An Ensemble Deep Learning Framework with Adaptive Windowing for EEG-Based Brain Disorder Detection

University of Staffordshire

School of Digital, Technology, Innovation & Business

Dina Mohamed Abooelzahab

dina.abouldahab@research.staffs.ac.uk

A thesis submitted in partial fulfilment of the requirements of University of Staffordshire for the degree of Doctor of Philosophy

Abstract

Electroencephalogram (EEG)-based signal analysis serves as an important tool for evaluating brain dynamics and diagnosing neurological conditions, yet traditional interpretation remains difficult due to noise, variability, and the dependence on specialist expertise. Such challenges underscore the importance of developing automated diagnostic approaches that are more consistent and clinically dependable. This thesis introduces a multi-stage computational framework designed to improve the detection of abnormal EEG activity through refined feature selection, deep neural modelling, and ensemble decision strategies.

The first stage of the framework applies a windowing method, which adaptively segments each EEG recording into its most informative temporal interval, based on Generalised Morse Wavelet (GMW) analysis, reducing irrelevant fluctuations and strengthening signal relevance. From these refined windows, five alternative signal representations max pooling, min pooling, average pooling, and two decimation schemes are generated to capture different temporal and amplitude characteristics. Each version is processed using an 18-layer CNN–LSTM model capable of learning combined spatial and temporal patterns in the EEG.

To address the shortcomings of individual models, particularly their variable sensitivity (the ability of a model to correctly detect abnormal cases), a meta-classification layer is introduced. Three ensemble techniques: Support Vector Machines, Linear Regression, and Random Forest are used to integrate the outputs of the five CNN–LSTM branches. Among these, the Random Forest ensemble delivered the strongest performance, reaching 81.8% accuracy and a sensitivity of 92.86% on the TUH Abnormal EEG Corpus, demonstrating a substantial reduction in missed abnormal events compared with existing approaches.

The results confirm that combining wavelet-based windowing, diverse feature extraction, and ensemble fusion can significantly enhance the robustness and clinical reliability of automated EEG analysis. The proposed framework provides a foundation for future systems capable of real-time monitoring and broader diagnostic functionality, with potential extensions including multi-class classification, transformer-based learning, explainability techniques, and cross-dataset validation.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Professor AbdelHamid Soliman, Dr Nawal Zaher, and Professor Claude Chibelushi, for their invaluable guidance, continuous support, and constructive feedback throughout my PhD journey. Their expertise, encouragement, and patience have been instrumental in shaping this research and in supporting my academic development. I am deeply thankful for their commitment and mentorship during all stages of this work.

I am also profoundly grateful to my father, my husband, and my family for their unwavering support, understanding, and encouragement throughout this period. Their belief in me and their constant presence provided strength and motivation during both the academic and personal challenges of this journey.

Finally, I would like to dedicate this work to the memory of my mother, who passed away during my PhD. Her love, values, and unwavering belief in the importance of education continue to inspire me every day. Although she is no longer with us, her spirit and support remain a constant source of strength throughout this achievement.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AD	Alzheimer's Disease
AUC	Area Under the Curve
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANFN	Adaptive Neural Fuzzy Network
ApEn	Approximate Entropy
AR	Autoregression / Autoregressive Model
ASD	Autism Spectrum Disorder
BCI	Brain Computer Interface
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DL	Deep Learning
DWT	Discrete Wavelet Transform
EEG	Electroencephalogram / Electroencephalography
ECoG	Electrocorticogram / Electrocorticography
F1	F1-Score
FFNN	Feed Forward Neural Network
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy

GMW	Generalised Morse Wavelet
GRU	Gated Recurrent Unit
HMM	Hidden Markov Model
ICA	Independent Component Analysis
IEDs	Interictal Epileptiform Discharges
IRA	Inter-Rater Agreement
LR	Logistic Regression
LSTM	Long Short-Term Memory
MEG	Magnetoencephalography
ML	Machine Learning
MRMR	Minimum Redundancy Maximum Relevance
PET	Positron Emission Tomography
PCA	Principal Component Analysis
PE	Permutation Entropy
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
TUH	Temple University Hospital EEG Corpus

List of Symbols

$\Psi_{\beta,\gamma}(\omega)$	Frequency domain representation of the Generalised Morse Wavelet (GMW).
ω	Angular frequency variable in the frequency domain.
$U(\omega)$	Unit step function (0 for $\omega < 0$, 1 for $\omega \geq 0$).
$a_{\beta,\gamma}$	Normalisation constant ensuring correct wavelet scaling and energy.
β	Time–domain decay parameter controlling temporal localisation.
γ	Frequency domain decay parameter controlling spectral localisation (must satisfy $\gamma > 0$).
α	Positivity parameter in the GMW admissibility condition ($\alpha > 0$).
e	Euler's number (≈ 2.71828).
X	EEG signal segment containing N samples.
X_i	The i-th sample of the EEG signal.
N	Total number of samples in the EEG sequence.
k	Window size used for averaging or pooling ($k = 2$).
i	Window index, ranging from 1 to N/2.
$X_{avg[i]}$	Average of the i-th sample pair: $(X_{(2i-1)} + X_{(2i)}) / 2$.
$X_{max[i]}$	Maximum value of the i-th sample pair: $\max(X_{(2i-1)}, X_{(2i)})$.
$X_{min[i]}$	Minimum value of the i-th sample pair: $\min(X_{(2i-1)}, X_{(2i)})$.

$X_{even[i]}$ Even-indexed sample during decimation: $X_{(2i)}$.

$X_{odd[i]}$ Odd-indexed sample during decimation: $X_{(2i-1)}$.

List of Publications

Journal Articles

1- Abooelzahab, D., Zaher, N., Soliman, A.H., & Chibelushi, C. (2025). *A Combined Windowing and Deep Learning Model for the Classification of Brain Disorders Based on Electroencephalogram Signals.*AI, 6(3), 42.

2- Abooelzahab, D., Zaher, N., Soliman, A.H., & Chibelushi, C. (2026). *A Voting-Based Ensemble Approach for Brain Disorder Detection Using Random Forest.* Computers, 15, 18.

Conference Papers

1- Abooelzahab, D.M., Zaher, N., Soliman, A.H., & Chibelushi, C. (2024). *Deep Learning vs. Conventional Techniques for Processing and Classifying EEG Brain Disorders: A Survey*. In Proceedings of the 2024 International Congress on Human–Computer Interaction, Optimization and Robotic Applications (HORA 2024).

Chapter 1 Introduction

1.1 Background

In hospitals, physicians rely on electroencephalogram (EEG) recordings to diagnose neurological disorders, including epilepsy. The popularity of EEG in clinical settings stems from its non-invasive nature, high temporal resolution, and relatively low cost compared to other neuroimaging techniques (R.K. Desai Marg, et al., 2022; Jui et al., 2023). Despite these advantages, manual analysis of EEG data poses several challenges. It requires highly trained clinicians or neurophysiologists, and the sheer volume and acquisition rate of EEG recordings make the interpretation process time-consuming, resource-intensive, and expensive (Rivas-Carrillo et al., 2023; Yun, 2024). Automated EEG analysis has the potential to address these limitations by reducing human error, shortening the time to diagnosis, and enabling the automatic detection of clinically significant events (Shang et al., 2024). In recent years, machine learning has been incorporated into EEG research to identify relevant features for classifying brain disorders (Uyanik et al., 2025). More recently, deep learning architectures, particularly Convolutional Neural Networks (CNNs) (MohiudDin & Jayanthy, 2021; Khademi, Ebrahimi & Kordy, 2022; Rathod et al., 2025; Stefanou et al., 2025) and Recurrent Neural Networks (RNNs), have gained traction for their ability to learn complex temporal and spatial features directly from EEG signals (Craik et al., 2019; Najafi et al., 2022).

However, most of these approaches rely on idealised datasets collected under controlled laboratory conditions, typically involving homogeneous patient groups of similar ages and a limited range of neurological disorders (Obeid and Picone, 2018). Consequently, their generalisability to real-world clinical environments remains limited. To address this issue, researchers have increasingly utilised the Temple University Hospital (TUH) EEG Corpus, the largest publicly available clinical EEG database (Obeid and Picone, 2018). Unlike controlled laboratory datasets, TUH contains recordings acquired under authentic hospital conditions across

diverse age groups, a wide spectrum of brain disorders, and varying patient health states. This diversity makes TUH a reliable benchmark for developing clinically relevant automated models (Harati et al., 2013). Nevertheless, early benchmark studies using large-scale clinical EEG datasets reported modest performance, highlighting the challenges inherent in developing clinically dependable automated classification systems and the need for further methodological innovation (Wu et al., 2022). Despite subsequent progress, achieving high diagnostic accuracy particularly sensitivity remains a significant challenge (Siddiqui, 2025). This creates a clear need for more advanced and robust classification frameworks, particularly those capable of handling the heterogeneity of clinical EEG data to improve reliability and support real-world adoption in clinical workflows.

1.2 Problem Statement

Although EEG is widely used in clinical practice for diagnosing neurological disorders such as epilepsy, its manual interpretation remains a specialised, resource-intensive task subject to inter-observer variability (Dişli et al., 2025; Vinutha et al., 2025). Automated EEG classification has emerged as a promising solution; however, existing approaches continue to face critical limitations. Many earlier studies were developed using controlled laboratory datasets with limited demographic and diagnostic diversity, thereby constraining their applicability to real-world clinical conditions (Obeid and Picone, 2018). The TUH EEG Corpus provides a large-scale, realistic alternative, yet automated classification on TUH remains highly challenging due to signal variability, noise, and the inherent heterogeneity of clinical data (Harati et al., 2013). Benchmark studies on TUH have demonstrated relatively low sensitivity and specificity, where sensitivity is defined as the proportion of true abnormal cases correctly identified by the model and specificity is defined as the proportion of true normal cases correctly identified, reflecting the difficulty of achieving reliable brain disorder detection under real-world conditions. (Wu et al., 2022).

While deep learning techniques such as CNNs (Rathod et al., 2025; Stefanou et al., 2025) and LSTMs (Chekhmane and Benali, 2025; Kim et al., 2025) have improved performance by modelling the temporal and spatial dependencies of EEG signals, they still exhibit sensitivity limitations; an issue of critical importance in clinical diagnosis where missed detections can have severe consequences. Ensemble learning offers a potential solution by combining complementary strengths of multiple classifiers to enhance robustness and accuracy. However, its application in

EEG-based brain disorder classification remains limited (Bunterngchit et al., 2025; Manafian et al., 2025).

Therefore, there is a pressing need for a robust, sensitivity-focused EEG classification framework that leverages deep learning feature representations alongside ensemble decision-making to achieve clinically reliable performance. The use of large, trustworthy, and heterogeneous real-world datasets such as TUH is essential to validate and ensure the generalisability of such models.

1.3 Research Aim

The aim of this research is to develop and evaluate an advanced automated framework for brain disorder detection using EEG signals, with a focus on improving diagnostic accuracy and sensitivity on realistic and heterogeneous clinical datasets such as the TUH corpus. The proposed framework seeks to enhance the reliability and efficiency of EEG-based decision-making in clinical settings.

1.4 Research Objectives

The main objective of this thesis is to enhance EEG-based brain disorder detection by introducing novel methods for feature selection, feature extraction, and classification to achieve improved performance metrics. The specific research objectives are as follows:

1. To develop a novel windowing-based feature selection method that improves the representativeness of EEG data by selecting the most informative temporal segments for analysis.
2. To implement and evaluate multiple feature extraction strategies, each designed to capture different aspects of EEG signal characteristics and enhance the diversity of the learned representations.
3. To propose and design a deep learning model specifically tailored for both feature extraction and classification of EEG signals, enabling automatic learning of spatial and temporal dependencies within the data, and to evaluate its performance across all feature extraction approaches.
4. To propose and validate an ensemble Random Forest stacking model that integrates the outputs of the proposed classifiers, thereby improving robustness and maximising sensitivity, particularly in detecting abnormal EEG patterns.

5. To benchmark the proposed framework against state-of-the-art approaches in the literature using a trusted dataset, the TUH EEG Corpus, highlighting the superiority and clinical applicability of the developed framework.

1.5 Thesis Contribution to knowledge

In response to the limitations identified in previous EEG classification approaches, this thesis comprises the following key contributions to knowledge:

- Design and Implementation of a Robust hybrid 18-Layer CNN-LSTM Model

Proposing a customised deep learning architecture that efficiently learns both spatial and temporal EEG features. The novel 18-layer CNN-LSTM model demonstrates superior automatic feature extraction and classification capabilities compared to other available techniques in literature, and its effectiveness is validated for each feature extraction approach.

- Novel Windowing-Based Feature Selection Approach

Introducing a new temporal windowing-based feature selection method is introduced to provide more representative and informative EEG segments for subsequent analysis. This approach surpasses traditional global or static selection techniques by dynamically focusing on clinically relevant signal regions, improving the interpretability and discriminative quality of extracted features.

- Design and Validation of a Feature-Diverse Ensemble Learning Framework

Proposing a comprehensive ensemble learning framework that systematically integrates multiple feature extraction strategies namely Maximum, Minimum, Average, Odd, and Even sampling to capture complementary EEG signal characteristics. Five independently trained CNN–LSTM classifiers are developed based on these feature representations and subsequently combined using a Random Forest–based stacking mechanism. This ensemble strategy enhances model robustness and sensitivity by exploiting the complementary strengths of diverse temporal–spatial feature perspectives, leading to improved performance in detecting abnormal EEG events, addressing key reliability issues in clinical applications.

`1.6 Thesis Structure

This thesis is structured into eight chapters. Chapter 2 provides a comprehensive review of the literature on EEG-based brain disorder detection, with particular emphasis on existing methodologies, their strengths, and their limitations. Chapter 3 outlines the methodological framework of the proposed ensemble approach, detailing the underlying design principles, architectural components, and integration strategy adopted throughout the study. Chapter 4 describes the development of the deep learning framework that integrates convolutional and recurrent layers for effective brain disorder detection. Chapter 5 introduces a novel windowing-based feature selection technique designed to enhance EEG signal representation and improve model input quality. Chapter 6 presents an in-depth description of five distinct feature extraction strategies, each independently implemented within the proposed CNN-LSTM architecture. It further details the design and evaluation of a Random Forest–based stacking model that integrates the outputs of multiple classifiers to enhance overall system performance. Chapter 7 presents a comprehensive discussion of the experimental findings, analysing the effectiveness of the proposed integrated framework, including the windowing-based feature selection, multi-feature CNN–LSTM modelling, and ensemble learning strategy. The results are critically compared with previous work, practical implications are discussed, and the limitations of the proposed approach are identified. The chapter concludes by summarising the overall contributions and key findings and outlining potential directions for future research.

Chapter 2
Literature Review

2.1 Introduction

Understanding brain activity is fundamental to the diagnosis, monitoring, and treatment of neurological disorders. The brain functions through highly interconnected electrochemical processes, where billions of neurons communicate via electrical impulses to regulate cognition, behaviour, motor functions, and physiological responses (Yen et al., 2023). These electrical and hemodynamic signals form the basis of measurable brain activity, which can be captured non-invasively from the scalp or invasively through direct cortical access.

Advancements in signal acquisition and computational analysis now enable clinicians and researchers to examine neural dynamics with increasing precision. Brain measurement techniques have become essential tools for evaluating cognitive functions, detecting abnormalities such as epileptic seizures, assessing sleep architecture, monitoring neurodegenerative progression, and supporting brain–computer interface (BCI) applications (Yen et al., 2023). Each technique provides unique insights into neuronal behaviour by quantifying different aspects of brain activity whether electrical, magnetic, or metabolic.

Given the diversity of available technologies, selecting an appropriate brain measurement method depends on several factors, including spatial and temporal resolution, invasiveness, susceptibility to artefacts, portability, cost, and clinical feasibility. A comprehensive understanding of these modalities is therefore crucial when designing diagnostic systems or research protocols aimed at analysing complex neural phenomena.

2.2 Brain Function Measurement Techniques

Brain activity can be measured using a range of techniques that vary in complexity, accuracy, and clinical applicability. Broadly, these methods fall into two major categories: invasive (Chaudhary, 2025a) and non-invasive approaches (Chaudhary, 2025b). Invasive methods involve direct access to the cortical surface and offer high spatial resolution at the cost of surgical risk, whereas non-

invasive techniques allow safe and repeated measurement with varying levels of precision and sensitivity.

Each modality carries specific strengths and limitations with respect to its resolution, anatomical coverage, signal quality, and practicality for long-term monitoring (Yen et al., 2023). The following sections provide an overview of both invasive and non-invasive brain-measurement techniques, highlighting their operational principles, clinical utility, and relevance to neurological assessment.

2.2.1 Invasive Techniques

Invasive techniques involve surgical implantation of electrodes or devices directly onto the surface of the brain (cortex), into deep brain structures, or sometimes within the brain tissue itself. They are primarily used in clinical settings where precise localisation of activity is critical, and the benefits outweigh the risks of surgery (Chaudhary, 2025a) . Although the invasive techniques are very accurate, they are risky and can cause brain damage. The main technique utilised is the Electrocorticogram ECoG (Chaudhary, 2025a) .

Electrocorticogram (ECoG)

Electrocorticography (ECoG) is recorded by positioning electrodes beneath the skull, either on top of the dura mater (epidural) or directly beneath it (subdural). Because the electrodes lie close to the cortical surface, ECoG provides highly accurate measurements of brain activity (Todaro et al., 2019). There are several problems with the ECoG that prevent it from being used in long-term applications. First, the biocompatibility of microelectrodes must be considered, and several designs incorporating neurotrophic materials have been proposed to encourage neuronal integration and reduce tissue rejection. Second, reliable communication between the implanted microelectrode and external hardware is essential; wireless interfaces are often recommended to minimise infection risk associated with transcutaneous connections. Third, repeated mechanical stress from plugging and unplugging wired recording systems can cause tissue irritation or even device failure (Moon et al., 2024).

2.2.2 Non-invasive Techniques

Non-invasive brain measurement techniques are safer, more practical, and widely used in both clinical and research settings (Smith, Chen and Davis, 2023). These methods do not require surgical intervention and are typically used for long-term or large-scale studies. The sensors are placed on the scalp to measure the electrical potential produced by the brain. The main non-invasive techniques are Functional Near-Infrared Spectroscopy (fNIRS), Magnetoencephalography (MEG), Functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), and EEG Electroencephalography (Smith, Chen and Davis, 2023).

a) Functional near-infrared spectroscopy (fNIRS)

FNIRS is a portable method to measure blood oxygenation and blood volume in the cortex. It is useful in bedside monitoring but limited in depth and resolution (Huff et al., 2025). Functional near-infrared spectroscopy (fNIRS) has been used to examine how pain and neuropsychiatric symptoms relate to cortical hemodynamic activity in individuals with Alzheimer's disease and related dementias as well as those with knee osteoarthritis, with analyses stratified according to cognitive status (Huff et al., 2025). Figure 2.1: Illustration of a functional Near-Infrared Spectroscopy (fNIRS) system setup, showing the optode cap with light sources and detectors, data cables, processing unit, and brain activity monitor displaying left (L) and right (R) hemisphere activation maps.

Figure 2.1: Illustration of an fNIRS system setup showing the optode cap, data acquisition unit, and real-time brain activation display.

b) Magnetoencephalography (MEG)

Magnetoencephalography (MEG) measures the brain's oscillatory magnetic fields generated by neuronal activity, providing excellent temporal and moderate spatial resolution. Mandal et al. (2018) reviewed MEG analysis techniques and their use in studying neuronal rhythm modulation in Alzheimer's disease (AD). The review emphasised MEG's potential for connectivity analysis and its integration with other neuroimaging modalities to identify early diagnostic biomarkers for AD. Figure 2.2 Illustration of a MEG system setup showing a participant seated under the sensor helmet with brain activation maps displayed on the monitor.

Figure 2.2: Illustration of a participant undergoing MEG recording with brain activation maps displayed on the monitor.

c) Functional Magnetic Resonance Imaging (fMRI)

fMRI measures neural activity indirectly by detecting changes in blood oxygenation, using the so-called blood-oxygen-level-dependent contrast. The underlying principle is that neural activity during cognitive tasks increases metabolic demand, primarily through glucose consumption, which subsequently requires an elevated supply of oxygen. As a result, greater concentrations of oxygenated blood within a specific brain region indicate increased metabolic support and therefore suggest that neural activation has occurred in that area. The main disadvantages of fMRI-based brain computer interface are its high cost, complexity of development and delay in data acquisition process (Meitzler et al., 2005, Mhanna et al., 2025).

d) Positron Emission Tomography (PET) Technology

Positron Emission Tomography (PET) enables researchers to visualise functional activity within the living brain and obtain relatively high-resolution measurements of regional metabolic changes associated with cognitive processes and pathological conditions. The technique involves administering trace amounts of a short-lived radiotracer, which emits positrons as it decays. These emissions are detected by the scanner, and regions showing higher levels of radioactivity correspond to increased neural activity (Xie et al., 2024).

e) Electroencephalography EEG

In 1924, Hans Berger achieved the first recording of human brain activity using electroencephalography (EEG). A decade later, in 1934, Adrian and Matthews confirmed the presence of what they called “human brain waves” and described a regular oscillatory pattern at approximately 10–12 Hz, which they named the alpha rhythm (Tassinari et al., 2019). As noted by Bickford (1987), EEG research and its clinical applications in both humans and animals are employed to:

- Monitor alertness, coma and brain death
- Locate areas of damage following head injury
- Test afferent pathways of strokes and tumors (by evoked potentials),
- Monitor levels of cognitive engagement by analysing alpha-rhythm patterns
- Generate controlled biofeedback conditions
- Manage anesthesia depth through automated servo-anesthesia control.
- Examine epileptic activity and determine the focal source of seizures
- Assess the effectiveness of anti-epileptic medications.
- Aid in planning and guiding experimental cortical resections to eliminate epileptic focus areas.
- Track developmental changes in the human brain.
- Assess whether medications induce convulsive responses.
- Analyse sleep-related physiological processes and the neural characteristics of sleep disorders.

It is worth mentioning that EEG is widely used in BCI due to its high temporal resolution, non-invasiveness, and relatively low cost. Figure 2.3 illustrates the placement of a 10-20 electrode cap to measure brain signals.

Figure 2.3:Illustration of an EEG system setup showing a participant wearing an electrode cap connected to a data acquisition unit displaying multi-channel EEG signals.

2.3 Advantages and Disadvantages of Brain Measurement Techniques

To provide a clearer comparison of the methods reviewed, this section summarises the principal advantages and limitations of major brain-measurement techniques, including EEG, fMRI, fNIRS, MEG, and ECoG. Each modality differs substantially in terms of temporal and spatial resolution, cost, portability, susceptibility to artefacts, and level of invasiveness factors that strongly influence their suitability for clinical diagnosis, experimental research, and real-time monitoring. While techniques such as EEG and MEG offer excellent temporal resolution, others like fMRI and ECoG provide superior spatial precision at the expense of increased cost or invasiveness. By consolidating these core characteristics, Table 2.1 presents a structured overview that highlights the strengths and drawbacks of each technique, enabling direct comparison and supporting the selection of an appropriate modality for specific neurological applications.

Table 2.1: Comparison of common brain imaging modalities and their advantages and limitations, adapted from established literature (Naseer M et al, 2015, Tassinari et al., 2019, Smith, Chen and Davis, 2023, Mhanna et al., 2025).

Data Acquisition Method	Advantage	Disadvantage	Invasive / Non-Invasive
EEG	-High temporal resolution -Safe and easy technique -Affordable cost -Portable	-Prone to artefacts arising from ocular (EOG) and cardiac (ECG) activity, electromyographic noise, and power-line interference. -Low spatial resolution -Nonstationary signal	-Non-Invasive
fMRI	-High spatial resolution	- Presence of delays during data acquisition. -Expensive -Low temporal resolution	-Non-Invasive
fNIRS	-Moderate temporal resolution -Inexpensive	-Low spatial resolution -Hinders transformation rates -poorer performance -Susceptible to motion artefacts	-Non-Invasive

MEG	-Wider frequency range -Excellent temporal resolution	- Requires a bulky setup -Expensive experimental setup -Good spatial resolution but not as -high as fMRI	-Non-invasive
ECOG	- Offers high-precision acquisition and superior signal quality for human brain activity. -High spatial resolution	-Causes brain damage since it is invasive	-Invasive

2.4 Electroencephalography (EEG)

EEG measures the brain's electrical activity using electrodes placed on the scalp. Highly conductive silver/silver-chloride electrodes and gold cup electrodes are commonly used to obtain accurate and reliable measurements. Electrode placement typically follows the internationally standardised 10–20 system, established by the American Electroencephalographic Society (Yuan et al., 2021). This system defines electrode positions using two anatomical reference points on the head: the nasion, located at the bridge of the nose at eye level, and the inion, the bony prominence at the base of the skull. Transverse and midline planes are drawn between these points, and electrode locations are marked at intervals of 10% and 20% along these planes, as illustrated in Figure 2.4.

In this system, each electrode label corresponds to a specific brain region: A denotes the ear lobe, C the central region, Pg the nasopharyngeal area, P the parietal region, F the frontal region, Fp the frontal polar area, and O the occipital region.

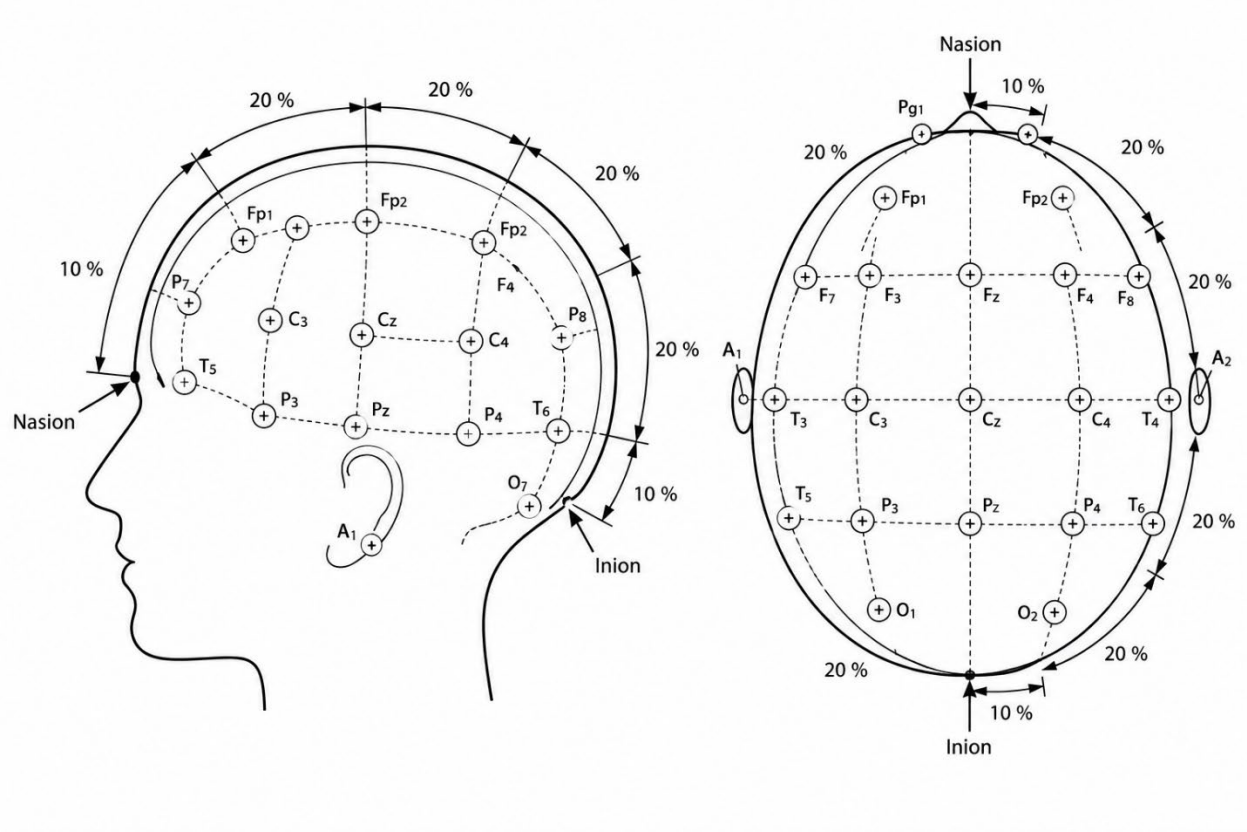


Figure 2.4: Illustration of the international 10-20 EEG electrode placement system showing lateral and top views of the head with standard electrode positions

There are two types of recording techniques for the signals received from these electrodes, namely Monopolar and Bipolar. In a monopolar montage, one active electrode is placed over the area of interest and referenced to an electrode located at an electrically inactive site, along with a ground terminal(Trinka et al., 2015). In contrast, a bipolar montage records the voltage difference between two active electrodes placed over cortical areas of interest. Figure 2.5 represents EEG recording showing 19 scalp channels arranged according to the international 10–20 electrode placement system. The recording is displayed using a longitudinal bipolar montage, where each trace represents the voltage difference between adjacent electrode pairs

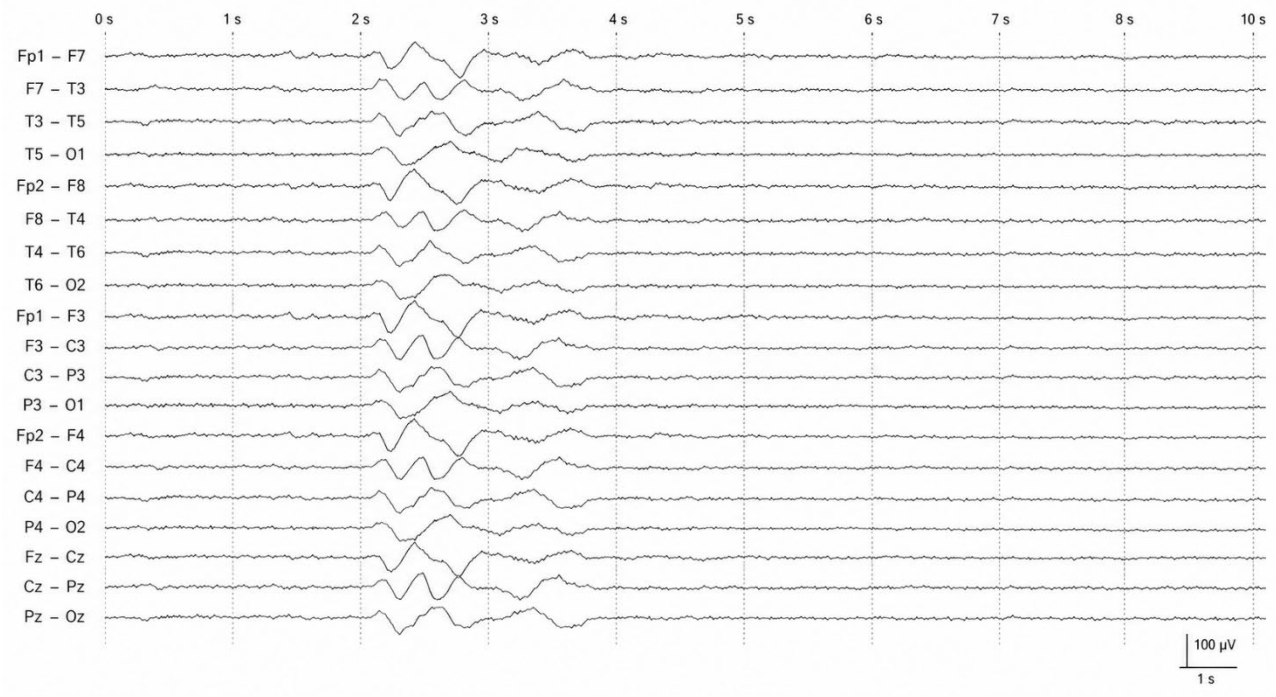


Figure 2.5: Example of a 19-channel EEG recording displayed using the international 10–20 electrode placement system in a longitudinal bipolar montage.

2.5 Challenges in EEG-Based Brain Disorder Detection

Although electroencephalography (EEG) provides a non-invasive and temporally precise method for monitoring neural activity, several inherent challenges limit its reliability and diagnostic performance in brain disorder detection. The primary challenge lies in its low signal-to-noise ratio (SNR) (He, Wu & Wu, 2023; Abdelfattah, Abdelrahman & Wang, 2022). EEG recordings are easily contaminated by artefacts arising from muscle activity, ocular movements, cardiac rhythms, and external electrical interference. Moreover, EEG signals are highly non-stationary, meaning their statistical properties such as mean and variance change over time, exhibiting temporal fluctuations that complicate consistent feature extraction. Significant inter- and intra-subject variability further hinders model generalisation, as EEG patterns can differ substantially between individuals and even within the same subject under varying physiological or cognitive conditions. The limited spatial resolution of scalp EEG restricts its ability to localise deep cortical or subcortical sources, while overlapping neural signatures across disorders such as epilepsy, Alzheimer's disease (Aviles et al., 2024), and schizophrenia (Sairamya, Subathra & George, 2022; Bhardwaj & Kumar, 2024) further complicate diagnostic differentiation.

Additionally, most available EEG datasets are small, imbalanced, and heterogeneous, constraining the training of robust and generalisable machine-learning models. The lack of standardisation in acquisition protocols, electrode configurations, and preprocessing pipelines also compromises the reproducibility of reported results. Finally, although deep learning has improved classification accuracy, these models often operate as black-box systems, offering limited interpretability and reducing clinical trust. Addressing these limitations through advanced signal-processing techniques, robust feature extraction strategies, and explainable AI frameworks remains essential for building reliable, transparent, and clinically meaningful EEG-based diagnostic systems (Abdelfattah, Abdelrahman & Wang, 2022).

Apart from these methodological and computational challenges, the clinical interpretation of EEG recordings remains demanding and resource-intensive, often relying on expert judgement and manual inspection. This highlights the need for automated and standardised analysis approaches to enhance diagnostic reliability and efficiency in clinical environments.

2.6 Clinical Interpretation of EEG

In clinical practice, the diagnosis of brain-related disorders such as epilepsy often relies on the expert visual interpretation of EEG signals. However, this manual interpretation requires extensive expertise and is associated with low inter-rater agreement (IRA, meaning that even experienced clinicians may interpret the same recording differently (Halford et al., 2015; Qiu, Yan & Liu, 2023). The large volume of EEG data generated during both short-term and long-term monitoring further makes interpretation time-consuming, resource-intensive, and costly (Lemoine et al., 2024). These limitations highlight the increasing need for automated diagnostic systems capable of reducing human error, accelerating decision-making, and improving overall patient care (Tveit et al., 2023).

EEG remains widely used due to its high temporal resolution, non-invasive nature, and relatively low equipment cost. Nevertheless, accurate diagnosis often requires multiple recording sessions or prolonged monitoring, as disorder-specific abnormalities may not appear during a single short session. For example, only around 50% of patients with epilepsy exhibit interictal epileptiform discharges (IEDs), which are abnormal electrical patterns occurring between seizures that are characteristic of epilepsy, in their first EEG. Consequently, clinicians must review large amounts of data to identify relevant pathological events, which further emphasises the importance of

automated analysis frameworks capable of detecting clinically significant patterns efficiently and objectively (Roy et al., 2018).

2.7 Available EEG Datasets

Although electroencephalography (EEG) is one of the most widely used modalities for acquiring brain activity signals, there remains a serious shortage of large-scale, publicly available datasets (Wong et al., 2023). The existing open-access datasets vary considerably in terms of recording duration, sampling frequency, and number of subjects, with collections ranging from a few dozen to several thousand recordings (Khan et al., 2022).

The Children's Hospital Boston–MIT (CHB-MIT) Scalp EEG Database is one of the earliest and most extensively used datasets for seizure detection research. It includes EEG recordings from pediatric patients with intractable epilepsy. Data were collected after the withdrawal of anti-seizure medication to monitor seizure activity over several days and evaluate potential candidates for surgical intervention. The dataset comprises 23 cases from 22 subjects (5 males aged 3–22 years and 17 females aged 1.5–19 years). Each recording exhibits various abnormal EEG patterns such as spikes, polyspikes, sharp waves, and slow waves that support the study of epileptic activity (Shoeb and A.H., 2009).

The Karunya University Dataset contains EEG recordings from individuals aged 1 to 107 years, collected at a diagnostic facility in Coimbatore, Tamil Nadu, India. Each record includes the primary classification of the disorder, patient details, digitised EEG data, and diagnostic information. The database also features a search interface to facilitate data retrieval and exploration (Selvaraj et al.,2014).

The European Epilepsy Database (EU) provides annotated EEG recordings from over 250 patients with epilepsy, including 50 cases with intracranial recordings of up to 122 channels. The sampling rate varies between 250 Hz and 2500 Hz, and each dataset contains an average of 150 hours (over five days) of continuous EEG data. This extensive repository enables in-depth studies on long-term seizure monitoring and intracranial analysis (Ihle et al., 2012).

The Neonatal Intensive Care Unit (NICU) Dataset described by (Stevenson et al., 2019) includes EEG recordings from 79 term newborns hospitalised at Helsinki University Hospital. Multi-channel EEGs were recorded for an average duration of 74 minutes (range: 64–96 minutes). Three independent clinical experts visually annotated the presence of seizures, resulting in approximately 460 annotated seizure events per expert. By consensus, 39 neonates exhibited seizures while 22

were seizure-free. This dataset is highly valuable for developing automated neonatal seizure detection algorithms, studying inter-rater agreement (IRA), and establishing reference benchmarks for clinical research (Stevenson et al., 2019).

The University of Bonn Dataset developed by (Andrzejak et al., 2001) is among the most commonly used benchmark datasets for EEG signal analysis. It consists of five subsets (A–E), each containing 100 single-channel EEG segments of 23.6 seconds in duration, resulting in a total of 4097 samples. The simplicity of this dataset, with clearly defined normal and epileptic segments, has made it standard for testing new classification algorithms (Andrzejak et al., 2001).

The Temple University Hospital (TUH) EEG Corpus is the largest publicly available clinical EEG database and is continually being expanded (Obeid, I. and Picone, J., 2016). Collected in a real-world hospital environment over a 14-year period, it contains EEG data and corresponding physician reports from a wide variety of patients, covering different ages, seizure types, and comorbidities. The complete corpus includes 16,986 sessions from 10,874 unique subjects, with recordings sampled primarily at 250 Hz (87 %), while others use 256 Hz, 400 Hz, or 512 Hz. The number of EEG channels per recording ranges between 20 and 31, and some sessions also include supplementary channels such as EKG, EMG, and photic stimuli. Subjects range in age from infancy to over 90 years, with an average of 1.56 sessions per patient, and up to 37 sessions for long-term follow-up cases (Obeid, I. and Picone, J., 2016). A comparative summary of the available datasets, including the number of recordings, recording duration, subjects, and channels, is presented in Table 2.2.

Table 2.2: A comparison between the available datasets.

Dataset	Number of recordings	Duration of each recording	Number of subjects	Number of channels
CHB-MIT (Shoeb and A.H., 2009).	664	60 mins	22	23
Karunya University (Selvaraj et al.,2014)	175	10 seconds	175	16
European Epilepsy Database (Ihle et al., 2012)	250	Up to 150 hours	30	122
NICU (Stevenson et al., 2019).	79	74 mins	79	19
BONN University (Andrzejak et al., 2001)	500	23.6 seconds	5	1
TUH Dataset (Obeid, I. and Picone, J., 2016)	16,986	Approximately 20–60 minutes	10,874	31

Among all publicly accessible EEG repositories, the TUH- EEG Corpus stands out as the largest and most comprehensive clinical dataset available to date. Its extensive size, diversity of patient demographics, and inclusion of detailed physician reports make it an invaluable benchmark for research on automated EEG interpretation. The corpus covers a wide spectrum of seizure types and neurological conditions recorded under real clinical conditions, thereby enhancing its relevance for developing and validating deep learning based diagnostic systems. Owing to its rich variability, standardised annotation structure, and large subject pool, the TUH dataset provides an ideal foundation for this thesis, serving as the primary source for training, testing, and evaluation of the proposed EEG classification models (Obeid, I. and Picone, J., 2016).

2.8 Towards Intelligent EEG-Based Brain Disorder Classification

The last two decades have seen a significant surge in research on automated seizure detection using machine-learning techniques, as well as growing interest in EEG-based analysis for Alzheimer's disease. Numerous algorithms have been proposed to highlight key characteristics of EEG signals

during seizure and non-seizure states and increasingly, to distinguish Alzheimer's-related neural changes based on extracted features (Kim, Youn & Paik, 2023; Amer N.S. et al., 2023; Saminu S. et al., 2023; Gadgil A. et al., 2025; (Aviles et al., 2024).

Despite the widespread use of conventional pattern-recognition techniques, manual feature selection remains a challenge, as it relies heavily on domain knowledge and expert experience. To improve pattern-recognition accuracy, feature-extraction and classification algorithms have been developed using various optimisation objectives. Deep learning networks are capable of extracting more distinct and meaningful representations directly from EEG signals. Unlike classical approaches, deep-learning methods integrate feature extraction and classification into a single framework, preventing potential information loss that may occur when these processes are performed separately. Consequently, deep-learning-based EEG classification has gained substantial attention as a promising research direction (Rahul J. et al., 2024).

Building upon the TUH EEG Corpus as the foundation dataset (Obeid, I. and Picone, J., 2016), the effectiveness of automated brain-disorder detection largely depends on the feature extraction and classification strategies applied to the raw EEG signals. EEG data are high-dimensional, non-stationary, and often contaminated by noise; therefore, transforming these raw recordings into compact, discriminative features is a crucial step in the analysis pipeline (Yap HY et al., 2023). The following subsection discusses the EEG feature extraction techniques, outlining both traditional handcrafted and deep learning-based approaches that underpin modern automated classification frameworks.

2.8.1 EEG Feature Extraction and Classification Techniques

Feature extraction from EEG signals plays a fundamental role in accurately classifying neural activity associated with various brain disorders (Vega-Escobar et al., 2015; Singh & Krishnan, 2023). Numerous feature-extraction methods have been proposed, including the Fast Fourier Transform (FFT) (Fisher, 1936;), Independent Component Analysis (ICA) (Rejer and Górski, 2015; Tharwat, 2020), Principal Component Analysis (PCA) (Bugli and Lambert, 2007), the Wavelet Transform (Chaovalit et al., 2011; Upadhyay, 2013), Autoregression (AR) (Huan and Palaniappan, 2004), Wavelet Packet Decomposition (WPD), and Local Fisher's Discriminant Analysis (LFDA) (García-Laencina et al., 2014).

PCA and ICA reduce dimensionality, which is particularly advantageous for lengthy EEG recordings (Matin et al., 2019; Tharwat, 2020). ICA often provides a more informative representation than PCA for analysing event-related potentials (ERPs), which reflect cognitive or pathological brain responses (Artoni et al., 2018).

EEG classification algorithms can generally be divided into linear (Al-Fahoum & Al-Fraihat, 2014) and non-linear categories (Saneesh Cleatus & Thungamani, 2021). Linear classifiers assign EEG samples to classes using linear decision boundaries. Common examples include Logistic Regression (LR) (Guerrero et al., 2021), Naïve Bayes (Sharmila and Geethanjali, 2016), and Linear Discriminant Analysis (LDA) (Atangana et al., 2020). Non-linear classifiers include Support Vector Machines (SVM) (Chandaka et al., 2009; Antony et al., 2022), Neural Networks (NN) (Bhatti et al., 2019), Multi-Layer Perceptron (MLP) (Narang et al., 2018), Random Forests (RF) (Antoniou et al., 2021), Decision Trees (DT) (Duan et al., 2015), k-Nearest Neighbours (KNN) (Sha'Abani et al., 2020) Quadratic Discriminant Analysis (QDA). Figure 2.6 illustrates a hierarchical representation of common EEG classification techniques, categorised into linear and non-linear methods. This taxonomy illustrates the major algorithmic families used in EEG-based brain disorder detection.

