

Chapter 5

Epileptic Seizures Occurrence Prediction

5.1 Introduction

It is observed that 50.0% of epileptic patients experience vigorous seizures. These seizures negatively influence the patients as they experience seizure-related uneasiness to varying degrees in their regular routines. Epileptic patients are at a greater danger of accidents and wounds because of the seizure-related occurrences alone, just as having high depression rates [14]. The unpredictable nature of epilepsy, such as the time duration of a seizure, the time of occurrence, the degree of intensity, and others, makes it undeniably challenging for patients to perform ordinary tasks. Sometimes, the seizure events can be dangerous to the life of individuals. As indicated by the WHO, millions of individuals worldwide experience the ill effects of epilepsy, with around 33.0% of these patients being resilient to anti-epileptic medicine. Thus, the early determination of these seizures before they happen will be helpful for the patients to go for on-time medical treatment and enable the on-time application of the preventive therapies [22]. An EEG is an examination that records the brain's activities with the help of numerous electrodes placed on the scalp. Even while sleeping, brain cells interact via electrical impulses and remain

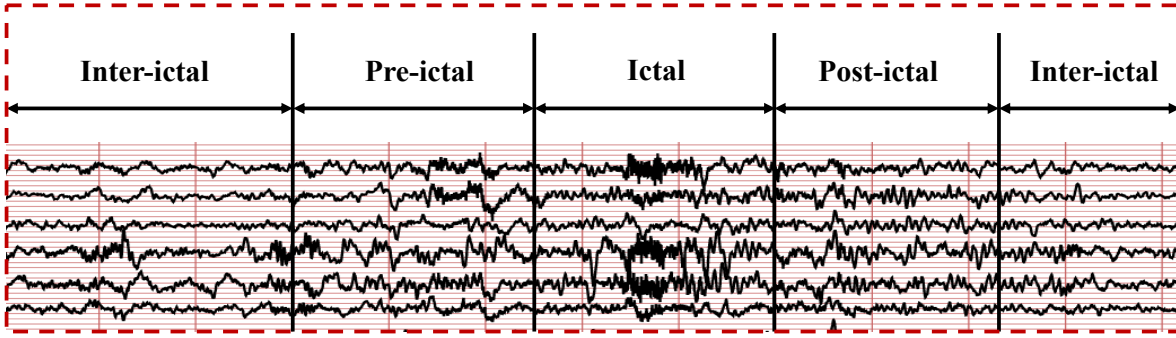


Figure 5.1: Various stages in an Electroencephalogram signal when a seizure occurs. This sample is from one of the edf files of case chb01 from the DB1 database [1, 2].

continually active. On an EEG recording, this activity appears as wavy lines. There are various stages of a seizure, as shown in Figure 5.1. The stage that begins with the onset of the seizure and terminates with it is called an ictal stage [197]. The post-ictal stage follows the ictal stage and proceeds for a few minutes. The inter-ictal stage determines the normal function of the brain. The pre-ictal stage refers to the aberrant activity of the brain that occurs 60 to 90 minutes before the commencement of a seizure [99]. To summarize, the pre-ictal stage defines the brain activity before a seizure, the post-ictal stage establishes the state of the brain just after a seizure, and the inter-ictal is the state of the brain without any seizures.

The literature suggests that significant effort has been made to achieve preferable performance through various extracted attributes from signals and various classifiers to understand EEG. Besides, a good potential has been shown by DL methods by directly or indirectly (via feature space) learning the EEG signal. However, the preferable combination of feature and classifier is still unknown, and most proposed approaches are PS. As a result, the proposed techniques are not generalized and need appropriate modification with the new data availability. Apart from this, the seizure characteristics vary with time and with individuals. Also, significant variation in duration is observed between the inter-ictal and pre-ictal stages. Thus, this work aims to develop an ES prediction model that forecasts the occurrence of ESs by analyzing the

EEG signals gathered from epileptic patients. The signals are preprocessed to remove the noise, which increases the quality of the signal. From the denoised signal, the most significant features are extracted to enhance the prediction accuracy of the proposed framework. The prediction of the ES is carried out through the proposed novel optimized SLSTM. The optimization algorithm called FB-SARO is developed by integrating the investigation and searching characteristics of human beings. The classifier's parameters are tuned with the help of this novel optimization algorithm to classify ES and predict their occurrences. The following are the primary contributions of this research work:

1. This work proposes an approach to forecast the occurrence of a seizure episode and raise the alarm for it. It is achieved by training the classifier to find the optimal SPH and achieve early prediction.
2. This work addresses the issue of PS seizure prediction models by putting forth a DL-based solution. A broad input set is created and prepared as a proof of concept to train the classifier using generic patient data.
3. As proof of concept, the data scarcity issue of EEG data sets is alleviated by providing a feature augmentation strategy that employs several adaptive signal decomposition algorithms to enhance the quantity of data. A hybrid feature space is created to boost performance and capture potential non-linear patterns for seizure classification.
4. This study provides an improved optimization-based DL model based on the steps taken during a criminal investigation and the dynamics of a group that has lost a comrade.

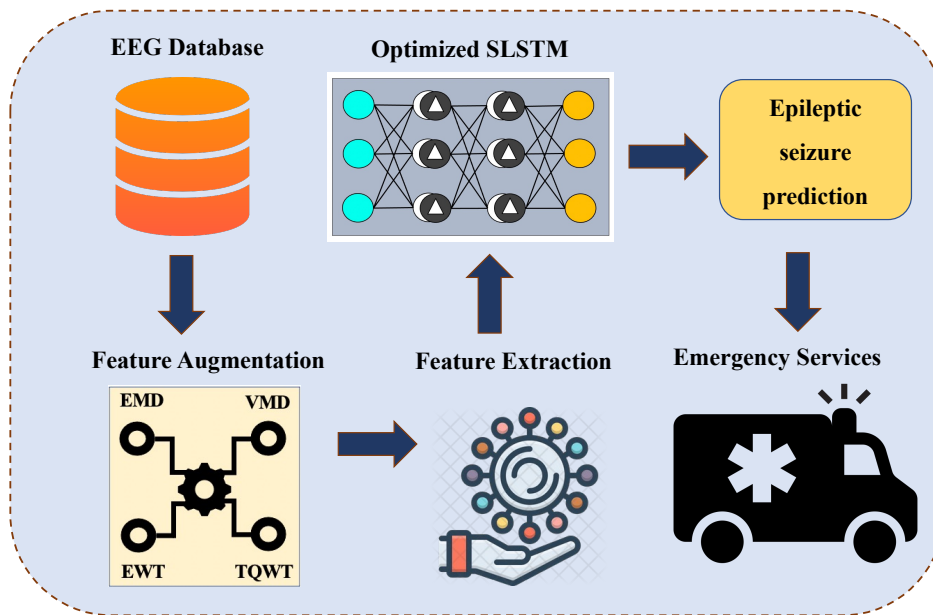


Figure 5.2: Proposed method.

5.2 Proposed Methodology

The proposed methodology designed for ES prediction has been elaborated on in this section. This methodology uses EEG signals, optimization algorithms, and DL techniques. It comprises various steps illustrated in Figure 5.2. Further, each part has been described in detail in the following subsections.

5.2.1 Preprocessing of Input Signals

It is a vital step in designing and developing a BCI-based application. Since the input EEG signal taken from an epileptic patient may contain noise or artifacts, these unwanted artifacts can be in the form of interference from the recording device, eye blinking, limb movement, and others. Since these signals are not essential in ES prediction, the input EEG signal is refined through a band-pass filter [20, 21, 169]. Generally, a band-pass filter filters in signals of a particular frequency range from the input signal and filters out irrelevant frequencies.

For this research work, a frequency range of 0.25-25.0 Hz has been considered for constructing an FIR band-pass filter. This specific range has been used to preprocess the input EEG signals because when any cerebral activity is noticed during an sEEG recording, it is observed to fall in this range [1, 2, 169]. The band-pass filter has a 'firwin' design because it employs a time-domain design method, often resulting in better attenuation with fewer samples and a 'hamming' window.

5.2.2 Feature Augmentation

The next step in this methodology is feature augmentation, which is a derivative of the data augmentation step applied to combat the problem of small data sets [18]. It is used to overcome the challenge of overfitting issues due to the data set size. Data augmentation has shown promising results in computer vision applications by increasing the data set size using various techniques. For non-stationary, non-linear time series data, like EEG signals, the proposed solution employs a variety of adaptive decomposition algorithms. The adaptiveness of these methods makes them highly efficient and helpful [177]. The signal augmentation techniques used and their descriptions are given as follows.

5.2.2.1 Empirical Mode Decomposition

In 1998, Huang et al. [178] proposed EMD, a signal decomposition technique that adapts to the characteristics of a time series. EMD is operated in power signal analysis, bearing fault detection, seismic signals, and others [198]. It is also utilized for the analysis of biomedical data. EMD decomposes the signals $S(t)$ into several IMFs ($imf_i, i = \{1, \dots, n\}$) [36] which must follow certain constraints (see [178]). These finite-size IMFs can be combined linearly along with a residual $r(t)$ to reconstruct the original signal as :

$$S(t) = \sum_{i=1}^n imf_i + r(t) \quad (5.1)$$

Lastly, EMD derives an elementary spectral trend by eliminating the pattern of $S(t)$.

5.2.2.2 Empirical Wavelet Transform

This approach allows for the analysis of non-stationary signals [179]. Unlike other WTs, which employ predefined Wavelets Filter Bank (WFB), the core principle of EWT is to generate a WFB by first segmenting the Fourier spectrum. This technique is operated to decompose the processed signal into IMFs [120, 179]. To perform EWT, the following steps are performed :

- Firstly, the frequency spectrum of $S(t)$ is drawn using the FFT.
- The spectrum gives a set of local maxima P_i , where $i = 1, 2, \dots, n$ to set boundaries between two succeeding P_i .
- Let the fixed number of modes be Q , then $Q + 1$ frequency boundaries are to be identified. Amongst these, ω_0 and ω_n are already defined.
- Next, extraction of modes is executed by segmenting the spectrum. These segments give the empirical WFB, which helps in attaining the modes.

Lastly, the benefit of Fourier analysis and wavelet theory is recognized by the EWT approach to investigate $S(t)$ in multi-component.

