CORONARY HEART DISEASE CLASSIFICATION USING IMPROVED PENGUIN EMPEROR OPTIMIZATION-BASED LONG SHORT TERM MEMORY NETWORK

RAJESHWARI MARAMGERE RAMAIAH[1*] AND KAVITHA KUNTAEGOWDANALLI SRIKANTEGOWDA[2]

[1]*Department of Computer Science & Engineering, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India*
[2]*Department of Computer Science and Engineering, Dayananda Sagar college of Engineering, Bengaluru, Karnataka 560078, India*

[*]Corresponding author: rajeshwari_maya1@yahoo.com

ABSTRACT: Ventricular fibrillation (VF) is the most life-threatening and dangerous type of Cardiac Arrhythmia (CA), with a mortality rate of 10-15% in a year. Therefore, early detection of cardiac arrhythmia is important to reduce the mortality rate. Many machine learning algorithms have been proposed and have proven their usefulness in the classification and detection of heart problems. In this research manuscript, a novel Long Short Term Memory (LSTM) classifier with Improved Penguin Optimization (IPEO) is implemented for VF classification. The IPEO is used in finding optimal hyperparameters that overcome the overfitting problem. The presented model is tested, trained, and validated using two standard datasets that are available publicly: Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) and the China Physiological Signal Challenge (CPSC) 2018 dataset. Both of them consist of ECG recordings for five seconds of coronary heart disease (CHD) patients. Furthermore, Fuzzy C-Means and Enhanced Fuzzy Rough Set method (FCM-ETIFRST) are used for feature selection to extract informative features and to cluster membership degree, non-membership degree, and hesitancy degree. On the MIT-BIH dataset, the proposed model achieved accuracy, sensitivity, specificity, precision, and Matthews's correlation coefficient (MCC) of 99.75%, 98.29%, 98.39%, 98.35%, and 97.79% respectively. On the CPSC 2018 dataset, the proposed model achieved accuracy of 99.79%, sensitivity of 99.11%, specificity of 98.20%, precision of 99.43%, and MCC of 98.57%. Hence, the results proved that the proposed method provides better results in the classification of VF.

ABSTRAK: Pemfibrilan Ventrikel (VF) adalah ancaman nyawa nombor satu dan jenis Aritmia Jantung (CA) berbahaya dengan kadar kematian 10-15% setahun. Oleh itu, pengesanan awal Aritmia Jantung sangat penting bagi mengurangkan kadar kematian. Terdapat banyak algoritma pembelajaran mesin yang telah dicadangkan dan terbukti berkesan dalam pengelasan dan pengesanan sakit jantung. Kajian ini mencadangkan kaedah baru pengelasan Memori Ingatan Jangka Panjang Pendek (LSTM) dengan Pengoptimuman Penambahbaikan Penguin (IPEO) yang dilaksanakan bagi klasifikasi VF. IPEO digunakan bagi mencari hiperparameter yang dapat mengatasi masalah padanan berlebihan. Model yang dicadangkan diuji, dilatih dan disahkan menggunakan dua dataset piawai yang dapat diperoleh secara terbuka: Institut Teknologi Hospital Massachusetts-Beth Israel (MIT-BIH) dan Cabaran Signal Psikologi Cina 2018 (CPSC). Kedua-dua data ini mempunyai rakaman ECG selama lima saat daripada pesakit Penyakit Jantung Koronari (CHD). Malah, kajian itu turut menggunakan Purata-C Kabur dan Kaedah Set

Kasar Kabur Dipertingkat (FCM-ETIFRST) untuk pemilihan bagi mengekstrak ciri-ciri dan mengelaskan kelompok tahap keahlian, bukan ahli dan tahap keraguan. Bagi dataset MIT-BIH, model yang dicadangkan mencapai ketepatan, tahap sensitif, tahap spesifik, kejituan dan pekali kaitan Matthews (MCC) sebanyak 99.75%, 98.29%, 98.39%, 98.35%, dan 97.79% masing-masing. Bagi dataset CPSC 2018 pula, model yang dicadangkan mencapai ketepatan sebanyak 99.79%, 99.11% tahap sensitif , 98.20% tahap spesifik, 99.43% kejituan dan 98.57% MCC. Oleh itu, dapatan kajian membuktikan kaedah yang dicadangkan menunjukkan keputusan lebih baik dalam pengelasan VF.

1. INTRODUCTION

As per the World Health Organization (WHO), coronary heart disease (CHD) is a substantial global epidemic and the most common cause of death worldwide, accounting for one in every three deaths. Coronary heart disease (CHD) is a substantial global epidemic and the most common cause of death worldwide, accounting for one in every three deaths. In 2002, 16.7 million cardiac disease (CD)-related deaths were reported, and this is expected to rise to 23.3 million by 2030 [1]. Age, obesity, smoking, exercise, hypertension, diabetes, and high blood cholesterol are all risk factors for CD. Heart disease diagnosis is conducted using a variety of medical tests, a medical history, and an examination of the patient's lifestyle. There are many variables to consider when diagnosing heart diseases, thus a specialist is usually involved [2-3]. The contraction of the ventricles that exceeds 100 bpm causes ventricular tachycardia (VT), and when the ventricles' contraction exceeds 500 bpm, it is referred to as ventricular fibrillation (VF), which is the most abnormal cardiac arrest. VT and VF are the two subsets of ventricular arrhythmias (VAs) [4]. Nearly all countries face a significant coronary artery disease (CAD) burden, due to limited resource availability for providing comprehensive health care of CAD patients, and insufficient awareness campaigns on such diseases [5]. Various machine learning methods are developed to detect CHD at an early stage using different software libraries and platforms to extract huge amounts of information from large datasets. The predictive analysis of CHD includes pre-processing the data collected from datasets, preparing the data for implementing suitable algorithms, training, testing, and validating the chosen model, and making final predictions with fine-tuned parameters [6,7]. Atrial fibrillation (AF) detection has been done in many types of research depending on the activities of the atrial region of the heart, which also includes the detection of P-wave, the entropy of wavelet samples, and the detection of F-waves [8]. To detect P-wave or R-R wave irregularities with manual instructions on long-term ECG recordings, AF uses currently available algorithms. Any change to P waveform in shape or blunt from Lead 2 and 4 results in irregularity, and it directly relates to the condition of atrial activity since it is known as a depolarization wave of the atrial region [9-10]. In this work, a design of methodology to detect and classify CHD is developed by using ECG signals of MIH-BIT and CPSC 2018 dataset. The major contributions of the paper are described as follows:

- To resolve the problem of class imbalance and dimensionality curse, the method of FCM-ETIFRST has been included in the process of feature selection.

- For accurate classification of MIT-BIH arrhythmia and CPSC 2018 datasets with reduced overfitting issues, an efficient LSTM classification has been considered.

- An Improved Penguin Emperor Optimization (IPEO) has been proposed in this research for finding the optimal hyperparameters to operate LSTM without overfitting issues. This is done by improving the exploitation and exploration stage of PEO by levy flight and Gaussian mutation mechanism.

