Elevated Seizure Detection Performance Using Advanced EEG Signal Processing

MILIND NATU [1*], MRINAL BACHUTE [2], KETAN KOTECHA[3]

[1*]Research Scholar, Department of Electronics and Telecommunication, Symbiosis Institute of Engineering and Technology, Lavale, Pune, India.

[2]Associate Professor, Department of Electronics and Telecommunications, Symbiosis Institute of Engineering and Technology, Lavale, Pune, India.

[3]Department of Electronics and Telecommunications, Symbiosis Institute of Engineering and Technology, Lavale, Pune, India.

ORCID ID: 0000-0002-4080-4949

Abstract:

Recurrent seizures profoundly impact well-being, necessitating timely detection for effective management. Deep learning, particularly with Electroencephalogram (EEG) signals, holds promise for automated seizure detection. EEG directly captures brain electrical activity, aiding early intervention and treatment monitoring. Challenges include limited data and variability in seizure patterns. Key contributions involve advanced feature extraction, innovative selection algorithms, and deep learning models. This study proposes a comprehensive approach to advance EEG-based seizure detection with deep learning method.

Methods:

The study initially focuses on feature extraction, utilizing a range of techniques including Power Spectral Density (PSD), Weighted Raised Cosine-Window-based Short-Time Fourier Transform (W-RCW-STFT), Discrete Curvelet Transform (DCT), and Wigner-Ville Distribution. These techniques are employed to capture various aspects of EEG signals. Subsequently, a novel feature selection algorithm known as the Gravitational Seagull Optimization Algorithm (GSOA) is introduced to identify the most informative features. Finally, a Multi-Head Attention Bi-directional LSTM model is employed for deep feature extraction, leveraging its capabilities to discern intricate patterns and relationships within EEG signals.

Results:

The hybrid approach employed for extracting temporal and spatial features from EEG signals significantly enhances the performance of the seizure detection system. Achieving a specificity of 100% along with an accuracy, sensitivity, and F1 score of 99% underscores the effectiveness of seizure detection.

Conclusions:

This study presents a novel and effective approach to seizure detection from EEG signals, leveraging innovative feature extraction, selection methods, and advanced deep learning models. These findings underscore the effectiveness of the proposed methodology in accurately identifying seizures, holding promise for improved clinical diagnosis and treatment of neurological disorders.

Keywords: Epilepsy, Deep Learning, Seizure Detection, EEG Signals, MHA-BiLSTM Model.

1. Introduction

Recurrent seizures is a neurological disorder that causes epilepsy, profoundly affects the well-being of those impacted. It is important to detect seizures timely and accurately for effective management [1]. Deep learning holds promise in automating seizure detection through diverse model developments. Electroencephalogram (EEG) signals, measuring brain electrical activity, are commonly employed in epilepsy diagnosis. Neurologists scrutinize EEG recordings for specific seizure-indicative patterns, such as spike-and-wave discharges, sharp waves, slow waves, and epileptiform abnormalities. Analysis encompasses pattern presence, duration, frequency, and spatial distribution, considering clinical history and diagnostic tests like MRI or PET for a conclusive epilepsy diagnosis and informed treatment decisions [2].

EEG is vital for seizure detection, directly measuring brain electrical activity in real time with millisecond resolution [3]. Unlike MRI or CT scans, EEG captures seizures' dynamic nature. Its non-invasive, portable, and cost-effective nature makes it suitable for long-term monitoring. EEG's distinctive patterns, like spikes and sharp waves, enhance epileptic activity identification. It aids early intervention and treatment monitoring [4], improving patient outcomes. Integrated with other diagnostics, EEG offers a nuanced understanding of seizure causes, leading to tailored care. In summary, EEG is indispensable for diagnosing and managing seizures, offering precision and practicality in medicine.

During seizures, EEG exhibits notable changes like increased neural synchronization, high-frequency spikes, and interictal epileptiform discharges. These alterations serve as markers for abnormal brain activity, aiding diagnosis and classification of seizures [5]. EEG patterns' evolution over time and alterations in background rhythm contribute to understanding seizure origin and propagation. EEG monitoring, sensitive to seizure type and brain location, is invaluable for diagnosis, classification, and treatment planning in epilepsy and seizure disorders.

The work presented in this paper highlights the importance of leveraging deep learning for automated seizure detection, particularly in the context of epilepsy. It emphasizes the challenges in obtaining and handling limited and imbalanced labeled data for training models. The variability in seizure patterns among individuals, noise artifacts in EEG signals and the need for real-time processing pose significant hurdles. The key contributions of the work presented in this article can be succinctly summarized as follows:

1. Feature Extraction: This work advances the field by extracting candidate features from preprocessed EEG signals. Utilizing techniques such as Power Spectral Density (PSD), Weighted Raised Cosine-Window-based Short-Time Fourier Transform (W-RCW-STFT), Discrete Curvelet Transform (DCT), and Wigner-Ville Distribution, the extraction process is designed to capture diverse aspects of the EEG signal.
2. Innovative Feature Selection Algorithm: The development of the Gravitational Seagull Optimization Algorithm (GSOA) introduces a novel feature selection approach. This algorithm is designed to effectively identify the most informative features, enhancing the model's discriminative power.
3. Deep Feature Extraction Model: The introduction of a Multi-Head Attention Bi-directional LSTM model signifies a step forward in deep feature extraction. This model is tailored to extract intricate patterns and relationships from the EEG signals, providing a deeper understanding of the underlying data.
4. Hybrid Seizure Detection Model: The proposed hybrid model integrates various components, including the feature extraction techniques, the GSOA feature selection, and the Multi-Head Attention Bi-directional long short term memory (BiLSTM) model. This comprehensive approach contributes to an effective and holistic model for seizure detection from EEG signals.
5. End-to-End Methodology: The work presents an end-to-end methodology that encompasses dataset preparation, involving the compilation of benchmark datasets. The performance evaluation includes not only the proposed model but also a comparative analysis with state-of-the-art methods, ensuring a comprehensive assessment of its effectiveness.

In addition to the introduction in Section I, Section II offers concise information about the related work in this field. Section III outlines the proposed methodology, including details on dataset preparation, model design, and mathematical aspects of the model. Section IV delves into the specifics of results and analysis for the respective EEG signal datasets, accompanied by a comparative analysis with state-of-the-art methods. Lastly, Section V presents the concluding remarks for the work presented in this article.

2. Related Work

In the extensive body of literature dedicated to seizure detection, a myriad of models has been strategically employed, leveraging diverse algorithms to meticulously evaluate their performance across an array of distinct parameters. One noteworthy contribution comes from Atal et al. [6], who introduced a sophisticated method for precise seizure detection through the amalgamation of MBBF-GPSO and CNN. In this innovative approach, MBBF efficiently mitigates noise, while GPSO meticulously optimizes features, culminating in a dependable classification output.

In a parallel vein, Bhandari et al. [7] orchestrated a comprehensive strategy by integrating different processing approaches. The first approach was Empirical Mode Decomposition (EMD) in which important modalities of the signal are extracted. The frequency dependent features were extracted using Discrete Wavelet Transform (DWT). The model's accuracy was further augmented through the application of Principal Component Analysis (PCA). Also, additional method of feature discrimination, Linear Discriminant Analysis (LDA) and a meticulously crafted weighted feature selection approach optimized by J-CSO was applied.

Ghembaza et al. [8] undertook a meticulous exploration of EEG signals across frequency bands. The classification was performed using Support Vector Machine (SVM). The comparative analysis with k-Nearest Neighbors (kNN) classifier was carried out in which SVM found better. Their focus centered on the nuanced detection of signal features spanning multiple frequency bands.

In a different facet, Visalini et al. [9] introduced deep learning model for identification of seizure and seizure free segments of EEG signals. A Deep Belief Network (DBN) model was evaluated using multi-channel EEG recordings which showed AUC score of 97.1% and 98.7% accuracy [6].

Exploring the potential of Capsule Networks (CapsNet), Jana et al. [10] embarked on an investigation into its efficacy for classifying seizures and non-seizures. Their meticulous approach involved normalizing input data using the L2 normalization technique and subjecting the CapsNet to training and fine-tuning with normalized data. The comprehensive evaluation of the CapsNet and comparative analysis highlighted its substantial mean accuracy of 93.50% across diverse subjects and EEG data types, underscoring its robustness and generalizability.

Further enriching the landscape of seizure detection methodologies, Hossain et al. [11] employed CNN model. The results seen by this model were 90%, 91.65% and 98.05% of sensitivity, specificity and accuracy respectively. Yuan et al. [12] explored feed-forward neural network. This single layered model showed 96.5% accuracy. The Extreme Learning Machine (ELM) classifier clubbed with existing model provided improved accuracy of 97%.

