Journal of Theoretical and Applied Information Technology
15th July 2023. Vol.101. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DESIGN AND PSO BASED OPTIMIZATION OF BILSTM-CNN STACK OF NEURAL NETWORKS FOR ECG SIGNAL CLASSIFICATION

TATA BALAJI[1] , N.JAYA[2] , G.VENKATA HARI PRASAD[3]

[1]Research Scholar, Department of EIE, Faculty of Engineering & Technology, Annamalai University, Chidambaram, Tamil Nadu-608002, India.

[2]Professor, Department of EIE, Faculty of Engineering & Technology, Annamalai University, Chidambaram, Tamil Nadu-608002, India

[3] Professor, Department of ECE, Anurag Engineering College, Kodad, Telangana-508206, India

Email.:[1]balu170882@gmail.com,[2] jayanavaneethan@rediffmail.com, [3]drgvhariprasad.ece@anurag.ac.in

ABSTRACT

Classifying ECG signals and analyzing the likelihood of cardiac arrest are essential aspects of the medical field. In recent years, artificial intelligence methods like Artificial Neural Networks have been used to create models for ECG data classification. More sophisticated deep learning methods are effective, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. A CNN was implemented to extract the characteristics from the images to forecast the percentage of porosity in the successive layers of simulated porosity images. Yet formerly, most LSTM-CNN hyperparameters were determined by experience, which frequently did not result in the most outstanding results. The BiLSTM-CNN hyperparameter was modified using the Particle Swarm Optimization (PSO) technique to improve the ability to learn the properties of data sequences. The BiLSTM-CNN-PSO model offers better prediction stability and accuracy. This study showed that the use of CNNs and Bi-LSTM networks can provide accurate classification of ECG signals. The optimized model achieved a categorization accuracy of 99.2%, which is significantly higher than the mean accuracy of 85.3% achieved by the unoptimized model. The results could contribute to long-term cardiac arrest prevention in the study area.

Keywords: *ANN, CNN, Bi-LSTM, hyperparameters and PSO.*

1. INTRODUCTION

The Electrocardiogram (ECG) analysis and associated cardiac diagnosis procedures are the fundamental healthcare approaches that offer quick insights into future health issues through either automatic identification of suspected cardiac irregularities or simple viewing and interpretation by physicians[1]. ECG signal classification is a crucial task in the field of cardiology as it helps in identifying various cardiac abnormalities and diagnosing heart diseases. The ECG signal reflects the electrical activity of the heart, and the analysis of this signal provides important diagnostic information that helps in determining the health of the heart. ECG signal classification is important for the early detection of cardiac abnormalities, which can help in the timely treatment of heart disease and improving patient outcomes. ECG signal classification is used in various medical applications, including arrhythmia detection, ischemia detection, heart rate variability analysis, and QT interval analysis. Arrhythmia detection is particularly important, as it helps in identifying irregular heartbeats, such as atrial fibrillation, which can increase the risk of stroke and other serious health problems. In particular, for non-linear systems, neural networks have been extensively used in system identification due to their non-linear solid mapping capabilities. The hyperparameters' values can considerably impact the performance of the neural network. Yet, there must be a theoretical foundation for its value, and various systems frequently demand multiple values [2]. Therefore, studying a technique for automatically optimizing a system identification algorithm's hyperparameters is extremely practical. Following is a summary of the paper's contribution in light of the information provided: PSO was used to find the best hyperparameters, like the number

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of neurons, to improve the categorization of ECG signals. The CNN-BiLSTM-PSO model suggested in this study outperformed standard approaches regarding statistical error metrics using ECG data for evaluation.

2. RELATED WORK

The number of hidden layers and the number of neurons in each hidden layer of the DNN are two hyperparameters specific to DL models that could affect the model's performance and learning outcome [3].

One of the simplest methods is grid search, which explores the grid in a feature space determined by the hyperparameters. Grid search is simple to use and can quickly identify a decent set of hyperparameters that are not strongly connected. Unfortunately, it is greatly plagued by the dimensionality curse and rarely finds an ideal point not situated at a grid point. Because of this grid search constraint, Bergstra and Bengio thought about random search, developed a random point generation strategy to improve speed, and developed a more organized Bayesian optimization of hyperparameters. Since they may achieve both efficacy and efficiency, metaheuristics like evolutionary algorithms are receiving more attention [4]. A genetic algorithm (GA) was created to select the ideal topology of a CNN, including the size of the filter and the number of layers. A cutting-edge evolutionary approach for minimizing the mean absolute error and finding an appropriate RNN structure [5-6].

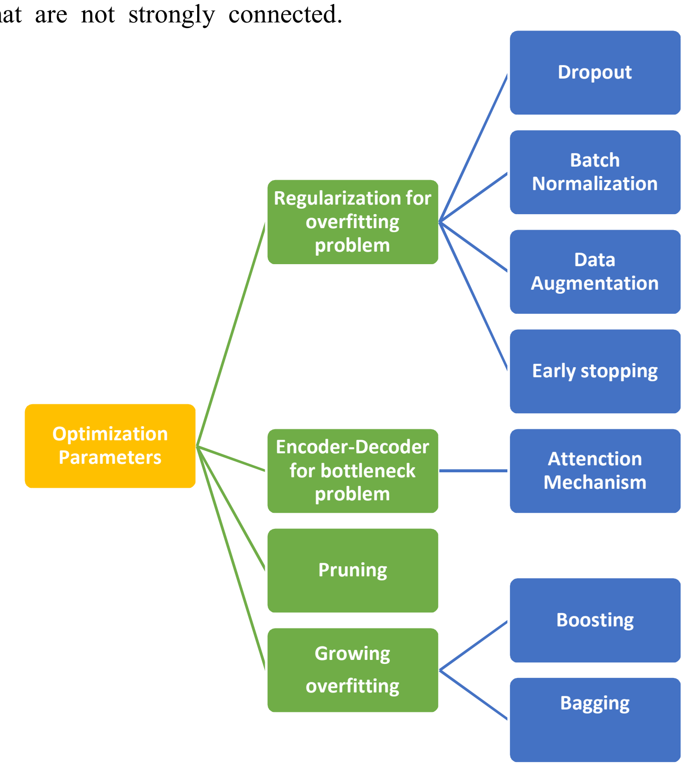


Figure 1. Network Tuning/Optimization Parameters.