Support Vector Machines (SVMs) have been widely adopted as binary EEG classifiers, with early demonstrations by Bhuvaneswari and Satheesh Kumar (2013). More recent advances were introduced by Miao Shi et al. (2020), who optimised SVM parameters using an Improved Squirrel Search Algorithm (ISSA), combining preprocessing, time-domain feature extraction, and enhanced initialisation strategies to improve exploration capability and convergence speed.

Turker et al. (Tuncer et al., 2021) further extended SVM-based methods through Wavelet Packet Decomposition and Chaotic Local Binary Patterns (CLBP), selecting features using the Iterative Minimal-Redundancy Maximum-Relevance (ImRMR) method and applying SVM with post-processing to reduce false alarms.

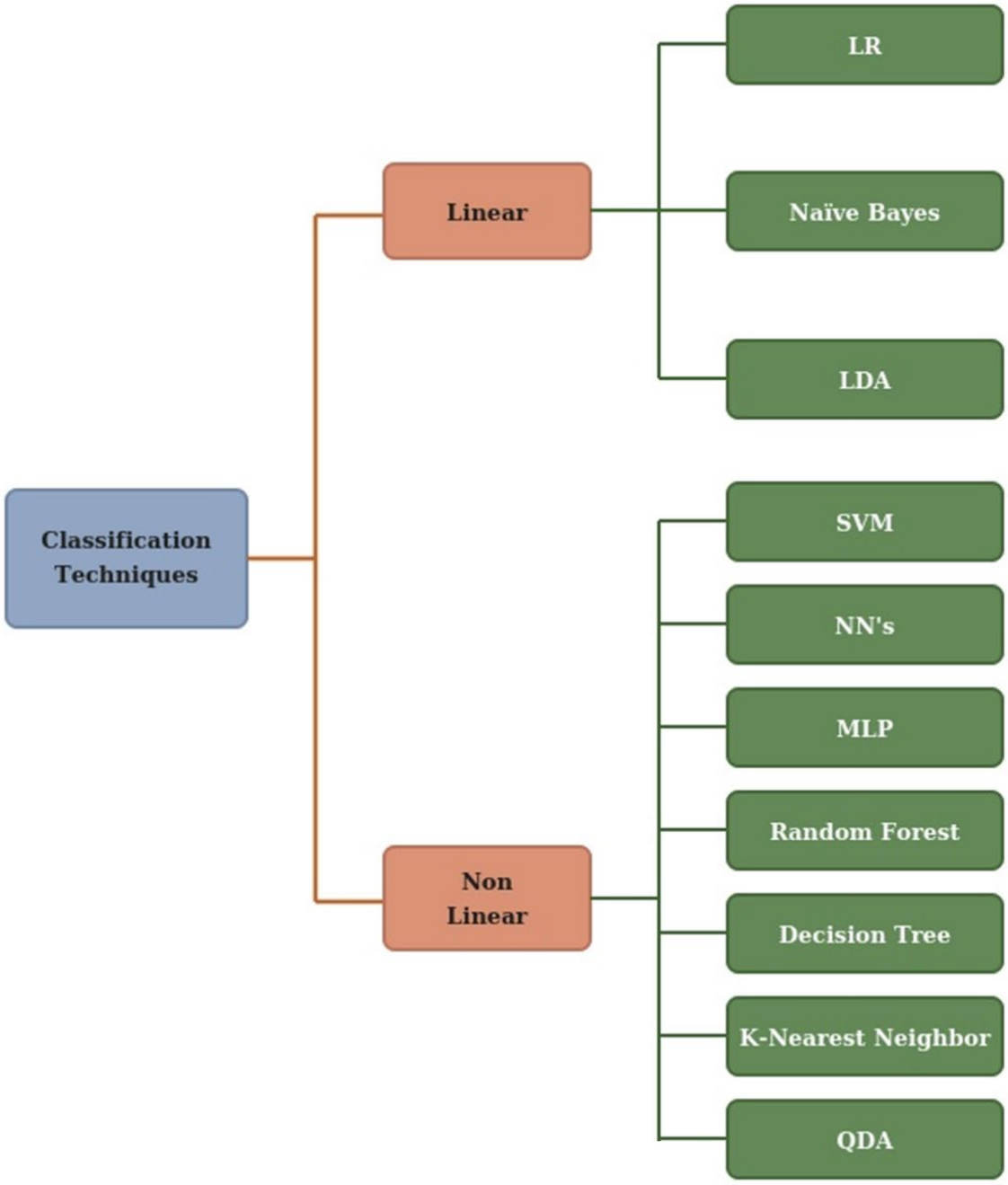


Figure 2.6: Major Algorithmic Families Used in EEG-Based Brain Disorder Classification

Linear Discriminant Analysis (LDA) has also been extensively integrated with EEG pipelines. Subasi and Gursoy (2010) combined PCA, ICA, and LDA for dimensionality reduction prior to SVM classification, improving separation between epileptic and non-epileptic EEG signals. Khan et al. (2012) incorporated wavelet-based features with LDA to distinguish seizure from non-seizure activity, while Hasan et al. (2015) applied LDA to motor-task EEG analysis and noted that its underlying assumptions of normality and equal covariance are frequently violated in clinical data (Martišius, 2016). Using Stationary Wavelet Transform features, Orosco et al. (2016) achieved high specificity (99.9%) and sensitivity (87.5%) with LDA. Fathima and Bedeeuzzaman (2016) compared LDA and SVM within a 1D-LBP seizure prediction pipeline, showing that SVM yielded superior sensitivity and fewer false positives.

Artificial neural networks have also been widely adopted. Huan and Palaniappan (2004) designed a two-state ANN using AutoRegressive (AR) features, while Hindarto et al. (2018) applied wavelet-derived features with an ANN to classify mental-state EEG patterns. Hengjin et al. (Ke et al., 2018) used global synchronisation features with a CNN for EEG classification. Mohan and Perumal (2023) analysed EEG data from patients with depression and anxiety using a Fractional Order Butterworth Filter and feature extraction across multiple frequency bands; feature selection via MRMR followed by deep CNN, ANN, and KNN classification showed CNN achieving 97.6% accuracy, demonstrating its suitability for cost-effective mental-health detection.

Comparative studies have also evaluated classifier performance. Thieu and Yang (2015) analysed FFT-based features using a linear neural network and benchmarked it against Naïve Bayes and Logistic Regression (LR). Alturki et al. (2020) developed a unified diagnostic framework for epilepsy and autism spectrum disorder using LDA, SVM, KNN, and ANN, reporting accuracies up to 99.9% with SVM. Among non-linear models, Arabi et al. (Aarabi, Wallois and Grebe, 2006) used a three-layer Backpropagation Neural Network (BNN) optimised via Fast Supervised Backpropagation Training Algorithm (FSBPTA) for neonatal seizure detection. Nasehi and Pourghassem (2013) proposed an Improved Particle Swarm Optimisation (IPSO) neural network for patient-specific seizure onset detection, enhancing speed and accuracy. Fergus et al. (2016) classified seizure and non-seizure segments using KNN and achieved balanced performance on the CHB-MIT dataset, while Almoustafa (2020) showed that Random Forest outperformed KNN, Naïve Bayes, LR, and Decision Trees for EEG classification.

Hybrid and adaptive approaches have continued to expand capability. Subasi (2007) introduced a Mixture-of-Experts (ME) neural model that combined multiple linear classifiers through a gating network. Fiscon et al. (2018) used Fourier and Wavelet Transforms to classify cognitive impairment conditions including MCI and AD showing that wavelet features offer superior discriminatory ability. Upadhyay (2013) applied Daubechies-2 wavelets with a one-step-secant ANN training algorithm, achieving accuracies up to 98%. Kannathal et al. (2005) utilised entropy and approximate entropy with ANFIS for epilepsy detection, while Sadati et al. (2006) introduced an Adaptive Neural Fuzzy Network (ANFN) to distinguish normal from epileptic EEGs. Hendel et al. (2016) combined a Self-Organising Map (SOM) with Probabilistic Quadratic Loss SVM, achieving accuracies between 81.7% and 91.9% across five mental-state categories. Nunes et al. (2014) introduced the Optimum Path Forest (OPF) classifier, outperforming SVM-RBF, Bayesian, and ANN-MLP models in accuracy and computational efficiency. Bilal et al. (2019) improved seizure-detection robustness under class imbalance using the RASBoost algorithm, which unites resampling with boosting.

Entropy-based methods have also been extensively used for characterising EEG complexity. Measures such as Permutation Entropy (PE), Approximate Entropy (ApEn), and Sample Entropy effectively capture the transition between normal and ictal states (Yuan et al., 2011; Nicolaou & Georgiou, 2012; Hariharan et al., 2014; Abdelhameed & Bayoumi, 2021). Amin et al. (2017) combined DWT, LFDA, and PCA to extract discriminative and stable features for pathological EEG classification.

Overall, a wide spectrum of feature-extraction and classification approaches from traditional statistical and frequency-based techniques to advanced neural, fuzzy, and hybrid architectures has been explored for EEG-based brain disorder detection. Despite considerable progress, challenges such as non-stationarity, inter-subject variability, and class imbalance persist, reinforcing the need for deep learning frameworks capable of jointly learning feature representations and performing end-to-end classification.

2.8.2 Deep Learning- Based EEG Features Extraction and Interpretation

Deep learning is a subfield of machine learning that models complex data through layered neural architectures, with each layer learning increasingly abstract and discriminative representations. Unlike traditional approaches that rely on handcrafted features, deep learning networks can

automatically extract hierarchical patterns directly from raw EEG data. These models can be trained in supervised, unsupervised, or reinforcement-learning settings and are highly effective for both feature extraction and classification tasks (Nour, Senturk & Polat, 2024).

Artificial Neural Networks (ANNs) process input signals through multiple hidden layers, each of which encodes specific characteristics of the data. This hierarchical learning structure enables deep models to capture non-linear, high-dimensional EEG patterns associated with various neurological disorders (Sairamya, Subathra & George, 2022). A key advantage of deep learning lies in its ability to minimise manual preprocessing and feature engineering while achieving high classification accuracy (Alom et al., 2019).

a) Recurrent and Convolutional Neural Networks

Among deep learning architectures, Recurrent Neural Networks (RNNs) and their advanced variants Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are widely used for EEG analysis due to their capacity to capture temporal dependencies in sequential data (Ocak, 2009; Hussein, 2016; Bablani et al., 2020; León et al., 2021; Phutela et al., 2022; Choi et al., 2022; Najafi et al., 2022). These architectures are also referred to as Feedback Neural Networks (FNNs), as they maintain internal memory states that allow contextual information from previous time steps to influence future predictions.

Deep learning methods have been extensively applied to seizure detection using major EEG repositories such as the Temple University Hospital (TUH) EEG Corpus and the CHB-MIT dataset (Schirrmeister et al., 2017; L.S., 2017; Golmohammadi et al., 2019; Iešmantas and Alzbutas, 2020; Sharma, Patel and Acharya, 2020; Yıldırım, Baloglu and Acharya, 2020). Early approaches explored nonlinear preprocessing filters and diagnostic ANNs (Ocak, 2009), while later studies introduced LSTM-based sequence models for robust temporal modeling (Hussein, 2016).

More recent research has harnessed Convolutional Neural Networks (CNNs) for end-to-end seizure classification, eliminating the need for handcrafted features (Schirrmeister et al., 2017; Li et al., 2020). Hybrid architecture has also gained traction: Sharma, Parashar and Joshi (2021) proposed DepHNN, which integrates CNN-based temporal feature extraction with LSTM-based sequential modelling. Similarly, Iešmantas and Alzbutas (2020) demonstrated that transforming raw EEG signals into 2D representations such as phase-locking value, entropy, and energy enables CNN models to capture complex spatial and temporal patterns effectively.

b) Recent Architectures and Attention Mechanisms

Recent advances have extended deep-learning capabilities through attention mechanisms, multi-view learning, and transformer-based architectures (Vafaei and Hosseini, 2025). Earlier classical machine-learning approaches focused on engineered features. For example, Nigam and Graupe (2004) used multi-stage nonlinear filtering with an adaptive neural network (LAMSTAR), while Polat and Güneş (2007) developed a hybrid FFT–Decision Tree framework that achieved reliable seizure detection.

Chiang et al. (2011) introduced an online retraining system for continuous seizure prediction, and Zandi et al. (2013) proposed a variational Gaussian Mixture Model (GMM) that effectively distinguished preictal and interictal states. Rasekhi et al. (2013) further improved seizure prediction by integrating multiple linear univariate features into a high-dimensional space and classifying them using machine-learning algorithms.

The rise of deep learning brought significant improvements. Truong et al. (2018) applied CNNs with STFT features for robust seizure prediction, while Yuan et al. (2018) proposed ChannelAtt, a multi-view deep model incorporating channel-aware attention for multi-channel EEG representation. Golmohammadi et al. (2019) combined Hidden Markov Models (HMMs) with deep-learning-based post-processing to achieve high sensitivity. Ozal (Yıldırım, Baloglu and Acharya, 2020) developed a 1D CNN for abnormal EEG classification using the TUH dataset. Likewise, Iešmantas and Alzbutas (2020) demonstrated that converting EEG signals into 2D feature maps (entropy, energy, PLV) enhances CNN-based seizure prediction.

More recent work integrates multi-scale frequency representations and attention mechanisms. Qi Xin et al. (2022) introduced a Wavelet CNN with attention for multi-scale feature extraction, while Wu et al. (2022) enhanced discrimination performance using age information and DWT-derived statistical features with ensemble classifiers. Yan et al. (2022) developed a transformer-based seizure prediction model that uses STFT features and a three-tower self-attention architecture to overcome sequence length limitations. Najafi et al. (2022) proposed an LSTM-based framework using longitudinal bipolar montage EEG for distinguishing focal and generalised epilepsy.

A comprehensive summary of EEG-based classification studies is provided in Table 1 (Appendix A), detailing the datasets used, feature extraction methods, classifiers, and corresponding accuracy

reported in the literature. The distribution of feature extraction and classification techniques derived from these studies is visualised in Figures 2.7 and 2.8, respectively.

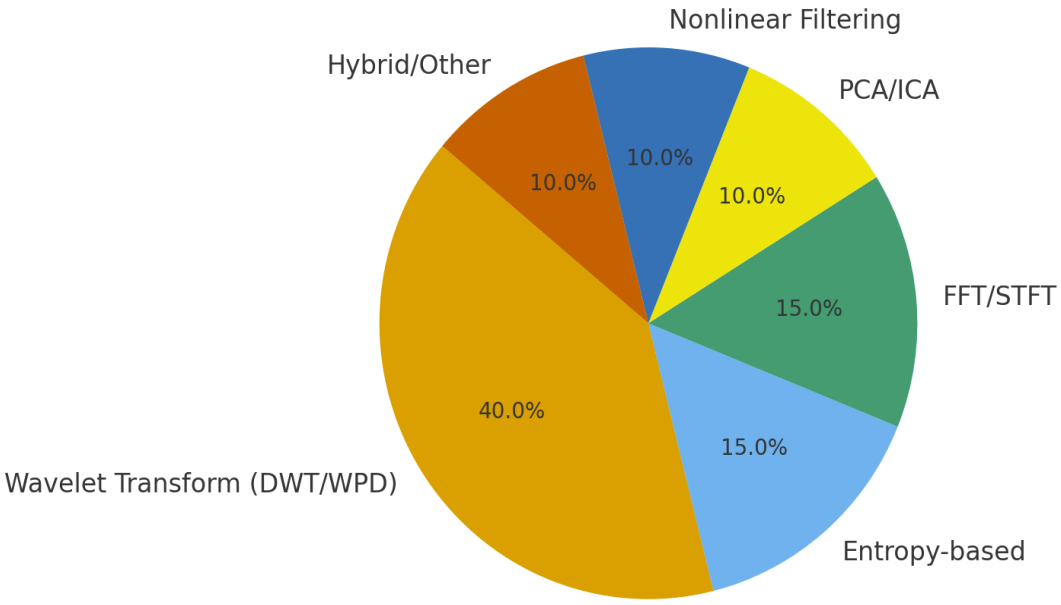


Figure 2.7: Distribution of Feature Extraction Techniques used in EEG Classification.

Figure 2.7 presents the distribution of feature-extraction techniques commonly employed in EEG-based brain disorder classification studies. Wavelet-based methods, including Discrete Wavelet Transform (DWT) and Wavelet Packet Decomposition (WPD), dominate the literature, accounting for approximately 40% of all studies due to their strong ability to capture both time- and frequency-domain characteristics of non-stationary EEG signals. Entropy-based approaches and FFT/STFT methods each represent around 15% of reported techniques, reflecting their usefulness in quantifying signal complexity and spectral content. PCA and ICA contribute an additional 10%, primarily supporting dimensionality reduction and artefact removal. Nonlinear filtering methods and hybrid feature-extraction strategies each comprise roughly 10%, often appearing in frameworks that combine multiple transformations for enhanced discriminative representation. Overall, the distribution highlights the continued prominence of time–frequency analysis in EEG research and the growing interest in hybrid and nonlinear methods for improved robustness in real-world clinical applications.

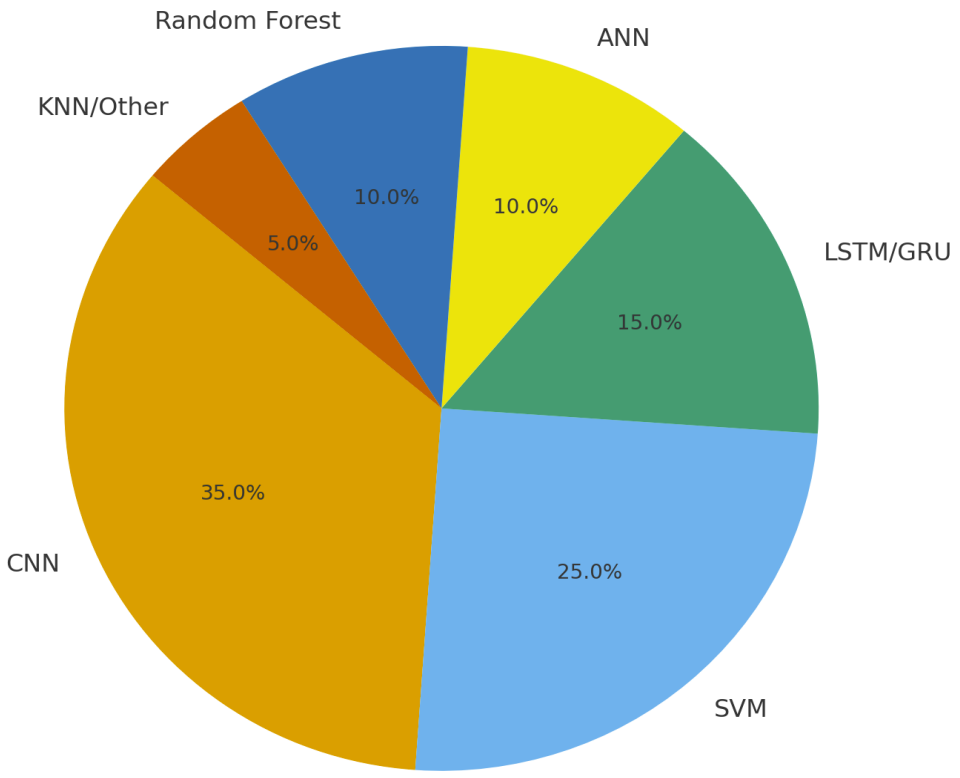


Figure 2.8: Distribution of Classification Techniques used in EEG Classification

Figure 2.8 illustrates the proportional use of major classification techniques in EEG-based brain disorder detection. Convolutional Neural Networks (CNNs) dominate the landscape, representing 35% of reported studies, underscoring their strong capability for automated spatial temporal feature learning directly from raw or minimally processed EEG signals. Support Vector Machines (SVMs) account for 25%, reflecting their continued relevance as robust baseline classifiers. Recurrent architectures such as LSTM and GRU represent 15%, highlighting their suitability for modelling temporal dependencies within sequential EEG data. Artificial Neural Networks (ANNs) and Random Forest classifiers contribute 10% each, typically appearing in traditional machine-learning pipelines or ensemble methods. KNN and other miscellaneous techniques account for the remaining 5%. Collectively, the figure demonstrates a clear research shift toward deep-learning-based models, with CNNs emerging as the predominant choice in contemporary EEG classification studies.

2.9 Summary and Discussion

This chapter has presented an extensive review of the literature on EEG-based brain disorder detection, with particular emphasis on the evolution of feature-extraction, classification, and deep-learning methodologies. The discussion began with an overview of the major challenges inherent in EEG signal analysis, including low signal-to-noise ratio, non-stationarity, and substantial inter- and intra-subject variability factors that significantly complicate reliable pattern recognition. These challenges highlight the need for advanced preprocessing approaches and robust feature-selection strategies capable of extracting meaningful information from noisy EEG recordings.

A detailed comparison of publicly available EEG datasets showed that although several options exist such as CHB-MIT, Bonn, and NICU the Temple University Hospital (TUH) EEG Corpus stands out as the most comprehensive and clinically diverse resource. Its scale, long recording durations, and inclusion of heterogeneous neurological conditions make it particularly well-suited for developing and evaluating deep-learning models at large scale.

The review of feature-extraction and classification techniques demonstrated a clear shift from traditional handcrafted features (e.g., Wavelet Transform, PCA, ICA) and classical machine-learning models (e.g., SVM, LDA, Logistic Regression) toward more adaptive deep-learning architectures. While conventional approaches have achieved acceptable accuracy, their performance is often restricted by the reliance on manual feature engineering, imbalanced datasets, and limited generalisability.

In contrast, modern deep-learning models including CNNs, RNNs, LSTMs, GRUs, and hybrid CNN-LSTM frameworks have shown superior performance by automatically learning rich spatial–temporal representations of EEG signals. Recent advancements incorporating attention mechanisms and transformer-based architectures have further enhanced temporal modeling and interpretability. Nonetheless, the black-box nature of deep learning, along with high computational demands and limited clinical explainability, continues to pose significant challenges for real-world deployment.

Building on these insights, the next chapter introduces the methodological framework developed in this thesis to overcome the limitations identified in existing EEG-based classification approaches. Chapter 3 presents the proposed ensemble methodology, detailing the end-to-end pipeline that integrates EEG preprocessing, windowing-based feature selection, multi-strategy feature extraction, the design of an 18-layer CNN-LSTM architecture, and an ensemble Random

Forest stacking mechanism. This framework forms the foundation of the experimental work that follows, providing a structured and robust approach for enhancing diagnostic accuracy, sensitivity, and generalisability in automated EEG-based brain disorder detection.

Chapter 3
Methodological Ensemble Framework

3.1 Introduction

This chapter presents the methodological framework developed for automated EEG-based brain disorder detection. The proposed pipeline integrates all essential components of the analysis process, including signal preprocessing, window-based feature selection, feature extraction, deep learning model design, and ensemble decision-making. Each stage of the framework was formulated to address the inherent challenges of EEG signals, such as high dimensionality, temporal variability, noise contamination, and the need to optimise sensitivity and specificity for clinically meaningful detection. The methodology outlined in this chapter establishes a systematic and robust foundation for the experimental design and evaluation presented in the subsequent chapters. Figure 3.1 presents the conceptual overview of the proposed ensemble framework, highlighting the integration of multi-model feature extraction with deep learning and a stacking-based decision mechanism.

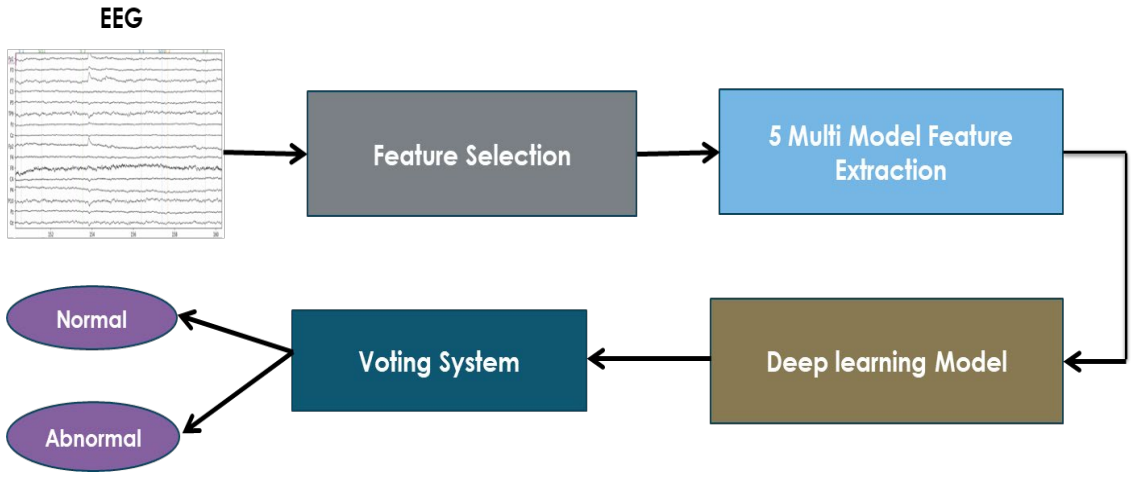


Figure 3.1: EEG-Based Brain Disorder Detection Framework Using Multi-Model Feature Extraction and stacking System.

The pipeline begins with EEG data acquisition and preprocessing, where artifacts are removed and the signals are standardised for analysis. A novel windowing-based feature selection strategy is then introduced to segment EEG signals into clinically meaningful temporal intervals. This is followed by multi-model feature extraction, in which statistical representations such as Maximum, Minimum, Average, Odd-indexed, and Even-indexed sampling are derived to highlight complementary signal characteristics.

Building on these inputs, a deep CNN–LSTM hybrid model is constructed to capture both spatial and temporal dynamics in the EEG data. The CNN layers are employed to extract discriminative spatial features, while the LSTM layers model sequential dependencies across time. Finally, to enhance robustness and overcome the limitations of individual classifiers, an ensemble stacking mechanism based on Random Forest is proposed, integrating the outputs of multiple CNN–LSTM models trained on different feature extraction strategies.

The remainder of this chapter is organised as follows. Section 3.2 describes the datasets used and the preprocessing procedures applied. Section 3.3 presents the proposed windowing-based feature selection method. Section 3.4 details the five multi-model feature extraction strategies. Section 3.5 illustrates the evaluation of ML Models. Section 3.6 introduces the design and training of the CNN–LSTM architecture. Section 3.7 explains the ensemble stacking mechanism, while Section 3.8 summarises the chapter and transitions to the experimental evaluation in Chapter 4.

3.2 Data Acquisition and Preprocessing

3.2.1 TUH Dataset

The experimental framework adopted in this thesis is based on the TUH- EEG Corpus, which is currently the largest publicly available clinical EEG dataset. The TUH corpus consists of real-world EEG recordings collected from a heterogeneous patient population presenting a wide range of neurological conditions, making it a suitable benchmark for evaluating automated EEG analysis and abnormality detection methods.

All EEG recordings in the TUH corpus follow the international 10–20 electrode placement system and were acquired under routine clinical conditions. Depending on the acquisition protocol, the signals are sampled at rates typically ranging between 250 Hz and 500 Hz. The dataset includes recordings labelled as normal or abnormal.

3.2.2 Dataset Selection and Preprocessing

The TUH EEG signals were pre-processed to ensure that the input to the feature selection and classification models was consistent and standardised. Since the dataset was acquired in a clinical environment, several filtering steps were already applied at the hardware level. The preprocessing procedure in this study therefore consisted of the following stages:

- Channel Selection: A total of 19 electrodes corresponding to the standard international 10–20 system were retained from each recording. This ensured consistency across patients and captured the most clinically relevant channels.
- Filtering (Dataset-Level Acquisition): The TUH recordings were acquired with standard clinical filters embedded in the acquisition system. A bandpass filter of 0.5–70 Hz was applied to reduce baseline drift and high-frequency artefacts, while a 60 Hz notch filter was used to suppress mains interference. As these filters were applied during data collection, no additional digital filtering was performed in this study. Figure 3.2 shows multichannel EEG segment recorded over a 10 second interval, illustrating signals from 19 bipolar EEG channels derived according to the international 10–20 electrode placement system. The signals are shown in microvolts (μV) and demonstrate typical background activity with transient abnormal patterns across different electrode pairs.
- Segmentation: Continuous EEG signals were divided into fixed-length windows according to the proposed windowing method (see Section 3.3). This segmentation enabled temporal feature learning and supported consistent input lengths for the CNN–LSTM architecture.

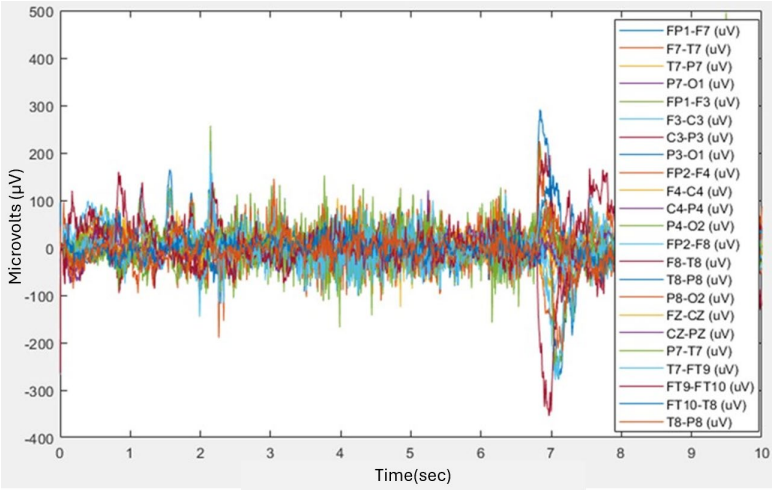


Figure 3.2: EEG Data representation of one patient.

3.3 Proposed Windowing-Based Feature Selection Method

EEG signals are non-stationary and contain time-varying frequency components, which makes the selection of informative temporal segments critical for accurate classification. To address this challenge, a novel windowing-based feature selection strategy was developed in this study. Unlike conventional fixed segmentation methods, the proposed approach adaptively identifies the most informative interval for each electrode by analysing the time–frequency representation of the signal.

3.3.1 Time Frequency Analysis with Generalised Morse Wavelets

Time frequency analysis is employed as the primary technique for processing the EEG signals in this work. Specifically, the analysis was conducted using the continuous wavelet transform (Gurrala, Yarlagadda and Koppireddi, 2019) with Generalised Morse Wavelets (GMW) (Olhede and Walden, 2002). GMWs form a family of analytic wavelets whose Fourier transforms are supported exclusively on the positive real axis, making them particularly suitable for complex-valued wavelet representations. These analytic wavelets are highly effective for examining signals whose amplitude and frequency vary over time, commonly referred to as modulated signals. Non-analytic wavelets are typically used to enhance sharp transitions within the time frequency representation, whereas analytic wavelets are better suited for characterising oscillatory behaviour and frequency transients. Generalised Morse Wavelets (GMWs) address limitations associated with non-analytic wavelets, particularly the presence of interference patterns and artefacts in the time–frequency plane (Lilly and Olhede, 2012; Lilly, 2017; Suwansawang and Halliday, 2017; Martinez-Ríos et al., 2022). The frequency-domain formulation of GMWs, expressed in terms of ω-representation, is given in Eq. (1).

$$\Psi_{\beta,\gamma}(\omega) = U(\omega)\, a_{\beta,\gamma}\, \omega^{\beta} e^{-\omega\gamma} \tag{3.1}$$

Where $\Psi\beta,\gamma$ denotes the frequency-domain representation of the GMWs, U is the unit step function, $a\beta, \gamma$ is the normalisation constant. The parameter βcontrols the decay of the wavelet in the time domain, whereas γ governs its decay in the frequency domain. For a GMW to be valid, the conditions $\gamma > 0$ and $\alpha > 0$must be satisfied. The normalisation constant, with e representing Euler's number, the normalisation constant is governed by $a\beta, \gamma \geq 2(e\gamma/\beta)\ \beta/\gamma$. When $\gamma = 1$, the GMW family corresponds to a solution of the Schrödinger equation (Olhede and Walden, 2002). For $\gamma = 2$, the GMW reduces to the symmetric Derivative-of-Gaussian wavelet family. Owing to

these properties, GMWs are widely recommended as a suitable general-purpose wavelet choice (Olhede and Walden, 2002; Suwansawang and Halliday, 2017).

In this study, the parameter γ was set to 3 and the time–bandwidth product was fixed at 60. These Generalised Morse Wavelet settings were selected based on both theoretical justification and empirical performance. Choosing $\gamma = 3$ yields a wavelet that is highly symmetric, closely approximates a Gaussian shape, and provides strong time–frequency concentration properties that make it one of the most effective members of the GMW family. This configuration also offers a favourable trade-off between temporal and spectral localisation, similar to the Airy wavelet, which is nearly symmetric in both the time and frequency domains (Suwansawang and Halliday, 2017). Such symmetry is particularly advantageous when analysing the oscillatory characteristics of EEG signals.

3.3.2 Windowing Procedure

The proposed windowing method adaptively selects the most informative temporal interval of the EEG signal by analysing its time–frequency representation. The detailed steps of the procedure, illustrated in Figure 3.3, are described as follows:

1. Time–Frequency Transformation: For each electrode, the continuous EEG signal is transformed into the time–frequency domain using the continuous wavelet transform (CWT) with Generalised Morse Wavelets (GMW). This produces a magnitude scalogram that captures both oscillatory and transient characteristics of the EEG.

2. Identification of Maxima: The absolute maximum of each wavelet transform is located, corresponding to the time interval where the signal exhibits the strongest oscillatory activity. This point represents the most distinctive behaviour within the EEG, potentially linked to abnormal neural patterns.

3. Outlier Rejection: To avoid misleading selections caused by artifacts or isolated peaks, extreme maxima (far outliers) are removed. This ensures that the chosen maxima reflect consistent, physiologically relevant activity rather than noise.

4. Averaging Across Electrodes: Since EEG is recorded simultaneously across multiple electrodes, the average position of significant maxima is computed across all 19 channels.

This step aligns the analysis spatially, ensuring that the selected window reflects global brain activity rather than channel-specific fluctuations.

5. Window Selection: A temporal window is centred around the average maximum position. To ensure informative features are preserved, additional samples immediately before and after the maximum are included, preventing truncation of relevant oscillatory patterns.
6. Final EEG Segment: The resulting segment constitutes the optimal window for downstream feature extraction and classification, providing a focused and representative input to the learning models.

To evaluate the effect of window length on classification accuracy, three different window sizes were tested:

-Narrow window: 500 samples (~2 seconds at 250 Hz), providing high temporal resolution.

-Medium window: 2,000 samples (~8 seconds), balancing temporal resolution with frequency context.

-Wide window: 8,000 samples (~32 seconds), offering extensive contextual information but at the risk of including non-informative data.

The windowing procedure is formally presented as a pseudocode algorithm in Algorithm 3.1, which provides a step-by-step mathematical formalisation of the proposed feature selection method.

These experiments were conducted on all 19 electrodes for each patient, enabling a systematic comparison of the effect of window size on classification performance. The outcomes of these experiments are presented in Chapter 5, where the trade-offs between temporal resolution, feature richness, and classification accuracy are analysed in detail.

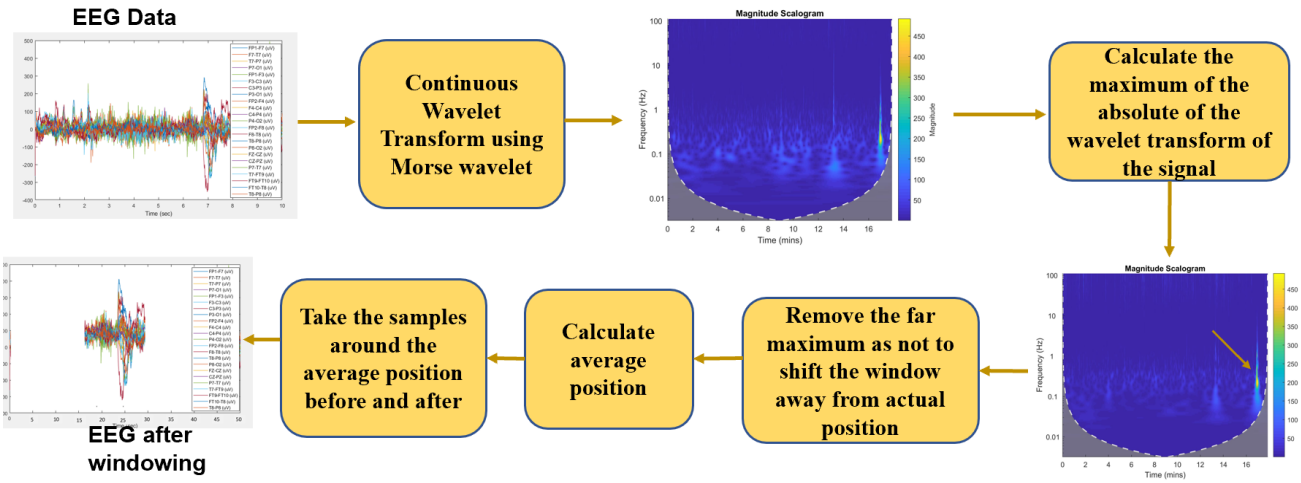


Figure 3.3: Workflow of the proposed windowing-based feature selection method.

Algorithm 3.1: Windowing-Based Feature Selection

Input / Output

Input: EEG signal X with N samples across E electrodes

Output: Optimal windowed segment W of length L

Notation

`X_e`	EEG signal of electrode e
`CWT`	Continuous Wavelet Transform using Generalised Morse Wavelets (GMW)
`M_e`	Magnitude scalogram of electrode e
`t_e`	Time index of maximum wavelet coefficient for electrode e
`t_avg`	Mean peak position across all electrodes
`σ`	Standard deviation of all peak positions
`3σ`	Outlier threshold: peaks beyond 3 standard deviations from the mean are considered artefacts and excluded
`L`	Window length ∈ {500, 2000, 8000} samples
`W`	Final selected temporal window

Steps

```
1. FOR each electrode e = 1 TO 19 DO
    2. Compute CWT of X_e using GMW
    3. Extract magnitude scalogram M_e = |CWT(X_e)|
    4. Identify peak index t_e = argmax(M_e)
5. END FOR
-----------------------------------------------------------------------------------------
6. Compute mean peak position: t_avg = (1/E) × Σ t_e
7. Remove outliers: exclude t_e WHERE |t_e − t_avg| > 3σ
8. Recompute t_avg from remaining electrodes
-----------------------------------------------------------------------------------------
9. Select window W centred at t_avg with length L
10.RETURN W
```

3.3.3 Strengths of the Proposed Method

The proposed wavelet-based windowing method presents notable strengths over conventional fixed-length segmentation approaches. By adaptively selecting the window based on actual signal dynamics rather than arbitrary divisions, it ensures that the analysis focuses on intervals carrying the most clinically relevant information. This adaptivity improves the discriminability of abnormal EEG patterns, as the method targets the period of maximum oscillatory activity where pathological signatures are most likely to appear. In addition, robustness to noise is achieved by discarding

extreme maxima, thereby reducing the risk of selecting spurious or artifact-driven intervals. A further strength of the method lies in its consistency across electrodes; the averaging step ensures that the selected window represents coordinated brain activity rather than isolated fluctuations in individual channels. Collectively, these strengths enhance the representativeness of the EEG features and contribute to improved classification accuracy in brain disorder detection.

3.3.4 Rationale for Selecting the Proposed Windowing Strategy

During the development of the proposed framework, two feature selection strategies were investigated. The first employed correlation-based electrode selection to reduce input dimensionality by selecting subsets of five and eight minimally correlated electrodes. The second introduced the proposed adaptive windowing strategy, which preserves information from all nineteen EEG channels while identifying the most informative temporal segment within each recording. The adaptive windowing approach was ultimately adopted because it retained the complete spatial information provided by the electrode configuration while focusing the analysis on the most discriminative temporal interval. The comparative experimental evaluation of these two feature selection strategies is presented in Chapter 5.

3.4 Multi Feature Extraction

Following the selection of the optimal temporal window using the proposed wavelet-based method, a set of feature extraction strategies was applied to capture diverse representations of the EEG signal. Since EEG recordings are complex, highly variable, and prone to noise, no single extraction method can fully characterise all relevant signal properties. To address this limitation, five distinct feature extraction strategies are employed: Minimum (Min), Maximum (Max), Average (Avg), odd-sample selection, and even-sample selection. Each strategy emphasises different aspects of the underlying signal and, when combined, provides complementary perspectives that enhance discriminability.

The workflow of the proposed multi-model feature extraction framework is illustrated in Figure 3.4. The input EEG segment is processed using five distinct techniques to generate complementary feature representations. Mean pooling computes local averages to smooth high-frequency noise, max pooling emphasises transient high-amplitude events such as epileptiform spikes, and min pooling highlights inhibitory phases and low-amplitude activity. In addition, odd and even decimation subsample the signal into two complementary representations, ensuring that

temporally localised features are preserved. Collectively, these five strategies provide diverse and complementary views of the EEG signal, enhancing feature discriminability and supporting robust classification in the subsequent CNN–LSTM modelling stage. The individual feature extraction strategies are described in detail in the following subsections.

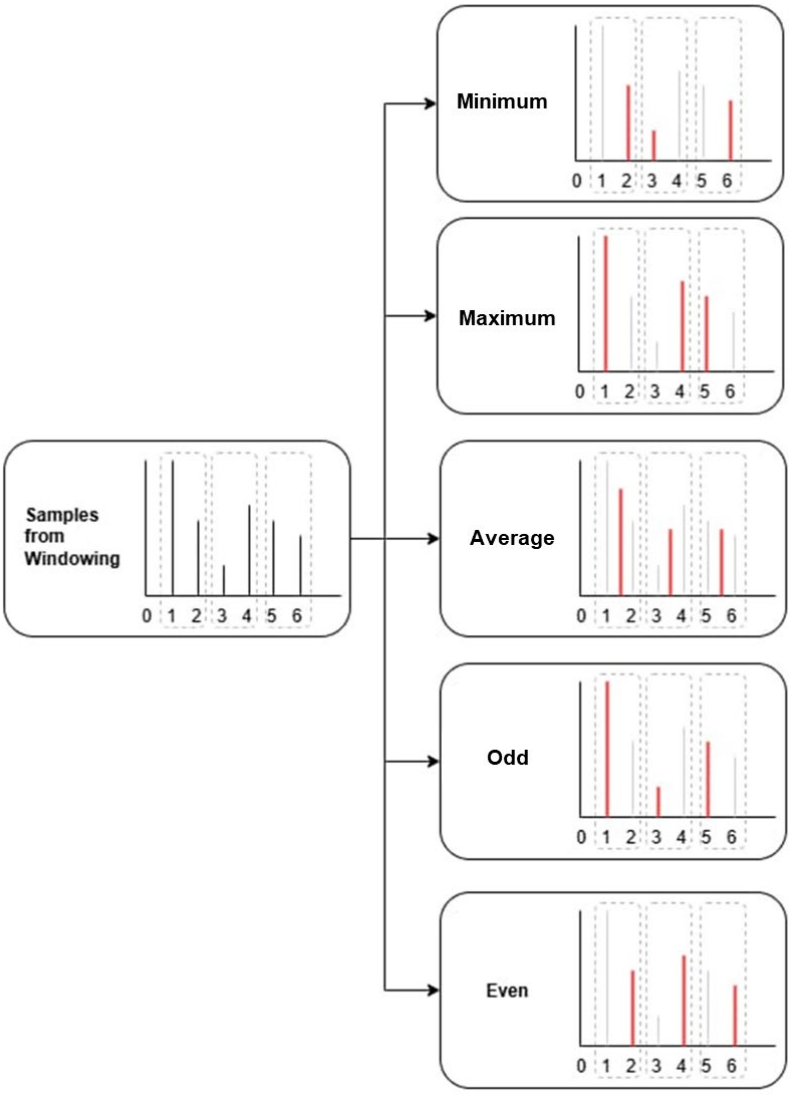


Figure 3.4: Multi-model feature extraction strategies applied to the optimally selected EEG window

Initially, each feature extraction strategy was evaluated independently to assess its ability to represent abnormal EEG characteristics. Although individual feature representations demonstrated useful discriminative capability, each captured different aspects of the EEG signal. Consequently, the proposed framework combines all five complementary feature extraction strategies to

maximise feature diversity and provide richer information for the subsequent ensemble learning stage. The comparative evaluation between single-feature and multi-feature approaches is presented in Chapter 6.

