hospital,” “Dr. Abbas Nariman” Clinics, and “Dr. Sarwer Jamal Al-Bajalan” Clinics, Sulaymaniyah, Kurdistan region, Iraq. There were 53 participants aged 56 to 85, including 2 in their 50s, 9 in their 60s, 32 in their 70s, and 10 in their 80s. Further, there were 9 male participants and 44 female participants. Additional information, such as gender and disease stage, is contained in Table 1.

**Table 1:** Participant Description. In the Gender column, F indicates female and M indicates male

| | Participant | Gender | Age | Group |
|---|---|---|---|---|
| **1** | Mild1 | F | 65 | Mild |
| **2** | Mild2 | F | 70 | Mild |
| **3** | Mild3 | M | 81 | Mild |
| **4** | Mild4 | F | 71 | Mild |
| **5** | Mild5 | F | 70 | Mild |
| **6** | Mild6 | F | 75 | Mild |
| **7** | Mild7 | F | 71 | Mild |
| **8** | Mild8 | F | 77 | Mild |
| **9** | Mod1 | F | 74 | Moderate |
| **10** | Mod2 | F | 72 | Moderate |
| **11** | Mod3 | F | 82 | Moderate |
| **12** | Mod4 | F | 70 | Moderate |
| **13** | Mod5 | F | 74 | Moderate |
| **14** | Mod6 | F | 74 | Moderate |
| **15** | Mod7 | F | 72 | Moderate |
| **16** | Mod8 | M | 82 | Moderate |
| **17** | Mod9 | F | 70 | Moderate |
| **18** | Mod10 | F | 82 | Moderate |
| **19** | Mod11 | F | 79 | Moderate |
| **20** | Mod12 | F | 60 | Moderate |
| **21** | Sv1 | F | 56 | Severe |
| **22** | Sv2 | F | 74 | Severe |
| **23** | Sv3 | F | 72 | Severe |
| **24** | Sv4 | F | 85 | Severe |
| **25** | Sv5 | M | 79 | Severe |
| **26** | Sv6 | F | 74 | Severe |
| **27** | Sv7 | F | 70 | Severe |
| **28** | Sv8 | F | 72 | Severe |
| **29** | Sv9 | F | 70 | Severe |
| **30** | Sv10 | F | 79 | Severe |
| **31** | CN1 | M | 66 | Cognitive Normal |
| **32** | CN2 | F | 88 | Cognitive Normal |
| **33** | CN3 | M | 77 | Cognitive Normal |
| **34** | CN4 | F | 81 | Cognitive Normal |
| **35** | CN5 | F | 62 | Cognitive Normal |
| **36** | CN6 | F | 77 | Cognitive Normal |
| **37** | CN7 | F | 83 | Cognitive Normal |
| **38** | CN8 | M | 75 | Cognitive Normal |
| **39** | CN9 | F | 59 | Cognitive Normal |
| **40** | CN10 | F | 66 | Cognitive Normal |
| **41** | CN11 | F | 69 | Cognitive Normal |
| **42** | CN12 | M | 81 | Cognitive Normal |
| **43** | CN13 | F | 62 | Cognitive Normal |
| **44** | CN14 | F | 71 | Cognitive Normal |
| **45** | CN15 | M | 67 | Cognitive Normal |
| **46** | CN16 | F | 74 | Cognitive Normal |
| **47** | CN17 | F | 76 | Cognitive Normal |
| **48** | CN18 | F | 62 | Cognitive Normal |
| **49** | CN19 | M | 71 | Cognitive Normal |
| **50** | CN20 | F | 74 | Cognitive Normal |
| **51** | CN21 | F | 70 | Cognitive Normal |
| **52** | CN22 | F | 83 | Cognitive Normal |
| **53** | CN23 | F | 68 | Cognitive Normal |

### 3.3 EEG Recordings

In the present study, EEG recordings were made on patients with AD to extend the dataset and enable in-depth analysis. For the research study, strict adherence to EEG recording standards was always followed to ensure that the data gathered was valid and that the methodology outlined a plan for acquiring reliable EEG signals.

EEG activity was recorded utilizing a "Compumedics Profusion" EEG system coupled with a saline-based EEG cap; electrode placement followed the International 10–20 System. Saline electrodes have advantages over other types in clinical studies: they are easy to operate, their placement is reliable, and they cause less discomfort to the subjects. The scalp was mildly chemically abraded before recording to ensure good electrode-to-skin conductivity, ensuring optimal impedance levels. After fitting the cap, connection cables were checked for stability, and subjects were allowed a rest period of 15-30 minutes to adapt to the setup, which is essential to minimize movement artifacts and ensure relaxed brain activity [11].

We employed a 40-channel EEG system comprising standard 10-20 electrodes and additional custom-made electrodes to provide complete scalp coverage. This results in a wider

recording of the brain activity and greater spatial detail than that provided by the standard configuration. The electrode array consisted of 19 conventional 10-20 scalp electrodes positioned at the following locations [10]:

- Frontal electrodes: Fp1, Fp2, F7, F3, Fz, F4, and F8.
- Temporal electrodes: T3, T4, T5, T6
- Central electrodes: C3, Cz, C4
- Parietal electrodes: P3, Pz, P4
- Occipital electrodes: O1, O2

Besides these standard positions, 21 additional reference electrodes were placed to increase the spatial resolution and informational richness. This allows data to be captured from the standard and custom electrodes for high-resolution EEG in a broader range of scalp regions. Therefore, the subtler, more localized neural activity that might be an early sign of Alzheimer's disease would be more easily detected. Enhanced spatial resolution can allow a better investigation of more complex waves, especially in frequency bands related to cognitive decline, hence enabling the possibility of an earlier diagnosis [12]. With wider spacing, there is reduced noise and increased signal clarity, hence reliable data for clinical and analytic purposes. These extra electrodes provide excellent coverage over the brain, particularly extending over scalp areas that were previously poorly represented. This allows the precise mapping of changes that involve Alzheimer's disease concerning the posterior parietal and temporal lobes[13].

This setting provides several advantages in investigating Alzheimer's disease and many other neurodegenerative disorders; thus, it is a handy tool for future research. Recordings were made at a sampling rate of 250 Hz with an amplitude resolution of 10 μV/mm, and sessions were about 30 minutes long. This careful setup provides one example of how the study of the detailed EEG biomarkers of Alzheimer's disease may enable model improvements for the diagnosis of neurodegenerative disorders.

### *3.4 Dataset Structure*

This dataset is collected and prepared for EEG-based Diagnosis-related research on Alzheimer's Disease. It includes four categories: Healthy, Mild, Moderate, and severe AD. The dataset is provided in four separate folders. Each folder represents one category, and several subject EEGs are recorded within that specific folder. The dataset consists of 15,820,760 samples, thus offering broad coverage for different tasks in machine learning and deep learning.

The recordings were exported to. edf (European Data Format) for the standardization of neuroimaging; after that, these signals needed to be filtered and prepared for analysis in CSV format.

The data was further refined into a structured format and saved in a file named “HMMS”, which is very helpful in machine learning and deep learning processes. This unified structure permits higher consistencies along many categories, making integrating the data into the machine and deep learning pipelines relatively easy.

The dataset is organized in the following structure:

1. Healthy Folder: Data from 23 subjects representing cognitively normal controls.
2. Mild Folder: Recordings from 8 subjects’ representatives of Mild AD.
3. Moderate Folder: These are data from 12 subjects, representing those with Moderate AD.
4. Severe Folder: Recordings from 10 subjects’ representative of subjects with Severe AD.

The data for every subject consists of the EEG recordings already preprocessed into a uniform format readily usable for classification.

The HMMS file provided to be used to study brain activity patterns across the stages of AD. It has become a valuable resource for various applications, ranging from identifying EEG biomarkers of AD to creating diagnostic models. The availability of structured HMMS file allows for easy integration with machine learning and deep learning pipelines.

## 4. Methods

### *4.1. Preprocessing*

This work applies a proper preprocessing and analysis pipeline to the extracted EEG signals to guarantee high-quality, clean, and

well-structured data for machine learning and deep learning models. A structured overview of the EEG preprocessing workflow is illustrated in Figure 2.