The accuracy of 95.96% was seen by Chandaka et al. [13] with SVM classifier. The mixed expert model and wavelet features used by Subasi et al. [14] achieved an accuracy of 94.5%. With 93% accuracy, Aarabi et al. explored the application of Bayesian Neural Networks (BNN) [12].

Hengjin et al. [15] used VGG16 model along with the global maximal information coefficient (MIC). The results on CHB-MIT dataset shown accuracy of 98.13%.

Aarabi et al. [16] achieved 98.7% accuracy using fuzzy rule based classification model. Themethod was meant for patterns of interictal and ictal EEG. Waseem et al. [17] used auto encoder approach with inclusion of BiLSTM on CHB-MIT dataset which achieved classification accuracy of 99.7%. Andrzejak et al. [18] employed a pre-analysis step based on the weak stationarity criterion. They observed that the most pronounced evidence of nonlinear deterministic dynamics was associated with seizure activity, while the results of the other sets fell somewhere between these two extremes.

Siddiqui et al. [19] delved into both black box and non-block box oriented techniques in their review. The authors provided a detailed examination of seizure detection methods and discussed their future prospects.

Panda et al. [20] used particle swarm optimization and LDA to extract and reduce relevant characteristic features from EEG data. The ensemble extreme learning machine based method provided accuracy of 99.32% on single BONN dataset.

Seizure detection methods from EEG face limitations, often evaluated on single datasets, lacking generalizability. Many fixate on specific sampling frequencies, neglecting broader variations, while handcrafted feature extraction gives way to deep learning. This article introduces a novel approach addressing these challenges by integrating multiple datasets and employing augmentation techniques for robustness. It extends focus beyond seizure zones, considering critical features in both seizure and non-seizure regions. Innovative strategies are detailed, offering potential to enhance EEG-based seizure detection efficacy and reliability.

3. Proposed Work

The proposed methodology of seizure detection model development, involves the EEG dataset preparation, Model Design, Training and Validation stages as shown in Figure 1. The step by step procedure of each stage is detailed further in this section.

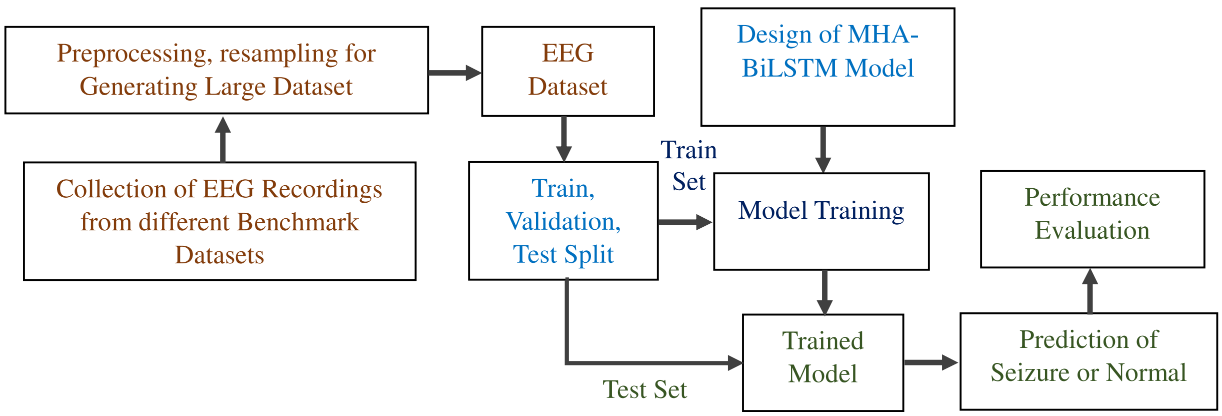


Figure 1: Stages of Proposed Work

Figure 1 depicts the proposed architecture combining conventional and deep feature extraction for Epileptic Seizure detection. Processing stages include (a) pre-processing, (b) feature extraction, (c) feature selection, and (d) deep features extraction, followed by (e) Classification. Pre-processing involves noise reduction and artifact removal through sliding window segmentation and Gaussian noise filtering. Feature extraction encompasses Power Spectral Density, Weighted Raised Cosine-Window-based Short-Time Fourier Transform, Discrete Curvelet Transform, and Wigner-Ville Distribution. Feature selection is conducted via Gravitational Seagull Optimization Algorithm. Seizure detection employs the MHA-BiLSTM model fused with hybrid features from ConvCaps, determining the presence or absence of Seizure.

3.1. Pre-Processing

Sliding window segmentation and Gaussian noise are employed as pre-processing techniques to prepare the collected raw EEG signals in this research work.

3.1.1. Sliding window segmentation

Sliding window segmentation is a widely used technique [21] in the pre-processing of biomedical signals such as EEG, ECG, and blood samples. Sliding window segmentation involves dividing a signal into fixed-length segments, analyzed separately to eliminate noise and artifacts, improving quality. In seizure detection, it aids in analyzing gene or protein expression levels in blood samples to identify disease biomarkers. The window size incrementally increases on time series data until the closest value of the short-term estimate is obtained, repeating until all data is segmented.

3.1.2. Gaussian noise

Gaussian noise is used as a pre-processing technique [22] in this study to remove unwanted noise from the collected raw EEG signals. It involves applying a Gaussian filter to the collected data to reduce the impact of noise and improve the quality of the signal. A non-uniform low pass filter or the Gaussian filter function is given in Eq. (1).

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{1}$$

Where, σ represents the standard deviation. By applying Gaussian noise removal as a pre-processing step, the quality of the EEG signals is improved, which can lead to better feature extraction and classification results in subsequent steps of the analysis.

3.2. Feature Extraction

The pre-processed signal is further processed using several methods for feature extraction including Power Spectral Density (PSD), Weighted Raised Cosine-window based short-time Fourier transform (W-RCW-STFT), Discrete Curvelet Transform (DCT), and Wigner-Ville Distribution.

3.2.1. Power spectral density (PSD)

First signal is transformed in frequency domain using Fourier transform in this process [23]. As PSD provides power distribution among different frequency bands, Seizure impacts in EEG signals are easily captured with the use of these features. The power spectral density is thus calculated with the use of Welch method.

3.2.1.1. Welch Method

A smoother PSD estimate is obtained with the use of Welch method [24]. Inn this process first the signal is converted into overlapping segments. These segments are then processed for average estimates of the periodogram. The details of steps involved are represented mathematically as,

1. Consider, N is the length of the discrete signal x(n),
2. Prepare L overlapping segments of the input signal. The length of each segment is m, then apply window w(m) on each segment.
3. The periodogram k for each segment is calculated as,

$$P(k,f) = |FFT(x\_k(m)w(m))|^2 / (M * |W(f)|^2) \tag{2}$$

Where mth sample of segment k is denoted as x_k(m), The w(m) is applied with FFT using w(f) the squared magnitude to obain PSD is thus denoted by $|.|^2$.

After estimating the periodograms of each segment, for overall PSD, the average is taken as,

$$PSD(f) = (1/L) * sum\_k = 1^{L} P(k,f) \tag{3}$$

Where the sum of all segments periodograms is denoted by, $sum\_k = 1^{L}$.

3.2.2. Weighted Raised Cosine-window based short-time Fourier transform (W-RCW-STFT)

Short-time Fourier transform (STFT) is a technique used for time-frequency analysis of a signal [25]. STFT involves Fourier transforming short signal segments, ideal for analyzing non-stationary signals with changing frequency content. It offers fixed frequency resolution but may suffer from spectral leakage and distortion. Trade-offs between time and frequency resolution impact accuracy. To overcome these problems, this work proposed the Weighted Raised Cosine-window based Short-Time Fourier Transform (W-RCW-STFT) is a modified version of the STFT that combines a raised cosine window with an additional weight function to improve spectral resolution and reduce spectral leakage.

The W-RCW-STFT is computed as follows:

Step 1: Divide the signal into overlapping segments of length N, with a percentage of overlap between adjacent segments.

Step 2: Apply the raised cosine window to each segment. The raised cosine window used in the W-RCW-STFT is defined as per Eq. (4).

$$w(n) = 0.5 * (1 - \cos\left(\frac{2\pi n}{N-1}\right) \tag{4}$$

where N is the length of the window and n is the index of the sample within the window.

Step 3: Multiply each sample in each windowed segment by the corresponding value of an additional weight function. The additional weight function used in the W-RCW-STFT can vary depending on the specific application. One commonly used weight function is the Hamming weight function, which is defined as per Eq. (5)

$$h(m) = 0.54 - 0.46 * \cos\left(\frac{2\pi m}{M-1}\right) \tag{5}$$

where M is the length of the Fourier transform and m is the index of the frequency component.

