



Optimizing epileptic seizure recognition performance with feature scaling and dropout layers

Ahmed Omar¹ · Tarek Abd El-Hafeez^{1,2}

Received: 29 July 2023 / Accepted: 21 October 2023 / Published online: 24 November 2023
© The Author(s) 2023

Abstract

Epilepsy is a widespread neurological disorder characterized by recurring seizures that have a significant impact on individuals' lives. Accurately recognizing epileptic seizures is crucial for proper diagnosis and treatment. Deep learning models have shown promise in improving seizure recognition accuracy. However, optimizing their performance for this task remains challenging. This study presents a new approach to optimize epileptic seizure recognition using deep learning models. The study employed a dataset of Electroencephalography (EEG) recordings from multiple subjects and trained nine deep learning architectures with different preprocessing techniques. By combining a 1D convolutional neural network (Conv1D) with a Long Short-Term Memory (LSTM) network, we developed the Conv1D + LSTM architecture. This architecture, augmented with dropout layers, achieved an effective test accuracy of 0.993. The LSTM architecture alone achieved a slightly lower accuracy of 0.986. Additionally, the Bidirectional LSTM (BiLSTM) and Gated Recurrent Unit (GRU) architectures performed exceptionally well, with accuracies of 0.983 and 0.984, respectively. Notably, standard scaling proved to be advantageous, significantly improving the accuracy of both BiLSTM and GRU compared to MinMax scaling. These models consistently achieved high test accuracies across different percentages of Principal Component Analysis (PCA), with the best results obtained when retaining 50% and 90% of the features. Chi-square feature selection also enhanced the classification performance of BiLSTM and GRU models. The study reveals that different deep learning architectures respond differently to feature scaling, PCA, and feature selection methods. Understanding these nuances can lead to optimized models for epileptic seizure recognition, ultimately improving patient outcomes and quality of life.

Keywords Epileptic seizure recognition · EEG · Conv1D · LSTM · Feature scaling · PCA · Chi-square feature selection · BiLSTM · GRU

1 Introduction

Epileptic Seizure Recognition is a critical area of research in the field of medical diagnostics and neuroscience. Epilepsy is a neurological disorder characterized by recurrent and unprovoked seizures, which are abnormal electrical discharges in the brain. Detecting and predicting epileptic

seizures accurately is of utmost importance for providing timely medical interventions and improving the quality of life for patients [1].

One of the primary tools used in Epileptic Seizure Recognition is Electroencephalography (EEG). EEG measures the electrical activity of the brain using electrodes placed on the scalp. EEG signals contain valuable information about brain activity and can help in identifying abnormal patterns associated with seizures [2].

Deep Neural Networks have gained popularity in many applications in medicine, energy, and environmental studies [3–6]. Deep Learning has emerged as a powerful approach for EEG-based seizure recognition. Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that can process time-series data, making it particularly suitable for analyzing EEG signals, which are

✉ Ahmed Omar
ahmed.omar@mu.edu.eg

✉ Tarek Abd El-Hafeez
tarek@mu.edu.eg

¹ Department of Computer Science, Faculty of Science, Minia University, EL-Minia, Egypt

² Computer Science Unit, Deraya University, EL-Minia, Egypt

sequential in nature. LSTMs can capture long-term dependencies in the data and have shown promising results in seizure detection and prediction tasks [7].

Before feeding EEG data into LSTM or any other deep learning model, pre-processing steps are essential to enhance the model's performance. Feature scaling is one such step that normalizes the input data to bring all features to a similar scale. This process ensures that no single feature dominates the learning process due to its larger magnitude [8].

Principal Component Analysis (PCA) is another pre-processing technique that helps in dimensionality reduction. EEG data may have a high number of features, and PCA can compress this data into a lower-dimensional space while retaining most of its important characteristics. This not only reduces computational complexity but also prevents overfitting and enhances generalization [9].

Feature selection is crucial to select the most relevant EEG features for seizure recognition. Chi-Square Feature Selection is one such method that assesses the statistical significance of each feature's relationship with the target variable (seizure or non-seizure). By selecting features with high chi-square values, the model focuses on the most informative aspects of the EEG data, leading to improved accuracy [10].

Deep Learning Models are a class of algorithms that utilize neural networks with multiple layers to learn intricate patterns and representations from the data. Convolutional Neural Networks (CNNs) have also been applied in EEG-based seizure recognition, especially when dealing with spectrogram or image-like representations of EEG signals [11].

The pipeline for Epileptic Seizure Recognition typically involves data preprocessing, including feature scaling, PCA, and feature selection. Next, the selected EEG features are used to train LSTM or other deep learning models for seizure detection or prediction. The model is then evaluated using metrics such as accuracy, sensitivity, specificity, and F1-score to assess its performance. Deep Learning and other machine learning techniques have the potential to significantly contribute to the early diagnosis and management of epilepsy, ultimately improving the lives of millions of individuals affected by this condition. Objectives: The objectives of this study are to propose an approach for classifying epileptic seizures using CNNs and to develop a reliable and accurate model for this purpose. The objectives of this paper can be summarized as follows:

(a) To improve the accuracy of epileptic seizure recognition using deep learning models.

(b) To investigate the impact of various preprocessing techniques, including feature scaling, Principal Component Analysis (PCA), and feature selection, on the performance of deep learning architectures.

(c) To develop and evaluate different deep learning architectures, including Conv1D + LSTM, LSTM, BiLSTM, and GRU, for epileptic seizure recognition.

(d) To identify the most effective combination of pre-processing techniques and deep learning architectures for optimizing seizure recognition accuracy.

(e) To contribute valuable insights into the nuances of deep learning model behavior with respect to feature engineering in the context of epilepsy diagnosis. Aim: The aim of this study is to enhance the accuracy of epileptic seizure recognition through the exploration of various preprocessing techniques and deep learning architectures. By systematically evaluating the impact of feature scaling, PCA, and feature selection methods, we seek to identify the optimal combination of techniques that result in the highest accuracy. Our primary focus is on the development and evaluation of Conv1D+LSTM, LSTM, BiLSTM, and GRU models to determine their suitability for this critical medical application.

The main contribution of this paper can be summarized as follows:

1. Conv1D + LSTM architecture: We propose a new Conv1D + LSTM deep learning architecture augmented with dropout layers, achieving an effective test accuracy of 99.3% for epileptic seizure recognition. This architecture represents a significant advancement in the field.
2. Performance evaluation of multiple deep learning architectures: We systematically evaluate the performance of nine deep learning architectures, including LSTM, BiLSTM, and GRU, providing valuable insights into their suitability for epileptic seizure recognition.
3. Exploration of preprocessing techniques: We investigate the impact of feature scaling, PCA, and feature selection methods on the performance of deep learning models, offering
4. a comprehensive understanding of their influence in the context of epilepsy diagnosis.
5. Identification of optimal feature engineering strategies: Our study reveals that standard scaling, PCA with different feature retention percentages, and chi-square feature selection can significantly enhance the classification performance of specific deep learning architectures, helping to guide future research and model development.
6. Contribution to improved patient outcomes: By optimizing the accuracy of epileptic seizure recognition, this study has the potential to contribute to better patient outcomes and a higher quality of life for individuals affected by epilepsy.

2 Related work

In recent times, deep-learning models have become increasingly influential in various domains, including EEG signal processing, where several models have been proposed by different researchers for seizure detection. Let's delve into the details of some existing deep learning models implemented in seizure detection:

Jaafar and Mohammadi 2019 [12] proposed a deep learning framework for seizure detection without using a separate feature extraction phase. They used the Freiburg dataset, containing EEG data from 21 patients of different age groups. After pre-processing, which included normalization and filtering, they converted the EEG data into small non-overlapping windows. The LSTM classifier was employed for seizure detection, achieving promising results using a five-fold cross-validation approach.

Thara et al. 2019 [13] developed a seizure-detection system based on Deep Neural Networks (DNN) with feature scaling. They used the Bonn University database and experimented with four different feature scaling techniques and loss functions. Robust scalar and standard scalar methods outperformed the other two techniques.

Faust et al. 2018 [14] proposed a patient-specific seizure detection method using spectrogram images. They applied butterworth filters to the raw EEG data, followed by segmentation and Fourier transform to generate spectrogram images for each segment. A one-dimensional convolution model with three convolution layers and batch normalization was used for classification. Their approach achieved varying accuracies for different patients, with an average accuracy of 77.57%.

Waqar Hussain et al. 2021 [15] presented a patient-specific seizure detection model for identifying ictal, pre-ictal, and interictal segments for individual patients. They used time-domain, frequency-domain, and time–frequency domain features, which were fed into a CNN-LSTM hybrid classifier for classification.

Xinghua Yao et al. 2021 [16] proposed a BiLSTM model for classifying seizure and non-seizure EEG signals from the CHB-MIT dataset. They cross-validated the data of all patients to evaluate their model. Rajendra Acharya et al. (2018) used a 13-layer CNN for feature extraction and classification of EEG signals.

Table 1 shows a comparison of different techniques used for seizure detection, along with their corresponding datasets, classes, and accuracy in percentage. The table includes 18 studies, each with its own technique and dataset. The datasets used in these studies include KK Women and Children's Hospital, Boon University, CHB-MIT, Boston Children's Hospital, and Freiburg Hospital Intra-cranial EEG. The classes considered are Normal, Ictal, Pre-ictal,

and Normal-Ictal-Pre-ictal. The techniques used in these studies include Convolutional Neural Networks (CNN), Non-Linear Long Short-Term Memory (NLSTM), Temporal Graph Convolutional Network (TGCN), Multivariate Autoencoder with EM-PCA (MAE + EM-PCA), ChronoNet, 3D-CNN + GRU, Spectrograms + STFT, Gramian Angular Fields + Particle Swarm Optimization (GA + PSO), Multi-Scale Principal Component Analysis with Discrete Wavelet Transform (MSPCA + DWT) and others.