5.2.2.3 Variational Mode Decomposition

In signal processing, VMD is one of the versatile tools that nearly correlates with the Wiener filter and, as a result, consummately overcomes the noise in $S(t)$ [180]. Generally, a non-recursive

approach is applied in VMD to foster band-limited IMFs through decomposed $S(t)$. The appropriate band is flexibly located and simultaneously determines the relevant modes for balancing the errors between them. A constrained variational optimization problem is an underlying principle adopted by VMD to produce the modes from $S(t)$. The predefined quantity of modes and the alternate direction method of multipliers are adopted to resolve the problem [181, 182]. In VMD, Lagrangian multipliers ϕ , first modes u_k^1 , and central frequency w_k^1 are initialized. After that, (λ, u_k^1, w_k^1) are recursively improved till the convergence criteria is satisfied [180]. Lastly, the approach assists in generating the principal modes of $S(t)$, and most importantly, specific characteristics of each mode are observed due to the perseverance of a particular frequency range.

5.2.2.4 Tunable-Q Wavelet Transform

The TQWT provides the conditions for investigating the oscillatory signals through tunable quality factor (Q-factor) [14, 199]. Thus, TQWT is widely employed in biomedical signal processing. The $S(t)$ decomposed to generate several high-pass sub-band and a single low-pass sub-band. The required adjustable attributes to decompose $S(t)$ are the number of levels in the decomposition stage, redundancy (oversampling rate), and Q. The TQWT approach uses the two-channel filter banks to generate several high-pass sub-band and a single low-pass sub-band based on several levels in the decomposition stage. The redundancy assists in restricting the wavelets to preserve their shape in the time domain, and a low Q-factor transform supports highly non-oscillatory $S(t)$. Lastly, the TQWT technique assists in performing the precise decomposition of $S(t)$ for optimum time-frequency representation.

5.2.3 Feature Extraction

The next stage in the proposed methodology is feature extraction. After performing feature augmentation of $S(t)$, the feature extraction process is executed to generate an elaborate and communicative feature set. It enhances the learning being performed by the classifier and, thus, enhances the classification performance [15, 16, 90, 96, 100, 103, 170, 183, 184]. Consequently, the current work identifies several features like HE, FD, entropy, HP, and so on of time and frequency domain based on literature review [185]. Besides, the most frequently observed attributes, like the number of zero crossings, the number of extrema, median frequency, peak frequency, minimum amplitude, maximum amplitude, and root mean square amplitude, are determined and nominated as attributes in the developed feature set.

The proposed descriptive feature set has 275 attributes. The heterogeneous and rich feature space is introduced to express more differentiating and unique patterns for the diagnosis. In addition, the feature set supports and encourages the investigation of the feature space elements' potential to reveal the relevant and vigorous patterns/facts of ES patients.

5.2.4 Proposed Optimization Technique based SLSTM

The proposed ES prediction method is based on EEG signals recorded for various seizure patients. The recording was performed for a particular period. Also, the brain performs certain complex activities continuously, and thus, it can be observed as a time series. In this regard, the classifier can learn the brain EEG data signature through the robust and intelligent method to explore the time-series properties of the data set. Thus, sequential DL models can be valuable for understanding complex brain activities through EEG data sets. Hence, LSTM is employed in the proposed work for ES prediction. Also, it has been successfully operated on multiple

time series data sets [200]. However, it is worth noting that a remarkable deviation in time is observed between various seizure stages. It becomes challenging due to considerable variation in seizure properties with different patients and time stamps. Therefore, it is essential to investigate the EEG data to provide an environment for learning at different timescales. As a result, the proposed seizure prediction methodology utilizes SLSTM, a variant of LSTM with multiple LSTM layers.

Furthermore, the parameters of a classifier play an essential role in the learning of the classifier and directly affect the classifier's performance. As an effect, the current work proposes a novel optimization algorithm for tuning the classifier's parameter. Consequently, it boosts the learning being performed by the classifier, thus achieving superior performance compared to state-of-the-art techniques. The novel FB-SARO algorithm combines two optimization techniques, i.e., FBIO [201] and SARO [188]. These algorithms are based on human behavior. The former is based on the procedure police officers and detectives follow to investigate a crime, tracking down the suspect and the suspect's location—the latter deals with the steps taken by a TL who has lost an inmate from their team. The team collectively searches for the location to find the lost team member. It was observed that both these algorithms had an intersecting step to find the best location or the most probable location where the suspect (in the case of FBI) or the lost team member (in the case of SARO) would be found. This step worked as a merging point between these algorithms and gave us the novel optimization algorithm. This proposed novel algorithm effectively converges to a globally optimal solution.

5.2.4.1 Motivation

The proposed novel algorithm combines two behaviors seen in humans and their intellectual characteristics. Since homo-sapiens are a social species, they cooperate to thrive and survive.

On one end, when a team of people loses a member, the team members and the TL work together to look for that teammate and rescue them as soon as possible. For this task, the team searches for clues and analyzes them to extract information. This activity helps narrow down the best location where their lost aide can be found during the search operation. On the other end, when a crime occurs, humans form groups to investigate the crime and collect clues that provide information regarding the suspect. Further, information regarding various locations is analyzed to narrow down the suspect. Another team uses this information to pursue the suspect. These characteristics motivated the research direction to create the proposed novel FB-SAR optimization algorithm.

5.2.4.2 Mathematical Model

Assume that an LSTM architecture is being utilized to predict ESs and that $L1$ is the name of the hidden layer in this LSTM design. There are W weights of dimensions, let us say $Y \times Z$, for layer $L1$. This gives $W \times Y \times Z$ weight values. This aids in the population formation process for the optimization method, where each member is assigned a parameter called position initialized by a uniform distribution. This parameter helps evaluate the fitness of each member/solution in the population.

An investigation starts once a crime is reported. It includes identifying the physical evidence, collecting information, and interrogating the suspects. All possible suspects and their positions are determined, and these positions are established within the search space. The locations and the evidence are stored in a location matrix l and a memory matrix M , respectively. The dimensions of the l and the M are the same and stored in a solution matrix. It is mathematically represented as,

$$S = \begin{bmatrix} l \\ M \end{bmatrix} = \begin{bmatrix} l_{11}l_{12}l_{13} \dots l_{1d} \\ \vdots \\ l_{N1}l_{N2}l_{N3} \dots l_{Nd} \\ M_{11}M_{12}M_{13} \dots M_{1d} \\ \vdots \\ M_{N1}M_{N2}M_{N3} \dots M_{Nd} \end{bmatrix} \quad (5.2)$$

where l_{N1} denotes the position of the 1st dimension for N^{th} suspect and M_{1d} represents the location of the d^{th} dimension for first memory. The teams are formed to evaluate the feasible hiding position of the suspects inside the search space, and these positions are known as suspected locations. The search area is characterized by determining the location with the highest probability based on the classification accuracy of the DL technique used, which is the objective function of the proposed optimization algorithm. Then, the tracking team is organized. Now, all the search agents in the tracking team are directed toward the predetermined position to detain the suspect.

The tracking team must follow the headquarters' directions and move toward the suspected locality. In the meantime, it keeps reporting all the information to the headquarters. The exploring team, the tracking team, and the pursuit team communicate throughout various stages. The exploring team directs the tracking team to explore the localities so that the tracking team can upgrade the information and enhance its accuracy for the succeeding evaluations. The proposed algorithm comprises two stages: the exploring stage (A) and the tracking stage (B). The explorers carry out the exploring stage, and the search agents carry out the pursuit stage. In the exploring stage, the i^{th} suspected locality to be explored is denoted as l_{A_i} , where $i = 1, 2, \dots, N_A$ and N_A is the number of suspected locality need to be explored. In the tracking stage, l_{B_i} denotes

the agent's location i , and N_B is the number of search agents in the tracking process. In this research, let us assume that $N_A = N_B = N$, in which N is the population size. The investigation is a recurrent process and terminates when the current iteration I attains the maximum number of iterations I_{gmax} .

- Step A1: This process demonstrates the 'explication of findings' The exploring team evaluates all the gathered information and determines the probable suspect localities. Every suspected locality is investigated for prior observation. In this research, each individual is assumed to progress under the influence of other individuals. The general formula for the progression is given as follows:

$$l_{A1ij} = l_{Aij} + ((ra - 0.5) * 2) * \left(\sum_{a=1}^{a_1} l_{Aaj} \right) / a_1 \quad (5.3)$$

where, $j = 1, 2, \dots, d$, d denotes the number of dimensions, ra represents the random number within the range of $[0, 1]$; the $a_1 \in 1, 2, \dots, N - 1$, denotes the number of individual that effects l_{A1ij} 's movement. For the numerical evaluations of Eq. 5.4, $a_1 = 2$ provides the best outcome within a short period of execution time. The newly suspected locality l_{A1ij} is demonstrated in the Eq. 5.4, ρ_{A_i} demonstrates the possibility of the suspect being at the location l_{A_i} . The explorers evaluate the locations with the best chance (objective value) of finding the suspect to keep it, while the others will be abandoned.

$$l_{A1ij} = l_{Aij} + ((ra1 - 0.5) * 2) * (l_{Aij} - (l_{A_{kj}} + l_{A_{hj}}) / 2) \quad (5.4)$$

The k, h and i are the suspected location, where $k, h, i \in 1, 2, \dots, N$, $ra1$ in the range $[-1, 1]$.

- Step A2: This step is known as the 'direction of inquisition' Investigations compare the possibility of the suspected locations to those of others. The worst and the best (lowest and the highest probability) are represented as ρ_{worst} and the ρ_{best} . The probability of each

location, $Prob(l_{A_i})$, is estimated using Eq. 5.5, and a high value of $Prob(l_{A_i})$ corresponds to a high probability for the location.

$$Prob(l_{A_i}) = (\rho_{worst} - \rho_{A_i}) / (\rho_{worst} - \rho_{best}) \quad (5.5)$$

The position of other suspected locations influences the search location update. It is to be noted that the change is only made in some directions. Some randomly selected directions are changed in the updated location. This is done to boost the diversity of the search spaces. In this step, the best individual and other random individuals affect the movement of l_{A_i} . The move is mathematically represented as the following:

$$l_{A2_i} = l_{best} + \sum_{b=1}^{a_2} \varepsilon_b * l_{A_{bj}} \quad (5.6)$$

where l_{best} is the best location, a_2 indicates the number of people that have an impact on the movement of l_{A2_i} , $a_2 \in 1, 2, \dots, N - 1$, $b = 1, 2, \dots, a_2$, the effective coefficient of the individual to move is represented as $\varepsilon_b = [-1, 1]$. Computational evaluations give $a_2 = 3$. The new suspected locality is generated, and a probability is then calculated to determine whether the suspected locality is to be updated or not.

$$l_{A2_{ij}} = l_{best} + l_{A_{dj}} + ra2 * (l_{A_{ej}} - l_{A_{fi}}) \quad (5.7)$$

where, d, e, f, i denotes the suspected locations and $ra2$ is in the range $[0, 1]$. $d, e, f, i \in 1, 2, \dots, N$, and d, e, f , and i are chosen randomly.