Preprocessing first included artifact removal and signal conditioning, which had already been performed directly on the recording EEG device. The FIR filter was used with cutoff frequencies of 1 Hz and 30 Hz, allowing neural activity to be in focus within this range. This was followed by a 50 Hz notch filter to remove power-line interference. For further enhancement of signal clarity, RMS windowing was used to measure the signal's average power in small, sliding windows, thereby helping reduce noise and hence offering a greater clarity signal. For visualization, the temporal scale was set to 10 seconds per page so that a detailed analysis of specific segments of interest in the EEG was possible, and the required uniform epoch windows were provided throughout the complete analysis [10]. These are hardware-oriented preprocessing steps, thus serving as a basis for subsequent software-oriented preprocessing.

The total samples for the entire dataset were 15,820,760, taken from a total of four categories: Healthy, Mild, Moderate, and Severe AD. First, the preprocessed signals were exported in the ".edf " file format and then in CSV format for further processing.

Afterward, cleansing and column alignment were done with the obtained EEG signals. Each was checked to confirm whether they had the same structure: each file should contain 40 features. Those EEG files with less than 40 columns were padded with columns of zeros to make them reach the predefined dimensionality. Further, index columns, duplicate entries, and samples containing missing values were filtered out to keep the data clean. Label encoding was performed for categorical labels "Healthy," "Mild," "Moderate," and "Severe" into their numerical labels 0, 1, 2, and 3, respectively.

Following the preprocessing steps above, Actual ICA was performed using the following steps using the "fastICA" algorithm from the sci-kit-learn library [14]: EEG signals were analysed without the label column to extract 40 independent components using the "fastICA" model with a fixed random state for reproducibility, set as 42. This was the step to segregate the electrical and mechanical artifacts like eye blinks and muscle movements from neural activity. The first five features were plotted against the first five ICA components, which would give a good visual inspection of the original signals and their corresponding ICA-transformed components, as clear in Figure 3. This transformation helps separate neural activity from artifacts such as eye blinks and muscle movements, ensuring cleaner EEG signals for analysis. Further, to compare the structure of the original EEG signals and the ICA-transformed components, some insight into the spatial distribution of independent components has been obtained by creating circular polar plots for visualization. The shape of the original EEG data and the ICA-transformed data has been printed to validate the correctness of the decomposition process. Figure 4 illustrates how ICA decomposition captures distinct spatial patterns across EEG electrodes, helping differentiate between neural activity and noise sources.

After the ICA step, ASR was applied to refine the EEG signals further and remove persistent artifacts [15]. First, the ASR process normalizes the EEG signals with "StandardScaler" to have zero mean and unit variance. Then, the covariance matrix is computed on the first window of clean EEG data, and the eigenvalues and corresponding eigenvectors are calculated to build the spatial filter. Afterward, the EEG data were processed in sliding windows, each projected into the subspace defined by the eigenvectors. Then, Z-scores were calculated for each component, and components with a Z-score higher than 20 were considered artifacts and assigned a value of zero. To smooth transitions across windows, the overlapping segments were linearly interpolated to ensure coherence in the reconstruction of the EEG signal.

Then, the cleaned data was converted back to the original format of the EEG, and the "StandardScaler" inverse transformation was applied to bring it to its actual scale. This step also involved removing duplicate entries and filling in missing values to clean and quality the

final EEG signals. The cleaned EEG data was further used in machine learning and deep learning analysis.

The EEG signals were preprocessed, followed by further segmentation and storage of the data in a file named "HMMS." Each record of the individual was structured so that the label column was kept separate and distinct from the input features. For the HMMS file, specific sample ranges, such as rows 15,000 to 30,000, were chosen to have equal representation across all the other three classes except "Healthy." there will be an equal amount of samples so that problems concerning class imbalance can be minimal. The final HMMS file was maintained to easily be integrated into various machine learning workflows and deep learning architectures.

Finally, a correlation matrix was performed on the preprocessed dataset to check feature relationships. The heatmap representation in the Seaborn library was used, which highlights potential associations between EEG features. As shown in Figure 5, this correlation matrix helps in understanding how different EEG channels interact, which is crucial for feature selection. Identifying strongly correlated features can reduce redundancy and improve classification performance in machine learning models, leading to more efficient EEG-based AD diagnosis [16].

This whole pipeline for preprocessing and preparation involved device-level filtering, RMS windowing, structuring into epochs, data alignment, ICA decomposition, ASR-based artifact removal, feature engineering, and correlation analysis. It prepared a clean, structured, and high-quality EEG dataset. Such standardized data is critical to developing sophisticated classification models that aspire to differentiate between the healthy, mild, moderate, and severe stages of Alzheimer's disease.

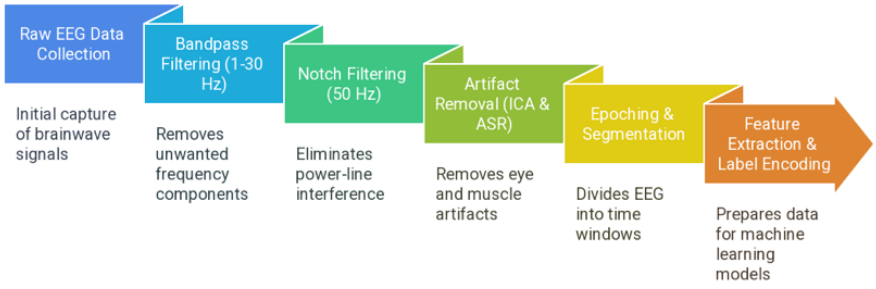


**Figure 2**. EEG processing workflow

### *4.2. Benchmark Classification*

Different simple classification techniques are used here that any other researcher can reproduce, verify, or extend quite easily to set a classification benchmark for the dataset. Thus, other, more advanced approaches might turn out even better. Of course, the substantial interest here was to provide an easily observable and reproducible baseline.

In this work, we have experimented with 20 different ML/DL models to classify the subjects belonging to four classes-Mild, Moderate, and Severe AD and CN subjects.

This work evaluates ten different baseline ML models. Some of these baseline models are RF, Gradient Boosting, AdaBoost, XGBoost, CatBoost, LightGBM, Naive Bayes, Decision Tree, SVM, and KNN, whereas from another side, the DL techniques comprise DNN, LSTM, CNN, GRU, Bi-LSTM, VGG16, LSTM-XGBoost, CNN-LSTM, CNN-SVM, and CNN-DT. These are selected as they have been proven to work effectively in disease diagnosis. Accordingly, all the aforementioned baseline models were exploited in this paper as a TL model by modifying the last output layer to be suitable for the number of classes being used in the experiment [17]. As a result, all models undertook 10-fold cross-validation training and testing processes to assure the robustness of the obtained models .

*4.3 Model Selection and Justification*

This research implements a broad range of Machine Learning and Deep Learning models for the classification of EEG signals to predict the stages of AD. These models have been chosen because they handle the complexity of EEG data well and are proven to perform well in similar classifications. These models perform better with many dependencies on parameters and hyperparameters. Further model performance improvements are made based on the results from hyperparameter tuning. Table 2 and Table 3 elaborately overview the different parameters and hyperparameters used in any ML/DL model. These elements must be properly tuned and selected to enhance model accuracy and efficiency.

*4.3.1 Machine Learning Models:*

Following are some of the Machine Learning methods utilized to classify EEG signals into AD stages effectively:

- **Random Forest** (RF) that makes multiple decision trees for better performance and improved generalization. RF handles the variations in EEG signals to maintain the consistency of the classification.
- **K-Nearest Neighbours** (K-NN) - a non-parametric algorithm that uses proximity among the features to classify EEG data. K-NN is well-known for its simplicity and interpretability.
- **Support Vector Machine (SVM)** - This is well-known to extract optimal hyperplanes to separate classes in multi-dimensional feature spaces, which is very effective in the classification of EEG signals [18].
- **XGBoost -** Optimized the gradient boosting framework by reducing overfitting and speeding up the training.
- **LightGBM** - This is a gradient boosting algorithm designed for handling large-scale data with superior efficiency and accuracy.
- **CatBoost**: Particularly able to handle categorical features and thus more convenient for EEG data with a mix of feature types.
- **AdaBoost**: It aims to improve the classification for the misclassified instances when creating an ensemble of weighted weak classifiers.
- **Naive Bayes:** Probabilistic model with an assumption of independence among the EEG features, and therefore useful for comparisons of baselines.
  **Decision Tree**: Simple and easy-to-interpret method for classifying EEG signals based on predefined criteria.