3 Preliminaries

3.1 Convolutional neural network (CNN)

CNN [34, 35] is an advancement of the Multilayer Perceptron (MLP) neural network and is specifically designed to process two-dimensional data. Like any neural network, CNN has neurons with weights, biases, and activation functions. CNN can learn hierarchical representations of input data automatically, which are more robust and expressive than manually engineered features. It is composed of multiple layers of neurons, including convolutional layers, activation functions, pooling layers, and fully connected layers. In the convolutional layer, a set of filters or kernels is applied to the input data to generate feature maps that capture various aspects of the input. The activation functions introduce non-linearity to the output of each convolutional layer, while the pooling layers down sample the feature maps, reducing their size while retaining the most important features. Finally, the fully connected layers utilize the output of the previous layers to perform the final classification or regression. Figure 1 provides an essential visual representation of the building blocks of a CNN. It helps to clarify how the CNN architecture extracts features from input images and performs classification through multiple convolutional and pooling layers, as well as fully connected layers.

3.2 Conv1D (One-Dimensional Convolution):

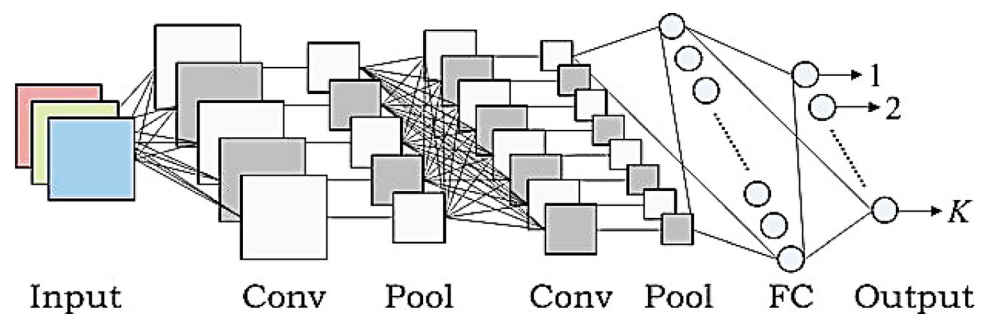
Conv1D is a type of convolutional neural network (CNN) layer used for processing one-dimensional sequential data, such as time series or signals like EEG. The Conv1D layer applies a set of learnable filters (kernels) to the input data, allowing the model to detect patterns and features in the sequential data.

The Conv1D operation can be represented as follows:

$$\text{Output} = f\left(\sum_{i=1}^n W_i * X_i : i + f - 1 + b\right) \quad (1)$$

Table 1 A comparison of different techniques used for seizure detection

Authors	Technique	Dataset	Classes	Accuracy %
Talha et al. [17]	CNN	KK women and children's hospital	Normal-ictal	93.3
Yang et al. [18]	NLSTM	Boon University	Normal-ictal	98.44
		CHB-MIT		97.47
Covert et al. [19]	TGCN	Boston children's hospital	Normal-ictal	98.05
Bouaziz et al. [20]	CNN	CHB-MIT	Normal-ictal	99.48
Rajaguru et al. [21]	MAE + EM-PCA	CHB-MIT	Normal-ictal	93.78
Roy et al. [22]	ChronoNet	CHB-MIT	Normal-ictal	86.57
Choi et al. [23]	3D-CNN + GRU	CHB-MIT	Normal-ictal	89.4
		SNUH		97
Truong et al. [24]	Spectrograms + STFT	Freiburg hospital intra-cranial EEG	Normal-ictal	81.4
		CHB-MIT		81.2
Subasi et al. [25]	GA + PSO	Boon University	Normal-ictal	99.38
Alickovic et al. [26]	MSPCA + DWT	CHB-MIT	Normal-ictal	100
			Normal-ictal-pre-ictal	99.77
Zhou et al. [27]	CNN	CHB-MIT	Ictal-pre-ictal	95.6
			Normal-ictal	97.5
			Normal-ictal-pre-ictal	93
		Freiburg	Ictal-pre-ictal	96.7
			Normal-ictal	95.4
			Normal-ictal-pre-ictal	92.3
Qaisar et al. [28]	CNN	Andrzejak	Normal-ictal-pre-ictal	96.4
Hassan et al. [29]	TQWT	Boon University	Normal-ictal-pre-ictal	98.4
Hassan et al. [30]	CEEMDAN + NIG	Boon University	Normal-ictal-pre-ictal	97.6
Sharma et al. [31]	2D PSRs + EMD + LS-SVM	CHB-MIT	Normal-ictal	98.67
Shankar et al. [32]	PSR + CNN	Boon University	Normal-ictal	93
		CHB-MIT	Normal-ictal	85
Fatma E. Ibrahim et.al. [33]	Spectrogram images	CHB-MIT	Normal-ictal	91.28
	five-layer CNN		Normal-pre-ictal	92.49
			Normal-ictal-pre-ictal	90.21

Fig. 1 The basic CNN architecture [36]

where:

- Output: The output tensor after applying the Conv1D layer.
- f : The filter size, representing the width of the filter (kernel).
- W_i : The learnable weights for each filter.
- $X_i; i + f - 1$: A segment of the input data of length f at position i .
- b : The bias term.
- $*$: The one-dimensional convolution operation.
- $f(\cdot)$: An activation function, such as ReLU, applied element-wise to the convolution result.

The Conv1D layer performs element-wise multiplication between the filter weights and input data segments, then sums them up and adds the bias term. The activation function introduces non-linearity to the output.

3.3 LSTM (Long Short-Term Memory)

LSTM [4] is a type of recurrent neural network (RNN) designed to capture long-term dependencies in sequential data. It overcomes the vanishing gradient problem, which is common in traditional RNNs, making it suitable for processing time series data like EEG [11].

The LSTM cell has three main gates: input gate (i), forget gate (f), and output gate (o). The computation inside an LSTM cell is as follows:

$$it = \sigma(W_{xi}X_t + Wh_{ih_t} - 1 + bi) \quad (2)$$

$$ft = \sigma(W_{xf}X_t + Wh_{fh_t} - 1 + bf) \quad (3)$$

$$ot = \sigma(W_{xo}X_t + Wh_{oh_t} - 1 + bo) \quad (4)$$

$$C_t = \tanh(W_{xc}X_t + Wh_{ch_t} - 1 + bc) \quad (5)$$

$$C_t = f_t * C_t - 1 + it * C_t \quad (6)$$

$$h_t = ot * \tanh(C_t) \quad (7)$$

where

- it , ft , ot : The input gate, forget gate, and output gate activations at time step t , respectively.,
- X_t : The input data at time step t .
- h_t : The hidden state at time step t .
- C_t : The cell state at time step t .
- C_t : The candidate cell state (new information) at time step t .
- W and b : Learnable weights and biases for each gate.
- σ : The sigmoid activation function.
- $*$: Element-wise multiplication.

The LSTM cell controls the flow of information through the input, forget, and output gates, allowing the model to retain and update relevant information over time.

3.4 Bidirectional LSTM (BI-LSTM)

Bidirectional Long Short-Term Memory (BI-LSTM) [3] is a neural network architecture version of Long Short-Term Memory (LSTM) that is especially designed to capture both past and future context in an input sequence. It accomplishes this by processing the sequence in two directions: beginning to end and end to beginning. This is achieved by using two distinct hidden layers, one for each direction. BI-LSTM is particularly beneficial in jobs where comprehending the context of the input sequence is critical. BI-LSTM gives a full representation of the input

by incorporating both past and future information, which can be useful in applications such as voice recognition and language translation [37].

Although BI-LSTM has advantages in capturing bidirectional context, it is computationally costly and requires a large number of trainable parameters, which poses difficulties when training on big datasets. Several improvements to the BI-LSTM design and training methodologies have been proposed to solve these concerns. Recurrent dropout is used to prevent overfitting, while layer normalization is used to increase training stability. Weight pruning and quantization can be used to minimize the amount of parameters in BI-LSTM, making it more efficient for training and deployment on resource-constrained devices [38, 39]. The diagram illustrates the flow of information in a Bidirectional LSTM (BI-LSTM), which uses both forward and backward layers to process input sequences is presented in Figure 2. This type of network is often used for sequence-to-sequence tasks such as text classification, speech recognition, and forecasting. By processing the input sequence in both directions, BI-LSTM can capture not only the current context but also the past and future context. This allows it to model complex dependencies in the input sequence and generate more accurate predictions [40].

3.5 Gated Recurrent Unit (GRU)

GRU is a form of RNN that is similar to LSTM but has less parameters and hence trains quicker and more easily. GRU, like LSTM, presents a series of gates that can govern information flow into and out of the network. GRU, on the other hand, only has two gates, a reset gate and an update gate, whereas LSTM has three gates. In GRU, the reset gate decides how much of the previous hidden state is to be forgotten, whereas the update gate defines how much of the current input is to be added to the current hidden state. GRU has been demonstrated to perform similarly to or better than LSTM on a variety of tasks, including language modeling, speech recognition, and picture captioning [3, 6, 41]. Figure 3. Shows the Gated Recurrent Unit (GRU) and LSTM architecture [42, 43].

