AI

Article

A Combined Windowing and Deep Learning Model for the Classification of Brain Disorders Based on Electroencephalogram Signals

Dina Abooelzahab [1,*], Nawal Zaher [1], Abdel Hamid Soliman [2] and Claude Chibelushi [3]

[1] Arab Academy for Science and Technology and Maritime Transport, Cairo 2033, Egypt; nawalzaher@aast.edu
[2] School of Digital, Technologies and Arts, Staffordshire University, Stoke-on-Trent ST4 2DF, UK; a.soliman@staffs.ac.uk
[3] Unit Qa Beacon House, Semantics 21 Ltd., Beacon Business Park, Weston Road, Stafford ST18 0WL, UK
* Correspondence: dinaabouldahab@aast.edu

Abstract: Background: The electroencephalogram (EEG) is essential for diagnosing and classifying brain disorders, enabling early medical intervention. Its ability to identify brain abnormalities has increased its clinical use in assessing changes in brain activity. Recent advancements in deep learning have introduced effective methods for interpreting EEG signals, utilizing large datasets for enhanced accuracy. Objective: This study presents a deep learning-based model designed to classify EEG data with better accuracy compared to existing approaches. Methods: The model consists of three key components: data selection, feature extraction, and classification. Data selection employs a windowing technique, while the feature extraction and classification stages use a deep learning framework combining a convolutional neural network (CNN) and a Long Short-Term Memory (LSTM) network. The resulting architecture includes up to 18 layers. The model was evaluated using the Temple University Hospital (TUH) dataset, comprising data from 2785 patients, ensuring its applicability to real-world scenarios. Results: Comparative performance analysis shows that this approach surpasses existing methods in accuracy, sensitivity, and specificity. Conclusions: This study highlights the potential of deep learning in enhancing EEG signal interpretation, offering a pathway to more accurate and efficient diagnoses of brain disorders for clinical applications.

Keywords: EEG; CNN; LSTM; MWD; deep learning

Academic Editor: Ioannis Kakkos

Received: 30 December 2024
Revised: 30 January 2025
Accepted: 13 February 2025
Published: 20 February 2025

Citation: Abooelzahab, D.; Zaher, N.; Soliman, A.H.; Chibelushi, C. A Combined Windowing and Deep Learning Model for the Classification of Brain Disorders Based on Electroencephalogram Signals. *AI* **2025**, *6*, 42. https://doi.org/10.3390/ai6030042

1. Introduction

Electroencephalograms (EEGs) are analyzed by medical professionals at hospitals to identify brain illnesses. The great temporal resolution of EEG recordings, non-invasiveness, and relatively inexpensive equipment all add to the popularity of EEG data among medical professionals. However, manual analysis of EEG data is a time-consuming, resource-intensive, and expensive process, due to the rather large volume of data, the relatively high rate at which new data are produced, and the need for specialist expertise to analyze these data [1]. On the other hand, automated analysis has the potential to enhance patient care by speeding up the diagnosis process, minimizing human error, minimizing diagnostic variability caused by physician subjectivity, and automatically spotting potentially life-threatening conditions [2]. Recent work on the automated classification of brain disorders has highlighted the pivotal role of Artificial Intelligence (AI) [3,4], particularly machine learning [5] and deep learning [6], in extracting and processing relevant data to categorize brain illnesses. Deep learning architectures, such as Long Short-Term Memory (LSTM) networks [7], utilize a type of Recurrent Neural Network (RNN) particularly adept at handling sequential data. LSTM models are widely used for EEG signal classification tasks, such as seizure detection [7], motor imagery [8], and emotion recognition [9], due to their ability to capture temporal dependencies in brain activity patterns. Hybrid models combining LSTM with other techniques, like Convolutional Neural Networks (CNNs) [10] or Support Vector Machines (SVM), have also been explored to enhance classification accuracy by leveraging both spatial and temporal features of EEG signals. These AI techniques enable the interpretation of EEG data in discerning normal or abnormal brain function, facilitating the diagnosis of neurological disorders. For instance, deep learning models can analyze EEG wave patterns to detect specific conditions, while machine learning algorithms can refine the classification process by identifying key features in EEG signals. In [11], deep learning was used to develop an end-to-end automatic brain disorder detection system, which did not necessitate feature engineering or intensive EEG data preprocessing. CNN was employed in [8,12] to predict seizures, with minimum feature engineering. In addition, a method was proposed in the aforementioned work to convert raw EEG data into a format that could be applied to CNN. The criteria that were employed to create visualizations of brain activity included the phase-locking value, entropy, and energy, with the CNN serving as the primary emphasis [12]. A new method for diagnosing epilepsy using EEG signals is presented in the study by Dişli and Gedikpınar [13], which makes use of a Depthwise Convolutional Neural Network model based on Continuous Wavelet Transform (CWT). By efficiently extracting time–frequency information from EEG data, the suggested technique improves the detection accuracy of epilepsy. With encouraging results in differentiating between normal brain signals and epileptic activity, the study shows the model's potential as a trustworthy diagnostic tool. The authors of [14] offer a novel method for precisely identifying patients' interictal epileptiform discharges using transformer-based deep learning. This approach shows great promise for improving self-limited epilepsy diagnostic skills and offering a more effective and dependable substitute for conventional detection methods. According to the results, the transformer model can analyze EEG data efficiently, enhancing automated analysis and overall epilepsy care. Due to the skewed structure of data pertaining to seizure detection, the authors of [15] proposed an enhanced transformer network called Inresformer for seizure detection, which combines Inception and Residual networks to extract multi-scale features from electroencephalography (EEG) signals, thereby improving feature representation. To improve the model's nonlinear representation, the architecture uses two half-step Feedforward layers instead of typical Feedforward layers. Furthermore, it uses the discrete wavelet transform (DWT) to decompose EEG signals, selects three sub-bands for reconstruction, and uses the Co-Mix-up method to address data imbalance before processing the signals through the Inresformer network for effective seizure recognition, resulting in a discriminant fusion of the three EEG sub-signals [15]. Although various researchers have demonstrated the potential of deep neural networks approaches in offering a significant potential for improving the automated analysis of EEG signals, there remain many research challenges which need to be overcome before these networks can be widely adopted for use in clinical settings [16]. Among these challenges are the high variability between samples of the utilized data set, due to variations in the patients whose EEG data were under analysis, or in the physicians who performed the analysis, and the high dimensionality of the data, which is composed of multiple EEG electrodes, each carrying time series information. These factors result in several significant implications, which make classification more complicated [11].

The first of these implications is the high dimensionality of the data. A short EEG recording of a few seconds, using a single electrode, might contain up to 30,000 points. These samples are the source data from which discriminative features might be extracted and used for classification and diagnosis. Usually, recordings are longer than a few seconds and there might be multiple recordings per subject, using multiple electrodes. As a result of a large number of subjects, a large number of electrodes, multiple recordings per subject, and a fairly long-time span, the dimensionality curse arises and often presents a real problem [4]. The nature of the data itself, as a time series, is yet another problem. This implies that the value at a given point in the signal is not as significant as the temporal pattern of values sampled along the given signal. In addition, the information is distributed over multiple electrodes. The important information to distinguish between classes might not only lie in the patterns contained in one electrode, but also in the joint patterns of activity over multiple electrodes [10].