*4.3.2 Deep Learning Models:*

Several DL techniques applied attempt to exploit the complexities of EEG data at both sequential and spatial levels [19]:

- **Deep Neural Network(DNN):** Multi-layer perceptron that mines complex patterns of EEG signals to classify them as accurately as possible.
- **CNN**: Attempts to exploit the convolutional layers in the extraction of the spatial features from EEG data and identify crucial biomarkers of AD.
- **LSTM**: A type of recurrent network that tries to model temporal dependencies in EEG sequences that are significant to the differentiation between AD stages.
  For LSTM-based models, the EEG data is structured as a 3D tensor with dimensions (samples, time steps, features), where each sequence maintains temporal dependencies essential for deep learning models to learn from EEG patterns over time [17].
- **Bi-LSTM**: An extension of LSTMs by including past and future dependencies and thus modeling temporal relationships and improving the subtle variations in EEG.
- **GRU**: A simplified architecture of RNNs that processes the EEG sequential data using fewer parameters than the LSTMs.
- **CNN-LSTM**: This integrates CNN feature extraction with LSTM modeling

to capture effectively the spatial temporal relationships in EEG.

- **CNN-SVM**: It uses CNN for feature extraction, but the classification is done by SVM, thus increasing the performance with simultaneous interpretability.
- **LSTM-XGBoost**: Couples LSTM learning with gradient boosts from XGBoost to provide a strong and robust classification.
- **CNN-DT**: Integrates CNN feature extraction with the rule-based classification of DT for interpretability.
- **VGG16:** This is the deep CNN adapted to this work to allow hierarchical feature extraction for the diagnosis of AD using EEG signals [20].