- Step B: This step demonstrates the actions taken after the best locality of the suspect is obtained from the exploring team. Each agent B_i from the pursuit team now approaches the best locality. Thus, the newly approached location is mathematically represented as,

$$l_{B_{ij}} = l_{best} * ra3 + l_{A_{dj}} + ra4 * (l_{best} - l_{B_j}) \quad (5.8)$$

The l_{best} represents the best-updated location provided by the exploring team in step A1, and $ra3, ra4$ are in the range $[0, 1]$. The Eq5.8 determines the best-updated solution obtained from the investigation skills of humans.

- Step C: Conversely, the search agent provides an advanced exploring behavior to discover the lost team member. They analyze the evidence they obtained on the search space to determine the best position and update it based on the evidence they collected during the exploration stage. The revised position based on the exploring characteristics is mathematically represented as,

$$l_{exp}^{t+1} = E_{kj} + ra5 \times (l_{kj}^t - E_{kj}) \quad (5.9)$$

E_{kj} represents the position of evidence, l_{kj}^t denotes the position of human at instance t and $ra5$ is randomly generated using a uniform distribution ranged between $[-1, 1]$.

- Step D: This step is known as the integration step. In this, the best position of the suspect attained by the investigating skill of the human is integrated with the best position of the lost team member acquired by the exploring behavior of the humans based on [189]. The newly obtained position is given as,

$$l_{global}^{t+1} = 0.5l_{Bij} + 0.5l_{exp}^{t+1} \quad (5.10)$$

Substituting the Eq. 5.8 and 5.9 in Eq. 5.10, the following equation is obtained

$$l_{global}^{t+1} = 0.5(l_{best} * ra3 + l_{A_{dj}} + ra4 * (l_{best} - l_{B_j})) + 0.5(E_{kj} + ra5 \times (l_{kj}^t - E_{kj})) \quad (5.11)$$

$$l_{global}^{t+1} = 0.5ra6 \left[(2 * l_{best}) - l_{B_j} + \frac{1}{ra6} * (2E_{kj} + l_{A_{dj}} + l_{kj}^t) \right] \quad (5.12)$$

Here, $ra6$ is a random number generated from a uniform distribution between $[0,1]$. Integrating the two characteristics provides the best global solution as it combines the advantage of the exploring and investigation features of humans in Eq. 5.12. All these steps are repeated until it attains the maximum iteration to obtain an optimal global solution for tuning the parameters of the SLSTM classifier.

Finally, the novel FB-SAR Optimization algorithm is a hybrid of FBI and SAR optimization algorithms and has been used to optimize the parameters of an SLSTM classifier. FBIO is a meta-heuristic algorithm based on the process carried out by federal agents and detectives to solve a crime. On the other hand, SARO is also a metaheuristic algorithm based on humans' search and rescue behavior when a team loses one of its members. FBIO has an outstanding ability to explore, and it converges faster. These properties are compared to various algorithms, some of which are symbiotic organisms search, teaching–learning-based optimization, FPA, and others [201]. This algorithm takes minimal computational effort to identify solutions for continuous non-linear functions while ensuring high accuracy. On the other hand, SARO finds solutions to constrained optimization problems [188]. It has proven to be efficient for the same. Therefore, these two algorithms were chosen for hybridization for parameter tuning in this research work. To the best of our knowledge, this is the first time FBIO and SARO have been combined. The corresponding flowchart (see Figure 5.3) has also been provided.

5.3 Experimental Results and Discussion

5.3.1 Result Analysis

This section discusses the experiments executed in the current work, their respective outcomes, and their reviews. The four publicly available EEG data sets are employed to conduct the exper-

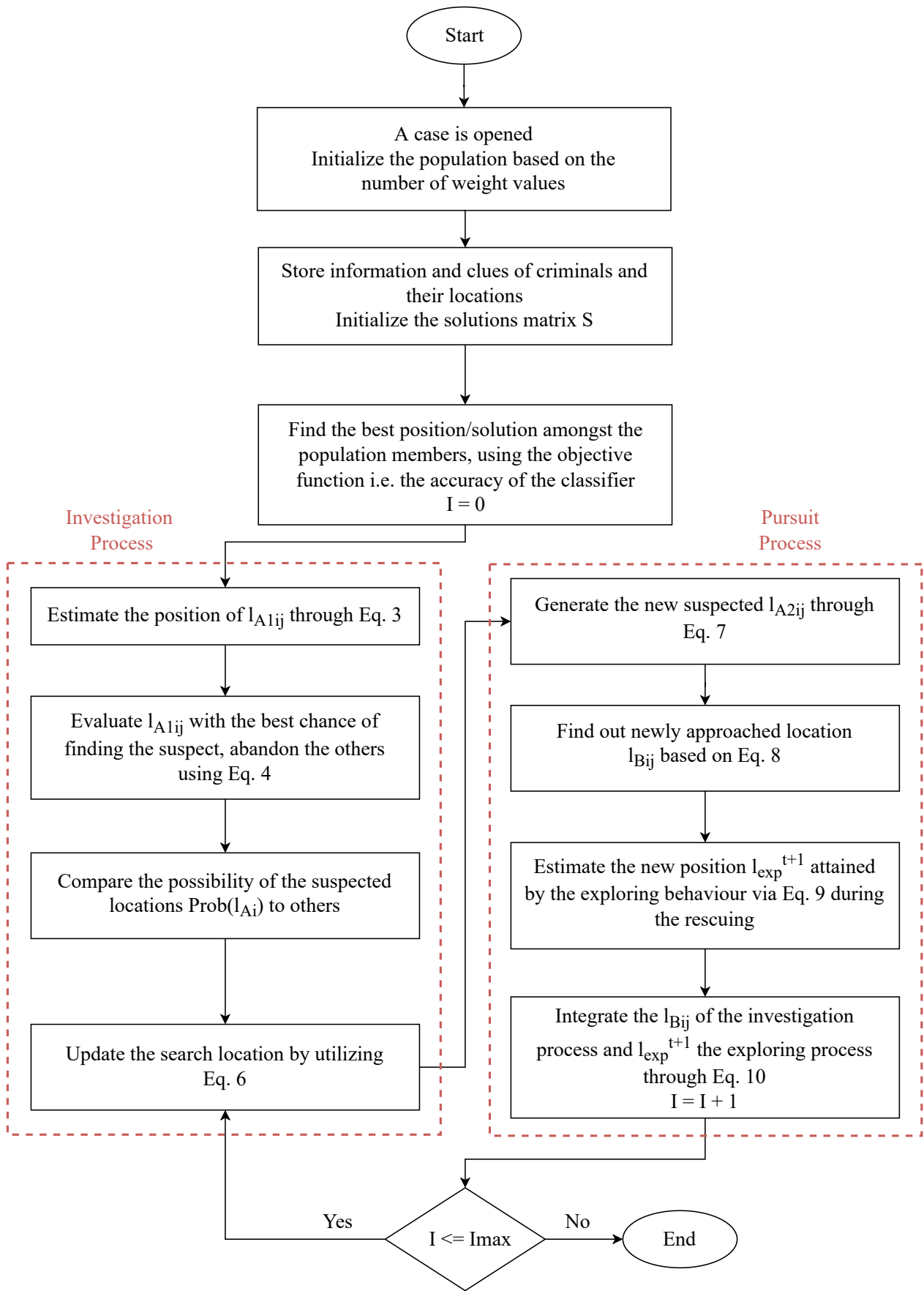


Figure 5.3: The flowchart of the Proposed Optimization Algorithm.

iments. To compare the proposed work, Linear Regression (LR) and NN from the ML family and Stacked RNN (SRNN), Stacked CNN (SCNN), and SLSTM from the DL family are selected. Various performance metrics like accuracy, sensitivity, specificity, and FPR are utilized to gauge the classifiers' performance. The rationale behind choosing these metrics, particularly sensitivity and specificity, is the data set being used, i.e., clinical data. When handling clinical data, the number of FP and FN is vital. The value of FP and FN should be minimum; therefore, this becomes a good measure to assess the performance of a classifier used to build an automated system [17, 18, 138, 190, 191]. The same goes for FPR; the number of false alarms must be minimal. The accuracy, sensitivity, and specificity are defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.13)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (5.14)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5.15)$$

$$FPR = \frac{FP}{FP + TN} \quad (5.16)$$

The input EEG data of epilepsy patients are utilized to develop the paradigm. For creating a general model for all epileptic patients of the data sets used in this research, the proposed approach leverages the data obtained from all the channels of the patient's data for all the patients, making this work patient-generic. In the current study, epoch duration is considered to be 100 seconds [192]. The comprehensive feature collection aids improved learning and helps prevent overfitting [185]. The 10-fold CV approach was used to conduct the studies.