Another important challenge is methodological in nature, which is tied to insufficient diversity of EEG data used in research. Researchers who conducted investigations of deep neural networks in EEG classification used various input formulations and network designs [6,7]. An important caveat about the findings of the said investigations concerned the generalizability of their associated performance assessment results. Notably, most of the reported findings, for classifier performance, were obtained using what was called "good" EEG data [17], whereby EEG data were collected under controlled experimental conditions. This typically meant that data were recorded in a laboratory setting where brain disorders were frequently of the same type or occurred in groups of patients who were of similar ages [17].

To obtain more reliable and trustful results, it was necessary to find a dataset that contained more realistic and diverse data. The TUH EEG Corpus [18], being among the largest publicly accessible databases of clinical EEG data in the world, was deemed to fill the need, as the data were collected at different clinical conditions, spanning a variety of age groups, different types of brain disorders, and patient health conditions. Thus, this corpus offered a better platform for potentially generalizable results. Results demonstrated in [16], using this database, were given in terms of sensitivity of around 0.292 and specificity of around 0.667. Using the TUH dataset, a new deep one-dimensional convolutional neural network (1DCNN) model was created, to differentiate between normal and pathological behavior [19]. The model categorized the EEG signals as a full end-to-end structure, without the necessity for handcrafted feature extraction. The developed model achieved a 20.66% classification error rate, for differentiating between normal and pathological EEG data. However, there is still a need for more accurate classification techniques, in which more realistic datasets (such as TUH) are used in order to obtain more generalizable results.

The aim of this paper is to propose a more accurate EEG signal classification model for initial screening of brain disorders, as either "normal" or "abnormal". Such a model would help in saving medical resources, and aid in faster and more efficient diagnosis of critical cases. In this respect, a labelling model is adopted which comprises three sequentially connected modules: the first performs data selection, the second carried out feature extraction, and the last performs classification, yielding "normal" or "abnormal" output. The main contributions of this work are as follows:

- Proposing an end-to-end deep learning-based architecture and model for labelling EEG signals, involving the combination of a data selection module and a deep learning module.
- Incorporating a windowing technique in the architecture for data selection to overcome the dimensionality curse, using the Morse wavelet transform.
- Designing the deep learning module, comprising a CNN–LSTM combined network, for feature extraction and classification.

The rest of this paper is organized as follows. The methodology and architecture of the proposed model for EEG signal classification is presented in Section 2, with a detailed discussion of the proposed feature selection, extraction, and classification techniques. Section 3 presents the evaluation of the proposed model in classifying brain abnormalities. Finally, Section 4 concludes this article.

2. Methodology

In this paper a model is adopted for brain disorder classification to label EEG signals, as representing normal or abnormal brain function. The model comprises three building blocks, which sequentially process EEG signals: a feature selection technique, a feature extraction technique, and a classification technique, as illustrated in Figure 1 [7].

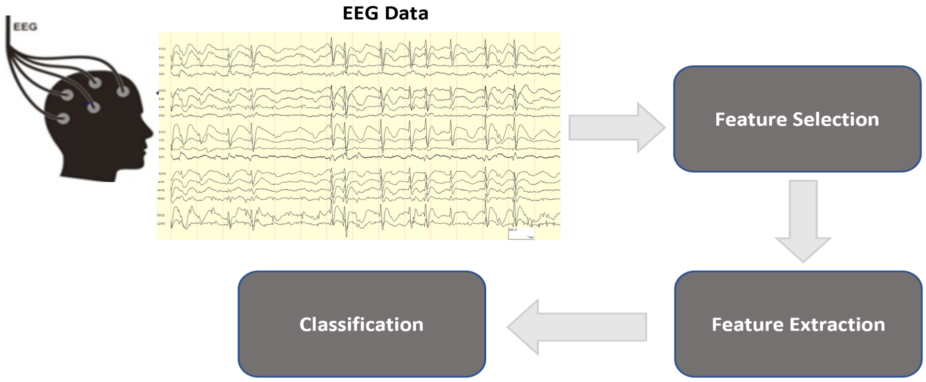


Figure 1. The structural layout of the EEG labelling model.

In the work reported herein, EEG brain signals are taken from a number of non-invasive electrodes on the skull, typically 19 electrodes [20]. These signals are analyzed and processed using alternative techniques, to select the most suitable. Different combinations of signal processing techniques and deep learning architectures are tested for feature selection, extraction and classification. The best combination is selected based on classification results.

2.1. Architecture of the Proposed Model

An architecture is proposed with two main components: one for data selection, and the other for both feature extraction and classification. In this respect, two approaches are adopted: introducing a windowing technique and combining CNN with LSTM networks. First, the windowing technique is applied to each EEG electrode, aiming to select the best time interval that contains the most significant data, related to distinctive features of abnormality. The data obtained within the selected window is then fed to a deep learning model containing a CNN [21] coupled to LSTM [22] network. CNN first extracts discriminative features from the data, and then it relays these features to the LSTM, to classify the EEG as “normal” or “abnormal”. CNN is selected for feature extraction because it extracts features at different scales without the need for hand-picking or manually crafting features. In addition, an LSTM is selected for the classification task, since it can recall crucial information over a fairly long period of time, along a temporal sequence of data. Figure 2 illustrates the layout of the proposed architecture.

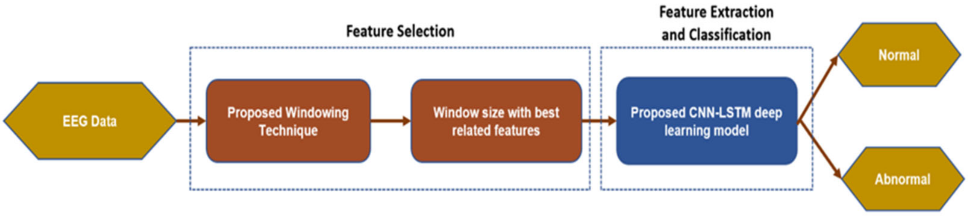


Figure 2. Layout of the proposed EEG labelling architecture.

This shows how EEG data are processed by the proposed windowing technique, followed by the proposed deep learning CNN–LSTM model, to give a decision as to whether the EEG recording for the patient is normal or abnormal.

