Chapter 4

Epileptic Seizures Classification

4.1 Introduction

The electrical activity within the brain is tracked by installing electrodes on the scalp of the brain. Yet, the EEG-based prediction model requires well-experienced neurologists to diagnose epilepsy accurately from the EEG recordings. Besides, the determination of epilepsy from the EEG signals is a time-consuming activity. The hand-operated ES classification process requires an entire day to recognize post-ictal, inter-ictal, and pre-ictal patterns from the EEG signals [16–18, 20, 21]. The classification mainly relies on categorizing brain activities between the states of EEG patterns mentioned above. The predetermination of ESs enhances the lifetime of epileptic victims in various circumstances as it generates an alarm before the onset of a stroke [16–18, 20, 21, 23]. Various existing methods are available to restrain the issues related to seizure classification, which provides high classification accuracy. Most seizure classification methods are PS, as the EEG signals vary across patients due to the variety of seizure patterns and positions. In the PS strategies, supervised learning procedures are utilized through two principal stages: extraction and characterization between pre-ictal and ictal states. Here, the term PS systems refers to those research works where techniques are evaluated by training the classifier on individual patient information and presenting a patients-wise analysis of the approach. In such cases, the methodology is tailor-made to fit the specifications of the patient's condition. Also, the seizure properties vary with patients and duration. Thus, it becomes challenging to capture the varying properties of information and learn them to attain the desired performance [16–18, 20, 21].

Most methods considered feature extraction and classification/prediction phases for seizure diagnosis. The feature extraction process mainly targeted the localization (univariate/bivariate) and linearity (linear/non-linear) characteristics [16–18, 20, 21]. With the advancement in signal processing, various research works have introduced multiple/hybrid features to express the brain dynamics better. Various classifiers have been utilized to learn such features for seizure recognition. In some cases, the feature dimensionality reduction method was adopted and found suitable for the seizure diagnosis [16, 20]. Generally, time, frequency, and time-frequency domain features are extracted to develop the feature space. After that, various ML algorithms, like NN and SVM, are operated to improve the diagnosis performance [117, 120, 121, 123, 125, 170, 171]. From the literature, it can be inferred that EEG signals are quite sensitive to noise. The addition of artifacts to these signals during recording makes it even more challenging to interpret them. Thus, it is important to have a higher-order complex feature space to express the seizure characteristics for the classification. Hence, this work constructs a hybrid feature set of non-linear features using EMD, EWT, and VMD.

Furthermore, the seizure recognition process has been revolutionized today by the DL-based approaches. They can potentially find complex relationships between the EEG data values to construct useful patterns for classification [16, 18, 20, 21, 169]. Amongst the various DL-based approaches, a CNN has also been adopted for seizure classification [18, 20, 21, 24, 113, 169, 172]. CNN has manifested its extraordinary performance on image data sets [173]. Hence, most researchers have represented raw EEG data sets as image/2D matrices and then used them as inputs for CNN to perform seizure recognition. They have used STFT [122], WT [118], Hjorth, statistical and spectral band power parameters [124] for generating an image/2D matrix. The literature informs that most of the approaches aim to use CNN for classification. Therefore, they have concentrated on converting the raw data into images without focusing on the property of raw EEG data sets. As a result, valuable temporal information of the EEG signal is lost [24]. Also, it is worth mentioning that the CNN architecture can identify descriptive features from the input and then perform classification on the data set. Also, the variation in the raw EEG signal length affects the proper image representation process.

Further, the DL models were trained on raw EEG data sets for the feature extraction/classification to overcome the issue mentioned above [20, 23, 126, 127]. The authors in [23] have used CNN for feature extraction and RNN for seizure detection. Also, a 13-layer deep CNN has been introduced in [126], which performs the automated detection and diagnosis of seizures using raw EEG signals. The critical issue with the abovementioned approaches is that they require a significantly large architecture that works directly on the raw signal. Thus, substantial computational resources are required to train a massive number of parameters and hyperparameters. It also adds overhead to tuning these parameters and hyperparameters. Besides CNN, LSTM has been operated on to understand the temporal characteristics of EEG signals [20, 169]. LSTM has been used to predict ESs by categorizing pre- and inter-ictal EEG signals [174]. This work gives a patient-wise analysis of the methodology's performance by providing various pre-ictal window segments as inputs. The authors in [175] have proposed double-layered LSTM for multi-class EEG signal classification through Auto-Regressive Moving Average with Hurst Exponent (HE) features. A BLSTM architecture has been introduced to extract before and after information of moments from the EEG signals by constructing a statistical feature space [114]. Similarly, WPT and a BLSTM-based PS approach have been reported in [119] for ES prediction during sleep. Identifying human activities and ES conditions have been performed by determining the correlation-based Q-score of features to train LSTM [176]. Various challenges are faced by the existing DL methods. One such example is [126]. In this work, the authors did not consider the feature extraction stage that minimizes the computational complexity. The 1D-CNN method in [127] has been employed to detect the seizure after the occurrence but fails to predict it. Similarly, a technique developed by [174] did not consider any preprocessing or post-processing methods that may assist in elevating the accuracy of prediction due to the AR and enhancing the quality of the image, respectively. Besides, the literature informs that the DL architectures proposed in various works are developed on a trial-and-error basis, as many hyperparameters are utilized in those architectures. As a result, the performance of proposed frameworks is influenced by non-optimized architectures. Thus, a novel optimization technique is proposed in this work to achieve a compact DL architecture.

This work provides a generalized methodology that can be used for ES classification while utilizing the information from the various phases of the ES. The innovative part of the proposed work is to generate a hybrid feature space for expressing the non-linear seizure information and then train the sequential DL model. In addition, a novel optimization algorithm has been proposed to optimize the DL model. This optimization algorithm is developed using BRO and SARO and is called Battle Royale Search and Rescue Optimization (BRRO). BRRO integrates fortification and analyzing skills of the defender and the TL, respectively, which is proposed in this research to enhance the classifier's performance. The proposed BRRO algorithm is used to tune the classifier's parameters to restrain the model's optimization issues. Thus, BRRO is a swarm intelligence-based algorithm advanced by amalgamating the defender's fortification

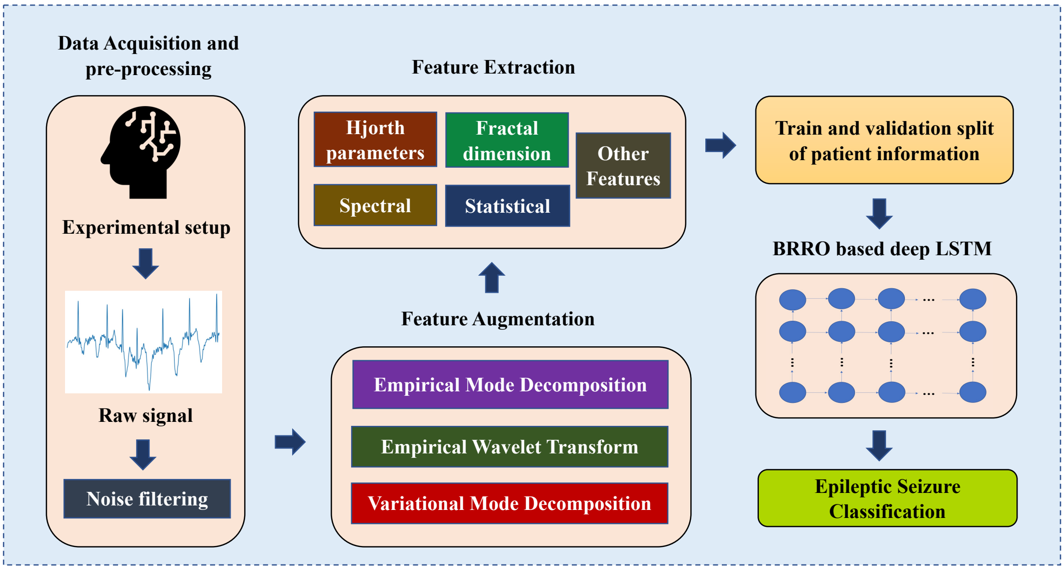


Figure 4.1: Illustration of the proposed methodology.

characteristics and the TL's analyzing skills for rescuing the abandoned inmates. Therefore, the main contributions of this research work are as follows:

1. In this work, a DL-based approach is proposed to mitigate the problem of PS seizure classification models. As proof of principle, a general data set is designed and developed to train the classifier based on non-specific patient information.