3.4.1 Averaging (Mean Pooling)

This method computes the arithmetic mean of adjacent samples within a sliding window, effectively smoothing high-frequency noise and baseline fluctuations while preserving the overall signal morphology. Averaging is widely used in EEG preprocessing to enhance the signal-to-noise ratio (SNR) by suppressing random artefacts without distorting underlying neural oscillations (Sanei & Chambers, 2013). By reducing stochastic variability, this technique improves the robustness of subsequent machine learning models to intersession and inter-subject variability (Repovs, 2010).

Given an EEG signal segment $X= [X_1, X_2, \ldots, X_N]$, where N=16,000, we apply non-overlapping sliding windows of size k=2. We computed mean of each sample pair as:

$$X_{avg[i]} = (X_{2i-1} + X_{2i})/2 \qquad i=1, 2,\ldots\ldots\ldots\ldots\ldots\ldots, N/2 \tag{3.2}$$

This approach effectively smooths high-frequency noise such as muscle artefacts and amplifier interference while preserving important low-frequency trends like delta and theta oscillations. It functions equivalently to a moving-average finite impulse response (FIR) filter with a boxcar kernel (Subasi, 2019).

3.4.2 Max-Pooling

Max-pooling retains the highest amplitude value within each segment, thereby emphasising transient events such as epileptic spikes, sharp waves, and other pathological discharges. This technique is particularly justified by prior studies in seizure detection, which have shown that max-pooling enhances the visibility of these high-amplitude transient events that are often clinically significant. By preserving abrupt changes in neural activity during down sampling, max pooling ensures that critical features necessary for detecting disorders like epilepsy are effectively maintained ((Roy, Kiral-Kornek & Harrer, 2019; Acharya et al., 2018). If we apply non-overlapping sliding windows of size k=2, we compute the maximum among each pair as:

$$X_{max[i]} = \max(X_{2i-1}, X_{2i}) \qquad i=1, 2, \ldots\ldots\ldots\ldots, N/2 \tag{3.3}$$

Max-pooling emphasises peak amplitudes, which is crucial for detecting epileptiform spikes, while also reducing sensitivity to baseline drifts.

3.4.3 Min-Pooling

Conversely, min-pooling extracts the lowest amplitude value within each temporal window, thereby capturing inhibitory phases, suppression bursts, and troughs in oscillatory EEG activity. This operation is particularly useful for representing deep brain states characterized by reduced cortical activity, such as burst-suppression patterns observed during anaesthesia or coma. Although direct applications of min-pooling in EEG analysis are relatively limited, related pooling and down sampling approaches have been shown to preserve critical signal components associated with inhibitory neural dynamics and low-amplitude events (Tveitstøl et al., 2024).

By complementing max-pooling, min-pooling ensures that both excitatory and inhibitory neural dynamics are retained within the extracted feature set, resulting in a more comprehensive representation of brain activity. When non-overlapping sliding windows of size k=2 are applied, the min-pooling operation is mathematically defined in Equation (4), where the minimum value of each sample pair is computed.

$$X_{min[i]} = \min (X_{2i-1}, X_{2i}) \qquad i=1, 2, \ldots\ldots\ldots\ldots, N/2 \qquad (3.4)$$

3.4.4 Even Decimation

Even decimation subsamples the EEG signal by selecting every even-indexed sample, effectively reducing the temporal resolution by half while maintaining a sparse yet evenly distributed representation of the data. This operation is mathematically defined in Equation (5), where the decimated signal X_{even} is obtained by selecting samples at even indices from the original signal. This approach significantly reduces computational overhead without substantial loss of diagnostic information, particularly in scenarios where high-frequency details are less critical. Consequently, even decimation provides a computationally efficient strategy that is well suited for real-time EEG processing applications requiring faster data handling and analysis (Roy et al., 2019).

$$Xeven_{[i]} = X_{2i} \qquad i=1, 2, \ldots\ldots\ldots\ldots\ldots, N/2 \qquad (3.5)$$

3.4.5 Odd Decimation

Similarly, odd decimation subsamples the EEG signal by retaining every odd-indexed sample, thereby producing an alternative reduced-resolution representation of the original signal. This

operation is formally defined in Equation (6), where the decimated signal X_{odd} is obtained by selecting samples at odd indices. When used in conjunction with even decimation, this strategy helps mitigate the risk of losing temporally localized features that may occur exclusively at either even or odd sampling points. By preserving complementary temporal information, the approach yields two structurally similar yet distinct representations of the same EEG signal, which can enhance feature retention, model generalization, and overall robustness (Wang, Yan & Oates, 2017).

$$Xodd_{[i]} = X_{2i-1} \qquad i=1, 2,\ldots\ldots\ldots\ldots\ldots, N/2 \qquad (3.6)$$

3.5 Evaluation of ML Models

The initial experiments were conducted on a selected subset of patients from the TUH dataset to establish baseline performance. Three models were evaluated: a Support Vector Machine (SVM), representing a traditional machine-learning approach; a Feed-Forward Neural Network (FFNN), serving as a simple neural baseline; and an LSTM network, designed to capture the temporal dependencies inherent in EEG signals. The results showed that the LSTM model consistently outperformed both the SVM and FFNN, achieving higher accuracy and sensitivity. This outcome underscored the importance of temporal sequence modelling in EEG analysis and motivated the adoption of LSTM-based deep learning for subsequent experiments. The comparative evaluation supporting this architectural choice is presented in Chapter 4.

Building on these findings, a more advanced CNN–LSTM hybrid architecture was developed to combine the spatial feature-extraction capability of convolutional layers with the temporal learning strength of LSTM units.

3.6 Deep CNN–LSTM Model Development

Following the selection of the LSTM as the baseline deep learning model, several network architectures were investigated to further improve classification performance. These included the standard LSTM, a Dropout-LSTM architecture designed to reduce overfitting, and a hybrid CNN–LSTM model combining convolutional feature extraction with sequential learning. The CNN–LSTM architecture was ultimately selected because it integrates the complementary strengths of

both convolutional and recurrent networks. The comparative evaluation supporting this architectural choice is presented in Chapter 4.

3.6.1 Motivation for a Hybrid CNN–LSTM Architecture

EEG signals contain both spatial patterns (across electrodes) and temporal dynamics (over time). Convolutional Neural Networks (CNNs) are highly effective for learning spatially localised features, while Long Short-Term Memory (LSTM) networks excel at modelling sequential dependencies and long-range temporal patterns. To leverage the strengths of both architectures, this study proposes a deep CNN–LSTM hybrid model. The CNN layers act as feature extractors, transforming raw EEG windows into compact spatial representations, which are then passed to stacked LSTM layers to capture temporal relationships crucial for distinguishing between normal and abnormal brain activity. Detailed description of the proposed model along with the relevant hyper parameters setting are given in chapter 4.

3.6.2 Architecture Overview

The proposed CNN-LSTM architecture is shown in Figure 3.5. The proposed network is developed to capture both local temporal features and long-range dependencies in EEG signals. The model begins with 1-D convolutional layers, each followed by non-linear activation and average pooling, which extract short-term temporal features across the 19 EEG channels while progressively reducing noise and dimensionality.

The extracted feature maps are then passed to stacked LSTM layers. These recurrent layers are designed to capture sequential dependencies and temporal dynamics that span beyond the receptive field of the convolutional filters, ensuring sensitivity to both transient and sustained abnormalities. The temporal outputs are aggregated using a global max pooling layer, which provides a compact, translation-invariant representation of the most salient activations across time.

The pooled representation is processed by fully connected layers, each with Rectified Linear Unit (ReLU) activations, which form a multi-layer perceptron (MLP) head that projects the features into a discriminative space. The final classification stage consists of a fully connected layer, followed by a SoftMax output that produces probability distributions over the two classes (normal vs. abnormal EEG).

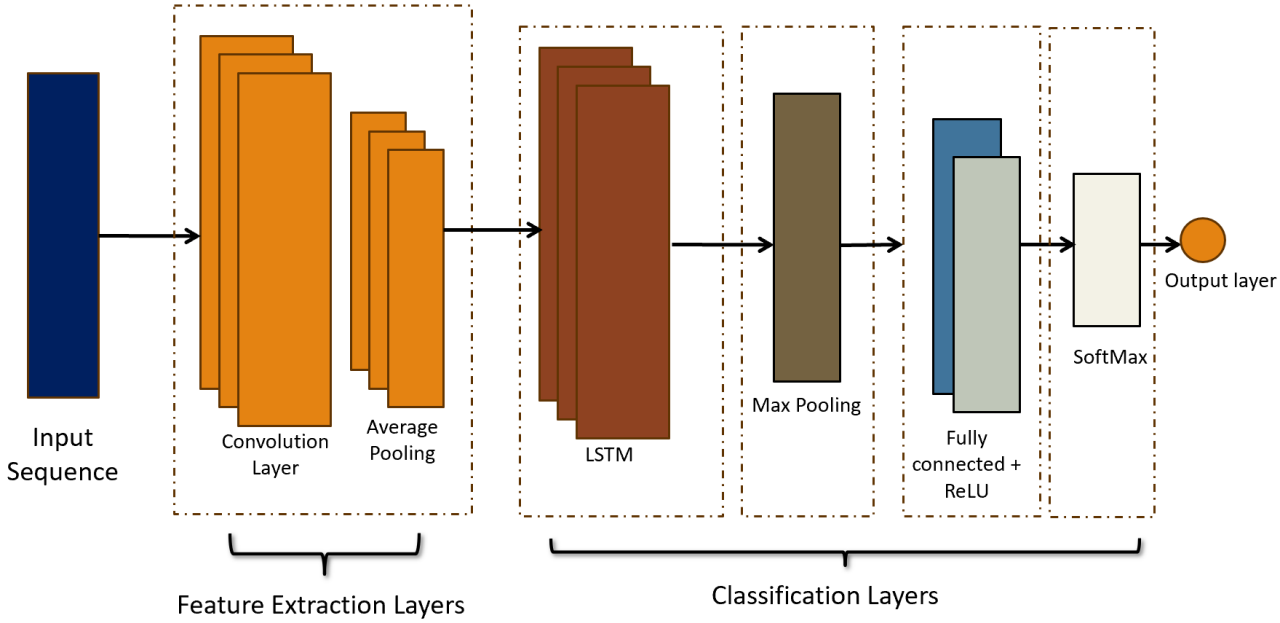


Figure 3.5: Overview of the CNN–LSTM Architecture for EEG Classification

3.6.3Advantages of the Proposed Architecture

The hybrid CNN–LSTM model offers several advantages over standalone CNNs or LSTMs. CNN layers enhance the robustness of learned features by reducing noise sensitivity, while LSTM layers capture long-term temporal dependencies often critical in abnormal EEG patterns such as seizures or transient discharges. By combining these two architectures, the proposed model achieves improved sensitivity to abnormal events while maintaining high specificity in distinguishing normal brain activity.

3.7 Ensemble Decision-Making via Random Forest-Based stacking Mechanism

3.7.1 Motivation for Ensemble Learning

Although the CNN–LSTM model described in Section 3.5 achieved strong classification performance, EEG signals remain highly variable and complex, often leading to trade-offs between sensitivity and specificity across different feature extraction strategies. For example, models trained on maximum-pooled features may emphasise transient spikes but overlook inhibitory patterns, whereas models based on averaging may generalise well but miss fine-grained oscillatory details. To address these limitations, an ensemble decision-making strategy was developed. By combining the predictions of multiple models trained on different feature sets, the ensemble aims

to leverage their complementary strengths, thereby achieving more robust and clinically reliable classification.

3.7.2 Random Forest as a Meta-Learner

The ensemble employs a Random Forest (RF) classifier as the meta-learning mechanism. The input to the RF is formed from the predicted class probabilities of the five CNN–LSTM models trained on the distinct feature sets: Min, Max, Avg, Odd, and Even sampling. Each CNN–LSTM produces a probability distribution over the two classes (normal vs. abnormal), resulting in a five-dimensional probability vector for each sample. These vectors serve as the input features to the RF, which learns non-linear decision boundaries across the combined outputs.

Several meta-learning algorithms were investigated for integrating the outputs of the five CNN–LSTM models, including Support Vector Machine (SVM), Linear Regression (LR), and Random Forest (RF). Based on the comparative evaluation presented in Chapter 6, Random Forest was selected as the final meta-classifier because it provided the most effective combination of classification accuracy and sensitivity while modelling complex non-linear relationships among the outputs of the base classifiers. RF was selected as the meta-classifier for several reasons:

- Non-linear aggregation: Unlike simple majority voting or linear meta-learners, RF can model complex relationships between the outputs of individual models.
- Robustness to overfitting: By averaging across multiple decision trees, RF maintains stability and generalisation even with correlated input features.
- Flexibility and interpretability: RF provides feature importance measures and can handle heterogeneous input distributions, making it suitable for combining the outputs of multiple base learners.

3.7.3 Workflow of the stacking Mechanism

The ensemble decision-making workflow is designed to proceed as follows:

1. Each EEG input window is processed by the five CNN–LSTM models trained on Min, Max, Avg, Odd, and Even feature representations.
2. Each model outputs a probability distribution over the two classes.

3. The five probability vectors are concatenated to form the input to the Random Forest (RF) meta-classifier.
4. The RF classifier combines these inputs through an ensemble of decision trees to produce the final class prediction.

This framework ensures that the final decision reflects both excitatory and inhibitory dynamics, global trends, and complementary temporal representations of the EEG signals.

The workflow of the proposed ensemble decision-making strategy is illustrated in Figure 3.6. The figure illustrates how the raw EEG signal is first segmented into an optimal window, processed by five CNN–LSTM models based on distinct feature extraction strategies, and then aggregated through a Random Forest meta-classifier to yield the final classification outcome.

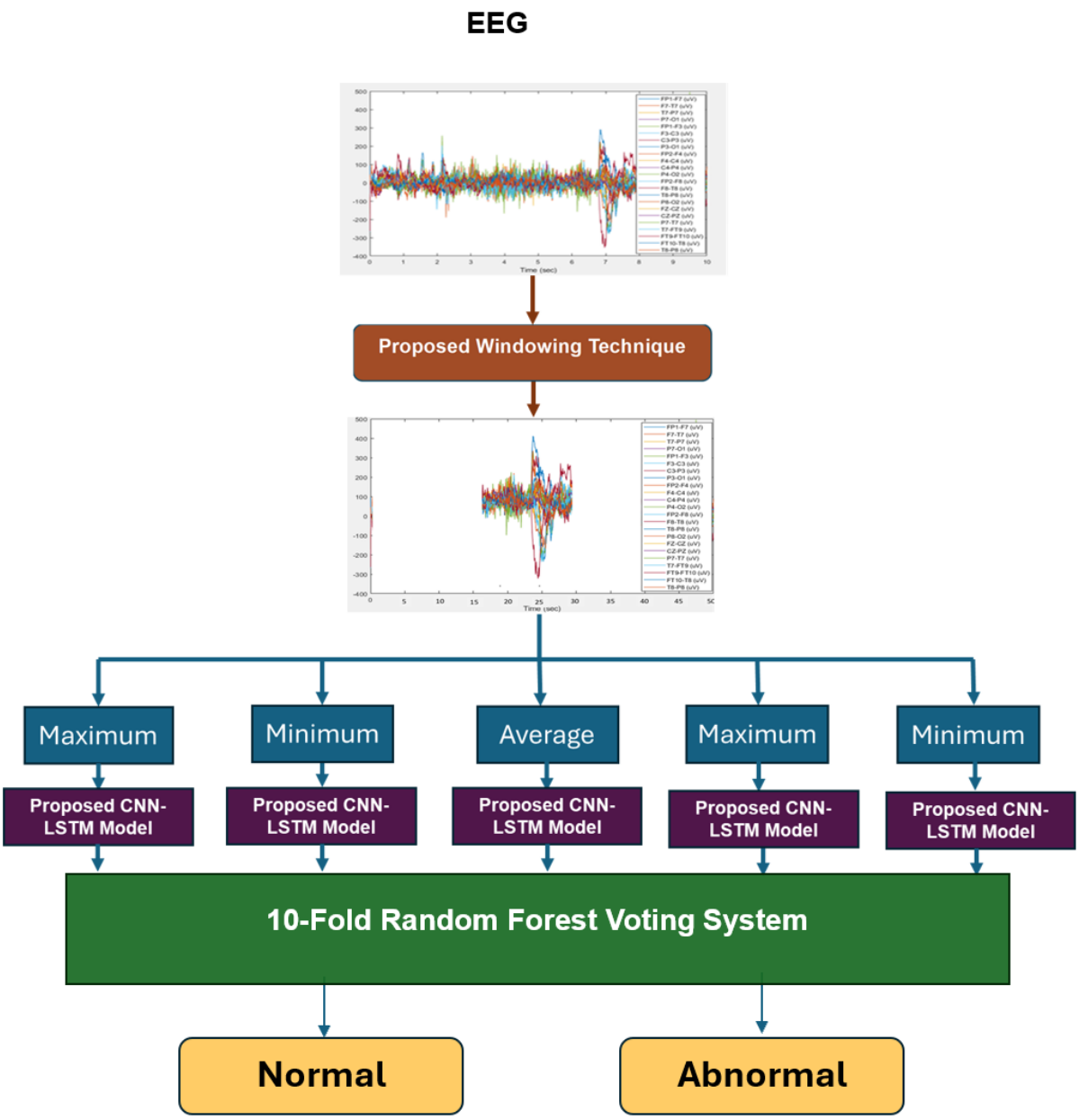


Figure 3.6: Random Forest–based ensemble framework integrating five CNN–LSTM models with different feature extraction strategies for EEG classification.

3.8 Summary

This chapter presented the full methodology developed for automated EEG-based brain disorder detection. The process began with dataset preparation, where raw EEG signals from the TUH Abnormal EEG Corpus were described, and preprocessing steps such as channel selection and dataset-provided filtering were clarified. A wavelet-based windowing technique was then

introduced to adaptively select the most informative temporal segments, ensuring that distinctive pathological features were retained.

Following window selection, five distinct feature extraction strategies Minimum, Maximum, Average, Odd-sample, and Even-sample pooling were applied to capture complementary perspectives of EEG dynamics. These features were processed by a custom-designed 18-layer CNN–LSTM architecture, which combined convolutional layers for local temporal feature extraction with stacked LSTMs for sequential modelling, ultimately yielding robust classification outputs. To address performance variability across feature-specific models, an ensemble decision-making mechanism was implemented using a Random Forest meta-classifier. This approach integrated the probability outputs of the five CNN–LSTM models, leveraging their complementary strengths to improve sensitivity, specificity, and overall diagnostic reliability.

In brief, the methodology combines adaptive windowing, multi-model feature extraction, deep CNN–LSTM classification, and ensemble-based decision fusion into a unified pipeline. The next chapter presents a comprehensive experimental evaluation in which traditional machine learning models and deep learning baselines are systematically compared on both small-scale and full-scale subsets of the TUH Abnormal EEG Corpus. These experiments provide the necessary evidence to assess the contribution of each methodological component and establish the foundation for the final CNN–LSTM and ensemble models proposed in this research.

Chapter 4
Experimental Evaluation of Baseline and Hybrid CNN-LSTM Architectures for EEG-Based Brain Disorder Classification

4.1 Introduction

This chapter presents the experimental evaluation of machine learning and deep learning models for EEG-based brain disorder detection. The investigation was conducted progressively, beginning with exploratory experiments on a limited subset of patients and subsequently extending to the full TUH Abnormal EEG Corpus. This systematic approach enabled a clear understanding of the strengths and limitations of different modelling strategies and ultimately guided the development of the proposed CNN–LSTM architecture.

The initial experiments were performed on a subset of 82 patients from the TUH dataset to establish baseline performance. Three models were evaluated: a Support Vector Machine (SVM), representing a traditional machine learning approach; a Feed-Forward Neural Network (FFNN), serving as a simple neural baseline; and a Long Short-Term Memory (LSTM) network, designed to capture temporal dependencies in EEG signals. The evaluation was later extended to the larger TUH dataset, where only SVM and LSTM were retained for comparison as representative baselines for classical machine learning and deep learning, respectively. This progression further shaped the design of the final CNN–LSTM architecture.

4.1.1 Purpose of Comparative Experiments

The comparative experiments were designed to benchmark the performance of traditional machine learning models (SVM, FFNN) against deep learning approaches (LSTM, CNN–LSTM) and to determine which category is most suitable for EEG classification. Results from the initial 82-patient subset demonstrated the superiority of LSTM over SVM and FFNN, thereby directing subsequent analyses toward sequential deep learning models. Experiments on the full TUH dataset confirmed these findings, while an electrode-reduction study provided additional insights into the importance of spatial resolution in EEG analysis.

4.1.2 Experimental Setup Overview

All experiments were conducted using data from the Temple University Hospital (TUH) Abnormal EEG Corpus, as described in Chapter 3. Two experimental stages were carried out.

- Stage 1: An exploratory study on a subset of 82 patients, comparing SVM, FFNN, and LSTM models to determine which category of algorithms is best suited for learning and generalizing from EEG data.
- Stage 2: A large-scale evaluation on the full TUH dataset (2,674 patients). Based on stage 1 findings, only SVM and LSTM were retained as representative baselines for classical ML and DL, respectively.

4.2 Stage 1: Baseline Performance Analysis on a Limited Patient Subset

Classical machine learning models were first implemented to establish baseline performance for EEG-based brain disorder detection, providing a reference point for assessing the improvements achieved by more advanced deep learning architectures in later experiments. This preliminary evaluation was conducted using a balanced subset of 82 patients from the TUH Abnormal EEG Corpus to identify the most promising modelling strategy. Three representative models were selected: a Support Vector Machine (SVM) as a traditional machine learning baseline, a Feed-Forward Neural Network (FFNN) as a simple neural baseline, and a Long Short-Term Memory (LSTM) network to capture the temporal dynamics inherent in EEG signals. The objective of this stage was to benchmark these models under consistent experimental conditions and determine which approach is most suitable for EEG-based brain disorder classification. All models were implemented and trained in MATLAB version 9.12.0.1884302 (R2022a) on an Intel® Core™ i7-8700 CPU @ 3.20 GHz system equipped with 128 GB of RAM.

4.2.1 SVM Baseline Configuration

For the SVM, classification was performed using a one-vs-all configuration for binary discrimination. A Gaussian (RBF) kernel with a width parameter of 50 was used to capture nonlinear class boundaries. A high regularization constant (C = 1000) penalized misclassifications, while an additional regularization term ($\lambda = 1\times10^{-7}$) controlled model complexity. Training was conducted with silent output (verbose = 0), and evaluation was performed using a 10-fold procedure with independent training and testing splits.

4.2.2 FFNN Baseline Configuration

The FFNN consisted of a single hidden layer with sigmoid activation and a SoftMax output layer for binary classification. The architecture illustrated in Figure 4.1, where w denotes weights and b denotes biases was optimised by varying the number of hidden nodes from 1 to 19. Training was performed using the scaled conjugate gradient backpropagation algorithm with automatic early stopping triggered by increases in validation cross-entropy loss, ensuring improved generalisation. A custom MATLAB implementation was developed to train and assess the FFNN for EEG-based abnormality detection.

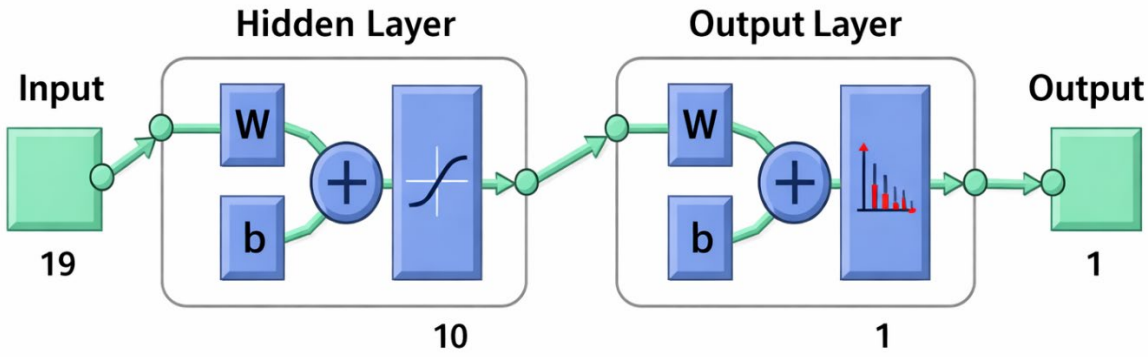


Figure 4.1: Two-layer feed forward neural network.

4.2.3 LSTM Baseline Configuration

The LSTM model, which processes EEG samples as temporal sequences, was trained using the Adam optimiser with a learning rate of 1×10^{-4}, L2 regularisation of 0.001, and a gradient threshold of 1 to prevent gradient explosion. The network was trained for up to 500 epochs with a mini-batch size of 150, using the longest sequence in each batch to preserve temporal structure.

Model performance was evaluated using training and testing accuracy to assess both learning behaviour and generalisation capability. This consistent experimental setup enabled a fair comparison between traditional and deep learning approaches, providing a solid foundation for identifying the most suitable architecture for EEG-based brain disorder detection.

4.2.4 Comparative Performance Analysis

The results are summarised in Table 4.1. The SVM achieved perfect training accuracy (100%) but a low testing accuracy (50%), suggesting overfitting , which occurs when a model learns the training data too closely and fails to generalise to unseen data, resulting in poor generalisation. The FFNN exhibited similar testing accuracy (50%) with a lower training accuracy (79.3%), indicating difficulty in learning complex signal patterns. In contrast, the LSTM achieved a balanced performance, with 91% training accuracy and 65% testing accuracy, demonstrating improved generalisation and the ability to capture temporal dependencies more effectively than the other models.

Table 4.1: Comparison of Deep Learning Model Performance.

Method	***Training Accuracy***	***Testing Accuracy***	***Conclusion***
SVM	100%	50%	Overfitting observed; poor generalization
FFNN	79.3%	50%	Unstable convergence; limited learning capacity
LSTM	91%	65%	Balanced performance; smooth training convergence

Building on these preliminary findings, the next phase of the evaluation expanded the analysis to the full TUH dataset to validate the baseline performance of traditional machine-learning models at a larger scale.

4.3 Stage 2: Large-Scale Evaluation on the Full TUH Abnormal EEG Corpus

Following the baseline evaluation using classical machine learning models, this section presents the large-scale experimental assessment of deep learning architectures for EEG-based brain disorder detection. To validate the findings from the initial exploratory study and to assess the

scalability of the baseline models under realistic clinical conditions, this second experimental stage was conducted using the full TUH Abnormal EEG Corpus, which is the largest publicly available clinical EEG dataset. The selected subset comprises 2,674 patient recordings, representing a heterogeneous population with both normal EEG patterns and abnormal EEG events, including epileptiform discharges, generalized slowing, and other pathological signatures.

The dataset was partitioned into 2,400 recordings for training and 274 recordings for testing, ensuring a balanced representation of normal and abnormal cases, as summarized in Table 4.2. This balanced composition is essential for evaluating the ability of deep learning models to discriminate between healthy and disordered brain activity without bias toward the dominant class. The objective of this experimental stage is to investigate how effectively different deep learning architectures namely a standalone LSTM, an LSTM with dropout regularization (50%), and the proposed CNN–LSTM hybrid model can learn discriminative spatial–temporal representations directly from raw EEG signals. All models were trained and tested using identical EEG input sequences to ensure a fair comparison, with hyperparameters kept consistent across experiments.

Table 4.2: Distribution of Normal and Abnormal Cases in Training and Testing Datasets by Gender

Training		**Testing**	
Normal	**Abnormal**	**Normal**	**Abnormal**
1150	1150	126	148
49.4% Male	43.9% Male	50% Male	43.2% Male
50.6% Female	56.1% Female	50% Female	56.8% Female

4.3.1 LSTM Architecture

Based on the results observed in the initial 82-patient experiments, the Long Short-Term Memory (LSTM) network was selected as the first deep learning baseline. LSTMs are particularly suited for EEG analysis because they can retain long-range temporal dependencies, making them effective for identifying patterns associated with abnormal neural activity. Figure 4.2 illustrates the architecture of the LSTM based deep learning model for EEG brain disorder detection.

The model begins with a Sequence Input Layer, which receives the pre-processed EEG time-series data. The subsequent LSTM layer learns the temporal dependencies within the EEG signals by maintaining memory of past activations through its internal gating mechanisms. This capability

allows the model to effectively capture dynamic variations in neural activity over time. Following the LSTM layer, a ReLU (Rectified Linear Unit) activation function introduces non-linearity to the network, enhancing its ability to learn complex patterns and reduce the risk of vanishing gradients. The output from this layer is then passed to a Fully Connected Layer, which transforms the extracted temporal features into a condensed representation suitable for classification. A Softmax Layer converts these representations into a probability distribution across the target classes (normal or abnormal EEG). Finally, the Output Layer assigns each EEG sequence to its predicted class based on the highest probability.

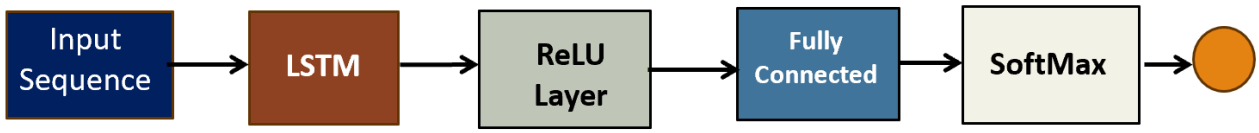


Figure 4.2: The architecture of LSTM based deep learning model.

Training was performed for 30 epochs using the Adam optimiser, a learning rate of 1×10^{-4}, L2 regularisation = 0.001, and a gradient threshold = 1. The LSTM achieved a notable improvement in testing accuracy, however, further improvement was still possible, as the model relied solely on sequential processing without local spatial feature extraction.

4.3.2 LSTM with Dropout Regularisation (50%)

To improve generalisation and reduce overfitting, a dropout layer (50%) was added after the LSTM layers, where dropout is a regularisation technique that randomly disables a fraction of neurons during each training iteration, forcing the network to learn more robust and generalisable representations. Figure 4.3 shows the architecture of the LSTM-dropout based deep learning model. The model begins with a Sequence Input Layer, which receives the pre-processed EEG time-series data. This input is then processed through an LSTM layer, which captures the temporal dependencies of the EEG signals using its gated memory mechanism. Following the LSTM, a ReLU activation function introduces non-linearity to enhance the model's ability to learn complex temporal patterns. A Dropout layer (50%) is then applied to randomly deactivate half of the neurons during training, thereby preventing co-adaptation of features and improving robustness. The output is passed through a Fully Connected Layer, which transforms the learned temporal

representations into a lower-dimensional space suitable for classification. The SoftMax layer converts these representations into probability distributions across the two classes (normal and abnormal EEG), and the Output Layer assigns the final classification based on the maximum probability.

This Dropout-LSTM configuration enhances the generalisation ability of the baseline LSTM model by mitigating overfitting, an essential consideration when training on noisy and non-stationary EEG signals. The modification increased the training accuracy to 68%, while the testing accuracy slightly decreased to 47.87%. This decline in testing performance, despite improved training results, suggests that the applied dropout rate was excessively high, leading to underfitting and reduced capacity for feature representation.

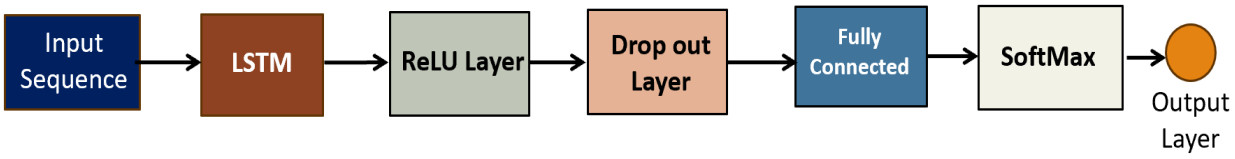


Figure 4.3: The architecture of LSTM-dropout based deep learning model.

4.3.3 Hybrid CNN-LSTM Architecture

To enhance the ability of deep learning models to capture both spatial and temporal characteristics of EEG signals, a hybrid Convolutional Neural Network–Long Short-Term Memory (CNN–LSTM) architecture was developed and evaluated. The CNN layers served as feature extractors, learning localised temporal–spatial patterns within EEG segments through one-dimensional convolutional filters, while the LSTM layers modelled the sequential dependencies across time. Figure 4.4 shows the architecture of the 18-layer CNN-LSTM based deep learning model. The network architecture consisted of two 1D convolutional layers, each with 50 filters of size 5, followed by ReLU activation functions and average pooling layers to reduce noise and prevent overfitting. The extracted feature maps were then passed to two stacked LSTM layers with 100 and 120 hidden units, respectively, to capture the long-term temporal dependencies of EEG dynamics. A global max-pooling layer followed the recurrent layers to condense the temporal information into a fixed-length representation. The final classification block comprised three fully connected layers (50, 60, and 60 neurons) with ReLU activations, and a Softmax output layer for

binary classification (normal vs. abnormal EEG). Training was performed using the Adam optimiser, with a learning rate of 1×10^{-4}, L2 regularisation = 0.001, and gradient threshold = 1 to prevent gradient explosion. The model achieved a training accuracy of 98 % and a testing accuracy of 67.06 %, demonstrating superior generalisation compared to single-model LSTM configurations. This improvement highlights the advantage of combining spatial feature extraction through convolutional layers with temporal sequence modelling via LSTM units.

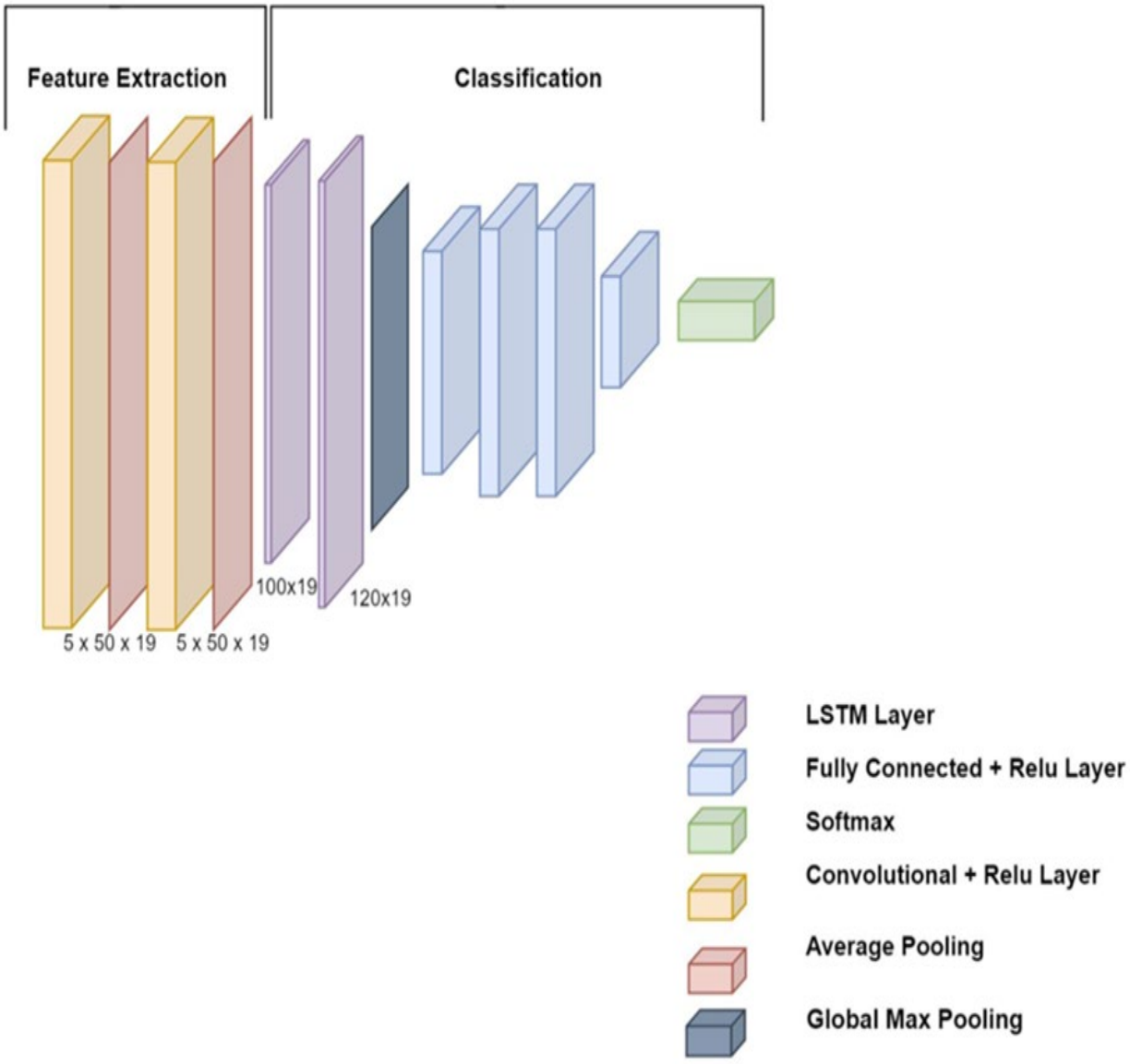


Figure 4.4: 18-layer Model architecture based on CNN and LSTM

- *Hyper Parameter Settings*

The model was trained using the Adam optimiser with an initial learning rate of 10^{-4}. L2 regularisation (0.001), and a mini-batch gradient descent algorithm. Training was conducted for 30 epochs with a gradient threshold of 1 to prevent exploding gradients. The combination of convolutional feature extraction, sequential modelling via LSTMs, and a dense classification head enables the network to balance sensitivity and specificity, making it well-suited for EEG-based brain disorder detection. Hyper parameters, optimisation and regularisation techniques are detailed in Table 4.3.

Table 4.3: Hyper Parameters, Optimisation and Regularization of The Proposed Architecture.

Title 1	Type	Parameter
Architecture	Number of Layers	18
	Number of CNN filters	5
	Size of each CNN filter	50
	Total number of CNN layers	2
	Total number of LSTM Layers	2
	1st LSTM Hidden Layers	100
	2nd LSTM hidden Layers	120
Hyper Parameters	Solver name/Optimiser	ADAM
	Learning rate	10^{-4}
	L2Regularization	0.001
	Mini batch algorithm	Gradient decent
	Input Size	19
	Number of Epochs	30
	Gradient Threshold	1

4.3.4 Comparative Analysis of Deep Learning Models

Table 4.4 presents a comparative evaluation of three deep learning models designed for EEG-based brain disorder classification. The baseline LSTM model, which relies solely on temporal learning, achieved a modest training accuracy of 65% and a testing accuracy of 48.9%, indicating limited ability to generalise beyond the training data. Incorporating dropout regularisation at a rate of 50% resulted in the Dropout-LSTM model, which slightly improved training accuracy to 68% but achieved a lower testing accuracy of 47.7%. This decline suggests that excessive regularisation hindered the model's representational capacity, causing underfitting. In contrast, the hybrid CNN–LSTM architecture demonstrated a substantial improvement, achieving 98% training accuracy and 67.06% testing accuracy. These results highlight the effectiveness of combining convolutional

feature extraction with temporal modelling, enabling the CNN–LSTM model to capture both spatial and sequential EEG patterns and ultimately providing the best generalisation performance among the evaluated approaches.

Table 4.4: Comparison of Deep Learning Architectures Based on Feature Extraction, classification Strategy, and Model Performance.

Model	***Feature Extraction***	***Classification***	***Training Accuracy***	***Testing Accuracy***	***Observation***
LSTM	-	LSTM	65	48.9	Limited temporal learning; poor generalisation
Dropout-LSTM	Drop out 50%	LSTM	68	47.7	Over-regularised; reduced representational capacity
CNN-LSTM	CNN	LSTM	98	67.06	Best generalisation; effective spatial–temporal learning

The progressive improvement across models demonstrates the importance of hierarchical feature learning. While standalone LSTM architectures capture only partial aspects of EEG dynamics, the hybrid CNN–LSTM model effectively integrates both spatial and temporal information, resulting in the strongest overall performance. Figure 4.5 visually illustrates these results, showing a clear performance gap between hybrid CNN–LSTM and the other models. The chart highlights the superior generalisation ability of the hybrid CNN–LSTM, which achieves the highest testing accuracy while avoiding the underfitting seen in the LSTM and the over-regularisation effects observed in the Dropout-LSTM model.

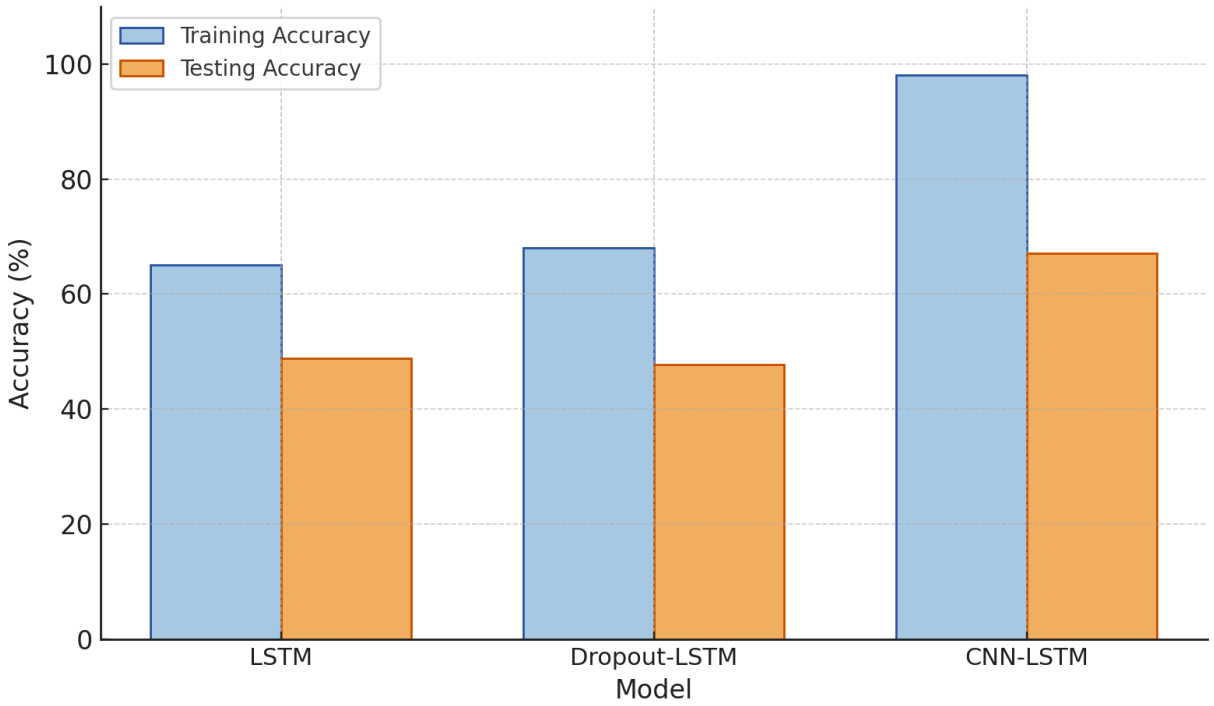


Figure 4.5: Training and testing for LSTM based architectures.

a) Evaluation Metrics

To comprehensively evaluate the performance of the proposed EEG classification framework, several standard quantitative performance metrics are employed. These metrics provide a multidimensional understanding of the model's behaviour across accuracy, sensitivity, specificity, and predictive balance factors that are particularly critical in medical diagnosis tasks such as EEG-based brain disorder detection (Baratloo et al., 2015).