2.1.1. TUH Data Set

Despite EEG's crucial role in brain signal acquisition, publicly available datasets are limited and vary significantly in the number of patients they encompass. For example, the CHB–MIT database includes recordings from 22 juvenile patients [23]. In contrast, the Karunya University dataset offers EEG recordings from a wider demographic, spanning ages 1 to 107 years [24]. The EU database goes further, with EEG records from over 250 epilepsy patients, providing extensive data for in-depth analysis [25]. The NICU dataset focuses on 79 newborns, contributing critical early-life data to the field [26]. Additionally, the BONN database, though structured around brief EEG segments, encompasses a variety of cases [27]. Most notably, the TUH EEG Corpus stands out as a vast resource, compiled over 14 years and including data from 10,874 unique patients, making it an invaluable asset for neurological research and diagnostics. This array of datasets highlights the diverse patient populations involved in EEG studies, underscoring the need for expansive data to drive advancements in medical science [18]. It offers a broad range of patient ages, medication, diagnoses, electrode counts, and sampling rates. In this paper, the data of 2300 patients from the TUH dataset corpus are taken, including 1150 normal and 1150 abnormal cases. Each participant data consists of 300,500 readings for a maximum of 22 EEG electrodes [18] and a minimum of 19 electrodes. Table 1 shows the distribution of cases between normal and abnormal categories in both training and testing datasets, along with gender percentages.

Table 1. Distribution of Normal and Abnormal Cases in Training and Testing Datasets by Gender.

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

2.1.2. Data Selection

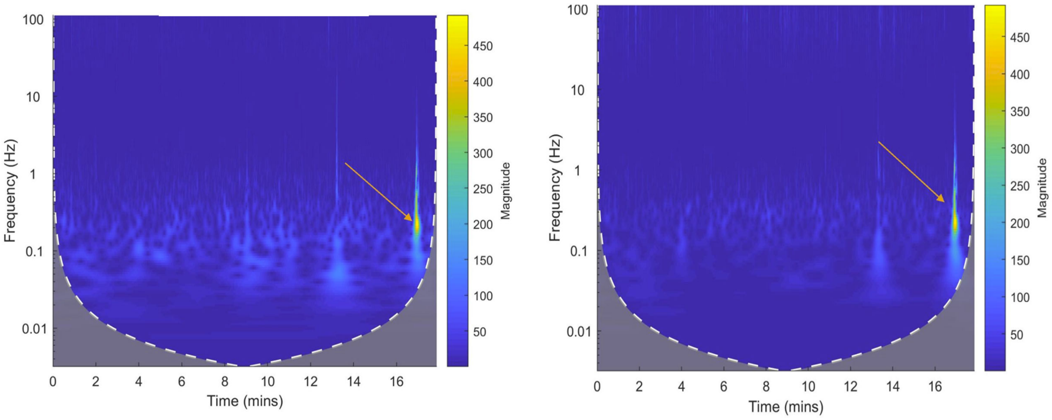
Time–frequency analysis is selected as a tool for processing EEG signals. The time–frequency analysis adopted in the work is based on a continuous wavelet transform [28], using Generalized Morse Wavelets (GMWs) [29]. GMWs are a family of analytic wavelets where only the positive real axis is supported for Fourier transformations of complex-valued wavelets. The analytic wavelets are efficient at the analysis of signals with time-varying amplitude and frequency, known as modulated signals. Nonanalytical wavelets are used for enhancing sharp signal transitions in the time–frequency plane, whereas analytical wavelets analyze the oscillatory behavior of frequency transients. GMWs overcome problems associated with non-analytical wavelets, such as the occurrence of interferences and artefacts in the time–frequency plane [30–33]. The frequency domain in terms of w representations of GMWs is given in Equation (1)

$$\Psi_{\beta,\gamma}(\omega) = U(\omega)\, a_{\beta,\gamma}\, \omega^{\beta} e^{-\omega\gamma} \tag{1}$$

where $\Psi_{\beta,\gamma}$ represents the frequency domain components of GMWs, U is the unit step function, and $a_{\beta,\gamma}$ is the normalization constant. The wavelet's time domain decay is governed by the β parameter, but the frequency domain decay is controlled by γ. The conditions $\gamma > 0$ and $\alpha > 0$ must be satisfied in order for a GMW to be considered genuine. Furthermore, with e denoting Euler's number, the normalization constant is governed by $a_{\beta,\gamma} \geq 2(e\gamma/\beta)\ \beta/\gamma$. When $\gamma = 1$, GMWs correspond to a solution of the Schrödinger equation [29]. For $\gamma = 2$, GMWs are the same as the Derivative of Gaussian family of wavelets and the gaussian and symmetric. GMWs are suggested as a good place to start for general purpose use because of these qualities [29,31].

In this paper, γ is set to 3 and the time-bandwidth product is equal to 60. The parameters for the Generalized Morse Wavelets were chosen based on both theoretical considerations and empirical evidence. The value of $\gamma = 3$ was selected because it offers a favorable balance between time and frequency localization, as seen in the Airy wavelet, which is symmetric in the time domain and nearly symmetric in the frequency domain [31]. This symmetry is beneficial for analyzing oscillatory behaviors in EEG signals.

The time-bandwidth product of 60 was chosen to achieve a high level of spectral resolution while maintaining computational efficiency. A larger time-bandwidth product generally enhances frequency resolution but at the cost of reduced time resolution [29,33]. Our preliminary experiments indicated that this value provided the best trade-off for our specific application, allowing us to effectively capture the frequency components of interest in the EEG signals. These choices align with established practices in the field, where similar parameters have been successfully used for analyzing time-varying signals [32]. By selecting these parameters, we aimed to optimize the performance of the GMWs for our specific task of analyzing EEG signals. Thus, the time-domain data obtained from each of the electrodes are transformed into the frequency domain. Figure 3 illustrates the magnitude scalogram for two different electrodes of the same patient, where the frequency is plotted on a logarithmic scale. The cone of influence [31], showing where edge effects become significant, is also plotted. Gray regions outside the dashed white line describe regions where edge effects are significant.



(**a**) (**b**)

Figure 3. The magnitude scalogram of two different electrodes for the same patient, (**a**) electrode number 4, (**b**) electrode number 2. The arrow highlights a strong signal component, indicating a specific event in the data at a particular frequency and time.

It is observed that the most significant (or maximum) components lie within nearly the same time interval for the different EEG signals produced, in parallel, by the electrodes. This is taken as the basis for selecting the appropriate window size. The maximum value of the wavelet-transformed signals occurs at the time interval where potentially distinctive features are present [28,32,34,35].

Hence, the time instants where the maximum value for the resulting transformed signal of each electrode occurs, is obtained for each patient individually. The average of these positions in time, across all electrodes, is obtained for each patient, excluding outliers, which are positions that are more than 3 standard deviations away from the average position. A time interval around the average position is selected for the patient and the remaining data are discarded. This time interval represents a portion of the data where defining features exist in the majority of the electrodes. This step of the windowing process is illustrated in Figure 4. Narrow, medium and wide window sizes around the average value are tested to check the accuracy. Three window sizes are tested, containing 500, 2000, and 8000 samples. The experiments are carried out using the windowing technique on all 19 electrodes for each patient.