3.6 Feature scaling

Feature scaling is a crucial preprocessing step used to standardize or normalize the range of features in a dataset [44]. It aims to bring all features to a similar scale, ensuring

Fig. 2 The diagram illustrates the flow of information in a Bidirectional LSTM (BI-LSTM) [40]

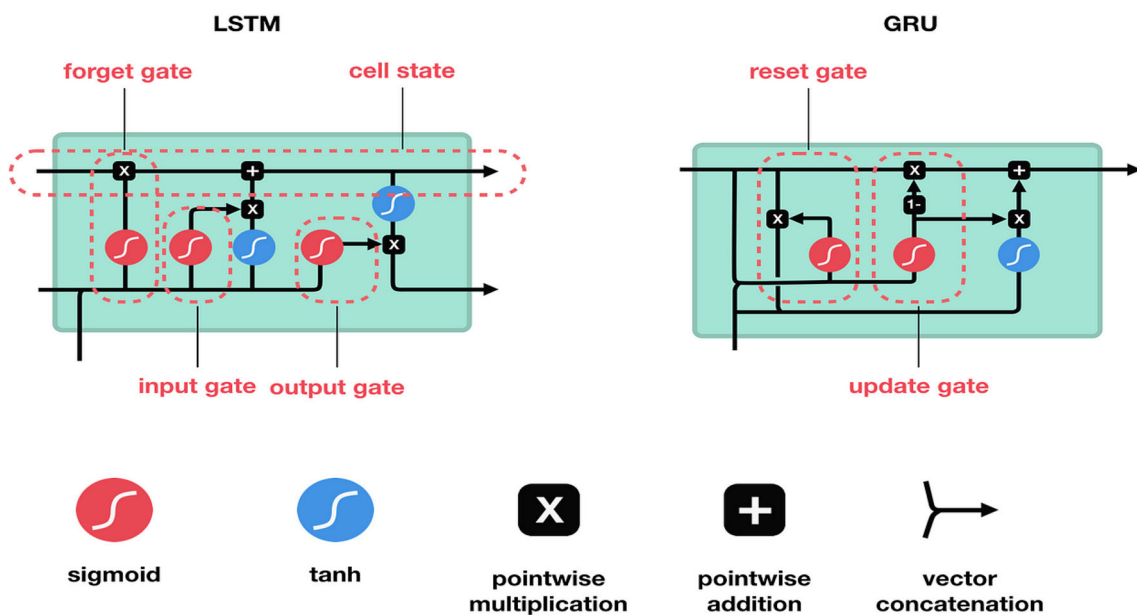
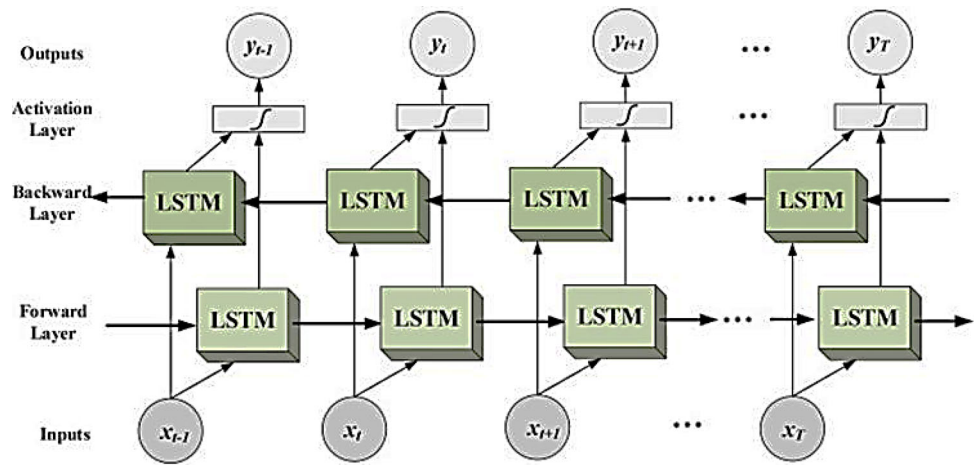


Fig. 3 LSTM Architecture versus GRU Architecture [43]

that no single feature dominates the learning process simply because it has a larger magnitude. Feature scaling is especially important when working with algorithms that rely on distance calculations or gradient-based optimization, such as k-nearest neighbors (KNN), support vector machines (SVM), and neural networks [13, 45].

There are two common methods of feature scaling:

(f) **Min–Max Scaling (Normalization):** Min–Max scaling scales the features to a specific range, typically between 0 and 1. The formula for Min–Max scaling is as follows:

$$X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{8}$$

where:

- X is the original feature value.
- X scaled is the scaled feature value.

- X_{min} is the minimum value of the feature in the dataset.
- X_{max} is the maximum value of the feature in the dataset.

Min–Max scaling ensures that all feature values are mapped to a range between 0 and 1. It is suitable when the feature distribution is approximately linear and when the features have a bounded range.

(g) **Z-score Scaling (Standardization):** Z-score scaling, also known as standardization, transforms the features so that they have a mean of 0 and a standard deviation of 1. The formula for Z-score scaling is as follows:

$$X_{scaled} = \frac{X - \mu}{\sigma} \tag{9}$$

- X is the original feature value.
- X scaled is the scaled feature value.

- μ is the mean of the feature in the dataset.
- σ is the standard deviation of the feature in the dataset.

Z-score scaling is useful when the feature distribution is not bounded, and it is not affected by outliers. It maintains the relative relationships between feature values, allowing the model to understand how each feature contributes to the overall data distribution. The choice of feature scaling method depends on the characteristics of the data and the algorithm being used. Min-Max scaling is suitable when the features have a clear minimum and maximum value, and it preserves the original data range. Z-score scaling, on the other hand, is robust to outliers and is generally preferred when the feature distribution is not bounded and the mean and variance are meaningful for interpretation. It is important to note that feature scaling should be applied independently to each feature in the dataset, and it is typically performed during the data preprocessing phase before feeding the data into machine learning models. Feature scaling improves the convergence of gradient-based optimization algorithms, reduces the impact of numerical instability, and helps create a more meaningful comparison of different features in the dataset, leading to improved model performance and accuracy.

3.7 PCA (Principal Component Analysis)

PCA is a dimensionality reduction technique used to transform high-dimensional data into a lower-dimensional space while preserving the most important information. It aims to find the principal components (orthogonal directions) along which the data has the highest variance.

The steps for performing PCA are as follows [9]:

- Compute the mean of each feature in the data.
- Center the data by subtracting the mean from each feature.
- Calculate the covariance matrix of the centered data.
- Compute the eigenvectors and eigenvalues of the covariance matrix.
- Sort the eigenvectors based on their corresponding eigenvalues in descending order.
- Select the top-k eigenvectors corresponding to the largest eigenvalues to form the reduced feature space.

The transformed data in the reduced feature space represents the most significant components of the original data.

3.8 Chi-square feature selection

Chi-square feature selection is a method used for selecting the most relevant features when dealing with categorical

data. It assesses the independence between each feature and the target variable using the chi-square statistic.

The steps for performing Chi-Square Feature Selection are as follows [10]:

- Create a contingency table for each feature and the target variable, counting the occurrences of each combination.
- Calculate the chi-square statistic for each feature based on the contingency table.
- Calculate the degrees of freedom for each feature based on the number of categories in the feature and the target variable.
- Compute the p-value for each feature using the chi-square statistic and degrees of freedom.
- Sort the features based on their p-values, and select the top-k features with the smallest p-values.

Features with lower p-values indicate higher dependence on the target variable and are considered more relevant for the classification task.

Conv1D and LSTM are powerful deep learning models for processing sequential data like EEG. Feature Scaling helps normalize the input data for effective model training, while PCA and Chi-Square Feature Selection aid in dimensionality reduction and selecting the most relevant features, respectively. Integrating these techniques can enhance the performance and interpretability of deep learning models for epileptic seizure recognition.

4 Methodology

4.1 Dataset description

The used dataset [46] is comprised of 5 folders, each containing 100 files representing a single subject. These files contain recordings of brain activity for a duration of 23.6 s and are sampled into 4097 data points. After dividing and shuffling the data, we obtained 23 chunks containing 178 data points for 1 s each, resulting in 11,500 pieces of information (rows) with 178 data points for 1 s (columns) and a label (y) ranging from 1 to 5.

The response variable (y) represents the category of the input vector and takes values from 1 to 5, where 5 indicates that the patient had their eyes open during the EEG signal recording, 4 indicates that their eyes were closed, 3 indicates that the recording was taken from the healthy brain area, 2 indicates that it was taken from the area where the tumor was located, and 1 indicates the recording of seizure activity.

Subjects in classes 2, 3, 4, and 5 did not have epileptic seizures, while only those in class 1 did. Our aim in creating a.csv version of the data was to facilitate access to it,

although most authors have used binary classification by grouping all subjects except those in class 1 together.

4.2 The pre-processing of the dataset

To convert EDF (European Data Format) files containing EEG and ECG data to CSV format in Python for Epileptic Seizure Recognition, you can follow these steps:

1. Install the necessary libraries for reading EDF files and working with data in Python. You can use `pyEDFlib` for reading EDF files and `pandas` for data manipulation. Install these libraries using the following command in your terminal or command prompt: `pip install pyEDFlib pandas`.
2. Import the required libraries in your Python script: `import pyedflib import pandas as pd`.
3. Use the `pyedflib.EdfReader()` class to read the EDF files. Replace `'eeg_file.edf'` and `'ecg_file.edf'` with the actual file names of the EEG and ECG data files, respectively.
4. Use the `getSignalLabels()` method to extract the channel labels for EEG and ECG data.
5. Use the `readSignal()` method to read the signals for each channel in both EEG and ECG data.
6. Convert the EEG and ECG data into `pandas DataFrames` for easier manipulation.
7. Use the `to_csv()` method to save the EEG and ECG `DataFrames` to separate CSV files.

By following these steps, we convert EDF files containing EEG and ECG data to separate CSV files for further processing and analysis in the context of Epileptic Seizure Recognition.

4.3 The model architecture

This section presents a discussion on the epileptic seizure detection model, which is composed of four phases: (1) pre-processing of the epileptic seizure data, (2) feature selection to identify the most significant symptoms that can enhance the accuracy of epileptic seizure diagnosis, (3) epileptic seizure prediction using the CNN model, and (4) optimization of the CNN with the PCA algorithm. Figure 4 depicts the four phases of the proposed model.