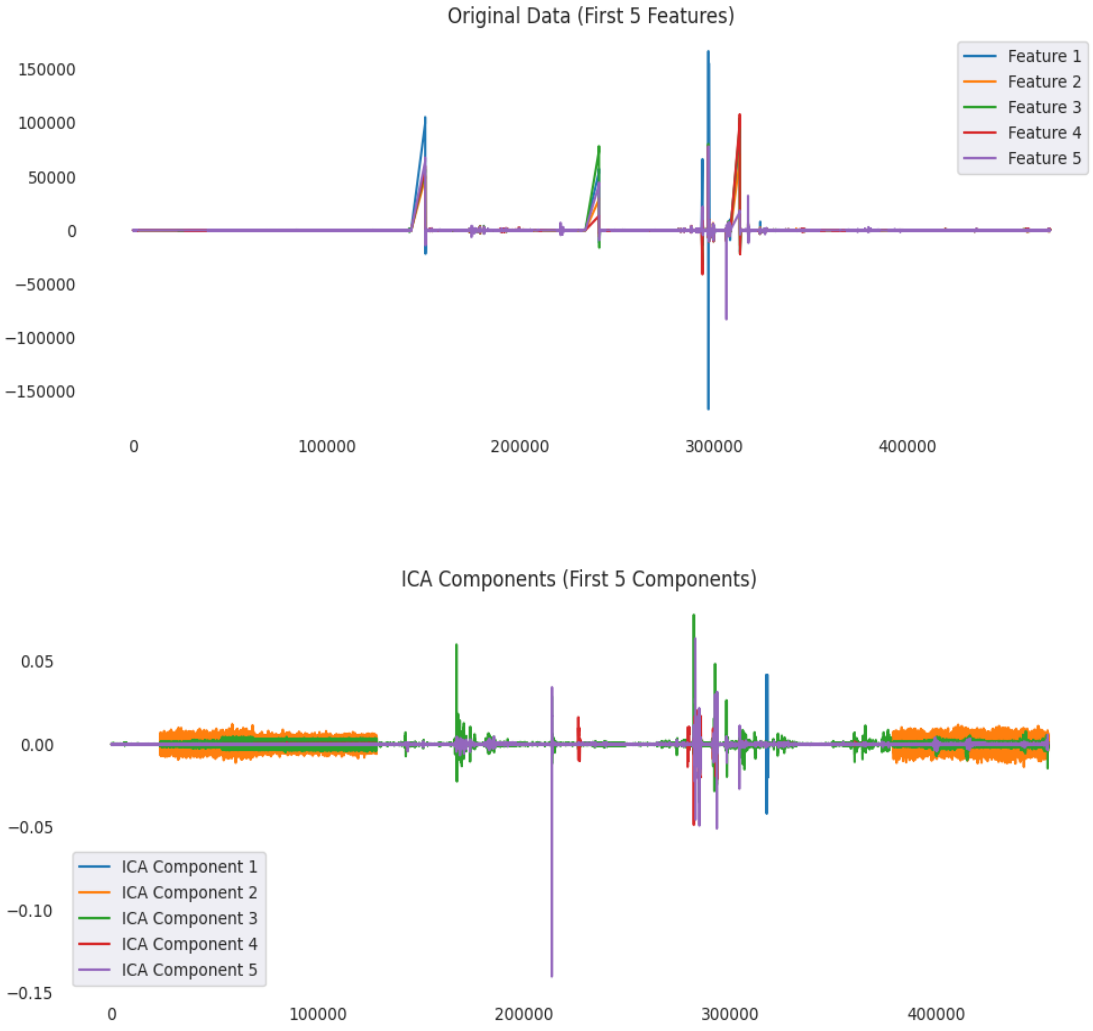


**Figure 3**. The original signals and their corresponding ICA-transformed components

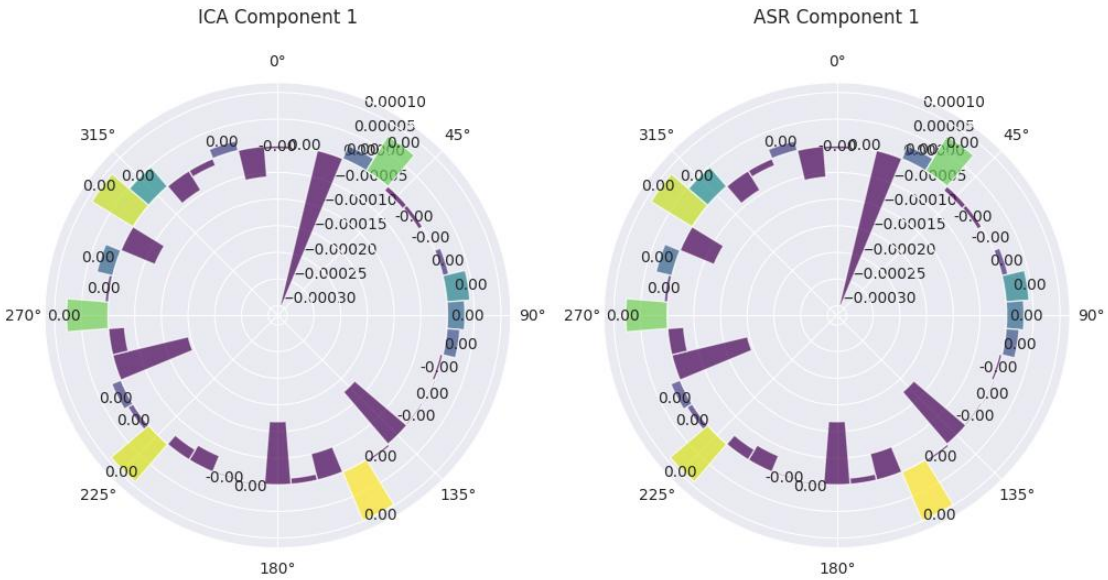


**Figure 4.** Circular polar plots of the original signals and ICA-transformed components

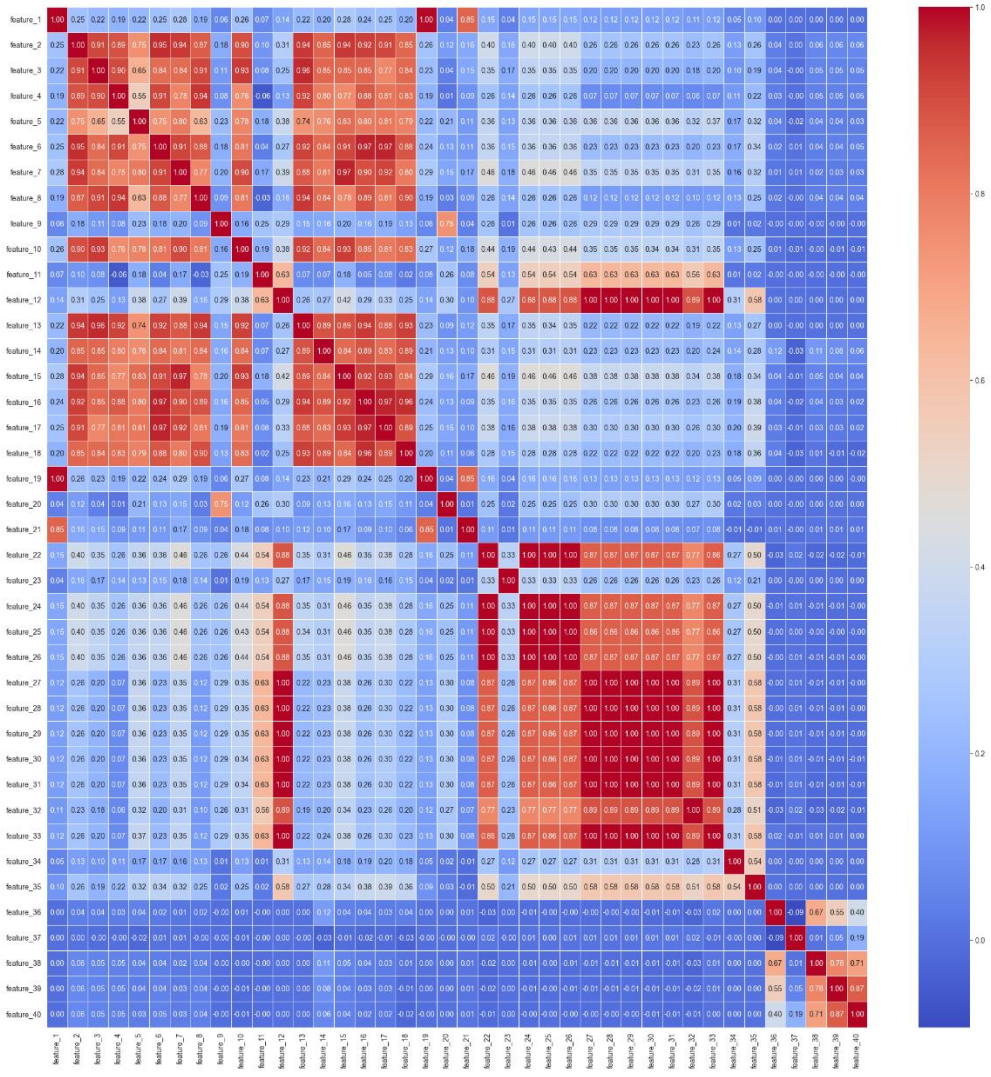

**Figure 5.** Correlation matrix for the data

**Table 2:** The parameters and values of the various implemented ML models

| Model | Parameter | Value(s) |
|---|---|---|
| **RF** | Max Depth<br>Number of Trees | 8<br>4000 |
| **LightGBM** | Max Depth<br>Number of Trees | 8<br>100 |
| **AdaBoost** | Max Depth<br>Learning Rate | 8<br>0.01 |
| **SVM** | Cost factor (C)<br>Cache_size<br>Max_iter | 25<br>200<br>-1 |
| **DT** | Max Depth | 8 |
| **Logistic Regression** | Max Iter | 100 |
| **K-NN** | N-neighbors | 10 |
| **Naive Bayes** | N/A | N/A |
| **XGBoost** | Max Depth<br>Number of Trees<br>Learning Rate | 8<br>100<br>0.01 |
| **CatBoost** | Max Depth<br>Number of Trees<br>Learning Rate | 8<br>100<br>0.01 |

**Table 3:** The parameters and values of the various implemented DL models

| Model | No. of Layers | No. of Units | Loss Function | Optimizer | LR | No. of Epochs |
|---|---|---|---|---|---|---|
| **DNN** | 6 | 64, 128, 256, 512, 256, 128 | MAP | Adam | 0.001 | 15 |
| **LSTM** | 6 | 64, 128, 256, 128, 64, 32 | MSE | Adam | 0.001 | 15 |
| **GRU** | 5 | 64, 128, 256, 128, 64 | MSE | Adam | 0.001 | 15 |
| **CNN** | 5 | 32, 64, 128 | MAP | Adam | 0.001 | 15 |
| **VGG16-like** | 16 | Configured based on layers | MAP or Categorical Crossentropy | Adam | 0.001 | 15 |
| **Bi-LSTM** | 7 | 64, 128, 256, 512, 256, 128, 64 | MSE | Adam | 0.001 | 15 |
| **CNN-LSTM** | 5 (CNN) + 1 (LSTM) | 32, 64, 128 (CNN), 64, 128, 256 (LSTM) | MSE | Adam | 0.001 | 15 |
| **LSTM-XGBoost** | 6 (LSTM) + 1 (XGBoost) | 64, 128, 256, 128, 64, 32 (LSTM) + XGBoost hyperparameters | MSE | Adam | 0.001 | 15 |
| **CNN-SVM** | 5 (CNN) + 1 (SVM) | 32, 64, 128 (CNN) | MAP (CNN), Hinge loss (SVM) | Adam (CNN) | 0.001 | 15 |
| **CNN-DT** | 5 (CNN) + 1 (DT) | 32, 64, 128 (CNN) | MAP (CNN), Gini Index or Entropy (DT) | Adam (CNN) | 0.001 | 15 |

## 5. Model evaluation

### *5.1 Experimental Setup*

The experiments have been conducted using Google Colab - a Google cloud-based platform with GPUs, which provide an efficient environment for operating Machine Learning and Deep Learning techniques. The appropriate computing capabilities of the platform have allowed the usage and training of different models without the need for a dedicated local setup.

Due to the diversity of the ML and DL models used in this work, some of the training parameters-such as the mini-batch size, the learning rate, and the number of epochs-change depending on the model requirements. In any case, these parameters were optimized during training to reach the best results for each model. This flexible approach would ensure the models are tuned enough to cope with such a challenging task of EEG and to classify the different stages of Alzheimer's Disease accurately.

### *5.2 Confusion Matrix*

The confusion matrix employed to assess the performance of the proposed network architectures, as it is one of the most critical tools summarizing the outcome of classification tasks. It is an ideal tool to have in terms of counting correct predictions against incorrect predictions for highlighting the model's real performance [21].

The confusion matrix classifies predictions into four key components:

**True Positive (TP):** Both predicted and actual outcome are positive.

**False Positive (FP):** The Predicted has come out to be positive and actual outcome as Negative.

**True Negative (TN):** When prediction has come out as negative along with the actual outcome as negative.

**False Negative (FN):** When prediction has come out as negative, and actual outcome as Positive.

The confusion matrix enables a clear visualization for the accuracy of classification by the model. In addition, distinguishing capability for AD stages v/s Cognitively Normal has been shown to be useful to help in deciding upon strengths and weaknesses of the model[22].

### *5.3 Performance Evaluation Metrics*

The models' performance was assessed using four key metrics: accuracy, precision, recall, and F1 score[23]:

- **Accuracy**: provides the overall correctness of the classification.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

- **Precision**: expresses the proportion of true positives to the overall number of positive predictions, showing the model's ability to avoid false positives.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

- **Recall**: The proportion of true positives among all actual positive cases correctly identified measured the models' sensitivity in detecting AD stages.

$$Recall = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

- **MAE** is calculated by averaging the individual differences between true and predicted values. It gives a straight idea about the predicting performance.

$$MAE = \frac{1}{U}\sum_{t=1}^{U}|F_t - A_t| \quad (4)$$

- **MSE** - represents the averaged squared differences between the prediction and real values, representing an emphasis on large errors [24].

## 6. Results and discussion

This section debates the performance of the implemented ML and DL models for EEG signals classification and predicting of AD stages. For this work, the dataset used consisted of EEG recordings for different cognitive states,

namely Healthy, Mild, Moderate, and Severe AD. These were further tested using the metrics such as Accuracy, Precision, Recall, MAE, and MSE.

### *6.1 Comparative Analysis*

Among the running ML models, the RF model was the best performer. It gave the best accuracy of 96.85% and precision of 97.08%, hence showing its skill at handling these complex interdependencies in the EEG data. RF can generalize to different AD stages without falling into overfitting or underfitting because of its ensemble learning nature.