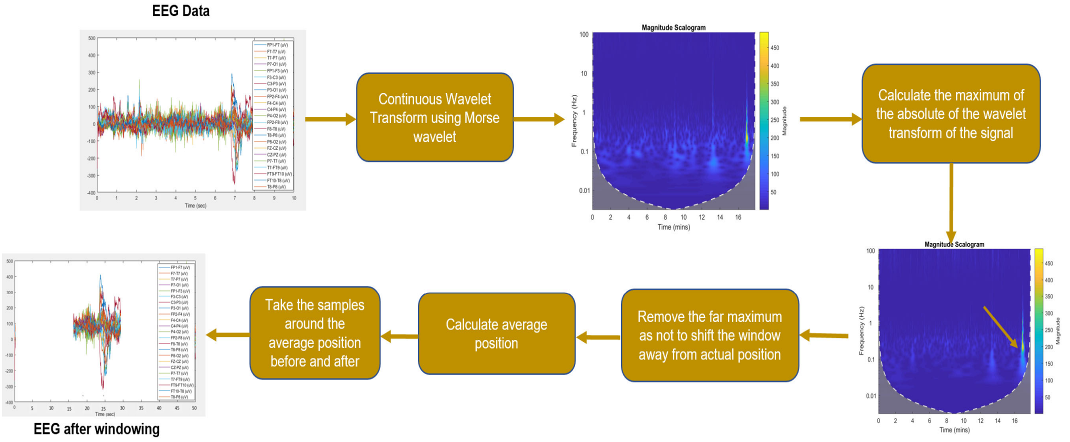


Figure 4. Steps of the proposed windowing technique.

2.1.3. Feature Extraction and Classification

Several designs are taken into consideration to assess the efficiency of CNNs and LSTMs for this particular problem and to overcome their weaknesses. By looping over time steps and discovering long-term dependencies between them, an LSTM network analyses sequence data [21,22]. In initial experiments, LSTM is used without prior data selection or feature extraction. Performance of the LSTM network was satisfactory on training data, but unsatisfactory during testing, indicating overfitting [22]. In addition, the long sequence of features, though shorter than the original data, poses a problem for the LSTM. This is mainly because LSTMs do not possess a long enough memory span to process long sequences [22].

Dropout and feature extraction are investigated as potential remedies against overfitting. Dropout is used to pick neurons at random and discard them during training; nonetheless, overfitting persists. A comparison of the model with and without windowing shows that the accuracy increased after applying data selection, yet the gap between training and testing performance is still substantial, indicating significant overfitting.

Hence, convolutional layers are added to extract important features embedded in the EEG, and then these features are passed to the LSTM. Average pooling and max-pooling layers are utilized, in the CNN, to focus on the important features after each convolution step. The CNN–LSTM deep learning model is designed to comprise 18 layers that start with a 1D convolutional layer followed by average pooling, then another 1D convolution layer with average pooling, and these layers are used for feature extraction. After feature extraction through convolutional layers, LSTM with the fully connected layers is used for classification. A process of 10-fold cross-validation is employed in all experiments, to ensure that performance is accurate and generalizable.

3. Performance and Evaluation

3.1. Procedure

All simulations are carried out using MATLAB version 9.12.0.1884302 (R2022a), running on an Intel I core™ i7-8700 CPU @3.20 GHz with 128 GB of RAM. The data are originally sampled at a frequency of 250 Hz [19]. The signals are first processed by the windowing technique for data selection, and then the proposed CNN–LSTM feature extraction and classification model is used to label them into two classes: "normal" and "abnormal". The TUH dataset [18] is used to assess the performance of the proposed model. A total of 2574 patient records are used for training and testing. The L2 regularization approach is employed to reduce overfitting, since it distributes the weights among the features more evenly [36]. Hyper parameters, optimization and regularization techniques are detailed in Table 2. Various filters and window sizes were tested to evaluate their impact on accuracy. The results revealed that using a single filter resulted in 69.67%, while with 5 filters the accuracy increased to 70.3%, and 7 filters yielded to a lower accuracy of 62.67%. Additionally, tests showed that a window size of 30 provided an accuracy of 70%, a window size of 50 improved accuracy to 74.50%, and a window size of 60 resulted in an accuracy of 70.67%. These findings highlight the influence of filter selection and window size on performance. The architecture of the proposed 18-layer CNN–LSTM network is illustrated in Figure 5. Cross-validation is performed on the data, where each fold comprises 90% of the data for training, with the remaining 10% being used as hold-out data for validation. The average and maximum values of sensitivity, specificity [36–38] and accuracy are calculated for 10 folds.

Table 2. Hyper Parameters, Optimization 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/Optimizer	ADAM
	Learning rate	10^{-4}
	L2Regularization	0.001
	Mini batch algorithm	Gradient decent
	Input Size	19
	Number of Epochs	30
	Gradient Threshold	1

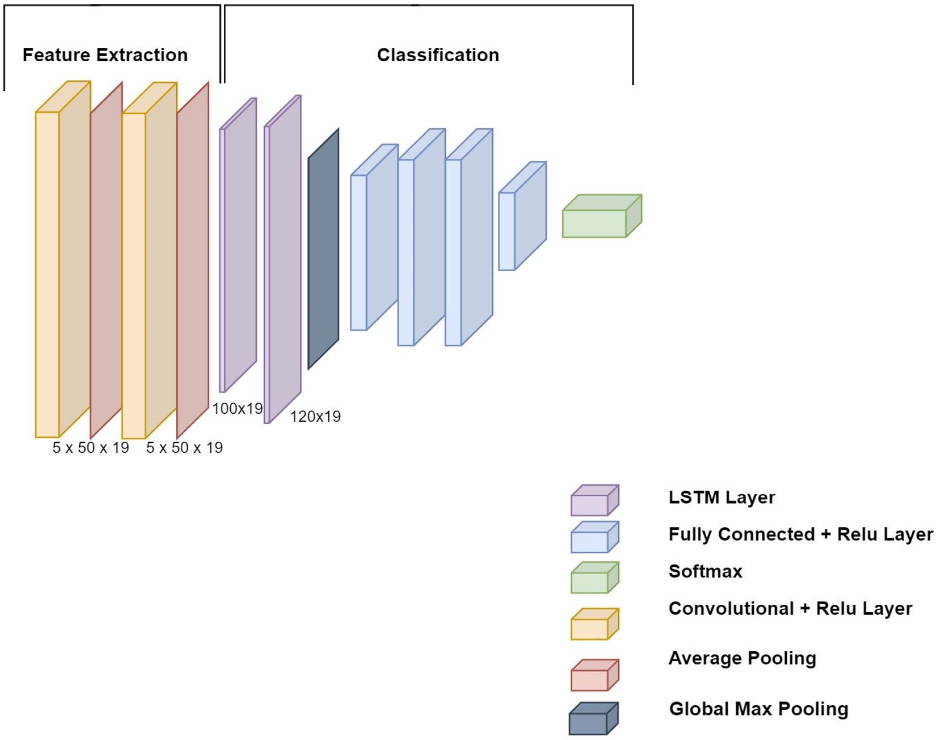


Figure 5. Illustration of the proposed 18-layer CNN–LSTM deep learning module.