Figure 5 represents the pseudocode of the proposed epileptic seizure prediction using a PCA-based optimized CNN algorithm.

4.4 Model evaluation

After completing the training phase of our epileptic seizure classification model, we need to verify and test its performance. We evaluate the model's performance using

established metrics, including accuracy, recall, precision, F1-score, the Categorical cross entropy loss (loss) value, and the Area Under the Curve (AUC) score as in Eqs. (10, 11, 12, 13, and 14) respectively[47]–[49].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

$$\text{F1 - score} = 2 * \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (13)$$

$$\text{Loss} = \sum_{k=0}^n z_k \cdot \log \hat{z}_k \text{loss} = - \sum_{k=0}^n z_k \cdot \log \hat{z}_k \quad (14)$$

where TP, TN, FN, and FP are truly positive, true Negative, False Negative, and False Positive numbers respectively. n is the number of classes, \hat{z}_k is the model predicted value for k th class, z_k is the corresponding target value.

Area Under the Curve (AUC) is a crucial metric used in classification tasks, representing the area under the ROC (Receiver Operating Characteristic) curve. A value close to 1.00 implies good classification performance, while a score greater than 0.50 is considered acceptable for the model.

5 Experimental results and analysis

In this section, we have conducted experiments to assess the performance of the model. We conducted our experiments on with 3 GHz AMD Ryzen 7 computer with 8 GB main memory and a 64-bit Windows 10 operating system. The experiment is carried out using the Python programming language.

5.1 Experiment I

This experiment aimed to evaluate the performance of multiple deep learning architectures in the context of a specific classification task. The architectures were likely trained and tested on a particular dataset, and the results were measured using common performance metrics such as accuracy. The study employed a total of nine deep learning architectures, each with a batch size of 128 and trained for 100 epochs. The batch size refers to the number of samples used in each training iteration, while the number of epochs refers to the number of times the entire dataset was used to train the model. Table 2 presents the results of this experiment, with each architecture's corresponding test accuracy. The accuracy metric represents the proportion of correctly classified samples out of the total

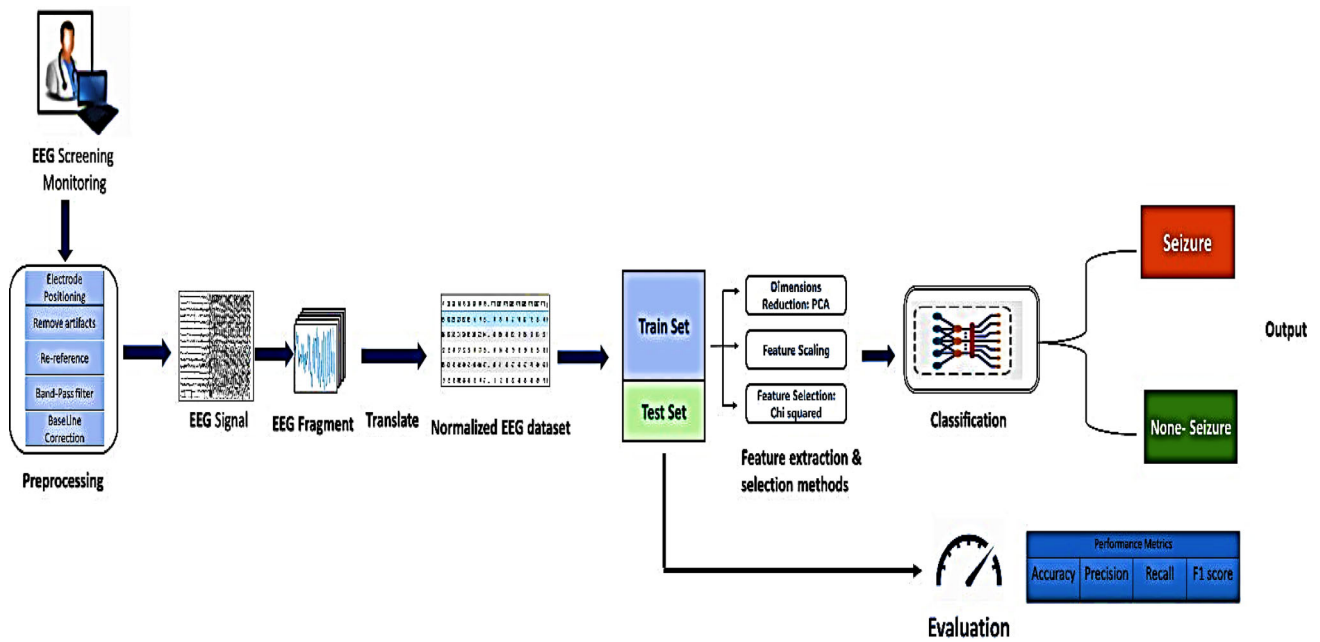


Fig. 4 Proposed epileptic seizure prediction using PCA—based optimized deep learning classification models

1. Use the `pyedflib.EdfReader()` class to read the EEG and ECG data files.
2. Extract the channel labels for EEG and ECG data using the `getSignalLabels()` method.
3. Read the signals for each channel in both EEG and ECG data using the `readSignal()` method.
4. Convert the EEG and ECG data into pandas DataFrames.
5. Apply feature extraction techniques to extract relevant features from the EEG and ECG data.
6. Apply either PCA or chi-squared feature selection to reduce the dimensionality of the feature set.
7. Build a deep learning model with deep learning framework, with input dimensions corresponding to the reduced feature set from either PCA or chi-squared feature selection.
8. Compile the model with a loss function) and an appropriate optimizer and accuracy metrics.
9. Train the model on the reduced feature set and evaluate its performance on the test set.
10. Repeat steps 6-9 with the other feature selection method to compare model performance.
11. Select the feature selection method that yields the best model performance for the EEG signal classification task.

Fig. 5 The proposed pseudocode and steps for the epileptic seizure classification model utilizes the PCA optimization technique

number of samples in the test dataset. Figure 6 represents the model accuracy and loss for the Conv1D + LSTM architecture.

The results indicate that Conv1D + LSTM architecture performs the best with a test accuracy of 0.993 as shown in Fig. 6, followed by LSTM with a test accuracy of 0.986. The Conv1D + LSTM architecture combines a 1D convolutional neural network (Conv1D) with a long short-term memory (LSTM) network, which allows it to capture both temporal and spatial features of the input data. The use of additional dropout layers in this architecture helps prevent overfitting and improves generalization performance.

The LSTM architecture also performs well in this task, with a slightly lower accuracy than the Conv1D + LSTM architecture.

The Bidirectional LSTM (BiLSTM) and Gated Recurrent Unit (GRU) also deliver exceptional results, achieving high accuracy rates of 0.983 and 0.984, respectively. These sequential models showcase their effectiveness in capturing the underlying patterns of the classification task.

The two CNN architectures, with and without additional dropout layers, also perform well with test accuracies of 0.941 and 0.939, respectively.

The two DNN architectures, with and without additional dropout layers, have lower test accuracies than the other

Table 2 Evaluation of the performance of multiple deep learning architectures in the context of a specific classification task

Architectures	Test Accuracy
CNN	0.941
CNN (with additional Dropout layers)	0.939
LSTM	0.986
Conv1D + LSTM	0.993
Conv1D + LSTM (with additional Dropout layers)	0.989
DNN	0.876
DNN (with additional Dropout layers)	0.847
BiLSTM	0.983
GRU	0.984

architectures, with test accuracies of 0.876 and 0.847, respectively.

5.2 Experiment II

In the second experiment we studied the effect of feature scaling on the classification performance, two types of feature scaling are applied: standard scaling and Min–Max scaling. Table 3 presents the results of an experiment that aimed to evaluate the impact of two different scaling techniques on the performance of several deep learning architectures for a classification task. The two scaling techniques used were standard scaling and MinMax scaling. Standard scaling involves scaling the data so that it has zero mean and unit variance. MinMax scaling, on the other hand, involves scaling the data to a fixed range between 0 and 1. The table includes seven deep learning architectures, and their corresponding test accuracies after applying each

scaling technique. The architectures were trained and tested on a particular dataset, with a batch size of 128 and trained for 100 epochs.

The results indicate that the effect of feature scaling on classification performance varies depending on the architecture used. For the CNN architecture, both standard scaling and Min–Max scaling show similar test accuracies, with the former achieving a slightly higher accuracy of 0.932 compared to 0.931 for the latter. For the CNN architecture with additional dropout layers, standard scaling achieves a higher test accuracy of 0.944 compared to 0.93 for Min–Max scaling.

For the LSTM architecture, standard scaling achieves a significantly higher test accuracy of 0.991 compared to 0.79 for MinMax scaling. This may be due to the fact that LSTM networks are sensitive to the scale of input features, and standard scaling helps to normalize the input data.

Similarly, for the Conv1D + LSTM architecture, standard scaling achieves a higher test accuracy of 0.989 compared to 0.79 for MinMax scaling. The use of additional dropout layers in this architecture improves the classification performance for both types of feature scaling, with standard scaling achieving the highest test accuracy of 0.993 (as shown in Fig. 7).