Gradient Boosting Models such as XGBoost, CatBoost, and LightGBM do feature rather good performance since they can handle nonlinearities of data. These models achieved accuracies of 89.50%, 89.70%, and 89.30%, respectively. On the contrary, Naive Bayes and Logistic Regression showed low performances recorded at 63.60% and 65.90%, respectively, proving that such algorithms are not capable of describing the complexity behind EEG signals.

Results regarding machine learning models are shown in Table 4. Comparing the actual values to the predicted ones is represented in Figure 6, where the confusion matrix is shown, and in Figure 7, where ROC curves are depicted.

**Table 4:** The results obtained using ML Models

| Models | Accuracy | Precision | Recall | MAE | MSE |
|---|---|---|---|---|---|
| RF | 0.9685 | 0.9708 | 0.9685 | 0.0315 | 0.0315 |
| AdaBoost | 0.7130 | 0.7145 | 0.7130 | 0.4585 | 0.9065 |
| XGBoost | 0.8950 | 0.8956 | 0.8950 | 0.1655 | 0.3135 |
| CatBoost | 0.8970 | 0.8974 | 0.8970 | 0.1550 | 0.2850 |
| LightGBM | 0.8930 | 0.8936 | 0.8930 | 0.1680 | 0.3210 |
| Naive Bayes | 0.6360 | 0.6377 | 0.6360 | 0.5845 | 1.1485 |
| Decision Tree | 0.6645 | 0.6656 | 0.6645 | 0.5430 | 1.0760 |
| SVM | 0.9050 | 0.9057 | 0.9050 | 0.1500 | 0.2900 |
| K-NN | 0.7850 | 0.7920 | 0.7850 | 0.3460 | 0.6820 |
| Logistic Regression | 0.6590 | 0.6589 | 0.6590 | 0.5260 | 0.9890 |

**Table 5:** The results obtained using DL Models

| Models | Accuracy | Precision | Recall | MAE | MSE |
|---|---|---|---|---|---|
| DNN | 0.9605 | 0.9606 | 0.9605 | 0.0645 | 0.1255 |
| LSTM | 0.9240 | 0.9246 | 0.9240 | 0.1195 | 0.2245 |
| GRU | 0.9275 | 0.9281 | 0.9275 | 0.1100 | 0.2020 |
| CNN | 0.8270 | 0.8274 | 0.8270 | 0.2690 | 0.5100 |
| VGG16-like | 0.5815 | 0.5932 | 0.5815 | 0.6595 | 1.2575 |
| Bi-LSTM | 0.9360 | 0.9362 | 0.9360 | 0.1010 | 0.1930 |
| CNN-LSTM | 0.7720 | 0.7772 | 0.7720 | 0.3535 | 0.6625 |
| LSTM-XGBoost | 0.6615 | 0.6625 | 0.6615 | 0.5455 | 1.0645 |
| CNN-SVM | 0.8395 | 0.8401 | 0.8395 | 0.2495 | 0.4775 |
| CNN-DT | 0.8655 | 0.8657 | 0.8655 | 0.2195 | 0.4365 |

Out of these, while some of the DL models performed with varied efficiencies, DNN was able to capture intricate spatial features from EEG data and gave an accuracy of 96.05%. However, the consideration of temporal dependencies from input data allowed both Bi-LSTM and GRU to outperform most of the considered DL architectures by achieving classification accuracies of 93.60% and 92.75%, respectively. Among different hybrid models, CNN-SVM and CNN-DT reported competitive performances and yielded accuracies of 83.95% and 86.55%, respectively, proving that

combining feature extraction and classification steps enhances efficiency.

Performances of the implemented DL models are given in Table 5. Further, the details are shown in the confusion matrix presented in Figure 8, ROC curves in Figure 9 and training loss curves in Figure 10. Among these models, the DNN has a high capacity of spatial learning but it does not learn the temporal dependencies as much as the LSTM-based models.

Therefore, all these results indicate that both the ML and DL methods were quite effective for the classification of AD at different stages using EEG data. A reproducible classification benchmark is thus formed, allowing further studies of EEG-based classification for AD and establishing a basis of more advanced algorithms with the aim of increasing the accuracy of diagnosis.

### *6.2. Limitations and considerations*

This work introduces a new dataset on AD in Iraq. However, there are several limitations that must be considered when interpreting these results to help guide future studies better. Major limitations of this dataset include regional and demographic specificity, where the AD dataset is mainly composed of people from the Kurdistan region in Iraq. This situation may affect the generalizability of models developed on this dataset to other regions or populations due to different lifestyle factors, genetic predisposition, or the healthcare environment. The dataset is dominantly based on resting-state EEG recordings and may fail to capture dynamic neural processes associated with cognitive decline. The inclusion of task-based EEG paradigms or other neuroimaging modality together with clinical biomarkers may further help improve our understanding of brain changes associated with AD.

Other considerations include the possibility of model overfitting. Powerful, advanced architectures like CNNs and LSTMs may learn to recognize dataset-specific patterns rather than generalizable features. This limitation underscores the need for further rigorous cross-validation and testing on external datasets to confirm the robustness of such models.

Finally, ICA and ASR are two preprocessing steps applied to the EEG signals, efficient for noise removal though they are, which may actually remove very small neural signals associated with AD. Further research is, therefore, recommended to investigate other pre-processing methods and to determine their effects on successive analysis.

In conclusion, though this study represents an improvement of previous EEG-based AD studies, overcoming these limitations through the use of larger datasets, methodological novelties, and collaborative efforts will be essential to fully realize the EEG's potential as a diagnostic modality for Alzheimer's Disease.