Despite having many neurons, the few LSTM layers cannot detect complicated and latent data patterns. On the other hand, overfitting and slow convergence may result from too-deep LSTM layers. Similar to the LSTM layer, hidden layer neurons have a similar function. The values of the hyperparameters have a significant impact on how well the CNN architecture performs [7]. It is critical for the proper design of the GA to determine whether a population's chromosome is excellent or poor using a fitness function. As a result, we propose the fitness function below that is proportional to Acc, where Acc denotes accuracy and c denotes a constant [8].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$fitness\_function = \frac{1}{C} Acc \quad (1)$$

Swarm intelligence algorithms and evolutionary algorithms are the two main categories of metaheuristic algorithms. Genetic algorithms, evolutionary strategies, and evolutionary programming (EP) are the three subclasses of evolutionary algorithms (EA), which are driven by biological evolution (GA) [9]. The interaction between individuals locally and with their environment gives rise to collective intelligence. By combining the benefits of two or more metaheuristic algorithms or by adding new operations to the existing metaheuristics, hybrid metaheuristic algorithms can produce faster convergence rates, improved accuracy, and better balance between the exploration and exploitation processes [10].

CNN and RNN are two types of neural networks. GRU and LSTM are sub-classified in RNN with adopted bi-directional practice offering better efficiency. The CNN provides high accuracy and Fast training, but it is complex, and max-pooling leads to losing the feature position/order. RNN offers high reliability and can process sequential data. But RNN is difficult, expensive and high training time. The attention is naturally used to advance Bi-LSTM/Bi-GRU efficiency with automatic learning.Multi-LSTM employs a recursive LSTM unit to construct structured representations from predefined parsing trees [11]. This model should be able to capture the structured data that influences the subsequent prediction. The stack of CNN, max-pooling and LSTM offer high efficiency with long sequence processing capability. This stacking approach does not separate sections, and LSTM is used to process the order of features from max-pooling [12].

Table 1. Optimization of neural networks for ECG signal classification as per state of art.

Ref.	Neural Network stack	Optimizer	Batch size	Epochs	Learning rate	Loss function	Drop out	Training time in minutes	Accuracy in %
[13]	Bi-LSTM+ FFNN	Minimum batch gradient descent	120	70	0.2	Cross entropy	0.5	40	91.78
[14]	CNN+ LSTM	Adam	32	20	0.0012	Cross entropy	0.25	30	90.5
[15]	CNN + Bi-LSTM + Attention (SLCABG)	Adam	128	8	0.2	Cross entropy	0.4	80	91.9
[16]	CNN+ Bidirectional GRU + Attention	Minimum batch gradient descent	56	50	0.001	--	0.5	35	91.5
[17]	CNN+RNN	Adam	82	56	0.2	Cross entropy	0.32	40	91.4
[18]	RNN+ Attention	Adam	128	49	0.14	--	0.46	48	89.56

3. PROBLEM STATEMENT

The problem addressed in this study is the classification of cardiac dysfunctions using ECG signals. The study uses a combination of CNNs and Bi-LSTM networks to achieve high accuracy in the classification of ECG signals. The study aims to demonstrate that extracting features from the unprocessed ECG signal and converting it into spectral characteristics can improve the accuracy of classification models. To optimize the performance of the classification models, the study uses PSO to find the optimal values for various hyperparameters. The hyperparameters include the number of CNN and LSTM layers, the learning rate, batch size, activation function, dropout rate, optimizer, and number of epochs. PSO is used to find the optimal set of hyperparameters that result in the highest accuracy of the classification models. The study demonstrates that using a combination of CNNs and Bi-LSTM networks and optimizing the hyperparameters using PSO can significantly improve the accuracy of ECG signal classification. The results of the study can potentially lead to better diagnosis and treatment of heart diseases, improving patient outcomes and quality of life.

4. METHODOLOGY

An ECG is a quick bedside examination that records the electrical activity the contracting heart produces. Many cardiac illnesses like arrhythmia, cardiomyopathy, coronary heart disease, cardiovascular disease, and many others are frequently recognized using this method. Doctors and other medical professionals have always conducted the inspection process, which is a time-consuming operation that needs a lot of medical and human resources to process the massive amount of ECG data. On the other hand, various problems may occur due to the diversity of ECG signals, making this ECG inspection process much more difficult. For instance, two healthy individuals' ECGs cannot be identical. Moreover, two people with the same heart condition may exhibit various symptoms in their ECGs. Another problem can be the similarity of the ECG signals between two distinct disorders. There are no set guidelines that must be followed when making a diagnosis. AI approaches are required to identify hidden patterns that humans cannot find. These methods can be learned through collected experience.

.

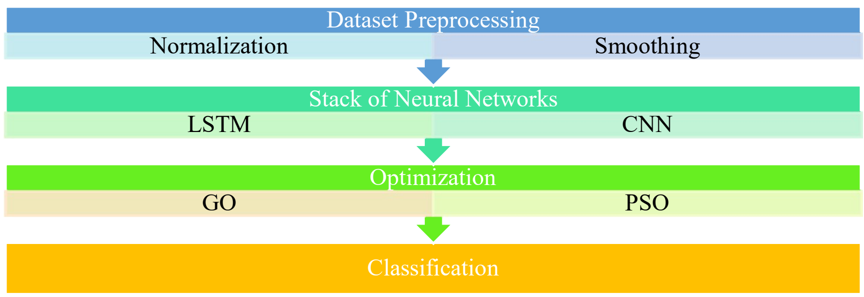


Figure 2. Detailed research.