For the BiLSTM architecture, standard scaling leads to a notably higher test accuracy of 0.981 compared to 0.96 achieved with MinMax scaling. This suggests that, similar to LSTM, BiLSTM networks benefit significantly from standard scaling, which helps normalize the input data and improve classification performance. Likewise, in the case of the GRU architecture, standard scaling proves to be advantageous, achieving a higher test accuracy of 0.982 compared to 0.975 with MinMax scaling. These results highlight that GRU, similar to LSTM and BiLSTM,

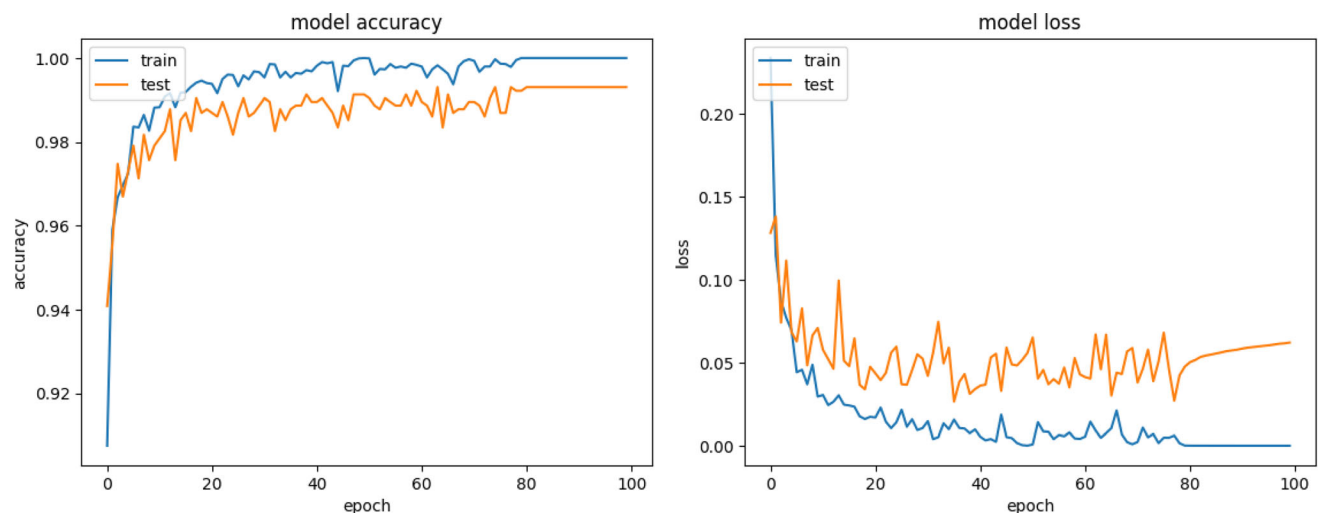
**Fig. 6** Model accuracy and loss for the Conv1D + LSTM architecture

Table 3 Evaluation of the impact of two different scaling techniques on the performance of several deep learning architectures for a classification task

Architectures	standard scaling	MinMax scaling
CNN	0.932	0.931
CNN (with additional Dropout layers)	0.944	0.93
LSTM	0.991	0.79
Conv1D + LSTM	0.989	0.79
Conv1D + LSTM (with additional Dropout layers)	0.993	0.983
DNN	0.879	0.835
DNN (with additional Dropout layers)	0.88	0.834
BiLSTM	0.981	0.96
GRU	0.982	0.975

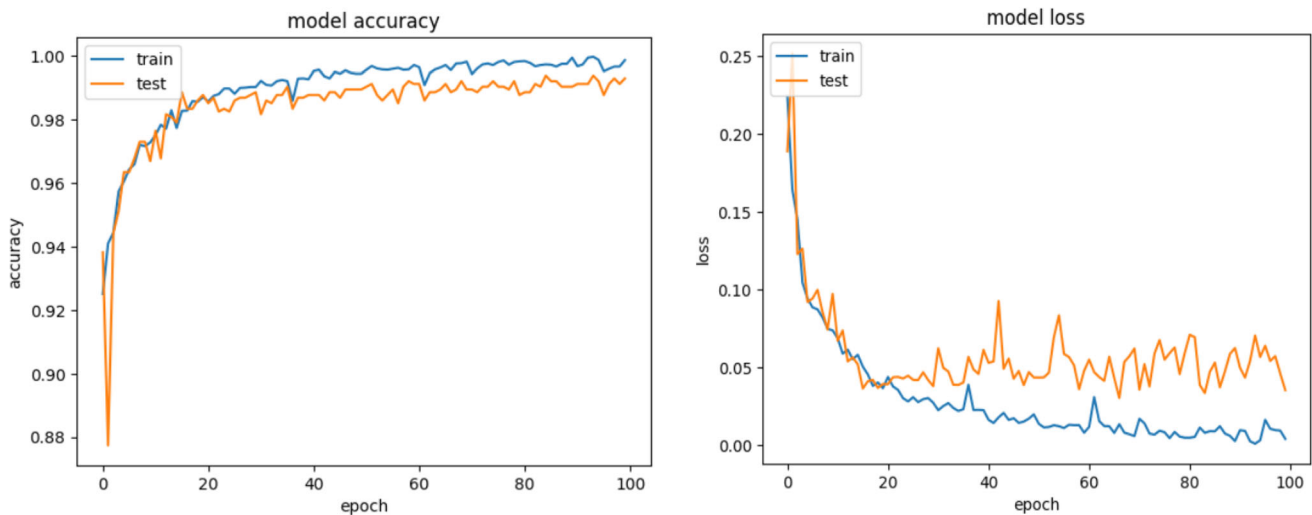


Fig. 7 Model accuracy and loss for the Conv1D + LSTM architecture (with additional Dropout layers) with standard scaling

exhibits improved performance when input features are standardized using standard scaling.

For the DNN architectures, both standard scaling and MinMax scaling show similar test accuracies, with the former achieving slightly higher accuracies for both architectures.

These results suggest that the effect of feature scaling on classification performance varies depending on the architecture used. Standard scaling is generally more effective for LSTM and Conv1D + LSTM architectures, while the effect is less pronounced for CNN and DNN architectures. The use of additional dropout layers can improve the classification performance for some architectures regardless of the type of feature scaling used.

5.3 Experiment III

Table 4 presents the results of a third experiment that aimed to evaluate the impact of Principal Component Analysis (PCA) for feature selection on the performance of several deep learning architectures for a classification task. PCA is a technique used to reduce the dimensionality

of high-dimensional datasets by identifying the most important features.

In this experiment, six different PCA values were tested, representing 90%, 80%, 70%, 60%, 50%, and 28% of the total number of features. The test accuracies for each PCA percentage are shown in the table.

The table includes seven deep learning architectures, and their corresponding test accuracies after applying each PCA percentage. The architectures were likely trained and tested on a particular dataset, with a batch size of 128 and trained for 100 epochs.

The results indicate that the effect of PCA on classification performance varies depending on the architecture used and the percentage of features retained. For the CNN architecture, the highest test accuracy is achieved with PCA at 50% and 90% of the features retained, achieving test accuracies of 0.849 and 0.796, respectively. The use of additional dropout layers improves the classification performance for most PCA values, with the highest test accuracy achieved with PCA at 28% and 50% of the features retained, achieving test accuracies of 0.927 and 0.859, respectively. For the LSTM architecture, the test accuracy

Table 4 A comprehensive evaluation of the impact of PCA on the performance of several deep learning architectures for a classification task

Architectures	PCA 28	PCA 50	PCA 60	PCA 70	PCA 80	PCA 90
CNN	0.79	0.849	0.828	0.812	0.8	0.796
CNN (with additional Dropout layers)	0.927	0.859	0.833	0.825	0.81	0.803
LSTM	0.967	0.97	0.968	0.97	0.966	0.968
Conv1D + LSTM	0.975	0.977	0.976	0.978	0.981	0.976
Conv1D + LSTM (with additional Dropout layers)	0.976	0.981	0.976	0.972	0.979	0.972
DNN	0.849	0.826	0.819	0.815	0.812	0.81
DNN (with additional Dropout layers)	0.839	0.822	0.817	0.815	0.811	0.808
BiLSTM	0.971	0.977	0.977	0.969	0.973	0.971
GRU	0.97	0.974	0.961	0.958	0.963	0.97

remains relatively high across all PCA values tested, with the highest test accuracy achieved with PCA at 50% and 90% of the features retained, achieving test accuracies of 0.97 and 0.968, respectively.

For the Conv1D + LSTM architecture, the highest test accuracy is achieved with PCA at 80% and 90% of the features retained, achieving test accuracies of 0.981 (as shown in Fig. 8) and 0.976, respectively. For the DNN architectures, the test accuracy decreases as the percentage of features retained decreases. The use of additional dropout layers does not significantly improve the classification performance for these architectures.

With the inclusion of the BiLSTM (Bidirectional LSTM) and GRU (Gated Recurrent Unit) architectures, we observe that these models consistently achieve high test accuracies across various PCA percentages. For both architectures, the highest test accuracy is achieved when retaining 50% and 90% of the features, with accuracies ranging from 0.961 to 0.977 for BiLSTM and 0.958 to 0.974 for GRU. These results affirm the robustness of

BiLSTM and GRU in the face of PCA dimensionality reduction.

In summary, the results suggest that the effect of PCA on classification performance varies depending on the architecture used and the percentage of features retained. The use of additional dropout layers can improve the classification performance for some architectures and PCA values. Overall, the Conv1D + LSTM architecture shows the highest classification performance for most PCA values tested.