3.2. Experimental Results

An ablation study was conducted with four different models: Model 1 utilized dropout regularization to prevent overfitting, Model 2 employed an LSTM-only architecture for sequential data analysis, Model 3 combined CNN and LSTM layers to capture spatial and temporal features, and Model 4 enhanced the CNN–LSTM framework with an 18-layer structure and a windowing technique for improved feature extraction and processing. The study aimed to evaluate the individual contributions of each component to the overall performance of the brain disorder diagnosis model. The study aimed to compare the performance of the different models in terms of ROC curve, classification accuracy, F1 score, recall, and precision, evaluating the individual contributions of each component to the overall performance of the diagnosis system.

The Area Under the Curve (AUC) of the ROC plot serves as a key metric for assessing model performance, with an AUC of 1 indicating a flawless classifier and a score of 0.5 suggesting a random classifier. This makes the AUC a reliable indicator for evaluating the efficacy of score classifiers, as it encapsulates the model's performance across the entire range of thresholds. Figure 6 shows the ROC curve for the four different architectures: Model 1 (50% dropout architecture), Model 2 (LSTM only), Model 3 (CNN and LSTM deep learning model), and Model 4 (18-layer CNN LSTM deep learning model with windowing technique); the proposed model with the selected windowing technique (Model 4) demonstrates a significant distance from the line of no discrimination. This indicates that the model achieves a good level of class separation performance, suggesting its effectiveness in distinguishing between positive and negative classes.

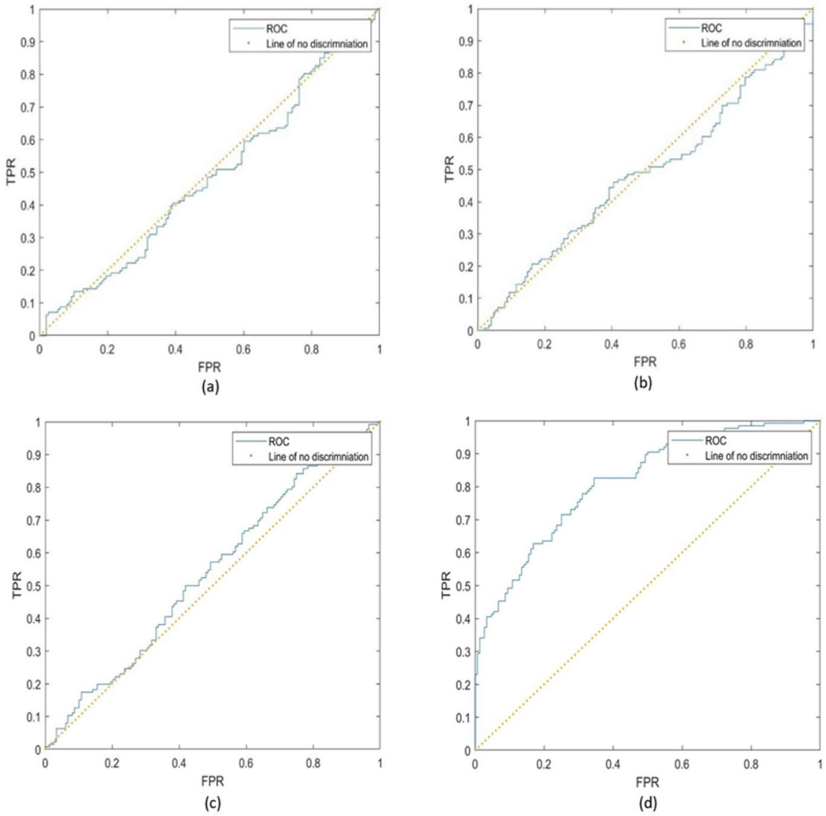
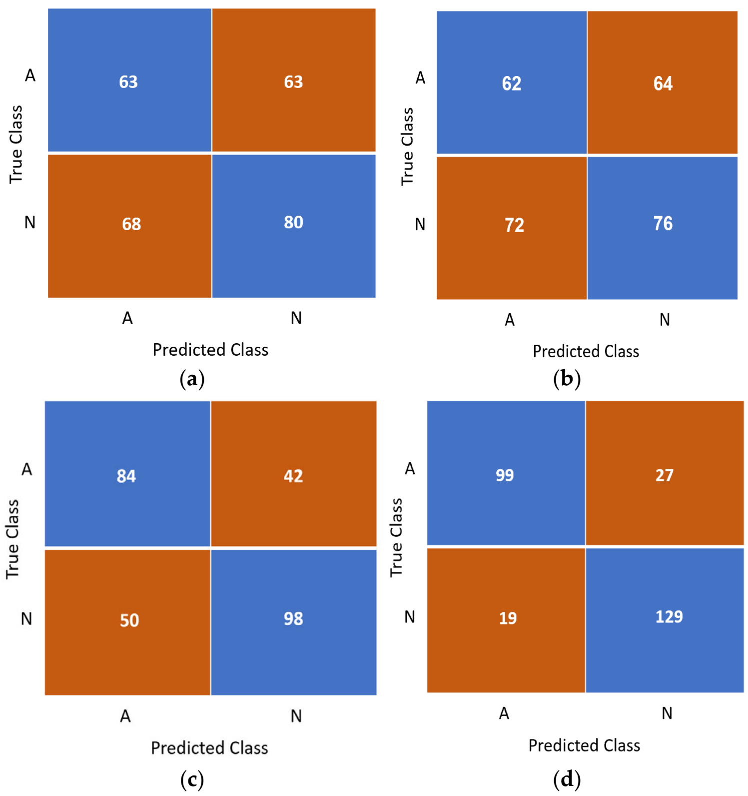


Figure 6. ROC plots for different architectures: (**a**) Model 1 (**b**) Model 2. (**c**) Model 3. (**d**) Proposed Model.

Classification accuracy alone can be a misleading indicator of classification performance; for this reason, a confusion matrix is used, since it provides a detailed breakdown of the model's predictions against the actual outcomes, offering a clearer picture of what the model is getting right and what types of errors it is making. Specifically, it helps identify True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). Figure 7 shows the confusion matrix for different architectures: 50% dropout architecture, LSTM only, CNN and LSTM, and 18-layer CNN LSTM deep learning model with windowing. It is clear that the proposed CNN–LSTM model with windowing achieves the best results when compared to the different architectures. The confusion matrix of the model shows comparable accuracies in detecting normal and abnormal cases.