5. IMPLEMENTATION

This work primarily describes the implementation of optimized stack of neural network model using CNN and Bi-LSTM for ECG signal classification.

4.1. Convolution Neural Network

The CNN algorithm was used to reduce the complexity of the function's space and to excerpt applicable text properties. CNN is particularly well-known for image categorization. However, CNN's ability in SA has been demonstrated via word-embedding, which is capable of converting text to an embedded vector sentence matrix [19-20]. The input text can be denoted as,

$$Ti{:}k = T1 \oplus T2 \oplus T3 \text{ --------- } \oplus Tp \qquad (2)$$

Sentence length(t), WE size(v). The text matrix 'T' size is t × v. Let Tj ∈ Rv be the v-dimensional word vector equivalent to the jth word. After including necessary padding, 'p' is the text length. Convolution contains a filter w ∈ Rev, which is executed on a window of 'e' words to create a novel feature. Kj feature is created from a window of words T(j : j + e − 1) i.e.,

$$Kj = Y\,(w \,.\, T(j : i + e - 1) + B) \qquad (3)$$

The feature map f ∈ Rt−e+1 is designed by including a bias (B) and activation (Y). Filter is applied to each achievable window of words in the sentence i.e., {T1:e,T2:e+1,- - - - - - - -Tf-e+1:f} to create feature map i.e.,

$$k=[k1,k2,k3\text{------}kf\text{-}e\text{+}1] \qquad (4)$$

If the text matrix size is t × v, filter size is a × b and feature map will be (t - a +1) × (v− b + 1). The max-overtime pooling procedure is to take the full significance, i.e., for every feature map that has been put through a pooling feature to start creating a fixed time vector.

$$\hat{k} = \max(k) \qquad (5)$$

Dropout avoids co-adaptation of hidden units by randomly setting to zero and offer regularization.

The last but one layer, $X = [\hat{k}_1, \ldots, \hat{k}_m]$, in its place of using,

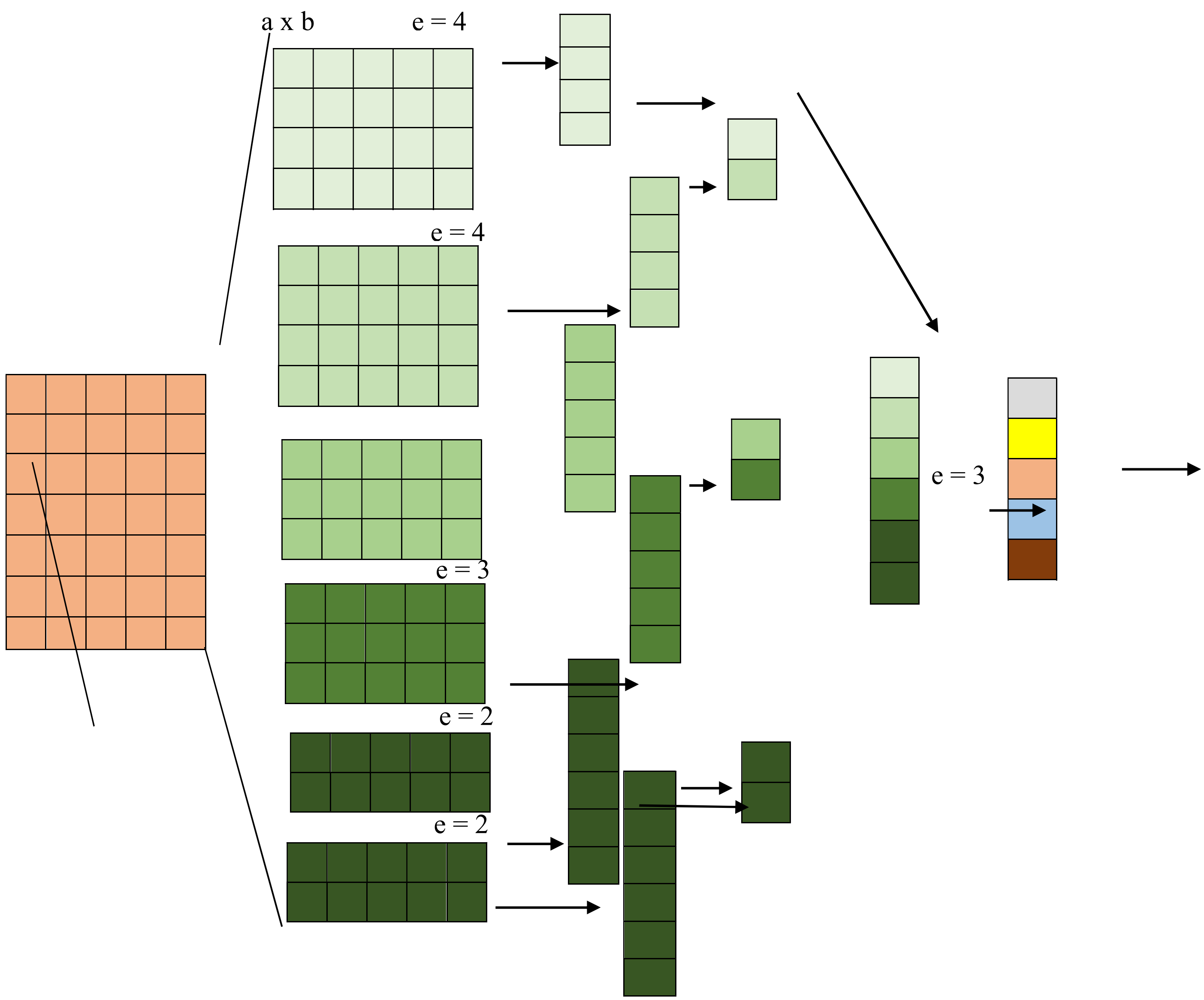
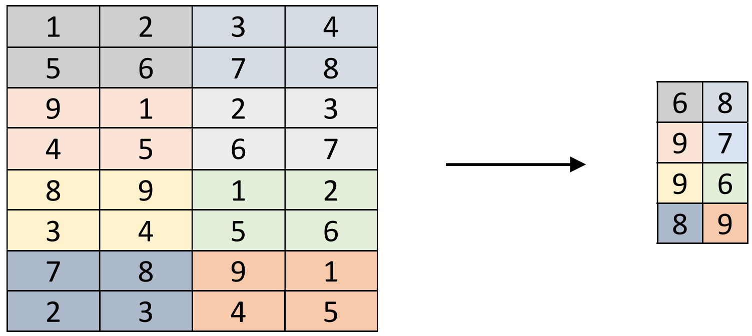
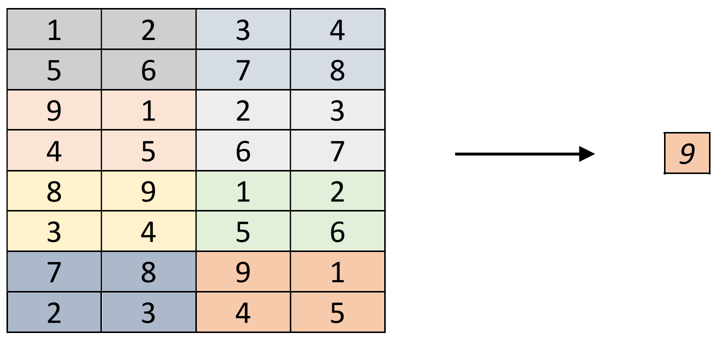