2. A feature augmentation approach is presented to handle the data scarcity issue in EEG data sets. As proof of concept, various advanced signal processing techniques were employed to enrich the quantity of data. Consequently, a hybrid feature space is generated to capture the potential non-linear patterns for seizure classification to improve performance.

3. This work presents an optimized DL model based on defender and TL characteristics to identify a better solution with a fast convergence rate.

4.2 Proposed Methodology

4.2.1 Signal Preprocessing

Signal preprocessing, the first and foremost step, removes the artifacts from the signals. Generally, the EEG signal is analyzed by neurologists or physicians to determine the brain's activity. However, the analysis and processing of the EEG signals become complex due to the artifacts present in them. These artifacts may be from noise sources, any irrelevant biological signal occurring while recording the EEG signal, or interference in the capturing device. Hence, removing such additive noises from the EEG signal is essential to reduce the adversarial effect on the performance of seizure classification [20, 21, 169].

Based on the above discussion, the proposed ES classification model utilizes a well-adapted band-pass filter. The band-pass filter allows the signal to pass through the desired frequency range while rejecting the other signal frequencies deviating from the designated boundary values. Thus, a FIR filter of 0.25 to 25.0 Hz is considered in the proposed work for preprocessing EEG signals. The specified frequency range has been used because most of the cerebral activity observed from sEEG falls under this range [1, 2, 169]. The FIR design used is 'firwin' since it uses a time-domain design technique that generally gives improved attenuation using fewer samples, and the window used for filter design is 'hamming'.

4.2.2 Feature Augmentation

The feature augmentation techniques are widely adopted in ML approaches to expand the signal/data by applying transformations over the sampled signal to enhance the available quantity of training data [18]. The data scarcity is greatly minimized using the feature augmentation process, improving the classifier's classification accuracy. Besides, the EEG signal is non-linear and non-stationary. As a result, the feature augmentation in this work has been performed through well-known adaptive time series decomposition methods [177], namely, EMD, EWT, and VMD, to enrich the feature space. The beauty of these methods is that they provide a simple and effective mechanism to analyze the signal's non-linear behavior. An insight into the augmentation methods is enumerated below.

4.2.2.1 Empirical Mode Decomposition

Proposed by Huang et al. [178], EMD is a posteriori and intuitive decomposition method often employed in analyzing biomedical signals, information detection, and other purposes. It is used for the decomposition of signals $S(t)$ into IMFs ($imf_i, i = \{1, ..., n\}$) [36] of finite size n. IMFs generated based on the local attribute time scale of the signal along with a residual $r(t)$ can be combined linearly to reconstruct the original signal as:

$$S(t) = \sum_{i=1}^{n} imf_i + r(t) \tag{4.1}$$

A constant or monotonic trend is presented by $r(t)$ and an oscillatory function, i.e., IMF. The IMF must adhere to the following conditions [36, 178]:

1. The number of zero crossings, as well as the number of extrema, should be equal or differ at most by one.

2. The mean value of the envelope characterized by the maxima and the envelope marked by the minima should be zero at any instance.

In conclusion, the underlying spectral pattern is elicited by EMD by removing the trend of $S(t)$.

4.2.2.2 Empirical Wavelet Transform

The EWT [179] is employed for feature augmentation in the proposed method to enable the analysis of non-stationary signals. The underlying principle of EWT is to segment the spectrum and develop a corresponding WFB. It is utilized to extract principle modes based on the segments derived from the Fourier Spectrum. Unlike other WT methods, it does not use predefined filter banks but operates upon the spectral information from the signal in an adaptive manner to explicitly build the filter bank.

In EWT, different modes of the EEG signal are extracted. Firstly, the FFT operation is performed to obtain the frequency spectrum of $S(t)$. Secondly, a set of local maxima (P_i, $i = \{1,...,n\}$) is identified from the spectrum to set boundaries between the two consecutive P. Let Q be the fixed number of modes, and then it is essential to determine the $Q+1$ frequency boundaries. Amongst these, ω_0 and ω_n are predefined boundaries. Then, spectrum segmentation is performed to extract modes and utilize the segments to construct the empirical WFB. After that, the filter bank is applied to the signal to attain the principal modes [120, 179].

In conclusion, EWT assists in analyzing the signal in multi-component by utilizing the advantages of both wavelet theory and Fourier analysis.

4.2.2.3 Variational Mode Decomposition

VMD is a novel tool for signal processing that closely relates to the Wiener filter and, thus, optimally mitigates the noise in the signal [180]. In VMD, various band-limited IMFs are generated non-recursively by decomposing the $S(t)$. It adaptively identifies the suitable band and concurrently calculates the respective modes to stabilize errors between them. The VMD technique uses a constrained variational optimization problem to extract modes from a signal. The problem is solved using the alternate direction method of multipliers [181, 182], and the

Table 4.1: Description of the features from the attribute set.

Feature	Parameters	Domain	Indicates
Hjorth	Activity	Time	Variance of time function
	Mobility	Time	Mean frequency or portion of SD of power spectrum
	Complexity	Time	Change in the frequency
Entropy	Tsallis	Time	Dynamics of signal
	Spectral	Frequency	Randomness of signal
Fractal dimension	Petrosian	-	Fractional dimensions of signal
	Higuchi	Time	Characterizing primary waves
Hurst Exponent	-	Time	Index of long-range dependence

number of modes is predefined. In this technique, the first modes u_k^1, central frequency w_k^1, and Lagrangian multipliers ϕ are initialized and then updated iteratively until the function is converged based on the convergence criteria.

In conclusion, VMD decomposes the signal into its principal modes, and thus, each mode has a particular property by preserving a specific frequency range.

4.2.3 Feature Extraction

The augmented signals were then subjected to feature extraction for the generation of highly informative features, which directly boosts the classification performance of the classifier [15, 16, 90, 96, 100, 103, 170, 183, 184]. Therefore, various feature extraction approaches were applied to the augmented signal. As a result, in this work, based on literature, multiple features like HP, entropy, FD, and HE were considered to generate a descriptive feature set for deep investigation [185]. The informative analysis of these features is given in Table 4.1. Besides, common features like root mean square amplitude, maximum amplitude, minimum amplitude, peak frequency, median frequency, number of extrema, and number of zero crossings were derived and utilized as a feature set member.

The developed hybrid feature set is rich in size and contains 164 features in the feature set. This diverse feature set is designed through various attributes and is expected to provide more descriptive and significant information for the classification. Also, using existing features to develop a hybrid feature set assists in evaluating each feature's capability to express the ES patient's meaningful information.