2. LITERATURE REVIEW

Li et al. [11] developed a novel feature fusion framework that used phonocardiogram (PCG) signals of 175 subjects to identify CAD features. A sum of 110 features from different domains were extracted, reduced, and selected. The obtained images were taken as CNN input for feature learning. The feature selection and the features of deep learning were then combined and led to a multilayer perceptron for classification purposes. This proposed framework outperformed deep learning features and multi-domain features with the highest accuracy, sensitivity, and specificity. Too many features inhibit the model learning process and result in poor generalization, which is a limitation of this work. Future work can be focused on improving the detection accuracy of CAD by using dynamic features and multi-channel PCG signals. Nguyen et al. [12] developed a stacking method to predict atrial fibrillation (AF) from ECG signals and perform statistical segment-based feature recognition, which was done by Support Vector Machine (SVM), the segmentation units produced by a convolutional neural network. To validate this method, the ECG dataset from the PhysioNet/Computing in Cardiology Challenge 2017 was used, which contained 8528 ECG recordings. The proposed method applied to the same dataset and metric outperformed state-of-the-art methods with a high F1 score. The limitation of this method was the nature of the dataset that was chosen. The AF signal segment was not included with AF due to the absence of information about AF or the occurrence of other ECG signals. The proposed method was applied to other medical signal-related problems. Jahmunah et al. [13] proposed an automated system (AS) using automated categorization of ECG signals with CNN, which classified signals into normal, CAD, myocardial infarction (MI), and congestive heart failure (CHF) classes. The proposed GaborCNN achieved better performance with 98.5% accuracy. The use of GaborCNN resulted in low computational complexity which was an added advantage for this work. Furthermore, the system was validated with a large database and was seen to exhibit the highest potential to assist clinicians in screening for CVDs using ECG signals. Only a few subjects and smaller datasets were used for CAD and CHF, which was the limitation of this work. To improve the classification accuracy of GaborCNN, more data needs to be trained to the network in the future, so that the onset of CAD is detected early and prevented from progressing to MI or CHF.

Tseng et al. [14] developed a method for VF prediction which was a novel deep-learning method. For training and validation of the data, ECGs from MIT-BIH datasets were used. The estimated results demonstrated that the proposed two-dimensional short-time Fourier transform (2D STFT)/continuous wavelet transform (CWT) convolutional neural network (CNN) model achieved the highest recall and accuracy. The proposed model was also compared with 1D CNN and 2D time-domain models and again achieved the highest accuracy. The database did not identify concomitant diseases, and the influence of these diseases on ECG signals was unknown, which was one of the limitations of this work. In the future, larger datasets should be used to build a more accurate model. Additional data from conventional and portable hospital vital sign monitors, as well as wearable devices should be collected. Panigrahy et al. [15] proposed a novel approach to detect the VF rhythm that involved algorithms namely: Support Vector Machine (SVM), Adaptive Boosting (AdaBoost) and Differential Evolution (DE). These algorithms were implemented with the help of optimal variable combinations. The training and test datasets were taken and

validated from three databases namely the arrhythmia database, CUDB database, and MIT-BIH malignant ventricular arrhythmia databases for this methodology. The proposed approach achieved the highest accuracy, sensitivity, and specificity in detecting VF rhythm. Further, the proposed method was suitable for implementation during real time detection of VF rhythm. Çınar et al. [16] developed a deep learning architecture with SVM for the classification of Normal Sinus Rhythm (NSR) which was based on hybrid Alex-net along with ECG signals of Abnormal Arrhythmia (ARR) and Congestive Heart Failure (CHF). This approach was implemented on 192 ECG signals, out of which 96 were arrhythmias, 30 were CHFs, and 36 were NSRs. SVM and KNN algorithms were implemented on ARR, CHR, and NSR signals for classification purposes. Later on, these signals were classified using Long Short-Time Memory (LSTM). Chen et al. [17] proposed a CNN model for CA classification from ECG signals of the China Physiological Signal Challenge (CPSC) 2018 dataset. The dataset provided 6,877 recordings with a 12-lead ECG dataset, in which different types of diagnoses for 476 patients were predicted. The proposed technique obtained an overall highest F1-score, and first rank in the classification-and-challenge competition. This approach still needed to be filtered to achieve an equal level of performance on other datasets, which is a limitation of this work. Various CA types and wearable ECG devices are possible to be developed for use with this approach in the future.

3. METHODOLOGY

The framework for the proposed work is a combination of feature selection with FCM-ETIRST and LSTM classification with Improved Emperor Penguin Optimization (IEPO) hyperparameter optimization for reducing the problems of dimensionality curse and overfitting in CHD analysis and it is illustrated in Fig. 1. Because of the nature of time-step order in CNN with its several hidden layers, it is used to extract the features from the complex data of ECG signals. The characteristics of time will be taken by the layers of LSTM while classification is skipped at the stage of feature extraction. To optimize the computational resources and make more efficient detection of CHD, the proposed IPEO algorithm is utilized to optimize hyperparameters of LSTM. The steps of the proposed methodology workflow are designed as shown in Fig. 1.
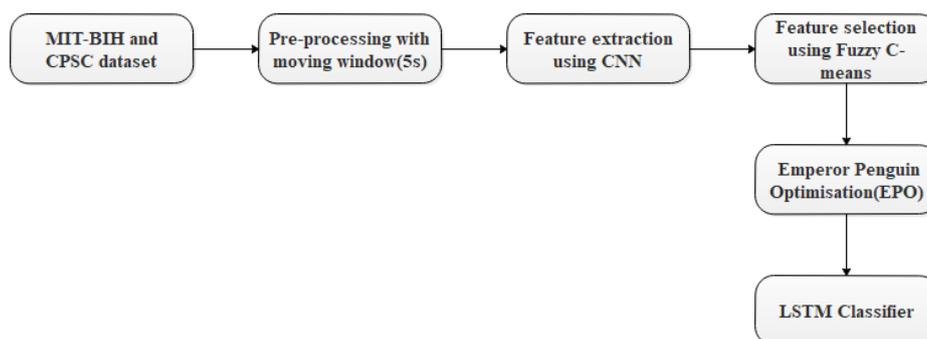
Fig. 1: Flowchart of the proposed IEPO based LSTM model.

3.1 Data Pre-processing

ECG recordings of VF patients are considered from two datasets namely MIT-BIH and CPSC 2018. The MIT-BIH dataset consists of 23 publicly available ECG recordings which include four types of rhythms: AF, VF, AV junction, and Normal. However, in this work, only VF signals are classified and sampled at a frequency of 250Hz. The proposed algorithm is implemented for detecting VF by pre-processing ECG signals in a window length of 5