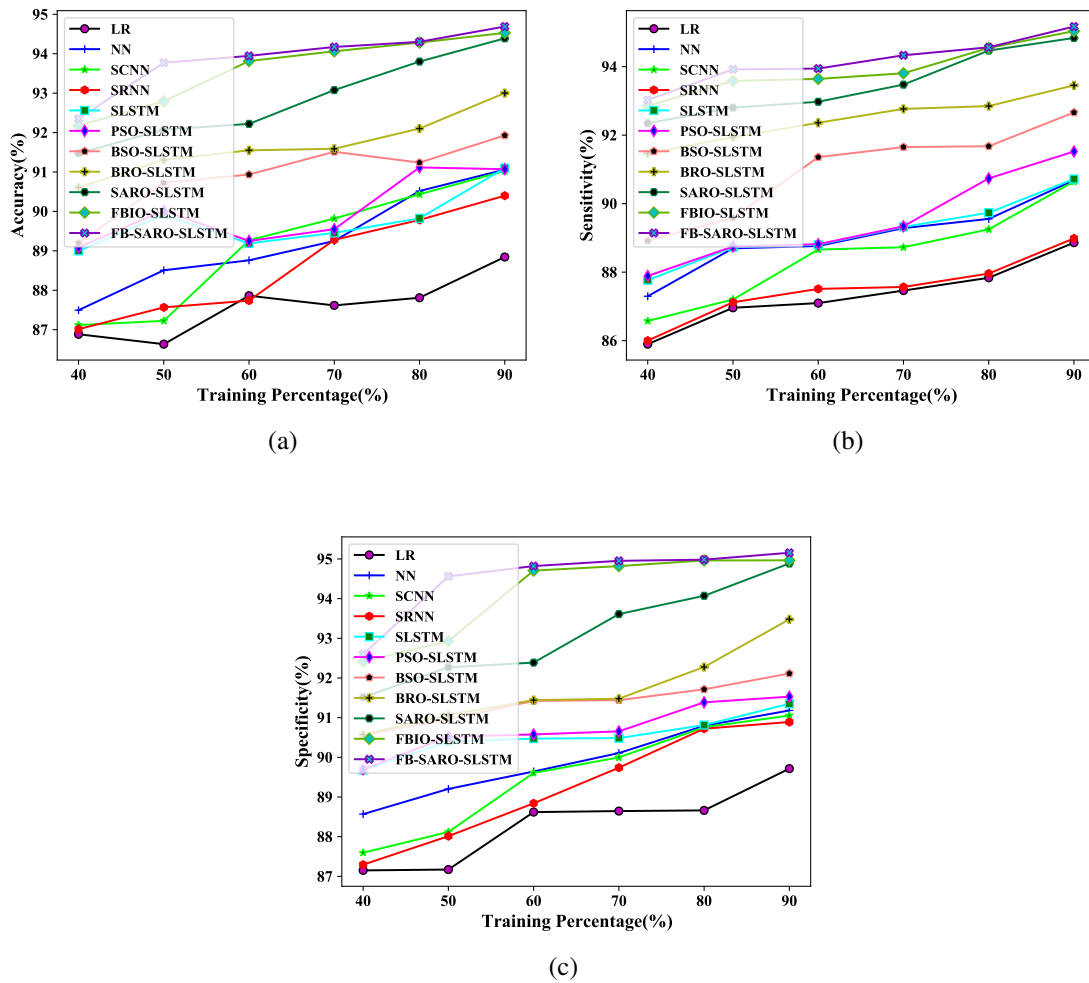


Figure 5.4: Comparison of the (a) Accuracy, (b) Sensitivity, and (c) Specificity of the classifiers' performances on DB1 for various values of the training percentages.

5.3.1.1 Performance Evaluation Based on Usage of Augmentation Techniques and Machine Learning Techniques

In this research, experiments were performed with several decomposition techniques to augment the data sets. These augmentation techniques help increase the data set size and handle the problem of overfitting. For this work, EMD, EWT, VMD, and TQWT were utilized for augmentation. The augmentation techniques were operated separately and together in this research work to analyze their effect. The ML techniques used to classify the derived feature set gave accuracy ranging from 83.90% to 88.39% for LR and 85.16% to 89.40% for NN for individual

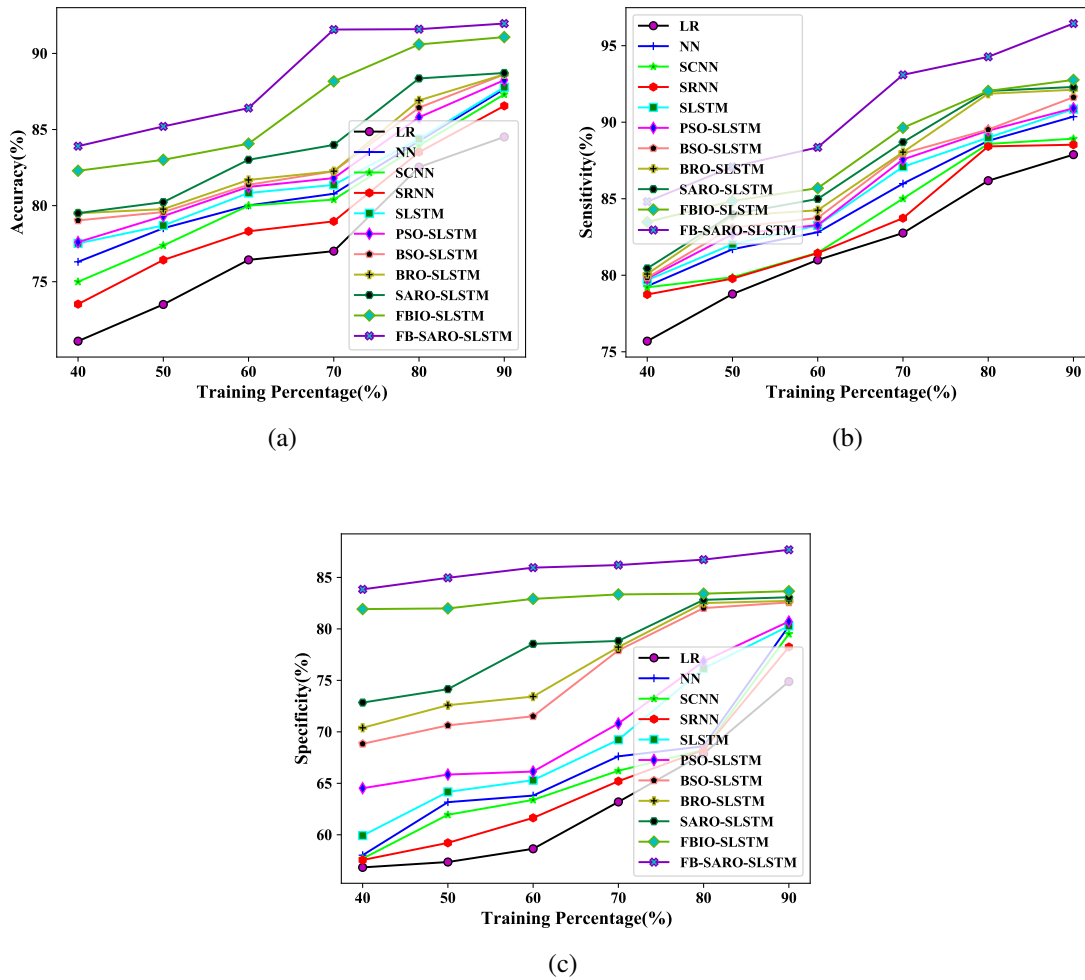


Figure 5.5: Comparison of the (a) Accuracy, (b) Sensitivity, and (c) Specificity of the classifiers' performances on DB2 for various values of the training percentage.

augmentation techniques, as shown in Table 5.1. The ML approach performs almost similarly on all data sets with different augmentation techniques. However, DB1's performance is still slightly better than DB2's due to the comparatively large data set size; other than that, DB4 shows the best results. Also, it indicates that the NN performs better than LR by learning the features. The variation in the performance is approximately 5.0%, 4.0%, 3.0%, and 5.0% for DB1, DB2, DB3, and DB4 with various augmentation techniques and ML approaches. Also, the performance achieved in terms of accuracy is above 83.0%. The features of the EMD techniques are superior to those of the other methods for both classifiers. It achieves the performance of more than 87.0%, 86.0%, 89.0%, and 87.0%, for DB1, DB2, DB3, and DB4, respectively.

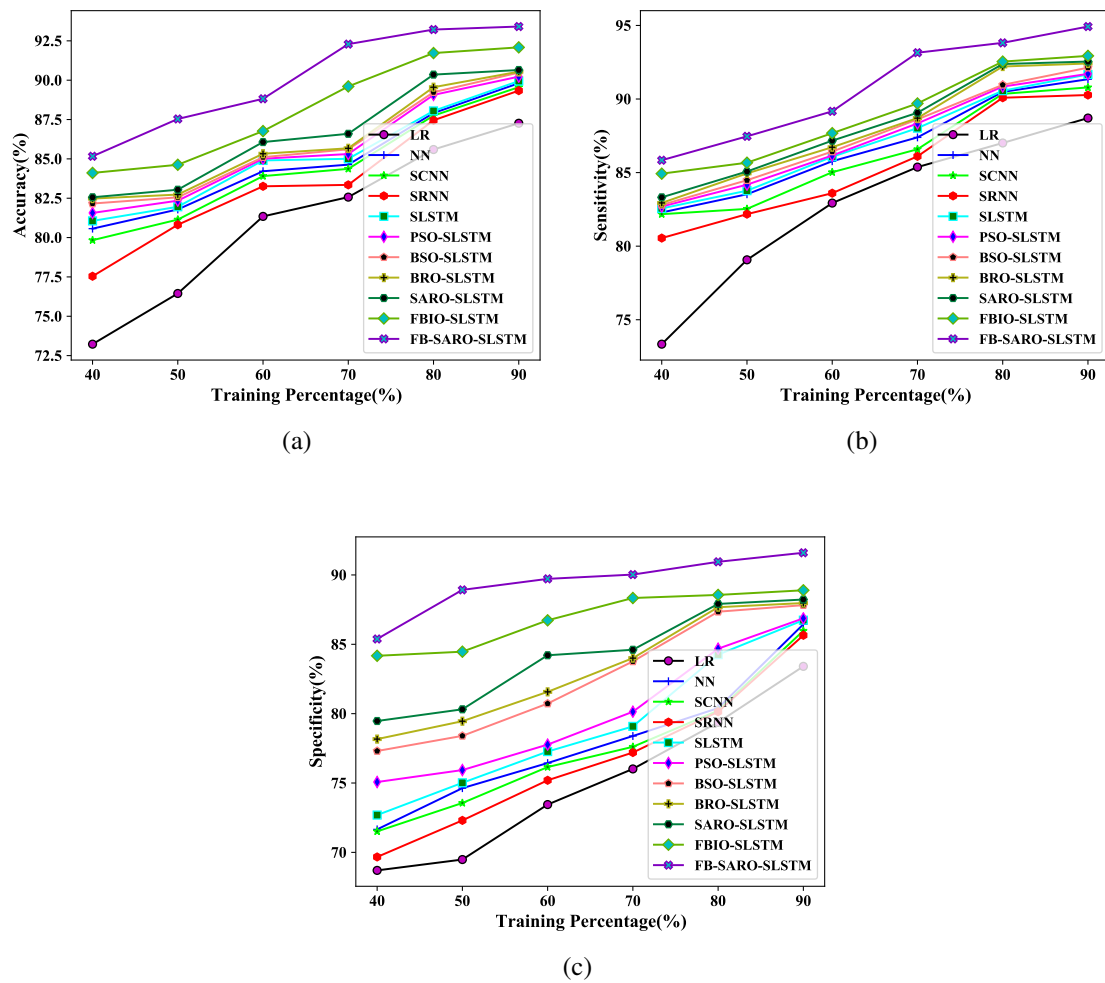


Figure 5.6: Comparison of the (a) Accuracy, (b) Sensitivity, and (c) Specificity of the classifiers' performances on DB3 for various values of the training percentage.

In contrast, the performance of the TQWT and VMD techniques is poor for DB1 and DB2, respectively.