The classification results are assessed using the following metrics derived from the confusion matrix parameters: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

- Sensitivity (Recall or True Positive Rate): Measures the ability of the model to correctly identify abnormal EEG signals. High sensitivity is particularly vital in clinical applications, as it minimises the risk of missed detections (false negatives) (Baratloo et al., 2015).

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \qquad (4.1)$$

- Specificity (True Negative Rate): Reflects the model's ability to correctly classify normal EEG recordings, minimising false alarms (Baratloo et al., 2015).

$$\text{Specificity} = \frac{TN}{TN+FP} \times 100 \quad (4.2)$$

- Precision (Positive Predictive Value): Indicates how many of the detected abnormal cases are actually true positives, helping to assess the reliability of positive predictions (Baratloo et al., 2015; McDonald et al., 2021).

$$\text{Precision} = \frac{TP}{TP+FP} \times 100 \quad (4.3)$$

- F1 Score: Serves as the harmonic mean of precision and sensitivity, yielding a comprehensive performance measure that simultaneously reflects the effects of false-positive and false-negative classifications (Baratloo et al., 2015)

$$F1score = 2 \times \frac{(Precision \times Sensitivity)}{(Precision + Sensitivity)} \quad (4.4)$$

These metrics together enable a robust evaluation of the classification framework's diagnostic performance. In EEG-based clinical applications, sensitivity is often prioritised over other metrics, as failing to detect an abnormal event (such as a seizure or pathological rhythm) can have more severe clinical implications than issuing a false alarm. Therefore, while overall accuracy and specificity are also reported, achieving high sensitivity with acceptable specificity remains the main objective of this work.

To provide a deeper evaluation beyond overall accuracy, several clinical performance metrics were calculated, including precision, recall (sensitivity), specificity, and F1 score. These metrics offer a more comprehensive assessment of each model's ability to correctly classify EEG signals into normal and abnormal categories, which is crucial in clinical contexts where false negatives can lead to missed diagnoses. The three deep-learning architectures LSTM, LSTM with Dropout, and CNN–LSTM were evaluated on the full TUH Abnormal EEG Corpus. Table 4.5 presents a detailed performance comparison of the three evaluated models using key classification metrics, namely

True Positives (TP), False Negatives (FN), True Negatives (TN), False Positives (FP), Precision, Sensitivity, Specificity, and F1-Score. Both the LSTM and Dropout-LSTM models exhibit similar behaviour, with moderate TP counts (62 and 63, respectively) but relatively high FN values (64 and 63), indicating difficulty in correctly identifying abnormal EEG events. Their TN and FP values further show limited discrimination between normal and abnormal cases, resulting in modest Sensitivity (≈0.49–0.50), Specificity (≈0.51–0.54), and F1-Scores below 0.50. In contrast, the CNN–LSTM model demonstrates substantially stronger performance, achieving the highest TP (84), the lowest FN (42), and superior TN (98) while reducing FP (50). These improvements translate into markedly better Sensitivity (0.667), Specificity (0.662), Precision (0.627), and F1-Score (0.646). Overall, the CNN–LSTM achieves the most balanced and clinically meaningful performance across all evaluated metrics.

Table 4.5: Performance Comparison of LSTM, Dropout-LSTM, and CNN–LSTM Models Based on True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) and Derived Scores.

	TP	**FN**	**TN**	**FP**	**Precision**	**Sensitivity**	**Specificity**	**F1 Score**	**Total Normal**	**Total Abnormal**
LSTM	62	64	76	72	0.463	0.492	0.514	0.463	148	126
Dropout-LSTM	63	63	80	68	0.481	0.5	0.540	0.490	148	126
CNN_LSTM	84	42	98	50	0.627	0.667	0.662	0.646	148	126

Comparison of performance metrics for the evaluated deep-learning architectures is presented in the bar chart in Figure 4.6. The figure illustrates the precision, recall (sensitivity), specificity, and F1 score achieved by the LSTM, LSTM with Dropout, and CNN–LSTM models. A clear upward trend is observed across all metrics, with the CNN–LSTM architecture demonstrating the most balanced and superior performance. This confirms that integrating convolutional feature extraction with sequential temporal modelling enhances the model's discriminative ability and clinical reliability for EEG-based brain disorder classification.

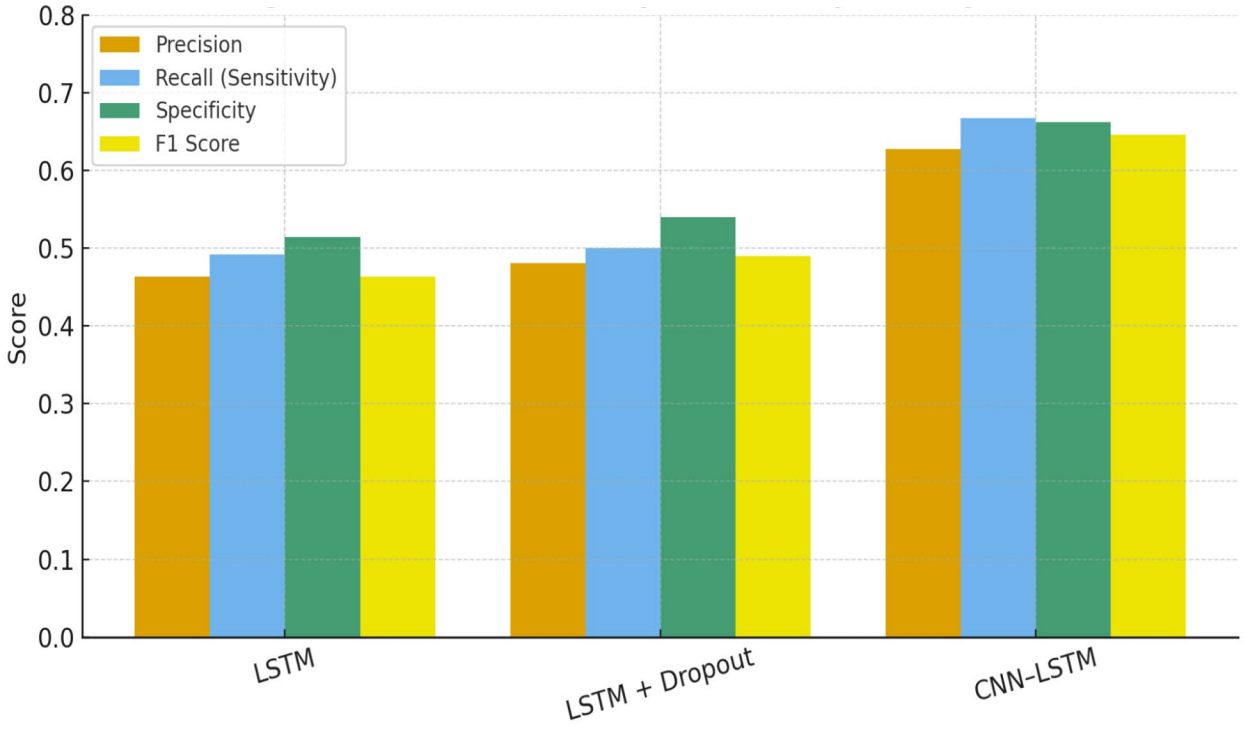


Figure 4.6: Performance Comparison between deep learning models.

The results in Table 4.5 and figure 4.6 highlight the performance evolution across deep-learning architectures where the CNN–LSTM model achieved the best overall performance, with precision = 0.627, recall = 0.667, specificity = 0.662, and F1 score = 0.646, indicating a strong balance between sensitivity and reliability. However, the LSTM and LSTM + Dropout models performed moderately, showing that while temporal modelling captures sequential EEG information, it benefits significantly from convolutional feature extraction. Overall, the CNN–LSTM hybrid clearly demonstrated superior discriminative power, combining spatial and temporal analysis to achieve the best clinical performance balance among the tested models.

A confusion matrix is employed because it provides a comprehensive summary of the model's predictions relative to the true labels, offering insight into both correct classifications and the specific types of errors made. In particular, it highlights the distribution of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Figure 4.7 presents the confusion matrices for several architectures: (a) the standard LSTM model, (b) the 50% Dropout-LSTM model, and (c) the proposed CNN–LSTM architecture, including an 18-layer deep CNN–LSTM variant. The results clearly demonstrate that the windowing-based CNN–LSTM model

delivers the strongest performance compared to the other architectures. Its confusion matrix indicates balanced and reliable detection of both normal and abnormal EEG cases.

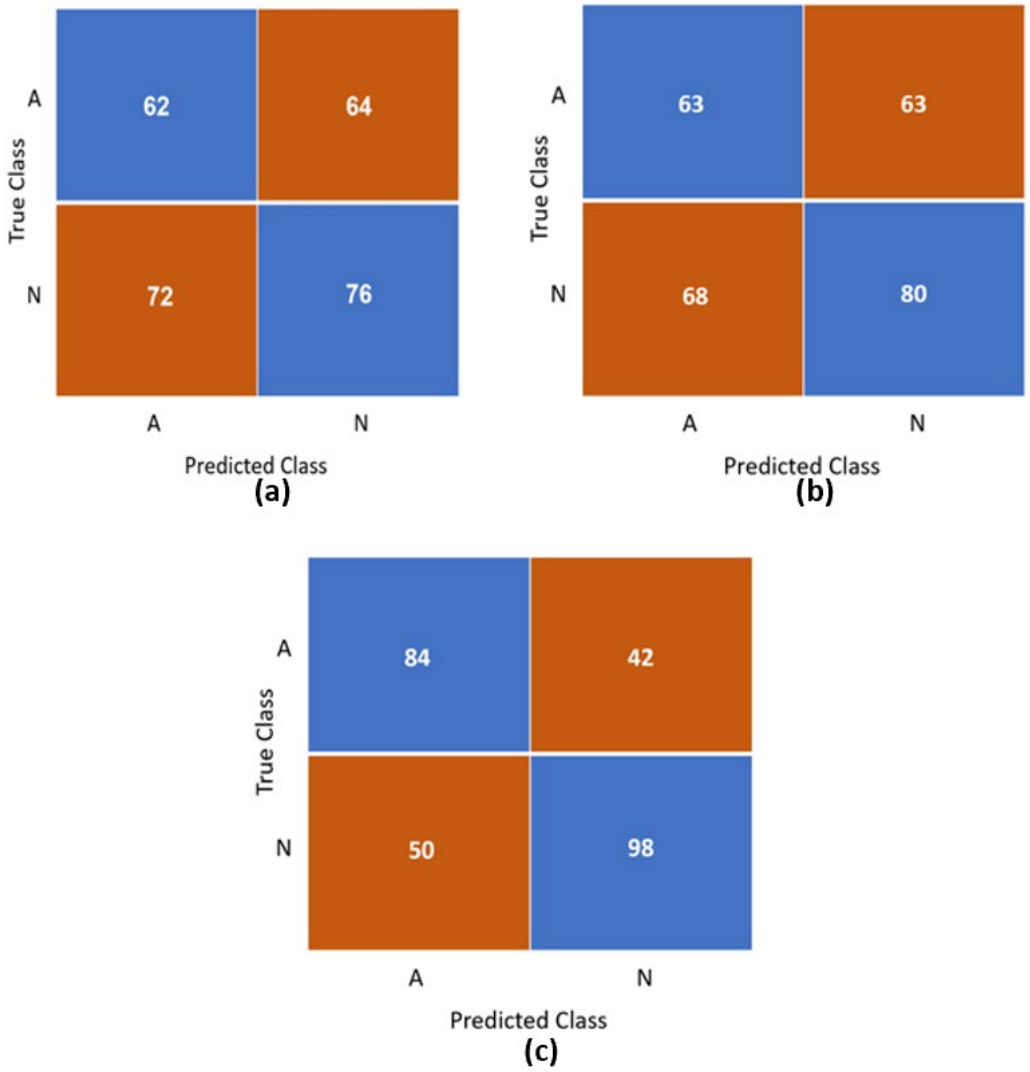


Figure 4.7: Confusion Matrices for the Three Evaluated Models: (a) Dropout-LSTM, (b) LSTM only, and (c) 18-layer CNN–LSTM.

The Receiver Operating Characteristic (ROC) curve is a graphical plot that illustrates the diagnostic ability of a binary classifier by showing the relationship between the true positive rate and the false positive rate across different classification thresholds. The Area Under the Curve (AUC) of the ROC plot is a fundamental metric for evaluating classifier performance. An AUC value of 1 represents a perfect classifier, whereas a value of 0.5 reflects performance equivalent to random guessing. Because it summarises the model's behaviour across all possible decision thresholds, the AUC provides a robust and comprehensive

measure of a classifier's overall effectiveness. Figure 4.8 presents the ROC curves for the three evaluated deep learning architectures, illustrating their ability to discriminate between normal and abnormal EEG recordings. Subfigure (a) shows the ROC curve of the standalone LSTM model, which demonstrates limited discriminatory power and follows closely to the line of no discrimination. Subfigure (b) corresponds to the 50% dropout CNN–LSTM configuration, which shows a slight improvement over the LSTM-only model but still exhibits suboptimal separability between classes. In contrast, subfigure (c) displays the ROC curve of the full 18-layer CNN–LSTM deep learning model, which achieves a markedly superior performance with a curve rising substantially above the diagonal reference line, indicating stronger true-positive rates across a wide range of false-positive thresholds. Overall, Figure 4.8 highlights the clear performance advantage of the deeper CNN–LSTM architecture over the simpler LSTM and dropout-based variants.

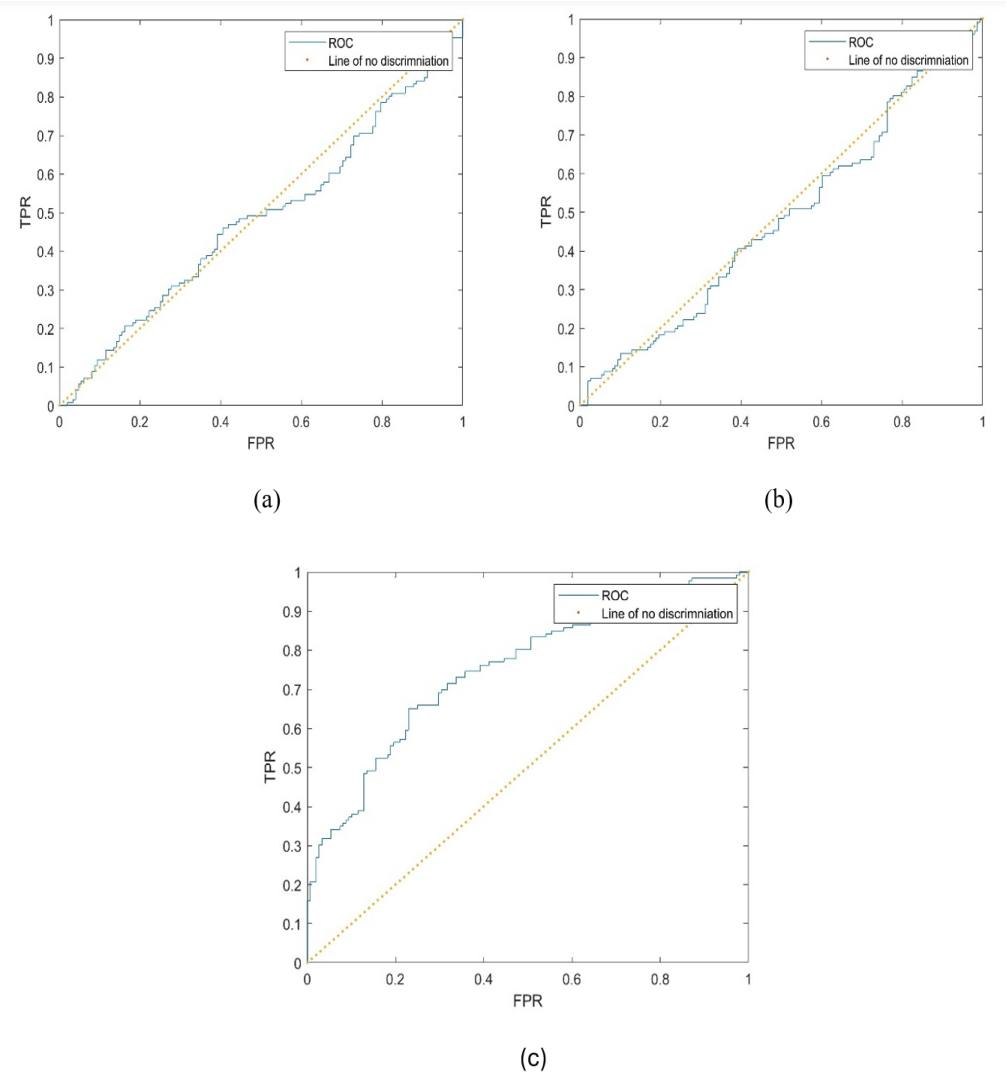


Figure 4.8: ROC plots for different architectures (a) Dropout-LSTM, (b) LSTM only, and (c) 18-layer CNN-LSTM.

4.3.5 Discussion of Deep Learning Results

The comparative analysis of the four deep-learning architectures provides several key insights into the effects of network design on EEG classification performance.

1. Temporal modelling alone (LSTM) offered a meaningful baseline but exhibited limited generalisation, as evidenced by the gap between training and testing accuracies (65 % → 48.9 %). This outcome indicates that while LSTM networks can learn temporal dependencies, they require more discriminative feature inputs to achieve robust classification across subjects.
2. Introducing dropout regularisation (50%) improved training stability and mitigated overfitting, but the high dropout rate slightly degraded testing accuracy (47.87%), suggesting that excessive regularisation reduced the model's representational capacity.
3. The hybrid CNN–LSTM model successfully combined the strengths of both architectures. The CNN layers captured short-term localised features, while the LSTM layers modelled their temporal evolution, resulting in the best overall testing accuracy (67.06 %). This hybrid approach achieved both high learning capacity and improved generalisation, validating its suitability for EEG-based brain-disorder detection.

In summary, the progression from stand-alone LSTM model to the hybrid CNN–LSTM architecture demonstrates that integrating spatial and temporal learning mechanisms substantially enhances classification performance. The CNN–LSTM achieved the most balanced trade-off between feature richness, training stability, and generalisation capability, making it the optimal framework for subsequent ensemble modelling explored in later chapters.

4.4 Combined Model Evaluation

After independently evaluating classical and deep learning models, this section presents a direct comparison across all approaches to assess their relative effectiveness for EEG-based brain-disorder detection. The goal was to determine how model complexity and architecture type affect both training performance and generalisation ability when applied to the full TUH Abnormal EEG Corpus.

4.4.1 Comparative Overview

The experimental workflow was conducted in two stages as follows:

The first stage involved an exploratory analysis on a subset of 82 patients from the TUH Abnormal EEG Corpus, where three models: Support Vector Machine (SVM), Feedforward Neural Network (FFNN), and Long Short-Term Memory (LSTM) were compared. This preliminary phase aimed to determine the most promising classification approach for EEG signal learning and generalisation under limited data conditions. Among the three, the LSTM model achieved the best overall performance, with a training accuracy of 91% and testing accuracy of 65%, demonstrating smoother convergence and better generalisation than SVM and FFNN. Based on these encouraging results, LSTM was selected as the baseline model for subsequent large-scale experiments.

In the second stage, the analysis was extended to the full dataset of 2,300 patients to validate and expand upon the earlier findings. Four deep-learning architectures LSTM, LSTM with Dropout, and CNN–LSTM were evaluated under consistent preprocessing, training, and validation settings. This allowed for a fair comparison of temporal, spatial, and hybrid spatial temporal learning strategies.

4.4.2 Results Interpretation

The results in Table 4.6 demonstrate a clear developmental trajectory in model performance as both data scale and architectural complexity increased.

During the small-scale experiments (82 patients), the SVM and FFNN exhibited overfitting and weak generalisation, both reaching only 50% testing accuracy. The LSTM, however, achieved a notably higher testing accuracy of 65%, confirming that temporal modelling is better suited for EEG data due to its ability to capture sequential dependencies.

When the dataset was expanded to 2,300 patients, the simple LSTM model struggled to maintain performance (testing accuracy dropped to 48.9%), indicating that while temporal modelling helps, it alone is insufficient for larger and more complex EEG data. The LSTM with Dropout showed slightly more stable training but no significant gain in accuracy, suggesting mild underfitting caused by excessive regularisation.

By integrating both spatial and temporal processing, the CNN–LSTM hybrid achieved the best performance overall, reaching 67.06% testing accuracy while maintaining high training stability.

This confirms that EEG signal interpretation benefits architectures capable of learning both localised patterns and long-term dynamics.

Table 4.6: Evaluation of Multiple Classification Architectures Using Feature Extraction Methods and Training/Testing Accuracy.

Model	Number of patients	Category	Feature Extraction	Classification Technique	Training Accuracy (%)	Testing Accuracy (%)	Observation
SVM	82	Machine Learning	-	RBF Kernel	100	50.0	Overfitted; poor generalisation
FFNN	82	Machine Learning	-	Single Hidden Layer	79.3	50.0	Unstable training; limited learning capacity
LSTM	82	Deep Learning	-		91%	65%	Balanced performance; smooth training convergence
LSTM	2300	Deep Learning	–	LSTM	65.0	48.9	Temporal modeling only; weak generalisation
LSTM + Dropout	2300	Deep Learning	–	LSTM (50 % Dropout)	68.0	47.87	Regularised but slightly underfitted
CNN–LSTM	2300	Deep Learning	Convolutional + Sequential	CNN + LSTM	98.0	67.06	Best overall performance; captures spatial–temporal features

4.4.3 Clinical and Computational Interpretation

From a clinical standpoint, the CNN–LSTM model's enhanced testing accuracy suggests a more reliable ability to detect subtle abnormalities within EEG recordings reducing the risk of false negatives in diagnostic settings. The model's performance demonstrated potential for aiding clinical screening processes where rapid and accurate differentiation between normal and abnormal EEG patterns was essential.

4.4.4 Summary of Comparative Evaluation

The combined experimental results reveal a clear progression:

- Machine-learning models (SVM, FFNN) are limited by shallow representation capacity and poor scalability.
- Sequential models (LSTM, LSTM + Dropout) capture temporal dependencies but exhibit weak generalisation on large data.
- Hybrid models (CNN–LSTM) deliver the highest generalisation accuracy and stability by learning both spatial and temporal EEG characteristics.

In summary, the experimental findings demonstrated that the CNN–LSTM architecture provided the strongest overall performance among the evaluated machine-learning and deep-learning models, offering superior accuracy, sensitivity, and generalisation.

4.5 Summary

In this chapter, an extensive experimental evaluation of machine-learning and deep-learning models for EEG-based brain disorder detection was conducted using the TUH Abnormal EEG Corpus. The investigation followed a progressive strategy, beginning with an exploratory study on a subset of 82 patients to establish baseline performance and then extending to large-scale experiments on the full dataset. Initial comparisons between SVM, FFNN, and LSTM models demonstrated that traditional machine-learning approaches (SVM, FFNN) suffered from overfitting and poor generalisation, whereas the LSTM achieved smoother convergence and higher testing accuracy, confirming the importance of temporal modelling for EEG signals.

Building on these findings, the second stage focused on deep-learning architectures at scale, comparing a standalone LSTM, an LSTM with dropout regularisation, and a hybrid 18-layer

CNN–LSTM model. While the LSTM and Dropout-LSTM configurations provided only moderate performance and struggled to generalise on the larger cohort, the CNN–LSTM architecture consistently outperformed them, achieving the highest testing accuracy and the most favourable balance across sensitivity, specificity, precision, and F1-score. Detailed analyses using confusion matrices and ROC curves further highlighted its superior discriminative capability and clinical relevance.

Overall, the findings confirmed that integrating convolutional feature extraction with temporal sequence modeling provides clear advantages over both classical machine-learning approaches and purely recurrent deep-learning models. The CNN–LSTM architecture proved to be the most robust and scalable solution, forming a solid foundation for the ensemble and windowing-based methods explored in the subsequent chapters. Nonetheless, despite its strong performance, the CNN–LSTM framework does not fully overcome challenges associated with the high dimensionality, redundancy, and non-stationary nature of EEG signals. Further improvements in classification accuracy and sensitivity therefore depend on refining the input representations supplied to the model. In response, Chapter 5 introduces advanced feature-selection strategies that target both channel relevance and temporal informativeness, with the aim of enhancing the discriminative strength of the input features and enabling more efficient and reliable EEG-based brain disorder detection.

Chapter 5
Experimental Evaluation of Electrode Correlation and Windowing-Based Feature Selection within the CNN-LSTM Framework

5.1 Introduction

The comparative evaluation presented in Chapter 4 demonstrated that hybrid deep-learning architectures, particularly the CNN–LSTM, achieved superior classification accuracy and generalisation compared to traditional machine-learning and sequential models. Despite these promising results, the CNN–LSTM framework still faces challenges associated with the high dimensionality, redundancy, and non-stationary nature of EEG signals, which can lead to computational inefficiency and potential overfitting.

To further enhance model performance, this chapter investigates feature-selection strategies designed to reduce irrelevant information while preserving the most discriminative features of the EEG data. Two complementary approaches were developed and evaluated:

- Electrode Correlation-Based Feature Selection, which identified and retained the most informative EEG channels based on inter-electrode correlation analysis, aiming to reduce redundancy and noise.
- Windowing-Based Temporal Feature Selection, which segmented the EEG signals into shorter temporal windows to capture local dynamic patterns and isolate the most informative time intervals for classification.

Both techniques were integrated into the CNN–LSTM architecture and tested using the Temple University Hospital (TUH) EEG Corpus under consistent experimental settings.

This chapter details the rationale, methodology, and performance evaluation of each approach, ultimately establishing windowing-based feature selection as the most effective technique for improving the reliability and efficiency of EEG-based brain-disorder classification.

5.2 Rationale for Feature Selection in EEG Analysis

Electroencephalographic (EEG) signals are inherently high-dimensional, noisy, and non-stationary, presenting substantial challenges for automated classification. Each EEG recording typically comprises multiple electrodes capturing overlapping neural activity, leading to a considerable volume of redundant and correlated information. Within deep learning models, particularly convolutional and recurrent architectures such redundancy can increase computational complexity, promote overfitting, and weaken generalisation across subjects and recording sessions. Feature selection therefore plays a crucial role in mitigating these issues by identifying the most informative, discriminative, and stable components of the signal. By removing irrelevant or weakly correlated features, feature selection enhances model interpretability and accelerates the training process without compromising accuracy. In the context of EEG analysis, effective feature selection contributes to dimensionality reduction by limiting the number of input channels or temporal segments, thereby suppressing noise and improving the signal-to-noise ratio for downstream models. This results in more efficient learning, enabling deep learning architectures to focus on the most salient temporal–spatial patterns, and strengthens generalisation by supporting more reliable classification across varying subjects and recording conditions.

In this study, two complementary feature-selection strategies were investigated. The first employs electrode-correlation analysis to eliminate redundant spatial information, whilst the second utilises a windowing-based temporal segmentation approach to emphasise time intervals containing the most meaningful neural activity. Together, these techniques aim to optimise the CNN–LSTM framework by improving accuracy, robustness, and computational efficiency for EEG-based brain disorder detection.

5.3 Electrode Correlation-Based Feature Selection

The electrode correlation-based feature selection technique was developed to reduce redundancy across EEG channels by identifying the most independent and informative electrodes. Owing to the high dimensionality and strong interdependence between EEG electrodes, signals originating from neighbouring or functionally related regions often exhibit high correlation. Such redundancy increases computational load and may obscure subtle discriminative features that are critical for accurate classification.

A full Minimum Redundancy Maximum Relevance (MRMR) (Hu et al., 2025) approach was not adopted in this study, as estimating feature relevance would require repeated model training on large data volumes, resulting in substantial computational overhead. Instead, a similarity-based redundancy minimisation strategy was employed to identify EEG electrodes that contribute maximally complementary spatial information. For this purpose, a normalised maximum cross-correlation similarity measure was computed between all pairs of electrode signals for each patient, capturing the strongest temporal similarity across time lags. The resulting 19×19 similarity matrices were then averaged across the entire dataset of 2300 patients to obtain a global representation of inter-electrode similarity, while self-similarity terms were removed by setting the diagonal elements to zero.

The electrode selection process was initialised by identifying electrodes associated with the global maximum similarity and selecting, from these candidates, the electrode exhibiting the lowest total similarity (row-sum) to all others. Subsequently, electrodes were iteratively added using a greedy strategy that minimised the cumulative similarity between each candidate electrode and the already selected set. This procedure yielded a ranked list of electrodes ordered by their contribution to reducing spatial redundancy and maximising complementary information.

Two subsets of electrodes were derived for experimental evaluation:

Set 1: the five least correlated electrodes.

Set 2: the eight least correlated electrodes.

These subsets were individually used to train and evaluate the CNN–LSTM classification model to assess the impact of spatial channel reduction on overall performance. The complete procedure is formally presented in Algorithm 5.1, which provides a step-by-step mathematical formalisation

of the proposed electrode correlation-based feature selection method, detailing the cross-correlation computation, similarity matrix averaging, and greedy electrode selection process. Figure 5.1 illustrates the electrode selection process based on the proposed correlation-minimisation technique.

Algorithm 5.1: Electrode Correlation-Based Feature Selection

Input / Output

Input: EEG dataset with D0 = 2400 patients, E = 19 electrodes
Output: Reduced electrode subset S of size K ∈ {5, 8}

Notation

`D0`	Total number of patients (D0 = 2400)
`X_e^p`	Signal of electrode e for patient p
`C_ij^p`	Normalised cross-correlation between electrodes i and j for patient p
`C_avg`	Average similarity matrix across all patients (19×19)
`S`	Selected electrode subset
`K`	Number of electrodes to select, K ∈ {5, 8}

Steps

```
1. FOR each patient p = 1 TO D0 DO
    2. FOR each pair of electrodes (i,j) WHERE i ≠ j DO
        3. Compute normalised cross-correlation C_ij^p
              between X_i^p and X_j^p
    4. END FOR
5. END FOR
-----------------------------------------------------------
6. FOR each electrode pair (i,j) WHERE i ≠ j DO
      C_avg(i,j) = (1/D0) × Σ_p C_ij^p
7. END FOR
8. Set diagonal elements of C_avg to zero
-----------------------------------------------------------
9. Identify electrode pair with global maximum similarity
10.From that pair, select electrode with lowest row-sum
       as first element of S
-----------------------------------------------------------
11.WHILE |S| < K DO
    12.FOR each candidate electrode c NOT IN S DO
        13.Compute cumulative similarity of c to S
    14.END FOR
    15.Add electrode with minimum cumulative
          similarity to S
16.END WHILE
-----------------------------------------------------------
17.RETURN S
```

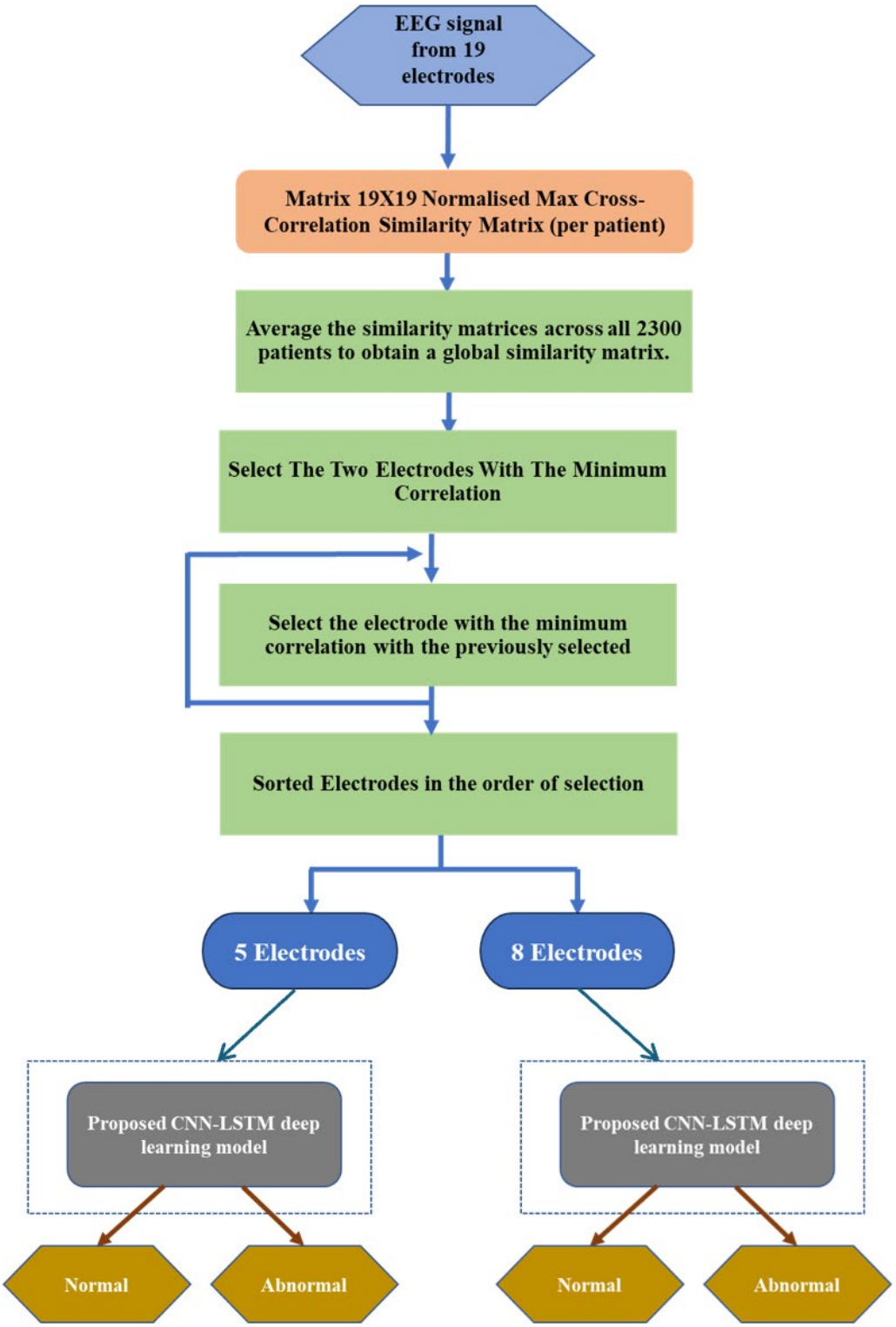
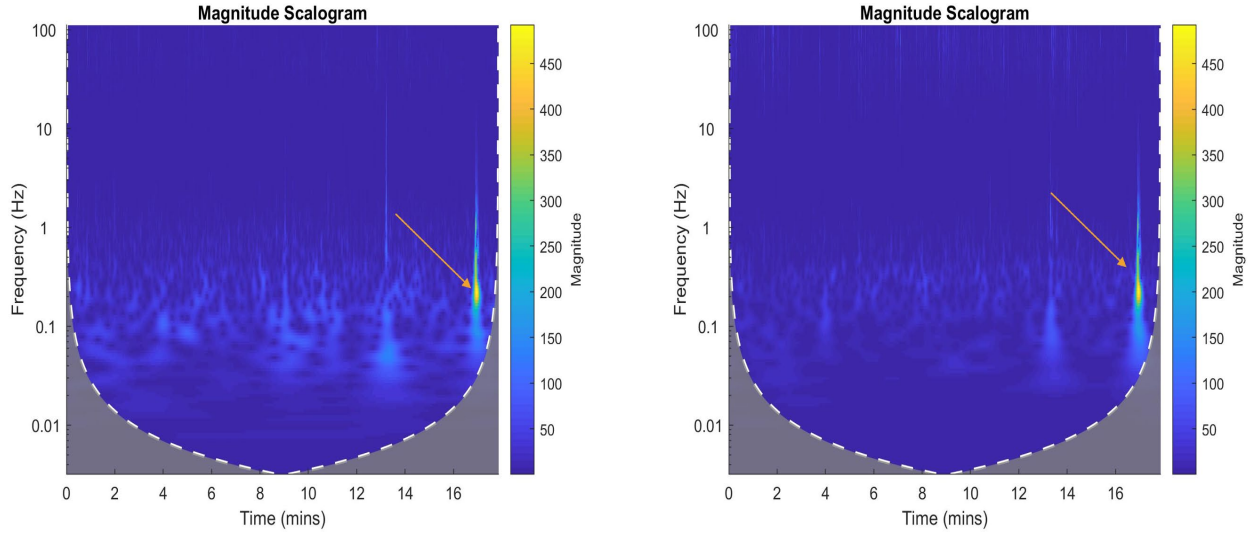


Figure 5.1: Block diagram of the proposed correlation-based EEG electrode selection and CNN–LSTM classification framework using reduced 5- and 8-channel subsets.

5.4 Windowing-Based Temporal Feature Selection

Time–frequency analysis was selected as the primary method for processing EEG signals in this study. The analysis is based on the Continuous Wavelet Transform (CWT) (Lilly, 2017; Gurrala, Yarlagadda & Koppireddi, 2019) using Generalised Morse Wavelets (GMWs) (Olhede & Walden, 2002; Martinez-Ríos et al., 2022), as previously described in Section 3.3.1.

Each electrode's time-domain signal was transformed into the frequency domain using the GMW-based CWT. Figure 5.2 presents example magnitude scalograms for two electrodes from the same subject, with frequency displayed on a logarithmic scale. The cone of influence (Suwansawang & Halliday, 2017), indicating regions affected by edge effects, is shown by the dashed white line, while the grey areas outside the cone represent unreliable regions of the transform. Across all electrodes, it was observed that the most significant wavelet coefficients representing the highest energy or maximum magnitude occurred within similar time intervals. This observation formed the basis of the windowing feature-selection technique, whereby the most informative temporal segment of the EEG signal is automatically identified.



(a) (b)

Figure 5.2: Illustration of magnitude scalograms for two distinct electrodes recorded from the same patient: (a) electrode 4 and (b) electrode 2.

For each subject, the time instants corresponding to the maximum CWT magnitude were extracted from each electrode. The average peak position across electrodes was then computed, excluding outliers more than three standard deviations from the mean. A temporal window centred on this average position was selected, representing the period containing the most distinctive EEG features. Three window lengths namely, narrow (500 samples), medium (2000 samples), and wide (8000 samples) were evaluated to determine the optimal duration for capturing clinically meaningful information. The windowing process was applied to all 19 EEG electrodes per subject, and classification experiments were conducted using the CNN–LSTM model. Figure 5.3 illustrates the windowing procedure and the selected temporal regions for analysis.

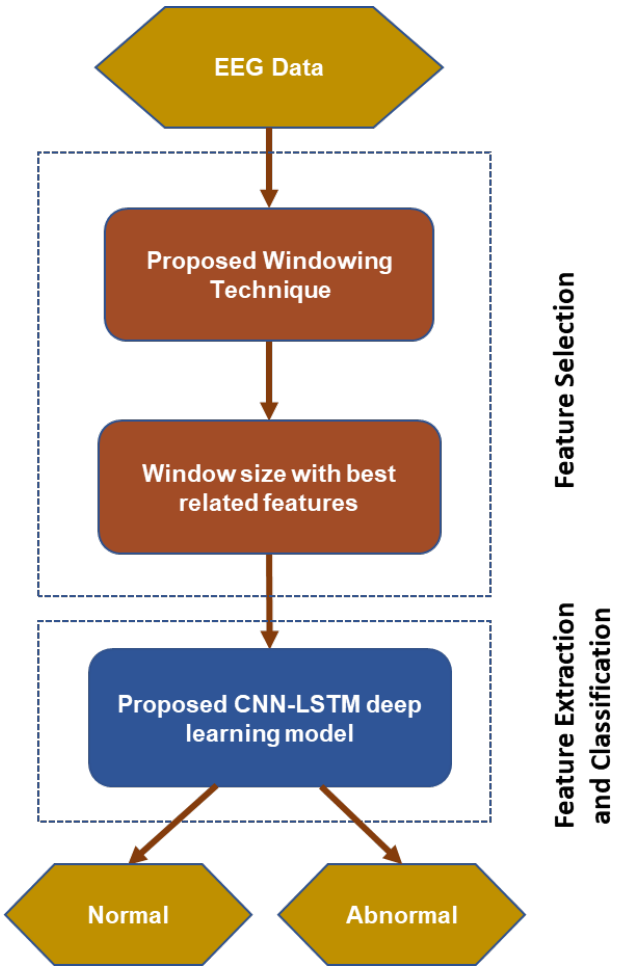


Figure 5.3: Block diagram of the proposed windowing-based feature selection and CNN–LSTM classification framework for distinguishing normal and abnormal EEG signals.

5.5 Results and Comparative Analysis

The comparative evaluation of the two feature-selection strategies correlation minimisation and windowing-based temporal segmentation was conducted using the 18-layer CNN–LSTM deep learning model developed in this study. Both techniques were applied to the TUH EEG Corpus under identical preprocessing and training conditions to ensure a fair and controlled comparison. The results, summarised in Table 5.1, present the model configuration, the applied feature-selection method, the corresponding feature-extraction and classification strategy, and the resulting performance metrics.