seconds. The CPSC 2018 also consists of ECG signals in which all forms of arrhythmias are available. All signals are segmented into a number of episodes, where each episode is labeled by the clinician and divided into training, test, and validation datasets. Two leads are selected for this work from the dataset, and the data is divided into VF and non-VF categories. The ECG signals from the CPSC dataset are sampled at a frequency of 500Hz and segmented into 5-second episodes. A lead 1 sample signal from the CPSC dataset is considered, as shown in Fig. 2, and is segmented into normal signal and VF signal as shown in Fig. 3(a) and Fig. 3(b). Lead 2 normal signal and VF signal are shown in Fig. 4(a) and Fig. 4(b) respectively.
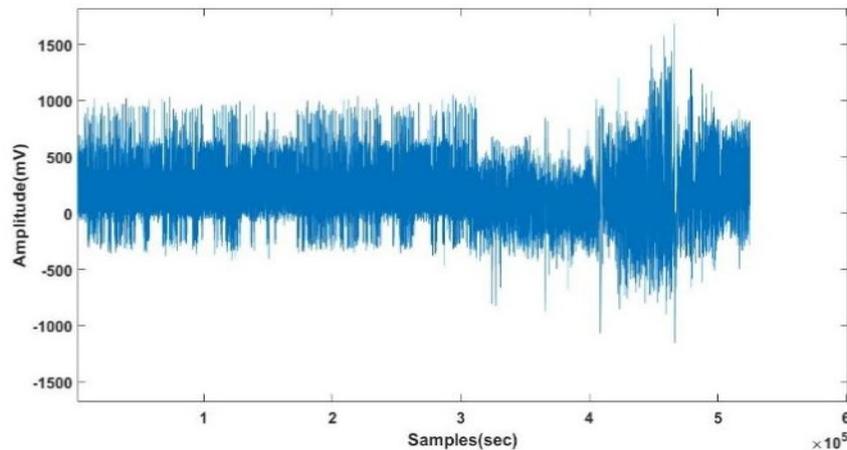
Fig. 2: Sample signal from MIT-BIH dataset.
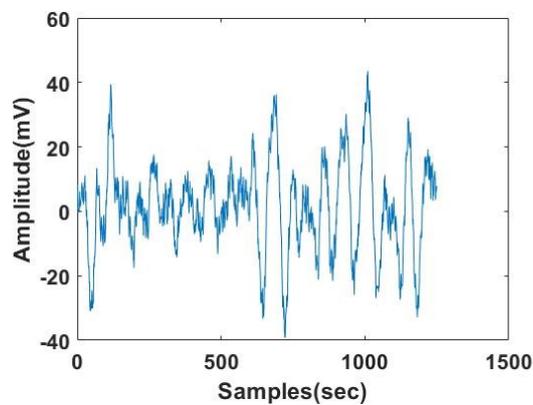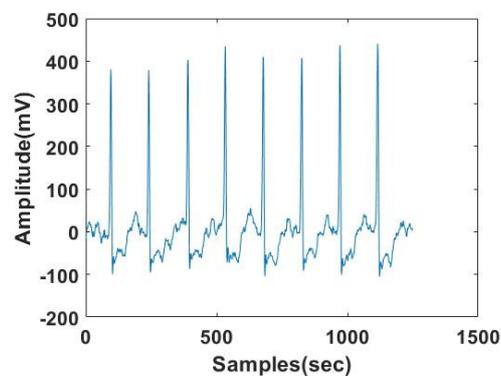
(a)   (b)

Fig. 3. Lead 1 sample signal (a) Normal, (b) VF.

<div align="center">(a)   (b)</div>
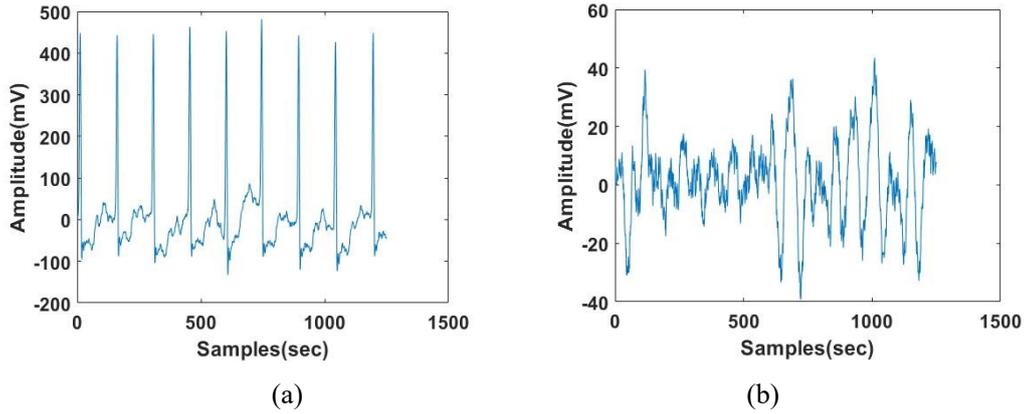
<div align="center">Fig. 4: Lead 2 sample signal (a) Normal, (b) VF.</div>

3.2 Feature Extraction

Here, the extraction of features is done using CNN layer architecture which is based on the 1-D framework. Timing characteristics and sequence of data can be retained with the representation of CNN-related features in the proposed framework. The following features are to be extracted from the proposed CHD framework.

3.2.1 1-D Approximation and Details Coefficients

By use of low pass and high pass filtering operations on the transform coefficients group, a set of coefficients namely, 'approximation from low pass' and 'detail from high pass' is obtained. Consider the parameters of coefficients for approximation and details to be CA_j and CD_j correspondingly, in case of a stationary signal, the transform their properties for the level of j as mentioned in Eq. (1) and Eq. (2).

$$E(CA_j) = \frac{\sqrt{2^{-j}}}{N} \int f(t) \tag{1}$$

$$E(CD_j) = \sqrt{2^{-j}} \int f(t) E_n[\varphi(2^{-j}t - n)]dt \tag{2}$$

Where CA_j and CD_j are parameters of coefficients at the level of j, E_n represents the wavelet transform of the signals at n-level coefficients, $\varphi(t)$ and $f(t)$ represents the wavelet coefficient transformation functions. N represents the total number of wavelets obtained from the signal.

3.2.2 Hjorth Activity

Signal power is represented using an activity parameter that is time function variance, as given in Eq. (3). This activity denotes the power spectrum surface in the frequency domain.

$$Activity = var(y(t)) \tag{3}$$

where the signal is denoted using $y(t)$.

3.2.3 Hjorth Mobility

The mean frequency or power spectrum of standard deviation proportion is represented as a mobility parameter. A first derivative variance of the square root of signal $y(t)$ divided by signal variance $y(t)$, which is explained in Eq. (4).

$$Mobility = \sqrt{\frac{var\left(\frac{dy(t)}{dt}\right)}{var(y(t))}} \qquad (4)$$

3.2.4 Hjorth Complexity

The frequency change is denoted using the complexity parameter. The signal's similarity and a pure sine wave are compared using the parameter, where the value converges to 1 if the signal is similar, as given in Eq. (5).

$$Complexity = \frac{Mobility\left(\frac{dy(t)}{dt}\right)}{Mobility(y(t))} \qquad (5)$$

3.2.5 Mean Curve Length (MCL)

Katz fractal dimensions are measured using MCL and this measures EEG signals activity, as given in Eq. (6).

$$CL[n] = \frac{1}{N}\sum_{m=k-N+1}^{k}|x[m]-x[m-1]| \qquad (6)$$

where the epoch last sample is denoted as k, the window length is denoted as N, and the EEG time series is denoted as $x[m]$.

3.2.6 Mean Teager Energy (MTE)

EEG research was widely carried out using MTE feature, as given in Eq. (7).

$$MTE[k] = \frac{1}{N}\sum_{m=k-N+3}^{k}(x[m-1]^2 - x[m]x[m-2]) \qquad (7)$$

where the last epoch sample is denoted as k, the window length is denoted as N, and the EEG time series is denoted as $x[m]$.

3.2.7 Zero Crossing Rate

The rate of signal changes from negative to zero to positive or positive to zero to negative is denoted as Zero-Crossing Rate (ZCR), as in Eq. (8).

$$ZCR = \frac{1}{T-1}\sum_{t=1}^{T-1}1_{R<0}(S_t S_{t-1}) \qquad (8)$$

where the indicator function is denoted as $1_{R<0}$ and the length of the signal is denoted as S.

The above-mentioned features are extracted and forwarded to the process of feature selection for further enhancement of dimensionality and to enable better output on the overfitting of classifiers.

3.3 Feature Selection

The FCM – ETIFRST feature selection method is applied to the extracted features to select relevant features for arrhythmia classification.