Figure 3. CNN for ECG Features Extraction.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(a)



(b)

Figure 4. Pooling, (a) Max-pooling, (b) Global max-pooling.

$$y = w \cdot X + B \tag{6}$$

for output unit 'S' in forward propagation, dropout uses

$$S = w \cdot (X \circ r) + B \tag{7}$$

The result of the max-pooling layer was the largest value in a specific subarea.

4.2. Bi-LSTM

Bi-LSTM involves a memory cell (ct) that maintains its state through time intervals of any length [21-23]. The input gate (it), forget gate (ft), and output gate (o) are the three non-linear gates that make up the LSTM unit (ot). The data flow regularization in ct is the main aim of it, ft, and o. At time t, input vector is 'xt' and hidden state vector is 'ht'. The ft chooses what information needs to be forgotten by output a number in [0, 1] i.e.,

$$ft = \sigma(Wf\ \ ht\text{-}1 + Uf\ xt + bf) \tag{8}$$

The it will decides what new information must be stored by computing it and $\widetilde{c}_t$ and uniting i.e.,

$$it = \sigma(Wi\ ht\text{-}1 + Ui\ xt + bi) \tag{9}$$

$$\widetilde{c}_t = \tanh(Wc\ ht\text{-}1 + Uc\ xt + bc) \tag{10}$$

$$ct = ft \odot ct\text{-}1 + it \odot \widetilde{c}_t \tag{11}$$

The ot gate will decides which parts of the cell state should be output i.e.,

$$ot = \sigma(Wo\ ht\text{-}1 + Uo\ xt + bo) \tag{12}$$

$$ht = ot \odot \tanh(ct) \tag{13}$$

The output of Bi-LSTM, hidden vector matrix (H ∈ Rd×N), size of hidden layers (d) and length (N). H contains of output vectors [h1, h2,..., hN]. To execute attention, we calculate α as follows:

$$z = \tanh(W1 * H) \quad (14)$$

$$\alpha = \mathrm{softmax}(wTz) \quad (15)$$

$$r = H\alpha T \quad (16)$$

Where W1 ∈ Rd×d, w∈Rd, z ∈ Rd×N, α ∈ RN. tanh used for sentence weighted representation r ∈ Rd. The hidden depiction (hfinal), is written as,

$$hfinal = \tanh(W2 * r + b2) \quad (17)$$

Where W2 ∈ Rd×d, b2 ∈ Rd. A softmax classification is used to predict the name $\hat{y}$ for a sentence S from a set of discrete 27 classes using the outputs of the final statement depiction.

$$\hat{y} = \mathrm{softmax}(W3 * hfinal + b3) \quad (18)$$

Where W3 ∈ R1×d , b3 ∈ R classes, W1, W2, W3 are weights and b1, b2 bias parameters. With the publication of a seminal study on machine translation, focus processes gained traction in NLP and signified a big step forward in the field. Models of emphasis are used to assign varying weights to phrases that connect to the sentiment of a text in different ways. A popular technique of giving varying weights to distinct phrases in a sentence is to use a weighted mixture of all hidden states, as follows.

$$\alpha_t = \frac{\exp(V^T.\ \tilde{h})}{\sum_t \exp(V.\ \tilde{h})}$$

(19)

$$S_{A_w} = \sum_t \alpha_t\ h_t$$

(20)