Table 5.1: Comparative results of feature-selection techniques using the TUH EEG Corpus

Model	Feature-Selection Technique	Feature Extraction and classification	Number of samples	No. of Electrodes	Training Accuracy (%)	Testing Accuracy (%)
Windowing (500 samples)	Windowing	18 Layer CNN-LSTM Deep learning Model	500	19	100	74.78
Windowing (2000 samples)	Windowing	18 Layer CNN-LSTM Deep learning Model	2000	19	100	78.26
Windowing (8000 samples)	Windowing	18 Layer CNN-LSTM Deep learning Model	8000	19	100	82.68
Correlation (5 electrodes)	Correlation	18 Layer CNN-LSTM Deep learning Model	3005000	5	68	54.38
Correlation (8 electrodes)	Correlation	18 Layer CNN-LSTM Deep learning Model	3005000	8	68	60.00

The results indicate that the windowing-based temporal feature-selection method consistently outperformed the correlation-based electrode reduction approach. In the correlation-based experiments, reducing the number of electrodes to five and eight channels yielded testing

accuracies of 54.38 % and 60.00 %, respectively. Although this reduction decreased computational complexity, it also removed spatially informative features, limiting the model's representational capacity and resulting in lower classification performance. Conversely, the windowing approach substantially improved generalisation. By applying wavelet-based temporal segmentation to identify the most informative time intervals, the CNN–LSTM model was able to focus on EEG segments containing distinctive neural patterns while discarding irrelevant or low-energy intervals. Testing accuracy increased progressively with window size from 74.78 % at 500 samples to 78.26 % at 2000 samples, reaching the peak value of 82.68 % at 8000 samples demonstrating that moderately extended temporal windows capture richer dynamic EEG features without introducing excessive redundancy. All windowing-based configurations achieved 100 % training accuracy, confirming strong model learning capability, whereas the correlation-based approach exhibited lower training performance (68 %), reflecting the loss of key spatial dependencies. Overall, the windowing-based feature selection technique provided an accuracy improvement of approximately 22–28 % compared with correlation minimisation, validating its effectiveness for enhancing both accuracy and generalisation in deep-learning-based EEG classification. The comparative testing accuracy of EEG feature-selection techniques using the 18-layer CNN–LSTM deep-learning model is shown in a bar graph in figure 5.4. The chart compares the performance of correlation-based electrode reduction and wavelet-based temporal windowing across different data sizes. Results show that correlation minimisation, despite reducing the number of electrodes and total samples (3,005,000), yielded lower accuracies (54-60 %), while temporal windowing with 500–8000 sample segments substantially improved generalisation, achieving the highest testing accuracy (82.68 %) at 8000 samples.

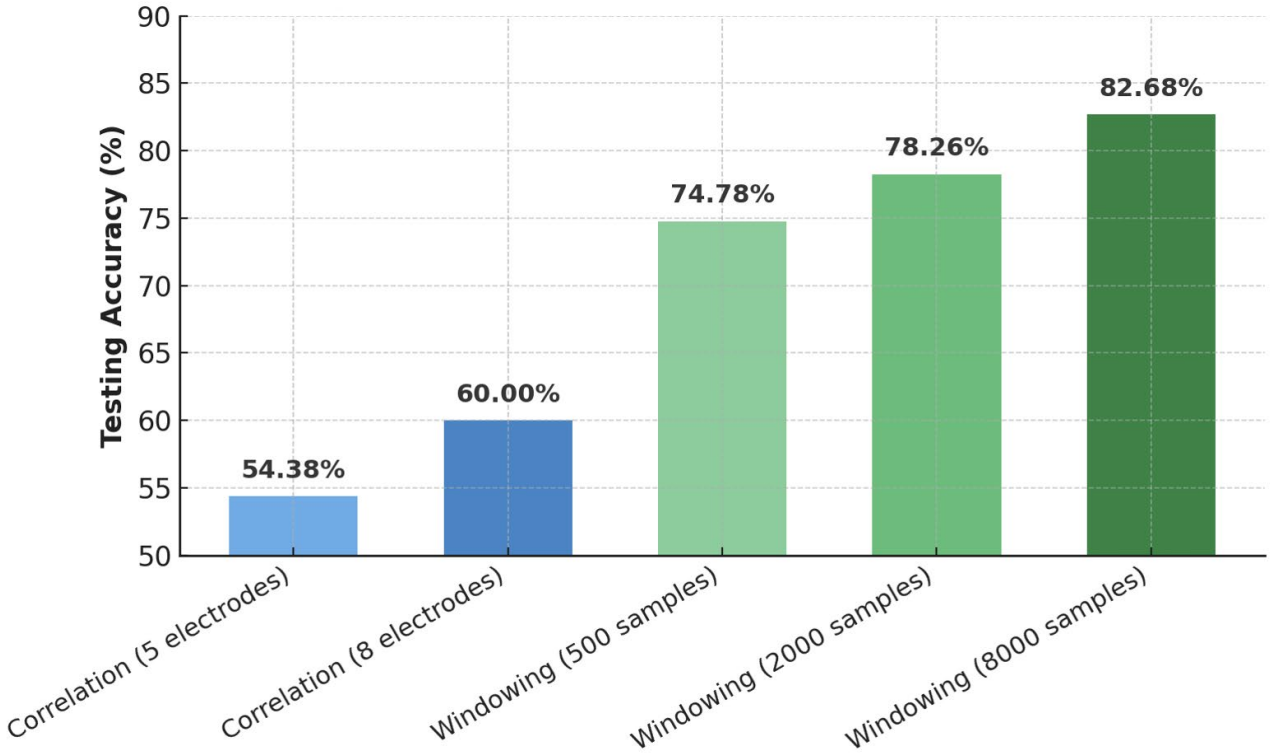


Figure 5.4: Comparative accuracy of feature selection techniques.

5.5.1. Detailed Performance Analysis of Windowing-Based Feature Selection

A more detailed evaluation of the windowing-based temporal feature-selection method was conducted to determine how different window lengths influence the diagnostic capability of the CNN–LSTM model. Performance was assessed using several clinically relevant metrics, including sensitivity (SV), specificity (SP), and overall accuracy. Three window sizes 500, 2000, and 8000 samples were analysed to identify the optimal temporal duration required to capture the most informative EEG activity.

The smallest window (500 samples) produced an average accuracy of 70.61% and a maximum accuracy of 74.78%, with sensitivity and specificity values of 70.77% and 70.58%, respectively. Although this short window captures the most immediate high-energy regions of the EEG, its limited temporal coverage restricts the model's ability to learn broader temporal–spatial patterns, leading to balanced but modest diagnostic performance.

Increasing the window length to 2000 samples resulted in improved performance across all metrics. The average accuracy rose to 73.40%, with a higher maximum accuracy of 78.26%. Sensitivity increased

slightly to 71.54%, while specificity improved more notably to 75.45%, indicating enhanced discrimination of normal EEG segments. This window size provides a more comprehensive representation of the underlying neurological dynamics without introducing substantial noise or irrelevant information.

The widest window (8000 samples) yielded the highest overall performance, achieving an average accuracy of 74.50% and a maximum accuracy of 82.68%. Sensitivity reached 78.5% and specificity increased substantially to 87.10%, reflecting the model's strongest capability to correctly identify both abnormal and normal EEG patterns. The superior performance of this window suggests that longer temporal segments enable the CNN–LSTM architecture to capture more extended abnormality-related patterns and contextual temporal variations that are essential for robust classification.

Overall, these results demonstrate a clear trend: larger window sizes enhance diagnostic performance, with the 8000-sample window providing the most reliable and discriminative features for EEG-based brain disorder detection. This confirms that meaningful temporal patterns in EEG signals often span longer intervals, and that capturing these broader dynamics is critical for improving sensitivity, specificity, and overall classification accuracy. Table 5.2 summarises the classification performance of the CNN–LSTM model across the three evaluated window sizes—500, 2000, and 8000 samples reporting true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), sensitivity, specificity, precision, and F1-score. For completeness, the corresponding fold-wise results obtained using 10-fold cross-validation are provided in Appendix B, where detailed performance variations across individual folds are reported. Collectively, these findings highlight a progressive improvement in diagnostic performance as the temporal window length increases.

Table 5.2: Classification Performance Across 500, 2000, and 8000 Sample Windows in terms of True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Sensitivity, Specificity, Precision, F1-Score for Different Window Sizes.

Window Size	TP	FN	FP	TN	Sensitivity	Specificity	Accuracy	Precision	F1 Score
500 samples	105	43	37	89	70.77%	70.58%	74.78%	0.7394	0.7232
2000 samples	106	42	31	95	71.54%	75.45%	78.26%	0.7737	0.7434
8000 samples	116	32	16	110	78.5%	87.10%	82.68%	0.8787	0.8292

Figure 5.5 presents the combined confusion matrices for the three evaluated window sizes 500, 2000, and 8000 samples illustrating how temporal segment length influences the diagnostic performance of the CNN–LSTM model. The 500-sample window, shown in the first panel, results in noticeably higher rates of misclassification, with a greater number of false negatives and false positives, reflecting limited discriminatory capacity due to the short temporal context. The 2000-sample window demonstrates moderate improvement, particularly in reducing false positives, indicating that a broader temporal view supports better recognition of normal EEG patterns. The third panel, corresponding to the 8000-sample window, achieves the most favourable distribution, with the highest true-positive and true-negative counts and markedly fewer misclassifications. This performance suggests that extended temporal segments capture richer spatial–temporal structure in the EEG, enabling more reliable identification of abnormal neural activity. Overall, the figure clearly demonstrates the progressive improvement in classification accuracy as window length increases, with the 8000-sample window providing the most robust and clinically meaningful results.

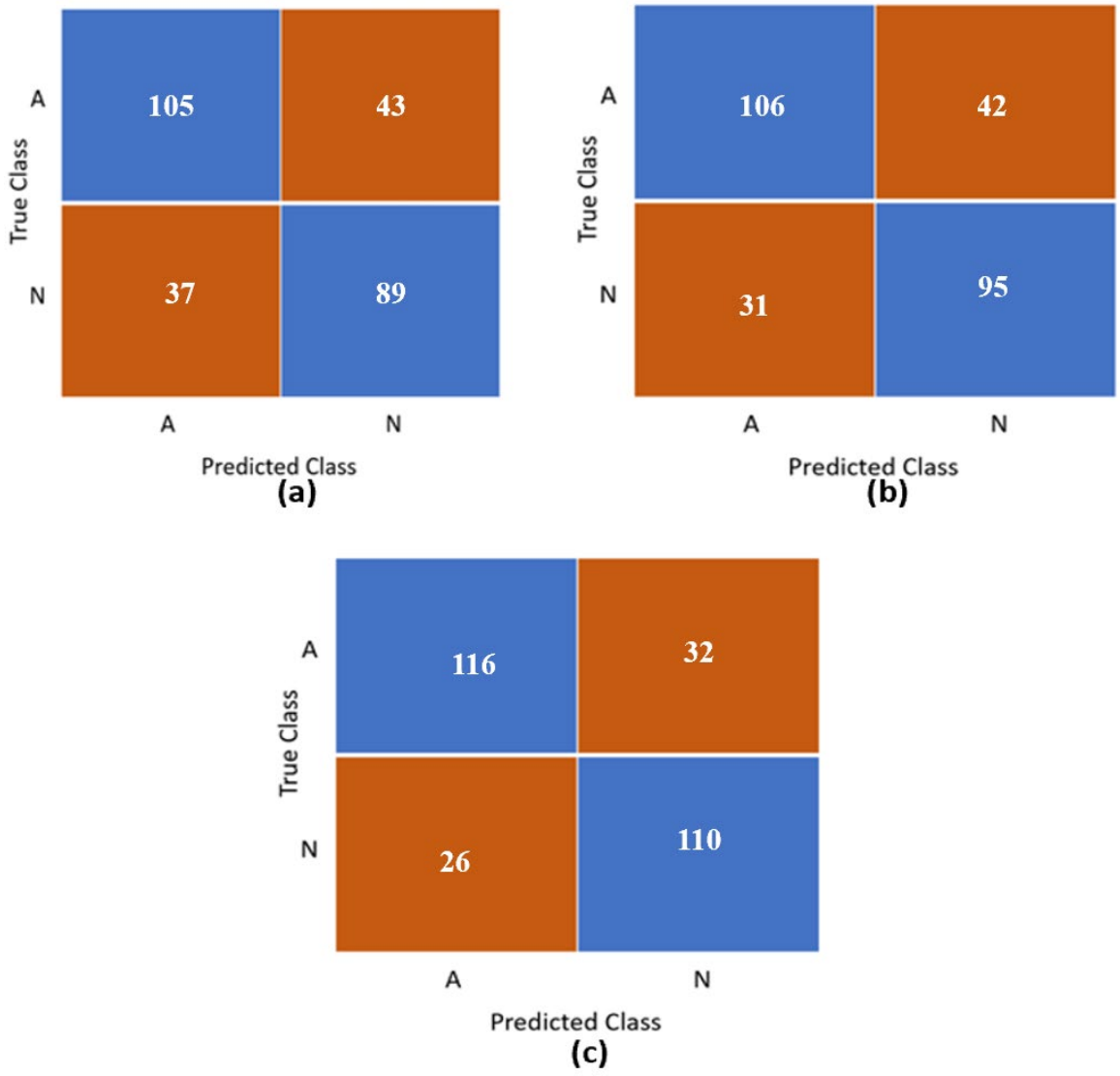


Figure 5.5: Confusion Matrices Comparing the Classification Performance Across Three Window Sizes (500, 2000, and 8000 Samples).

5.6 Discussion

The comparative evaluation of the two feature-selection approaches demonstrates that temporal windowing provides a substantially greater performance gain than electrode correlation-based reduction for EEG-based brain disorder classification. This outcome reinforces the importance of capturing time-localised dynamics in non-stationary EEG signals rather than focusing solely on spatial redundancy among electrodes. The correlation-minimisation method effectively reduced

the number of input channels and, consequently, the computational load. However, the decline in both training and testing accuracies (68% and 54–60%, respectively) indicates that removing highly correlated electrodes also eliminated useful spatial information essential for representing distributed brain activity. EEG signals are inherently correlated across neighbouring regions; therefore, minimising correlation does not necessarily correspond to maximising discriminative power. Although this approach improved efficiency, it compromised accuracy and generalisation, suggesting that spatial redundancy reduction alone is insufficient for reliable EEG classification.

Conversely, the windowing-based temporal feature-selection technique produced the highest testing accuracy, reaching 82.68% with an 8000-sample window. By segmenting EEG recordings into time-localised intervals, this method enabled the CNN–LSTM architecture to focus on transient events containing distinctive frequency–amplitude patterns, while excluding long stationary segments dominated by noise or baseline activity. This is particularly valuable for EEG, where discriminative features such as epileptic spikes, oscillatory bursts, or cognitive responses occur briefly within specific temporal regions. Moreover, windowing aligns naturally with the hierarchical structure of the CNN–LSTM model, wherein convolutional layers capture spatial–spectral patterns within each window, and LSTM layers model sequential dependencies across windows.

This hierarchical integration enhances both temporal resolution and contextual continuity, improving classification robustness without requiring additional preprocessing or electrode pruning. The observed improvement with larger windows (up to 8000 samples) suggests that moderately extended temporal contexts help preserve relevant spectral components while still excluding irrelevant background activity. However, excessively large windows may reintroduce redundancy and increase computational cost; therefore, the selected 8000-sample configuration represents an optimal balance between temporal coverage and feature compactness. In conclusion, the results confirm that temporal segmentation through wavelet-based windowing is a more effective feature-selection strategy than correlation-based spatial reduction for CNN–LSTM EEG classification. It enhances the model's ability to generalise across subjects and recording conditions while maintaining computational efficiency, thereby establishing a robust foundation for the ensemble framework presented in the following chapter.

5.8 Summary

This chapter presented and evaluated two distinct feature-selection strategies aimed at improving the performance and efficiency of EEG-based brain disorder classification using the proposed 18-layer CNN–LSTM deep learning model. The first approach, correlation-based electrode selection, reduced spatial redundancy by identifying the least correlated EEG channels. Although this method effectively lowered data dimensionality, it also removed valuable spatial information, resulting in limited testing accuracies between 54% and 60%. The second approach, windowing-based temporal feature selection using wavelet decomposition, focused on isolating the most informative time intervals within the EEG recordings. This technique substantially enhanced model generalisation and interpretability, achieving testing accuracies of 74.78%, 78.26%, and 82.68% for window sizes of 500, 2000, and 8000 samples, respectively. These findings demonstrate that temporal segmentation captures the dynamic and non-stationary characteristics of EEG signals far more effectively than spatial reduction alone.

Despite the improvement in overall accuracy, closer examination of the diagnostic metrics revealed important limitations. The correlation-based approach was not further evaluated using sensitivity or specificity measures, as its low performance made it unsuitable for detailed clinical interpretation. Consequently, diagnostic analysis including sensitivity, specificity, precision, and F1-score was conducted only for the windowing-based method. Across all window sizes, sensitivity remained moderate, indicating that the model continued to misclassify a considerable proportion of abnormal EEG recordings. This limitation is clinically significant, as high sensitivity is essential to minimise false-negative outcomes and prevent missed pathological events. These findings highlight the need for a more advanced modelling strategy that can specifically enhance sensitivity while preserving or improving overall classification accuracy.

Overall, the windowing-based approach proved markedly superior, offering an average accuracy improvement of more than 25% compared with correlation-based selection. While the 8000-sample window achieved the strongest performance, the sensitivity of the CNN–LSTM model remained insufficient for reliable clinical deployment. Given the critical importance of detecting abnormal EEG patterns with high confidence, further improvement is required to enhance the diagnostic robustness of the system. This limitation motivates the development of a more

sophisticated modelling framework. Accordingly, chapter 6 introduces an ensemble learning approach that integrates multiple CNN–LSTM classifiers to improve sensitivity, reduce model variability, and enhance generalisation across heterogeneous EEG patterns. Building upon the temporal windowing foundation established in this chapter, the ensemble framework aims to deliver more stable and clinically dependable predictions for automated brain disorder detection.

Chapter 6
Experimental Evaluation of Multi-Feature CNN-LSTM Sub-Models and Ensemble Stacking Mechanisms for EEG-Based Brain Disorder Classification

6.1 Introduction

The results from Chapter 5 demonstrated that the windowing-based CNN–LSTM model effectively captures the spatial temporal dynamics of EEG signals, achieving the highest testing accuracy of 82.68 %. However, the model still exhibited limited sensitivity in detecting abnormal EEG patterns, an important drawback for clinical applications, where missed pathological events can have serious consequences.

While aforementioned model represents a significant improvement over traditional and correlation-based methods, single deep-learning models often exhibit performance variability, overfitting tendencies, and sensitivity to initialisation and data imbalance. To address these limitations and further enhance the reliability of automated brain disorder detection, this chapter introduces an ensemble learning framework that integrates multiple CNN–LSTM classifiers. Ensemble learning combines the predictive strengths of several models to produce a final, more accurate and stable decision. By aggregating predictions from independently trained CNN–LSTM networks each optimised on distinct feature subsets or temporal windows, the ensemble framework

aims to reduce individual model bias, improve generalisation, and increase robustness across heterogeneous EEG patterns.

This chapter presents the rationale, design, and evaluation of the proposed ensemble classification framework, built upon the foundation of the windowing-based feature selection strategy introduced previously. The following sections describe the motivation for ensemble learning in EEG analysis, detail the proposed architecture, and compare the ensemble's performance against standalone models using the TUH EEG dataset.

6.2 Rationale for Ensemble Learning in EEG Classification

Although the CNN–LSTM architecture developed in Chapter 5 achieved strong overall accuracy and benefited from improved feature representation through windowing-based preprocessing, the results revealed a persistent imbalance between sensitivity and specificity. In clinical EEG analysis, this imbalance is of particular concern because high accuracy does not necessarily equate to high diagnostic reliability. A model may achieve a favourable overall accuracy while still failing to detect abnormal EEG patterns such as epileptic spikes or transient pathological discharges. Since such missed detections (false negatives) can carry serious clinical implications, enhancing sensitivity is often prioritised over raw accuracy in medical decision-support systems.

EEG data are inherently non-stationary, noisy, and highly subject-dependent, which contributes to significant variability across individuals and recording sessions. Even well-optimised CNN–LSTM architectures may develop biases towards dominant patterns in the training data, resulting in inconsistent generalisation when confronted with different noise levels, artefacts, or feature distributions. Single-model approaches, although effective under controlled conditions, often struggle to maintain performance across heterogeneous EEG characteristics.

Following the identification of the optimal temporal window using the wavelet-based segmentation technique, a multi-feature extraction stage was introduced to generate diverse representations of the EEG signal before the ensemble stacking stage. Because EEG recordings exhibit complex, rapidly changing temporal–spectral behaviour, no single transformation is capable of capturing all clinically relevant characteristics. To address this limitation, five complementary feature-extraction strategies were applied to each windowed segment: maximum pooling, minimum pooling, average pooling, odd-sample decimation, and even-sample decimation. These transformations highlight different dimensions of neural activity—max pooling

emphasises transient high-amplitude events, min pooling captures low-amplitude rhythms, average pooling smooths noise while preserving global temporal structure, and odd/even decimation provides two complementary subsampled views that retain temporally localised detail. The resulting five feature sets are processed independently by parallel CNN–LSTM sub-models, producing a diverse collection of predictive outputs that serve as the inputs to the ensemble stacking mechanisms.

To address the limitations observed in single-model performance particularly the moderate sensitivity and inconsistent generalisation, this study adopts an ensemble learning strategy, integrating multiple models to improve predictive stability and overall diagnostic capability. Ensemble learning leverages the diversity of several classifiers to mitigate the weaknesses inherent in individual models. Instead of relying on a single CNN–LSTM output, the predictions of multiple sub-models are aggregated through meta-classifiers that learn optimal decision boundaries across their collective outputs. This aggregation allows the system to exploit complementary representations of the EEG signal, thereby enhancing both sensitivity and specificity and improving robustness against variability and noise.

In this research, three distinct ensemble stacking approaches were investigated:

1. Support Vector Machine (SVM)-based stacking: Combines the probabilistic outputs of the CNN–LSTM models using a non-linear decision boundary that maximises the margin between normal and abnormal classes.

2. Linear Regression (LR)-based stacking: Applies a linear weighting scheme to combine sub-model outputs, offering simplicity and interpretability while capturing additive relationships between predictions.

3. Random Forest (RF)-based stacking: Employs multiple randomised decision trees to model non-linear dependencies among model predictions, offering robustness to noise, high-dimensional features, and class imbalance.

Each stacking mechanism presents unique advantages: SVM enhances discriminative power through non-linear boundary optimisation, LR provides transparency and computational efficiency, and RF offers resilience to overfitting and high sensitivity in complex signal domains.

By systematically comparing these three ensemble techniques, this research aims to determine the most effective strategy for integrating the diverse outputs of CNN–LSTM models in EEG-based brain disorder detection. The comparative evaluation focuses on identifying which meta-classifier provides the optimal balance between sensitivity, specificity, and overall accuracy, while maintaining interpretability and robustness to noise. This comparative approach ensures that the final framework is selected not only for its computational performance but also for its clinical reliability and diagnostic value. The next section introduces the proposed ensemble architecture, outlining how the three stacking-based systems SVM, Linear Regression, and Random Forest were implemented and integrated into the overall EEG classification pipeline.

6.3 Proposed Ensemble Framework

To further enhance EEG-based brain-disorder detection, a three-stage hierarchical architecture is proposed that integrates advanced preprocessing, multi-feature extraction, and ensemble classification. This design systematically combines signal selection, deep-feature learning, and meta-level fusion to achieve high sensitivity, robustness, and clinical interpretability as shown in Figure 6.1. The framework integrates temporal windowing, multi-feature CNN–LSTM modelling, and meta-level ensemble stacking to form a unified, sensitivity-optimised system for automated EEG-based brain-disorder detection.

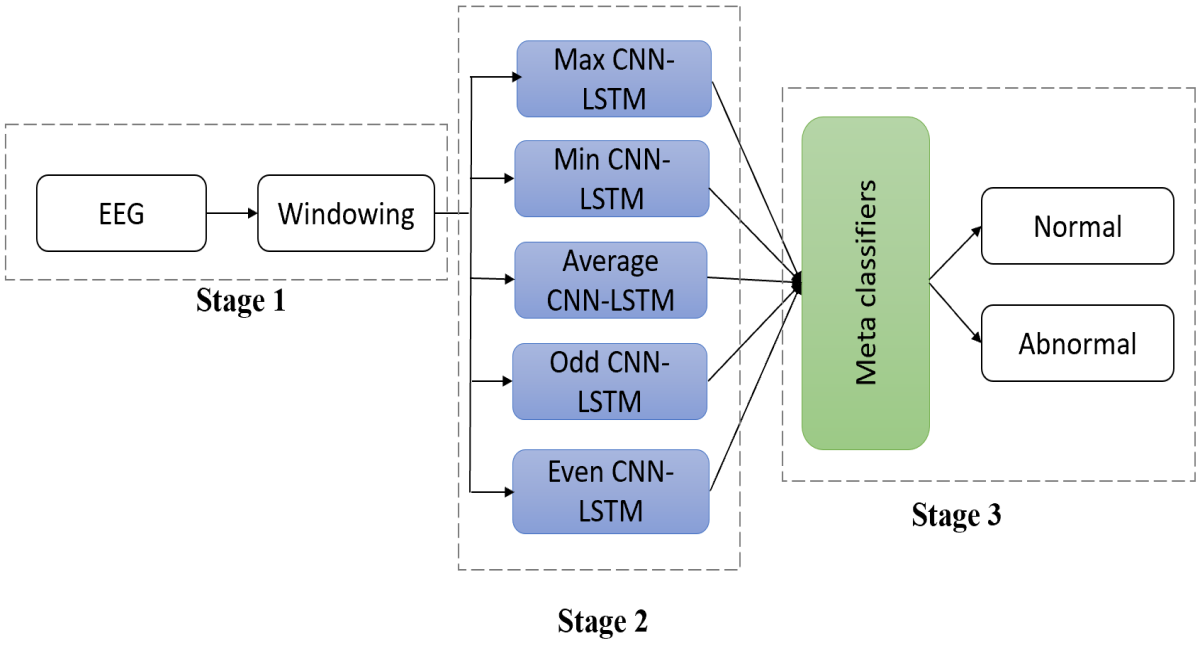


Figure 6.1: Three-Stage EEG Classification Framework Integrating Windowing, Multi-Feature Extraction, and Ensemble Learning.

Stage 1: EEG Windowing and Preprocessing

The first stage performs temporal windowing on the raw EEG recordings using the optimised time–frequency parameters established in Chapter 5. EEG signals are transformed via the Generalised Morse Wavelet (GMW) method to obtain time–frequency representations. Windows of 500, 2000, and 8000 samples were evaluated to identify the most informative temporal segments while suppressing redundant background activity. The analysis demonstrated that the 8000-sample window consistently yielded the highest discriminative performance, confirming that this temporal duration provides the optimal balance between contextual coverage and noise reduction for CNN–LSTM processing.

To ensure that the subsequent feature-extraction stage preserves this effective input length, a window of 16,000-sample EEG segment is first provided to the five feature extraction methods. Each method then reduces the segment to 8,000 samples, allowing all transformed signals to match the optimal window size identified previously. This design ensures that every CNN–LSTM sub-

model receives an input length already demonstrated through the Chapter 5 windowing experiments to produce superior classification accuracy.

Stage 2: Multi-Feature Extraction and CNN–LSTM Modelling

In this stage, we propose a novel multi-resolution feature extraction framework that systematically reduces the original EEG signal length from 16,000 samples to 8,000 samples while preserving diagnostically relevant information. This reduction is achieved through five distinct down-sampling techniques, each designed to capture complementary aspects of neural activity:

- Max pooling – captures transient high-energy spikes and epileptiform events.
- Min pooling – detects low-amplitude neural rhythms.
- Average pooling – preserves global temporal trends and smoother activity.
- Odd decimation – samples odd-indexed data points to emphasise sparse signal dynamics.
- Even decimation – complements odd sampling to ensure complete temporal coverage.

(These down-sampling techniques were previously detailed in Section 3.4.)

Each feature-transformed input is passed through an 18-layer CNN–LSTM network, which combines convolutional spatial filtering with recurrent temporal modeling. The CNN layers extract local spatial and spectral representations, while the LSTM layers capture sequential dependencies across time. Each of the five sub-models outputs class probabilities indicating normal or abnormal EEG activity.

Stage 3: Ensemble Stacking and Meta-Learning

Following the generation of diverse feature representations and the training of parallel CNN–LSTM sub-models, the third stage integrates their outputs through an ensemble stacking framework designed to enhance sensitivity, stability, and overall diagnostic reliability.

- Support Vector Machine (SVM) Stacking:

Employs a radial basis function (RBF) kernel to learn non-linear boundaries between the outputs of the CNN–LSTM sub-models. The SVM stacking mechanism maximises the separation margin between classes, thereby improving generalisation and reducing susceptibility to model-specific noise.

- Linear Regression (LR) Stacking:

Applies a linear weighting scheme to the five model outputs. This method emphasises simplicity and interpretability, allowing each CNN–LSTM prediction to contribute proportionally to the final decision according to its learnt weight.

- Random Forest (RF) Stacking:

Constructs an ensemble of randomised decision trees to model complex, non-linear interactions among the CNN–LSTM outputs. RF stacking is particularly effective in handling high-dimensional, noisy, and partially correlated features conditions typical of EEG data. It aggregates tree-level predictions through majority or probability-weighted stacking, providing high robustness and resistance to overfitting.

All ensemble models are trained and assessed using a 10-fold cross-validation procedure to enhance statistical robustness and reduce bias introduced by a single train–test split. The dataset was partitioned into ten equally sized folds, where nine folds were used for training and the remaining fold served as the test set in each iteration. This cycle was repeated until every fold had been used once for testing, and the final performance metrics were obtained by averaging the results across all iterations.

During training, the base CNN–LSTM models are frozen, and only the ensemble meta-learners are optimised. Training accuracy, validation accuracy, sensitivity, and specificity are monitored to assess convergence and performance stability. Finally, the three ensemble techniques (SVM, LR, and RF) are compared to determine the most effective stacking strategy for EEG-based brain disorder classification. The complete three-stage pipeline is formally presented in Algorithm 6.1, which provides a step-by-step mathematical formalisation of the proposed multi-feature extraction and ensemble stacking framework, detailing the feature extraction strategies, CNN-LSTM classification outputs, and meta-classifier decision fusion process.

Detailed fold-wise performance metrics for the Random Forest–based stacking ensemble model are provided in Appendix D.

Algorithm 6.1: Multi-Feature Extraction and Ensemble Stacking

Input / Output

Input: Windowed EEG segment W of length 16,000 samples, E = 19 electrodes
Output: Final classification label Y (Normal / Abnormal)

Notation

`W_e`	Windowed signal of electrode e
`F1_e`	Odd decimation output for electrode e
`F2_e`	Min pooling output for electrode e
`F3_e`	Max pooling output for electrode e
`F4_e`	Average pooling output for electrode e
`F5_e`	Even decimation output for electrode e
`n`	Sample index
`CNN-LSTM_k`	k-th CNN-LSTM sub-model, k = 1 TO 5
`P_k`	Probability output of k-th CNN-LSTM sub-model
`P`	Concatenated probability outputs [P1, P2, P3, P4, P5]
`Meta`	Meta-classifier ∈ {Random Forest, SVM, Linear Regression}
`Y`	Final predicted class label (Normal / Abnormal)

Steps

Stage 1: Multi-Feature Extraction

```
1. FOR each electrode e = 1 TO 19 DO
    2. F1_e = W_e(n) WHERE n is odd, n = 1,3,5,...,15999
    3. F2_e = min(W_e(2n-1), W_e(2n)) for n = 1 TO 8000
    4. F3_e = max(W_e(2n-1), W_e(2n)) for n = 1 TO 8000
    5. F4_e = (W_e(2n-1) + W_e(2n)) / 2 for n = 1 TO 8000
    6. F5_e = W_e(n) WHERE n is even, n = 2,4,6,...,16000
7. END FOR
```

Stage 2: CNN-LSTM Classification

```
8. FOR each feature set Fk, k = 1 TO 5 DO
    9. Feed Fk into CNN-LSTM_k
    10. Obtain probability output P_k = CNN-LSTM_k(Fk)
11. END FOR
```

Stage 3: Ensemble Stacking

```
12. Concatenate outputs: P = [P1, P2, P3, P4, P5]
13. Feed P into Meta-classifier
        WHERE Meta ∈ {Random Forest, SVM, Linear Regression}
14. Y = Meta(P)
```

```
15. RETURN Y
```

6.4 Results and Discussion

6.4.1 Performance of Individual Feature Extraction Techniques

Before introducing the ensemble stacking systems, the five distinct CNN–LSTM models were independently trained and evaluated to assess the influence of different feature extraction strategies on EEG classification performance. These five models namely, Model 1(Odd-CNN-LSTM), Model 2(Min-CNN-LSTM), Model 3(Max-CNN-LSTM), Model 4(Avg-CNN-LSTM), and Model 5(Even-CNN-LSTM) represent diverse down-sampling approaches applied to the same EEG data, each designed to emphasise specific temporal or amplitude-based signal characteristics. The experiments were conducted using the windowed EEG data (Section 5.6) and validated through 10-fold cross-validation to ensure statistical reliability.

An ablation study was carried out to systematically examine the discriminative contribution of each feature extraction method. The compared strategies included: (1) average pooling, which smooths high-frequency noise through mean aggregation of adjacent samples; (2) max pooling, which highlights transient high-amplitude events such as epileptic spikes; (3) min pooling, which captures inhibitory phases and low-amplitude signal troughs; (4) odd decimation, which preserves odd-indexed temporal samples; and (5) even decimation, which retains even-indexed temporal samples. Each method reduced the EEG signal dimensionality from 16,000 to 8,000 points per epoch while preserving unique information content.

The results for the five models based on TP, TN, FP, FN, precision, recall (sensitivity), and F1 score are given in Table 6.1. The evaluation was performed on a testing dataset of 274 patients, including 126 abnormal and 148 normal cases. The results reveal notable trade-offs between sensitivity and precision across the models. Model 1 demonstrates the highest recall (0.8889), indicating strong ability to identify positive cases, which is further supported by the highest F1 score (0.7782). However, this comes at the cost of a relatively high false positive count (FP = 50), which lowers its precision (0.6914). Model 2 shows a modest balance with lower precision (0.6494) and recall (0.7937), resulting in a lower F1 score (0.7141). Model 3, on the other hand, achieves the highest precision (0.7946) with a significantly lower FP rate (23), but suffers from the lowest recall (0.7063) among all models, reflecting a tendency to miss more actual positive cases (FN = 37). Model 4 presents the most balanced precision-recall combination (0.7049 and 0.6825), but with a slightly reduced overall F1 score (0.6945). Model 5 offers a middle ground,

with moderate precision (0.6812) and recall (0.7460), producing an F1 score of 0.7118. Overall, some models show high sensitivity but lower specificity, while others exhibit the opposite pattern. This variation highlights the strengths and weaknesses of individual models and motivates the proposal of a stacking system to leverage the complementary advantages of different models.

It is clear from the results that all models achieved high accuracy, whereas their sensitivity and specificity varied considerably, demonstrating that no single extraction approach could fully capture the diverse characteristics of EEG signals. Specifically, max pooling achieved superior sensitivity for detecting transient pathological events but exhibited reduced specificity, whereas min pooling displayed the opposite trend. Average pooling delivered balanced but less distinctive results, while odd and even decimation offered complementary perspectives with moderate overall performance. These findings confirm that each technique contributes unique and partially overlapping insights into the underlying EEG dynamics thereby justifying the need for an ensemble stacking mechanism that integrates their complementary strengths to achieve more robust and clinically reliable brain disorder classification.

Table 6.1 Comparative Analysis of Odd-, Min-, Max-, Avg-, and Even-CNN–LSTM Models Using Key Classification Metrics. The evaluation was conducted on a testing set comprising 274 patients. The results highlight the trade-offs between sensitivity (recall), specificity, and overall predictive balance across models.

	TP	**TN**	**FP**	**FN**	**Precision**	**Recall**	**F1 Score**	**Total Abnormal**	**Total Normal**
Model 1	112	98	50	14	0.6914	0.8889	0.7782	126	148
Model 2	100	94	54	26	0.6494	0.7937	0.7141	126	148
Model 3	89	124	24	37	0.7946	0.7063	0.7470	126	148
Model 4	86	112	36	40	0.7049	0.6825	0.6945	126	148
Model 5	94	104	44	32	0.6812	0.7460	0.7118	126	148

The confusion matrices for the five models (Model 1–Model 5) are presented in Figure 6.2. Model 1(Odd-CNN-LSTM) demonstrates strong sensitivity with 112 true positives, but also a relatively high number of false positives (50). Model 2(Min-CNN-LSTM) shows slightly reduced sensitivity compared to Model 1, with 100 true positives, and a similar level of false positives (54). Model 3 (Max-CNN-LSTM) achieves the best balance, with 89 true positives and the lowest false positives

(23), while also obtaining the highest true negatives (125), highlighting strong specificity. Model4 (Avg-CNN-LSTM) reflects a higher false negative count (40), though it compensates with a larger number of true negatives (112). Finally, Model 5 (Even-CNN-LSTM) provides moderate performance, with 94 true positives and 104 true negatives, but also a noticeable level of misclassifications (32 false negatives and 44 false positives).

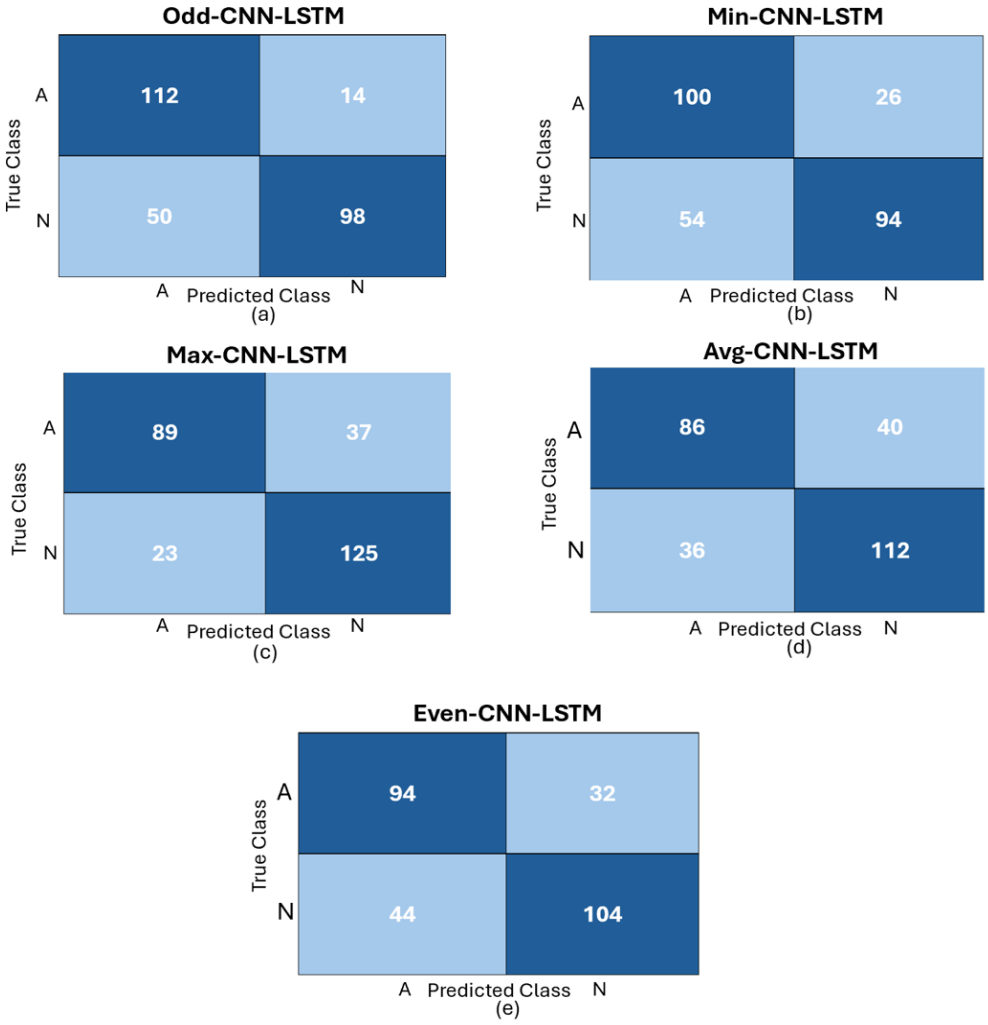
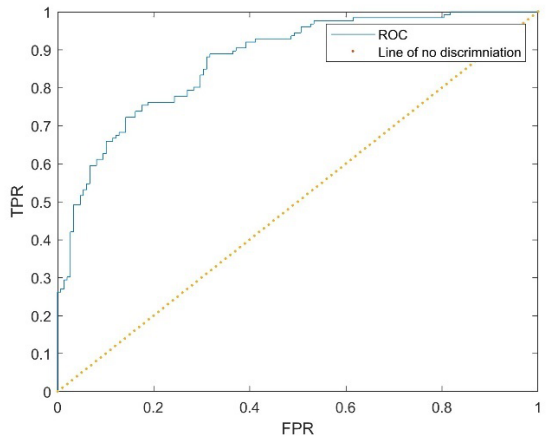


Figure 6.2: Confusion Matrices for the Five Windowing-Based CNN–LSTM Models Showing Classification of Abnormal (A) and Normal (N) EEG Signals.

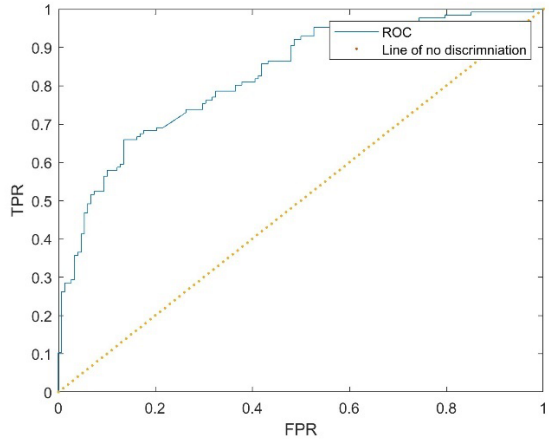
Overall, the comparison across models highlights varying trade-offs between sensitivity and specificity, suggesting that no single model dominates across all metrics, and ensemble integration may provide a more robust classification outcome.

Figure 6.3 presents the ROC curves for the five windowing-based CNN–LSTM models (a) Odd-CNN-LSTM, (b) Min-CNN-LSTM, (c) Max-CNN-LSTM, (d) Avg-CNN-LSTM, and (e) Even-CNN-LSTM illustrating their ability to discriminate between abnormal (A) and normal (N) EEG signals. All models demonstrate strong classification behaviour, with ROC trajectories rising well above the diagonal line of no discrimination. The Odd- and Min-based models ((a) and (b)) achieve the steepest initial ascent and highest true-positive rates across most false-positive ranges, indicating superior sensitivity. The Max- and Avg-CNN-LSTM models ((c) and (d)) show consistent and stable performance, with curves that maintain a favourable balance between sensitivity and specificity. The Even-CNN-LSTM model (e) also achieves competitive performance, though with a slightly more gradual rise at low false-positive rates.

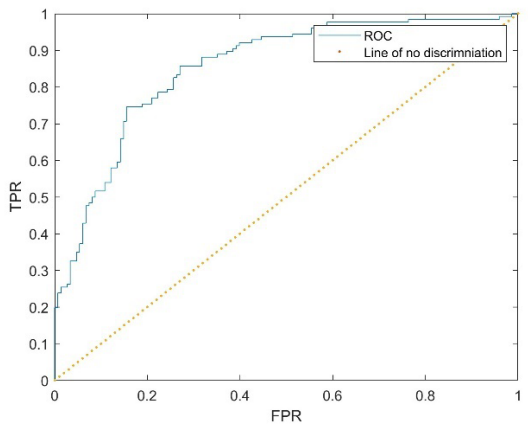
Detailed epoch-wise training accuracy and cross-entropy loss curves for each technique are provided in Appendix C.