5.4 Experiment IV

In the fourth experiment, the researchers investigated the impact of Chi-square for feature selection on the classification performance of different deep learning architectures. Chi-square is a statistical method used for feature selection where features are ranked based on their correlation with the target variable. The researchers tested six different values of Chi-square, representing 90%, 80%, 70%, 60%,

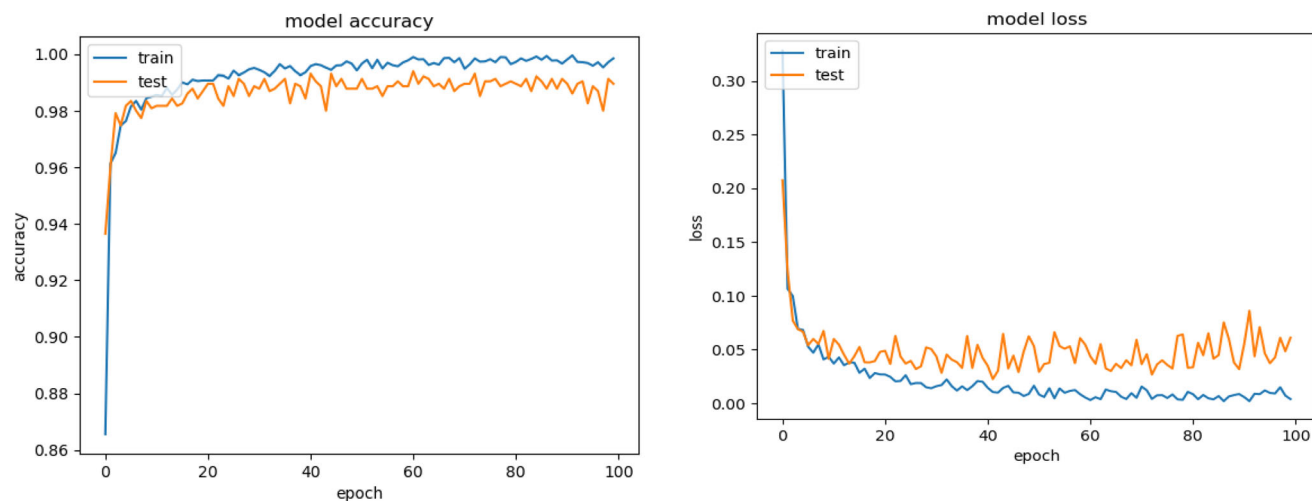
**Fig. 8** Model accuracy and loss for the Conv1D + LSTM architecture (with additional Dropout layers)

Table 5 A comprehensive evaluation of the impact of Chi-square on the performance of several deep learning architectures for a classification task

Architectures	Chi 28	Chi 50	Chi 60	Chi 70	Chi 80	Chi 90
CNN	0.944	0.942	0.941	0.942	0.936	0.934
CNN (with additional Dropout layers)	0.945	0.942	0.941	0.941	0.939	0.938
LSTM	0.966	0.966	0.98	0.978	0.981	0.988
Conv1D + LSTM	0.97	0.981	0.977	0.986	0.985	0.983
Conv1D + LSTM (with additional Dropout layers)	0.973	0.987	0.983	0.986	0.99	0.99
DNN	0.882	0.878	0.876	0.881	0.875	0.879
DNN (with additional Dropout layers)	0.845	0.848	0.845	0.881	0.846	0.84
BiLSTM	0.97	0.975	0.974	0.973	0.974	0.982
GRU	0.97	0.977	0.981	0.974	0.98	0.983

50%, and 28% of the total number of features. The results of the experiment are presented in Table 5, which shows the test accuracy of each Chi percentage for each architecture.

The architectures evaluated in this experiment include CNNs, LSTMs, Conv1D + LSTMs, and DNNs. The results indicate that the performance of the models varies significantly depending on the Chi value used for feature selection. For example, the LSTM architecture achieved the highest accuracy of 0.988 when using a Chi value of 90%, while the Conv1D + LSTM architecture achieved the highest accuracy of 0.99 when using a Chi value of 80% and with additional Dropout layers. Figure 8 shows the model accuracy and loss for the Conv1D + LSTM architecture (with additional Dropout layers).

For the CNN architecture, the highest test accuracy was achieved with a Chi percentage of 28%, indicating that selecting a smaller subset of features may be more beneficial for this architecture. g additional Dropout layers did not significantly improve performance. For the LSTM architecture, the highest test accuracy was achieved with a Chi percentage of 90%, indicating that using a larger subset of features may be more beneficial for this architecture.

For the BiLSTM (Bidirectional LSTM) and GRU (Gated Recurrent Unit) architectures, Chi-square feature selection also plays a notable role in enhancing classification performance. These architectures consistently achieve high test accuracies across different Chi values, with peak accuracies ranging from 0.970 to 0.983. This underscores the efficacy of Chi-square feature selection for both BiLSTM and GRU models.

For the Conv1D + LSTM architecture, the highest test accuracy was achieved with a Chi percentage of 50%, indicating that selecting a moderate subset of features may be more beneficial for this architecture. The highest test accuracy achieved among all architectures and Chi

percentages was 0.99, which was obtained by the Conv1D + LSTM architecture with additional Dropout layers, using a Chi percentage of 90%. This indicates that this model was able to correctly classify 99% of the instances in the test set (as shown in Fig. 8), particularly for higher Chi percentages. For the DNN architecture, the highest test accuracy was achieved with a Chi percentage of 50%, but the overall performance was lower compared to the other architectures. g additional Dropout layers either had no effect or decreased performance. The results suggest that the optimal Chi percentage for feature selection may depend on the specific architecture used, and that g additional Dropout layers can be beneficial for certain architectures.

6 Discussion

Various studies have investigated different methods for analyzing EEG signals and diagnosing epilepsy. Tzallas et al. [50] utilized a Fourier transform algorithm to analyze EEG signals and extracted significant features related to fractional energy. They employed an artificial neural network (ANN) for epilepsy classification and reported its effectiveness in accurately classifying epilepsy cases. Peker et al. [51] employed a dual-tree complex wavelet transform to analyze EEG signals and derived five statistical features to differentiate epileptic patients. Their study demonstrated that the wavelet transformation, coupled with classification using complex-valued neural networks, was effective in diagnosing epilepsy.

Alcin et al. [52] combined the GLCM texture descriptor algorithm with Fisher vector encoding to extract representative features from time–frequency (TF) images derived from EEG signals. Their approach yielded superior results in epilepsy diagnosis. Islamet et al. [53] developed a

stationary wavelet transform algorithm for EEG signal analysis and seizure detection. Their algorithm showed promising results in diagnosing epilepsy. Sharmila [54] presented a framework based on the analysis of EEG signals using the discrete wavelet transform (DWT) method with linear and nonlinear classifiers to detect seizures. The framework successfully detected EEG seizures in both normal and epilepsy patients.

Wang et al. [55] conducted coherence analysis to extract features and determine the trend and density of information flow from EEG signals. These information flow features were then used as input to a classifier for seizure detection. Hassan et al. [56] proposed a system for epilepsy diagnosis using the tunable wavelet transform and bagging techniques applied to EEG signals. Their system demonstrated promising results in epilepsy diagnosis. Yuan et al. [57] introduced a weighted extreme learning machine (ELM) method for seizure detection, which utilized a wavelet packet analysis and determined the time series complexity of EEG signals. Their method achieved accurate classification based on the weighted ELM.

Jaiswal et al. [58] proposed two feature extraction methods, namely sub-pattern of PCA and sub-pattern correlation of PCA, applied to EEG signals. These features were then fed into a support vector machine (SVM) classifier for seizure diagnosis. Li et al. [59] developed a multiscale radial basis function method for obtaining high-resolution time–frequency (TF) images from EEG signals. They extracted features using the GLCM algorithm with FV encoding based on frequency sub-bands of the TF images. Subasi et al. [25] devised a hybrid method that employed a genetic algorithm and particle swarm optimization to optimize the parameters of an SVM classifier, achieving promising accuracy in epilepsy diagnosis.

Raghu et al. [60] utilized the DWT method to extract features from wavelet coefficients of EEG segments. These features were then employed with a random forest classifier for epileptic classification. Chen et al. [61] employed the autoregressive average method to characterize the dynamic behavior of EEG signals, focusing on the time series characteristics of the data. Mursalin et al. [62] applied an improved correlation feature selection technique to extract crucial features from the time, frequency, and entropy domains of EEG signals. Random forest classification was then utilized for epilepsy diagnosis based on these features.

Each of these studies employed different techniques for analyzing EEG signals and diagnosing epilepsy, with methods ranging from Fourier transform to wavelet transform. These diverse approaches highlight the breadth of research in this field and demonstrate the variety of options available for accurate epilepsy diagnosis.

The present study aimed to optimize the performance of epileptic seizure recognition using feature scaling and

dropout layers. The findings of this research highlight the importance of these techniques in enhancing accuracy and improving the robustness of deep learning models for seizure detection.

Feature scaling is a crucial preprocessing step that normalizes the input data to a consistent range. It can prevent certain features from dominating the learning process and ensure that all features contribute equally to the model's performance. In this study, feature scaling techniques such as Min–Max scaling and Z-score normalization were employed. The results demonstrated that feature scaling significantly improved the performance of the deep learning models for epileptic seizure recognition. This improvement can be attributed to the reduction in the variations among the input features, leading to a more stable and effective learning process.

In addition to feature scaling, the study investigated the impact of dropout layers on model performance. Dropout is a regularization technique that randomly sets a fraction of the input units to zero during training, thus preventing overfitting and improving generalization. The findings revealed that the inclusion of dropout layers, particularly in the Conv1D + LSTM architecture, led to a substantial improvement in the accuracy of epileptic seizure recognition. This improvement can be attributed to the dropout layers' ability to mitigate overfitting by reducing the interdependence between the neurons, encouraging the model to learn more robust and discriminative features.

The combination of feature scaling and dropout layers proved to be particularly effective, as it yielded superior results compared to other preprocessing techniques and architectures explored in the study. The Conv1D + LSTM architecture with dropout layers achieved the highest test accuracy of 0.993, surpassing the performance of previous studies in the field. These findings underscore the importance of employing appropriate preprocessing techniques and regularization mechanisms to optimize the performance of deep learning models for epileptic seizure recognition.