3.3.1 Fuzzy C Means

Feature selection on the proposed method is done based on the clustering of Fuzzy C-Means and Enhanced Fuzzy Rough Set method. The FCM method has the advantage of producing better results for overlapped datasets than the k-means method. The data point is an assigned membership to every cluster to form the appropriate cluster [18]. The data point is applied with membership based on the distance between the data point and cluster center c. If more data is near to c, then the cluster has more data point membership.

The estimated membership is increased for every iteration. An enhanced fuzzy C-means algorithm is implemented in the proposed method. The algorithm is given as follows:

Step 1: Clusters number c varies from 2 to c_{max}. A certain value c is considered and initial class prototypes are selected and use $\epsilon > 0$.

Step 2: Image ς of new linearly weighted sum in terms of ς_k is calculated using Eq. (9).

$$\varsigma_k = \frac{1}{1+\alpha}\left(x_k + \left(\frac{\alpha}{N_R}\right)\sum_{j\in N_k} x_j\right) \tag{9}$$

where k^{th} pixel gray value of ς are ς_k and x_k (8-bit resolution is usually encoded for gray value) respectively. The neighbor x_j and the set of neighbours N_k is present in a window which is nearby x_k. The mean-filtered pixel value is $\sum_{j\in N_k} x_j$ where neighbor terms' effect controls α with a value approximately near to 1.

Step 3: Image matrix ς is updated using Eq. (10).

$$f_{il} = \frac{(\varsigma_l - v_i)^{\frac{-2}{s-1}}}{\sum_{j=1}^{c}(\varsigma_l - v_j)^{\frac{-2}{s-1}}} \tag{10}$$

The ith cluster prototype is v_i and f_{il} is gray value l of fuzzy membership that is with respect to cluster i. The l value varies from 1 to the number of gray levels denoted as d (gray level has a maximum value of 256 levels) of the image and each fuzzy membership of s weighted component parameter that estimates the final classification of fuzziness.

Step 4: Prototype v_i is calculated using Eq. (11).

$$v_i = \frac{\sum_{i=1}^{d} x_l f_{il}^{s} \varsigma_l}{\sum_{i=1}^{d} x_l f_{il}^{s}} \tag{11}$$

Convergence criteria are measured as $|v_{new} - v_{old}| < \epsilon$ by repeating step 3 and 4, cluster prototypes hold vectors v that is represented as $v = [v_1, v_2 \dots \dots v_c]$.

3.3.2 Enhanced Fuzzy Rough Set Feature Selection

The FCM has a wide range of applications and is significantly applied in numerous fuzzy clustering methods. Every sample point of membership degree to the class center is used as objective function optimization [19]. Euclidean distance is applied for the objective function of the sample point and clustering center. Solving the non-similarity index value function is the minimum value of every clustering center. The generalization is given in Eq. (12).

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^{m} d_{ij}^{2} \tag{12}$$

where the weighted index number is denoted as m, the jth sample point of ith clustering center of Euclidean distance is $d_{ij} = k_{ci} - x_{jk}$, the fuzzy set of clustering center, and u_{ij} is between 0 and 1. The constraint formula of the Lagrangian multiplier is constructed based on the derivate of input parameters in Eq. (13) and Eq. (14) to reach the minimum.

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{13}$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \tag{14}$$

The FCM outputs are centers c and membership matrix U. Every object degree contains U, which belongs to centers c. A fuzzy rough set of lower approximations is given in Eq. (15).

$$sim_a(x_i, x_j) = 1 - $$
$$\sqrt{\alpha \left(\mu_a(x_i) - \mu_a(x_j)\right)^2 + \beta \left(v_a(x_i) - v_a(x_j)\right)^2 + \gamma \left(\pi_a(x_i) - \pi_a(x_j)^2\right)} \tag{15}$$

where object x_i of hesitancy degree, non-membership, and membership are denoted as $\pi_a(x_i)$, $v_a(x_i)$ and $\mu_a(x_i)$, respectively. The weighted factors are α, β and γ.

Each pixel membership degree is normalized and initialized as in Eq. (16).

$$\mu_D(x_{ij}) = \frac{x_{ij} - (x_{ij})_{min}}{(x_{ij})_{max} - (x_{ij})_{min}} \tag{16}$$

where the image pixel of minimum and maximum intensity is denoted as $(x_{ij})_{max}$ and $(x_{ij})_{min}$. Complexity is reduced in the normalization process, which calculates in between 0 and 1.

Non-membership value is present in the new method in uncertain presence. A high grade of certainty is measured based on observations when the membership value is near 0 or 1 and the membership value is near 0.5 for a high grade of uncertainty. Non-membership value is measured in Eq. (17)

$$v_D(x_{ij}) = \left(1 - \mu_D(x_{ij})\right)\left(exp\left(\frac{-\mu_D(x_{ij})}{a\sigma}\right)\right)^{\frac{1}{a}} \tag{17}$$

where $(\mu_D(x_{ij}))$ membership value measures standard deviation σ, which is in the range of 0.39 – 0.41. The hesitation degree is measured using $\pi_D(x_{ij}) = 1 - \mu_D(x_{ij}) - v_D(x_{ij})$.

Every feature of fuzzy clustering gets every object membership degree, as given in Eq. (18).

$$\mu R_a(x) = U_{FCM(a)}[x] \tag{18}$$

Model equality is based on equivalence relations; approximate similarity or equality is measured based on fuzzy equivalence. Fuzzy equivalence relation R is used, as in Eq. (19).

$$R(x,y) = \frac{a - a|x-y| + bmin(x,y)}{a - (a-1)|x-y| + bmin(x,y)} \tag{19}$$

The size of the opening on one side is denoted as value b. If b is equal to 0, then the function makes sure to balance two objects, otherwise it is not able to balance the two objects.

3.4 Improved Emperor Penguin Optimization for Hyperparameter Optimization

In IEPO, three strategies are incorporated into EPO to maintain a suitable balance between exploration and exploitation to solve the drawback of the standard EPO [20]. These strategies are opposition-based learning, Levy flight, and Gaussian mutation. Equation (20) contains the fundamental formula for the EPO algorithm, which moves the penguins toward the center to locate the ideal value by measuring their distance from one another in the population. However, it is simple to get caught in the local optimal problem when the EPO algorithm optimizes the hyperparameters of LSTM. The strategy strengthens the bonds between penguins and improves each penguin's capacity for jumping to address this issue. The hyperparameters taken for LSTM have been given in the following Table 1.

The distance between the penguins is represented mathematically, as shown in Eq. (20), with the term Depth. The proposed method incorporates the Gaussian Mutation (GM) into Eq. (20) because the EPO's unpredictability is a vulnerability.