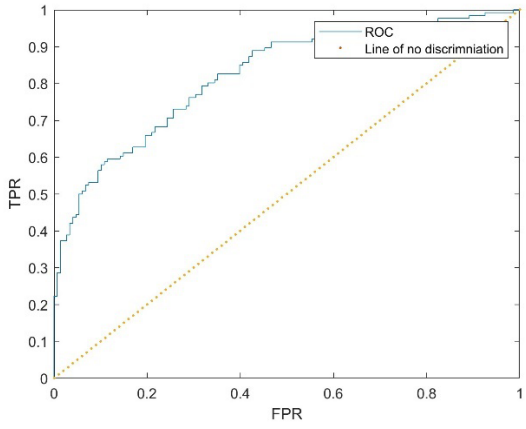
(a)



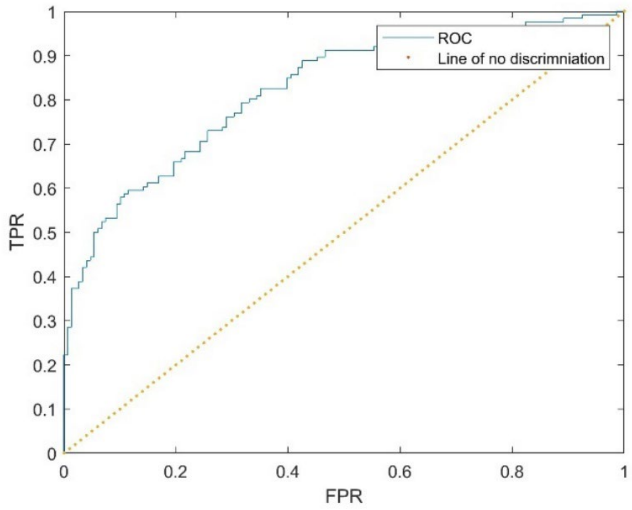
(b)



(c)



(d)



(e)

Figure 6.3: ROC curves for the Five Windowing-Based CNN–LSTM Models Showing Classification of Abnormal (A) and Normal (N) EEG Signals, illustrating their classification performance for abnormal (A) and normal (N) EEG signals. (a) Odd-CNN-LSTM, (b) Min-CNN-LSTM, (c) Max-CNN-LSTM, (d) Avg-CNN-LSTM, and (e) Even-CNN-LSTM.

The performance across the five models is shown in Figure 6.4 which demonstrates a clear trade-off between sensitivity and specificity. Model 1 (Odd-CNN-LSTM) achieves the highest sensitivity (88.9%), making it the most effective at detecting positive cases, but this comes with a lower specificity (66.2%), indicating a higher rate of false positives. Model 2 (Min-CNN-LSTM) shows a relatively high sensitivity (79.4%) but with the lowest specificity (63.5%) among all models, suggesting a similar bias toward identifying positives. In contrast, Model 3 (Max-CNN-

LSTM) stands out with the highest specificity (84.5%), reflecting its strength in correctly identifying negative cases, and it also achieves the highest accuracy (78.1%), despite having the lowest sensitivity (70.6%). Model 4 (Avg-CNN-LSTM) offers a balanced performance with moderate sensitivity (68.3%) and specificity (75.7%), leading to an accuracy of 76.4%. Model 5 (Even-CNN-LSTM), while showing the lowest accuracy (72.26%), maintains relatively balanced sensitivity (74.6%) and specificity (70.6%), suggesting a more neutral classification behaviour. It is obvious from these results that some models are more sensitive but less specific, while others are more conservative in detecting positives. To benefit from the strengths of each, a stacking system is proposed, where different ensemble strategies will be tested and evaluated to achieve a more robust and balanced classification performance. Figure 6.4 illustrates a bar chart that visually compares the accuracy, sensitivity, and specificity trends across all five models, helping to highlight the differences in their behaviour.

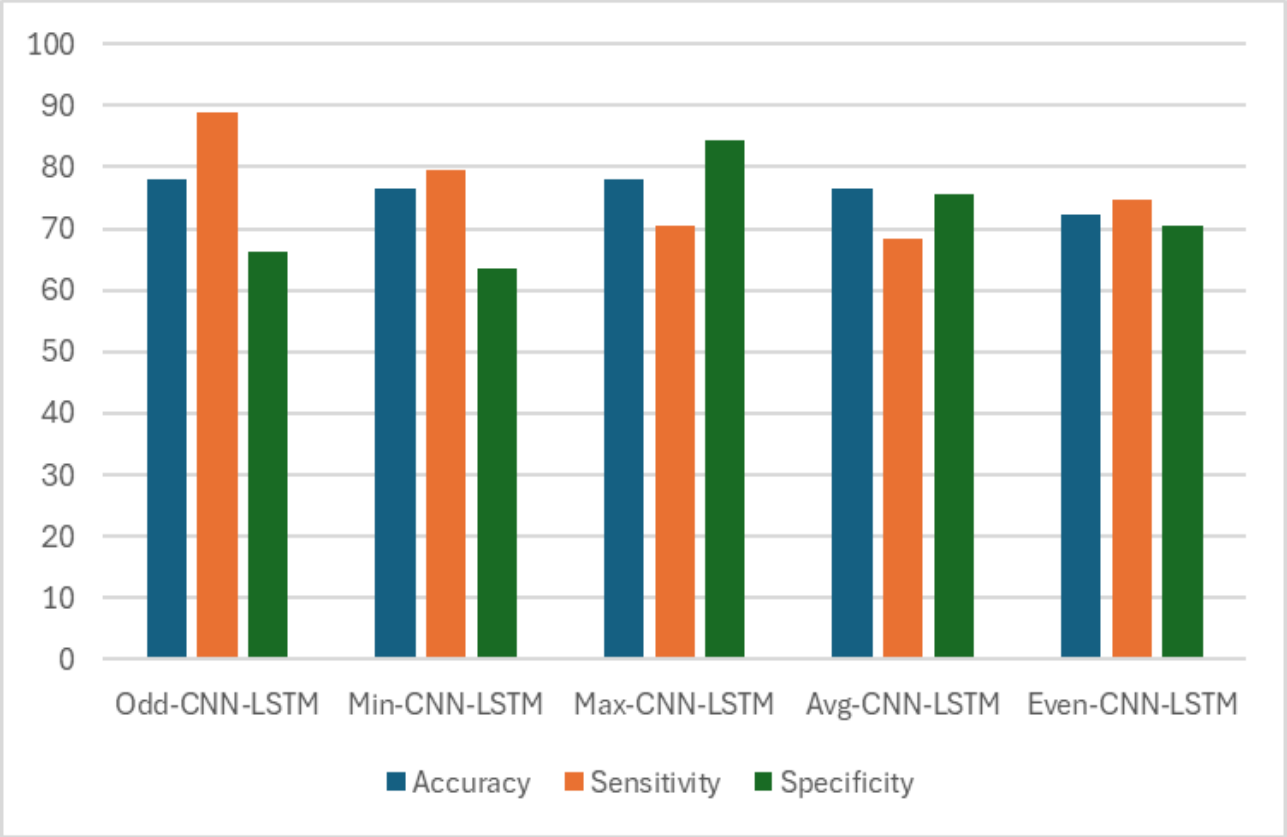


Figure 6.4: Bar Chart for comparison of accuracy, sensitivity, and specificity across the five models.

6.4.2 Performance Assessment of Ensemble Learning Approaches

To enhance EEG-based brain disorder classification, we propose a two-stage meta-learning framework, where the predictions of five base models (Model1-Model5) serve as inputs to the three distinct meta-classifiers, namely, Random Forest (RF), Support Vector Machine (SVM), and Linear Regression (LR). Each meta-classifier is trained and evaluated on accuracy, sensitivity, and specificity to determine which approach best improves upon the individual models' performance. This systematic comparison will identify whether a non-linear (RF/SVM) or linear (LR) meta-learner is most effective for consolidating diverse EEG classification outputs in clinical settings. Table 6.2 presents a comparison of three stacking system architectures SVM, logistic regression, and Random Forest applied to EEG-based brain disorder classification, using TP, TN, FP, FN, precision, recall (sensitivity), and F1 score as evaluation metrics. Among the models, the Random Forest-based stacking system achieved the highest recall (92.86%), indicating superior ability to correctly identify patients with brain disorders.

This is particularly critical in medical diagnosis, where missing a true positive case (i.e., a false negative) can lead to delayed or incorrect treatment. Although the Random Forest model exhibited the lowest precision (74.05%), it yielded the highest F1 score (82.37%), reflecting a strong balance between precision and recall. The logistic regression-based system achieved the highest precision (83.33%) but had the lowest recall (75.39%), potentially making it less suitable for medical applications where false negatives are more detrimental than false positives. The SVM-based system offered a balanced performance, with a precision of 77.14%, recall of 85.71%, and an F1 score of 81.2%. Overall, despite its slightly lower precision, the Random Forest-based stacking system stands out as the most appropriate choice in this medical context due to its exceptional sensitivity, making it the most reliable option for minimising missed diagnoses in brain disorder detection.

Table 6.2. Performance comparison of three stacking-based ensemble models SVM, LR, and RF evaluated using true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), precision, recall, and F1-score.

	TP	**TN**	**FP**	**FN**	**Precision**	**Recall**	**F1 Score**
SVM-based stacking Model	108	116	32	18	0.7714	0.8571	0.812
LR-based stacking Model	95	129	19	31	0.8333	0.7539	0.7917
RF-based stacking Model	117	107	41	9	0.7405	0.9286	0.8237

The confusion matrices reveal distinct performance characteristics for the three ensemble stacking models, as illustrated in Figure 6.5. The SVM-based model achieves 108 true positives and 116 true negatives, corresponding to a sensitivity of 77.1% and a specificity of 86.6%. This reflects balanced performance, though the 32 false negatives highlight some limitations in sensitivity. The Linear Regression-based model records 95 true positives and 129 true negatives, yielding a sensitivity of 75.4% and the highest specificity of 87.2% among the models. While it is strong at correctly identifying negative cases, its lower sensitivity indicates a higher risk of missed detections. In contrast, the Random Forest-based model achieves the highest sensitivity at 92.9% with 117 true positives and only 9 false negatives, making it the most effective at minimising missed cases. However, its specificity is lower at 72.3% due to a higher false positive rate of 41. Overall, the Random Forest model is most suitable for sensitivity-critical clinical applications, the Linear Regression model is preferable when specificity is prioritised, and the SVM model provides a moderate balance between the two.

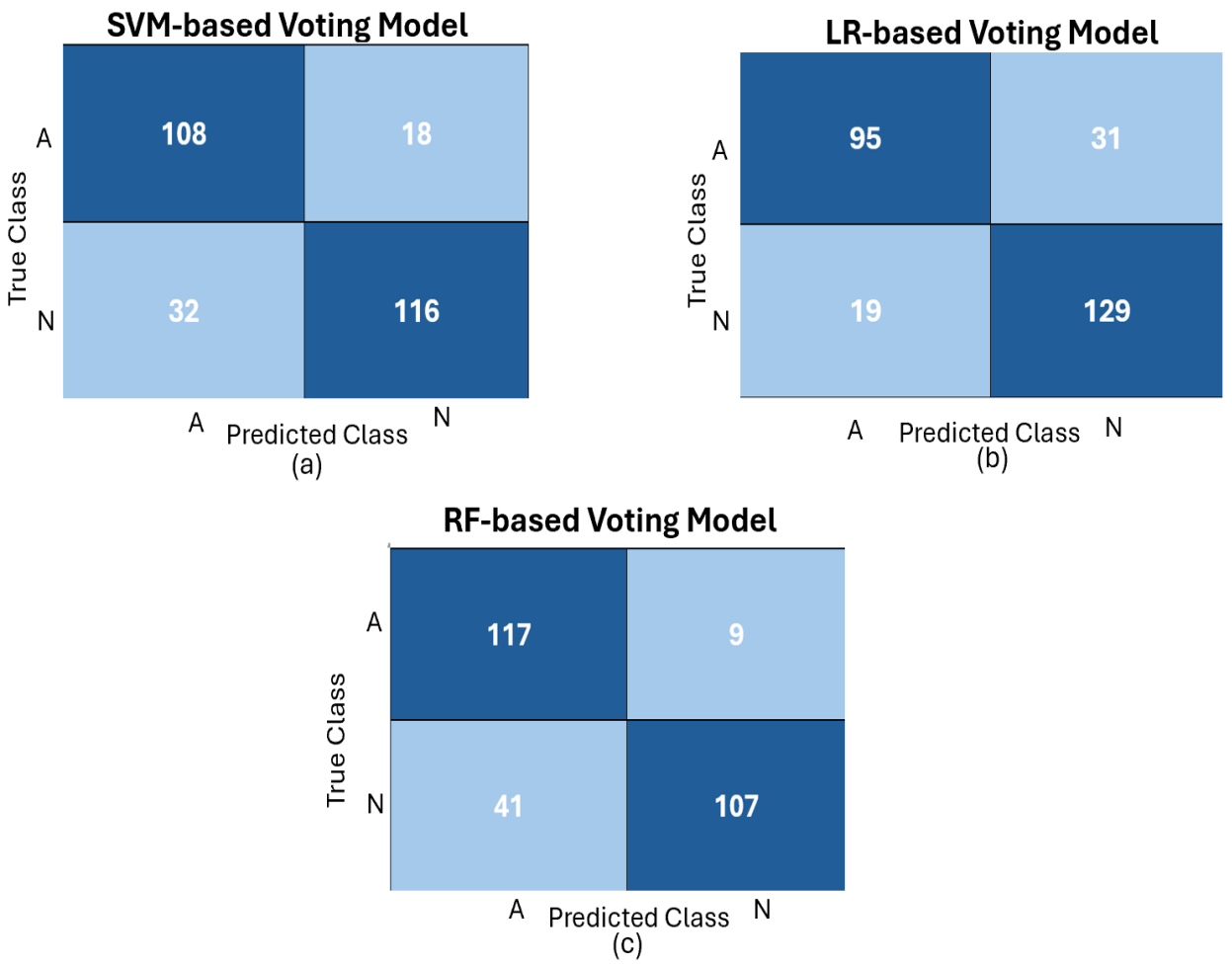
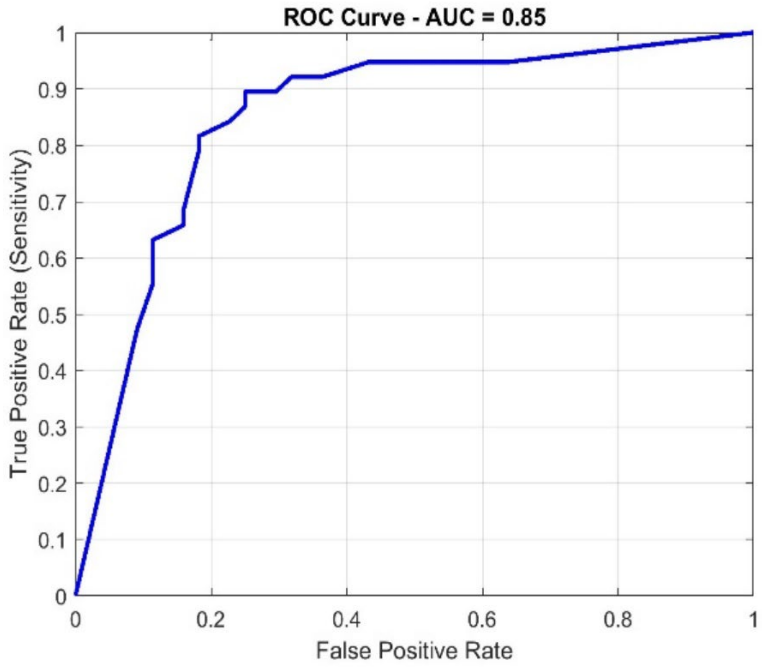


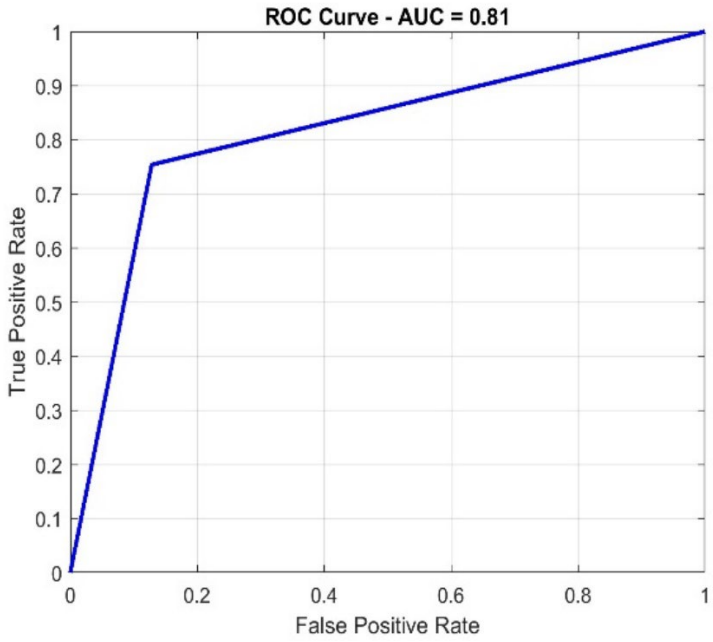
Figure 6.5: Confusion matrices of the three stacking-based ensemble models: (a) SVM-based, (b) Linear Regression-based, and (c) Random Forest-based. The matrices illustrate classification outcomes in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), highlighting the trade-offs between sensitivity and specificity for each technique.

The ROC curves in Figure 6.6 illustrate the classification performance of the three stacking-based ensemble models: (a) SVM-based stacking, (b) Linear Regression-based stacking, and (c) Random Forest-based stacking. All three models achieve strong discriminative capability between abnormal and normal EEG signals, with ROC curves rising well above the diagonal baseline. The SVM-based ensemble (a) demonstrates robust performance with an AUC of 0.85, indicating a good balance between sensitivity and false-positive rate across decision thresholds. In contrast, the Linear Regression-based model (b) shows a lower AUC of 0.81, reflecting reduced discriminative power and a more gradual ROC slope at low false-positive rates. The Random Forest-based ensemble (c) achieves the highest performance, with an AUC of 0.89 and a steeper initial ascent,

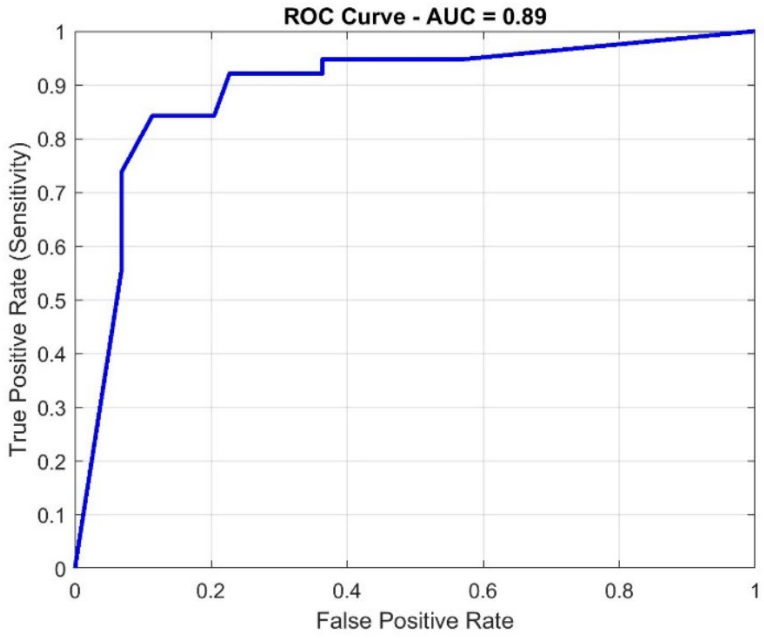
highlighting superior sensitivity and overall reliability. Comparatively, the Random Forest-based ensemble clearly outperforms both the SVM- and Linear Regression-based models, achieving the strongest ROC characteristics and the most effective decision-fusion strategy. Overall, these results confirm that integrating multiple CNN–LSTM base models through ensemble stacking enhances classification robustness, with the Random Forest meta-classifier providing the most effective decision-fusion strategy.



(a)



(b)



(c)

Figure 6.6. ROC curves for the three stacking-based ensemble models: (a) SVM-based, (b) Linear Regression-based, and (c) Random Forest-based, illustrating their discriminative ability in classifying Abnormal (A) and Normal (N) EEG signals.

The bar chart in Figure 6.7 compares the performance of three ensemble stacking techniques SVM, Linear Regression, and Random Forest used in EEG-based brain disorder classification, evaluated by accuracy, sensitivity, and specificity. The Random Forest-based system outperformed the others in terms of sensitivity (92.86%) and accuracy (81.8%), making it the most effective at correctly identifying patients with brain disorders. This high sensitivity is particularly valuable in medical applications where minimising false negatives is critical to avoid missed diagnoses. SVM showed a balanced performance with an accuracy of 81.2%, sensitivity of 85.7%, and specificity of 78.38%, suggesting it maintains a reasonable trade-off between detecting true positives and avoiding false positives. In contrast, the Linear Regression model achieved the highest specificity (87.1%) indicating a strong ability to correctly identify healthy individuals but at the cost of the lowest sensitivity (75.4%), which could lead to more undiagnosed cases. Given the medical importance of capturing all true disorder cases, the Random Forest approach is the most clinically appropriate due to its superior sensitivity and overall robust performance. In EEG classification, sensitivity is a critical performance metric, particularly in clinical applications where missing pathological activity such as epileptic seizures or abnormal brain rhythms can have serious consequences. High sensitivity ensures that the model detects the majority of true neurological events, reducing the risk of false negatives that could delay diagnosis or intervention (Aslam et al., 2022). While specificity and accuracy remain important for minimising false alarms and overall correctness, sensitivity is often prioritised in EEG analysis due to the high stakes of overlooking critical abnormalities. Striking a balance between these metrics is challenging, as overly sensitive systems may increase false positives, but in many medical contexts, the cost of missing a true event outweighs the cost of additional verification. Thus, optimising sensitivity without compromising practical utility is a key focus in EEG classification research.

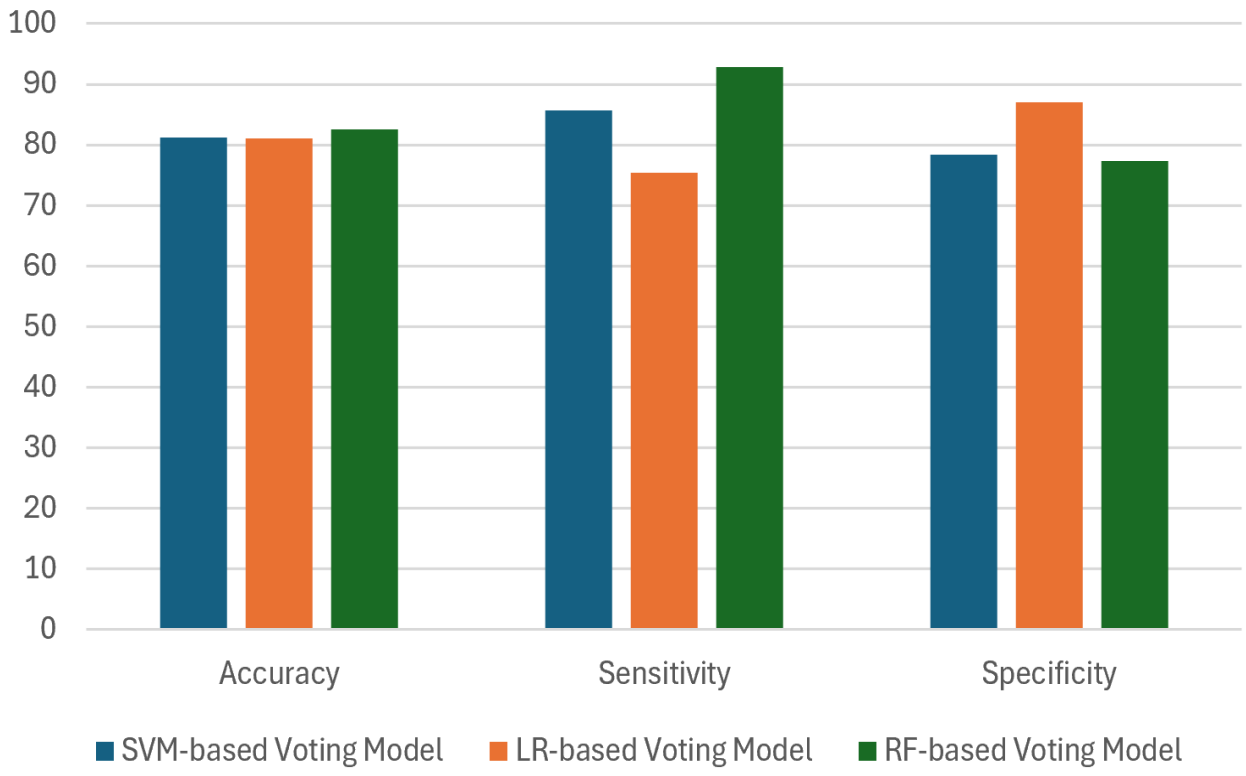


Figure 6.7: Bar Chart for the three stacking-based ensemble models: SVM-based, Linear Regression-based, and Random Forest-based. The Chart illustrate classification outcomes in terms of sensitivity and specificity and accuracy for each technique.

6.5 Progressive Development of the Proposed Classification Models

To provide a clear overview of how the proposed EEG classification framework evolved throughout this research, this section presents a structured comparison of all major model versions developed from the initial baseline to the final ensemble system. Each modeling stage introduced specific methodological enhancements aimed at improving diagnostic accuracy, sensitivity, and generalisation performance. Beginning with the proposed 18-layer deep learning architecture, subsequent stages incorporated wavelet-based windowing, followed by multi-feature extraction and, ultimately, ensemble stacking.

Table 6.3 provides a comparative overview of the proposed models, highlighting the effects of each successive enhancement. The baseline deep learning model, trained on full-length EEG recordings, achieved limited accuracy and exhibited insufficient sensitivity, demonstrating its difficulty in consistently detecting abnormal patterns within noisy, unsegmented signals. The

introduction of windowing-based feature selection substantially improved performance, increasing accuracy to 82.68% and markedly enhancing both sensitivity and specificity. The multi-feature CNN–LSTM stage further strengthened the model's ability to detect abnormalities, achieving the highest non-ensemble sensitivity of 88.90%, although its specificity was comparatively lower due to a greater number of false positives. The final Random Forest ensemble demonstrated the most clinically favourable balance, achieving a sensitivity of 92.86%, an accuracy of 81.8%, and a specificity of 72.30%. This progression clearly illustrates how each modeling refinement contributed to improved discriminative power and diagnostic reliability.

Table 6.3. Progressive evolution of the proposed classification models, detailing feature-processing strategies and performance metrics at each stage.

Model Stage	Model Description	Input Features	Accuracy (%)	Sensitivity (%)	Specificity (%)
Stage 1 Proposed Deep Learning Model	18-layer CNN–LSTM baseline model	Full-length EEG (3,005,000 samples)	67.06	—	—
Stage 2 Windowing-Based CNN–LSTM Model	CNN–LSTM trained on optimal 8000-sample temporal window	Windowed EEG (8000 samples)	82.68	78.50	87.10
Stage 3 Multi-Feature CNN–LSTM Model	odd-decimation feature set	Windowed EEG (8000 transformations)	76.60	88.90	66.20
Final Model RF Stacking Ensemble	Random Forest meta-classifier integrating 5 CNN–LSTM outputs	5 CNN–LSTM probability outputs	81.8	92.86	72.30

The performance trends outlined in Table 6.3 are further visualised in Figure 6.8, which presents a comparative bar chart of accuracy, sensitivity, and specificity across all model stages. The chart highlights the marked gains achieved through feature selection and ensemble integration, particularly the stepwise improvement in sensitivity, an essential metric for clinical EEG interpretation. While the multi-feature model attained the strongest sensitivity among the standalone CNN–LSTM architectures, only the ensemble model succeeded in achieving high sensitivity without sacrificing overall accuracy. Figure 6.8 therefore reinforces the progressive benefits of the proposed modeling strategy, demonstrating how each enhancement contributed to a more robust and clinically aligned classification framework.

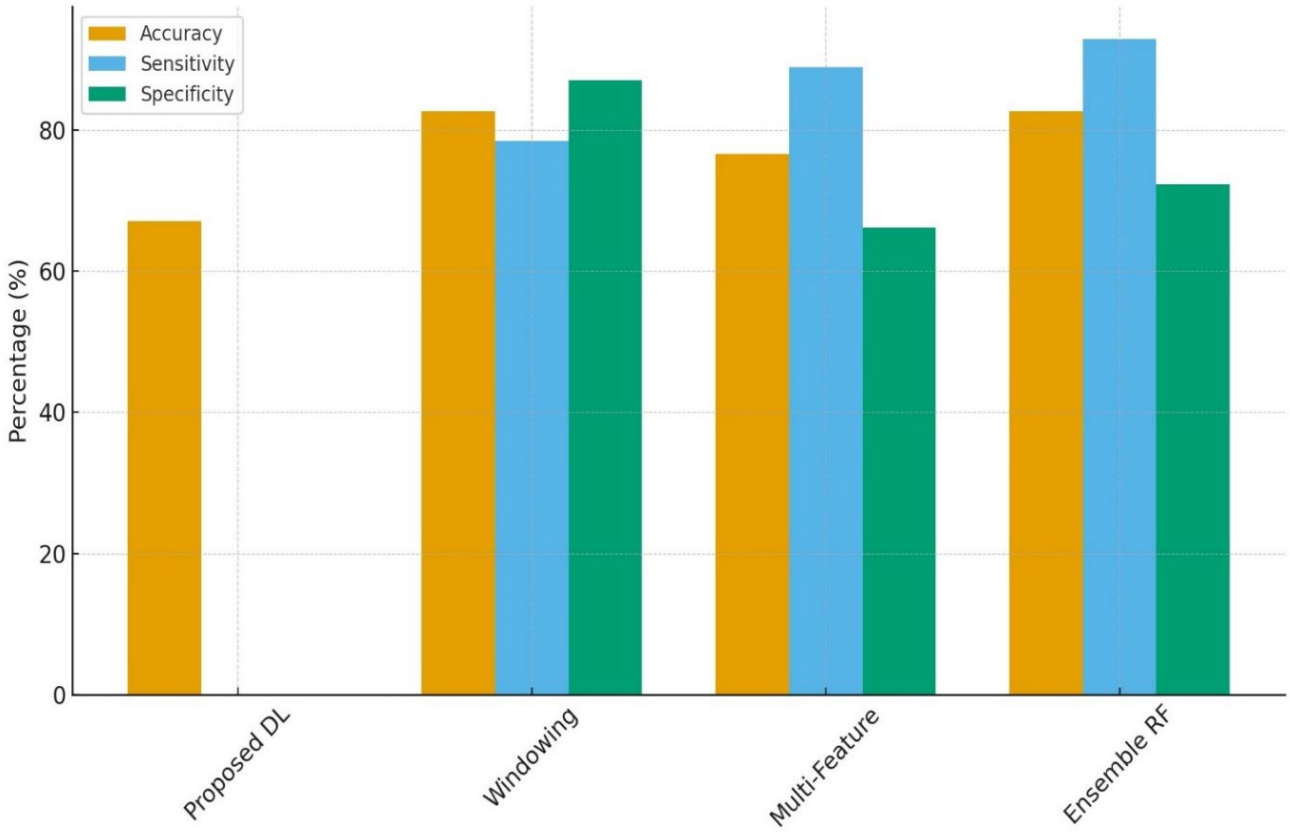


Figure 6.8: Performance comparison across model development stages.

6.6 Comparison with previous work

To contextualise the performance of the proposed framework, its results were compared with several studies from the literature that employed different EEG preprocessing strategies, feature extraction methods, and classification techniques. Table 6.4 summarises the preprocessing strategy, feature extraction approach, classification algorithm, and the reported accuracy and sensitivity for each study. Sharma, Patel and Acharya (2020) reported an accuracy of 79.34% and a sensitivity of 77.54% using first-minute EEG recordings with fuzzy entropy, logarithmic squared norm, and fractal dimension features classified by an SVM. Tomas et al. (2020) achieved an accuracy and sensitivity of 68% using non-overlapping one-second EEG epochs with power spectrum, phase locking value, and energy features classified using a Hidden Markov Model. Western et al. (2021) employed a deep CNN with automatic feature learning from raw EEG signals extracted from the second minute of each recording, reporting an accuracy of 81.88% and a sensitivity of 71%. Albaqami et al. (2021) combined Wavelet Packet Decomposition (WPD) with a LightGBM classifier, achieving an accuracy of 86.59% and a sensitivity of 81.74%, while Wu et al. (2022) utilised Discrete Wavelet Transform (DWT) features with a CatBoost classifier, obtaining the highest reported accuracy of 89.13% and a sensitivity of 84.92%. Ortega-Flores et al. (2025) proposed a Quaternion Convolutional Neural Network (QCNN) based on DWT preprocessing to preserve multi-channel relationships within EEG signals, reporting an accuracy of 77.17% and a sensitivity of 80%. Performance values are reported as presented in the respective publications. As the studies employ different preprocessing strategies, feature extraction methods, and evaluation protocols, the results are intended for qualitative comparison rather than direct performance benchmarking.

The proposed framework combines adaptive windowing, five complementary feature extraction strategies, CNN–LSTM classifiers, and a Random Forest stacking model, achieving an overall accuracy of 81.8% and the highest sensitivity of 92.86% among the studies considered. Although several existing approaches report higher overall accuracy, the proposed framework demonstrates superior sensitivity, indicating an enhanced ability to correctly identify abnormal EEG recordings. In clinical EEG screening, high sensitivity is particularly important because it reduces the likelihood of false-negative classifications, thereby decreasing the risk of missed pathological casesThe enhanced sensitivity observed in the proposed model is primarily attributed to its three-stage hierarchical framework, which integrates (1) optimised windowing-based feature selection,

(2) multi-resolution feature extraction through five CNN–LSTM models, and (3) ensemble learning using Random Forest Stacking. This multi-stage design ensures that critical temporal-spatial patterns are preserved, and complementary representations of EEG activity are fused for more reliable classification.

Furthermore, the incorporation of diverse down-sampling methods (max, min, average, odd, and even decimation) enables the network to capture both transient and persistent signal dynamics, while the Random Forest Stacking mechanism effectively reduces overfitting and enhances decision reliability. The result is a clinically aligned model that prioritises sensitivity, a metric often undervalued in machine learning research but indispensable for neurological disorder detection.

It is therefore apparent that while other recent studies have achieved slightly higher accuracies using specialised boosting or wavelet-based frameworks, none have reached the same level of diagnostic sensitivity as the proposed method. This underscores the model's potential for real-world clinical deployment, where early and accurate identification of brain abnormalities is critical for timely treatment and improved patient outcomes.

Table 6.4: Comparison of Model Performance with different architectures in literature in terms of data selection technique, feature extraction, classification technique, accuracy and sensitivity.

	Data Selection Technique	**Feature Extraction**	**Classification Technique**	**Accuracy**	**Sensitivity**
Sharma, Patel & Acharya (2020)	1st minute	Fuzzy Entropy+ Logarithmic Squared Norm+ Fractal Dimension	SVM	79.34%	77.54%
Tomas et al., (2020)	Non-overlapping 1 second epochs	PS + PLV + Energy	HMM	68%	68%
Western et al., (2021)	2nd minute selection	Automatic feature learning (raw EEG input)	CNN	81.88%	71%

Albaqami et al., (2021)	Standard EEG preprocessing	WPD	LightGBM	86.59%	81.74%
T Wu et al., (2022)	Non-overlapping 5 second window segmentation	DWT	CatBoost	89.13%	84.92%
Ortega-Flores *et al.* (2025)	First-minute removal, 23-channel selection.	DWT	Quaternionic CNN	77.17%	80%
Proposed Model	Adaptive Windowing Technique	5 Multi model + CNN LSTM model	Random Forest Stacking system	81.8%	92.86%

Comparison of sensitivity and specificity across different EEG classification studies is shown in the bar chart in Figure 6.9. The proposed Random Forest–based ensemble model demonstrates the highest sensitivity (92.86%) compared to existing approaches, confirming its superior ability to detect abnormal EEG patterns while maintaining competitive specificity.

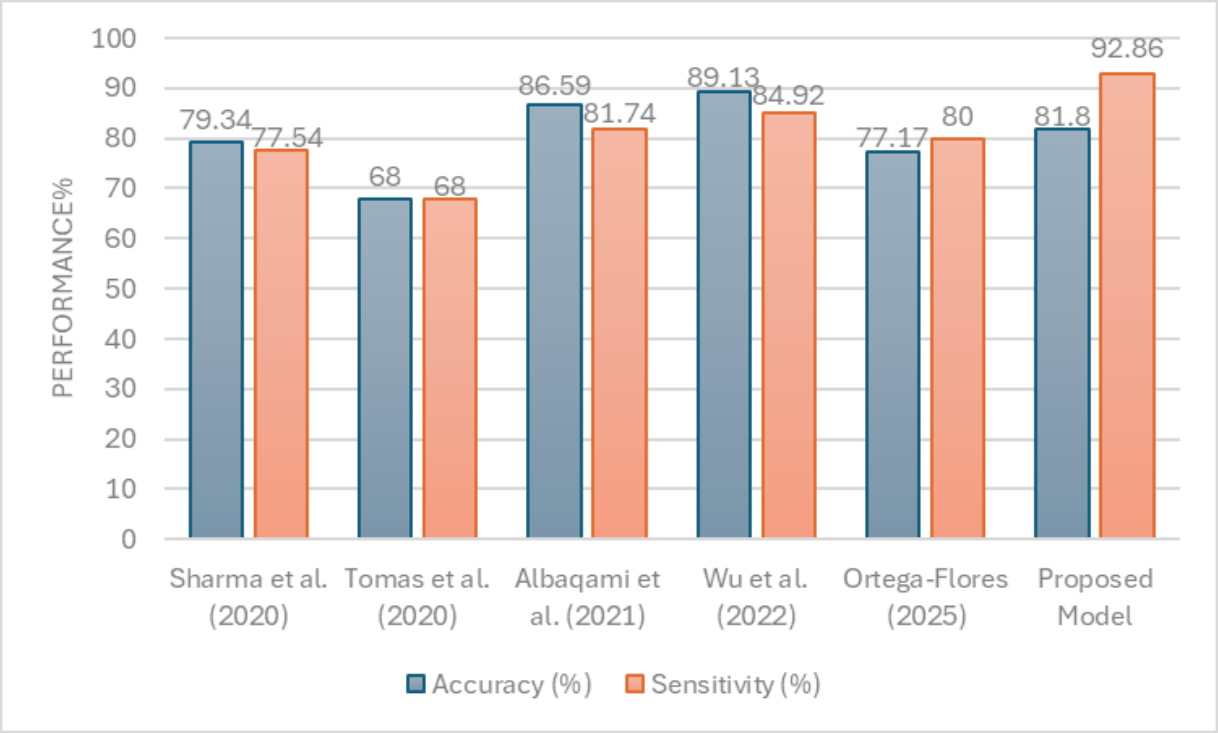


Figure 6.9: Comparison of Sensitivity and Specificity Across EEG Classification Studies

6.7 Summary

This chapter presented the development and comprehensive evaluation of the proposed EEG-based brain disorder classification framework, which integrates three main stages: (1) windowing-based feature selection, (2) multi-feature CNN–LSTM modeling, and (3) ensemble learning through stacking-based meta-classifiers.

Initially, five distinct feature extraction techniques: max pooling, min pooling, average pooling, odd decimation, and even decimation were independently tested with the CNN–LSTM model to investigate their individual contributions to classification performance. The results highlighted considerable variation among models in terms of sensitivity and specificity, confirming that no single technique could fully capture the diverse temporal spatial features of EEG signals.

To overcome these limitations, three ensemble learning approaches were explored: Support Vector Machine (SVM), Linear Regression (LR), and Random Forest (RF) stacking. These meta-learners combined the probabilistic outputs of the five CNN–LSTM sub-models to enhance overall robustness and generalisation. The comparative results revealed that while all ensembles improved stability and diagnostic consistency, the Random Forest Stackting system achieved the highest

sensitivity (92.86%) and overall accuracy (81.8%), demonstrating exceptional ability to detect abnormal EEG patterns.

Finally, the proposed framework was benchmarked against several state-of-the-art studies, where it achieved comparable accuracy and superior sensitivity, confirming its clinical applicability and reliability for automated EEG-based brain disorder detection. The strong performance achieved through this multi-stage approach validates the importance of integrating adaptive feature selection, deep temporal–spatial learning, and ensemble decision fusion in achieving sensitivity-critical medical diagnostics.

The next chapter presents a comprehensive discussion of these findings, examining the effectiveness of the proposed integrated framework and its individual components, comparing the results with existing studies, and addressing the practical implications and limitations of the proposed approach. Furthermore, it summarises the overall contributions and key findings of the research, and outlines potential directions for future work, including experimentation with deeper model architectures, real-time clinical implementation, and multi-modal signal integration.

Chapter 7
Discussion, Conclusion and Future work

7.1 Overview

This research aimed to develop a robust and clinically reliable framework for the automated detection of brain disorders using electroencephalogram (EEG) signals. Recognising the limitations of conventional EEG analysis such as high noise levels, significant signal variability, and heavy reliance on expert interpretation the study sought to enhance diagnostic performance through advanced feature selection, deep-learning modelling, and ensemble integration. The proposed system was designed to fulfil the following key objectives:

1- Feature Selection: The objective of developing a novel feature-selection method to improve the representativeness of EEG data was achieved by introducing a windowing-based technique that identified the most informative temporal segments for analysis.
2- Multi-Feature Extraction: The objective of implementing multiple feature-extraction strategies was achieved by selecting five distinct techniques: Maximum, Minimum, Average, Odd, and Even sampling each designed to capture different aspects of EEG signal characteristics.
3- Deep-Learning Architecture: The objective of designing a deep-learning model was met by proposing a novel 18-layer CNN–LSTM architecture, specifically tailored for both feature extraction and classification of EEG signals.
4- Ensemble Learning Framework: The objective of proposing a stacking-based ensemble learning strategy, leveraging the complementary strengths of the deep-learning models, was achieved by developing and validating a Random Forest ensemble stacking model. This integrated the outputs of the five CNN–LSTM classifiers, improving robustness and maximising sensitivity, particularly in detecting abnormal EEG cases.

5- Benchmarking Against State-of-the-Art Methods: To benchmark the proposed framework against recent literature, the TUH EEG Corpus a trusted and widely used dataset was employed to validate the methodology and demonstrate its clinical applicability.

This chapter provides an interpretive discussion of these findings, examining the influence of feature selection, CNN–LSTM modeling, and ensemble learning, and situating the results within the broader context of EEG-based brain disorder classification research. Furthermore, it highlights the clinical implications, methodological limitations, and future research directions that emerge from the outcomes of this study.

7.2. Effectiveness of the Proposed Integrated Framework

Initial experiments using a baseline CNN–LSTM deep-learning model trained on full-length EEG recordings demonstrated satisfactory training behaviour but limited generalisation to unseen data. While the model achieved reasonable classification accuracy, its performance indicated limitations in reliably capturing discriminative patterns associated with abnormal EEG activity. These observations motivated the introduction of two methodological enhancements: (1) a windowing-based feature selection strategy aimed at isolating the most informative EEG segments, and (2) the incorporation of multiple feature-extraction techniques designed to capture complementary temporal and amplitude-based signal characteristics.

While these improvements significantly enhanced classification accuracy and overall model stability, the sensitivity values remained suboptimal, revealing that individual models based on single feature-extraction methods could not fully capture the diverse and non-stationary nature of EEG dynamics. This limitation motivated the development of a third enhancement an ensemble learning framework to integrate the probabilistic outputs of multiple CNN–LSTM models using Random Forest, Support Vector Machine, and Linear Regression stacking systems.