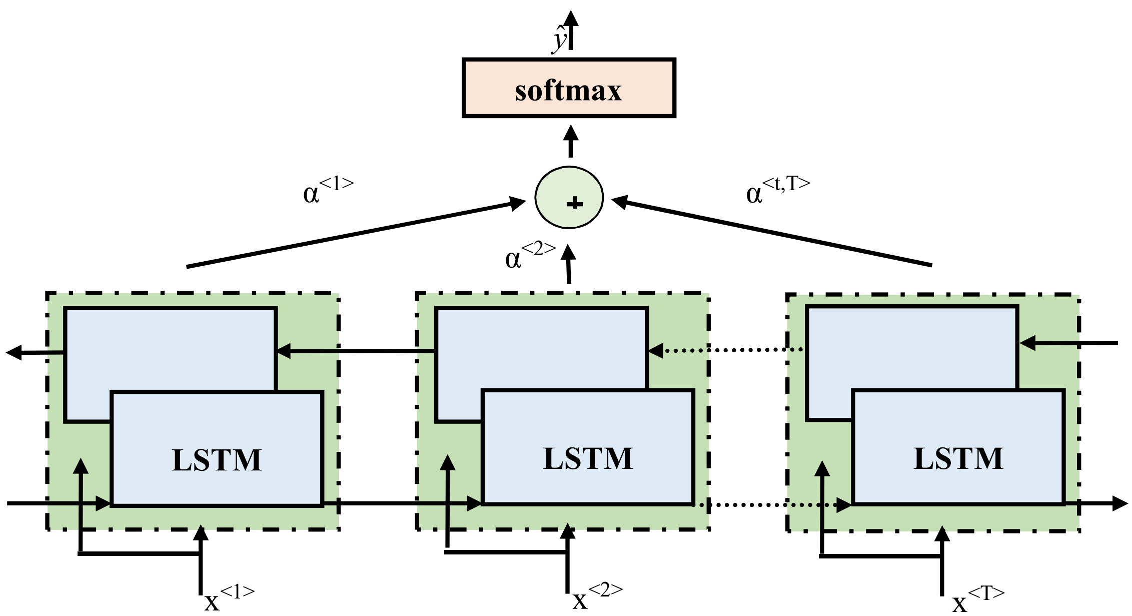


Figure 5. Bi-LSTM.

4.3. PSO

The number of hidden layers and the number of neurons in each hidden layer of the DNN are two hyperparameters specific to DL models that could affect the model's performance and learning outcome. One of the simplest methods is grid search, which explores the grid in a feature space determined by the hyperparameters. Three variables are used to describe the particle properties: location, velocity, and fitness value. The fitness function establishes the fitness value. The particle autonomously modifies its direction of travel and its distance based on the outstanding global fitness value, ultimately settling on the best option. To determine the optimum option, the PSO system is set up using random variables and updated after each cycle. In the multidimensional solution space, each possible solution, designated as a particle, is represented by a point. The particles enter the solution region at a predetermined speed and look for the best solution. Each particle modifies its location and velocity in response to its own experiences and those of its neighbors. The main idea behind PSO is to change each swarm's velocity as it approaches the pbest and gbest locations at each cycle. Suppose k particles form the group A = a1, a2, ak in an M-dimensional space, where ai = [ai1, ai2, aiM]. The i-th particle currently has the following characteristics:

$$A_i = (a_{i1}^t, a_{i2}^t, ...a_{iM}^t)^T$$
(21)

$$V_i = (v_{i1}^t, v_{i2}^t, ...v_{iM}^t)^T$$
(22)

$$P_i = (p_{i1}^t, p_{i2}^t, ...p_{iM}^t)^T$$
(23)

$$P_g = (p_{g1}^t, a_{g2}^t, ...a_{gM}^t)^T$$
(24)

In equations (21) and (22), at i and vti are the location and the velocity, where Pti and Ptg are the optimum values in the candidate solution. The position and velocity deviations per cycle using next situations (25) and (26).

$$v_i^{t+1} = wv_i^t + c_1 R_1^t (P_i^t - a_i^t) + C_2 R_2^t (P_g^t - a_i^t)$$
(25)

$$a_i^{t+1} = a_i^t + v_i^{t+1}$$
(26)

Here, vt+1i and at+1i shows the in what way velocity and position changes after each cycle. C1 and C2 represents the constants and R1 and R2 are arbitrary integers between (0, 1). To determine whether a particle swarm is fit, the initial output error of the LSTM is used.

$$fitness = \frac{numberofmissclassifications}{Total}$$
(27)