4.2.4 Proposed Optimization Technique based Deep LSTM

DL approaches have been widely operated in various domains like agriculture, health, automation, and others. Among different DL methods, Deep-LSTM has been employed successfully in various sequential data sets [186]. However, the performance of the classifier is significantly affected by the learnable parameters. Thus, the efficient tuning of the Deep-LSTM classifier is required in the proposed ES classification method to achieve decent performance. Hence, a pre-eminent optimization algorithm known as the BRRO algorithm is introduced to optimize the Deep-LSTM by combining the fortifying characteristics of the defender [187] and the exploring and rescuing characteristics of the leaders [188]. In the case of the standard BRO, the defender inherits the best position to fortify their extremities. In the SARO, the TL inherited the best position to rescue the abandoned inmates [188]. Hence, the hybridized BRRO is proposed, which inherits the defender and the TL's advantages to attain faster convergence. The characteristics of both the defenders and the TL boost the optimal global convergence rather than converging to the local optima points.

4.2.4.1 Motivation

Battle Royale is a widely played virtual game in which the surveying, fortifying, and competitive characteristics are widely explored. The individuals involved in the games are called defenders, whose aim is to reinforce their extremities from the enemy. The defenders must

Algorithm 1 Proposed Battle Royale - Search and Rescue optimization algorithm.

Input: I^t
Output: I^{t+1}

while $Iter \neq Iter_{max}$ **do**
    Randomly initialize the position of defender I^t_{def}
    Update the position of I^t_{def} by best position through Eq. 4.3.
    Gather evidence and form evidence matrix E
    Initialize the position matrix $[p]$ and Memory matrix $[M]$
    Initiate social phase to estimate the latest position I^{t+1}_{soc} using Eq. 4.5.
    Initiate independent phase to estimate the best position obtained by the team leader I^{t+1}_{TL} through Eq. 4.6.
    Integrate the best position I^{t+1} obtained by both the defender and the team leader by utilizing Eq. 4.8.
    Estimate the I^{t+1} via Eq. 4.11.
    $Iter = Iter + 1$
end while
Return I^{t+1}

select the best position to save their boundaries from the enemy and protect themselves. The main advantage of the standard BRO is that it provides better convergence compared with the other optimization algorithms [187]. On the other hand, the rescuing characteristic of human beings is finding abandoned inmates from their team. It is elucidated in the SRO optimization and is analyzed for the research. The TL of the team has to evaluate the directive about the abandoned inmates to find the best location to rescue those inmates. The prime advantage of SRO optimization is that it effectively handles single objective, continuous optimization issues [188]. In the proposed algorithm, these two characteristics of the defender and the TL were combined to develop the BRRO optimization algorithm to attain the best optimization solution.

4.2.4.2 Mathematical Model

This section provides a detailed mathematical model description of the proposed optimization algorithm for the DL model. Algorithm 1 demonstrates the pseudo-code of the proposed algorithm. Let us assume that LSTM architecture is used for ES classification. Assume that this LSTM architecture has one hidden layer called $L1$. For layer $L1$, there are W weights of dimensions, say $Y \times Z$. It gives us $W \times Y \times Z$ weight values. It helps us to form the population for the optimization algorithm.

a) Fortifying characteristics of the defender: The algorithmic procedure starts with the equal distribution of arbitrary populations among the search space. Every member of this population has a position and a destruction level. A uniform distribution initializes the positions p. In the next phase, each individual in the battleground, known as the defender, tries to defend themselves and fortify their extremities by attacking their opponent. It is assumed that the defender in the best position defeats their enemies based on an objective function, i.e., the accuracy of the classifier in this proposed method. The destruction level is enhanced by one of the defenders exterminating the opponent defender [187]. If the destruction level of the solution crosses a threshold, a new position is generated to respawn the defender. After experiencing the destruction, the defender is forced to shift their position to beat the opponent. The destructed defenders now move between the antecedent position and the best position. This is done to concentrate more on exploiting search space. These interactions were mathematically represented as follows:

$$I_{def}^{t+1} = I^t + r \cdot (I_{best} - I^t) \tag{4.2}$$

$$I_{def}^{t+1} = I^t (1 - r) + r \cdot I_{best} \tag{4.3}$$

where I_{best} is the position of the best solution obtained so far, I_{def} is the position of the destructed defender, and r is an arbitrarily generated number from a uniform distribution in the range [0,1]. If the destructed defender attacks the opponent in their next level, the destruction range is reorganized to zero. The defender is eliminated from the battlefield if the destruction level surpasses the predetermined threshold value. Hence, the best solution is obtained by removing the low-fitness solution. The current population with I_{best} is employed in the next stage.

b) Exploring and rescuing characteristics of the leaders: Further, the analyzing skill of the human to rescue abandoned inmates is explored in the algorithm to choose the best solution. The expedition behavior of the humans during the sort-out and rescue operation is highlighted in the algorithm. If an inmate was lost in the team, the team members formed a group under the TL to find and rescue the lost inmate. Hence, team cooperation is the preferred feature of the algorithm.

i) Gathering of evidence phase: The team members collected evidence on the lost inmates to begin the rescue operation. Hence, gathering evidence is an essential process in the search and rescue operation, which helps them to sort out the abandoned inmate. The evidence is of two types:

- Dominant evidence: The prevalent evidence gives vital information about the abandoned inmate. Hence, one of the team members has to search around the position of the evidence to find the inmate.
- Neglected evidence: The team members neglect this evidence to find more significant evidence to trigger the rescue operation. This is done based on the value of the objective function derived from this evidence. However, the information provided by the evidence is utilized to find the inmates.

To organize this information, firstly, a population of size N is initialized. However, for the exploring and rescuing characteristics, a population size of $2N$ is required to generate the evidence matrix [188]. A uniform distribution is used to create the rest of the N members of the population. Then, an evidence matrix E is created with two sections, p and M, which are position values of the previous stage, and M is the newly generated population. Here, a parameter called Failed Search (FS) is assigned for each member of the population. It indicates the number of times a member has failed to find significant evidence. The initial value of FS is 0 for each member. When a member encounters a critical clue (based on the objective function), its FS is assigned a value of 0, and else, it is incremented by 1. If the FS value goes beyond a threshold called Maximum FS (MFS), a new position is given to that member in the search space. E is generated according to the evidence collected by the members. The teammates' position is gathered in the position matrix p, and the neglected evidence is stored in the memory matrix M. Therefore, these matrices were updated in the search space.

$$E = \begin{bmatrix} p \\ \\ \\ M \end{bmatrix} = \begin{bmatrix} p_{11}p_{12}p_{13}\cdots p_{1q} \\ \vdots \\ p_{n1}p_{n2}p_{n3}\cdots p_{nq} \\ M_{11}M_{12}M_{13}\ldots M_{1q} \\ \vdots \\ M_{n1}M_{n2}M_{n3}\ldots M_{nq} \end{bmatrix} \tag{4.4}$$

The team members try to avoid exploring the same locations repeatedly to prevent wasting time. Hence, the exploration is done without interfering with the search space of the other members. Suppose the evidence gathered from the current position is more significant than the clues gathered from the previous position. In that case, the current status is updated, and the TL determines the best position to seek out the abandoned inmates. If the previous position is more effective than the current one, the situation will not change. Hence, the latest position is mathematically represented by:

$$I_{soc}^{t+1} = \begin{cases} E_{k,j} + r_1 \cdot (I_{i,j} - E_{k,j}), & \text{if} f_o(E_j) > f(I_i) \\ I_{i,j} + r_2 \cdot (I_{i,j} - E_{k,j}), & \text{otherwise} \end{cases} \tag{4.5}$$

where, I_{soc}^{t+1} is the new position of the human in the k^{th} human of the j^{th} dimension, and $E_{k,j}$ is the position of the evidence in the k^{th} evidence in the j^{th} dimension. r_1 is the arbitrarily generated number via uniform distribution in the range [-1, 1], and r_2 is the arbitrarily generated