The proposed multi-stage architecture, comprising (1) optimised windowing, (2) multi-feature CNN–LSTM modelling, and (3) ensemble meta-classification, demonstrated substantial improvement in both sensitivity and diagnostic reliability. Among the three ensemble approaches, the Random Forest stacking model achieved the highest sensitivity (92.86%) and accuracy (81.8%), confirming its superior ability to detect abnormal EEG activity.

7.2.1 Effectiveness of Windowing-Based Feature Selection

The results presented in Chapter 5 clearly demonstrated that windowing-based feature selection played a critical role in improving the reliability and efficiency of EEG classification. The initial CNN–LSTM model trained on full-length EEG signals showed high training accuracy but comparatively weaker testing performance. This indicated that large, unsegmented EEG recordings introduced redundant and noisy temporal information, leading to overfitting and reduced model generalisation.

To address this limitation, the windowing technique was introduced to identify and isolate the most informative temporal intervals within each EEG recording. By applying Generalised Morse Wavelet (GMW) decomposition, the method detected periods of maximal energy concentration across multiple electrodes, corresponding to the time windows where clinically relevant activity such as epileptiform discharges or abnormal oscillations was most likely to occur.

Three window sizes (500, 2000, and 8000 samples) were evaluated to examine the trade-off between temporal resolution and feature richness. The results revealed a consistent upward trend in performance as the window size increased: accuracy improved from 74.78% (500 samples) to 82.68% (8000 samples), accompanied by a corresponding rise in sensitivity from 70.77% to 78.5% and specificity from 70.58% to 87.1%. This progression confirmed that the chosen windowing strategy effectively concentrated learning on the diagnostically relevant segments of the EEG while filtering out redundant background activity.

Compared with the correlation-based electrode selection approach which achieved a maximum testing accuracy of 60% the windowing method produced substantially superior results while utilising the complete 19-electrode configuration. This demonstrated that temporal refinement of EEG segments was more beneficial than spatial reduction when the goal is to preserve inter-channel relationships essential for detecting distributed neural abnormalities.

In summary, the windowing-based feature-selection approach proved essential in enhancing the CNN–LSTM model's discriminative capacity by emphasising the most informative portions of the EEG signals. This stage not only improved accuracy and sensitivity but also reduced

computational complexity, laying a solid foundation for the subsequent multi-feature extraction and ensemble learning stages described in this chapter.

7.2.2 Contribution of Multi-Feature CNN–LSTM Modelling

Following the successful implementation of the windowing-based feature-selection strategy, the next stage of this research focused on further enhancing the model's representational capability through multi-feature CNN–LSTM modelling. EEG signals are inherently complex, characterised by non-stationary dynamics and overlapping frequency components that vary across both time and brain regions. Consequently, a single feature-extraction method is insufficient to capture the full spectrum of diagnostically relevant information embedded within these signals.

To overcome this limitation, five complementary down sampling and pooling techniques namely max pooling, min pooling, average pooling, odd decimation, and even decimation were introduced. Each technique emphasised a distinct aspect of neural activity: max pooling preserved transient high-amplitude events such as epileptiform spikes; min pooling captured inhibitory phases and low-amplitude fluctuations; average pooling smoothed local variations while maintaining global signal structure; and odd/even decimation retained temporal diversity by subsampling the signal from two complementary viewpoints. Together, these transformations generated five alternative yet complementary representations of the same EEG recordings, forming the basis for five corresponding CNN–LSTM sub-models.

Each CNN–LSTM model employed one-dimensional convolutional layers to extract localised spatial features from the down-sampled EEG inputs, followed by LSTM layers designed to capture temporal dependencies and evolving neural patterns. Independent evaluation of the five sub-models revealed that, although all achieved high overall accuracy, they exhibited notable variability in sensitivity and specificity. For example, max-pooling-based models demonstrated superior sensitivity by effectively identifying pathological transients yet tended to produce more false positives; conversely, min-pooling-based models favoured specificity but risked missing abnormal events. Average pooling delivered a balanced but less distinctive performance profile, while the decimation-based models produced more moderate yet stable generalisation.

These findings underscored the complementary strengths and weaknesses inherent to different feature-extraction strategies and provided the key rationale for developing an ensemble stacking mechanism to integrate their outputs. By combining the diverse CNN–LSTM models, it became

possible to leverage their collective discriminatory capabilities while compensating for the limitations of any single approach. Thus, the multi-feature CNN–LSTM framework served as a critical intermediate stage, linking signal-level diversity with high-level ensemble learning, and ultimately contributing to a more robust and clinically reliable EEG-based brain disorder classification system.

7.2.3 Effectiveness of Ensemble Learning and Meta-Classification

While the individual CNN–LSTM models derived from different feature extraction techniques achieved strong performance, the variation among their sensitivity and specificity metrics revealed that no single model could capture the full range of discriminative EEG features. This limitation motivated the adoption of an ensemble meta-classification strategy, designed to integrate the complementary strengths of all five CNN–LSTM sub-models into a unified decision framework.

The proposed three-stage ensemble system combined (1) windowing-based temporal feature selection, (2) multi-feature CNN–LSTM modelling, and (3) ensemble stacking through meta-classifiers. In this final stage, three distinct stacking mechanisms: Support Vector Machine (SVM), Linear Regression (LR), and Random Forest (RF) were implemented to aggregate the probabilistic outputs of the five CNN–LSTM models. Each stacking system represented a unique integration approach:

- SVM stacking, optimised class separation boundaries within a high-dimensional feature space, effectively balancing sensitivity and specificity.
- LR stacking, assigned linear weights to each model's predictions, emphasising interpretability and simplicity in the fusion process.
- RF stacking, on the other hand, leveraged multiple decision trees to capture nonlinear relationships between the outputs of the base models, thereby enhancing robustness and generalisation.

The results confirmed that the ensemble approach substantially improved classification reliability compared with any single CNN–LSTM model. Among the three stacking strategies, the Random Forest ensemble achieved the highest performance, with 92.86% sensitivity, 81.8% accuracy, and a strong F1 score of 82.37%, indicating a remarkable ability to identify abnormal EEG patterns with minimal false negatives. Although the Linear Regression and SVM ensembles demonstrated

competitive accuracy and superior specificity, they underperformed in sensitivity, which is the most critical metric for medical diagnosis.

This performance trend emphasises the importance of ensemble learning in EEG-based classification, where complex, noisy, and non-stationary signals often cause inconsistencies in individual models. By aggregating the probabilistic decisions of multiple CNN–LSTM models, the ensemble framework reduced model bias, mitigated overfitting, and captured diverse neural representations more effectively. The integration of Random Forest stacking, in particular, proved most advantageous for sensitivity-critical clinical applications, such as seizure detection or neurological disorder screening, where missing abnormal activity carries significant diagnostic risk.

In summary, the ensemble learning and meta-classification stage demonstrated the highest overall improvement in predictive reliability and clinical value, validating the effectiveness of combining feature diversity with hierarchical decision fusion for EEG-based brain disorder detection.

7.3 Discussion of results in Comparison with Previous Work

To contextualise the performance of the proposed framework, comparisons with previously reported EEG-based classification studies were conducted. The discussion highlights how the integration of windowing-based feature selection, multi-feature CNN–LSTM modelling, and Random Forest ensemble learning achieved superior diagnostic sensitivity while maintaining competitive overall accuracy.

The comparative results indicated that the proposed Random Forest stacking ensemble achieved a sensitivity of 92.86% and an accuracy of 82.68%, marking a substantial improvement in the model's ability to correctly identify abnormal EEG patterns. Although the accuracy is slightly lower than the best-reported CatBoost model, the sensitivity gain is clinically significant, as high sensitivity ensures the early and reliable detection of pathological events, minimising the risk of missed diagnoses. This trade-off is particularly important in neurological applications, where false negatives can have more severe consequences than false positives. Furthermore, the proposed model benefits from a hierarchical structure that combines three complementary stages: temporal refinement (windowing), spatial-temporal learning (CNN–LSTM), and decision fusion (ensemble stacking). This layered design distinguishes it from previous single-stage approaches, allowing it

to capture a broader spectrum of EEG dynamics while reducing model overfitting and enhancing generalisation.

The comparative results also revealed that the proposed Random Forest ensemble framework achieved the highest sensitivity compared to existing literature, highlighting its effectiveness in minimising false negatives and improving clinical reliability for EEG-based brain disorder detection.

7.4 Practical Implications

The findings of this research have clear clinical and practical significance. Achieving high sensitivity is critical in EEG-based diagnosis, as it minimises missed detections of pathological events such as epileptic seizures or abnormal rhythms errors that could delay treatment or lead to misdiagnosis. Previous studies have reported that human interpretation of EEG recordings is associated with error rates ranging from 25% to 30%, primarily due to inter-rater variability, clinician fatigue, and the inherent complexity of EEG patterns. Consequently, the proposed automated framework provides a reliable decision-support tool that complements neurologists' expertise, reducing diagnostic subjectivity and improving the accuracy and consistency of EEG-based assessments.

Furthermore, the framework could be integrated into computer-assisted diagnostic systems, providing neurologists with automated screening support that reduces manual interpretation time and variability. The inclusion of Random Forest ensemble learning enhances model interpretability by identifying the relative importance of extracted features, enabling clinicians to understand the basis of the model's decisions. This interpretability is particularly valuable in clinical environments, as it increases trust and facilitates regulatory approval for AI-based diagnostic systems.

Overall, by achieving high sensitivity while maintaining competitive specificity, the proposed model addresses a critical challenge in clinical EEG analysis minimising false negatives without excessively increasing false positives. This balance ensures that more true neurological events are detected early, improving the timeliness and reliability of patient care.

7.5 Limitations

Despite its promising results, the proposed framework had limitations that should be acknowledged:

1. Dataset Dependency: The model was trained and validated solely on the TUH EEG Corpus, a large yet single-source dataset. Although the dataset is heterogeneous and representative of real-world hospital conditions, it inherently reflects the acquisition protocols, annotation standards, and patient demographics of a specific clinical environment. As a result, model performance may vary when applied to EEG data collected using different hardware configurations, sampling rates, or electrode montages.

2. Methodological and Architectural Limitations

   The proposed CNN–LSTM–based framework integrates multiple down-sampling and pooling strategies to capture complementary temporal–spectral characteristics of EEG signals. This design choice increases the architectural complexity of the model and necessitates access to high-performance computing resources, such as GPUs or advanced workstations, during model development and training. While this requirement does not affect the conceptual validity or clinical relevance of the proposed approach, it may pose practical constraints for institutions with limited computational infrastructure. In such cases, the reproducibility and re-training of the framework could be restricted without access to specialised hardware or shared computing facilities. Nevertheless, this limitation reflects an implementation consideration associated with deep-learning–based methodologies rather than a fundamental limitation of the proposed

3. Framework Validation and Evaluation Constraints

   Although a 10-fold cross-validation strategy was employed to ensure statistical robustness and reduce data-partition bias, the absence of external validation on an independent dataset remains a limitation. Cross-validation provides strong internal reliability; however, external testing is necessary to fully assess generalisation across different clinical populations and recording conditions. Moreover, the task formulation focused on abnormal versus normal EEG detection rather than disorder-specific diagnosis. While this aligns well

with clinical screening applications, it limits the interpretability of the results in the context of precise neurological disorder classification.

4. Class Imbalance and EEG Heterogeneity: Although the TUH dataset includes a mixture of normal and abnormal EEG recordings, underlying class imbalance and heterogeneity between different abnormal event types may still influence model performance. Abnormal EEG patterns vary widely in morphology, frequency content, and duration, and the framework does not explicitly model these subcategories. As a result, the model may generalise less effectively to rare or a typical pathological event. Future work could involve dedicated modelling of abnormal subtypes, class-balancing strategies, or anomaly-detection approaches to improve robustness across diverse clinical presentations.

Building on the findings and critical analysis presented in this chapter, the following chapter concludes the thesis by summarising the main contributions of the proposed framework and outlining directions for future research in EEG-based automated brain disorder detection.

7.6 Conclusion

This research presented a comprehensive, multi-stage EEG-based brain disorder detection framework that integrates optimised windowing, multi-resolution feature extraction, and ensemble deep learning to enhance classification performance and clinical reliability. The proposed approach was developed to address key limitations in conventional EEG analysis, including low sensitivity, limited interpretability, and challenges in model generalisation across diverse patient presentations.

The first stage of the framework employed Generalised Morse Wavelet (GMW)-based windowing to localise the most informative temporal segments of each EEG recording. This method effectively reduced redundant background activity while preserving diagnostically relevant events. The second stage introduced five distinct feature-extraction strategies average pooling, max pooling, min pooling, odd decimation, and even decimation each offering a unique representation of EEG signal dynamics. These were independently processed using a hybrid CNN–LSTM architecture, enabling the model to exploit both spatial and temporal dependencies inherent in EEG signals.

In the final stage, an ensemble-based decision fusion strategy was implemented using a Random Forest classifier to integrate the outputs of the five CNN–LSTM sub-models. This Random Forest based stacking mechanism demonstrated the strongest overall performance, achieving an accuracy of 81.8% and a sensitivity of 92.86%. The high sensitivity reflects a substantial reduction in false negative detections, which is particularly critical in clinical EEG screening, where missed abnormalities may lead to delayed diagnosis or inappropriate clinical decision-making. These findings confirm the effectiveness of the proposed framework as a robust and sensitivity-oriented solution for automated abnormal EEG detection in heterogeneous clinical environments.

Overall, the study demonstrated that the combination of adaptive temporal windowing, diverse feature extraction, and ensemble learning substantially improves the sensitivity, robustness, and clinical utility of automated EEG classification. By reducing human error, enhancing diagnostic consistency, and supporting clinician decision-making, the proposed framework lays a strong foundation for future development of computer-assisted diagnostic systems capable of real-time monitoring and widespread clinical deployment.

7.7 Future Work

Building on the findings of this research, several directions may be pursued to further enhance the robustness, interpretability, and clinical applicability of the proposed EEG-based brain disorder classification framework.

While the proposed framework demonstrates strong performance and clinical relevance, several directions for future research can further enhance its applicability and robustness. First, external validation on independent EEG datasets acquired from different clinical centres would provide a more comprehensive assessment of the model's generalisability across varied recording protocols, patient demographics, and clinical conditions. Such validation would strengthen confidence in the proposed framework for broader clinical deployment.

Second, future work could extend the current binary abnormal-versus-normal formulation towards disorder-specific classification, enabling more fine-grained differentiation between neurological conditions such as epilepsy, encephalopathy, and other pathological EEG patterns. Incorporating explicit spatial modelling techniques, such as graph-based neural networks or attention mechanisms, may further improve the exploitation of inter-electrode relationships and enhance classification performance.

Although the proposed CNN–LSTM architecture achieved strong performance, future research may investigate deeper and more advanced deep learning architectures to further improve feature learning and classification accuracy. Potential directions include deeper CNN–LSTM variants, Transformer-based models, Vision Transformers (ViTs), and hybrid architectures capable of capturing more complex spatial-temporal representations of EEG signals. Such models may provide enhanced representation learning while maintaining the ability to model long-range temporal dependencies.

Another promising direction is the incorporation of Explainable Artificial Intelligence (XAI) techniques to improve the transparency and interpretability of the proposed framework. Methods such as SHAP (SHapley Additive exPlanations), Grad-CAM (Gradient-weighted Class Activation Mapping), or attention visualisation techniques could provide clinicians with greater insight into the decision-making process of the deep learning models by highlighting the EEG regions and temporal patterns that contribute most significantly to the final classification. Improved

interpretability would increase clinician confidence and facilitate the adoption of deep learning systems in clinical practice.

Future studies may also investigate multimodal learning frameworks that combine EEG with complementary neuroimaging modalities, such as functional Magnetic Resonance Imaging (fMRI) or functional Near-Infrared Spectroscopy (fNIRS). The integration of multiple physiological data sources has the potential to provide more comprehensive representations of brain activity, thereby improving diagnostic accuracy and supporting more reliable neurological assessment.

Although class weighting was employed in the present study to address class imbalance, future research could explore more advanced imbalance-handling strategies. These may include synthetic data generation techniques such as the Synthetic Minority Over-sampling Technique (SMOTE), Generative Adversarial Networks (GANs), or other data augmentation methods capable of generating realistic EEG signals. Such approaches may improve classifier robustness, particularly when detecting rare abnormal EEG patterns.

In addition, future work could investigate the development of patient-specific or personalised deep learning models. While the current framework is designed to generalise across a heterogeneous patient population, personalised models that adapt to individual EEG characteristics may further improve diagnostic performance by accounting for subject-specific variability in brain activity. Adaptive learning strategies and personalised calibration procedures may therefore represent valuable extensions of the proposed framework.

Finally, future studies may investigate strategies for optimising model retraining and deployment, including transfer learning, incremental learning, and model compression techniques, to facilitate adaptation to new datasets and evolving clinical requirements. Furthermore, implementing the proposed framework within real-time clinical environments would enable continuous EEG monitoring and rapid decision support, representing an important step towards routine clinical application. Collectively, these future research directions provide multiple opportunities to extend the proposed framework, further improving its robustness, scalability, interpretability, and clinical utility in automated EEG analysis.

Appendices

Appendix A: Comparative Summary of EEG Studies using Feature Extraction and Classification Methods

Table 1: A summary on EEG studies based on features extraction and classification techniques Summary of EEG Studies

Author	Dataset	Feature Extraction	Classifiers	Accuracy
(Chandaka et al., 2009)	BONN university	cross-correlation	SVM	95.96%
(Ocak, 2009)	BONN university	DWT	ANN	96%
(Alturki et al., 2020)	BONN university	DWT	KNN	100%
(Sadati, Mohseni & Maghsoudi, 2006)	BONN university	DWT	SNFN	86%
(Subasi, 2007)	BONN university	DWT	Mixture of expert model	94.5%
(Subasi & Gursoy, 2010)	BONN university	DWT	SVM	98%
(Kannathal et al., 2005)	BONN university	Entropy	ANFIS	92.2%
(Yuan et al., 2011)	BONN university	Entropy	ELM	96.5%
(Nicolaou & Georgiou, 2012)	BONN university	Entropy	SVM	93.8%
(Hussein, 2016)	BONN university	RNN	LSTM	100%
(Nigam & Graupe, 2004)	BONN university	Nonlinear filter	ANN	97.2%
(Almustafa, 2020)	BONN university	variance threshold feature selection method	Random forest classifier	97.08%
(Nunes et al., 2014)	BONN university	WT	Optimum path forest	89.2%
(Thieu & Yang, 2015)	BONN university	DTCWT	NN	100%

Polat et al, (Polat & Güneş, 2007)	BONN university	Permutation Entropy	SVM	93.8%
(Xin et al., 2022)	BONN university	Wavelet multi scale	CNN	98.8%
(Hendel, Benyettou & Hendel, 2016)	Boston University	DWT	SVM	84.73%
(Truong et al., 2018)	CHB-MIT	Short time Fourier Transform	CNN	81.40%
(Yan et al., 2022)	CHB-MIT	STFT	Three tower transformers	96.01%
(Abdelhameed & Bayoumi, 2021)	CHB-MIT	(2D-DCAE)	LSTM	98.79%
(Fathima & Bedeeuzzaman, 2016)	CHB-MIT	1D-LBP Based	SVM	96.15%
(Li et al., 2020)	CHB-MIT	CNN	LSTM	98.44%
(Nasehi & Pourghassem, 2013)	CHB-MIT	DWT	IPSONN	98%
(Khan et al., 2012)	CHB-MIT	DWT	LDA	91.8%
(Zandi et al., 2013)	CHB-MIT	Gaussian Mixture Model	SVM	83.8%
(Bilal et al., 2019)	CHB-MIT	multi-resolution dynamic mode decomposition	Other	93%
(Ke et al., 2018)	CHB-MIT	Maximal Information Coefficient	CNN	98%
(Fergus et al., 2016)	CHB-MIT	PCA	K-NN	93%
(Chiang et al., 2011)	CHB-MIT	Wavelet coherence values	SVM	57.14%
(Orosco et al., 2016)	CHB-MIT	WT	LDA	87.5%
(Rasekhi et al., 2013)	EUR	FFT	SVM	73.9%

(Aarabi, Wallois & Grebe, 2006)	Other	linear correlation-based feature selection methods + Relief method	BNN	93%
(Sharma, Parashar & Joshi, 2021)	Phycology department Arizona university	CNN	LSTM	99.1%
(Schirrmeister et al., 2017)	TUH	CNN	CNN	78%
(Iešmantas & Alzbutas, 2020)	TUH	No technique	CNN	0.68%
(Golmohammadi et al., 2019)	TUH	HMM	Deep Learning	90%
(L.S., 2017)	TUH	PCA	CNN	75.4%
(Sharma, Patel & Acharya, 2020)	TUH	Orthogonal wavelet Decomposition	SVM	79.34%
(Yıldırım, Baloglu & Acharya, 2020)	TUH	No technique	CNN	79.34%
(Wu et al., 2022)	TUH	DWT	CatBoost	89.13%

Appendix B: Detailed 10-Fold Cross-Validation Performance Results for Different EEG Window Sizes

B.1 Results Obtained for The Different Simulated Models

B.1.1 Performance Metrics for Different EEG Window Sizes (500, 2000, and 8000 Samples)

Table 1: Calculated percentage sensitivity, specificity and accuracy for a window size of 500 samples, along with the number of True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN).

	TP	FN	FP	TN	Sensitivity	Specificity	Accuracy
Fold 1	79	41	36	74	65.83%	67.27%	66.52%
Fold 2	78	32	42	78	70.91%	65.00%	67.83%
Fold 3	82	38	28	82	68.33%	74.55%	71.30%
Fold 4	78	32	27	93	70.91%	77.50%	74.35%
Fold 5	93	27	41	69	77.50%	62.73%	70.43%
Fold 6	92	18	56	64	83.64%	53.33%	67.83%
Fold 7	86	34	25	85	71.67%	77.27%	74.35%
Fold 8	82	28	42	78	74.55%	65.00%	69.57%
Fold 9	74	46	25	85	61.67%	77.27%	69.13%
Fold 10	69	41	17	103	62.73%	85.83%	74.78%

Table 2: Calculated sensitivity, specificity and accuracy for a window size of 2000 samples

	TP	FN	FP	TN	Sensitivity	Specificity	Accuracy
Fold 1	71	49	26	84	59.17%	76.36%	67.39%
Fold 2	81	29	27	93	73.64%	77.50%	75.65%
Fold 3	90	30	30	80	75.00%	72.73%	73.91%
Fold 4	79	31	19	101	71.82%	84.17%	78.26%
Fold 5	72	48	18	92	60.00%	83.64%	71.30%
Fold 6	78	32	20	100	70.91%	83.33%	77.39%
Fold 7	90	30	36	74	75.00%	67.27%	71.30%
Fold 8	94	16	39	81	85.45%	67.50%	76.09%
Fold 9	86	34	28	82	71.67%	74.55%	73.04%
Fold 10	80	30	39	81	72.73%	67.50%	70.00%

Table 3: Calculated sensitivity, specificity, and accuracy for a window size of 8000 samples

	TP	FN	FP	TN	Sensitivity	Specificity	Accuracy
Fold 1	83	47	30	94	63.85%	75.81%	69.69%
Fold 2	104	24	59	67	81.25%	53.17%	67.32%
Fold 3	96	28	19	111	77.42%	85.38%	81.50%
Fold 4	105	23	39	87	82.03%	69.05%	75.59%
Fold 5	94	36	27	97	72.31%	78.23%	75.20%
Fold 6	93	31	38	92	75.00%	70.77%	72.83%
Fold 7	118	8	68	60	93.65%	46.88%	70.08%

Fold 8	102	28	16	108	78.46%	87.10%	82.68%
Fold 9	98	28	24	104	77.78%	81.25%	79.53%
Fold 10	74	50	23	107	59.68%	82.31%	71.26%

Appendix C: Detailed Training Logs for CNN–LSTM Models Using Five Down-Sampling Strategies

Appendix C.1: Average Pooling Training Logs

As shown in Figures C.1 and C.2, the training curves obtained using the average pooling strategy exhibit a gradual and stable learning pattern. The epoch-wise average accuracy increases progressively after the initial epochs, indicating steady refinement of the learned representations. In parallel, the cross-entropy loss shows a smooth and consistent decrease, reflecting stable optimisation and convergence. This behaviour is characteristic of average pooling, which preserves global signal characteristics by smoothing local fluctuations, leading to reduced training noise and improved stability, albeit with a relatively slower convergence rate compared to more selective down-sampling techniques.

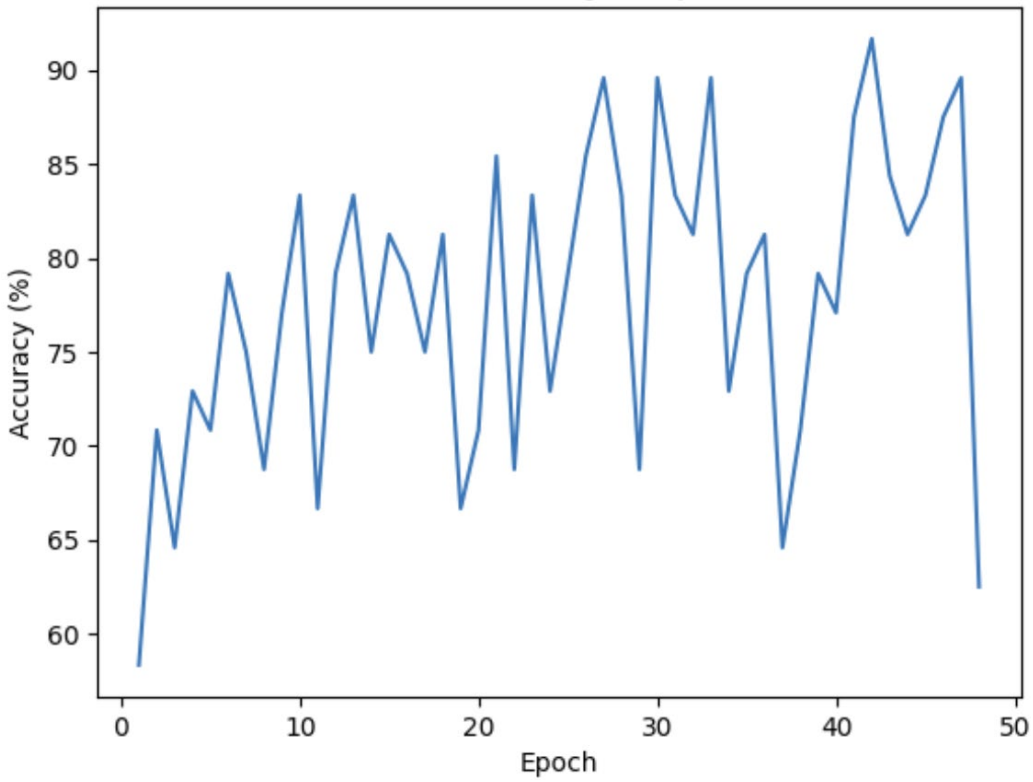


Figure C.1: Epoch-wise average training accuracy for the CNN–LSTM model using the average pooling down-sampling technique**.**

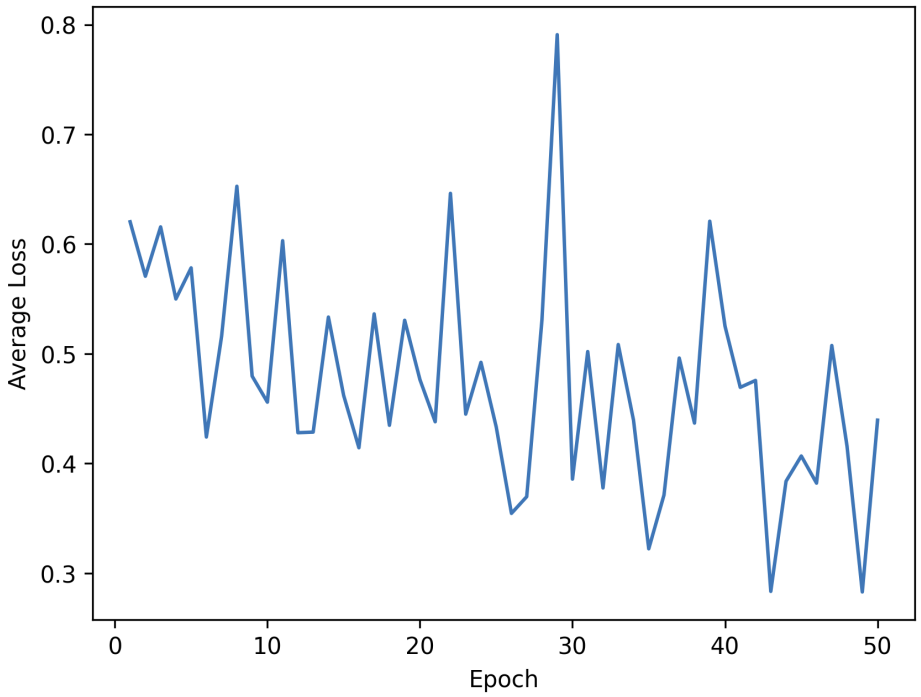


Figure C.2: Epoch-wise average cross-entropy training loss for the CNN–LSTM model using the average pooling down-sampling technique.

Appendix C.2: Max Pooling Training Logs

As illustrated in Figures C.3 and C.4, the max pooling model demonstrates rapid learning dynamics, evidenced by a sharp increase in training accuracy during the early epochs. This suggests that the model efficiently captures salient high-amplitude EEG features, such as transient spikes. Correspondingly, the cross-entropy loss decreases rapidly, indicating effective optimisation. Minor fluctuations in the loss curve reflect the sensitivity of max pooling to dominant activations, resulting in fast convergence with slightly increased variability.

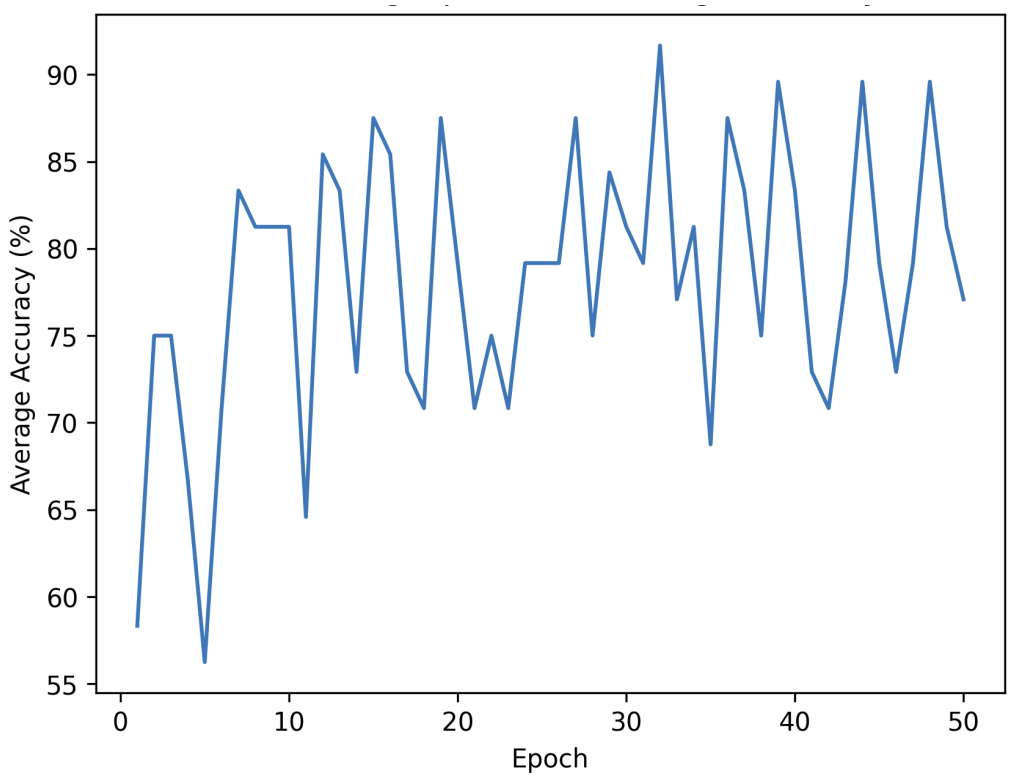


Figure C.3: Epoch-wise average training accuracy for the CNN–LSTM model using the max pooling down-sampling technique.

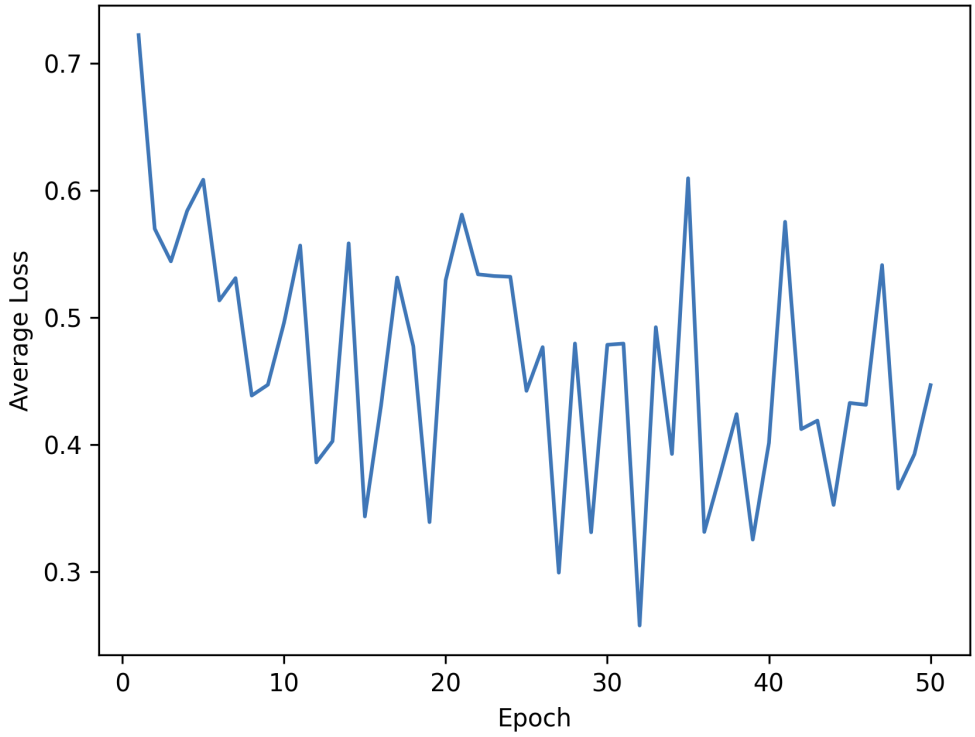


Figure C.4: Epoch-wise average cross-entropy training loss for the CNN–LSTM model using the max pooling down-sampling technique.

Appendix C.3: Min Pooling Training Logs

As shown in Figures C.5 and C.6, training with the min pooling strategy results in a more conservative convergence pattern. The accuracy curve shows steady but moderate improvement across epochs, while the cross-entropy loss decreases smoothly without pronounced oscillations. This behaviour reflects the emphasis of min pooling on low-amplitude and inhibitory signal components, which evolve more gradually during training, contributing to stable and robust optimisation.

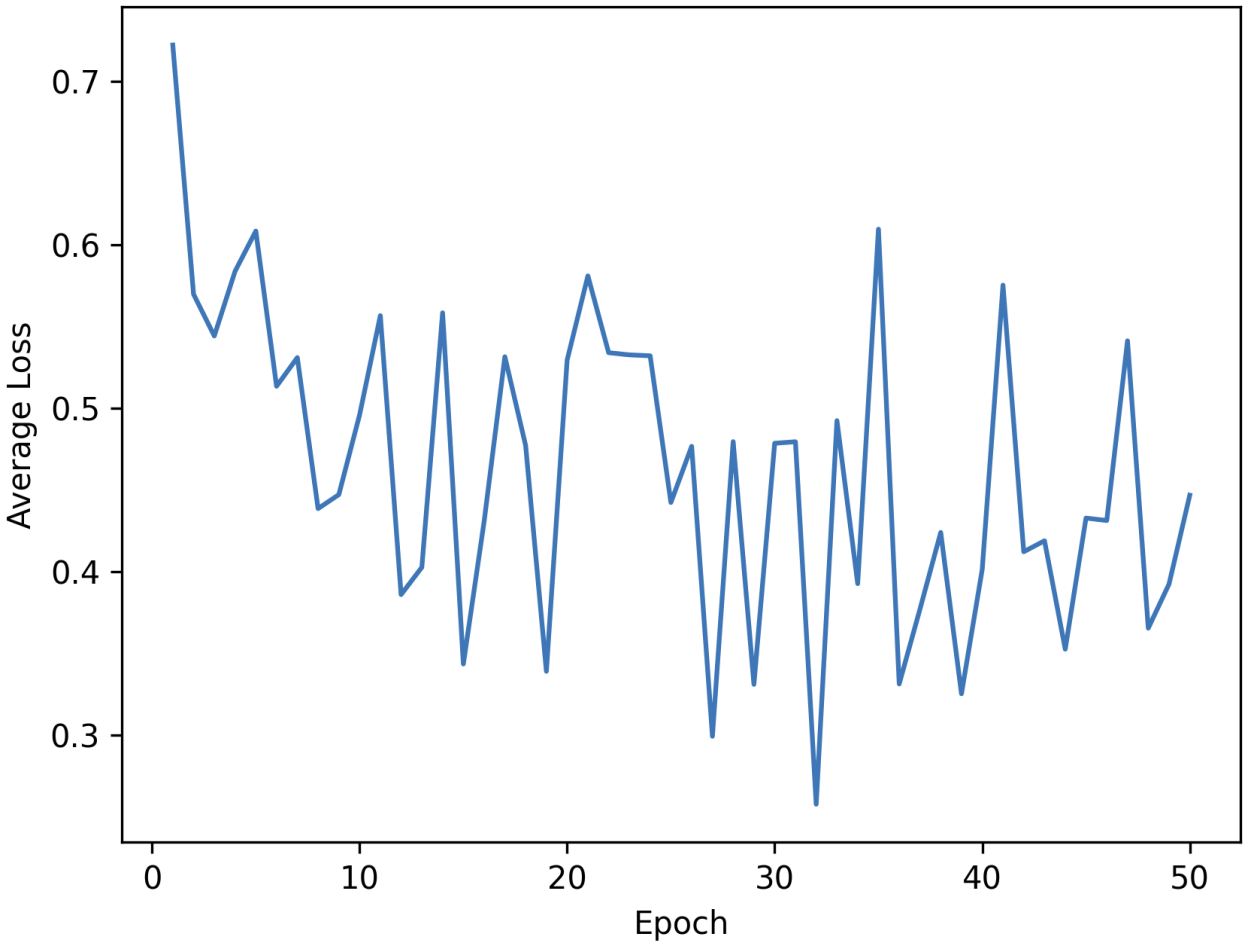


Figure C.5: Epoch-wise average training accuracy for the CNN–LSTM model using the min pooling down-sampling technique.

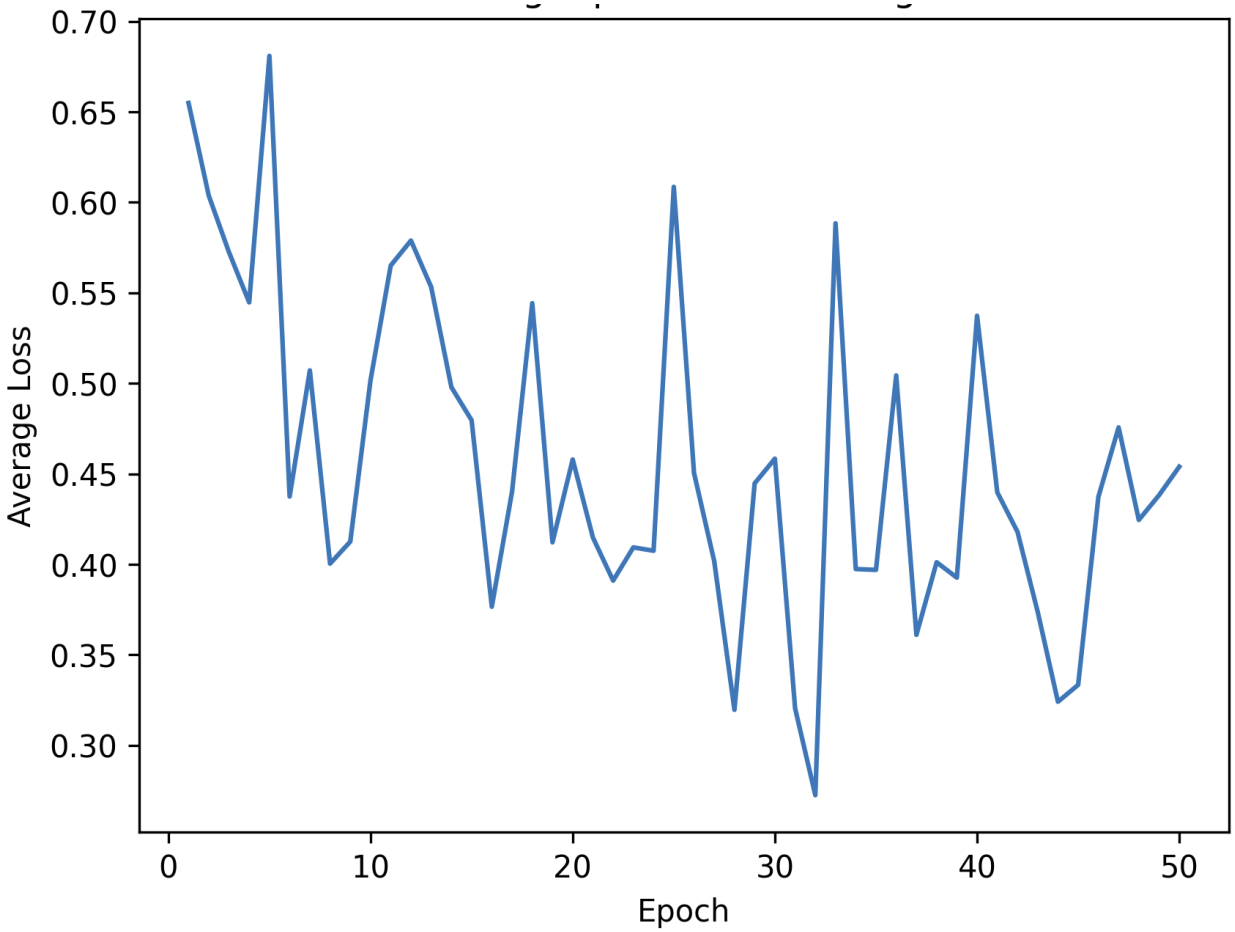


Figure C.6: Epoch-wise average cross-entropy training loss for the CNN–LSTM model using the min pooling down-sampling technique.

Appendix C.4: Odd Decimation Training Logs

As illustrated in Figures C.7 and C.8, the odd decimation approach exhibits balanced and consistent training behaviour. The accuracy curve increases steadily across epochs without abrupt changes, while the cross-entropy loss decreases gradually, confirming effective convergence. By retaining alternate temporal samples, odd decimation reduces redundancy while preserving essential temporal information, leading to smooth and reliable training dynamics.