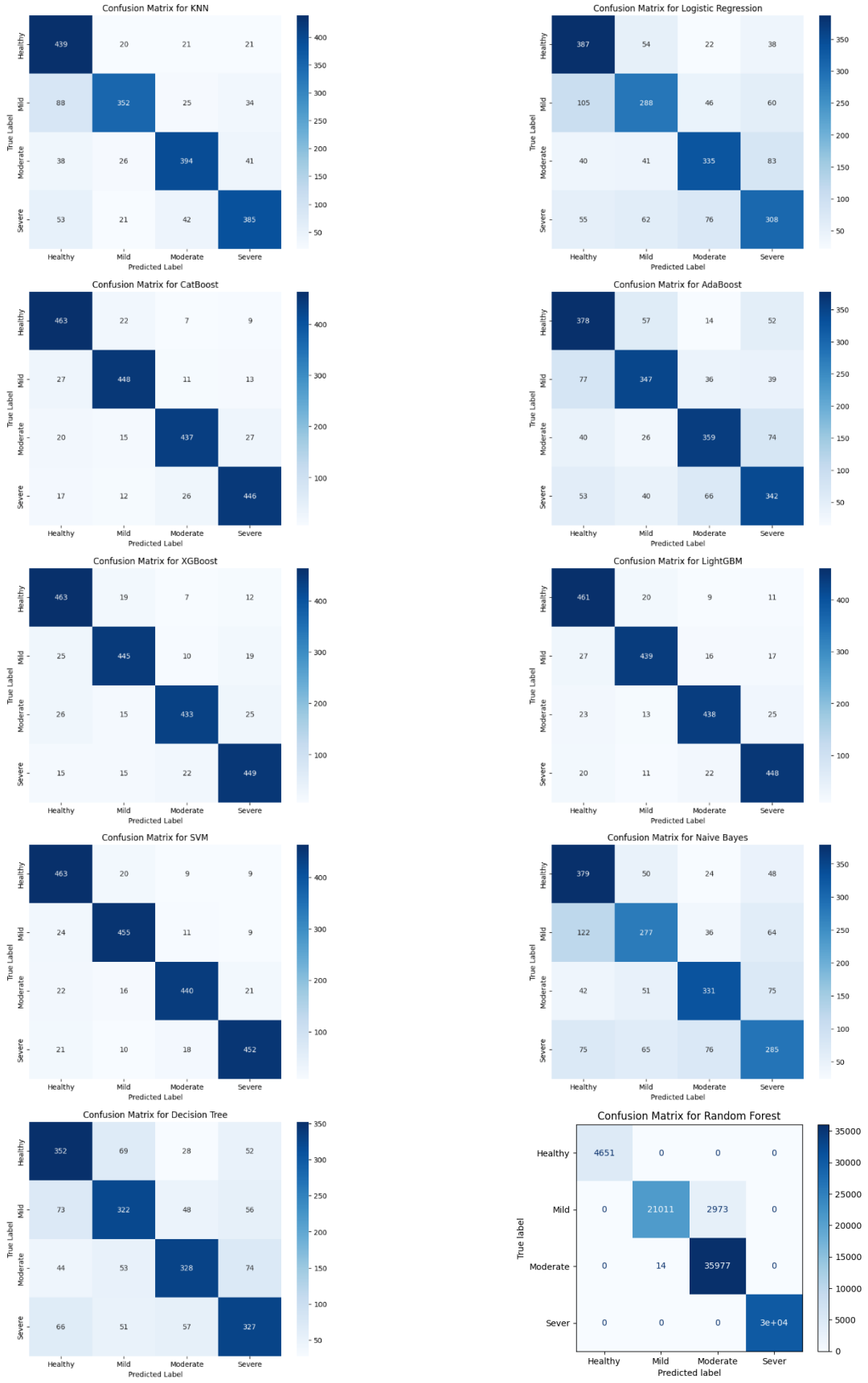


**Figure 6.** The confusion matrix result of the main dataset for the ML models

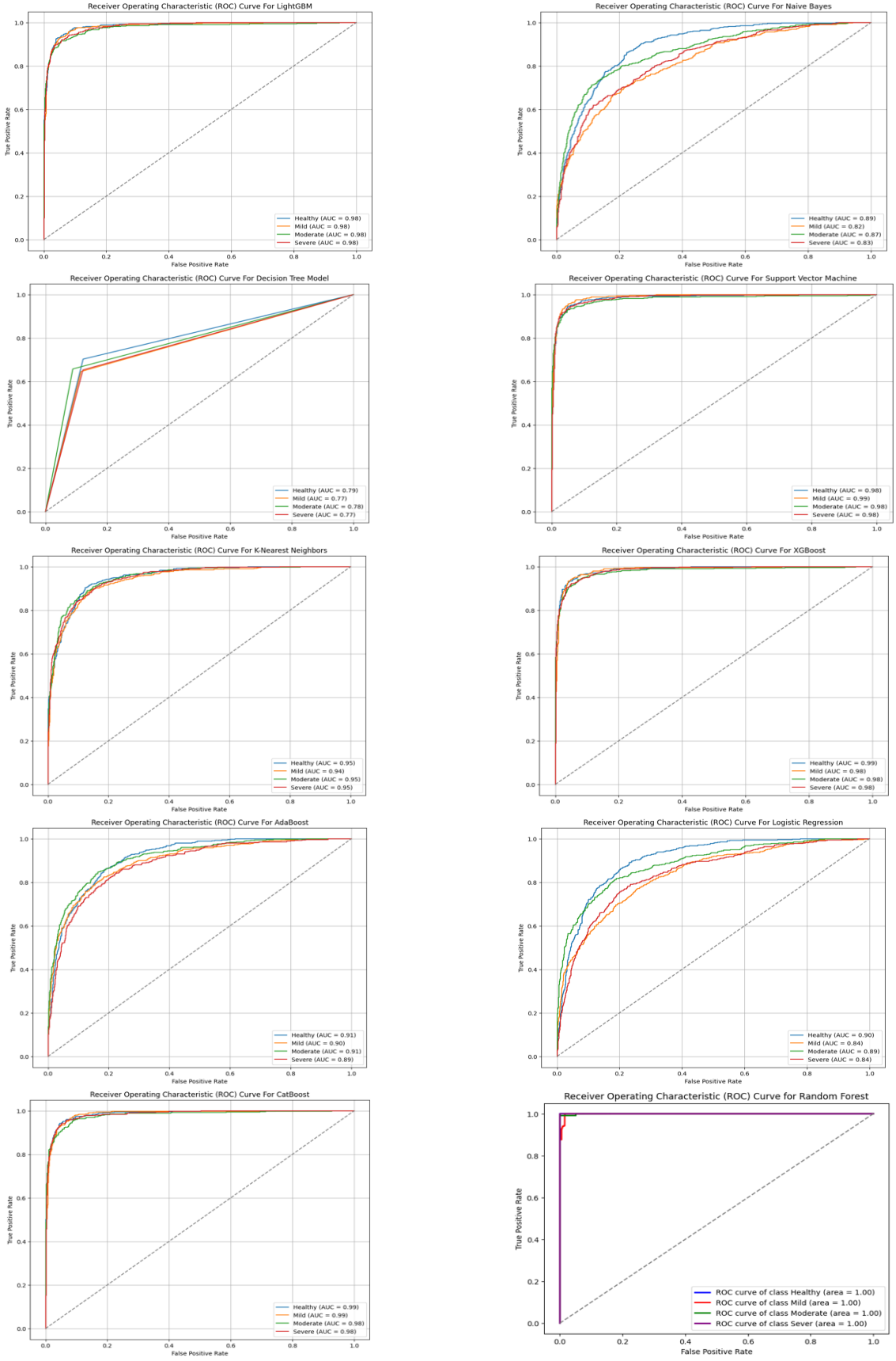


**Figure 7.** The ROC curves for the for the ML models

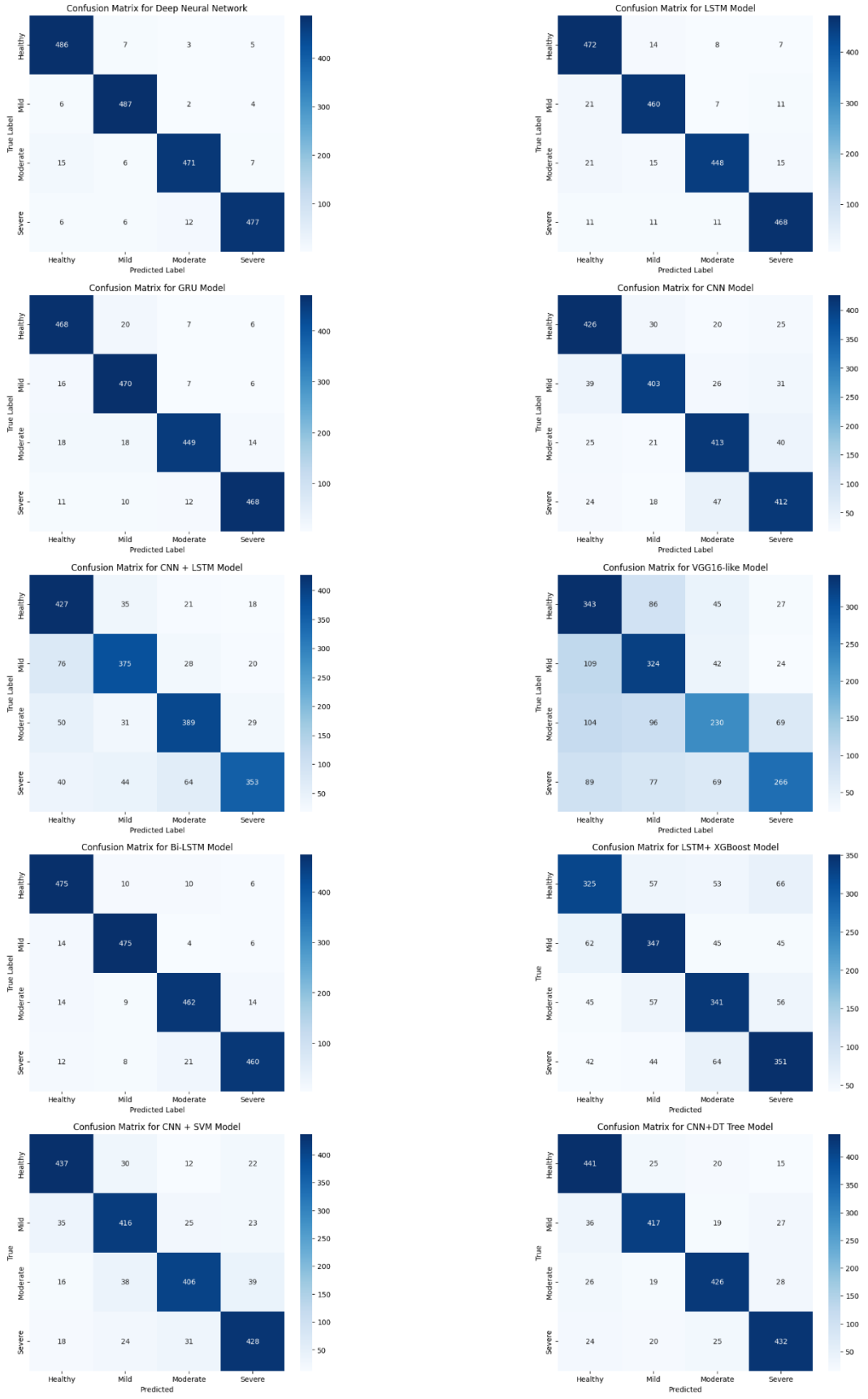


**Figure 8.** The confusion matrix result of the main dataset for the DL models

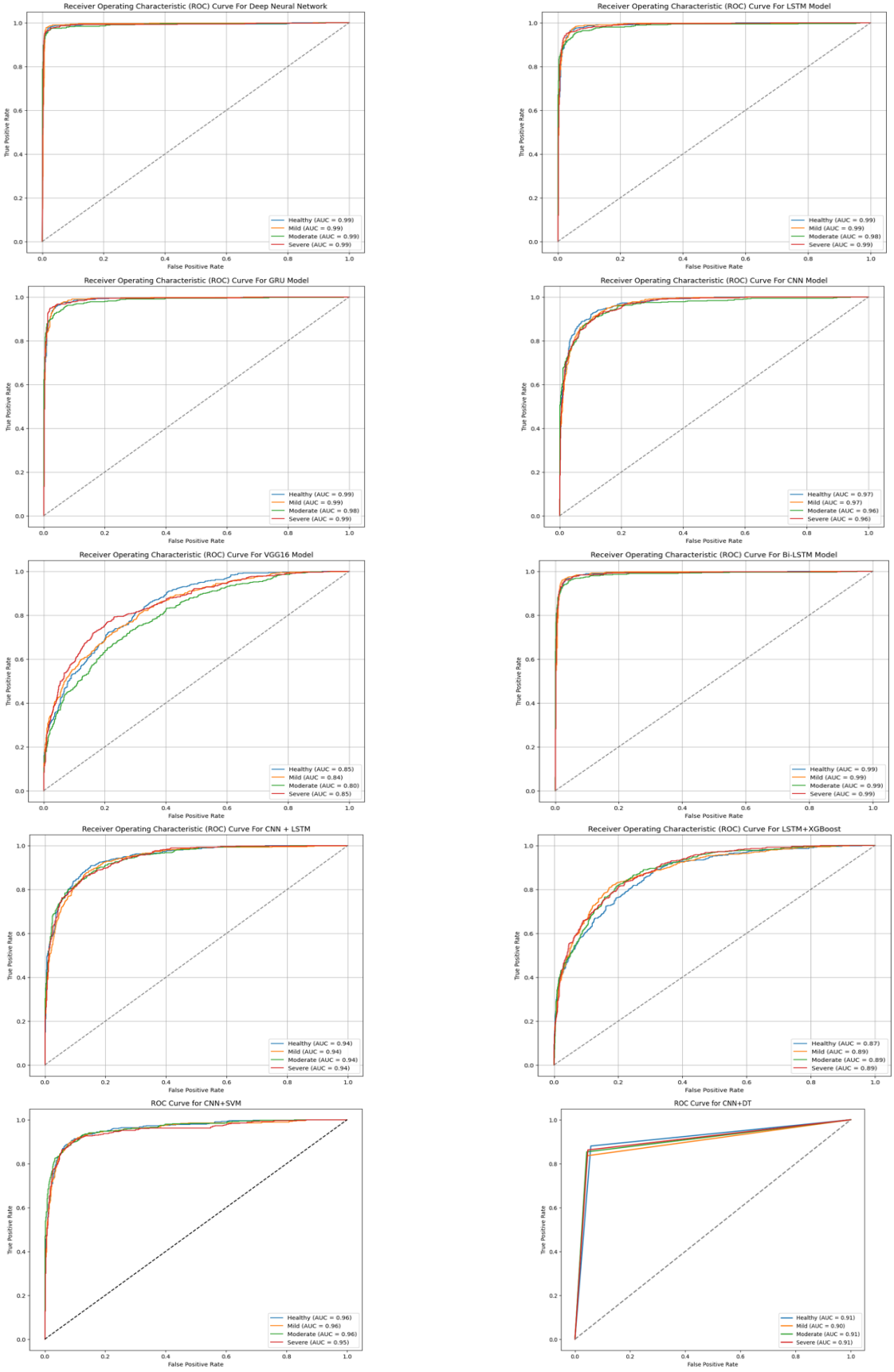


**Figure 9.** The ROC curves for the for the DL models

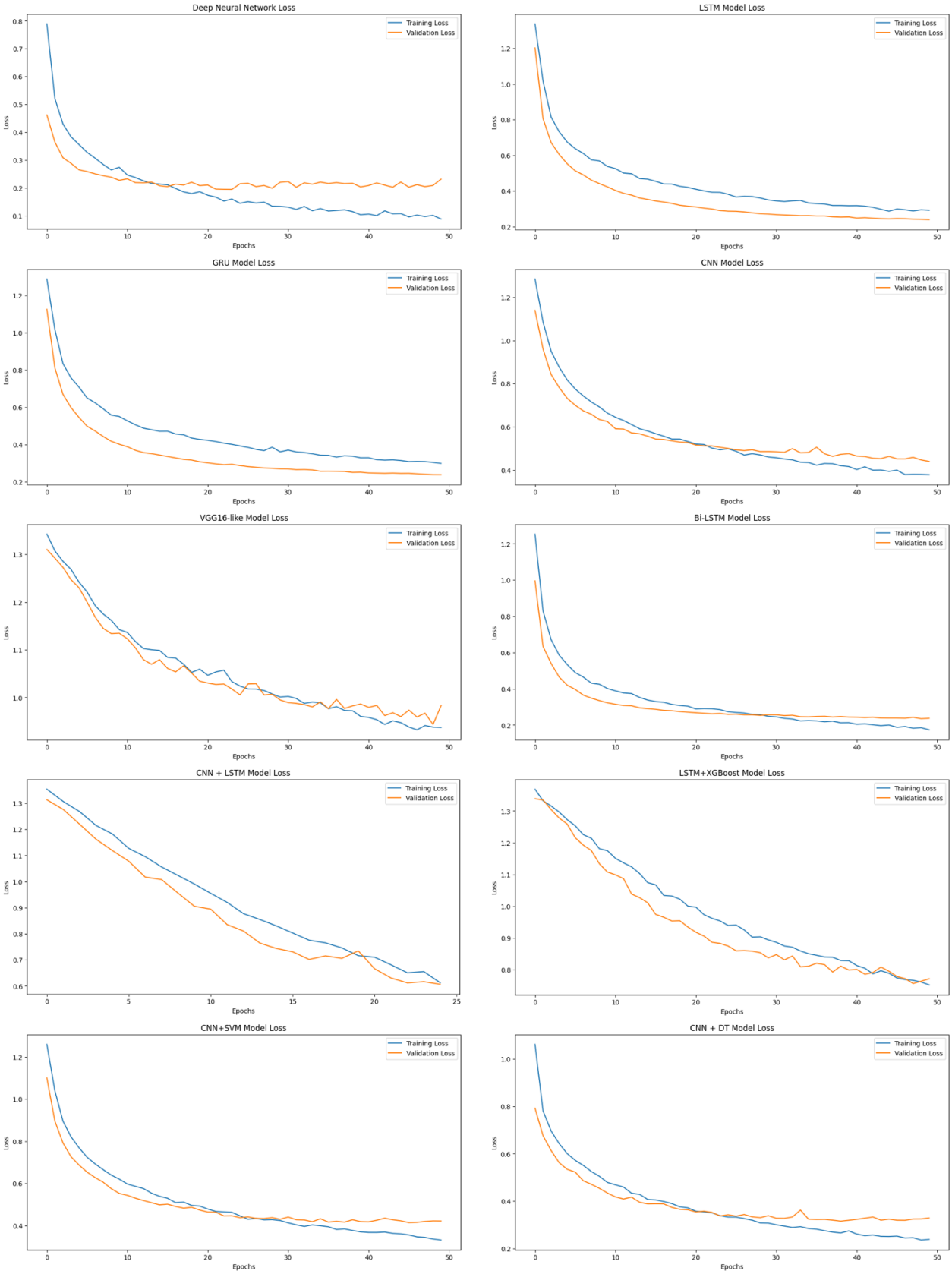


**Figure 10.** The training loss curves for the DL models

## 7. Conclusion

This research presents the first publicly available EEG dataset for Alzheimer's Disease (AD) in Iraq, contributing significantly to the international research community. The dataset includes 53 participants across four groups (Healthy, Mild, Moderate, and Severe AD), recorded using a 40-channel EEG system following the 10-20 electrode placement standard. Advanced preprocessing, including Independent Component Analysis (ICA) and Automatic Subspace Reconstruction (ASR), ensures high-quality signals suitable for machine learning (ML) and deep learning (DL) applications.