Step 4: Compute the Fourier transform of each weighted windowed segment.

Step 5: Concatenate the results of each weighted Fourier transform to form the W-RCW-STFT. The STFT is expressed mathematically as per Eq. (6).

$$STFT(n) = \int x(n)w(n).h(m).\ e^{(-2\pi i f \tau)} d\tau \tag{6}$$

where $x(n)$ is the signal, $w(n)$ is the window function (raised cosine window) and the $\int$ is taken over all time.

The W-RCW-STFT combines the benefits of the raised cosine window and the weight function to improve the accuracy of the spectral analysis. The raised cosine window reduces spectral leakage by smoothly tapering the edges of the windowed segments to zero, while the weight function enhances specific frequency components.

3.2.3. Discrete Curvelet Transform (DCT)

DCT is a multiscale and multidirectional signal processing technique used for feature extraction in various applications [26]. DCT decomposes a signal into different scales and directions, which makes it useful for capturing localized and directional features. DCT decomposes the signal, then filters to obtain Curvelet coefficients capturing localized and directional information. The formula for the DCT involves a series of operations including wavelet decomposition, directional filter banks, and local Fourier transforms. The DCT formula can be expressed as per Eq. (7).

$$C(u,v) = \sum W(u,v,k,l) * F(k,l) \tag{7}$$

where $C(u,v)$ represents the DCT coefficient at scale u and direction v, $W(u,v,k,l)$ represents the directional filter bank, $F(k,l)$ represents the local Fourier transform of the wavelet coefficients, and k and l represent the position in the wavelet domain. The DCT is a powerful feature extraction technique that can capture localized and directional information from a signal.

3.2.4. Wigner-Ville Distribution

The time frequency distribution, Wigner-Ville Distribution (WVD), represents the time-varying spectral content of a signal [27]. It was introduced by Eugene Wigner and is a popular tool in signal processing for analyzing non-stationary signals. Time-frequency analysis, promising for signal examination, aids in estimation, detection, and visualization, especially beneficial for EEG data analysis. In most cases, the WVD is provided as per Eq. (8) for a discrete signal $x(h)$ with H samples.

$$X(T,F) = \sum_{h=-H}^{H} \left(x\left(m + \frac{n}{2}\right) x^{*}\left(m - \frac{n}{2}\right) e^{\frac{-j2\pi Fh}{H}} \right) \tag{8}$$

where $j = \sqrt{-1}$ and T, F represents the time and frequency components, respectively. The complex conjugate of the number $x(n)$ is the amount $x^{*}(n)$. In the next step of the automated Seizure detection from EEG signal, the extracted features including PSD, STFT, DCT, and Wigner-Ville Distribution undergo a feature selection process.

3.3. Feature Selection via GSOA

The combination of two algorithms the Seagull Optimization Algorithm (SOA) [28] and the Gravitational Search Algorithm (GSA) [29] provides a new metaheuristic algorithm Gravitational Seagull Optimization Algorithm (GSOA). The method optimizes using seagulls' behavior. Seagulls vary in size and drink both freshwater and saltwater. They migrate in groups, avoiding collisions, moving toward survival. During migration, they attack in a spiral pattern. Inspired by gravity, the GSA algorithm guides particles representing individuals toward optimal solutions. Higher mass particles attract others, achieving global optimization.

Premature convergence occurs when an optimization algorithm settles for a suboptimal solution before reaching the best one. Combining SOA and GSA mitigates premature convergence by blending global exploration and local exploitation capabilities. This hybrid approach enhances diversity, prevents trapping in local optima, and reaches the global optimum effectively.

Step 1: Initialization: Set the population size N, maximum number of iterations T^{max}. Initialize the position (P), and velocity v of each seagull randomly within the search space. A population in GSA contains N particles. Each particle is expressed as $X_i = (x_i^1, \dots, x_i^{di}, \dots, x_i^{DI})\ i \in \{1,2,\dots,N\}$ where D is the dimension.

Step 2: Evaluation: Consider objective function and evaluate the fitness of each seagull. The objective function of this research work is: $obj = \max(A)$*(i.e. maximization of detection accuracy)*

Step 3: Leader Selection (proposed): Select the best seagull as the global best solution and the remaining top seagulls as the local best solutions (based on the fitness function).

Step 4: Movement (proposed): Update the velocity and position of each seagull based on the following equations: Velocity updated as per Eq. (9)

$$v^{(i,j)}(t) = w * v^{(i,j)}(t) + C1 * r1 * \left(p_{best}{}^{(i,j)}(t) - p^{(i,j)}(t)\right) + C2 * r2 * \left(g_{best}{}^{(i,j)}(t) - p^{(i,j)}(t)\right) \tag{9}$$

Position updated as per Eq. (10).

$$p^{(i,j)}(t+1) = p^{(i,j)}(t) + v(i,j) \tag{10}$$

where $v(i,j)$ is the velocity of the i^{th} seagull at the j^{th} dimension, w is the inertia weight, $C1$ and $C2$ are the acceleration coefficients, $r1$ and $r2$ are random numbers between 0 and 1, $p_{best}(i,j)$ is the local best position of the ith seagull at the jth dimension, and $g_{best}(i,j)$ is the global best position at the jth dimension.

Step 5: Avoiding the collisions: During the process of identifying the new search agent's position, a new extra variable A was introduced in Eq. (11) and used to stop collisions from happening with nearby seagull.

$$p^{(i,j)}(t) = A \times p^{q}(i,j) \tag{11}$$

Where the present iteration is denoted as t, A denotes the movement characteristics of the search agent on the available search space, p^q means the location of the search agent where it won't be able to collide with other search agents, and the search agent's current position is represented as $p^{(i,j)}(t)$. The value of A is given in Eq. (12),

$$A = f_c - \left(u \times (f_c / T^{max})\right) \tag{12}$$

Where: $u = 0,1,2,\dots, Max_{iteration}$

Where f_c was offered as a means of regulating the frequency of the employing variable, f_c was then linearly minimized towards 0. In the current task, 2 is the value set for f_c. In primary balancing requirements and traditional scenarios of SOA, value of f_c-factor is linearly decreased from 2 to 0.

Step 6: Moving in the direction of the best neighbor

Later, in order to minimize collisions between the neighboring herds, the herd's search agents travel closer to the other herds given in Eq. (13) in order to identify which herds are the most cooperative.

$$M = B \times \left(p_{best}{}^{(i,j)}(t) - p^{(i,j)}(t)\right) \tag{13}$$

Where the search agent p^t is positioned towards the search agent in the areas where M describes the search agents. The B behaviour was randomly generated, which is what has proven to be a successful balance between exploitation and exploration. The value of B is calculated using Eq. (14),

$$B = 2 \times A^2 \times rd \tag{14}$$

Where rd was a random number that would fall between [0, 1] approximately.

Step 7: Stay in close to the top search agent

The search agent is also be able to update its location in reference to the provided ideal search agent. The updated location is given in Eq. (15),

$$D = |C + M| \tag{15}$$

It is recognized as the best seagull, whose fitness value was minimal, where D was marked as the separation of the search agents and the distance to the most suitable search agent.

Step 8: Attacking (Exploitation)

Exploitation optimizes by leveraging past successes. Seagulls adjust attack angles and speed, using body weight and wings for altitude, spiraling to attack prey mid-flight. The u, v, and w planes were used to show the following properties in Eq. (16), Eq. (17), and Eq. (19) respectively.

$$u' = r \times \cos(k) \tag{16}$$

$$v' = r \times \sin(k) \tag{17}$$

$$w' = r \times k \tag{18}$$

$$r = q \times e^{ks} \tag{19}$$

Where r denoted the diameter of each spiral turn and k denoted a random value falling within the range of $[0 \leq k \leq 2\pi]$. Where e was used as the base for the natural logarithm, $q\ and\ s$ were the constants used to define the spiral shape. Eq. (20) is used to determine the search agents' most recent location.

$$p_{best}{}^{(i,j)}(t+1) = (D \times u' \times v' \times w') + p_{best}(i,j) \tag{20}$$

Where $p_{best}{}^{(i,j)}(t+1)$changes the position and then saves the best response from various search agents. The population is produced randomly at the start of the seagull optimization algorithm (SOA) that is being presented. The search agents were able to inform the positions according to the most effective search agent during iterations. The A was linearly minimized from f_c to 0. The seamless transition between exploitation and exploration was due to the variable B. The seagull optimization algorithm has earned a reputation as the global best optimizer due to its superior exploitation and exploration capabilities.