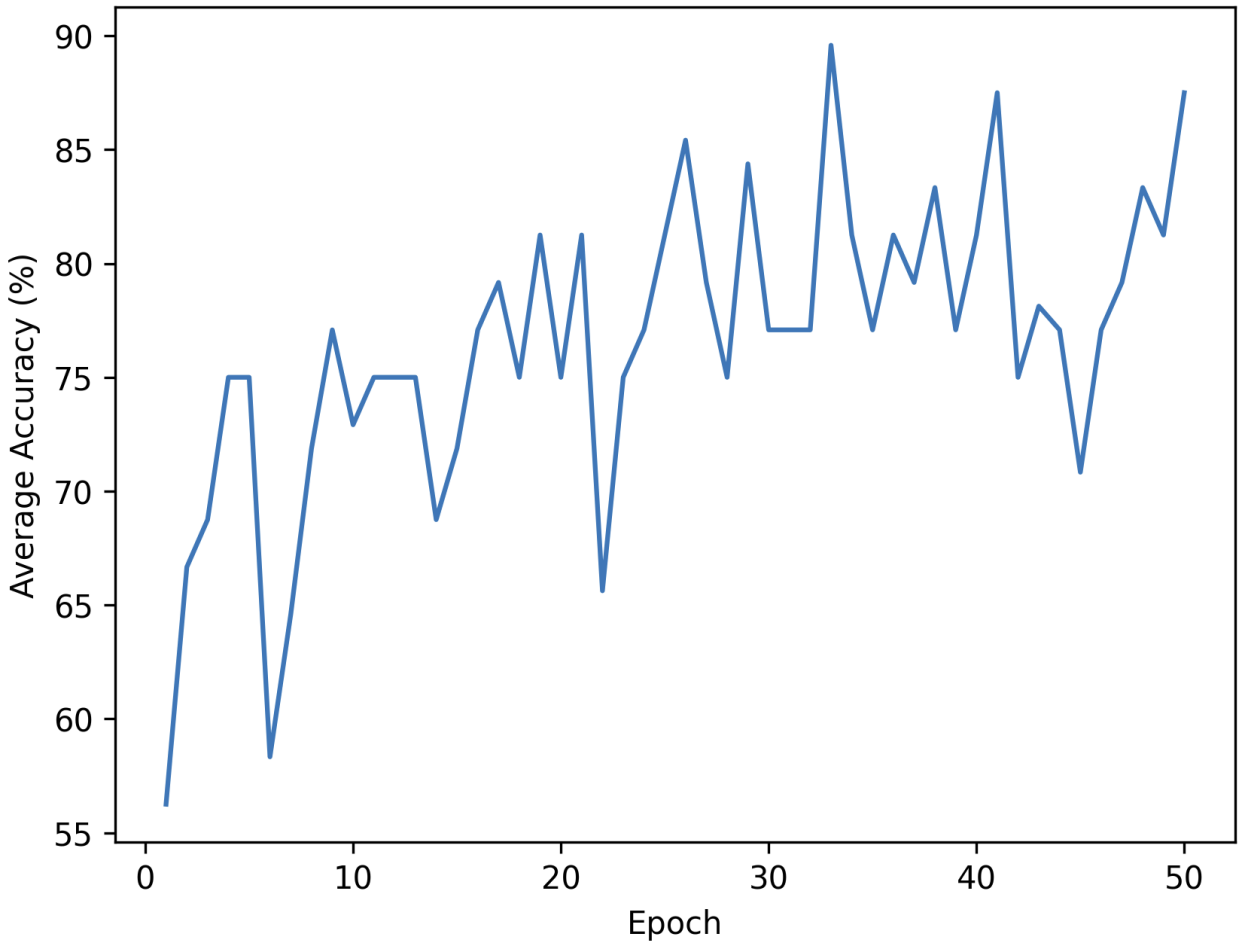


Figure C.7: Epoch-wise average training accuracy for the CNN–LSTM model using odd decimation.

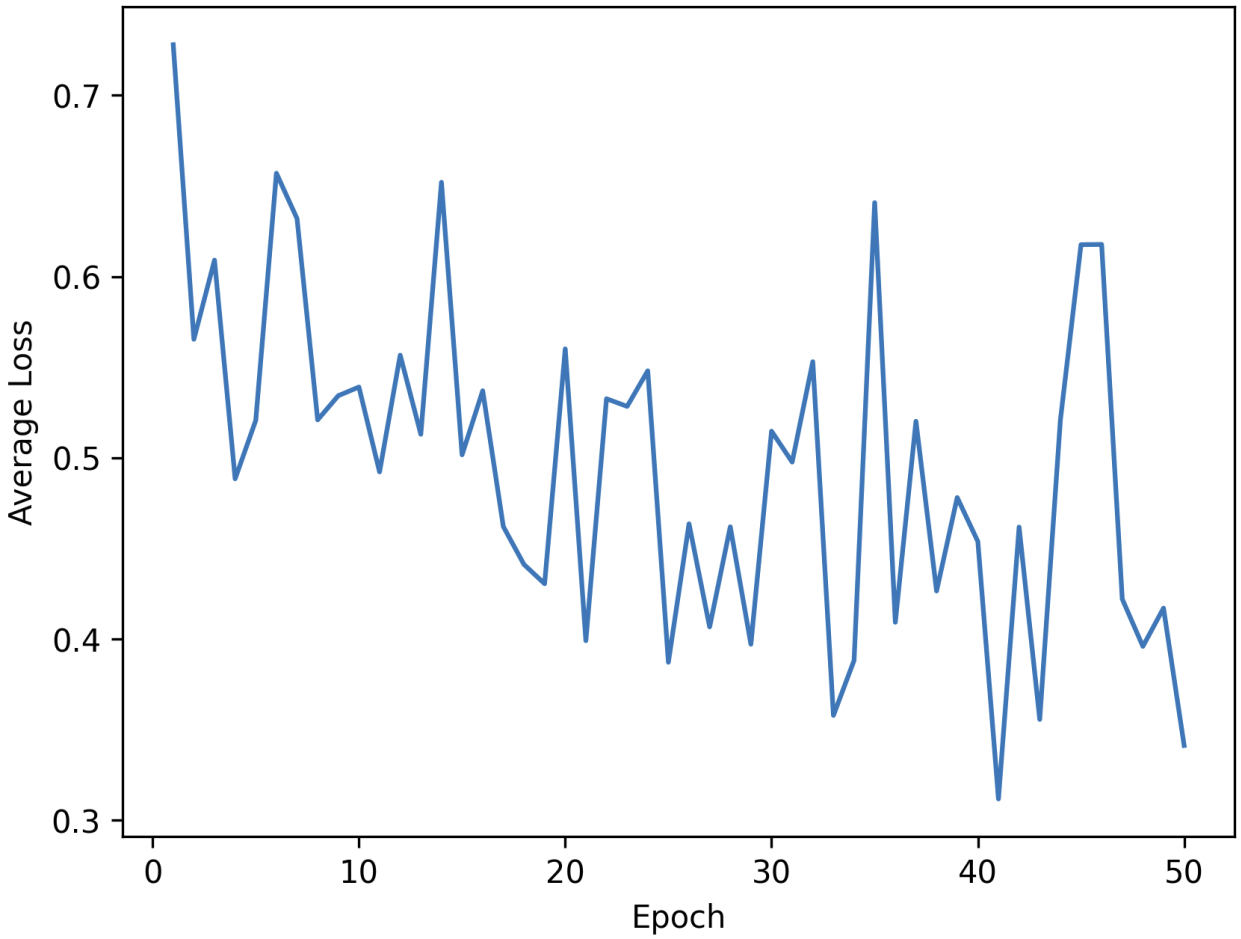


Figure C.8 Epoch-wise average cross-entropy training loss for the CNN–LSTM model using odd decimation.

Appendix C.5: Even Decimation Training Logs

As shown in Figures C.9 and C.10, the even decimation model shows training curves comparable to those of odd decimation, with a stable rise in accuracy and a monotonic decrease in cross-entropy loss. The smooth convergence observed across epochs suggests effective learning without instability. By selecting complementary temporal samples, even decimation captures alternative EEG dynamics and supports diversity within the overall modelling framework.

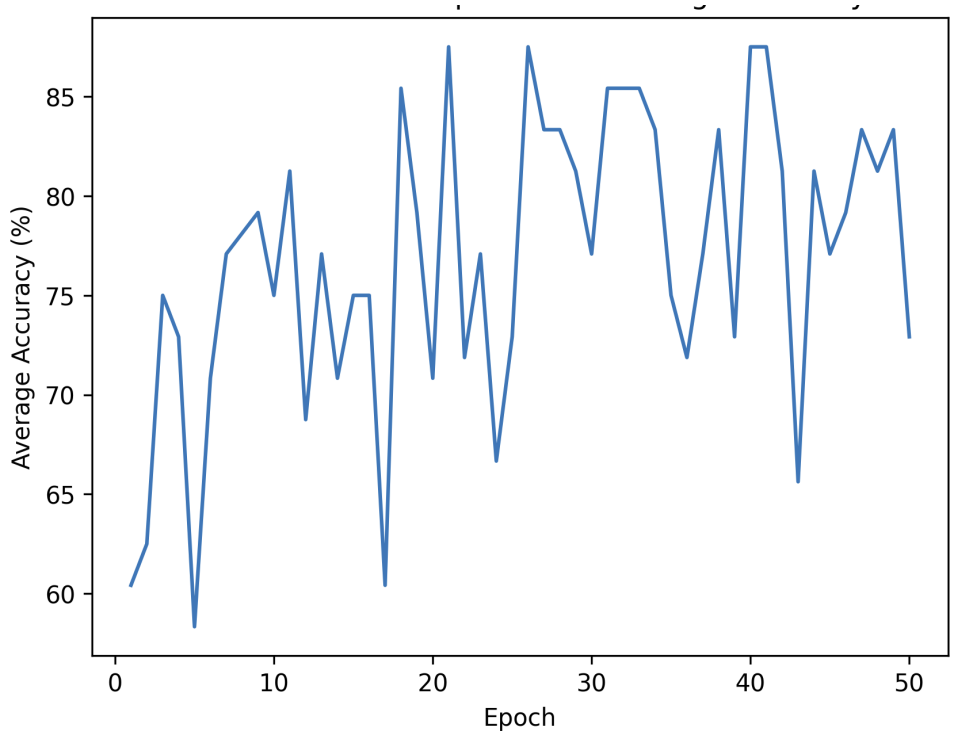


Figure C.9: Epoch-wise average training accuracy for the CNN–LSTM model using even decimation.

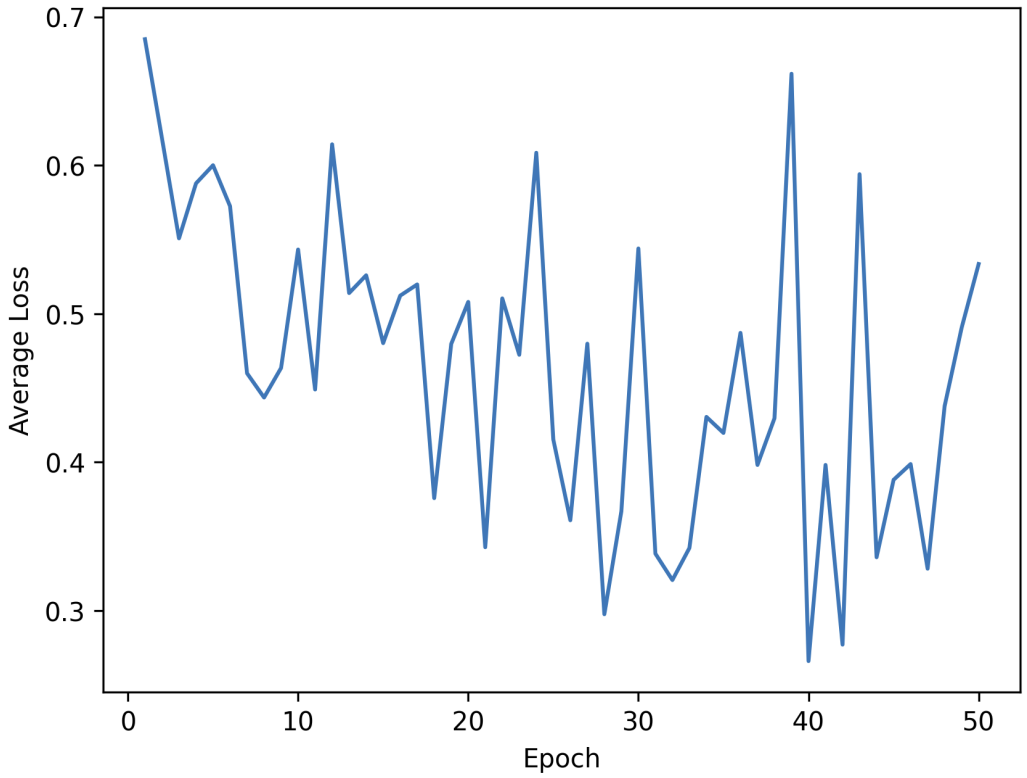


Figure C.10: Epoch-wise average cross-entropy training loss for the CNN–LSTM model using even decimation.

Appendix D: Detailed Cross-Validation Results of the Random Forest Based Stacking Ensemble.

Table 1: Accuracy, sensitivity, specificity, and AUC values for each fold of the Random Forest–based stacking ensemble evaluated using 10-fold cross-validation

Fold	Sensitivity	Specificity	Precision	F1-score	Balanced Accuracy	AUC
1	1	0.7143	0.7647	0.8667	0.8571	0.8407
2	1	0.4	0.5909	0.7429	0.7	0.7897
3	0.8462	0.7333	0.7333	0.7857	0.7897	0.8077
4	0.8462	0.8	0.7857	0.8148	0.8231	0.8282
5	1	0.5333	0.65	0.7879	0.7667	0.841
6	1	0.8	0.8	0.8889	0.9	0.9972
7	0.75	0.8	0.75	0.75	0.775	0.7639
8	0.9167	0.6667	0.6875	0.7857	0.7917	0.875
9	0.9167	1	1	0.9565	0.9583	0.9722
10	0.9231	0.7143	0.75	0.8276	0.8187	0.8462

References

Aarabi, A., Wallois, F. & Grebe, R. (2006) “Automated neonatal seizure detection via multistage classification through feature selection”, Clinical Neurophysiology, 117(2), pp.328–340.

Abdelfattah, S.M., Abdelrahman, G.M. & Wang, M. (2022) “Deep learning in EEG: A survey of recent advances and challenges”, IEEE Access, 10, pp.36219–36244.

Abdelhameed, A. & Bayoumi, M. (2021) “A deep learning approach for automatic seizure detection in children”, Frontiers in Computational Neuroscience, 15, p.650050.

Alom, M.Z. et al. (2019) “ state-of-the-art survey on deep learning theory and architectures”, Electronics, 8(3), p.292.

Almustafa, K.M. (2020) “Classification of epileptic seizure dataset using different machine learning algorithms”, Informatics in Medicine Unlocked, 21, p.100444.

Altaheri, H. et al. (2021) “Deep learning techniques for classification of EEG motor imagery signals: A review”, Neural Computing and Applications, pp.1–42.

Alturki, F.A., AlSharabi, K., Abdurraqeeb, A.M. & Aljalal, M. (2020) “EEG signal analysis for diagnosing neurological disorders using discrete wavelet transform and intelligent techniques”, Sensors, 20(9), p.2505.

Amer, N.S. & Belhaouari, S.B. (2023) “EEG signal processing for medical diagnosis, healthcare, and monitoring: A comprehensive review”, IEEE Access, 11, pp.143116–143142.

Amin, H.U. et al. (2017) “Classification of EEG signals via pattern recognition’, Frontiers in Computational Neuroscience, 11, p.103.

Andrzejak, R.G. et al. (2001) ‘Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state’, Physical Review E, 64(6), p.061907.

Antoniou, E. et al. (2021) ‘EEG-based eye movement recognition using brain–computer interfaces and Random Forests’, Sensors, 21(7), p.2339.

Antony, M.J. et al. (2022) ‘Classification of EEG using adaptive SVM classifier with CSP and ORICA’, Sensors, 22(19), p.7596.

Artoni, F., Delorme, A. & Makeig, S. (2018) ‘Applying dimension reduction to EEG by PCA reduces the quality of its subsequent ICA decomposition’, NeuroImage, 175, pp.176–187.

Atangana, R. et al. (2020) ‘EEG signal classification using LDA and MLP classifier’, Health Informatics International Journal, 9(1), pp.14–32.

Aviles, M., Sánchez-Reyes, L.M., Álvarez-Alvarado, J.M. & Rodríguez-Reséndiz, J. (2024) ‘Machine and Deep Learning Trends in EEG-Based Detection and Diagnosis of Alzheimer’s Disease: A Systematic Review’, Eng., 5(3), pp.1464–1484.

Bablani, A., Edla, D.R., Kuppili, V. & Ramesh, D. (2020) 'Multi-stage EEG classification using k-means and FFNN', Clinical Epidemiology and Global Health, 8(3), pp.718–724.

Bhardwaj, S. & Kumar, A. (2024) 'A systematic review of EEG based automated schizophrenia diagnosis using AI techniques', Frontiers in Human Neuroscience, 18, p.1347082.

Bari, M.F. & Fattah, S.A. (2020) 'Epileptic seizure detection in EEG using normalized IMFs in CEEMDAN domain and quadratic discriminant classifier', Biomedical Signal Processing and Control, 58, p.101833.

Bhatti, M.H. et al. (2019) 'Soft computing-based EEG classification by optimal feature selection and neural networks', IEEE Transactions on Industrial Informatics, 15(10), pp.5747–5754.

Bhuvaneswari, P. & Kumar, J.S. (2013) 'Support vector machine technique for EEG signals', International Journal of Computer Applications, 63(13).

Bickford, R.D. & Adelman, G. (1897) Encyclopedia of Neuroscience.

Bilal, M. et al. (2019) 'Automatic seizure detection using multi-resolution dynamic mode decomposition', IEEE Access, 7, pp.61180–61194.

Bugli, C. & Lambert, P. (2007) 'Comparison between PCA and ICA in EEG modelling', Biometrical Journal, 49(2), pp.312–327.

Bunterngchit, C., Baniata, L.H., Albayati, H., Baniata, M.H., Alharbi, K., Alshammari, F.H. & Kang, S., 2025. A hybrid convolutional–transformer approach for accurate electroencephalography (EEG)-based Parkinson's disease detection. Bioengineering, 12, 583. https://doi.org/10.3390/bioengineering12060583

Chaovalit, P. et al. (2011) 'DWT-based time series analysis and mining', ACM Computing Surveys, 43(2), pp.1–37.

Chandaka, S., Chatterjee, A. & Munshi, S. (2009) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', Expert Systems with Applications, 36(2), pp.1329–1336.

Chaudhary, U., 2025a. Invasive brain signal acquisition techniques: SUA, MUA, LFP, ECoG in neuroimaging and recording. In Expanding Senses Using Neurotechnology. Springer Nature Switzerland, Cham, pp. 81–118. https://doi.org/10.1007/978-3-031-76081-5_3.

Chaudhary, U., 2025b. Non-invasive brain signal acquisition techniques: Exploring EEG, EOG, fNIRS, fMRI, MEG, and fUS. In Expanding Senses Using Neurotechnology. Springer Nature Switzerland, Cham, pp. 25–80. https://doi.org/10.1007/978-3-031-76081-5_2.

Chekhmane, G. & Benali, R., 2025. Epileptic seizure detection in EEG signals using deep learning: LSTM and bidirectional LSTM. Computer Methods in Biomechanics and Biomedical Engineering, pp. 1–24. https://doi.org/10.1080/10255842.2025.2490136.

Chiang, C.Y. et al. (2011) 'Seizure prediction using EEG synchronisation patterns', in IEEE EMBC 2011. IEEE, pp.7564–7569.

Choi, W. et al. (2022) 'Deep convolutional GRU with attention for pre-ictal EEG classification', Journal of Personalized Medicine, 12(5), p.763.

Craik, A., He, Y. & Contreras-Vidal, J.L., 2019. Deep learning for electroencephalogram (EEG) classification tasks: a review. Journal of Neural Engineering, 16, 031001. https://doi.org/10.1088/1741-2552/ab0ab5.

Dişli, F., Gedikpınar, M., Fırat, H., Şengür, A., Güldemir, H. & Koundal, D., 2025. Epilepsy diagnosis from EEG signals using continuous wavelet transform-based depthwise convolutional neural network model. Diagnostics, 15, 84. https://doi.org/10.3390/diagnostics15010084.

Doll, T., Stieglitz, T., Heumann, A.S. and Wójcik, D.K., 2024. A case study on EEG signal correlation towards potential epileptic foci triangulation. Sensors, 24(24), p.8116. https://doi.org/10.3390/s24248116

Duan, L. et al. (2015) 'EEG feature selection method based on decision tree', Bio-Medical Materials and Engineering, 26(s1), pp.S1019–S1025.

Fathima, T. & Bedeeuzzaman, M. (2016) 'Epileptic seizure prediction in scalp EEG using one-dimensional Local Binary Pattern based features', in International Conference on Bio-inspired Systems and Signal Processing. SCITEPRESS, 5, pp.25–33.

Fergus, P. et al. (2016) 'A machine learning system for automated whole-brain seizure detection', Applied Computing and Informatics, 12(1), pp.70–89.

Fisher, R.A. (1936) 'The use of multiple measurements in taxonomic problems', Annals of Eugenics, 7(2), pp.179–188.

Fiscon, G. et al. (2018) 'Combining EEG signal processing with supervised methods for Alzheimer's classification', BMC Medical Informatics and Decision Making, 18(1), p.35.

Gadgil, A.A., Selvakumar, P., Gnanaselvi, G.S. & Malathi, G. (2025) 'AI in Neuroimaging and Brain Analysis', in Transforming Neuropsychology and Cognitive Psychology with AI and Machine Learning. IGI Global Scientific Publishing, pp.185–212.

García-Laencina, P.J., Rodríguez-Bermúdez, G. & Roca-Dorda, J. (2014) 'Exploring dimensionality reduction of EEG features in motor imagery task classification', Expert Systems with Applications, 41(11), pp.5285–5295.

Golmohammadi, M. et al. (2019) 'Automatic EEG analysis via hybrid deep learning', Frontiers in Human Neuroscience, 13, p.76.

Guerrero, M.C., Parada, J.S. & Espitia, H.E. (2021) 'EEG signal analysis using Logistic Regression, Artificial Neural Networks, Support Vector Machines, and Convolutional Neural Networks', Heliyon, 7(6), e07258.

Halford, J.J. et al. (2015) 'Inter-rater agreement on identification of electrographic seizures and periodic discharges in ICU EEG recordings', Clinical Neurophysiology, 126(9), pp.1661–1669.

Harati, A., Choi, S., Tabrizi, M., Obeid, I., Picone, J. & Jacobson, M.P., 2013. The Temple University Hospital EEG corpus. In 2013 IEEE Global Conference on Signal and Information Processing. IEEE, Austin, TX, pp. 29–32. https://doi.org/10.1109/GlobalSIP.2013.6736803

Hariharan, M. et al. (2014) ‘Classification of mental tasks using Stockwell transform’, Computers & Electrical Engineering, 40(5), pp.1741–1749.

Hasan, M.R., Ibrahimy, M.I., Motakabber, S.M. & Shahid, S. (2015) ‘Classification of multichannel EEG signal by linear discriminant analysis’, in Progress in Systems Engineering. Springer, pp.279–282.

He, H., Wu, D. & Wu, Z. (2023) ‘EEG-based classification of individuals with neuropsychiatric disorders: A review of deep neural network approaches’, Biomedical Signal Processing and Control, 83, p.104619.

Hendel, M., Benyettou, A. & Hendel, F. (2016) ‘Hybrid self-organizing map and probabilistic quadratic-loss SVM for mental task classification’, Informatics in Medicine Unlocked, 4, pp.1–9.

Hindarto, H., Muntasa, A. & Sumarno, S. (2018) ‘Feature extraction of EEG using wavelet transform for cursor movement’, IOP Conference Series: Materials Science and Engineering, 434, 012261.

Hu, W. et al. (2025) ‘Modified transformer network for seizure detection’, International Journal of Neural Systems, p.2550003.

Huan, N.J. & Palaniappan, R. (2004) ‘Neural network classification of autoregressive features from EEG signals for BCI design’, Journal of Neural Engineering, 1(3), pp.142–150.

Huff, A.J., Park, J., Montero-Hernandez, S., Park, L., Lee, C., Pollonini, L. & Ahn, H. (2025) ‘Functional near-infrared spectroscopy (fNIRS) detects brain changes for apathy and pain in patients with Alzheimer’s disease and related dementias: An exploratory study’, NeuroImage: Reports, 5(3), p.100266.

Hussein, R. (2016) Scalp and intracranial EEG quantitative analysis. PhD thesis. University of British Columbia.

Iešmantas, T. & Alzbutas, R. (2020) ‘CNN for seizure detection in clinical data’, Medical & Biological Engineering & Computing, 58(9), pp.1919–1932.

Ihle, M. et al. (2012) ‘EPILEPSIAE – A European epilepsy database’, Computer Methods and Programs in Biomedicine, 106(3), pp.127–138.

Joshi, V. & Nanavati, N., 2022. A review of EEG signal analysis for diagnosis of neurological disorders using machine learning. Journal of Biomedical Photonics & Engineering, 7. https://doi.org/10.18287/JBPE21.07.040201.

Jui, S.J.J., Deo, R.C., Barua, P.D., Devi, A., Soar, J. & Acharya, U.R., 2023. Application of entropy for automated detection of neurological disorders with electroencephalogram signals: A review of the last decade (2012–2022). IEEE Access, 11, pp. 71905–71924. https://doi.org/10.1109/ACCESS.2023.3294473.

Kannathal, N., Choo, M.L., Acharya, U.R. & Sadasivan, P.K. (2005) ‘Entropies for detection of epilepsy in EEG’, Computer Methods and Programs in Biomedicine, 80(3), pp.187–194.

Ke, H. et al. (2018) ‘Towards brain big data classification: Epileptic EEG identification with a lightweight VGGNet on global MIC’, IEEE Access, 6, pp.14722–14733.

Khan, H.A., Ul Ain, R., Kamboh, A.M., Butt, H.T., Shafait, S., Alamgir, W., Stricker, D. & Shafait, F. (2022) ‘The NMT scalp EEG dataset: An open-source annotated dataset of healthy and pathological EEG recordings for predictive modelling’, Frontiers in Neuroscience, 15, p.755817.

Khan, Y.U., Rafiuddin, N. & Farooq, O. (2012) 'Automated seizure detection in scalp EEG using multiple wavelet scales', in IEEE International Conference on Signal Processing, Computing and Control. IEEE, pp.1–5.

Kim, S.-K., Kim, J.B., Kim, H., Kim, L. & Kim, S.H., 2025. Early diagnosis of Alzheimer's disease using EEGConformer and attention-based LSTM during the short question task. Diagnostics, 15, 448. https://doi.org/10.3390/diagnostics15040448.

Kim, M.J., Youn, Y.C. & Paik, J. (2023) 'Deep learning-based EEG analysis to classify normal, mild cognitive impairment, and dementia: Algorithms and dataset', NeuroImage, 272, p.120054.

León, J. et al. (2021) 'RNNs for high-dimensional EEG classification', in Bioengineering and Biomedical Signal and Image Processing. Springer, pp.297–310.

Lemoine, É. et al. (2024) 'Computer-assisted analysis of routine EEG to identify hidden biomarkers of epilepsy: A systematic review', Computational and Structural Biotechnology Journal, 24, pp.66–86.

Li, Y. et al. (2020) 'Automatic seizure detection using fully convolutional nested LSTM', International Journal of Neural Systems, 30(4), p.2050019.

L.S. (2017) Automated identification of abnormal EEGs. MSc thesis. Temple University.

Manafian, J., Fazli, M., Ilhan, O., Zeynalli, S.M. & Tuama Ghafel, S., 2025. An ensemble fuzzy-based deep learning framework for automatic detection of children with ADHD from EEG signal. Journal of Artificial Intelligence Systems Modeling. https://doi.org/10.22034/jaism.2025.504722.1105.

Mandal, P.K., Banerjee, A., Tripathi, M. & Sharma, A. (2018) 'A comprehensive review of magnetoencephalography (MEG) studies for brain functionality in healthy aging and Alzheimer's disease (AD)', Frontiers in Computational Neuroscience, 12, p.60.

Martišius, I. (2016) Data acquisition and signal processing methods for brain–computer interfaces. PhD thesis. Kaunas University of Technology.

Matin, A. et al. (2019) 'A hybrid scheme using PCA and ICA based statistical feature for epileptic seizure recognition from EEG signal', in ICIEV/ICIVPR 2019. IEEE, pp.301–306.

Meitzler, T., Bryk, D., Sohn, E. & Hirsch, J. (n.d.) "fMRI for functional localization and task difficulty assessment during visual search for military vehicles". Unpublished.

Mhanna, H.Y.A. et al. (2025) 'Systematic review of functional magnetic resonance imaging (fMRI) applications in the preoperative planning and treatment assessment of brain tumors', Heliyon, 11(3).

Mohan, R. & Perumal, S. (2023) 'Classification and Detection of Cognitive Disorders like Depression and Anxiety Utilizing Deep Convolutional Neural Network (CNN) Centered on EEG Signal', Traitement Du Signal, 40(3).

MohiudDin, Q. & Jayanthy, A.K. (2021) 'Autism Spectrum Disorder classification using EEG and 1D-CNN', in 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), pp.01–05. IEEE.

Moon, H., Kwon, J., Eun, J., Chung, C.K., Kim, J.S., Chou, N. & Kim, S., 2024. Electrocorticogram (ECoG): Engineering approaches and clinical challenges for translational medicine. Advanced Materials Technologies, 9, 2301692. https://doi.org/10.1002/admt.202301692.

Najafi, T., Jaafar, R., Remli, R. & Wan Zaidi, W.A. (2022) 'A classification model of EEG signals based on RNN-LSTM for diagnosing focal and generalized epilepsy', Sensors, 22(19), p.7269.

Narang, A., Batra, B., Ahuja, A. & Yadav, J. (2018) 'Classification of EEG signals for epileptic seizures using Levenberg–Marquardt-based MLP neural network', Journal of Intelligent & Fuzzy Systems, 34(3), pp.1669–1677.

National Institute of Mental Health (HHS) (2009) MEG scanner with patient [image]. Public domain. Available at: https://commons.wikimedia.org/wiki/File:NIMH_MEG.jpg

Naseer, N. & Hong, K.S., 2015. fNIRS-based brain–computer interfaces: A review. Frontiers in Human Neuroscience, 9, 3.

Nasehi, S. & Pourghassem, H. (2013) 'Patient-specific epileptic seizure onset detection based on spectral features and IPSONN classifier', in International Conference on Communication Systems and Network Technologies. IEEE, pp.186–190.

Nicolaou, N. & Georgiou, J. (2012) 'Detection of epileptic EEG based on permutation entropy and SVM', Expert Systems with Applications, 39(1), pp.202–209.

Nigam, V.P. & Graupe, D. (2004) 'Neural-network-based detection of epilepsy', Neurological Research, 26(1), pp.55–60.

Nour, M., Senturk, U. & Polat, K. (2024) 'A novel hybrid model in the diagnosis and classification of Alzheimer's disease using EEG signals: Deep ensemble learning (DEL) approach', Biomedical Signal Processing and Control, 89, p.105751.

Nunes, T.M. et al. (2014) 'EEG signal classification for epilepsy diagnosis via optimum path forest', Neurocomputing, 136, pp.103–123.

Obeid, I. & Picone, J. (2016) 'The Temple University Hospital EEG data corpus', Frontiers in Neuroscience, 10, p.196.

Obeid, I. & Picone, J., 2018. Machine learning approaches to automatic interpretation of EEGs. In Sejdić, E. & Falk, T.H. (eds.) Signal Processing and Machine Learning for Biomedical Big Data. CRC Press, Boca Raton, pp. 271–300. https://doi.org/10.1201/9781351061223.

Ocak, H. (2009) 'Automatic seizure detection using wavelets and entropy', Expert Systems with Applications, 36(2), pp.2027–2036.

Orosco, L., Correa, A.G., Diez, P. & Laciar, E. (2016) 'Patient non-specific algorithm for seizure detection in scalp EEG', Computers in Biology and Medicine, 71, pp.128–134.

Ortega-Flores, G., Altamirano-Escobedo, G., Mercado-Ravell, D. and Bayro-Corrochano, E. (2025) Quaternion CNN in deep learning processing for EEG with applications to brain disease detection. Applied Sciences, 15(21), p. 11526.

Phutela, N. et al. (2022) ‘Stress classification using LSTM on EEG’, Computational Intelligence and Neuroscience, 2022.

Polat, K. & Güneş, S. (2007) ‘Hybrid decision tree–FFT system for epileptiform EEG classification’, Applied Mathematics and Computation, 187(2), pp.1017–1026.

Qiu, X., Yan, F. & Liu, H. (2023) ‘A difference attention ResNet-LSTM network for epileptic seizure detection using EEG signal’, Biomedical Signal Processing and Control, 83, p.104652.

Rahul, J., Sharma, D., Sharma, L.D., Nanda, U. & Sarkar, A.K. (2024) ‘A systematic review of EEG-based automated schizophrenia classification through machine learning and deep learning’, Frontiers in Human Neuroscience, 18, p.1347082.

Rasekhi, J. et al. (2013) ‘Preprocessing effects of univariate features on seizure prediction’, Journal of Neuroscience Methods, 217(1–2), pp.9–16.

Rathod, P., Naik, S. & Bhalodiya, J., 2025. Epileptic signal classification with CNN models and explainable AI. In 2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM). pp. 1–7. https://doi.org/10.1109/IMCOM64595.2025.10857540

Rejer, I. & Górski, P. (2015) ‘Benefits of ICA in the case of a few-channel EEG’, in EMBC 2015. IEEE, pp.7434–7437.

Rivas-Carrillo, S.D., Akkuratov, E.E., Valdez Ruvalcaba, H., Vargas-Sanchez, A., Komorowski, J., San-Juan, D. & Grabherr, M.G., 2023. MindReader: Unsupervised classification of electroencephalographic data. Sensors, 23, 2971. https://doi.org/10.3390/s23062971

Roy, S., Kiral-Kornek, I. & Harrer, S. (2018) ‘Deep learning enabled automatic abnormal EEG identification’, in 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp.2756–2759.

Sadati, N., Mohseni, H.R. & Maghsoudi, A. (2006) ‘Epileptic seizure detection using neural fuzzy networks’, in IEEE International Conference on Fuzzy Systems. IEEE, pp.596–600.

Sairamya, N.J., Subathra, M.S. & George, S.T. (2022) ‘Automatic identification of schizophrenia using EEG signals based on discrete wavelet transform and RLNDiP technique with ANN’, Expert Systems with Applications, 192, p.116230.

Saminu, S. et al. (2023) ‘Applications of artificial intelligence in automatic detection of epileptic seizures using EEG signals: A review’, in Artificial Intelligence and Applications, 1(1), pp.11–25.

Satam, I.A. (2023) ‘Safety and security aspects of implanted brain–computer interface (BCI) for human–robot interaction’, International Journal of Multidisciplinary Research and Publications (IJMRAP), 6(6), pp.268–276.

Saneesh Cleatus, T. & Thungamani, M. (2021) ‘Classification of EEG signals using nonlinear features and preprocessing techniques’, International Journal of Engineering and Advanced Technology (IJEAT), 10(5).

Schirrmeister, R.T. et al. (2017) ‘Deep learning with CNNs for EEG pathology’, arXiv preprint, arXiv:1708.08012.

Selvaraj, T.G., Ramasamy, B., Jeyaraj, S.J. & Suviseshamuthu, E.S. (2014) 'EEG database of seizure disorders for experts and application developers', Clinical EEG and Neuroscience, 45(4), pp.304–309.

Sha'Abani, M.N., Fuad, N., Jamal, N. & Ismail, M.F. (2020) 'KNN and SVM classification for EEG: a review', in ECCE 2019, pp.555–565.

Shang, S., Shi, Y., Zhang, Y., Liu, M., Zhang, H., Wang, P. & Zhuang, L., 2024. Artificial intelligence for brain disease diagnosis using electroencephalogram signals. Journal of Zhejiang University – Science B, 25, pp. 914–940. https://doi.org/10.1631/jzus.B2400103

Sharmila, A. & Geethanjali, P.J. (2016) 'DWT-based detection of epileptic seizure from EEG signals using Naïve Bayes and KNN classifiers', IEEE Access, 4, pp.7716–7727.

Sharma, G., Parashar, A. & Joshi, A.M. (2021) 'DepHNN for EEG-based depression screening', Biomedical Signal Processing and Control, 66, p.102393.

Sharma, M., Patel, S. & Acharya, U.R. (2020) 'Automated abnormal EEG detection using wavelet filter banks', Pattern Recognition Letters, 133, pp.188–194.

Siddiqui, S.T., 2025. Hybrid CNN-RNN deep learning framework for EEG-based mental health disorder diagnosis with explainable AI. In 2025 International Conference on Intelligent and Cloud Computing (ICoICC). pp. 1–6. https://doi.org/10.1109/ICoICC64033.2025.11052036

Shoeb, A.H. (2009) Application of machine learning to epileptic seizure onset detection and treatment. PhD thesis. Massachusetts Institute of Technology.

Smith, J.A., Chen, L. & Davis, R. (2023) 'Exploring the depths of the mind: A comprehensive overview of brain imaging techniques', Journal of Clinical Neuroscience, 45(2), pp.12–25.

Stefanou, K., Tzimourta, K.D., Bellos, C., Stergios, G., Markoglou, K., Gionanidis, E., Tsipouras, M.G., Giannakeas, N., Tzallas, A.T. & Miltiadous, A., 2025. A novel CNN-based framework for Alzheimer's disease detection using EEG spectrogram representations. Journal of Personalized Medicine, 15, 27. https://doi.org/10.3390/jpm15010027.

Stevenson, N.J., Tapani, K., Lauronen, L. & Vanhatalo, S. (2019) 'A dataset of neonatal EEG recordings with seizure annotations', Scientific Data, 6(1), pp.1–8.

Subasi, A. (2007) 'EEG signal classification using wavelet feature extraction and a mixture of expert models', Expert Systems with Applications, 32(4), pp.1084–1093.

Subasi, A. & Gursoy, M.I. (2010) 'EEG signal classification using PCA, ICA, LDA and support vector machines', Expert Systems with Applications, 37(12), pp.8659–8666.

Tassinari, C.A. (2019) 'An electroencephalographer recalls the history of the Federation on the 70th anniversary of its journal, Clinical Neurophysiology', Clinical Neurophysiology, 130(12), pp.2258–2263.

Teplan, M. (2002) 'Fundamentals of EEG measurement', Measurement Science Review, 2(2), pp.1–11.

Tharwat, A. (2020) 'Independent component analysis: An introduction', Applied Computing and Informatics.

Thieu, T.N. & Yang, H.J. (2015) 'Diagnosis of epilepsy based on EEG signal classification using fast Fourier transform', in Industrial, Engineering and Other Applications of Applied Intelligent Systems. Springer, pp.493–500.

Trinka, E. & Leitinger, M. (2015) 'Which EEG patterns in coma are nonconvulsive status epilepticus?', Epilepsy & Behavior, 49, pp.203–222.

Todaro, C., Marzetti, L., Valdés Sosa, P.A., Valdés-Hernandez, P.A. & Pizzella, V., 2019. Mapping brain activity with electrocorticography: Resolution properties and robustness of inverse solutions. Brain Topography, 32, pp. 583–598. https://doi.org/10.1007/s10548-018-0623-1

Truong, N.D. et al. (2018) 'CNNs for seizure prediction using EEG', Neural Networks, 105, pp.104–111.

Tuncer, T., Dogan, S. & Acharya, U.R. (2021) 'Automated EEG signal classification using Chaotic Local Binary Patterns (CLBP)', Expert Systems with Applications, 182, p.115175.

Tveit, J. et al. (2023) 'Automated interpretation of clinical electroencephalograms using artificial intelligence', JAMA Neurology, 80(8), pp.805–812.

Upadhyay, D. (2013) 'Classification of EEG signals under different mental tasks using wavelet transform and neural network with one-step secant algorithm', International Journal of Scientific Engineering and Technology, 2(4), pp.256–259.

Uyanik, H., Sengur, A., Salvi, M., Tan, R., Tan, J.H. & Acharya, U.R., 2025. Automated detection of neurological and mental health disorders using EEG signals and artificial intelligence: A systematic review. WIREs Data Mining and Knowledge Discovery, 15, e70002. https://doi.org/10.1002/widm.70002

Vega-Escobar, L. et al. (2015) 'Feature extraction schemes for BCI systems', in STSIVA 2015. IEEE, pp.1–6.

Vinutha, C.R., Subathra, M.S.P., George, S.T., Peter, G., Stonier, A.A., Sairamya, N.J., Prasanna, J. & Ganji, V., 2025. Automatic epilepsy seizure classification using EEG signals based on the CNN-LSTM model. IET Signal Processing, 2025, 7543401. https://doi.org/10.1049/sil2/7543401

Wong, S. et al. (2023) 'EEG datasets for seizure detection and prediction—A review', Epilepsia Open, 8(2), pp.252–267.

Wu, T., Kong, X., Zhong, Y. & Chen, L., 2022. Automatic detection of abnormal EEG signals using multiscale features with ensemble learning. Frontiers in Human Neuroscience, 16, 943258. https://doi.org/10.3389/fnhum.2022.943258

Xin, Q. et al. (2022) 'Attention-based wavelet CNN for epilepsy EEG', IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30, pp.957–966.

Vafaei, E. & Hosseini, M. (2025) 'Transformers in EEG analysis: A review of architectures and applications in motor imagery, seizure, and emotion classification', Sensors, 25(5), p.1293.

Xie, L., Zhao, J., Li, Y. & Bai, J. (2024) 'PET brain imaging in neurological disorders', Physics of Life Reviews, 49, pp.100–111.

Yan, J. et al. (2022) 'Seizure prediction using transformer models', Applied Sciences, 12(9).

Yap, H.Y., Choo, Y.H., Mohd Yusoh, Z.I. & Khoh, W.H. (2023) ‘An evaluation of transfer learning models in EEG-based authentication’, Brain Informatics, 10(1), p.19.

Yen, C., Lin, C.-L. & Chiang, M.-C., 2023. Exploring the frontiers of neuroimaging: A review of recent advances in understanding brain functioning and disorders. Life, 13, 1472. https://doi.org/10.3390/life13071472

Yıldırım, Ö., Baloglu, U.B. & Acharya, U.R. (2020) ‘Deep CNN for abnormal EEG detection’, Neural Computing and Applications, 32(20), pp.15857–15868.

Yuan, H., Li, Y., Yang, J., Li, H., Yang, Q., Guo, C., Zhu, S. & Shu, X. (2021) ‘State of the art of non-invasive electrode materials for brain–computer interface’, Micromachines, 12(12), p.1521.

Yuan, Q., Zhou, W., Li, S. & Cai, D. (2011) ‘Epileptic EEG classification using extreme learning machine’, Epilepsy Research, 96(1–2), pp.29–38.

Yuan, Y. et al. (2018) ‘Channel-aware attention for EEG seizure detection’, in IEEE BHI 2018. IEEE, pp.206–209.

Yun, S., 2024. Advances, challenges, and prospects of electroencephalography-based biomarkers for psychiatric disorders: a narrative review. Journal of Yeungnam Medical Science, 41, pp. 261–268. https://doi.org/10.12701/jyms.2024.00668.

Zandi, A.S. et al. (2013) ‘Seizure prediction via Bayesian GMM of zero-crossings’, IEEE Transactions on Biomedical Engineering, 60(5), pp.1401–1413.