Benchmark classification results demonstrate the dataset's effectiveness, with Random Forest achieving 96.85% accuracy in ML models and Deep Neural Networks (DNN) reaching 96.05% accuracy in DL models. These findings highlight EEG's potential as a noninvasive and cost-effective diagnostic tool for AD, particularly in regions with limited access to neuroimaging.

Despite its contributions, the dataset has limitations, including regional specificity, reliance on resting-state EEG, and potential overfitting in deep learning models. Future work will focus on expanding the dataset, incorporating task-based EEG, and integrating multimodal approaches such as neuroimaging and genetic data. Additionally, optimizing preprocessing pipelines and exploring advanced ML/DL architectures, including transformer networks and attention mechanisms, will enhance classification performance.

## References


[1] A. Miltiadous *et al.*, "Alzheimer's disease and frontotemporal dementia: A robust classification method of eeg signals and a comparison of validation methods," *Diagnostics*, vol. 11, no. 8, 2021, doi: 10.3390/diagnostics11081437.

[2] M. Ouchani, S. Gharibzadeh, M. Jamshidi, and M. Amini, "A Review of Methods of Diagnosis and Complexity Analysis of Alzheimer's Disease Using EEG Signals," *Biomed Res. Int.*, vol. 2021, 2021, doi: 10.1155/2021/5425569.

[3] X. Zheng *et al.*, "Diagnosis of Alzheimer's disease via resting-state EEG: integration of spectrum, complexity, and synchronization signal features," *Front. Aging Neurosci.*, vol. 15, no. November, pp. 1–9, 2023, doi: 10.3389/fnagi.2023.1288295.

[4] Y. Ma, J. K. S. Bland, and T. Fujinami, "Classification of Alzheimer's Disease and Frontotemporal Dementia Using Electroencephalography to Quantify Communication between Electrode Pairs," *Diagnostics*, vol. 14, no. 19, 2024, doi: 10.3390/diagnostics14192189.

[5] A. Miltiadous *et al.*, "A Dataset of Scalp EEG Recordings of Alzheimer's Disease, Frontotemporal Dementia and Healthy Subjects from Routine EEG," *Data*, vol. 8, no. 6, pp. 1–10, 2023, doi: 10.3390/data8060095.

[6] M. J. Sedghizadeh, H. Aghajan, and Z. Vahabi, "Brain electrophysiological recording during olfactory stimulation in mild cognitive impairment and Alzheimer disease patients: An EEG dataset," *Data Br.*, vol. 48, 2023, doi: 10.1016/j.dib.2023.109289.

[7] A. M. Pineda, F. M. Ramos, L. E. Betting, and A. S. L. O. Campanharo, "Quantile graphs for EEG-based diagnosis of Alzheimer's disease," *PLoS One*, vol. 15, no. 6, pp. 1–15, 2020, doi: 10.1371/journal.pone.0231169.

[8] K. Smith, D. Abásolo, and J. Escudero, "Accounting for the complex hierarchical topology of EEG phase-based functional connectivity in network binarisation," *PLoS One*, vol. 12, no. 10, pp. 1–21, 2017, doi: 10.1371/journal.pone.0186164.

[9] M. Cejnek, O. Vysata, M. Valis, and I. Bukovsky, "Novelty detection-based approach for Alzheimer's disease and mild cognitive impairment diagnosis from EEG," *Med. Biol. Eng. Comput.*, vol. 59, no. 11–12, pp. 2287–2296, 2021, doi: 10.1007/s11517-021-02427-6.

[10] M. L. Vicchietti, F. M. Ramos, L. E. Betting, and A. S. L. O. Campanharo, "Computational methods of EEG signals analysis for Alzheimer's disease classification," *Sci. Rep.*, vol. 13, no. 1, pp. 1–14, 2023, doi: 10.1038/s41598-023-32664-8.

[11] M. Soufineyestani, D. Dowling, and A. Khan, "Electroencephalography (EEG) technology applications and available devices," *Appl. Sci.*, vol. 10, no. 21, pp. 1–23, 2020, doi: 10.3390/app10217453.

[12] V. Aspiotis *et al.*, "Assessing Electroencephalography as a Stress Indicator: A VR High-Altitude Scenario Monitored through EEG and ECG," *Sensors*, vol. 22, no. 15, pp. 1–16, 2022, doi: 10.3390/s22155792.

[13] F. Duan *et al.*, "Topological Network Analysis of Early Alzheimer's Disease Based on Resting-State EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2164–2172, 2020, doi: 10.1109/TNSRE.2020.3014951.

[14] M. M. Hasan, S. Khandaker, N. Sulaiman, M. Mahfuj Hossain, and A. Islam, "Addressing Imbalanced EEG Data for Improved Microsleep Detection: An ADASYN, FFT and LDA-Based Approach," *Diyala J. Eng. Sci.*, vol. 17, no. 3, pp. 45–57, 2024, doi: 10.24237/djes.2024.17304.

[15] N. Hassan, A. S. M. Miah, and J. Shin, "A Deep Bidirectional LSTM Model Enhanced by Transfer-Learning-Based Feature Extraction for Dynamic Human Activity Recognition," *Appl. Sci.*, vol. 14, no. 2, 2024, doi: 10.3390/app14020603.

[16] R. Jamal Kolaib and J. Waleed, "Crime Activity Detection in Surveillance Videos Based on Developed Deep Learning Approach," *Diyala J. Eng. Sci.*, vol. 17, no. 3, pp. 98–114, 2024, doi: 10.24237/djes.2024.17307.

[17] M. Y. Shakor and N. M. S. Surameery, "CNN-Based Transfer Learning for 3D Knuckle Recognition," *Adv. Multimed.*, vol. 2023, pp. 1–12, 2023, doi: 10.1155/2023/6147422.

[18] V. Ronca *et al.*, "Optimizing EEG Signal Integrity: A Comprehensive Guide to Ocular Artifact Correction," *Bioengineering*, vol. 11, no. 10, 2024, doi: 10.3390/bioengineering11101018.

[19] N. M. S. Surameery, A. Alazzawi, and A. T. Asaad, "Intelligent Approaches for Alzheimer ' s Disease Diagnosis from EEG Signals : Systematic Review," vol. 17, no. 1, pp. 1–19, 2025.

[20] J. Shi, H. Yan, and Y. Wei, "Application of the denoising technique based on improved FastICA algorithm in cross-correlation time delay estimation," *J. Phys. Conf. Ser.*, vol. 1754, no. 1, 2021, doi: 10.1088/1742-6596/1754/1/012196.

[21] W. Xia, R. Zhang, X. Zhang, and M. Usman, "A novel method for diagnosing Alzheimer's disease using deep pyramid CNN based on EEG signals," *Heliyon*, vol. 9, no. 4, p. e14858, 2023, doi: 10.1016/j.heliyon.2023.e14858.

[22] M. Velazquez and Y. Lee, "Random forest model for feature-based Alzheimer's disease conversion prediction from early mild cognitive impairment subjects," *PLoS One*, vol. 16, no. 4 April, pp. 1–18, 2021, doi: 10.1371/journal.pone.0244773.

[23] H. Q. Flayyih, J. Waleed, and A. M. Ibrahim, "Indoor Air Quality Prediction in Sick Building Using Machine and Deep Learning: Comparative Analysis," *Diyala J. Eng. Sci.*, vol. 18, no. 1, pp. 203–218, 2025, doi: 10.24237/djes.2025.18112.

[24] K. AlSharabi, Y. Bin Salamah, M. Aljalal, A. M. Abdurraqeeb, and F. A. Alturki, "EEG-based clinical decision support system for Alzheimer's disorders diagnosis using EMD and deep learning techniques," *Front. Hum. Neurosci.*, vol. 17, 2023, doi: 10.3389/fnhum.2023.1190203.