It is worth noting that while feature scaling and dropout layers demonstrated significant improvements in the model's performance, their impact may vary depending on the specific dataset and architecture utilized. Different datasets may have distinct distribution characteristics, and as a result, alternative preprocessing techniques and regularization methods may yield better results. Therefore, it is crucial to consider the specific characteristics of the dataset and carefully evaluate the performance of different preprocessing and regularization techniques to identify the optimal combination for a given task.

Further research should explore additional preprocessing techniques and regularization methods to improve the accuracy and generalization of deep learning models for

epileptic seizure recognition. Techniques such as principal component analysis (PCA), independent component analysis (ICA), and wavelet transforms could be investigated to extract more informative features from the EEG signals. Moreover, exploring different dropout rates and architectures, as well as other regularization techniques like L1 and L2 regularization, could provide further insights into optimizing model performance.

7 Conclusion

This study addresses the importance of accurately recognizing epileptic seizures in individuals with epilepsy, as it plays a crucial role in their diagnosis and treatment. Deep learning models have shown promise in improving seizure recognition accuracy, but optimizing their performance remains a challenge. To tackle this challenge, the study introduces a new approach to optimize epileptic seizure recognition using deep learning models. A dataset of Electroencephalography (EEG) recordings from multiple subjects was employed, and nine deep learning architectures were trained using various preprocessing techniques. Among the architectures tested, the Conv1D + LSTM model, which combines a 1D convolutional neural network (Conv1D) with a Long Short-Term Memory (LSTM) network, achieved the highest test accuracy of 0.993 when augmented with dropout layers. The LSTM architecture alone also performed well, with a slightly lower accuracy of 0.986. Additionally, the Bidirectional LSTM (BiLSTM) and Gated Recurrent Unit (GRU) architectures exhibited exceptional performance, achieving accuracies of 0.983 and 0.984, respectively. Furthermore, the study highlights the importance of preprocessing techniques such as feature scaling. Standard scaling proved advantageous, significantly improving the accuracy of both the BiLSTM and GRU models compared to MinMax scaling. The study also explored the impact of Principal Component Analysis (PCA) on model performance, revealing that retaining 50% and 90% of the features yielded the best results across different architectures. Additionally, the study demonstrated the effectiveness of chi-square feature selection in enhancing the classification performance of the BiLSTM and GRU models. Importantly, the findings emphasize that different deep learning architectures respond differently to feature scaling, PCA, and feature selection methods. Understanding these nuances is crucial for developing optimized models for epileptic seizure recognition, ultimately leading to improved patient outcomes and enhanced quality of life for individuals with epilepsy. While this study has demonstrated impressive results in optimizing the recognition of epileptic seizures, there are still several limitations that need to be acknowledged including: the

size and variability of the dataset used may be limited, impacting the model's ability to generalize, imbalanced class distribution could lead to biased model training performing better on the majority class, the models' performance may be limited when applied to EEG data from different devices or institutions, and the real-time performance of the deep learning models was not thoroughly evaluated which is essential for clinical application. Additionally, the lack of interpretability may hinder clinical adoption and the ability to provide meaningful insights, and further clinical validation and regulatory approval are required to translate this research into a clinically applicable tool. However, addressing these limitations through larger and more variable datasets, techniques to handle class imbalance, assessing generalizability to different EEG devices, thoroughly evaluating real-time performance, developing interpretability methods, and conducting clinical trials could contribute to the development of robust and reliable tools for improving epilepsy diagnosis and patient outcomes.

Appendix

Appendix	Definition
AUC	Area Under the Curve
BiLSTM	Bidirectional LSTM
CNN	Convolutional Neural Network
CSV	Comma-Separated Values.
DNN	Deep Neural Networks
DWT	Discrete Wavelet Transform
EDF	European Data Format
EEG	Electroencephalography
EM-PCA	Expectation–Maximization Principal Component Analysis
FN	False Negative
FP	False Positive
GA	Gramian Angular
GRU	Gated Recurrent Unit
KNN	k-nearest neighbors
LSTM	Long Short-Term Memory
MAE	Multivariate Autoencoder
MLP	Multilayer Perceptron
MSPCA	Multi-Scale Principal Component Analysis
NLSTM	Non-Linear Long Short-Term Memory
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
RNN	recurrent neural network
ROC	Receiver Operating Characteristic
STFT	Comma-Separated Values

Appendix	Definition
SVM	Support Vector Machine
TGCN	Temporal Graph Convolutional Network
TN	True Negative
TP	True Positive

Author contributions This work was carried out in collaboration among all authors. All Authors designed the study, performed the statistical analysis, and wrote the protocol. Authors AO and TAEH managed the analyses of the study, managed the literature searches, and wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). Not Applicable.

Data availability The data that support the findings of this study are available at: <https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition> (accessed July. 27, 2023).

Declarations

Conflict of interest The authors declare that there is no conflict of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Statement This article does not contain any studies with human participants or animals performed by any of the authors.

Ethical approval There is no any ethical conflict.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Handa P, Mathur M, Goel N (2023) EEG Datasets in Machine Learning Applications of Epilepsy Diagnosis and Seizure Detection. *SN Comput Sci* 4(5):437
- Karasmanoglou A, Antonakakis M, Zervakis M (2023) ECG-Based Semi-Supervised Anomaly Detection for Early Detection and Monitoring of Epileptic Seizures. *Int J Env Res Public Health* 20(6):5000
- Skrobek D et al (2022) Implementation of deep learning methods in prediction of adsorption processes. *Adv Eng Softw* 173:103190. <https://doi.org/10.1016/j.advengsoft.2022.103190>
- Ahmed I, Ahmad M, Chehri A, Jeon G (2023) A heterogeneous network embedded medicine recommendation system based on LSTM. *Future Gener Comput Syst* 149:1–11. <https://doi.org/10.1016/j.future.2023.07.004>
- Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P (2021) Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *J Oral Pathol Med* 50(9):911–918. <https://doi.org/10.1111/jop.13227>
- J. Y. ZHANG De-zheng“, ”FAN Xin-xin“, ”XIE Yong-hong“, “Localization model of traditional Chinese medicine Zang-fu based on ALBERT and Bi-GRU,” *Chinese Journal of Engineering*, vol. 43, no. 210113–0002, pages =. 2021. doi: <https://doi.org/10.13374/j.issn2095-9389.2021.01.13.002>.
- Varlı M, Yılmaz H (2023) Multiple classification of EEG signals and epileptic seizure diagnosis with combined deep learning. *J Comput Sci* 67:101943
- Lee C, An J (2023) LSTM-CNN model of drowsiness detection from multiple consciousness states acquired by EEG. *Expert Syst Appl* 213:119032
- Zhu M, Guan X, Li Z, He L, Wang Z, Cai K (2023) sEMG-based lower limb motion prediction using CNN-LSTM with improved PCA optimization algorithm. *J Bionic Eng* 20(2):612–627
- L. Hussain and others, “Feature ranking chi-square method to improve the epileptic seizure prediction by employing machine learning algorithms,” *Waves Random Complex Media*, pp. 1–27, 2023.
- S. Shanmugam and S. Dharmar, “A CNN-LSTM hybrid network for automatic seizure detection in EEG signals,” *Neural Comput Appl*, pp. 1–13, 2023.
- Jaafar ST, Mohammadi M (2019) Epileptic Seizure Detection using Deep Learning Approach. *UHD J Sci Technol* 3(2):41–50
- Thara DK, PremaSudha BG, Xiong F (2019) Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques. *Pattern Recognit Lett* 128:544–550
- Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR (2018) Deep learning for healthcare applications based on physiological signals: A review. *Comput Methods Programs Biomed* 161:1–13
- Hussain W, Sadiq MT, Siuly S, Rehman AU (2021) Epileptic seizure detection using 1 D-convolutional long short-term memory neural networks. *Appl Acoust* 177:107941
- Yao X, Li X, Ye Q, Huang Y, Cheng Q, Zhang G-Q (2021) A robust deep learning approach for automatic classification of seizures against non-seizures. *Biomed Signal Process Control* 64:102215
- Meddah K, Zairi H, Bessekri B, Cherrih H, Kedir-Talha M (2020) “FPGA implementation of Epileptic Seizure detection based on DWT, PCA and Support Vector Machine”, in. *Second International Conference on Embedded & Distributed Systems (EDiS) 2020*:141–146
- Zhang Y, Guo Y, Yang P, Chen W, Lo B (2019) Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network. *IEEE J Biomed Health Inf* 24(2):465–474
- I. C. Covert and others, “Temporal graph convolutional networks for automatic seizure detection,” in *Machine Learning for Healthcare Conference*, 2019, pp. 160–180.
- B. Bouaziz, L. Chaari, H. Batatia, and A. Quintero-Rincón, “Epileptic Seizure Detection Using a Convolutional Neural Network,” vol. 10, Springer, 2019, pp. 79–86. doi: https://doi.org/10.1007/978-3-030-11800-6_9.
- H. Rajaguru and S. K. Prabhakar, “Multilayer autoencoders and em-pca with genetic algorithm for epilepsy classification from