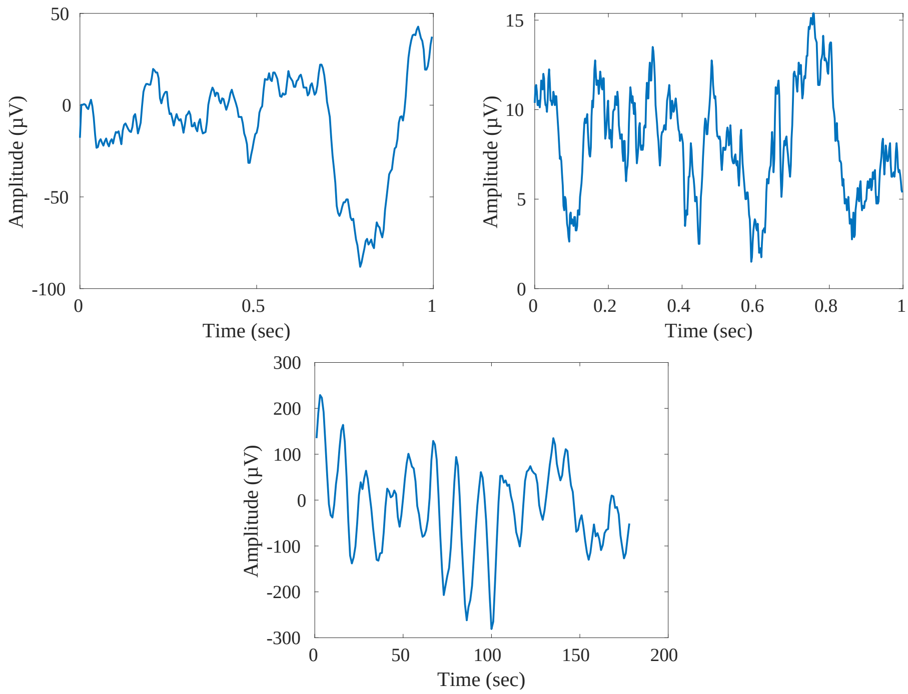


Figure 4.2: Sample EEG signals for (a) DB1, (b) DB2, and (c) DB3.

number through uniform distribution [0, 1].

ii) Independent phase: The group members were then distributed in the different positions determined by the TL to accelerate rescue operations. The ideas from the social phase relate to other evidence and were employed to attain the best position. Hence, the best place obtained by the TL for the rescue operation is mathematically expressed as:

$$I_{TL}^{t+1} = I^t + r_3 \cdot (E_k - E_v) \tag{4.6}$$

where r_3 is a random number generated from a uniform distribution in the range [0,1], and the position of evidence E_k is mathematically expressed as,

$$E_k = I^{t-1}(1-r) + r \cdot I_{best} \tag{4.7}$$

c) Integrating phase: The next phase employed in the proposed BRRO algorithm lies in sorting out the best solution by integrating the best position obtained from the defender in the battle royale and the TL in the rescue operation. The integrated best solution is mathematically expressed as in [189].

$$I^{t+1} = 0.5I_{def}^{t+1} + 0.5I_{TL}^{t+1} \tag{4.8}$$

From Eqs.4.3 and 4.6, the following equation is obtained:

$$I^{t+1} = 0.5I^{t}(1-r) + 0.5r \cdot I_{best} + 0.5I^{t} + 0.5r_{3} \times (E_{k} - E_{v}) \tag{4.9}$$

$$I^{t+1} = 0.5I^{t}(1-r+1) + 0.5r \cdot I_{best} + 0.5r_{3}(E_{k} - E_{v}) \tag{4.10}$$

$$I^{t+1} = 0.5I^{t}(2-r) + 0.5r \cdot I_{best} + 0.5r_{3}[I^{t-1}(1-r) + r \cdot I_{best} - E_{v}] \tag{4.11}$$

The above equation is the proposed update rule for BRRO that selects the optimal weights and biases for the Deep-LSTM classifier, exhibiting better convergence characteristics. In the proposed optimization, the tendency to converge towards the global optimal points is enhanced by minimizing the chances of converging towards the local optimal points. The proposed equation is developed through the hybrid characteristics of fortification and exploring the defender and the TL characteristics. Unlike the standard optimizations, the proposed algorithm updates the solution for every iteration. Also, based on the global best solutions, the best solution in the previous instances and optimization parameters were adapted to achieve diversification and intensification towards optimal global convergence.

d) Termination phase: The phases were repeated until the maximal iterations finally yielded the global best solution for fine-tuning the weights and biases of the Deep-LSTM classifier.

In conclusion, the novel BRRO algorithm is a hybrid algorithm developed using the BRO and SARO to optimize the Deep-LSTM to improve its performance. BRO is a virtual game-based optimization algorithm known to reach the global optima in less time and is well suited for high-dimensional optimization problems [187]. Similarly, SARO is proven to perform better in solving constrained optimization problems [188]. Therefore, these algorithms were chosen for hybridization and were expected to provide improved classification performance when used for parameter tuning of the Deep-LSTM classifier.

4.3 Result Analysis

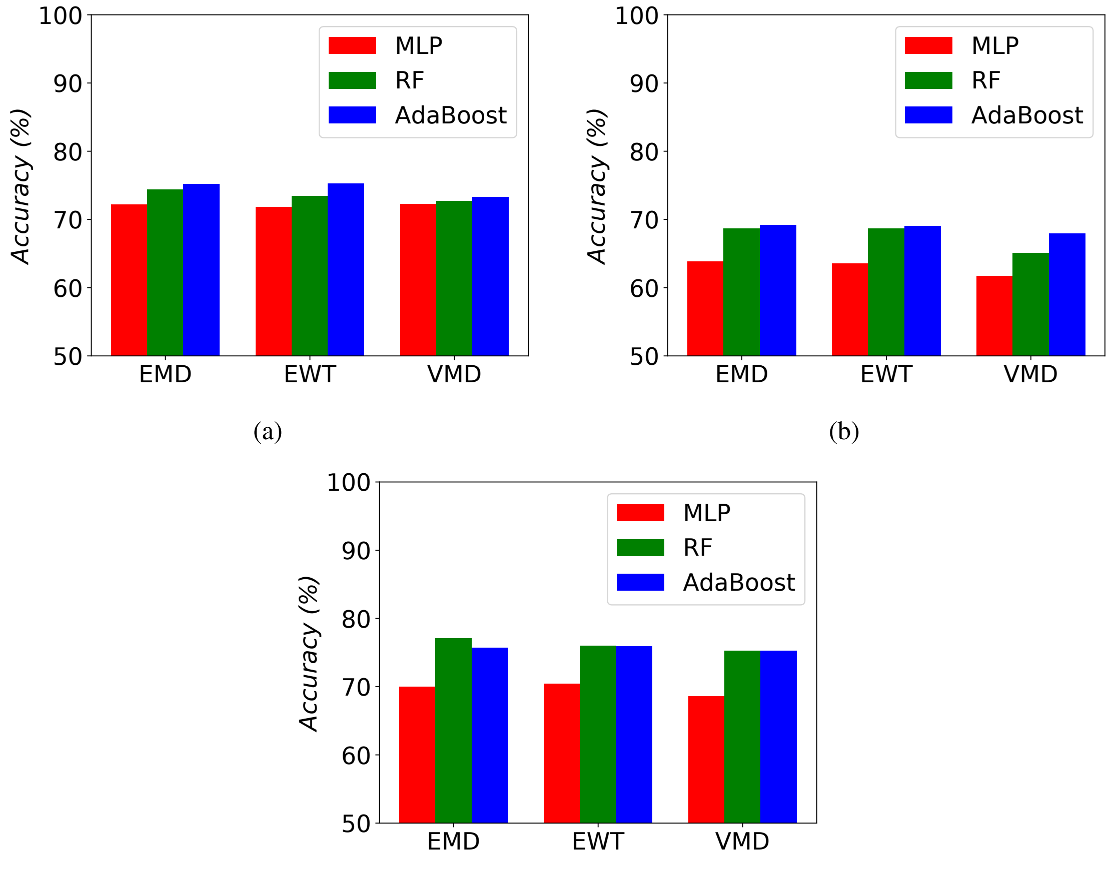
The proposed work has been compared with ML techniques such as MLP, RF, Adaptive Boosting (AdaBoost), and DL techniques such as Deep CNN and Deep-LSTM classifiers. The performance metrics used to evaluate the classification techniques are accuracy, sensitivity, and specificity. The reason for choosing these performance metrics was that when dealing with medical data, it is to be ensured that the number of False Positive (FP) and False Negative (FN) is the least [17, 18, 138, 190, 191]. These metrics have been calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.12}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{4.13}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4.14}$$