Table 1: Hyperparameters for Long Short Term Memory

Hyperparameters	Range
Number of hidden layers	3-10
Dropout	0.2 - 0.5
Decay rate	0.2-0.9
Learning rate	0.0 - 0.1
Momentum	0.5 - 0.9
Batch size	32,64,128,256
EPOCHS	10-100

The GM has excellent randomization capabilities, which it can use to strengthen the penguin relationships and accomplish the goal of leaving the local optimal. The following formula can be used to represent the updated mathematical model:

$$P_{eq}(x + 1) = P(x) - A \times Dep \oplus G(\alpha) \tag{20}$$

The individual penguins are then observed to move slowly while readily guiding the swarm to the regional ideal. The suggested strategy enhances the individual penguin's update position formula by utilizing the Levy flight's strong stochastics. The Levy flight expands the swarm's search area while simultaneously improving each penguin's capacity to jump. The best penguin can therefore swiftly arrive at the greatest answer. This is how it is formulated:

$$P\ levy\ eq(x + 1) = P(x) \times Levy - A \times Dep \oplus G(\alpha) \tag{21}$$

where, $P\ levy\ eq$ represents the updated position of the ith solution. Through the Levy flight mechanism, a new candidate solution is generated, which increases the knowledge diffusion across objects and yields a better solution, as Levy flight is a random process and the jump size leads to the Levy probability distribution function. The best penguin may also leave the search arena due to the addition of Levy flight and the strong randomness of Gaussian mutation. By doing this, the swarm might avoid falling into the infinite cycle and also the best solution from being updated. The Opposition Based Learning (OBL) is added to the EPO algorithm to reduce the upgraded algorithm's ability to jump. The OBL broadens the search domain's exploration, which in turn diversifies the swarm. It can be created using Eq. (22)

$$P_{eq} = P_{max} + P_{min} - P_{best} + r(P_{best} - P_{eq}) \tag{22}$$

where P_{eq} is the i_{th} opposite penguin's location within the search domain. The i_{th} variables lower and upper bounds are P_{max} and P_{min} respectively. P_{best} is the location of the best penguin, and r is an element-filled random vector within the range[0,1](0,1). P is the ith penguin in the population's position vector. Then, based on the fitness of the opposing sites, the best penguin is modified. To make the penguin change more quickly and discover the best value for the chosen hyperparameters, the new technique delivers greater performance and higher convergence velocity.

3.5 Classification using LSTM

Due to its outstanding performance in the extraction of temporal and spatial variables, the classification network used in the framework is an LSTM and its architecture is shown in Fig. 5. According to Fig. 5, the model is made up of an input layer, a single LSTM cell, a hidden layer, a normalization layer, and a classification layer.

Performance is improved by the idea of integrating the most recent data along with older data to predict the subjects one step ahead. With the benefits of a hidden layer self-feedback mechanism, the LSTM model solves the long-term dependence problem [21]. Gates of input state, output state, and forget state in the LSTM model are known to be three unique gates that are used to update data that are meant to be stored in the memory cell.
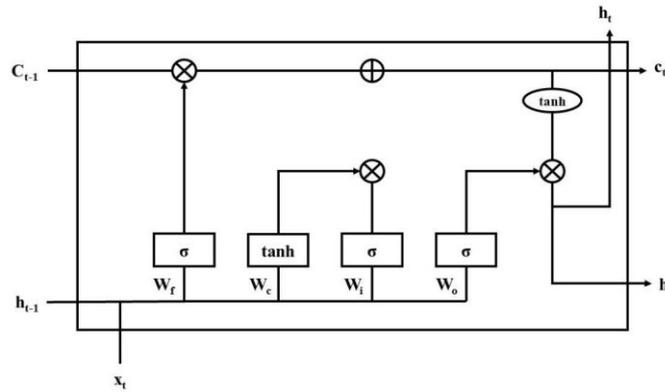
Fig. 5: LSTM architecture.

The input feature for LSTM is denoted and it contains the characteristics obtained via FCM-ETIRST. The previous instant output of the LSTM cell is marked as h_t, a cell of the memory unit is represented, and the output of the LSTM cell is indicated for the unit of LSTM input data at a time. The following steps illustrate the LSTM unit's process.

1) As shown in Eq. (1), the cell of memory for the candidate, the matrix of weight and the bias are assessed in (23).

$$\widetilde{c_t} = tanh\ (W_c.\ [h_{t-1}, x_t] + b_c) \qquad (23)$$

2) The gate of the input state modifies the memory state value depending upon the data input being used at the time; the function activation of sigmoid, matrix of weight, and bias are shown in Eq (24).

$$i_t = \sigma(W_i.\ [h_{t-1}, x_t] + b_i) \qquad (24)$$

3) By using past data, the gate of forget state modifies the memory state value; the bias gate is indicated by the weight matrix which is given in Eq (25).

$$f_t = \sigma(W_i.\ [h_{t-1}, x_t] + b_f) \qquad (25)$$

4) As shown in Eq (26), the value of the most recent LSTM unit is calculated by measuring the value of the current memory state.

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c_t} \qquad (26)$$

where the dot product is represented by "*". Depending upon the state values of the last and candidate cells, the input state and forget state gates regulate the updated memory cell.

5) The matrix of weight in the layer W_0, the bias b_0, and the output value of the gate controls the output of the memory cell value, as shown in Eq. (27)

$$Eq.o_t = \sigma(W_0.[h_{t-1}, x_t] + b_0) \tag{27}$$

6) As shown in Eq. (28), the output of the LSTM is measured as:

$$h_t = o_t * tanh (c_t) \tag{28}$$

According to the 3 gates of control and memory state, the LSTM continuously stores, reads, resets, and updates the data. Because the LSTM internal parameter is shared, the weight matrix's dimensions are changed to regulate the size of the output unit. The LSTM's input and feedback are separated by a significant amount of time. The internal state of the memory cell's architecture maintains a constant error flow, and the gradient doesn't explode or vanish [22]. The LSTM classifier predicts the type of arrhythmia present in the CHD and those results are evaluated in the following section.

4. RESULTS

In this work, the proposed modified model is replicated using enhanced LSTM with the system requirements. The performance of the modified model is analyzed by means of sensitivity, specificity, accuracy, F-score, and MCC on the actual educational dataset. The feature length of 506 is obtained after feature selection. Specificity is defined as the test to appropriately recognize the areas that are negatively less in learning skills. Sensitivity is defined as the test to correctly identify the regions that are positively high in learning skills. Further, accuracy is the most important performance measure that is utilized in the LMS, where it is the ratio of correctly predicted observations to the total observations. Accuracy, sensitivity, specificity, and MCC are mathematically defined in the Eqs. (29) to (33),

$$Specificity = \frac{TN}{FP+TN} \times 100 \tag{29}$$

$$Sensitivity = \frac{TP}{FN+TP} \times 100 \tag{30}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \tag{31}$$

F-score is determined as the harmonic mean of the model's recall and precision, and its mathematical equation is followed in Eq. (32).

$$F - score = \frac{2TP}{2TP+FP+FN} \times 100 \tag{32}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \times 10 \tag{33}$$

4.1 Quantitative Analysis

The performance analysis of enhanced LSTM without augmentation in accuracy, sensitivity, specificity, F-score, and MCC, with and without feature selection, are shown in Tables 2, 3, 4 and 5 respectively. Table 2 demonstrates the quantitative analysis of different classifiers with and without feature selection for the MIT-BIH dataset. Table 3 demonstrates the quantitative analysis of different classifiers with and without feature selection for the CPSC 2018 dataset.

The performance measure of classifiers on the MIT-BIH dataset with feature selection is highest compared to those without feature selection. The accuracy, sensitivity, specificity, precision, and MCC of Neural Network (NN), RNN (Recurrent Neural Network), GAN (Generative Adversarial Network), and DNN (Deep Neural Network) are measured and compared with the proposed IEPO-based LSTM. The obtained results show that the proposed IEPO based LSTM achieves the highest values with the accuracy of 99.75%, the

sensitivity of 98.29%, specificity of 98.29%, the precision of 98.35%, and MCC of 97.79% while considering feature selection compared to the other classifiers. The simulation results of the proposed method by varying different classifiers on MIT-BIH dataset, with and without feature selection are specified in Fig. 6 and Fig. 7.