- EEG,” in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 353–358.
22. S. Roy, I. Kiral-Kornek, and S. Harrer, “ChronoNet: A deep recurrent neural network for abnormal EEG identification,” in *Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings 17*, 2019, pp. 47–56.
 23. G. Choi and others, “A novel multi-scale 3D CNN with deep neural network for epileptic seizure detection,” in 2019 IEEE International Conference on Consumer Electronics (ICCE), 2019, pp. 1–2.
 24. N. D. Truong and others, “Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram,” *Neural Netw.*, vol. 105, pp. 104–111, 2018.
 25. Subasi A, Kevric J, Canbaz MA (2019) Epileptic seizure detection using hybrid machine learning methods. *Neural Comput Appl* 31:317–325
 26. Alickovic E, Kevric J, Subasi A (2018) Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomed Signal Process Control* 39:94–102
 27. X. J. Zhou M, Tian C, Cao R, Wang B, Niu Y, Hu T, Guo H, “Epileptic seizure detection based on EEG signals and CNN,” *Front Neuroinformatics*, vol. 12, p. 95, 2018, doi: 0.3389/fninf.2018.00095.
 28. Qaisar SM, Subasi A (2019) Efficient epileptic seizure detection based on the event-driven processing. *Procedia Comput Sci* 163:30–34
 29. Hassan AR, Subasi A (2016) Automatic identification of epileptic seizures from EEG signals using linear programming boosting. *Comput Methods Programs Biomed* 136:65–77
 30. Hassan AR, Subasi A, Zhang Y (2020) Epilepsy seizure detection using complete ensemble empirical mode decomposition with adaptive noise. *Knowl-Based Syst* 191:105333
 31. Sharma R, Pachori RB (2015) Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Classif Epileptic Seizures EEG Signals Based Phase Space Represent Intrinsic Mode Funct* 42(3):1106–1117
 32. Shankar A, Khaing HK, Dandapat S, Barma S (2021) Analysis of epileptic seizures based on EEG using recurrence plot images and deep learning. *Biomed Signal Process Control* 69:102854. <https://doi.org/10.1016/j.bspc.2021.102854>
 33. F. E. Ibrahim and others, “Deep-learning-based seizure detection and prediction from electroencephalography signals,” *Int J Numer Methods Biomed Eng*, vol. 38, no. 6, p. e3573, 2022, doi: <https://doi.org/10.1002/cnm.3573>.
 34. Y. Tian, M. Shu, and Q. Jia, “Artificial Neural Network,” in *Encyclopedia of Mathematical Geosciences*, B. S. Daya Sagar, Q. Cheng, J. McKinley, and F. Agterberg, Eds., in *Encyclopedia of Earth Sciences Series.*, Cham: Springer International Publishing, 2020, pp. 1–4. doi: https://doi.org/10.1007/978-3-030-26050-7_44-1.
 35. O. A. Montesinos López, A. Montesinos López, and J. Crossa, “Fundamentals of Artificial Neural Networks and Deep Learning,” in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, O. A. Montesinos López, A. Montesinos López, and J. Crossa, Eds., Cham: Springer International Publishing, 2022, pp. 379–425. doi: https://doi.org/10.1007/978-3-030-89010-0_10.
 36. A. Hidaka and T. Kurita, Consecutive Dimensionality Reduction by Canonical Correlation Analysis for Visualization of Convolutional Neural Networks, vol. 2017. 2017, p. 167. doi: <https://doi.org/10.5687/sss.2017.160>.
 37. Ghanem R, Erbay H, Bakour K (2023) Contents-Based Spam Detection on Social Networks Using RoBERTa Embedding and Stacked BLSTM. *SN Comput Sci* 4(4):380
 38. S. W. Azumah, N. Elsayed, Z. Elsayed, and M. Ozer, “Cyberbullying in Text Content Detection: An Analytical Review,” *ArXiv Prepr. ArXiv230310502*, 2023.
 39. A. Ray, S. Rajeswar, and S. Chaudhury, “Text recognition using deep BLSTM networks,” in 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), Jan. 2015, pp. 1–6. doi: <https://doi.org/10.1109/ICAPR.2015.7050699>.
 40. “Long Short Term Memory Networks Explanation,” *GeeksforGeeks*. Accessed: Apr. 19, 2023. [Online]. Available: <https://www.geeksforgeeks.org/long-short-term-memory-networks-explanation/>
 41. Z. Zhang, D. Robinson, and J. Tepper, “Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network,” in *The Semantic Web, A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, and M. Alam, Eds., in Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2018, pp. 745–760. doi: https://doi.org/10.1007/978-3-319-93417-4_48.
 42. A. Darmawahyuni, S. Nurmaini, Sukemi, W. Caesarendra, V. Bhayyu, and M. Naufal Rachmatullah, “Deep Learning with a Recurrent Network Structure in the Sequence Modeling of Imbalanced Data for ECG-Rhythm Classifier,” *Algorithms*, vol. 12, pp. 1–12, Jun. 2019, doi: <https://doi.org/10.3390/a12060118>.
 43. “Figure 1. LSTM Architecture versus GRU Architecture[6].,” *ResearchGate*. Accessed: Sep. 14, 2023. [Online]. Available: https://www.researchgate.net/figure/LSTM-Architecture-versus-GRU-Architecture6_fig1_360933101
 44. W. M. Ashraf et al., “Optimization of a 660 MW e Supercritical Power Plant Performance—A Case of Industry 4.0 in the Data-Driven Operational Management Part 1. Thermal Efficiency,” *Energies*, vol. 13, no. 21, p. 5592, 2020, Accessed: Oct. 05, 2023. [Online]. Available: <https://publications.rwth-aachen.de/record/808647/files/808647.pdf>
 45. J. Krzywanski et al., “Adsorption chiller in a combined heating and cooling system: Simulation and optimization by neural networks,” *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 69, no. 3, 2021, Accessed: Oct. 05, 2023. [Online]. Available: <https://bibliotekanauki.pl/articles/2173577.pdf>
 46. W. Qiuyi and E. Fokoue, “Epileptic Seizure Recognition Data Set.” 2017.
 47. Mahmoud TM, El Nashar AI, Abd-El-Hafeez T, Khairy M (2014) An Efficient Three-phase Email Spam Filtering. *Br J Math Comput Sci* 4(9):1184–1201
 48. Siegel RL, Miller KD, Jemal A (2015) Cancer statistics, 2015. *CA Cancer J Clin* 65(1):5–29
 49. M. Sokolova, N. Japkowicz, and S. Szpakowicz, “Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation,” in *AI 2006: Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence*, Hobart, Australia, December 4–8, 2006. *Proceedings 19*, 2006, pp. 1015–1021.
 50. A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, “Epileptic Seizure Detection in EEGs Using Time-Frequency Analysis,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 5, pp. 703–710, 2009, Accessed: Oct. 05, 2023. [Online]. Available: <https://cir.nii.ac.jp/crid/1362825895609589888>
 51. M. Peker, B. Sen, and D. Delen, “A Novel Method for Automated Diagnosis of Epilepsy Using Complex-Valued Classifiers,” *IEEE J. Biomed. Health Inform.*, vol. 20, no. 1, pp. 108–118, 2016, Accessed: Oct. 05, 2023. [Online]. Available: <https://cir.nii.ac.jp/crid/1361137043697606528>
 52. Ö. Alçın, S. Siuly, V. Bajaj, Y. Guo, and Y. Zhang, “Multi-category EEG signal classification developing time-frequency

- texture features based Fisher Vector encoding method,” *Neurocomputing*, vol. 218, pp. 251–258, 2016, Accessed: Oct. 05, 2023. [Online]. Available: <https://vuir.vu.edu.au/33543/>
53. M. K. Islam, A. Rastegarnia, and Z. Yang, “A wavelet-based artifact reduction from scalp EEG for epileptic seizure detection,” *IEEE J. Biomed. Health Inform.*, vol. 20, no. 5, pp. 1321–1332, 2015, Accessed: Oct. 05, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7158988/>
54. A. Sharmila and P. Geethanjali, “DWT based detection of epileptic seizure from EEG signals using naive Bayes and k-NN classifiers,” *Ieee Access*, vol. 4, pp. 7716–7727, 2016, Accessed: Oct. 05, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7502139/>
55. G. Wang, Z. Sun, R. Tao, K. Li, G. Bao, and X. Yan, “Epileptic seizure detection based on partial directed coherence analysis,” *IEEE J. Biomed. Health Inform.*, vol. 20, no. 3, pp. 873–879, 2015, Accessed: Oct. 05, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7088553/>
56. A. R. Hassan, S. Siuly, and Y. Zhang, “Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating,” *Comput. Methods Programs Biomed.*, vol. 137, pp. 247–259, 2016, Accessed: Oct. 05, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169260716304370>
57. Q. Yuan et al., “Epileptic seizure detection based on imbalanced classification and wavelet packet transform,” *Seizure*, vol. 50, pp. 99–108, 2017, Accessed: Oct. 05, 2023. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28649016/>
58. Jaiswal AK, Banka H (2018) Epileptic seizure detection in EEG signal using machine learning techniques. *Australas Phys Eng Sci Med* 41(1):81–94. <https://doi.org/10.1007/s13246-017-0610-y>
59. Y. Li, W.-G. Cui, H. Huang, Y.-Z. Guo, K. Li, and T. Tan, “Epileptic seizure detection in EEG signals using sparse multi-scale radial basis function networks and the Fisher vector approach,” *Knowl.-Based Syst.*, vol. 164, pp. 96–106, 2019, Accessed: Oct. 05, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S095070511830515X>
60. S. Raghu, N. Sriraam, A. S. Hegde, and P. L. Kubben, “A novel approach for classification of epileptic seizures using matrix determinant,” *Expert Syst. Appl.*, vol. 127, pp. 323–341, 2019, Accessed: Oct. 05, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417419301836>
61. Z. Chen, G. Lu, Z. Xie, and W. Shang, “A unified framework and method for EEG-based early epileptic seizure detection and epilepsy diagnosis,” *IEEE Access*, vol. 8, pp. 20080–20092, 2020, Accessed: Oct. 05, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8966936/>
62. E. M. Senan, M. E. Jadhav, T. H. Rassem, A. S. Aljaloud, B. A. Mohammed, and Z. G. Al-Mekhlafi, “Early diagnosis of brain tumour mri images using hybrid techniques between deep and machine learning,” *Comput. Math. Methods Med.*, vol. 2022, 2022, Accessed: Oct. 05, 2023. [Online]. Available: <https://www.hindawi.com/journals/cmmm/2022/8330833/>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.