Here, TP is True Positives, and TN is True Negatives. The proposed model is based on non-specific patient information, i.e., this approach is patient-generic; thus, the results presented are for training and testing accuracy. Another reason is to indicate that the proposed model does not suffer from overfitting the training data set. Thus, training and testing accuracies have been provided. For the data sets used in this research work, the approach is multichannel and patient-independent, i.e., the proposed methodology utilizes the data obtained from all the channels and aims to build a generic model for all epileptic patients. Also, in the proposed work, the ictal, i.e., seizure-active period, has been considered class 1, and other periods like pre-ictal, inter-ictal, and post-ictal as class 0, i.e., seizure-free, to perform the classification task. An epoch length of



(a)

(b)

(c)

Figure 4.3: Performance metrics for feature augmentation techniques on (a) DB1, (b) DB2, and (c) DB3.

100 seconds is assumed in the current work [192]. The elaborate feature set helps the classifier achieve better learning, and it also helps in avoiding overfitting [185].

4.3.1 Evaluation of Feature Augmentation Techniques

The most commonly used ML techniques are employed for comparison and analysis. Initially, the individual feature augmentation techniques are evaluated using ML methods, and the results are presented in Figure 4.3. The performance of DB1 for different ML methods with feature augmentation techniques ranges from 71.81% to 75.24%, and the best performance of 75.24% is reported for EWT with AdaBoost. Similarly, the accuracy of DB2 for various ML meth-

Table 4.2: Performance metrics for various machine learning methods used for classification of seizures.

Classifier	Training Accuracy(%)	Testing Accuracy(%)	Sensitivity (%)	Specificity (%)
		DB1		
MLP	75.29	75.04	75.53	75.29
RF	75.94	75.69	75.80	76.33
AdaBoost	79.18	77.87	77.95	78.57
		DB2		
MLP	65.69	64.93	68.28	69.00
RF	69.24	67.72	70.70	74.41
AdaBoost	76.31	70.36	88.01	89.43
		DB3		
MLP	70.86	70.18	71.82	70.62
RF	78.36	75.07	74.57	82.95
AdaBoost	79.85	79.59	76.82	83.69

ods with feature augmentation techniques scales from 61.73% to 69.17%, and the maximum accuracy of 69.17% is observed for EMD with AdaBoost. Likewise, the accuracy of DB3 for multiple ML methods with feature augmentation techniques spans from 68.59% to 77.1%, and the top performance of 77.1% is noted for EMD with RF. Besides, it has been discovered that the variation in the accuracy for different ML methods with feature augmentation techniques is significant for DB2 and DB3 compared to DB1. Also, AdaBoost is the best performing classifier on feature augmentation techniques. VMD is the worst-performing feature augmentation technique in comparison to EMD and EWT.

4.3.2 Evaluation using Machine Learning Techniques

Based on the performance reported in Fig 4.3, it can be concluded that the individual feature augmentation techniques could not capture meaningful information from the data set for seizure classification to achieve decent performance. Thus, a hybrid feature set is developed to learn the features for boosting the classifier performance. Table 4.2 indicates that the AdaBoost is the dominant classifier for all the data sets. The training and testing accuracy of the Adaboost

Table 4.3: Performance metrics for various DL methods used for classification of seizures.

Classifier	Training Accuracy(%)	Testing Accuracy(%)	Sensitivity (%)	Specificity (%)
		DB1		
CNN	87.37	82.06	82.97	81.96
LSTM	87.86	87.46	87.61	86.09
		DB2		
CNN	76.75	73.91	89.00	89.99
LSTM	81.53	77.16	89.32	90.21
		DB3		
CNN	86.20	85.86	89.13	84.13
LSTM	86.47	86.41	89.50	84.30

classifier is more than 76.0% with all the data sets. However, the best performance of 78.16% is recorded for the DB1. Also, MLP and RF provide an accuracy of more than 70.0% for DB1 and DB3. MLP's performance is inferior on all the data sets compared to the other classifiers, and the lowest accuracy of 64.93% is recorded for DB2. For DB1 and DB3, the performance of all the classifiers is more than 75.0% and 70.0%, respectively. The deviation in performance ranges from 2.83% to 9.41% among all the data sets. However, a notable difference is noticed for DB2 and DB3, which are 5.43% and 9.41%, respectively. Also, the error rate has been reduced by 10.62%, 3.85%, and 15.13% on the hybrid feature set on the AdaBoost classifier compared to the individual feature augmentation techniques.

The effect of the hybrid feature set is observed through sensitivity and specificity. The values are more than 69.0% for all the data sets. However, the best value is recorded for the Adaboost classifier, which is more than 76.0% for all the data sets. A decent result is observed for the DB2 with a sensitivity and specificity value of more than 88.0%. Also, for DB1 and DB2, the values are more than 75.0% and 70.0%, respectively.

4.3.3 Evaluation using Deep Learning Techniques

Based on the performance reported in Table 4.2, it can be concluded that the ML methods cannot capture the temporal information from the data set for seizure recognition. Thus, the DL models are employed to learn the non-linear features for performing the seizure diagnosis. In DL techniques, benchmark methods like CNN (Layers: 2, Filters: 32, Kernel: 3×3, Max pooling) and LSTM are selected for analyzing the hybrid feature set. The LSTM utilized in the proposed method has six layers: one output layer, four hidden layers, and one input layer. The batch size used is 32, the maximum epoch is 100, the dropout layer is 0.2, and MSE is considered as the loss function. The result depicts the testing output acquired by the proposed method. Along with the mentioned hyperparameters, CNN utilizes zero padding and ReLU activation function except for the last FC layer, which operates the sigmoid function. Similarly, the LSTM layer has 48 neurons, each with a learning rate 10^{-3}.

Table 4.3 represents the performance achieved by DL methods. The LSTM shows a slightly better performance than CNN. The difference between the performance ranges from 1.0% to 4.0% for all the data sets. The best performance of 87.46% is reported for the LSTM on DB1. Similar to the ML approach, lower performance is reported for DB2, and it is 73.91% with CNN. The DL approach achieves an accuracy of 82.0% or more for DB1 and DB2. Tables 4.2 and 4.3 indicate that the error rate is reduced due to DL methods. It is nearly 19.0%, 12.0%, and 31.0%, respectively for DB1, DB2, and DB3.

Besides, a significant improvement in the sensitivity and specificity values is noticed. This improvement is due to the construction of a hybrid complex feature set and the ability of the DL method to learn them. Also, it is worth mentioning that the feature set preserves the inherent temporal property of EEG signals. Hence, the LSTM can perform better than the other com-

Table 4.4: Performance metrics for various optimizations-based Deep-LSTM used for classification of seizures.