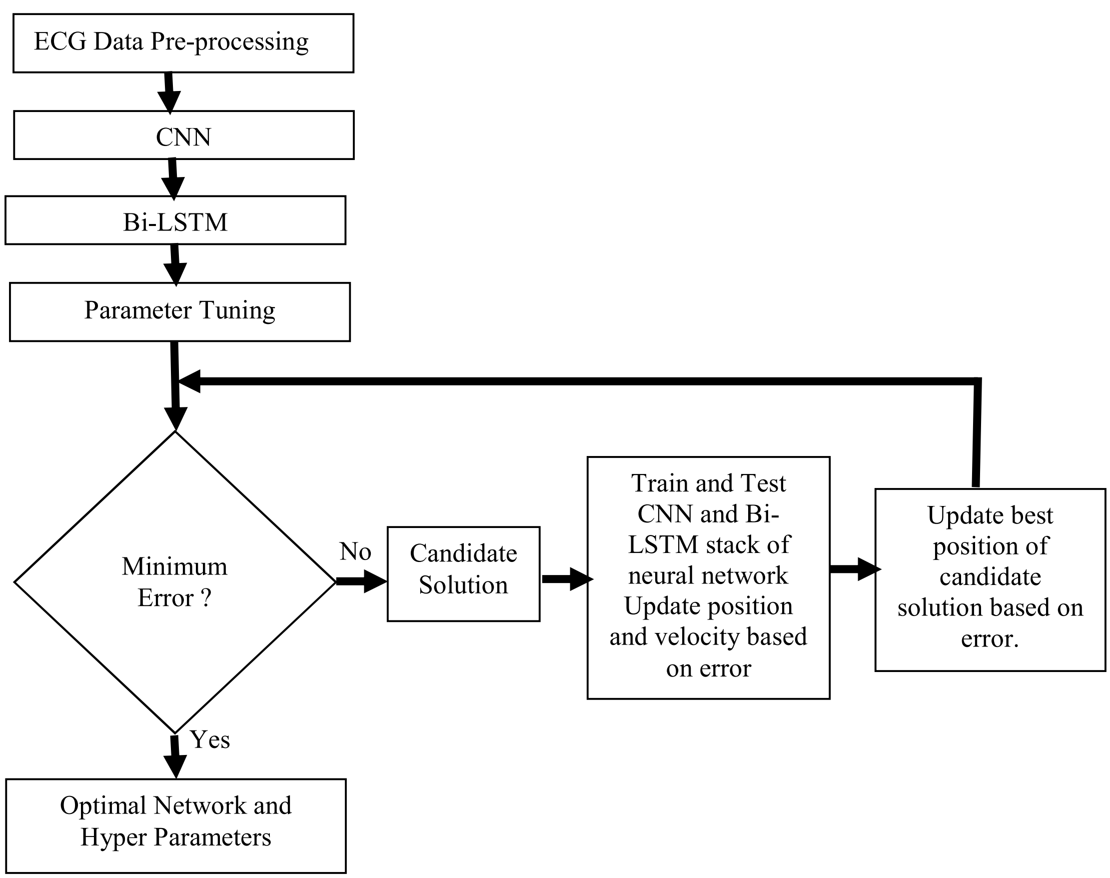


Figure 6. Presented BiLSTM-CNN-PSO model for ECG signal Classification.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.4. Cascaded Feed Forward Neural Network

In contrast to feed-forward networks from the input to each successive layer, as seen in Figure 7. If enough hidden neurons exist in a two-or-more layer cascade network, it can learn any finite input-output relationship arbitrarily, similar to feed-forward networks.

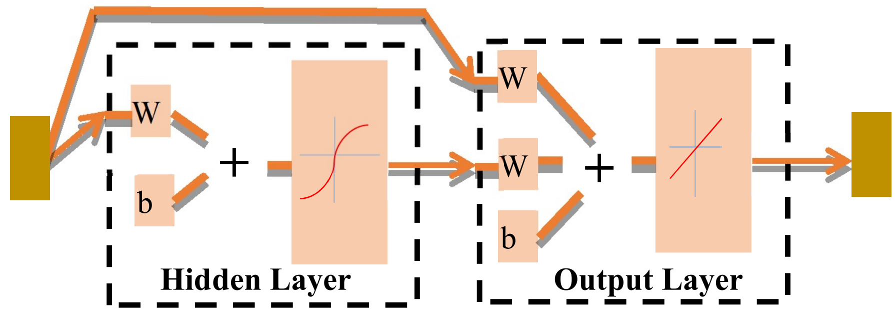


Figure 7. Cascade Forward Network.