The results achieved in terms of sensitivity show that EMD performed the best for all the data sets. The sensitivity is 85.35%, 90.89%, 89.40%, and 88.46% in the case of DB1, DB2, DB3, and DB4, respectively. These values were attained when NN was operated as a classifier. Except for DB1, all the data sets achieved a sensitivity of more than 80.0%. A sensitivity of more than 90.0% was reported for DB2 with the NN and EMD approach. TQWT performed the worst when used with LR at 78.61% sensitivity for DB1 and 85.76% for DB2 on the VMD method. A sensitivity of 85.89% for DB4 on LR with EWT and TQWT method

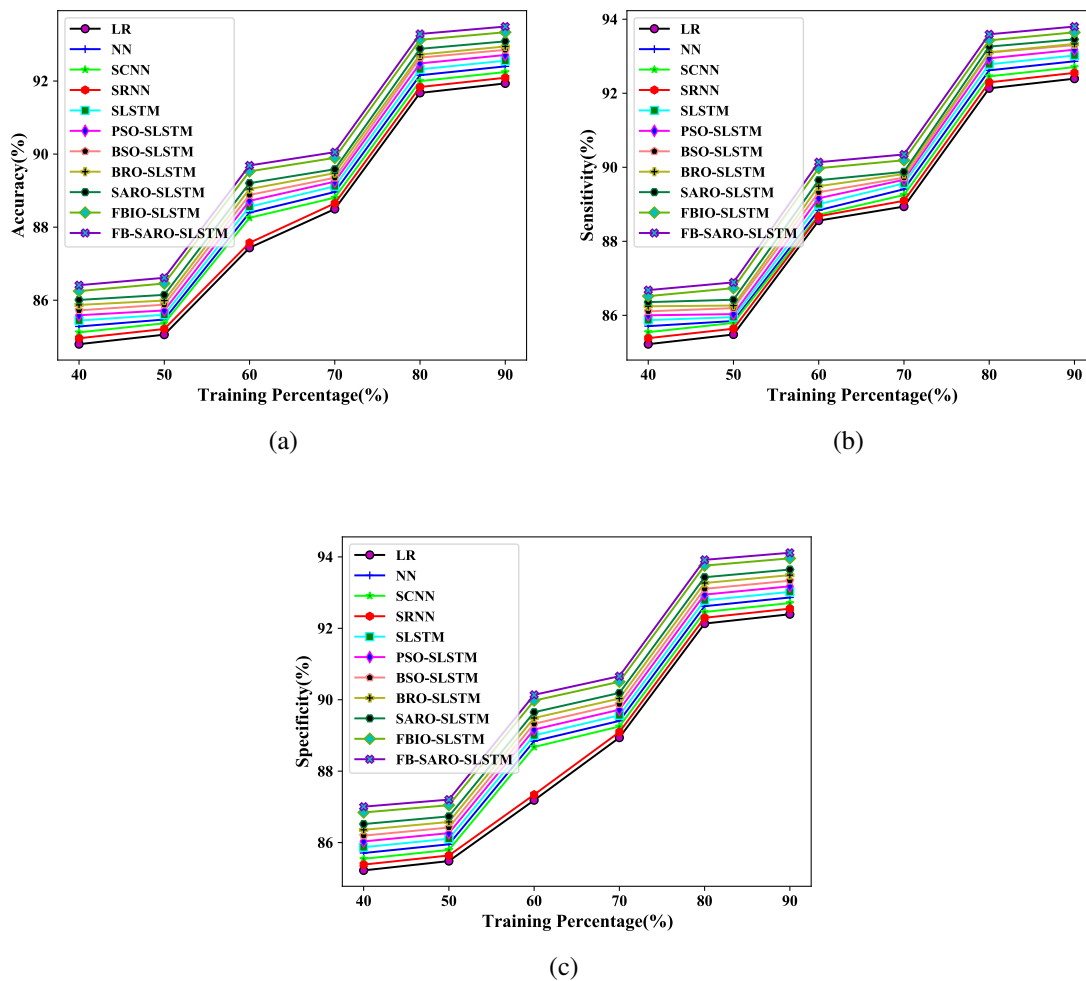


Figure 5.7: Comparison of the (a) Accuracy, (b) Sensitivity, and (c) Specificity of the classifiers' performances on DB4 for various values of the training percentage.

and 85.15% for DB3 on the TQWT method was attained. However, when all four techniques are employed together for augmentation, they give improved results compared to all of the individual implementation techniques. In Table 5.1, it can be seen that there is an increase in accuracy of 1.40% to 4.94% with LR. Also, an increase in accuracy of 1.03% to 3.63% with NN for DB1 was attained. As a result, the minimum reduction in the error rate is 11.14% with LR and 8.61% with NN for DB1. Significant improvement is noticed in specificity and sensitivity as well. A similar trend is observed for DB3 and DB4 for both ML classifiers. However, a similar trend is not noted for DB2 for both ML classifiers.

Table 5.2: Hyperparameter Settings for Deep Learning Techniques.

Classifier	Hyperparameter settings
CNN	Layers : 2
	Filters : 32
	Kernel : 3 X 3
	Pooling : Max pooling
RNN	Layers : 3
	Dropout : 10%
LSTM	Layers : 4
	Batch Size : 32
	Epochs : 100
	Dropout : 20%
	Learning rate : 0.001
	Loss Function : MSE

5.3.1.2 Performance Evaluation of Deep Learning Techniques

The results in the above section gave the motivation to explore DL techniques. DL techniques capture and interpret temporal features more efficiently than ML techniques. The data sets utilized in this research have been employed to create an elaborate feature set with temporal properties. Therefore, it led to the adaptation of CNN, RNN, and LSTM-based techniques for experimentation. The LSTM used to evaluate the proposed approach comprises six layers: one output layer, four hidden layers, and one input layer, with each LSTM layer having 48 neurons. In addition, the hyperparameters are listed in Table 5.2.

The performance is measured by increasing the TrP. Initially, it started at 40.0% and then subsequently increased by 10.0%. The continuous improvement has been noted in the performance of classifiers with an increase in TrP value, as shown in Table 5.3. Table 5.1 and 5.3 indicate an improvement in classifiers' performance when experiments are performed using SRNN, SCNN, and SLSTM. Compared with All-LR, the DL model performance has improved, ranging from 2.06% to 4.27% for all the data sets. As a result, the error rate has been reduced in the span, ranging from 13.97% to 29.46%. Similarly, for ALL-NN, the DL model performance

Table 5.3: Performance evaluation of various Deep Learning methods for the classification task.

TrP	Classifiers								
	SCNN			SRNN			SLSTM		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
DB1									
40	87.01	86.00	87.29	87.12	86.57	87.60	89.00	87.77	89.68
50	87.57	87.12	88.01	87.22	87.20	88.12	89.91	88.73	90.39
60	87.74	87.51	88.84	89.27	88.66	89.61	89.19	88.80	90.47
70	89.28	87.57	89.74	89.82	88.73	90.00	89.46	89.32	90.48
80	89.78	87.96	90.72	90.43	89.25	90.76	89.82	89.74	90.81
90	90.40	88.99	90.89	91.02	90.65	91.05	91.10	90.72	91.35
DB2									
40	73.53	78.74	57.55	74.99	79.21	57.69	77.52	79.68	59.92
50	76.43	79.78	59.21	77.38	79.87	61.94	78.69	82.02	64.17
60	78.31	81.45	61.64	80.00	81.47	63.39	80.83	83.26	65.30
70	78.97	83.73	65.20	80.40	85.00	66.22	81.36	87.09	69.22
80	83.51	88.42	68.24	83.94	88.58	68.24	84.38	88.99	76.15
90	88.55	89.53	78.24	89.31	90.92	79.50	89.78	90.86	80.28
DB3									
40	77.54	80.55	69.67	79.83	82.17	71.51	81.06	82.56	72.69
50	80.81	82.18	72.30	81.14	82.53	73.56	81.96	83.77	75.01
60	83.26	83.60	75.20	83.91	85.03	76.15	84.90	86.07	77.27
70	83.35	86.10	77.20	84.37	86.59	77.61	85.01	88.02	79.08
80	87.45	90.09	80.14	87.75	90.35	80.22	88.05	90.56	84.25
90	89.34	90.27	85.65	89.56	90.79	85.98	89.95	91.67	86.74
DB4									
40	84.96	85.38	85.38	85.12	85.55	85.55	85.44	85.87	85.87
50	85.21	85.64	85.64	85.37	85.80	85.80	85.60	85.95	86.11
60	87.57	88.68	87.34	88.26	88.72	88.68	88.56	89.00	89.00
70	88.65	89.09	89.09	88.81	89.25	89.25	89.12	89.56	89.56
80	91.84	92.30	92.30	92.00	92.46	92.46	92.32	92.78	92.78
90	92.09	92.55	92.55	92.25	92.71	92.71	92.56	93.02	93.02

is enhanced on the scale of 0.14% to 2.12% for all the data sets. Thus, the error rate has been reduced from 0.39% to 17.18%. Besides, positive refinement in the DL model was observed in terms of the sensitivity and specificity values compared to ALL-LR and ALL-NN.

Furthermore, the performance of the classifiers is improved due to the encoding of the data set. Another factor that adds to the performance is the capacity of DL techniques to learn from those attributes. Table 5.3 indicates that the SLSTM performance is better than SCNN and SRNN. Thus, the error rate has decreased from 5.73% to 10.74% for all the data sets compared to SCNN. Likewise, for SRNN, the error decreases on a scale of 0.89% to 4.39% for all the data sets. It is to be noted that this descriptive data set can maintain the temporal nature of the input EEG signals. Thus, the SLSTM can perform better than all these classifiers. These trends can be observed in Table 5.3. Therefore, extensive experimentation was done to build upon this learning ability of LSTM.

Table 5.4: Performance metrics for various optimizations-based Stacked Long Short-Term Memory classifiers used for classification of seizures for all the data sets.