Classifier	Training Accuracy(%)	Testing Accuracy(%)	Sensitivity (%)	Specificity (%)
		DB1		
PSO-Deep-LSTM	87.92	87.91	89.00	86.26
BSO-Deep-LSTM	89.23	88.32	89.53	87.21
BRO-Deep-LSTM	89.67	89.47	90.36	88.34
SARO-Deep-LSTM	89.81	89.59	91.35	89.07
Proposed method	**89.91**	**89.88**	**96.71**	**89.88**
		DB2		
PSO-Deep-LSTM	83.15	78.53	91.31	90.74
BSO-Deep-LSTM	88.96	88.39	92.69	91.39
BRO-Deep-LSTM	89.37	89.21	94.60	93.06
SARO-Deep-LSTM	90.37	90.29	94.72	94.44
Proposed method	**93.34**	**92.59**	**94.83**	**96.82**
		DB3		
PSO-Deep-LSTM	86.95	86.64	89.94	84.82
BSO-Deep-LSTM	89.29	87.24	90.04	89.44
BRO-Deep-LSTM	89.80	89.56	90.42	90.09
SARO-Deep-LSTM	90.56	90.42	91.09	90.91
Proposed method	**93.87**	**92.13**	**94.83**	**93.84**

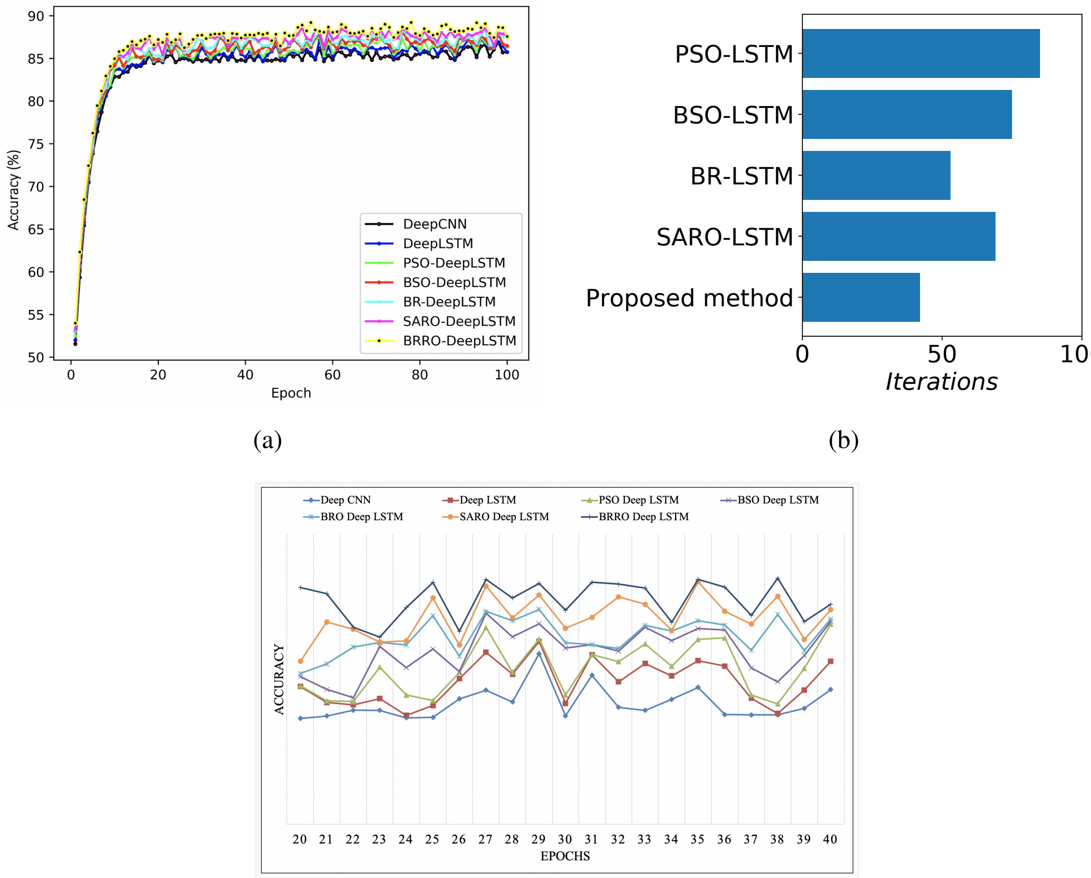
petitive methods. The error reduced due to LSTM compared to CNN is 30.1%, 12.45%, and 3.88%, respectively, for DB1, DB2, and DB3. Thus, in further work, LSTM has been explored more to improve seizure classification performance.

4.3.4 Evaluation using Optimization Techniques

The weights associated with each layer are the defenders and the leaders. Since these algorithms are human-based, the initial population is set according to the number of weight values. There are 51,227 trainable parameters in the utilized architecture. The various parameters of the proposed optimization algorithm are: population size = 25, uniform distribution range = [-100, 100], destruction level = 2, threshold = 3 [187], MFS = 50 [188].

Table 4.4 demonstrates the results of optimization algorithms on the Deep-LSTM classifier. The proposed optimization algorithm is compared with other state-of-the-art methods like PSO, BSO, BRO, and SARO. It is worth mentioning that the optimization algorithms on Deep-LSTM have helped improve the performance of seizure classification, and the enhancement ranges from 1.0% to 15.0% for all the data sets. Except for PSO-Deep-LSTM on DB2, all the optimization algorithms with Deep-LSTM have attained more than 86.0% performance. Also, PSO-Deep-LSTM performance is inferior on all the data sets compared to the other classifiers. Besides, the proposed optimization algorithm has shown an accuracy of 90.0% or more for all the data sets. The best performance of 92.59% is recorded for the proposed method on DB2. Also, SARO-Deep-LSTM performance is slightly inferior to the proposed method on DB1 and DB2. The various optimization algorithms on the Deep-LSTM demonstrated decent performance. However, the variation in the performance scale from 1.9% to 14.1% among all the data sets. A significant variation is observed in DB2 and DB3, i.e., 14.1% and 6.5%, respectively. Also, the proposed method has reduced the error rate by 19.0%, 68.0%, and 42.0%, respectively, for DB1, DB2, and DB3. Also, the fusion of BRO and SARO has reduced the error rate by 3.0%, 23.0%, and 17.0%, respectively, for DB1, DB2, and DB3. The results indicate that the proposed hybrid feature set and optimized Deep-LSTM have improved the diagnosis process.

Moreover, the performance of various optimization algorithms on Deep-LSTM for DB1 is observed while varying the number of epochs, as shown in Figure 4.4. A similar performance pattern is observed for DB2 and DB3 for various optimization algorithms on Deep-LSTM. Figure 4.4 and Table 4.4 indicate that BRRO is considered the leading algorithm compared to other optimization algorithms. Also, it can be noted that in Figure 4.4(c), the optimization-based classifiers follow an almost similar trend due to the common underlying classifier, i.e., Deep-LSTM. The variations that can be seen are due to the optimization algorithms. As can be seen from the results, BRRO can converge on a better accuracy in fewer epochs. Also, for the