The comparison of classification models reveals distinct performance characteristics. Model 1 and Model 2 exhibit relatively low precision and recall, with Model 1 achieving a precision of 0.481 and recall of 0.5, while Model 2 has a precision of 0.463 and recall of 0.492. Both models struggle with accuracy, as evidenced by their low F1 scores of 0.490 and 0.463, respectively. In contrast, Model 3 shows improved performance with a precision of 0.627 and recall of 0.667, resulting in an F1 score of 0.646. However, the Proposed Model outperforms all others, boasting a high precision of 0.839 and recall of 0.786, which translates to an impressive F1 score of 0.812. This suggests that the Proposed Model is significantly more effective at balancing accuracy and comprehensiveness, making it the most reliable choice among the options. High recall is crucial to ensure that most patients with the disease are correctly identified. However, high precision is also important to avoid unnecessary treatments. The F1 score provides a balanced measure of both, helping to evaluate the model's overall performance in this context. Table 3 compares the performance of the four models by presenting key metrics related to their classification effectiveness.



(**a**) (**b**) (**c**) (**d**)

Figure 7. Confusion matrix for the four different architectures, where 'A' indicates 'Abnormal' and 'N' indicates 'Normal' EEG signal. The red color presents the TP and TN, while the blue color corresponds to the FP and FN values. (**a**) Model 1 (**b**) Model 2 (**c**) Model 3 (**d**) Model 4 (Proposed Model).

Table 3. Comparison of Model Performance with different architectures using Precision, Recall, and F1 Score.

	TP	**FN**	**TN**	**FP**	**Precision**	**Recall**	**F1 Score**
Model 1	63	63	80	68	0.481	0.5	0.490
Model 2	62	64	76	72	0.463	0.492	0.463
Model 3	84	42	98	50	0.627	0.667	0.646
Proposed Model	99	19	129	27	0.839	0.786	0.812

The ablation studies conducted demonstrate the effectiveness of various models using different combinations of feature extraction methods and classification techniques, along with their respective training and testing accuracies. Model 1, which uses only an LSTM classifier without any feature extraction, achieved a training accuracy of 65% and a testing accuracy of 48.90%. Model 2 enhances feature robustness by incorporating a 50% dropout rate during the extraction phase and using an LSTM for classification. This resulted in a slight increase in training accuracy to 68%, although the testing accuracy decreased slightly to 47.87%. Model 3 significantly improved both training and testing accuracy to 98% and 67.06%, respectively, by combining a CNN for feature extraction with an LSTM classifier. The proposed model introduces a windowing technique for data selection, which, when combined with CNN-based feature extraction and LSTM classification, substantially enhances performance. This configuration achieved a perfect training accuracy of 100% and the highest testing accuracy of 82.68%. This highlights the superiority of the windowing technique in optimizing model efficacy, underscoring its crucial role in achieving the best performance metrics among the configurations studied. Table 4 compares various modules, making it evident that using the LSTM alone without feature extraction and feature selection or after incorporating dropout does not resolve overfitting. However, the results are improved by connecting a CNN upstream to the LSTM. This combined module produces the best results once the windowing technique is added.

Table 4. Ablation Study of Various Components of the Proposed Architecture.

Title 1	Data Selection	Feature Extraction	Classification Technique	Training Accuracy	Testing Accuracy
Model 1	-	-	LSTM	65%	48.90%
Model 2	-	Dropout 50%	LSTM	68%	47.87%
Model 3	-	CNN	LSTM	98%	67.06%
Proposed Model	Windowing Technique	CNN	LSTM	100%	82.68%

In this paper, the proposed model's performance is evaluated across various window lengths using 10-fold cross-validation, with a focus on four key metrics: Average Accuracy, Maximum Accuracy, Sensitivity (SV), and Specificity (SP). The evaluation is conducted for window lengths of 500, 2000, and 8000 samples, providing insights into how the model's performance scales with increasing data availability. The results demonstrate that the model's accuracy improves as the window length increases, highlighting its ability to leverage larger datasets for enhanced learning and pattern recognition. The Average Accuracy metric, which represents the mean accuracy across all 10 folds, shows a consistent upward trend with larger window lengths. Starting at 70.61% for 500 data points, it rises to 73.40% for 2000 data points and reaches 78.5% for 8000 data points. This improvement suggests that the proposed model benefits from more samples, as they enable the model to capture more nuanced patterns. Similarly, Maximum Accuracy shows a notable increase from 74.78% at 500 data points to 82.68% at 8000 data points, reflecting the highest accuracy achieved in any single fold, which also increases with larger window lengths, indicating that the model's peak performance potential is enhanced with more data. Sensitivity and Specificity provide further insights into the model's behavior. Sensitivity measures the model's ability to correctly identify positive cases, while Specificity evaluates its ability to correctly identify negative cases. At 500 data points, the sensitivity (70.77%) and specificity (70.58%) are nearly balanced, indicating comparable performance for both positive and negative cases. However, at 8000 data points, specificity (87%), exceeds the sensitivity (78.5%), suggesting that the model becomes more adept at identifying negative cases when trained on larger datasets. The use of 10-fold cross-validation is crucial for evaluation, ensuring a robust and reliable assessment of the model's performance. This method involves partitioning the dataset into 10 subsets and iteratively training and testing the model on each subset. By doing so, it mitigates the risk of overfitting and provides a more accurate estimate of the model's generalization ability. Additionally, 10-fold cross-validation allows for efficient utilization of the data, as every data point is used for both training and testing across the folds. Moreover, 10-fold cross-validation offers valuable insights into the stability of the proposed model. By analyzing the variation in accuracy across folds, such as the difference between the average and maximum accuracy, one can assess the consistency of the model's performance. This is particularly crucial in applications where reliability is paramount, such as medical diagnosis or fraud detection. The method also plays a vital role in hyperparameter tuning, ensuring that the selected parameters generalize effectively to unseen data, thereby enhancing the model's overall robustness and applicability. The simulation results for 10-fold cross validation, with different numbers of samples selected using windowing technique, are given in Table 5. The table includes the length of windows, the average accuracy, the maximum accuracy, sensitivity (SV) and specificity (SP). It can be observed that the highest accuracy is observed when a window size of 8000 samples is used.

Table 5. Accuracy of The Proposed Model for Various Window Lengths, using 10-Fold Cross Validation.