Table 2: Simulation results of IEPO-based LSTM by varying the classifiers for the MIT-BIH dataset

Without Feature Selection Results				
Methods	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**	**MCC (%)**
NN	88.99	87.53	89.44	88.91	86.5
RNN	90.49	88.63	90.84	90.71	89.4
GAN	91.99	90.53	93.74	92.81	90.4
DNN	93.39	91.73	95.64	94.81	92.1
IEPO based LSTM	96.29	95.83	95.74	95.51	94.3
With Feature Selection Results					
NN	90.69	89.53	92.14	91.31	87.7
RNN	92.32	94.18	95.61	92.71	92.34
GAN	92.25	92.28	89.98	91.43	89.91
DNN	90.86	94.53	93.21	93.04	90.56
IEPO based LSTM	99.75	98.29	98.29	98.35	97.79
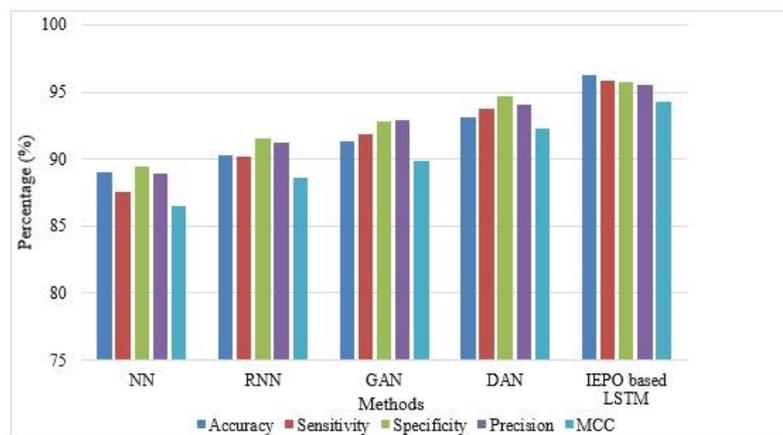
Fig. 6: Performance of MIT-BIH arrhythmia classification without feature selection.
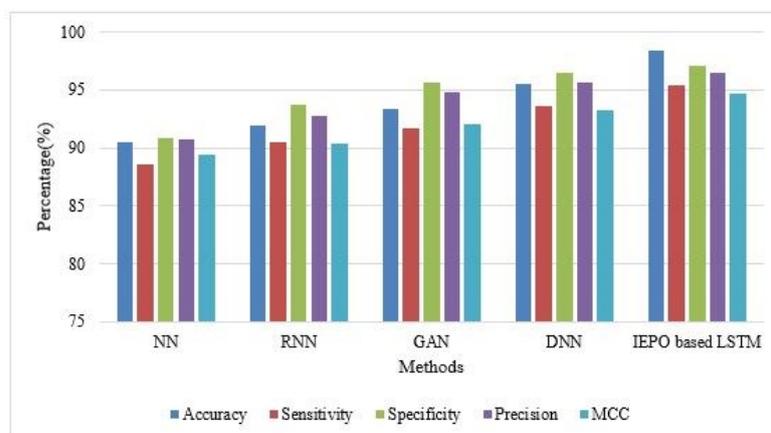
Fig. 7: Performance for MIT-BIH Arrhythmia classification with feature selection.

Ramaiah and Srikantegowda

Table 3: Simulation results of IEPO based LSTM by varying the classifiers for CPSC 2018 Dataset

Without feature selection results for CPSC 2018 Dataset					
Methods	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**	**MCC (%)**
NN	85.49	86.03	87.04	90.71	88.6
RNN	88.02	88.58	89.61	92.71	90.44
GAN	89.15	90.38	90.28	91.03	91.71
DNN	92.16	93.63	92.71	93.94	92.16
IEPO based LSTM	95.75	94.09	94.19	95.35	94.49
With Feature selection results					
Methods	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**	**MCC (%)**
NN	90.28	90.05	91.13	92.09	88.43
RNN	92.03	94.68	96.56	93.48	91.90
GAN	90.40	92.58	89.59	91.55	88.81
DNN	50.28	51.65	52.75	53.23	51.65
IEPO based LSTM	99.79	99.11	98.20	99.43	98.57

The performance measure of classifiers on CPSC dataset with feature selection is highest when compared with the values of performance metrics without feature selection. The accuracy, sensitivity, specificity, precision, and MCC of Neural Network (NN), RNN (Recurrent Neural Network), GAN (Generative Adversarial Network), and DNN (Deep Neural Network) are measured and compared with the proposed IEPO-based LSTM. The obtained results show that the proposed IEPO based LSTM achieves highest values with accuracy of 99.79%, sensitivity of 99.11%, specificity of 98.20%, precision of 99.43% and MCC of 98.57% while considering with feature selection compared to the other classifiers. The simulation results of the proposed method by varying different classifiers on CPSC dataset with and without feature selection are specified in Fig. 8 and Fig. 9.
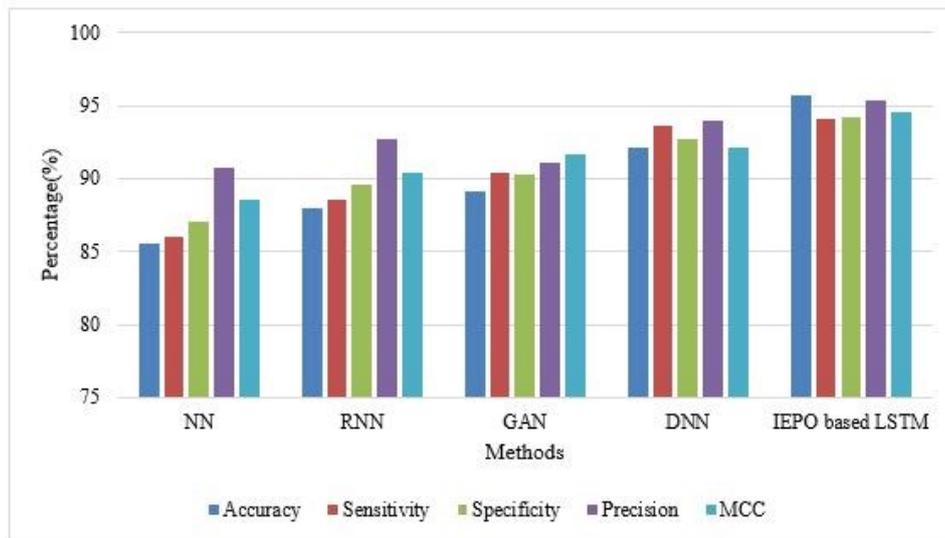
Fig. 8: Performance for CPSC Arrhythmia without feature selection.