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
DB1			
PSO-SLSTM	92.07	92.52	92.53
BSO-SLSTM	92.93	92.66	92.11
BRO-SLSTM	93.00	93.45	93.48
SARO-SLSTM	94.39	94.84	94.89
FBIO-SLSTM	94.53	95.04	94.96
Proposed FB-SARO-SLSTM	94.69	95.17	95.16
DB2			
PSO-SLSTM	90.25	90.91	80.72
BSO-SLSTM	90.62	91.63	82.57
BRO-SLSTM	90.62	92.12	82.72
SARO-SLSTM	90.71	92.30	83.07
FBIO-SLSTM	91.07	92.76	83.66
Proposed FB-SARO-SLSTM	91.96	96.45	87.69
DB3			
PSO-SLSTM	90.23	91.71	86.87
BSO-SLSTM	90.48	92.13	87.82
BRO-SLSTM	90.54	92.41	87.97
SARO-SLSTM	90.65	92.55	88.23
FBIO-SLSTM	92.09	92.93	88.89
Proposed FB-SARO-SLSTM	93.41	94.93	91.60
DB4			
PSO-SLSTM	92.71	93.18	93.18
BSO-SLSTM	92.85	93.30	93.33
BRO-SLSTM	92.95	93.33	93.49
SARO-SLSTM	93.09	93.46	93.65
FBIO-SLSTM	93.34	93.65	93.96
Proposed FB-SARO-SLSTM	93.49	93.80	94.12

5.3.1.3 Performance Evaluation of Optimization Techniques

Table 5.4 shows the performance of an SLSTM classifier with several optimization algorithms. The performance of the proposed methodology is compared with techniques based on PSO, BSO, BRO, SARO, and FBIO. It can be concluded from Table 5.4 that the optimization algorithms boost the performance of SLSTM and its scale from 1.0% to 4.0% for all the data sets. Also, all the optimization algorithms with classifiers have attained an accuracy of more than

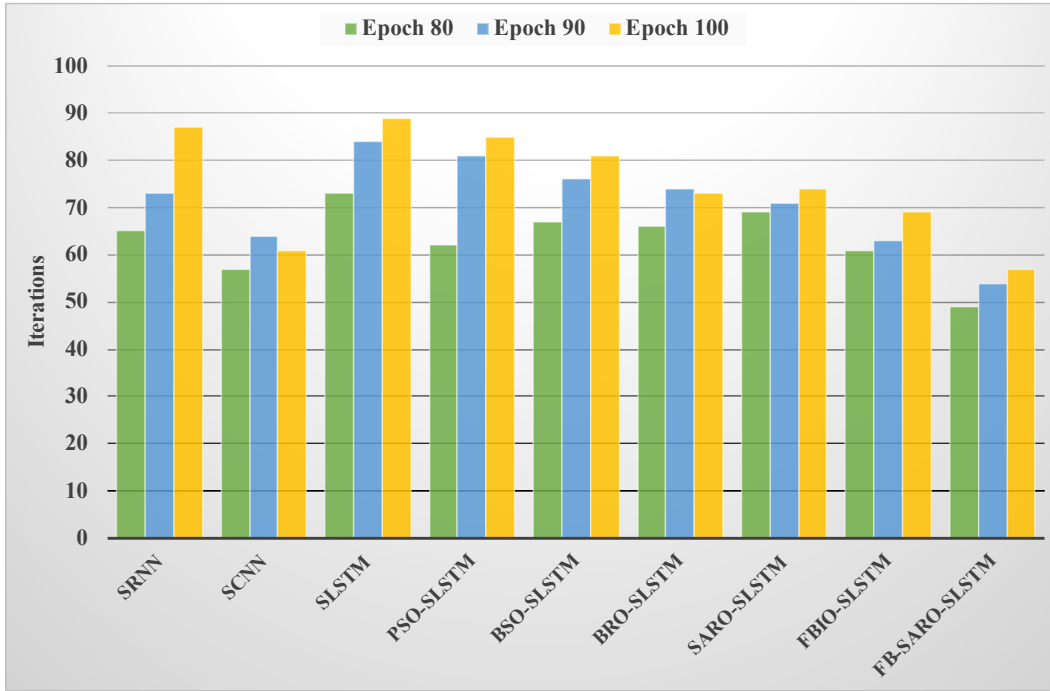


Figure 5.8: The proposed optimization algorithm converges faster than the other DL and optimization methods.

90.0%. Compared to all the optimization algorithms, PSO-SLSTM performance is lower on all the data sets. Besides, the proposed FB-SARO-SLSTM has shown an accuracy of 90.0% or more for all the data sets. The best is 94.69%, which is noted for DB1. Also, it has been observed that FBIO-SLSTM performance is slightly inferior to the proposed approach on DB1 and DB4. It is worth mentioning that the improvement is gradual with each optimization algorithm used for experimentation. Nevertheless, the variation in the performance on individual data sets ranges from 0.8% to 3.1%. However, a significant variation is observed in DB3 and then DB1. The 3.1% and 2.5% variation is recorded in DB3 and DB1, respectively. Also, the proposed method has reduced the error rate by 40.33%, 21.33%, 12.5%, and 34.43%, respectively, for DB1, DB2, DB3, and DB4. The results prove the superiority of the proposed optimization-based SLSTM due to its ability to converge in fewer iterations compared to the

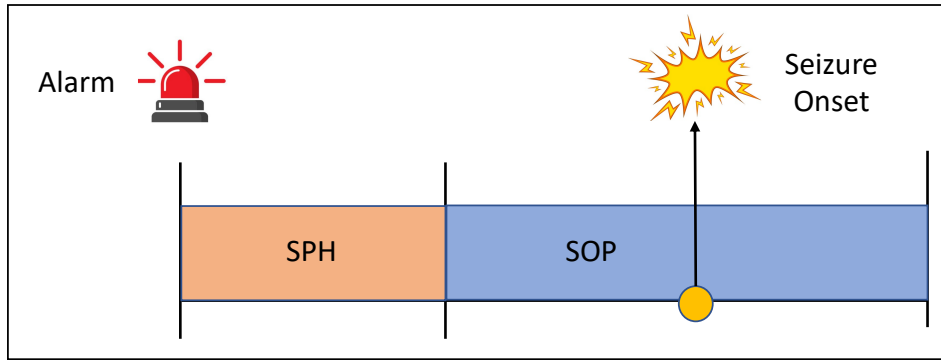


Figure 5.9: Seizure Prediction Horizon (SPH) starts after an alarm is raised to prepare the patient for the upcoming seizure. The Seizure Occurrence Period (SOP) starts just after the SPH. SOP is the time interval in which a seizure onset can occur at any point in time. SPH is the amount of time the patient spends preparing for the seizure.

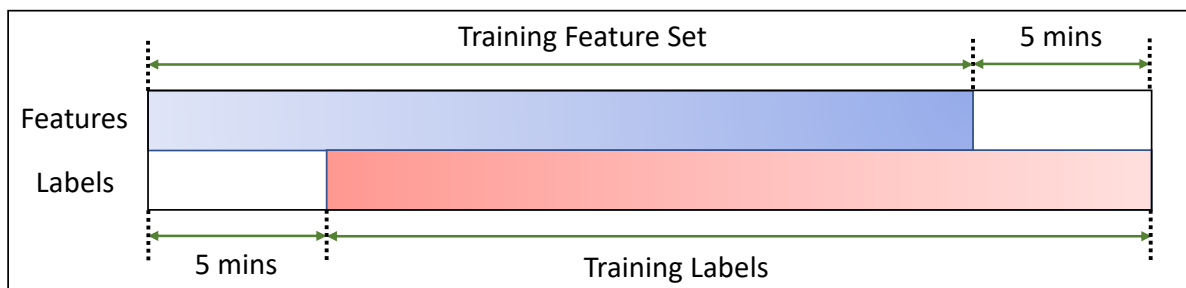


Figure 5.10: Training strategy for predicting Epileptic Seizure occurrences when Seizure Prediction Horizon is 5 mins.

other optimization techniques. The SLSTMs were trained for 100 epochs, and the optimization algorithms were executed for 100 iterations in each epoch. It can be seen from Figure 5.8 that the proposed FB-SARO took fewer iterations to converge to the optimal parameters. Also, it is to be noted that the comparison has been made with SCNN, SRNN, and *Vanilla* SLSTM as well, and these have Adam as the optimization method used. Therefore, according to the findings presented above, it can be inferred that the hybrid feature set and the proposed optimized SLSTM have enhanced the diagnosis procedure. Additionally, it is reported that the sensitivity and specificity values have significantly improved.

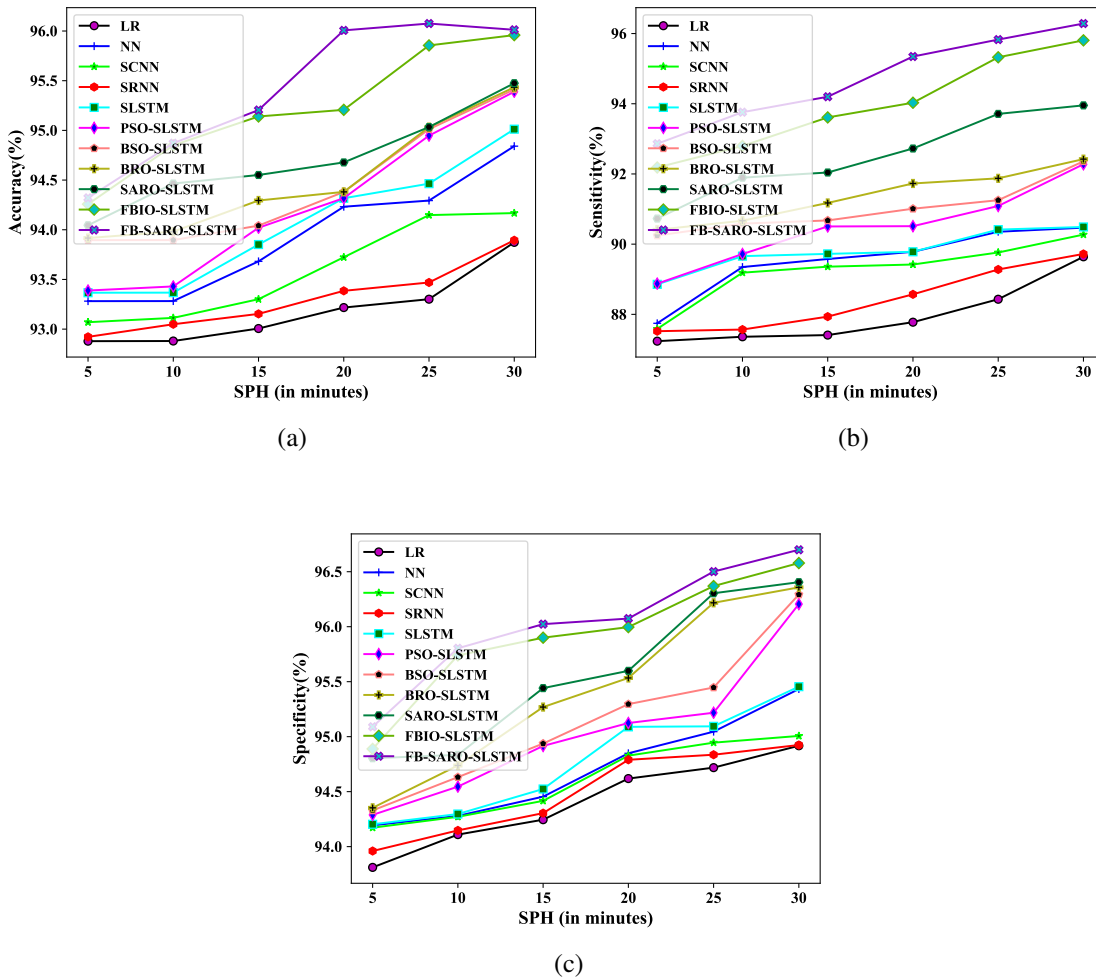


Figure 5.11: Classifiers’ performance on DB1 for varied Seizure Prediction Horizon values (in minutes) in terms of (a) Accuracy (b) Sensitivity (c) Specificity during the prediction of Epileptic Seizure occurrences.