Length of Sample/Data Points	Average Accuracy	Maximum Accuracy	SV	SP
500	70.61%	74.78%	70.77%	70.58%
2000	73.40%	78.26%	71.54%	75.45%
8000	74.50%	82.68%	78.5%	87.10%

The proposed model introduces a novel windowing technique that significantly enhances the accuracy of EEG data analysis compared to existing methods. This innovative approach, combined with a sophisticated 18-layer deep learning architecture incorporating both Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), yields the highest accuracy of 82.68%. This surpasses the accuracies reported by other studies, such as Sylvia et al. (78.65%) [39], Tomas et al. (68%) [16], A. Harati et al. (75.4%) [40], and D. Western et al. (81.88%) [41]. The superiority of the proposed windowing technique lies in its ability to effectively manage the non-linear and non-stationary nature of EEG signals, which are prone to noise and distortions [42]. Unlike traditional windowing methods, like Rectangular, Hamming, and Hann, which may introduce ripples and distortions during filtering [43], the novel technique optimizes signal processing by potentially reducing such issues. Additionally, it allows for more flexible time–frequency representation, similar to wavelet transforms, which can adapt to variable window sizes for better signal analysis [44]. This adaptability and precision in signal processing contribute to the model's superior performance in extracting meaningful features and improving classification accuracy. A comparison between the proposed architecture and others reported in the literature is given in Table 6. The table includes the data selection modules, classification modules incorporated in the architecture, and the accuracy achieved. The TUH data set is used with the same number of patients for comparison to existing architectures in the literature, in order to unify the comparison grounds. Table 6 shows that the proposed architecture outperforms existing architectures in the literature in terms of accuracy.

Table 6. Comparison between obtained results and others in the literature.

	Data Selection Technique	Feature Extraction	Classification Technique	Accuracy
Sylvia et al. [39]	1st minute	HMM	GMM	78.65%
Tomas et al. [16]	-	PS + PLV + Energy	HMM	68%
Harati et al. [40]	-	PCA	CNN	75.4%
Western et al. [41]	2nd minute	-	CNN	81.88%
Proposed Model	Proposed windowing technique	Proposed 18-layer deep learning CNN + LSTM		82.68%

4. Conclusions

The deep learning-based architecture presented in this paper can help in the diagnosis of brain disorders by labeling them as "normal" or "abnormal". It was designed with the objective of increasing the accuracy of results, by properly selecting the relevant data and enhancing the feature extraction and classification processes. A windowing technique was devised for data selection, whereas a combined CNN–LSTM module is proposed for feature extraction and classification. Multiple ablation studies were conducted to confirm the substantial impact of each element of the proposed architecture. The model was tested using the TUH data set and the relevant results showed improved accuracy as compared to other architecture reported in the literature and assessed on the same dataset. The results also showed that performance is stable for a large number of patients and generalizes well over different types of patients. This model overcomes previous biases in previous studies with a much lower number of patients and/or age, gender, acquisition conditions and disorder biases. The proposed model achieved an accuracy of 82.68% on the TUH EEG dataset. Furthermore, it showed strong true positive and low false negative rates with high sensitivity (78.5%) and specificity (87.10%) rates. In the future, we will employ random search as a strategy for hyperparameter tuning to enhance our model's performance. Furthermore, we intend to improve the architecture of the proposed model to support real time analysis of EEG data, which would significantly improve its practical medical application. This would allow for immediate diagnostic support during patient assessments.

Author Contributions: D.A.: writing the manuscript, conducting experiments and data analysis. N.Z.: Conducting experiments, data analysis, reviewing and approving the final manuscript. A.H.S.: Data analysis and approving the final manuscript. C.C.: Data analysis and approving the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The datasets used and analyzed during the current study are available in: https://isip.piconepress.com/projects/nedc/html/tuh_eeg/ (accessed on 13 May 2016).

Conflicts of Interest: The authors confirm that there are no conflicts of interest, and the research was carried out without any involvement of commercial or financial relationships.

References

1. Amer, N.S.; Belhaouari, S.B. EEG signal processing for medical diagnosis, healthcare, and monitoring: A comprehensive review. *IEEE Access* **2023**, *11*, 143116–143142. [CrossRef]
2. Abir, S.I.; Shoha, S.; Hossain, M.M.; Sultana, N.; Saha, T.R.; Sarwer, M.H.; Saimon, S.I.; Islam, I.; Hasan, M. Machine Learning and Deep Learning Techniques for EEG-Based Prediction of Psychiatric Disorders. *J. Comput. Sci. Technol. Stud.* **2025**, *7*, 46–63. [CrossRef]
3. Gadgil, A.A.; Selvakumar, P.; Gnanaselvi, G.S.; Malathi, G. AI in Neuroimaging and Brain Analysis. In *Transforming Neuropsychology and Cognitive Psychology with AI and Machine Learning*; IGI Global: Hershey, PA, USA, 2025; pp. 185–212.
4. Saminu, S.; Xu, G.; Zhang, S.; Ab El Kader, I.; Aliyu, H.A.; Jabire, A.H.; Ahmed, Y.K.; Adamu, M.J. Applications of artificial intelligence in automatic detection of epileptic seizures using EEG signals: A review. *Artif. Intell. Appl.* **2023**, *1*, 11–25. [CrossRef]
5. Ahmed, M.I.B.; Alotaibi, S.; Dash, S.; Nabil, M.; AlTurki, A.O. A Review on Machine Learning Approaches in Identification of Pediatric Epilepsy. *SN Comput. Sci.* **2022**, *3*, 437. [CrossRef]
6. Rahul, J.; Sharma, D.; Sharma, L.D.; Nanda, U.; Sarkar, A.K. A systematic review of EEG-based automated schizophrenia classification through machine learning and deep learning. *Front. Hum. Neurosci.* **2024**, *18*, 1347082. [CrossRef]
7. Bajaj, A.; Sharma, V. EEG-Based Epileptic Seizure Prediction Using Variants of the Long Short-Term Memory Algorithm. *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.* **2025**, *17*, 13.
8. Abdelhady, G. Enhancing Neuroprosthetic Control Using CNN-LSTM Models: A Simulation Study with EEG-Based Motor Imagery. *Artif. Intell. Inf. Secur.* **2025**, *3*, 17–35. [CrossRef]
9. Ma, Y. Application of LSTM emotion recognition based on deep learning and attention mechanism in patient relationship mediation. In Proceedings of the Fourth International Conference on Computer Vision, Application, and Algorithm (CVAA 2024), Chengdu, China, 11–13 October 2024; Volume 13486, pp. 229–234.
10. Cao, X.; Zheng, S.; Zhang, J.; Chen, W.; Du, G. A hybrid CNN-Bi-LSTM model with feature fusion for accurate epilepsy seizure detection. *BMC Med. Inform. Decis. Mak.* **2025**, *25*, 6. [CrossRef]
11. Kim, M.-J.; Youn, Y.C.; Paik, J. Deep learning-based EEG analysis to classify normal, mild cognitive impairment, and dementia: Algorithms and dataset. *NeuroImage* **2023**, *272*, 120054. [CrossRef]