The FCM – ETIFRST method selects the unique features to represent the class that solves the problem of imbalance and overfitting. The modified non-membership function helps to provide exploitation, and the modified membership function performs exploration. The similarity measure helps to maintain the exploration and exploitation of the feature

selection process. LSTM shows higher sensitivity than other classifiers due to its efficiency in handling the features.
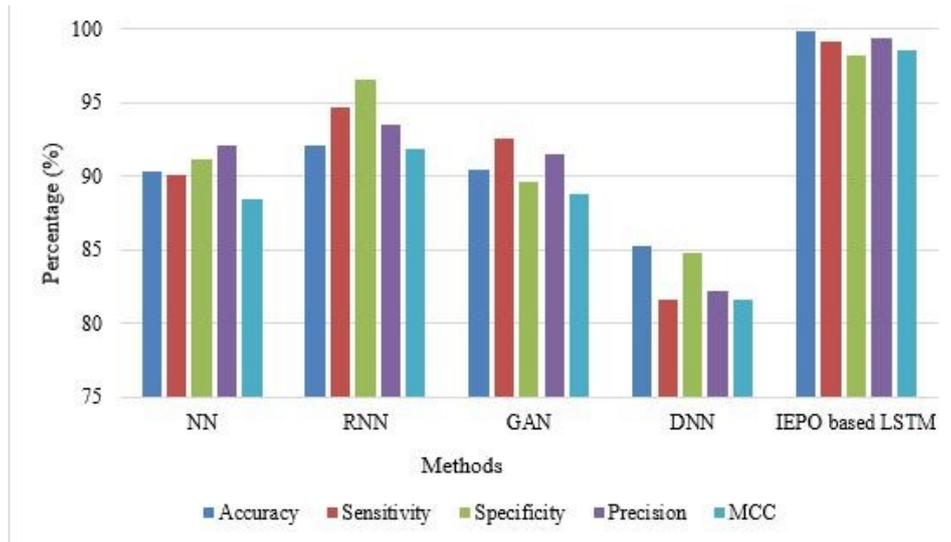
Fig. 9: Performance for CPSC Arrhythmia without feature selection.

Table 4 describes the performance of optimizing the hyperparameters of LSTM using various multi-objective optimization methods like Particle Swarm Optimization (PSO), Fruit Fly Optimization (FFO), Anti Colony Optimization (ACO), Salt Swarm optimization Algorithm (SSA) with the selected features obtained from FCM-ETIRST on MIT-BIH arrhythmia dataset. In the observation, the results infer that the proposed IEPO algorithm has chosen the best set of hyperparameters for LSTM to perform at its better stage, when compared to the resultant parameters obtained from other optimization algorithms. It also shows an improvement of around 3% in accuracy compared to the existing optimization algorithm SSA. The performance comparison of different hyperparameter optimization techniques on MIT-BIH dataset, with and without feature selection is graphically represented in Fig. 10 and Fig. 11.

Table 5 represents the performance analysis of K-fold validation in terms of performance metrics like accuracy, sensitivity, specificity, precision, and MCC on the MIT-BIH dataset and its graphical representation is specified in Fig. 12.

Table 4: Comparison of Hyper parameter optimization method on MIT-BIH dataset

Without Feature Selection Hyper Parameter Optimization method comparison					
Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	MCC (%)
PSO	85.89	84.69	86.14	80.59	86.38
FFO	83.33	84.23	84.06	83.24	86.00
ACO	96.17	92.03	94.83	97.08	97.54
SSA	95.31	90.67	96.00	94.04	96.30
IEPO based LSTM	98.95	99.49	99.49	98.45	97.99
With Feature Selection Hyper Parameter Optimization method					
PSO	85.89	84.69	86.14	80.59	86.38
FFO	83.33	84.23	84.06	83.24	86.00
ACO	96.17	92.03	94.83	97.08	97.54
SSA	95.31	90.67	96.00	94.04	96.30
IEPO based LSTM	98.95	99.49	99.49	98.45	97.99

Fig. 10: Performance of hyper parameter optimization with feature selection.
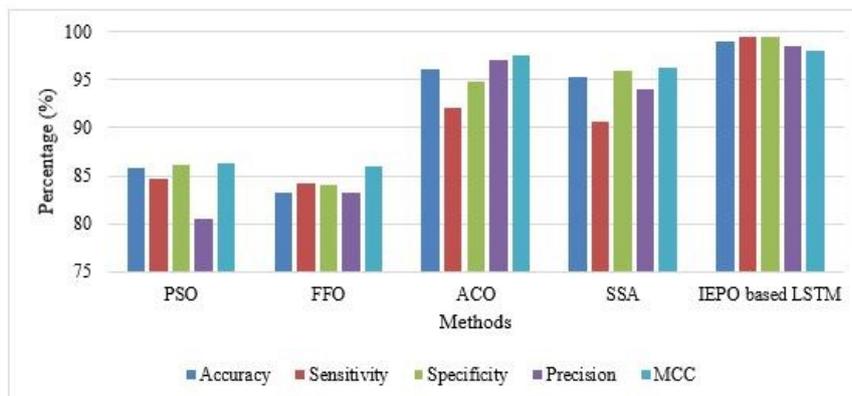
Fig. 11: Performance of hyper parameter optimization without feature selection.

Table 5: Performance analysis of K-fold validation on MIT-BIH dataset [13]

K-fold values	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	MCC (%)
4-fold	97.99	97.08	97.55	97.19	96.65
5-fold	98.95	99.49	99.49	98.45	97.99
7-fold	97.83	96.09	97.19	97.95	96.74
9-fold	97.47	96.28	96.15	96.37	95.45
10-fold	98.67	95.47	96.33	95.93	90.61
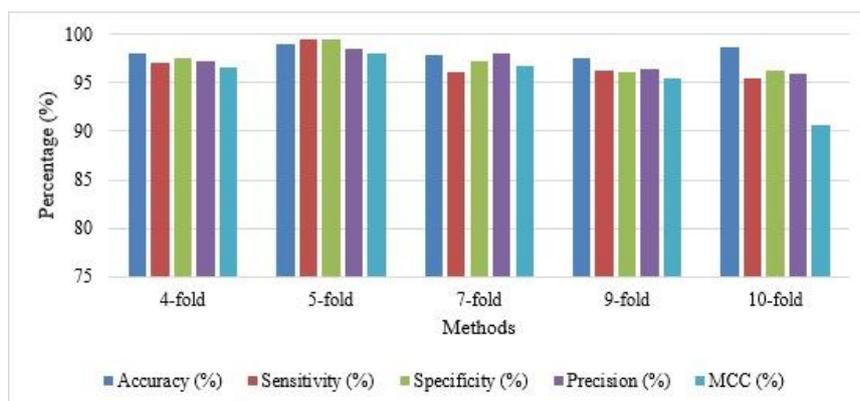
Fig. 12: K-fold validation performance.

5. COMPARATIVE ANALYSIS

The comparative analysis of the proposed VF classification using IPEO based LSTM to the existing Boosted SVM based DE algorithm and Hybrid CNN-SVM deep neural networks using the LSTM model in terms of accuracy, specificity, sensitivity, and precision is shown in Table 6. The proposed model is trained, tested, and validated using two publicly available standard datasets namely the Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) and the China Physiological Signal Challenge CPSC 2018 dataset and both consist of ECG recordings for 5 seconds of coronary heart disease (CHD) patients. The accuracy was improved to 99.75%, with a specificity of 98.29%, a sensitivity of 98.39% and the precision held was 98.35%. The proposed IPEO was evaluated for efficient classification using the VF technique by overcoming overfitting problems and dataset classification.