Step 9: Gravity Calculation: Calculate the gravitational forces between each seagull and update their velocities and positions based on the gravitational force, according to the GSA algorithm. Gravitational force $g^{(i,j)}(t)$ between two particles i_i and j are represented as per Eq. (21).

$$g^{(i,j)}(t) = g(t)\frac{pm_i(t) \times pm_j(t)}{r_{ij}(t)+\varepsilon}\left(g^i(t) - g^j(t)\right) \tag{21}$$

Where the gravitational constant $g(t)$ is used. The masses of two particles are $pm_i(t)$ and $pm_j(t)$. The Euclidean distance between them is $r_{ij}(t)$, and ε is a tiny constant. The definition of the gravitational constant $g(t)$ is given as per Eq. (22).

$$g(t) = g_0 \times e^{-\alpha\frac{t}{T^{max}}} \tag{22}$$

Where α is a coefficient and g_0 is the starting point. T^{max} is the maximum number of iterations. The mass $pm_i(t)$ of each particle is given as per Eq. (23) and Eq. (24).

$$PM_i(t) = \frac{Fit(t) - FitW(t)}{FitB(t) - FitW(t)} \tag{23}$$

$$pm_i(t) = \frac{PM_i(t)}{\sum_{l=1}^{N} PM_l(t)} \tag{24}$$

Where the fitness value is $Fit(t)$. The lowest and best fitness values are $FitW(t)$ and $FitB(t)$ respectively. Each particle i is drawn to other particles, hence its overall gravitational force $g^{(i,j)}(t)$ is calculated as per Eq. (25).

$$g^{(i,j)}(t) = \sum_{j \in KB, j \neq i} rand_j \cdot g^{(i,j)}(t) \tag{25}$$

Where $rand_j$ is a value chosen at random from the range (0,1). K is linearly reduced with iteration t, and KB denotes the first K best particles. Next, the acceleration $A^{(i,j)}(t)$, velocity $V^{(i,j)}(t)$, and location $p^{(i,j)}(t+1)$of each particle are calculated as per Eq. (26) to Eq. (28), respectively.

$$A^{(i,j)}(t+1) = \frac{g^{(i,j)}(t)}{pm_i(t)} \tag{26}$$
$$g^{(i,j)}(t+1) = rand_i \cdot v^{(i,j)}(t) + a^{(i,j)}(t) \tag{27}$$
$$p^{(i,j)}(t+1) = p^{(i,j)}(t) + v^{(i,j)}(t+1) \tag{28}$$

Where $rand_i$ represents the random value in the interval (0,1). Finally, a new approach for detecting Seizure has been proposed using a two-tier-deep learning model.

Step 10: Check Termination Criteria: Check whether the maximum number of iterations or the desired accuracy of the solution has been reached. If the termination criteria are not met, go to step 2. Otherwise, return the best solution found.

The GSOA algorithm can be modified by adjusting the parameters such as population size, maximum number of iterations, and acceleration coefficients, to improve the convergence rate and the quality of the solutions. Additionally, the GSOA algorithm can be extended to handle constraints, multi-objective optimization, and dynamic optimization problems.

3.4. Seizure Detection via hybrid model

The proposed method for Seizure detection combines the conventional features and deep features based Multi-Head Attention-based Bidirectional Long-Short Memory (MHA-BiLSTM) deep learning architectures. The final outcome, which determines the presence or absence of Seizure, is obtained by merging the outputs of both ConvCaps and MHA-BiLSTM.

3.4.1 BiLSTM based model with Multi head Attention

In terms of computational complexity, BiLSTM is twice complex than LSTM. The computation for each cell of LSTM is as per Eq. (25) to Eq. (30). The input Vector A_t is processed in LSTM cell with h_{t-1} as hidden state and C_t as current state.

$$Inp_t = S_a(I_{A.Inp}A_t + R_{h.inp}h_{t-1} + b_v(inp) \tag{29}$$
$$Fg_t = S_a(I_{A.Fg}A_t + R_{h.Fg}h_{t-1} + b_v(Fg) \tag{30}$$
$$Out_t = S_a(I_{A.Out}A_t + R_{h.Out}h_{t-1} + b_v(Out) \tag{31}$$
$$m_t = B(I_{At}A_t + R_{hc}h_{t-1} + b_v(c) \tag{32}$$
$$c_t = Fg_t c_{t-1} + m_t.Inp_t \tag{33}$$
$$h_t = Out_t.B(c_t) \tag{34}$$

In the given context, Inp_t, Fg_t, Out_t, and m_t refer to the "input, forget, output, and input modulation gates" at time t. The corresponding weights, denoted as Input weight I_A, R as recursive, and biased weight b_v. The sigmoid activation function, denoted as $S_a(A) = (1+e^{-A})^{-1}$, is used, while the function $B(A) = (e^A - e^{-A})/(e^A + e^{-A})$, is activated using hyperbolic tangent and is employed to derive the forget gate Fg_t. The memory cell values are thus updated using update gate and current state values. The memory cell contents optimization Is performed with control from input and forget gates. This depends on the decision of carrying forward or discarding the hidden states of previous steps. At time t, the output gate is influenced for future operations. The hidden state is thus obtained with element-wise multiplication of the values that are passed through tanh activation. The values that passed are taken from output and current memory cell. This way for each step, current memory cell is updated.

The MHA-BiLSTM architecture integrates BiLSTM and Multi Head Attention, advancing sequence processing bidirectionally. It captures contextual information and extended dependencies efficiently, enhancing temporal understanding. The output from the attention layer serves as the input to fully connected layers, providing the necessary foundation for the final stages of the network. These fully connected layers process the refined information, culminating in the ultimate output of the MHA-BiLSTM architecture (see Figure 2).

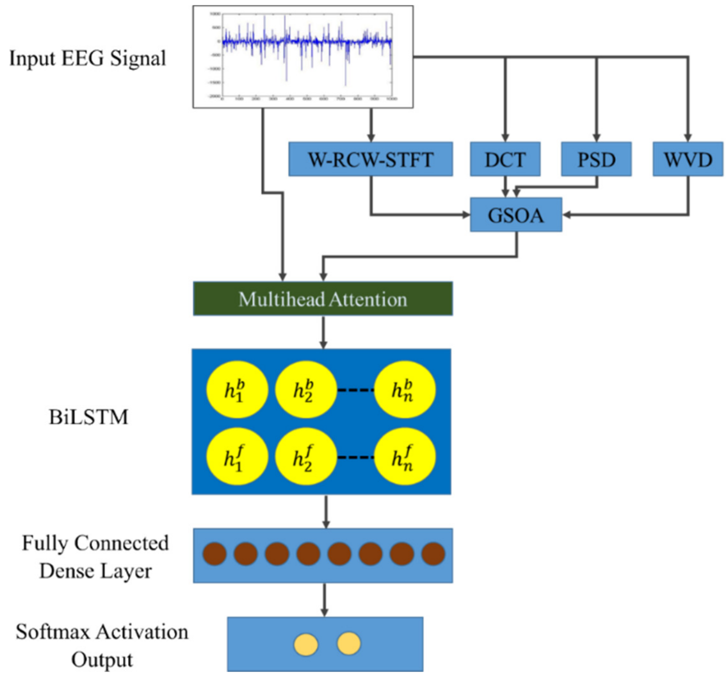


Figure 2: Proposed Model for Seizure Detection

4. Results and Analysis

4.1 Dataset Preparation

Segmenting EEG signals enhances seizure detection by breaking them into smaller, trainable segments. Optimizing dataset size is crucial for model performance, with careful consideration of segment length and accurate labeling. Automated or manual labeling based on seizure presence is vital. Integrating diverse datasets requires aligning sampling frequencies, as shown by Natu et al. [30]. This study adopts a similar approach, incorporating down-sampling and up-sampling. The dataset details are shown in Table 1.

Table 1: Recomposed Dataset

Details of Dataset	Classes	Samples	After Augmentation
(Kaggle [31])	Non-Seizure	261	1566
	Seizure	258	1548
	Total	**519**	**3114**
(Combined CHB-MIT [32] and BONN [33])	Non-Seizure	261	1566
	Seizure	258	1548
	Total	**519**	**3114**
3 (Combined)	Seizure	516	3096
	Non-Seizure	522	3132
	Total	**1038**	**6228**

4.2 Performance Parameters

The performance parameters used to evaluate the proposed model are shown in Table 2.