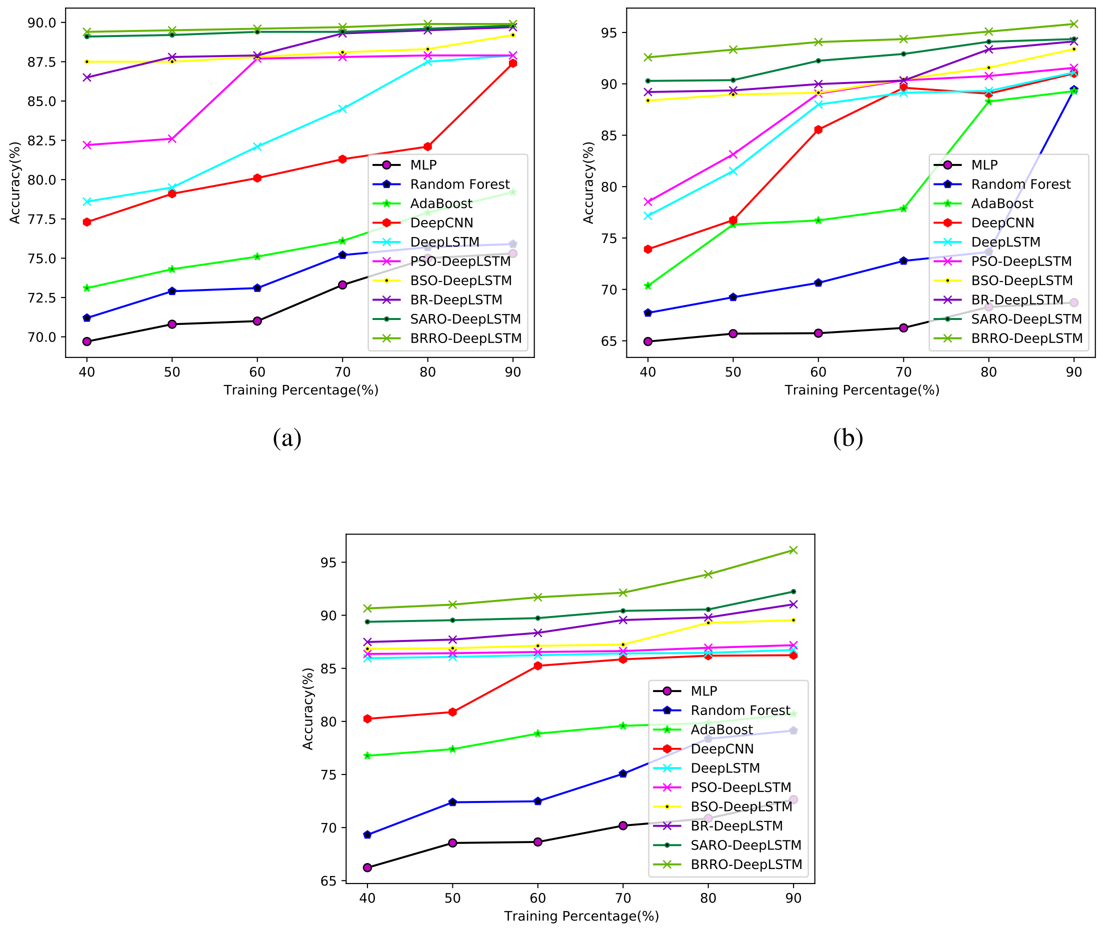
(a) (b)

(c)

Figure 4.4: (a) Performance of the classifiers based on accuracy for varying values of epochs achieved by the classifiers for DB1 and (b) a portion of curve (a) for epochs 20 to 40. (c) The number of iterations each optimization algorithm takes for one of the epochs for DB1.

optimization algorithms, the maximum number of iterations assumed is 100. It is observed that the proposed novel BRRO algorithm converged faster to the global optimum as compared to the other optimization algorithms, as seen in Figure 4.4(b).

Furthermore, the enhancement in the sensitivity and specificity has proven the effectiveness of the proposed paradigm. Table 4.4 shows that the proposed novel algorithm, i.e., BRRO-Deep-LSTM has the best performance in terms of specificity and sensitivity. The improved specificity and sensitivity achieved during classification suggest that when classifying seizure and non-seizure, there are fewer FN and FP compared to the classification performed by LSTM and other optimization-based classifiers. Figure 4.4 shows the performance of the classifier



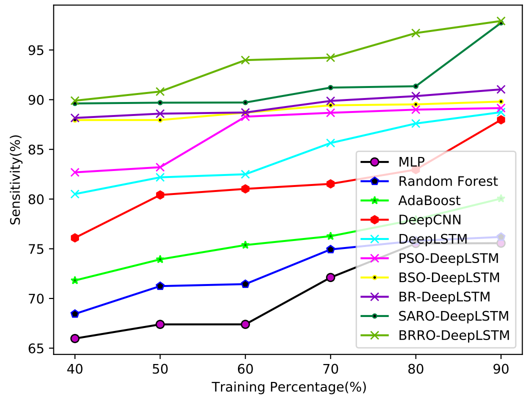
(a) (b)

(c)

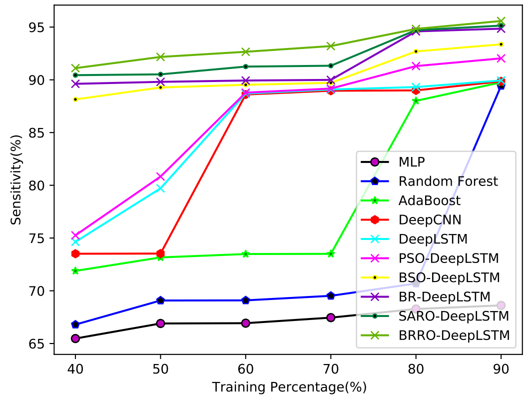
Figure 4.5: Performance comparison of the classifiers based on accuracy for varying values of training percentage achieved by the classifiers for (a) DB1, (b) DB2, and (c) DB3.

comparison for DB1. Also, a similar observation is reported for other data sets. The performance is enhanced with an increase in the epoch value for all the classifiers. It proves that the proposed algorithm can achieve better accuracy than other optimization methods in the same number of epochs. This is established by running the experiments for several epochs till 50, and it is observed that the novel optimization BRRO-Deep-LSTM gives the best accuracy. To confirm further accuracy improvements, the number of epochs was increased to 100. It was observed that the accuracy remains the same, ensuring that epoch runs are sufficient for all models.

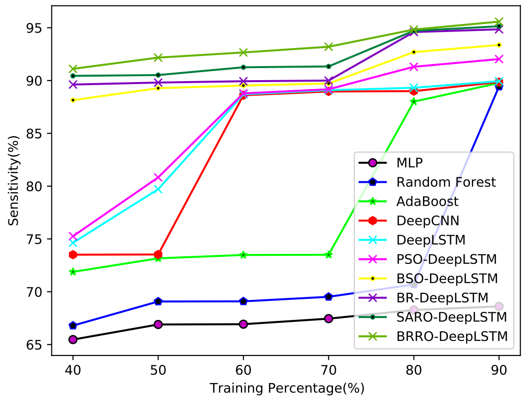
The training percentage was varied to assess the classifiers' performance. It was observed



(a)



(b)