Table 6: Comparative analysis of existing SVM classifier to the LSTM classifier

Technique	Dataset	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Boosted SVM based DE algorithm[15]	MIT-BIH	98.20	98.18	98.25	-
Hybrid CNN-SVM deep neural networks using LSTM [16]	CPSC	98.69	99.99	97.74	-
IEPO based LSTM	MIT-BIH	99.75	98.29	98.39	98.35
	CPSC	98.11	97.21	97.79	98.29

6. CONCLUSION

In this paper, a novel classification method for classifying VF rhythm has been developed on the MIT-BIH and CPSC datasets consisting of ECG recordings of CHD patients. The method uses an LSTM classifier to classify VF and FCM- ETIFRST for feature selection and performs clustering of membership, non-membership, and hesitancy degrees. The method uses a window size of 5 seconds that requires less memory and obtained the highest accuracy of 99.75%. The proposed algorithm, on the other hand, combines traditional machine learning algorithms and CNN algorithms. The deep CNN was able to successfully extract useful deep features for ECG classification from raw ECG signals without prior knowledge of ECG signals or cardiac rhythm disorders. The proposed method outperformed the existing classification models on CHD by improving the efficiency of arrhythmia classification with suitable hyperparameter optimization in the LSTM algorithm. The improved performance of the proposed method on different test datasets suggests that it has the potential for clinical application. Future work can be focused on improving classification accuracy for both VF and AF by considering a window size of more than 5 seconds and it will be analyzed in real-time events.

DATA AVAILABILITY STATEMENT

Dataset Links:

MIT BIH Arrhythmia Dataset : https://archive.physionet.org/cgi-bin/atm/ATM
CPSC 2018 Dataset: http://2018.icbeb.org/Challenge.html

REFERENCES

[1] Nilashi M, Ahmadi H, Manaf AA, Rashid TA, Samad S, Shahmoradi L, Aljojo N, Akbari E. (2020) Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates. Int. J. Fuzzy Syst., 22(4): 1376-1388. https://doi.org/10.1007/s40815-020-00828-7

[2] Khan MU, Aziz S, Iqtidar K, Zaher GF, Alghamdi S, Gull M. (2022) A two-stage classification model integrating feature fusion for coronary artery disease detection and classification. Multimedia Tools and Applications, 81(10): 13661-13690. https://doi.org/10.1007/s11042-021-10805-3

[3] Shah SMS, Shah FA, Hussain SA, Batool S. (2020) Support vector machines-based heart disease diagnosis using feature subset, wrapping selection and extraction methods. Comput. Electr. Eng., 84: 106628. https://doi.org/10.1016/j.compeleceng.2020.106628

[4] Mohanty M, Dash M, Biswal P, Sabut S. (2021) Classification of ventricular arrhythmias using empirical mode decomposition and machine learning algorithms. Prog. Artif. Intell., 10(4):489-504. https://doi.org/10.1007/s13748-021-00250-6

[5] Muhammad LJ, Algehyne EA. (2021) Fuzzy based expert system for diagnosis of coronary artery disease in Nigeria. Health and Technology, 11(2): 319-329. https://doi.org/10.1007%2Fs12553-021-00531-z

[6] Nasarian E, Abdar M, Fahami MA, Alizadehsani R, Hussain S, Basiri ME, Zomorodi-Moghadam M, Zhou X, Pławiak P, Acharya UR, Tan RS, Sarrafzadegan S. (2020) Association between work-related features and coronary artery disease: A heterogeneous hybrid feature selection integrated with balancing approach. Pattern Recognit. Lett., 133: 33-40. https://doi.org/10.1016/j.patrec.2020.02.010

[7] Apostolopoulos ID, Groumpos PP. (2020) Non-invasive modelling methodology for the diagnosis of coronary artery disease using fuzzy cognitive maps. Comput. Methods Biomech. Biomed. Eng., 23(12): 879-887. https://doi.org/10.1080/10255842.2020.1768534

[8] Ma C, Wei S, Chen T, Zhong J, Liu Z, Liu C. (2020) Integration of results from convolutional neural network in a support vector machine for the detection of atrial fibrillation. IEEE Trans. Instrum. Meas., 70(2504610): 1-10. https://doi.org/10.1109/TIM.2020.3044718.

[9] Chen X, Cheng Z, Wang S, Lu G, Xv G, Liu Q, Zhu X. (2021) Atrial fibrillation detection based on multi-feature extraction and convolutional neural network for processing ECG signals. Comput. Methods Programs Biomed., 202: 106009. https://doi.org/10.1016/j.cmpb.2021.106009

[10] Kumari CU, Murthy ASD, Prasanna BL, Reddy MPP, Panigrahy AK. (2021) An automated detection of heart arrhythmias using machine learning technique: SVM. Mater. Today Proc., 45(2): 1393-1398. https://doi.org/10.1016/j.matpr.2020.07.088

[11] Li H, Wang X, Liu C, Zeng Q, Zheng Y, Chu X, Yao L, Wang J, Jiao Y, Karmakar C. (2020) A fusion framework based on multi-domain features and deep learning features of phonocardiogram for coronary artery disease detection. Comput. Biol. Med., 120: 103733. https://doi.org/10.1016/j.compbiomed.2020.103733

[12] Nguyen QH, Nguyen BP, Nguyen TB, Do TT, Mbinta JF, Simpson CR. (2021) Stacking segment-based CNN with SVM for recognition of atrial fibrillation from single-lead ECG recordings. Biomed. Signal Process. Control, 68: 102672. https://doi.org/10.1016/j.bspc.2021.102672

[13] Jahmunah V, Ng EYK, San TR, Acharya UR. (2021) Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals. Comput. Biol. Med., 134: 104457. https://doi.org/10.1016/j.compbiomed.2021.104457

[14] Tseng LM, Tseng VS. (2020) Predicting ventricular fibrillation through deep learning. IEEE Access, 8: 221886-221896. https://doi.org/10.1109/ACCESS.2020.3042782

[15] Panigrahy D, Sahu PK, Albu F. (2021) Detection of ventricular fibrillation rhythm by using boosted support vector machine with an optimal variable combination. Comput. Electr. Eng., 91: 107035. https://doi.org/10.1016/j.compeleceng.2021.107035

[16] Çınar A, Tuncer SA. (2021) Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks. Comput. Methods Biomech. Biomed. Eng., 24(2): 203-214. https://doi.org/10.1080/10255842.2020.1821192

[17] Chen TM, Huang CH, Shih ES, Hu YF, Hwang MJ. (2020) Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model. Iscience, 23(3): 100886. https://doi.org/10.1016/j.isci.2020.100886

[18] Chang KC, Hsieh PH, Wu MY, Wang YC, Chen JY, Tsai FJ, Shih ES, Hwang MJ, Huang TC. (2021) Usefulness of machine learning-based detection and classification of cardiac arrhythmias with 12-lead electrocardiograms. Can. J. Cardiol, 37(1): 94-104. https://doi.org/10.1016/j.cjca.2020.02.096

[19] Cai W, Chen Y, Guo J, Han B, Shi Y, Ji L, Wang J, Zhang G, Luo, J. (2020) Accurate detection of atrial fibrillation from 12-lead ECG using deep neural network. Comput. Biol. Med., 116: 103378. https://doi.org/10.1016/j.compbiomed.2019.103378

[20] Rai HM, Chatterjee K. (2022) Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data. Appl. Intell., 52(5):5366-5384. https://doi.org/10.1007/s10489-021-02696-6

[21] Baral S, Alsadoon A, Prasad PWC, Al Aloussi S, Alsadoon OH. (2021) A novel solution of using deep learning for early prediction cardiac arrest in Sepsis patient: enhanced bidirectional long short-term memory (LSTM). Multimedia Tools and Applications, 80(21-23): 32639-32664. https://doi.org/10.1007/s11042-021-11176-5

[22] Sun L, Wang Y, He J, Li H, Peng D, Wang Y. (2020) A stacked LSTM for atrial fibrillation prediction based on multivariate ECGs. Health Inf. Sci. Syst., 8(1): 19. https://doi.org/10.1007/s13755-020-00103-x