Table 2: Performance Parameters

Parameter	**Formula**
Precision	TP/(TP+FP)
Sensitivity/Recall	TP/(TP+FN)
F1 Score	2*(Precision*Recall)/(Precision+Precision)
Specificity	TN/(TN+FP)
Accuracy	TP+TN/(TP+TN+FP+FN)

4.3 Choice of Loss Function

1. Binary Cross Entropy (BCE)

BCE is a loss function used for binary classification tasks [34]. In the current goal to categorize input instances into seizure or non-seizure, BCE is used. The fundamental idea behind Binary Cross Entropy is rooted in information theory and measures the dissimilarity between probability distribution of the predicted outcomes and the actual classes. For a single training instance, let's denote the true class label as y (either 0 for non-seizure or 1 for seizure) and the predicted probability of belonging to class 1 (seizure) as p. The Binary Cross Entropy loss is calculated as follows:

$$BCE = -y\,[ylog(p) + (1-y)\log(1-p)] \quad \text{...(35)}$$

BCE gauges the model's error by considering predicted probabilities and actual labels, minimizing overfitting and confidently incorrect predictions. Optimization algorithms iteratively adjust parameters to minimize BCE loss, ensuring model stability as illustrated in Figure 3.

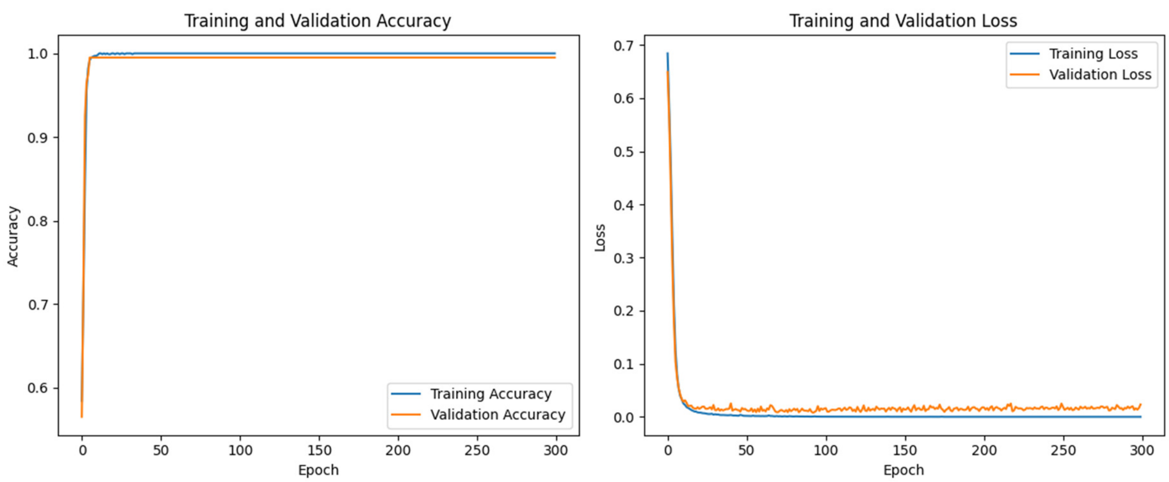


Figure 3: Training Analysis with Binary Cross Entropy Loss Function

2. Kullback-Leibler (KL)

KL Divergence [35], also known as relative entropy, is a measure of how one probability distribution diverges from a second, expected probability distribution. It is often used in information theory, statistics, and machine learning to quantify the difference between two probability distributions. The KL Divergence between two probability distributions P(x) and Q(x) is defined mathematically as follows:

$$D_{KL}(P||Q) = \sum_{x\,\epsilon\,support(P)} P(x)\ \log\left(\frac{P(x)}{Q(x)}\right) \quad \text{...(36)}$$

Or in the case of continuous distributions:

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} P(x)\ \log\left(\frac{P(x)}{Q(x)}\right) \quad \text{...(37)}$$

Where, P(x) is the true probability distribution, Q(x) is the estimated or approximate probability distribution, The sum or integral is taken over all possible values of xx in the support of P, The logarithm is typically taken to the base 2 or natural logarithm. In the proposed work, the loss function is replaced to analyze the impact on detection of seizures. Figure 4 shows the training accuracy and loss analysis.

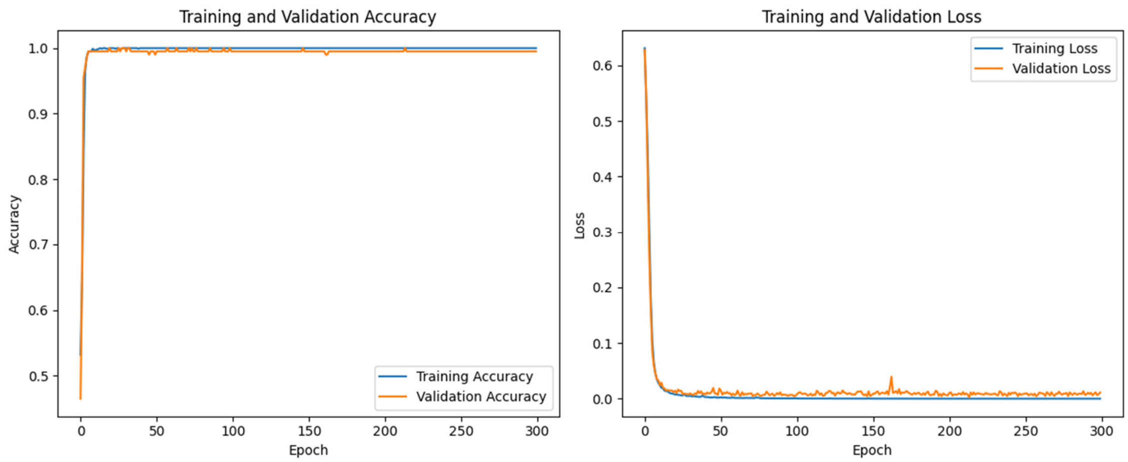


Figure 4: Training Analysis with KL Divergence Loss Function

With the observations of training performance, the binary cross entropy loss function is found most suitable in proposed model.

BCE proves superior to KL divergence in seizure classification due to its suitability for binary tasks. It computes cross-entropy loss independently for binary predictions, ideal for EEG data segment seizure identification. Conversely, KL divergence measures distribution divergence, better suited for multi-class problems. BCE is computationally efficient, crucial for timely medical intervention, while KL divergence may demand more resources. Additionally, BCE offers a straightforward measure of classification error, vital in medical contexts, whereas KL divergence may add complexity without aligning with interpretability needs. These factors underscore BCE's efficacy in seizure detection, ensuring quick, accurate predictions with direct interpretability, essential for medical applications.

4.4 Hyper Parameter Tuning

To assess model stability, the results of a 10-fold analysis were examined. From the 7th fold onward, the model consistently demonstrated stability in terms of accuracy as shown in **Figure 5**.

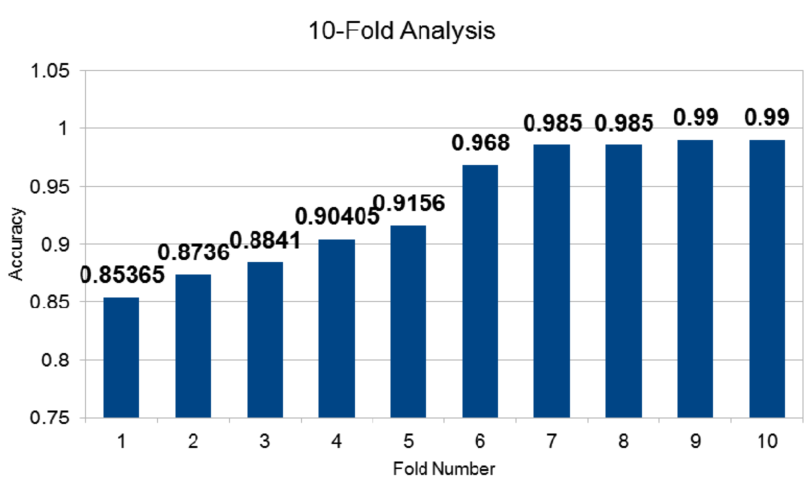


Figure 5: 10-Fold Analysis

The hyperparameter tuning with different batch sizes, the best results are obtained at batch size 32 as shown in Figure 6. Also, in second last dense layer hidden neurons analysis, best results are obtained at 128 neurons as shown in Figure 7.