12. Fergus, P.; Hussain, A.; Hignett, D.; Al-Jumeily, D.; Abdel-Aziz, K.; Hamdan, H. A machine learning system for automated whole-brain seizure detection. *Appl. Comput. Inform.* **2016**, *12*, 70–89. [CrossRef]
13. Dişli, F.; Gedikpınar, M.; Fırat, H.; Şengür, A.; Güldemir, H.; Koundal, D. Epilepsy Diagnosis from EEG Signals Using Continuous Wavelet Transform-Based Depthwise Convolutional Neural Network Model. *Diagnostics* **2025**, *15*, 84. [CrossRef] [PubMed]
14. Tong, P.F.; Dong, B.; Zeng, X.; Chen, L.; Chen, S.X. Detection of interictal epileptiform discharges using transformer-based deep neural network for patients with self-limited epilepsy with centrotemporal Spikes. *Biomed. Signal Process. Control* **2025**, *101*, 107238. [CrossRef]
15. Hu, W.; Wang, J.; Li, F.; Ge, D.; Wang, Y.; Jia, Q.; Yuan, S. A Modified Transformer Network for Seizure Detection Using EEG Signals. *Int. J. Neural Syst.* **2025**, *35*, 2550003. [CrossRef] [PubMed]
16. Iešmantas, T.; Alzbutas, R. Convolutional neural network for detection and classification of seizures in clinical data. *Med. Biol. Eng. Comput.* **2020**, *58*, 1919–1932. [CrossRef] [PubMed]
17. Obeid, I.; Picone, J. Machine learning approaches to automatic interpretation of EEGs. In *Signal Processing and Machine Learning for Biomedical Big Data*; CRC Press: Boca Raton, FL, USA, 2018; pp. 271–300.
18. Harati, A.; Choi, S.; Tabrizi, M.; Obeid, I.; Picone, J.; Jacobson, M.P. The temple university hospital EEG corpus. In Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing, Austin, TX, USA, 3–5 December 2013; pp. 29–32.
19. Yıldırım, Ö.; Baloglu, U.B.; Acharya, U.R. A deep convolutional neural network model for automated identification of abnormal EEG signals. *Neural Comput. Appl.* **2020**, *32*, 15857–15868. [CrossRef]
20. Alotaiby, T.; El-Samie, F.E.A.; Alshebeili, S.A.; Ahmad, I. A review of channel selection algorithms for EEG signal processing. *EURASIP J. Adv. Signal Process.* **2015**, *2015*, 66. [CrossRef]
21. Zulkifley, M.; Abdani, S.R. EEG signals classification by using convolutional neural networks. In Proceedings of the IEEE Symposium on Acoustics, Speech and Signal Processing (SASSP 2019), Kuala Lumpur, Malaysia, 20 March 2019.
22. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
23. Shoeb, A.H. Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
24. Selvaraj, T.G.; Ramasamy, B.; Jeyaraj, S.J.; Suviseshamuthu, E.S. EEG database of seizure disorders for experts and application developers. *Clin. EEG Neurosci.* **2014**, *45*, 304–309. [CrossRef]
25. Epilepsy Database. Available online: http://epilepsy-database.eu (accessed on 12 February 2025).
26. Stevenson, N.J.; Tapani, K.; Lauronen, L.; Vanhatalo, S. A dataset of neonatal EEG recordings with seizure annotations. *Sci. Data* **2019**, *6*, 190039. [CrossRef]
27. Andrzejak, R.G.; Lehnertz, K.; Rieke, C.; Mormann, F.; David, P.; Elger, C.E. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* **2001**, *64*, 061907. [CrossRef]
28. Gurrala, V.; Yarlagadda, P.; Koppireddi, P. Wavelet domain extraction of features from single channel sleep EEG. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 230–232.
29. Olhede, S.C.; Walden, A.T. Generalized morse wavelets. *IEEE Trans. Signal Process.* **2002**, *50*, 2661–2670. [CrossRef]
30. Lilly, J.M. Element analysis: A wavelet-based method for analysing time-localized events in noisy time series. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2017**, *473*, 20160776. [CrossRef] [PubMed]
31. Suwansawang, S.; Halliday, D. Time-frequency based coherence and phase locking value analysis of human locomotion data using generalized morse wavelets. In *International Conference on Bio-Inspired Systems and Signal Processing (BIOSTEC 2017)*; SciTePress: Setúbal, Portugal, 2017; Volume 5, pp. 34–41.
32. Martinez-Ríos, E.A.; Bustamante-Bello, R.; Navarro-Tuch, S.; Perez-Meana, H. Applications of the generalized morse wavelets: A review. *IEEE Access* **2022**, *11*, 667–688. [CrossRef]
33. Lilly, J.M.; Olhede, S.C. Generalized morse wavelets as a superfamily of analytic wavelets. *IEEE Trans. Signal Process.* **2012**, *60*, 6036–6041. [CrossRef]
34. Zhang, Z.; Moore, J. Intrinsic feature extraction in the COI of wavelet power spectra of climatic signals. In Proceedings of the 4th International Congress on Image and Signal Processing, Shanghai, China, 15–17 October 2011; Volume 5, pp. 2354–2356.
35. Newson, J.J.; Thiagarajan, T.C. EEG frequency bands in psychiatric disorders: A review of resting state studies. *Front. Hum. Neurosci.* **2019**, *12*, 521. [CrossRef]
36. Guerrero, M.C.; Parada, J.S.; Espitia, H.E. EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks. *Heliyon* **2021**, *7*, e07258. [CrossRef]
37. Baratloo, A.; Hosseini, M.; Negida, A.; El Ashal, G. Part 1: Simple definition and calculation of accuracy, sensitivity, and specificity. *Emergency* **2015**, *3*, 48–49. [CrossRef]
38. McDonald, A.C.; Gasperin Haaz, I.; Qi, W.; Crowley, D.C.; Guthrie, N.; Evans, M.; Bosnyak, D. Sensitivity, Specificity, and Accuracy of a Novel EEG-Based Objective Test, the Cognalyzer®, in Detecting Cannabis Psychoactive Effects. *Adv. Ther.* **2021**, *38*, 2513–2531. [CrossRef]

39. Lopez, S. Automated Identification of Abnormal EEGs. Master's Thesis, Temple University, Philadelphia, PA, USA, 2017.
40. Harati, A.; Golmohammadi, M.; Lopez, S.; Obeid, I.; Picone, J. Improved EEG event classification using differential energy. In Proceedings of the 2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 12 December 2015; pp. 1–4.
41. Western, D.; Weber, T.; Kandasamy, R.; May, F.; Taylor, S.; Zhu, Y.; Canham, L. Automatic report-based labelling of clinical eegs for classifier training. In Proceedings of the 2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 4 December 2021; pp. 1–6.
42. Singh, A.K.; Krishnan, S. Trends in EEG signal feature extraction applications. *Front. Artif. Intell.* **2023**, *5*, 1072801. [CrossRef]
43. Diana, N.; Kalsum, U.; Sabiq, A.; Jatmiko, W.; Mursanto, P. Comparing windowing methods on finite impulse response (FIR) filter algorithm in electroencephalography (EEG) data processing. *J. Theor. Appl. Inf. Technol.* **2016**, *88*, 558–567.
44. Al-Fahoum, A.S.; Al-Fraihat, A.A. Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. *Int. Sch. Res. Not.* **2014**, *2014*, 730218. [CrossRef] [PubMed]