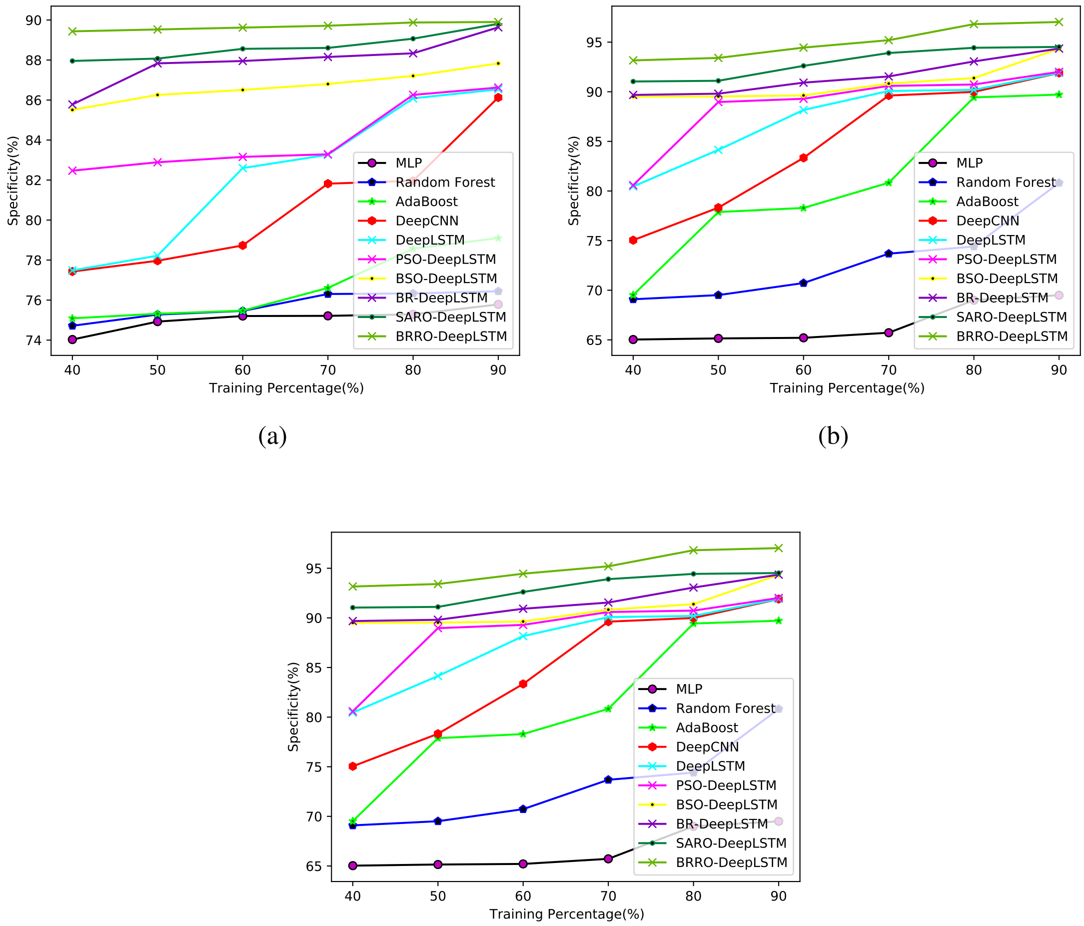


(c)

Figure 4.6: Performance of the classifiers based on sensitivity for varying values of training percentage achieved by the classifiers for (a) DB1, (b) DB2, and (c) DB3.

that the proposed methodology is less sensitive to variations in partition size and fold size. Initially, 40.0% of data is chosen randomly for the training of the classifiers and the remaining for testing. The training data percentage is gradually increased by 10.0% in each step to verify the effect on the proposed model. Figure 4.5 illustrates that the proposed BRRO-Deep-LSTM has better accuracy for different values of training percentage.

The plots in Figure 4.6 and 4.7 show that the novel algorithm, i.e., BRRO-Deep-LSTM, has the best performance compared to sensitivity and specificity, respectively. The BRO-based and SARO-based Deep-LSTM came close to the performance of the novel optimization BRRO-Deep-LSTM but could not beat it. The improved specificity and sensitivity achieved during



(a) (b) (c)

Figure 4.7: Performance of the classifiers based on specificity for varying values of training percentage achieved by the classifiers for (a) DB1, (b) DB2, and (c) DB3.

classification suggest that when classifying seizure and non-seizure, there are fewer FN and FP compared to the classification performed by LSTM and other optimization-based classifiers.

4.3.5 Comparison with State-of-the-art Methods

The proposed work is compared with the state-of-the-art methods reported in the literature, as shown in Table 4.5. The literature informs that DB1 is the most widely adopted data set for developing various seizure detection approaches. The proposed method is less superior to the other methods. However, the state-of-the-art methods are PS and thus, achieve performance almost nearer to 100.0%. One PS approach uses an event-based K-fold CV to identify each

Table 4.5: Comparison with state-of-the-art methods.

Study	Data set	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
[174]	DB1	LSTM	100.0	100.0	99.28
[172]	DB1	CNN	99.30	98.82	99.43
[113]	DB1	CNN	99.54	88.14	99.62
[193]	DB1	GDP	95.38	94.47	94.16
[193]	DB2	GDP	96.05	96.05	96.61
[194]	DB3	CNN	99.5	100.0	100.0
[195]	DB3	NN	95.0	-	-
[196]	DB3	Ensemble	90.0	-	-

patient's seizures. In this, K represents the number of seizures per patient. The model training has been performed in K rounds, if a subject had K seizures, $K-1$ seizures were selected for training in each round, and the remaining ones were used for testing. Thus, the results of previous methods are better than the proposed technique. Nevertheless, the problem of PS seizure classification models was mitigated in the proposed work. Since the seizure signature changes with the patient, the proposed method is generic and does not depend on the PS approach. As proof of principle, a general data set is designed and developed to train the classifier based on non-specific patient information [113]. The studies showcased in [195, 196] are patient-independent, similar to the work proposed in this chapter. The accuracy achieved on DB3 in [195] is 95.0%, and in [196], it is 90.0%. The ensemble approach in [196] uses classifiers like NN, SVM, KNN, and Bayes. Each classifier learns the feature space separately and overcomes the non-coverage seizure region problems due to a single classifier. As a result, it achieves a decent performance for seizure classification.

Also, most methods use the CNN approach for classification, trying to transform the data sets into image/2D matrices. The literature informs that the CNN architecture provides decent performance. In fact, in some cases, it has achieved results up to 100.0%. However, concerns for the CNN approaches are the size of the architecture, i.e., several layers used to achieve the accuracy. It is performed by capturing the specific information from lower to higher representations, moving from the first to the last layer. As the size increases, the computing resource requirement enhances simultaneously, and the time taken by the model is substantial. Also, with an increase in the size of architecture, the model becomes complex. As a result, the chances of overfitting increase consequently. These cases apply to both 2D- and 1D-CNN. In addition, it is worth mentioning that CNN generally tries to capture the specific information from lower to higher representations. Besides, the LSTM architecture's complexity is comparatively less, and as a result, the resource requirement is reduced. Also, various gates in the LSTM cell allow them to learn the time series data properly [24]. Thus, in the proposed work, LSTM is used to explore the temporal property of data without converting it into an image/2D matrix.

In one of the PS approaches event-based K-fold cross-validation was used for detecting all seizures of each patient. In this, K was the number of seizures per patient. If a subject had K seizures, the model training is performed K rounds. In each round, $K-1$ seizures are selected for training, and the remaining one is used for testing. Thus, the results of previous methods are better than ours. But, in the proposed work, we have mitigated the problem of PS seizure prediction models. As a proof of principle, a general data set is designed and developed to train the classifier based on non-specific patient information.[113]

4.4 Summary

This study suggests a unique hybrid optimization strategy for determining the optimized Deep-LSTM architecture to improve seizure detection performance. The non-linear properties of EEG signals may be expressed demonstratively using the hybrid feature set. After that, Deep-LSTM is used to classify seizures. The parameters control the classifier's performance; hence, Deep-LSTM is further optimized using BRRO. The hybrid optimization method combines the defender's fortification with the TL's analytical abilities. It assesses the appropriate Deep-LSTM parameters while achieving quicker convergence than other traditional optimization techniques. The performance and architecture of the classifier have benefited from the proposed technique. The numerous EEG data sets highlighted the need for a hybrid feature set. Additionally, a compact classifier architecture is achieved with good seizure classification performance.

In this chapter, it is seen that the classifiers can perform better due to the optimization algorithm and can successfully detect epileptic seizures from EEG signals. Taking this a step further, this technique can help warn patients about their upcoming seizures. Therefore, in the next chapter, a novel DL method is proposed. This approach combines a hybrid optimization algorithm and feature augmentation for predicting seizure occurrences using EEG data. The approach creates a patient-generic model to address data scarcity in epilepsy patients. A hybrid deep feature set is constructed and analyzed. The proposed approach efficiently selects optimal LSTM parameters, leading to faster convergence than conventional methods.