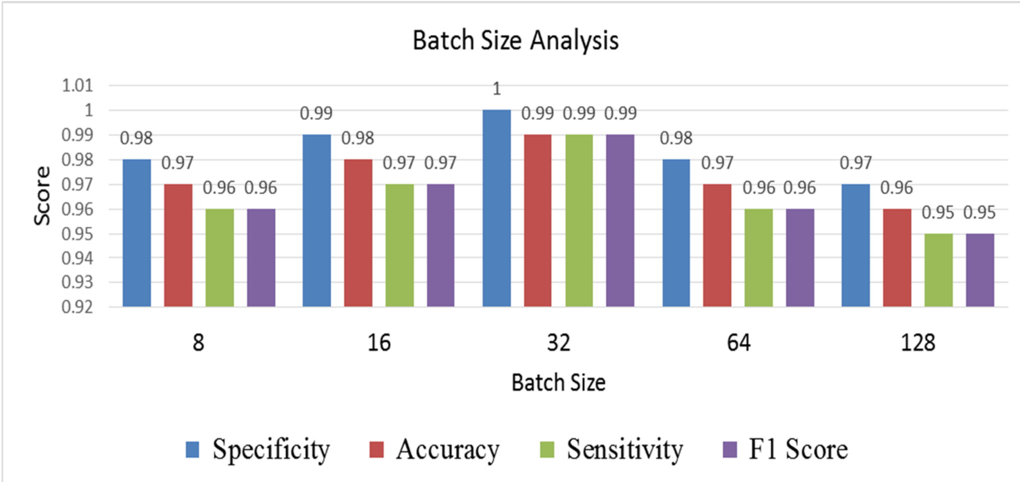


Figure 6: Hyper Parameter Batch Size Analysis

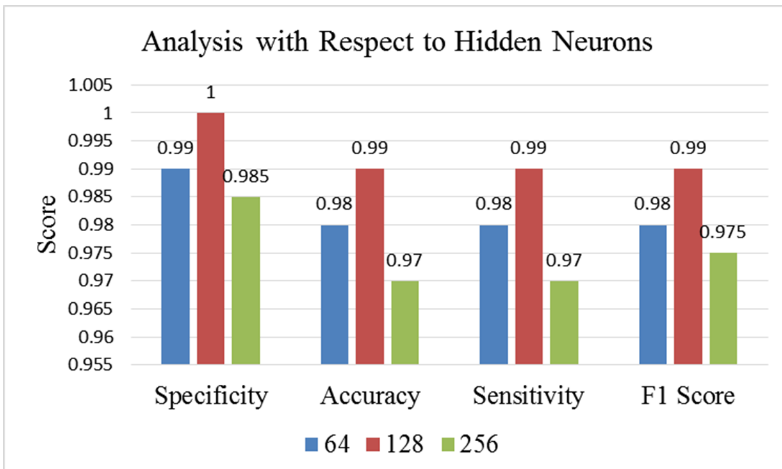


Figure 7: Hyper Parameter Hidden Neurons Analysis

4.5 Comparative Analysis

For the dataset 1 and dataset 2, the results are separately obtained to compare and identify the results of dataset 3 which shows significant impact on the result. Dataset 3 which is combination of datasets shows improved performance as shown in Figure 8. The confusion matrix analysis is shown in Figure 9 for the results in Figure 8.

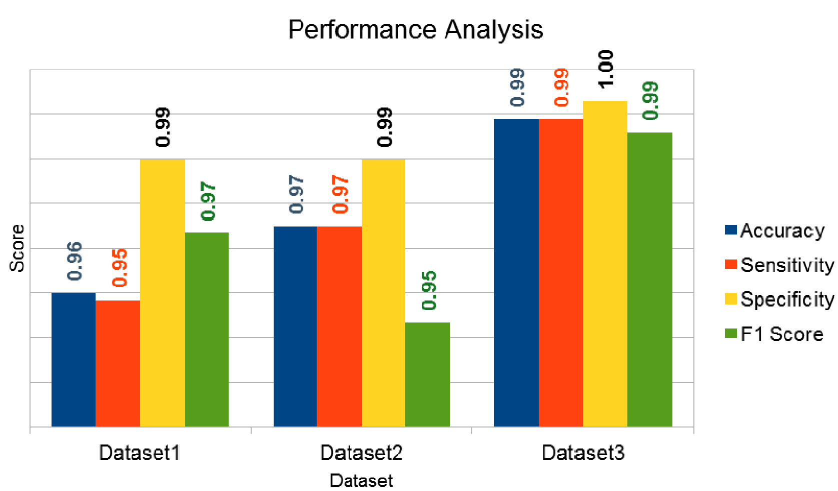
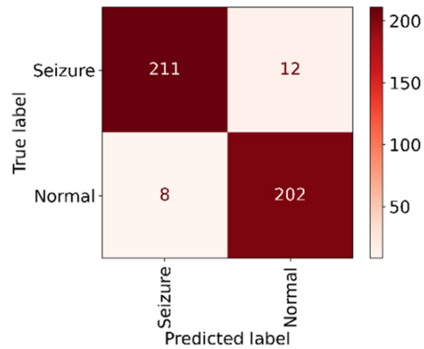
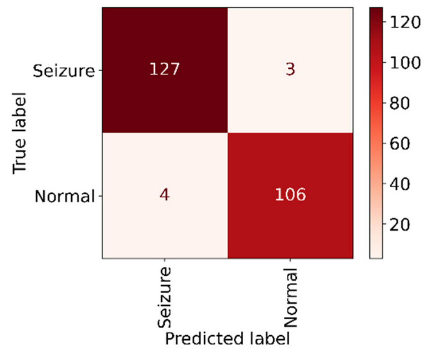


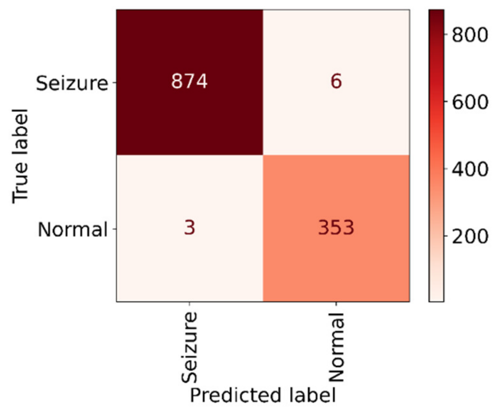
Figure 8: Performance on Different Datasets



(a) Confusion Matrix based on Dataset 1



(b) Confusion Matrix Based on Dataset 2



(C) Confusion Matrix on Dataset 3

Figure 9: Confusion Matrix Analysis with respect to Model trained on Datasets and Respective Test Sets

Table 3 presents a comprehensive comparative analysis of state-of-the-art methods for EEG-based seizure detection. It covers approaches employing conventional feature extraction with SVMs, as well as deep learning techniques like CNNs, LSTMs, and GRUs. The table not only details methodologies but also discusses utilized datasets, including their combinations. The proposed method stands out for its superior adaptability to diverse EEG signal sources and robust performance, surpassing existing approaches. Its effectiveness in addressing complexities in EEG-based seizure detection highlights its potential for practical deployment in real-world scenarios with varying sampling frequencies and data sources.

Table 3: Comparative Analysis

Method	Number of Datasets	Dataset Combined?	Dataset	Whether Augmentation is used?	Performance (% Accuracy)
Hand Crafted Feature Extraction and SVM [13]	2	No	1. Khyber Hospital Dataset [45] 2. CHB-MIT	No	95.96%
Hand Crafter Feature extraction and SLFN [12]	2	No		No	96.5%
Mean Extreme and Wavelet Features [36]	2	No		No	94.5%.
Retraining of VGGNet-16 [15]	2	No		No	98%
CNN model [11]	2	No		No	98.05%
Bayesian Neurla Network [16]	2	No		No	93%
Extreme Machine Learning [37]	2	No		No	96.5%
Deep Bayesian Network [9]	1	No	EEG data from 79 infants obtained from Helsinki	No	99.7%

			University Hospital		
CNN model with PSO and LDA for Feature Discrimination [6]	1	No	CHB-MIT	No	98.8%
SVM with RBF kernel with hand crafter rules of feature extraction [8]	2	No	1. TUH EEG Corpus [47] 2. CHB-MIT [23]	No	AUC= 0.9911 and 0.9701
Fuzzy Rule Based Classifier [7]	1	No	CHB-MIT	No	99.7%
CapsNet with deep features extraction [10]	2	No	1. BONN dataset 2. CHB-MIT	No	82.61%
Bidirectional Convolutive GRU model with optimization of attention weights [30]	3	Yes	1. BONN Dataset 2. CHB-MIT 3. Epileptic seizures dataset (Kaggle)	Yes	98.5%
Conventional Features of EEG and Multi Head Attention Model with Bidirectional LSTM and Optimization of Weights MHA-BiLSTM (Proposed)	**4**	**Yes**	**1. Epileptic seizures dataset (Kaggle)** **2. CHB-MIT** **3. ECoG Dataset, University of California** **4. BONN Dataset**	**Yes**	**99%**

Table 4 presents a comprehensive study on LSTM's application in EEG signal analysis, particularly in seizure detection. It highlights LSTM's efficacy in handling the intricate nature of EEG signals, crucial for pattern recognition and classification. The table details LSTM's performance metrics, methodologies, and outcomes, emphasizing its ability to capture temporal dependencies. LSTM's long short-term memory mechanism enables it to discern patterns effectively, vital for timely seizure detection. The study showcases the model's adaptability and efficiency in handling EEG complexities, contributing to both academic understanding and practical applications in healthcare. This novel approach underscores LSTM's advancements in EEG analysis, particularly in seizure detection, offering insights for further research and real-world implementations.