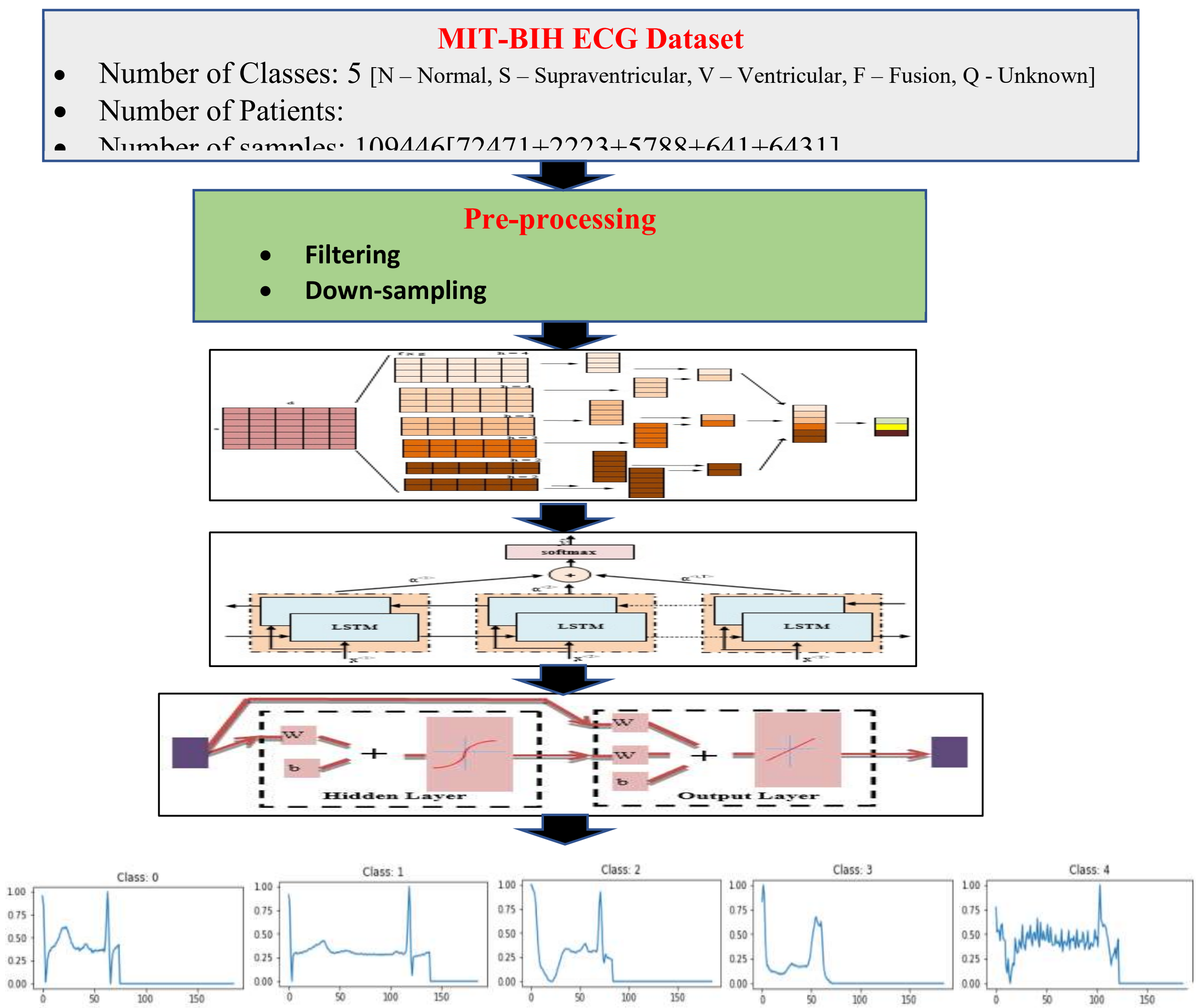


Figure 8. Proposed Optimized Stack of Neural Network for ECG signal Classification.

The convolution, pooling, and fully-connected layers are the fundamental components of CNNs. It can effectively complete various visual tasks by properly stacking these layers in a deep network. Factors including the number of convolutional layers and filters, filter size, and batch size greatly influence any CNN's performance. The CNN hyperparameters are the name given to these variables. The success rate of the CNN in addressing a specific problem depends on the values of the hyperparameters; only a CNN architecture consistently produces satisfactory results for all problem cases. One of the biggest problems in the CNN domain is determining the appropriate hyperparameter values for a given problem.

6. RESULTS AND DISCUSSIONS

TP is the count of sequences correctly identified as belonging to a specific class. TN is the count of arrangements accurately categorized as not belonging to a particular class. FP is the count of sequences that should have been classed as belonging to a different class but were instead classified as belonging to the class being tested. FN is the count of sequences corresponding to the tested class but wrongly identified as belonging to a different class. The accuracy of the forecasting model is measured in precision. The percentage of instances projected as positive that is positive is referred to as precision. The recall of a classifier measures how accurate it is. It denotes the percentage of cases from the positive class that were successfully anticipated. F1 Score is the harmonic mean of precision and memory; in other words, it reflects the balance of precision and recall.

$$Sensitivity = \mathrm{Re}call = \frac{TP}{TP+FN}$$

(28)

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

(29)

$$\Pr ecision = \frac{TP}{TP+TN}$$

(30)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (31)$$

Table 2. CNN and Bi-LSTM hyperparameter configuration.

Network	Parameter	Range	Value(s)
CNN	Number of CNN Layers	[1,2,3,4]	2
	Optimizer	[Adam, RMSprop, Adadelta, SGD, Adagrad]	Adam
	Learning rate (α)	[0.0001, 0.01, 0.1]	0.0001
	Activation Function	[ReLu, tanh, Softmax, linear]	tanh
	Batch size	[8, 16, 32, 54]	32
	Epoch number	[30, 50, 70, 110, 130]	50
	Padding valid Stride	[0,1,2]	1
	Loss function	---	MSE
LSTM	Number of LSTM Layers	[1,2,3,4]	2
	Learning rate	[0.0001, 0.01, 0.1]	0.01
	Batch size	[32, 54]	54
	Dropout	[0.1, 0.2, 0.4, 0.6]	0.2
	Activation function	[ReLu, tanh, Softmax, linear]	ReLu
	Optimizer	[Adam, RMSprop, Adadelta, SGD, Adagrad]	RMSprop
	Epochs	[30, 50, 70, 110, 130]	70
Fully Connected (FC)	Number of FC Layers	[1,2,3,4]	2
	Number of neurons in each layer	[40, 60, 80, 100, 120], [20, 40, 60, 80, 100]]	[120, 20]
	Size	--	[16,2048]

Optimization using PSO is a metaheuristic optimization algorithm that can be used to find the optimal values for the parameters of a neural network. In PSO, a swarm of particles represents candidate solutions, and each particle's movement is determined by its current position and the best position it has found so far, as well as the best position of the swarm. The algorithm iteratively updates the position of each particle based on its fitness value until a satisfactory solution is found. In the context of ECG signal classification, PSO can be used to optimize a wide range of network

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

parameters, including the number of CNN and LSTM layers, learning rate, activation functions, batch size, dropout, padding, stride, loss function, number of fully connected layers, and the number of neurons in each layer. By tuning these parameters, the PSO algorithm can help find the best combination of parameters that can achieve optimal classification accuracy.
Two convolutional layers were used to extract features from the ECG signals. Adam was chosen as the optimizer to update the weights of the neural network during training. A small learning rate of 0.0001 was used to avoid overshooting the optimal weights during training. The hyperbolic tangent (tanh) activation function was used in the CNN layers. A batch size of 32 was used during training, which represents the number of samples processed in each iteration. The neural network was trained for 50 epochs, which represents the number of times the entire training set was passed through the network during training. A stride of 1 was used with valid padding to ensure that the output feature maps have the same size as the input feature maps. Mean Squared Error (MSE) was used as the loss function to measure the difference between the predicted and actual ECG signal. Two LSTM layers were used to capture temporal dependencies in the ECG signal. A learning rate of 0.01 was used for the LSTM layers to update the weights during training. A batch size of 54 was used during training, which represents the number of samples processed in each iteration. A dropout rate of 0.2 was used to prevent overfitting during training. The rectified linear unit (ReLU) activation function was used in the LSTM layers. RMSprop was chosen as the optimizer to update the weights of the neural network during training. Two fully connected layers were used to map the features extracted by the CNN and LSTM layers to the output classes. The first fully connected layer had 120 neurons, while the second layer had 20 neurons. The input ECG signals were of variable size, ranging from 16 to 2048 samples.