5.3.1.4 Performance Evaluation for Execution of Prediction Task

This improved performance due to the proposed optimization algorithm led the research toward predicting ES occurrences. Two values were utilized to achieve this objective, i.e., SPH and Seizure Occurrence Period (SOP) [202], as shown in Figure 5.9. The optimal value of SPH is determined through the experiments performed on prediction. Every patient in the data set has European Data Format (edf) files and summary files associated with their seizure occurrences. The summary files of the data sets have been used to demarcate the seizure events. Labels such as normal, pre-ictal, seizure, and post-ictal have been assigned to various sections in the data

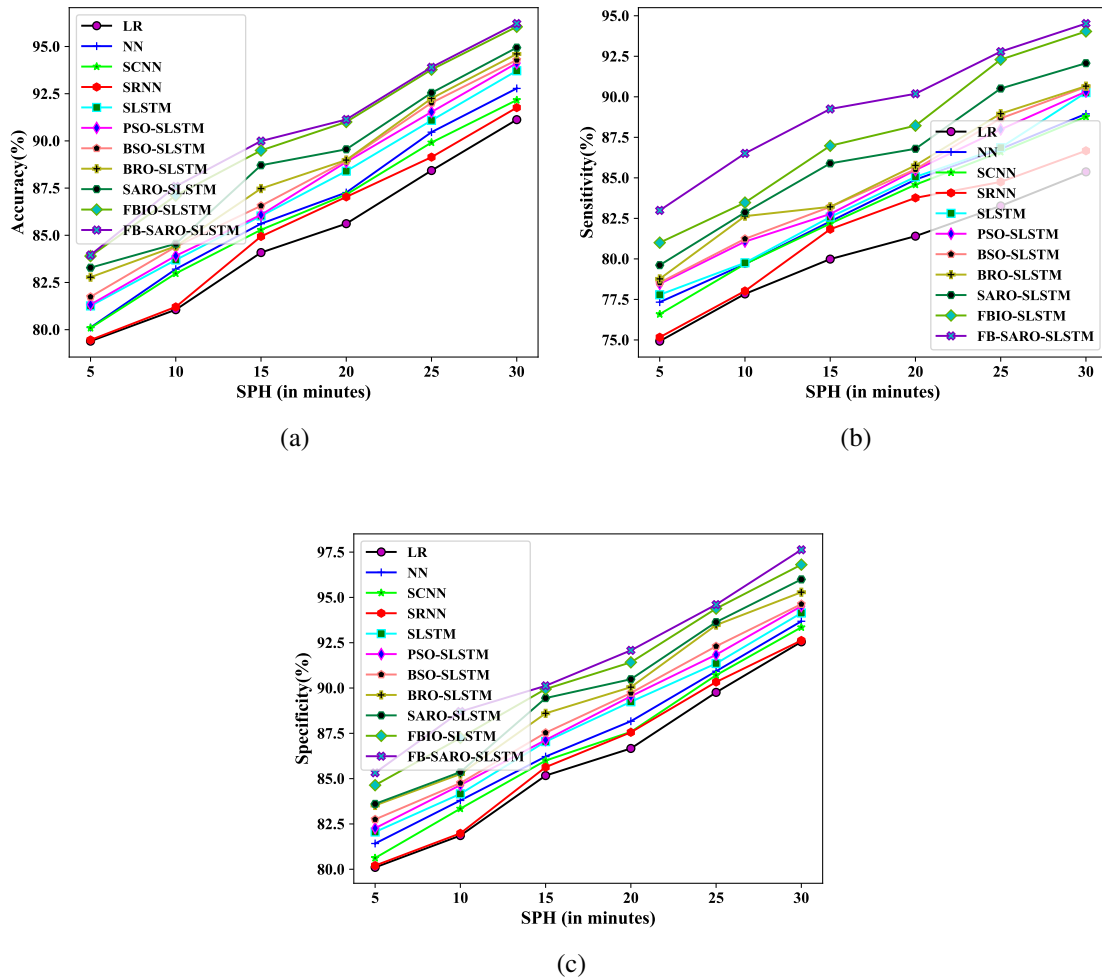


Figure 5.12: Classifiers’ performance on DB2 for varied Seizure Prediction Horizon values (in minutes) in terms of (a) Accuracy (b) Sensitivity (c) Specificity during the prediction of Epileptic Seizure occurrences.

corresponding to a patient. The seizure duration information, i.e., seizure start time and seizure end time, for each patient, corresponding to each EDF file, was provided for DB1 and DB2 only. The train/test split is 80/20 for the features to perform prediction. SPH determines the labels sent for the training, as shown in Figure 5.10. The SPH is increased gradually from 5 minutes to 30 minutes to attain the minimum value, which helps in efficiently identifying and predicting the occurrence of a seizure without compromising accuracy. Once the alert or alarm is raised at the start of the SPH, the patient has time to prepare herself or contact emergency services during the duration of the SPH. When the SPH ends, SOP starts. This means that the

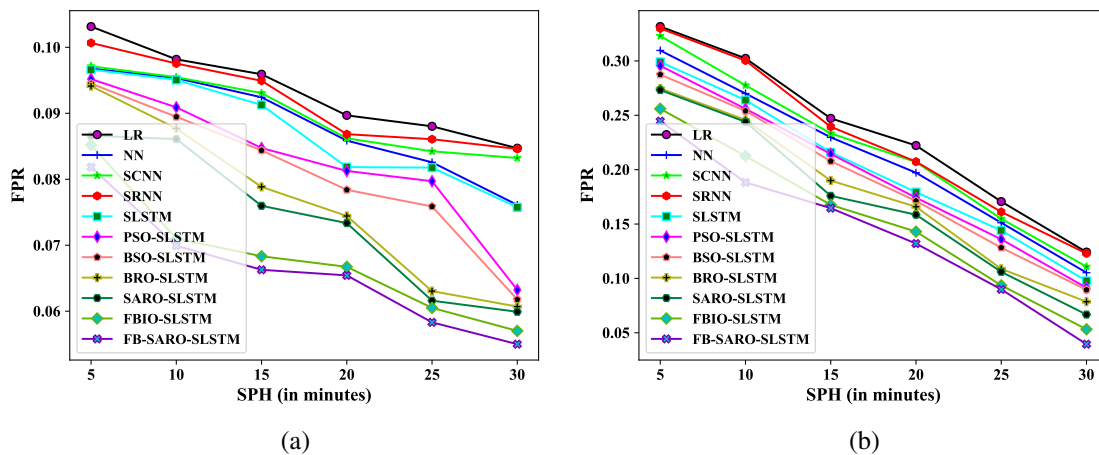


Figure 5.13: (a) Evaluation of False Prediction Rate for varied values of Seizure Prediction Horizon, when prediction is performed on DB1. (b) Evaluation of False Prediction Rate for varied values of Seizure Prediction Horizon, when prediction is performed on DB2.

seizure can occur anytime during those 30 minutes. This helps the patient and/or their attendant brace themselves for its impact.

The increase in the SPH provides better results with different optimization algorithms and SLSTM for DB1 and DB2. The effect is noted because more information is available for the learning of the classifier with an increase in SPH. The DB1 data set attains an accuracy of more than 93.0% for SLSTM optimized through various algorithms. Also, the variation is below 2.0%

5.3.1.5 Early Prediction Versus Accuracy Trade-Off

In this research work, an early prediction of ES occurrences is performed using the optimal value of SPH identified through experimentation on EEG signals of epileptic patients. While training the classifier, the SPH value played an important role. For example, if the SPH was taken as X minutes, then during training, the corresponding samples were sent in advance to the classifier. This helped in achieving the predictive property of the proposed model. However, early prediction is mostly achieved by compromising accuracy as well. Experimentation