Table 4: Study of use of EEG and LSTM/BiLSTM in Different Applications

Reference	Application	EEG Used?	Optimization Used?	LSTM or BiLSTM Used?	Conventional Features Used?
[38]	Music and Speech Classification	Yes	No	BiLSTM	No
[39]	Smart Wheel Chair using Brain Computer Interface	Yes	No	LSTM	No
[40]	Seizure Detection	Yes	No	LSTM	No
[41]	Emotion Recognition	Yes	No	BiLSTM	No
[42]	Seizure Detection	Yes	No	LSTM	No

[43]	Visual Scene Stimuli based EEG classification	Yes	No	LSTM	No
[44]	Stress Detection	Yes	No	LSTM	No
[45]	Emotion Recognition	Yes	No	LSTM	No
MHA-BiLSTM (Proposed)	**Seizure Detection**	**Yes**	**Yes**	**BiLSTM**	**Yes**

4.6 Discussions

The MHA_BiLSTM, a sophisticated deep learning architecture combining multi-headed attention with BiLSTM for sequence modeling, has proven to surpass both LSTM and Gated Recurrent Unit (GRU) in a spectrum of tasks. Several factors contribute to the superiority of MHA_BiLSTM:

1. **Better Performance:** MHA_BiLSTM has consistently demonstrated superior performance compared to LSTM and GRU across diverse tasks such as speech recognition, sentiment analysis, and video classification. Its ability to capture both short-term and long-term dependencies within input sequences enhances model accuracy.
2. **Faster Training:** Leveraging convolutional layers, MHA_BiLSTM expedites training compared to LSTM and GRU. The convolutional layers efficiently extract local features from input sequences, diminishing sequence dimensionality and accelerating the training process.
3. **Parameter Efficiency:** MHA_BiLSTM boasts superior parameter efficiency by necessitating fewer parameters than LSTM and GRU. This characteristic mitigates the risk of overfitting, enhancing overall model efficiency. The use of convolutional layers for feature extraction further contributes to reducing the model's parameter count.
4. **Bidirectional Modeling:** The incorporation of bidirectional modeling in MHA_BiLSTM enables the capture of both past and future information within input sequences. This comprehensive understanding of the input context enhances the model's predictive accuracy.
5. **Timed Analysis:** The model's performance evaluation includes an analysis of time complexity. The proposed model required one hour for training on a T4 GPU in the Google Colab implementation, significantly less than the 3.5 hours taken on a CPU. Conversely, model complexity becomes a significant concern when employing deep learning approaches.

5. Conclusion

The proposed MHA_BiLSTM model, designed for the detection of seizures using EEG data, has exhibited outstanding performance metrics, solidifying its position as a state-of-the-art solution in the field. A comprehensive evaluation, detailed in Table 3, accentuates the model's superiority over existing methods. With an impressive accuracy of 99%, sensitivity and specificity both at 99%, and an exceptional F1 score of 99%, the model stands as a robust and accurate tool for seizure detection. One of the key strengths of the model lies in its adaptability to diverse EEG sources, showcasing superior performance across multiple datasets with varying sampling frequencies. This adaptability addresses a crucial requirement in real-world scenarios, where EEG datasets often display significant variations due to different acquisition setups. Efficiency is another hallmark of the proposed model, as evidenced by the time complexity analysis. The training process on a T4 GPU in the Google Colab environment was completed within just one hour, marking a substantial reduction compared to the 3.5 hours required on a CPU.

In conclusion, the MHA_BiLSTM model not only achieves exceptional accuracy but also demonstrates efficiency and adaptability, making significant contributions to the field of EEG-based seizure detection. Its promising results pave the way for enhanced reliability and practical implementation of seizure detection systems in clinical settings, where timely and accurate assessments are critical for patient care.

Acknowledgement

We extend our sincere appreciation to the Symbiosis Institute of Technology, (Pune, India) for their invaluable support throughout this research endeavor. Their state-of-the-art laboratory facilities, technical expertise, and unwavering assistance have been instrumental in the successful execution of our experimental analyses.

Financial Disclosure:
This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Funding Declaration:
This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Contributions:
Author 1 have contributed by writing, drafting and experimentation of the work presented in the article under the supervision of Author 2. Author 3 contributed for editing and reviewing the draft.

Conflict of Interest:
The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

Ethical Declarations:
The datasets utilized in the experiments presented in the article have been recompiled using publicly available benchmark datasets. As a result, new ethical declarations are not applicable.

References:

[1] J. I. Sirven, “Epilepsy: A Spectrum Disorder,” *Cold Spring Harbor Perspectives in Medicine*, vol. 5, no. 9, p. a022848, 2015, doi: 10.1101/CSHPERSPECT.A022848.

[2] A. Abdelhameed and M. Bayoumi, “A Deep Learning Approach for Automatic Seizure Detection in Children With Epilepsy,” *Frontiers in computational neuroscience*, vol. 15, Apr. 2021, doi: 10.3389/FNCOM.2021.650050.

[3] G. A. Light *et al.*, “Electroencephalography (EEG) and Event-Related Potentials (ERP’s) with Human Participants,” *Current protocols in neuroscience / editorial board, Jacqueline N. Crawley ... [et al.]*, vol. CHAPTER, no. SUPPL. 52, p. Unit, Jul. 2010, doi: 10.1002/0471142301.NS0625S52.

[4] N. K. Al-Qazzaz, S. H. B. M. Ali, S. A. Ahmad, K. Chellappan, M. S. Islam, and J. Escudero, “Role of EEG as Biomarker in the Early Detection and Classification of Dementia,” *The Scientific World Journal*, vol. 2014, 2014, doi: 10.1155/2014/906038.

[5] M. Avoli, “Mechanisms of Epileptiform Synchronization in Cortical Neuronal Networks,” *Current medicinal chemistry*, vol. 21, no. 6, p. 653, Jan. 2014, doi: 10.2174/09298673206661311191511136.

[6] D. K. Atal and M. Singh, “Effectual seizure detection using MBBF-GPSO with CNN network,” *Cognitive Neurodynamics*, pp. 1–12, Feb. 2023, doi: 10.1007/S11571-023-09943-1/METRICS.

[7] V. Bhandari and M. D. Huchaiah, “A new design of epileptic seizure detection using hybrid heuristic-based weighted feature selection and ensemble learning,” *International Journal of Intelligent Robotics and Applications*, vol. 6, no. 4, pp. 668–693, Dec. 2022, doi: 10.1007/S41315-022-00233-3/METRICS.

[8] F. Ghembaza and A. Djebbari, “Novel quadratic time-frequency features in EEG signals for robust detection of epileptic seizure,” *Research on Biomedical Engineering 2023*, pp. 1–23, Mar. 2023, doi: 10.1007/S42600-022-00256-6.

[9] K. Visalini, S. Alagarsamy, and D. Nagarajan, “Neonatal seizure detection using deep belief networks from multichannel EEG data,” *Neural Computing and Applications*, pp. 1–11, Jan. 2023, doi: 10.1007/S00521-023-08254-0/METRICS.

[10] G. C. Jana, K. Swami, and A. Agrawal, “Capsule neural network based approach for subject specific and cross-subjects seizure detection from EEG signals,” *Multimedia Tools and Applications*, pp. 1–32, Mar. 2023, doi: 10.1007/S11042-023-14995-W/METRICS.

[11] M. Shamim Hossain, S. U. Amin, M. Alsulaiman, and G. Muhammad, “Applying Deep Learning for Epilepsy Seizure Detection and Brain Mapping Visualization,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 15, no. 1s, Feb. 2019, doi: 10.1145/3241056.

[12] Q. Yuan, W. Zhou, J. Zhang, S. Li, D. Cai, and Y. Zeng, “EEG classification approach based on the extreme learning machine and wavelet transform,” *Clinical EEG and neuroscience*, vol. 43, no. 2, pp. 127–132, Apr. 2012, doi: 10.1177/1550059411435861.

[13] S. Chandaka, A. Chatterjee, and S. Munshi, “Cross-correlation aided support vector machine classifier for

classification of EEG signals," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1329–1336, Mar. 2009, doi: 10.1016/J.ESWA.2007.11.017.

[14] A. Subasi, J. Kevric, and M. Abdullah Canbaz, "Epileptic seizure detection using hybrid machine learning methods," *Neural Computing and Applications*, vol. 31, no. 1, pp. 317–325, Jan. 2019, doi: 10.1007/s00521-017-3003-y.

[15] H. Ke, D. Chen, X. Li, Y. Tang, T. Shah, and R. Ranjan, "Towards Brain Big Data Classification: Epileptic EEG Identification with a Lightweight VGGNet on Global MIC," *IEEE Access*, vol. 6, pp. 14722–14733, Feb. 2018, doi: 10.1109/ACCESS.2018.2810882.

[16] A. Aarabi, R. Fazel-Rezai, and Y. Aghakhani, "A fuzzy rule-based system for epileptic seizure detection in intracranial EEG," *Clinical Neurophysiology*, vol. 120, no. 9, pp. 1648–1657, Sep. 2009, doi: 10.1016/J.CLINPH.2009.07.002.

[17] W. A. Mir, M. Anjum, Izharuddin, and S. Shahab, "Deep-EEG: An Optimized and Robust Framework and Method for EEG-Based Diagnosis of Epileptic Seizure," *Diagnostics (Basel, Switzerland)*, vol. 13, no. 4, Feb. 2023, doi: 10.3390/DIAGNOSTICS13040773.

[18] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 64, no. 6, p. 8, 2001, doi: 10.1103/PhysRevE.64.061907.

[19] M. K. Siddiqui, R. Morales-Menendez, X. Huang, and N. Hussain, "A review of epileptic seizure detection using machine learning classifiers," *Brain Informatics*, vol. 7, no. 1, pp. 1–18, 2020, doi: 10.1186/s40708-020-00105-1.

[20] S. Panda, S. Mishra, M. N. Mohanty, and S. Satapathy, "Seizure detection using integrated metaheuristic algorithm based ensemble extreme learning machine," *Measurement: Sensors*, vol. 25, p. 100617, Feb. 2023, doi: 10.1016/J.MEASEN.2022.100617.

[21] S. Garg, R. K. Patro, S. Behera, N. P. Tigga, and R. Pandey, "An overlapping sliding window and combined features based emotion recognition system for EEG signals," *Applied Computing and Informatics*, vol. ahead-of-print, no. ahead-of-print, 2021, doi: 10.1108/ACI-05-2021-0130/FULL/PDF.

[22] A. Harishvijey and J. Benadict Raja, "Automated technique for EEG signal processing to detect seizure with optimized Variable Gaussian Filter and Fuzzy RBFELM classifier," *Biomedical Signal Processing and Control*, vol. 74, p. 103450, Apr. 2022, doi: 10.1016/J.BSPC.2021.103450.

[23] L. Abou-Abbas, K. Henni, I. Jemal, A. Mitiche, and N. Mezghani, "Patient-independent epileptic seizure detection by stable feature selection," *Expert Systems with Applications*, vol. 232, p. 120585, Dec. 2023, doi: 10.1016/J.ESWA.2023.120585.

[24] M. M. Hasan, S. Rahman, A. Sarkar, F. Khan, and A. Seum, "Different Methods of EEG Signal Analysis Using Power Spectral Density, ChronoNet and ResNest," *European Journal of Electrical Engineering and Computer Science*, vol. 7, no. 5, pp. 20–27, Oct. 2023, doi: 10.24018/EJECE.2023.7.5.551.

[25] A. K. Singh and S. Krishnan, "Trends in EEG signal feature extraction applications," *Frontiers in Artificial Intelligence*, vol. 5, p. 1072801, Jan. 2023, doi: 10.3389/FRAI.2022.1072801/BIBTEX.

[26] M. Shen, P. Wen, B. Song, and Y. Li, "Real-time epilepsy seizure detection based on EEG using tunable-Q wavelet transform and convolutional neural network," *Biomedical Signal Processing and Control*, vol. 82, p. 104566, Apr. 2023, doi: 10.1016/J.BSPC.2022.104566.

[27] T. Islam, M. Basak, R. Islam, and A. D. Roy, "Investigating population-specific epilepsy detection from noisy EEG signals using deep-learning models," *Heliyon*, vol. 9, no. 12, p. e22208, Dec. 2023, doi: 10.1016/J.HELIYON.2023.E22208.

[28] G. Dhiman and V. Kumar, "Seagull optimization algorithm: Theory and its applications for large-scale industrial engineering problems," *Knowledge-Based Systems*, vol. 165, pp. 169–196, Feb. 2019, doi: 10.1016/J.KNOSYS.2018.11.024.

[29] H. Mittal, A. Tripathi, A. C. Pandey, and R. Pal, "Gravitational search algorithm: a comprehensive analysis of recent variants," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7581–7608, Feb. 2021, doi: 10.1007/S11042-020-09831-4/METRICS.

[30] M. Natu, M. Bachute, and K. Kotecha, "HCLA_CBiGRU: Hybrid convolutional bidirectional GRU based model for epileptic seizure detection," *Neuroscience Informatics*, vol. 3, no. 3, p. 100135, Sep. 2023, doi: 10.1016/J.NEURI.2023.100135.

[31] "Epileptic seizures dataset | Kaggle." https://www.kaggle.com/datasets/chaditya95/epileptic-seizures-dataset (accessed Feb. 26, 2023).

[32] A. H. 1981- Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," 2009, Accessed: Mar. 26, 2023. [Online]. Available: https://dspace.mit.edu/handle/1721.1/54669.

[33] "EEG Dataset: Nonlinear Time Series Analysis (UPF)." https://www.upf.edu/web/ntsa/downloads (accessed Feb. 26, 2023).

[34] X. Huang, G. Jin, and W. Ruan, "Loss Function and Gradient Descent," pp. 103–107, 2023, doi: 10.1007/978-981-19-6814-3_9.

[35] J. M. Joyce, "Kullback-Leibler Divergence," *International Encyclopedia of Statistical Science*, pp. 720–722, 2011, doi: 10.1007/978-3-642-04898-2_327.

[36] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, May 2007, doi: 10.1016/J.ESWA.2006.02.005.

[37] Q. Yuan, W. Zhou, S. Li, and D. Cai, "Epileptic EEG classification based on extreme learning machine and nonlinear features," *Epilepsy Research*, vol. 96, no. 1–2, pp. 29–38, Sep. 2011, doi: 10.1016/J.EPLEPSYRES.2011.04.013.

[38] I. Ariza, A. M. Barbancho, L. J. Tardón, and I. Barbancho, "Energy-based features and bi-LSTM neural network for EEG-based music and voice classification," *Neural Computing and Applications*, vol. 36, no. 2, pp. 791–802, Jan. 2023, doi: 10.1007/S00521-023-09061-3/TABLES/12.

[39] K. Martín-Chinea, J. Ortega, J. F. Gómez-González, E. Pereda, J. Toledo, and L. Acosta, "Effect of time windows in LSTM networks for EEG-based BCIs," *Cognitive Neurodynamics*, vol. 17, no. 2, pp. 385–398, Apr. 2023, doi: 10.1007/S11571-022-09832-Z/FIGURES/11.

[40] P. Kunekar, M. K. Gupta, and P. Gaur, "Detection of epileptic seizure in EEG signals using machine learning and deep learning techniques," *Journal of Engineering and Applied Science*, vol. 71, no. 1, pp. 1–15, Dec. 2024, doi: 10.1186/S44147-023-00353-Y/TABLES/3.

[41] Z. Huang, Y. Ma, R. Wang, W. Li, and Y. Dai, "A Model for EEG-Based Emotion Recognition: CNN-Bi-LSTM with Attention Mechanism," *Electronics 2023, Vol. 12, Page 3188*, vol. 12, no. 14, p. 3188, Jul. 2023, doi: 10.3390/ELECTRONICS12143188.

[42] S. M. Omar, M. Kimwele, A. Olowolayemo, and D. M. Kaburu, "Enhancing EEG signals classification using LSTM-CNN architecture," *Engineering Reports*, p. e12827, 2023, doi: 10.1002/ENG2.12827.

[43] Y. Deng, S. Ding, W. Li, Q. Lai, and L. Cao, "EEG-based visual stimuli classification via reusable LSTM," *Biomedical Signal Processing and Control*, vol. 82, p. 104588, Apr. 2023, doi: 10.1016/J.BSPC.2023.104588.

[44] S. A. M. Mane and A. Shinde, "StressNet: Hybrid model of LSTM and CNN for stress detection from electroencephalogram signal (EEG)," *Results in Control and Optimization*, vol. 11, p. 100231, Jun. 2023, doi: 10.1016/J.RICO.2023.100231.

[45] C. Fan *et al.*, "ICaps-ResLSTM: Improved capsule network and residual LSTM for EEG emotion recognition," *Biomedical Signal Processing and Control*, vol. 87, p. 105422, Jan. 2024, doi: 10.1016/J.BSPC.2023.105422.