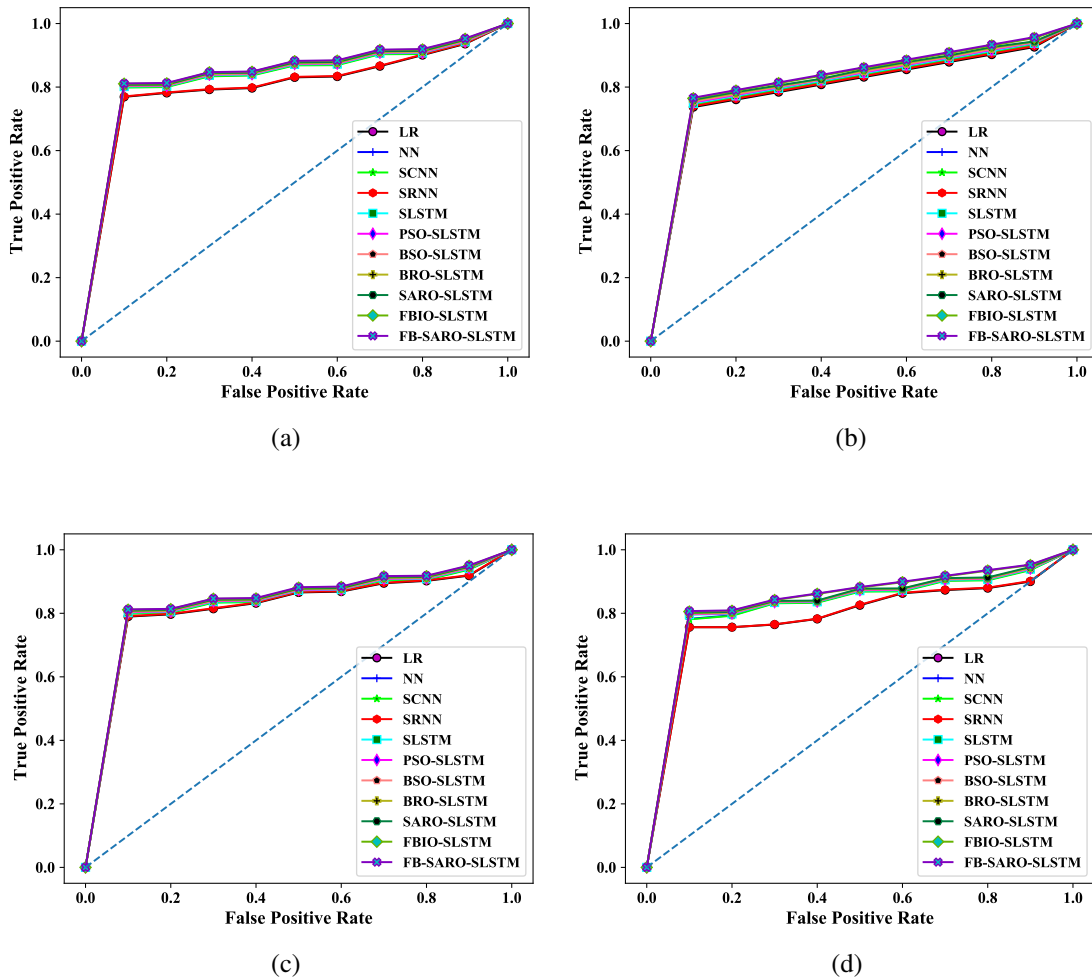


Figure 5.14: Receiver operating characteristic curves obtained for (a) DB1, (b) DB2, (c)DB3 and (d) DB4 showcasing the performance of proposed optimization-based SLSTM technique.

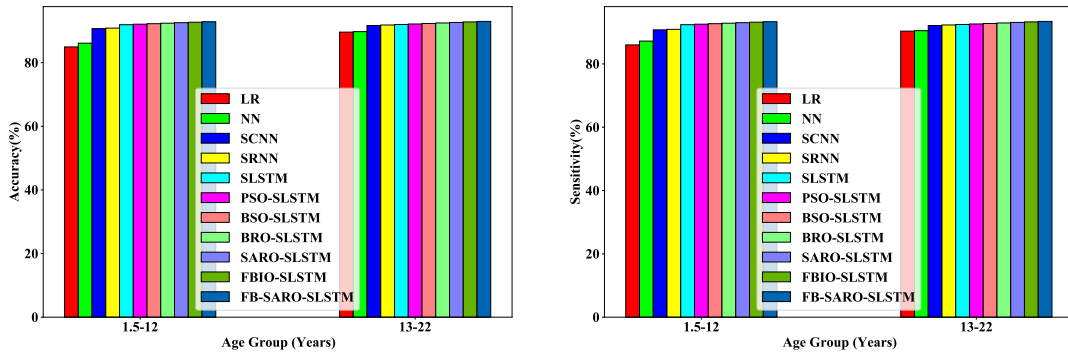
showed that an SPH of 30 minutes gave the maximum accuracy and the minimum FPR. To achieve this, one might be tempted to increase the SPH value, but this may lead to the failure of the system as it may lead to the overlapping of pre-ictal and ictal stages. This will nullify the purpose of the proposed model, which is to warn the epilepsy patient or their attendant about an upcoming seizure and give them ample time to prepare for it. The optimal value achieved through the proposed methodology ensures the system favors the patients without compromising much on accuracy. Thus, the proposed novel optimized model can predict an impending seizure with efficient accuracy and raises the alarm with minimum false alerts to warn the patient in time for them to get ready for it in terms of medical attention.

Table 5.5: Leave-one-out-database performance.

Models	Training (DB2, DB3, DB4); Testing (DB1)	Training (DB1, DB3, DB4); Testing (DB2)	Training (DB1, DB2, DB4); Testing (DB3)	Training (DB1, DB2, DB3); Testing (DB4)
	Performance (%)			
LR	91.24	90.66	90.13	92.99
NN	91.41	90.98	90.24	93.17
SCNN	93.75	91.31	93.36	93.78
SRNN	93.93	91.63	93.52	93.95
SLSTM	94.10	91.95	93.84	94.13
PSO-SLSTM	94.27	92.26	94.01	94.29
BSO-SLSTM	94.44	92.50	94.16	94.46
BRO-SLSTM	94.61	92.79	94.32	94.60
SARO-SLSTM	94.77	93.03	94.47	94.75
FBIO-SLSTM	95.12	93.68	95.60	95.03
Proposed				
FB-SARO-SLSTM	95.2738	93.9557	95.7612	95.1860

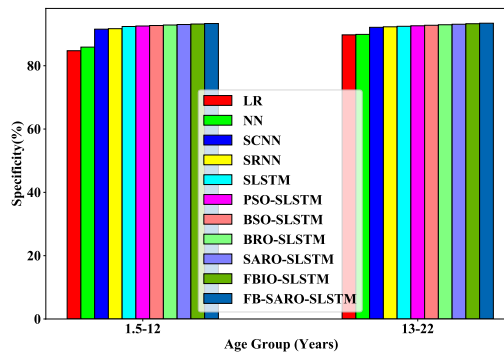
5.3.1.6 Leave-one-out-Database Validation

Further experimentation was performed to assess the performance of the proposed optimization-based SLSTM using Leave-One-Out-Database. Each database was taken as the test set, while the remaining three data sets were taken for the training set. This experimentation was done to judge the generalizability of the proposed method for a new patient. It can be observed in Table 5.5 that performance of more than 90.0% has been achieved with ML, DL, and optimized DL approach. The variation in the performance of the ML approach is almost 3.0%. The ML techniques performed best when DB3 was taken as the test set, giving an accuracy of 93.17%, and showed an inferior performance of 90.13% when DB4 was taken as the test set. A similar trend is reported for the DL method. The best performance of 94.12% is reported for the DL method. Also, SLSTM is a better-performing classifier with all combinations than other DL methods. The trend of continuously improving the performance from ML to optimized DL via the DL approach has been preserved, as shown in Tables 5.1 - 5.5. A notable reduction in the error rate has been observed, and it is at least 3.65% when moving from ML to the DL



(a)

(b)



(c)

Figure 5.15: Comparison of various techniques for prediction of seizure occurrence in terms of (a) Accuracy, (b) Sensitivity, and (c) Specificity showcasing the performance of the proposed optimization-based SLSTM for the two age groups.

approach. Similarly, at least a 2.6% reduction in error rate was achieved when moving from DL to an optimized DL approach. The proposed optimization-based SLSTM classifier performed the best in all the techniques compared. It achieves a performance of more than 95.0% except when DB2 was taken as the test set.

Moreover, it is worth mentioning that this work has a diversity in data collection. Thus, developing a generalized intelligent approach for seizure prediction is challenging. However, the hybrid feature set has provided a common platform to capture diversified information from various data sets. Besides, the leave-one-out data set concept has to assist in enriching the hybrid feature set for the learning of the classifier. Consequently, potential enhancement has

occurred in various seizure patterns and non-seizure pattern representations in feature space. Thus, it has supported capturing the maximum uncertainty in the data set and simultaneously improves the performance of seizure diagnosis, as shown in Table 5.5.

5.3.1.7 Evaluation Based on Age Groups

In the previous studies that have experimented on the DB1 database, it has been observed that the DL methods have been able to model epileptic patients' EEG data and given substantial results [203–206]. However, in this work, going a few steps ahead, experimentations were performed separately on the DB1 data set as the patients in this database are of varied ages, ranging from an infant of 1.5 years to an adult of 22 years. The patients were segregated into two age groups to assess the generalizability of the proposed approach; one group had infants and pre-teens, here referred to as Group-I, while the other had teenagers and adults, here referred to as Group-II.

In Figure 5.15, it can be seen that the proposed optimization-based LSTM achieved an accuracy of 93.0% in Group-I. An accuracy of 93.10% was achieved in Group II, which was the highest amongst all the other ML and DL techniques compared. It is observed that the ML approach achieves a performance of below 90.0% on both the group compared to DL and optimized DL. Besides, similar to previous results, the rise in the accuracy trend is observed from ML to optimized DL via DL.

5.3.1.8 Comparison with State-of-the-art Methods

Table 5.6 compares the proposed work to the state-of-the-art techniques described in the literature. It was observed that DB1 is the most often used data set for creating different seizure prediction methods. The proposed work beats most state-of-the-art techniques but is less superior to a few techniques. The main reason for this is the PS approach of these methods, i.e.,

Table 5.6: Comparison with state-of-the-art methods used to perform prediction of seizure occurrences.

Study	Data set	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	FPR
[22]	DB1	GAN	91.24	89.28	87.0	0.13
[69]	DB1	CNN	99.0	-	-	0.0
[133]	DB1	CNN	-	97.83	92.36	0.0764
[82]	DB1	ResNet	89.78	88.63	-	0.122

they analyze the EEG signals using classifiers tailored to the characteristics of the individual patient’s signals as they vary from patient to patient. The proposed methodology has been built keeping in mind the widespread nature of epilepsy as a disease and benefiting as many patients as possible. For this, a patient-generic approach has been adapted. Another reason is the underlying architecture of the proposed method, i.e., LSTM. Some recent methodologies utilize CNN to predict an ES by converting time-series EEG data into 2D images and spectrograms. However, these methods lack the exploration of the temporal nature of the data set, which can be done using LSTM. Lastly, the proposed approach utilizes multichannel data, which helps examine features to predict upcoming seizure occurrences accurately.

5.4 Summary

This study aims to provide a patient-generic approach to seizure occurrence prediction using EEG data. This is done by proposing a unique hybrid optimization algorithm-based DL technique. As a result, feature augmentation is used to create a hybrid deep feature set to address the issue of insufficient patient data. The SLSTM is then used to analyze the feature space. Additionally, by optimizing the parameters in accordance with the fundamentals of forensic investigations involving police personnel and the peculiarities of human search and rescue operations, the performance of the SLSTM architecture is enhanced. Given a seizure prediction horizon, the optimized SLSTM is utilized to forecast seizure episodes. The hybrid optimization

algorithm FB-SARO converges faster and selects the best LSTM parameters compared to traditional optimization techniques. These features improve the classifier's performance, and four publicly accessible data sets have shown that it performs better than state-of-the-art techniques. Seizure data from the data sets were used to achieve the prediction. The most effective SPH value to alert patients and their carers to an impending seizure has been determined. Finding the appropriate SPH enables people to seek medical attention and plan for it. The proposed approach can predict a seizure event early without compromising accuracy since seizure episodes are one of the time-critical concerns in healthcare.