Table 3. Performance Analysis of different stack of neural networks for SA.

Stack of Neural Networks	Training time in minutes	Epochs	Layers		Accuracy (%)
			Number	Type	
CNN+LSTM	40	72	1	Embedding	85.3
			1	Convolutional	
			1	LSTM	
			1	FC	
CNN+ Bi-LSTM	40	55	1	Embedding	92.35
			1	Convolutional	
			1	Bi-LSTM	
			2	FC	
CNN+ Bi-LSTM+ PSO	30	22	1	Embedding	99.2
			2	Convolutional	
			2	Dropout	
			1	Bi-LSTM	
			2	FC	

The study covered in this article demonstrates that utilizing a BiLSTM network to extract features from raw ECG data is a reliable method for improving classification accuracy. For a single raw ECG signal, the mean BiLSTM accuracy for the testing set was 85.3%. The same statistic for the CNN and BiLSTM using a twofold spectral ECG signal was 92.35%. Five types of cardiac dysfunctions were characterized to get to this finding. Bi-LSTM and CNN work better since the ECG signal can be translated into spectral properties. CNNs were designed to categorize. However, their memorylessness makes them ineffective in forecasting time series or signals. Bi-LSTM, which detects images after transforming a single input signal into a twofold spectral signal, combines signal memory and high performance. As a result, the network that was trained in this way has exceptional quality. The neural network stack that was demonstrated was optimized using PSO. The optimal values for most of the CNN and LSTM-related parameters were identified. The upgraded network has a 99.2% accuracy rate for categorizing ECG signals.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4. Comprehensive analysis of presented work with literature.

Parameters \ References	[24]	[25]	[26]	Presented work
Neural Network stack	CNN+GRU	Bi-LSTM+ Attention	RNN+LSTM	CNN+ Bi-LSTM+PSO
Optimizer	Minimum batch gradient descent	Adam	Minimum batch gradient descent	Adam
Drop out	0.4	0.2	0.5	0.2
Accuracy (%)	92.4	91.9	91.45	99.2

7. CONCLUSION

The research discussed in this article shows that extracting features from the unprocessed ECG data is a reliable technique for raising classification accuracy using a BiLSTM network. The mean BiLSTM accuracy for the testing set was 85.3% for a single raw ECG signal. Using a twofold spectral ECG signal, the same statistic for the CNN and Bi-LSTM was 92.35%. Five different forms of cardiac dysfunctions were categorized to get at this conclusion. Because of the ECG signal's translation into spectral characteristics, Bi-LSTM and CNN perform better. CNNs were created to classify, but they are useless for predicting time series or signals because they lack memory. Signal memory and high performance are combined in Bi-LSTM, which recognizes images after converting a single input signal into a twofold spectral signal. As a result, the network trained in this manner has extraordinarily high quality. Using PSO, the neural network stack that was shown was optimized. Most of the CNN and LSTM-related parameters were considered, and the best values could be found. The improved network has an ECG signal categorization accuracy of 99.2%.

REFERENCES

[1]. Kim, Do-Gyun, and Jin-Young Choi. "Optimization of design parameters in LSTM model for predictive maintenance." Applied Sciences 11, no. 14 (2021): 6450.

[2]. Ruma, Jannatul Ferdous, Mohammed Sarfaraz Gani Adnan, Ashraf Dewan, and Rashedur M. Rahman. "Particle swarm optimization based LSTM networks for water level forecasting: A case study on Bangladesh river network." Results in Engineering 17 (2023): 100951.

[3]. Wang, Dianrui, Junhe Wan, Yue Shen, Ping Qin, and Bo He. "Hyperparameter Optimization for the LSTM Method of AUV Model Identification Based on Q-Learning." Journal of Marine Science and Engineering 10, no. 8 (2022): 1002.

[4]. Alamri, Nawaf Mohammad H., Michael Packianather, and Samuel Bigot. "Optimizing the parameters of long short-term memory networks using the bees algorithm." Applied Sciences 13, no. 4 (2023): 2536.

[5]. Kłosowski, Grzegorz, Tomasz Rymarczyk, Dariusz Wójcik, Stanisław Skowron, Tomasz Cieplak, and Przemysław Adamkiewicz. "The use of time-frequency moments as inputs of lstm network for ecg signal classification." Electronics 9, no. 9 (2020): 1452.

[6]. Zang, Junbin, Juliang Wang, Zhidong Zhang, Yongqiu Zheng, and Chenyang Xue. "Improving ECG Classification Performance by Using an Optimized One-Dimensional Residual Network Model." Applied Sciences 12, no. 24 (2022): 12957.

[7]. Shankar, Kathiresan, Sachin Kumar, Ashit Kumar Dutta, Ahmed Alkhayyat, Anwar Ja'afar Mohamad Jawad, Ali Hashim Abbas, and Yousif K. Yousif. "An automated hyperparameter tuning recurrent neural network model for fruit classification." Mathematics 10, no. 13 (2022): 2358.

[8]. Ali, Yasser A., Emad Mahrous Awwad, Muna Al-Razgan, and Ali Maarouf. "Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity." Processes 11, no. 2 (2023): 349.

[9]. Yu, Yaoxiang, and Min Zhang. "Control chart recognition based on the parallel model of CNN and LSTM with GA optimization." Expert Systems with Applications 185 (2021): 115689.

[10]. Śmigiel, Sandra. "ECG Classification Using Orthogonal Matching Pursuit and Machine Learning." Sensors 22, no. 13 (2022): 4960.

[11]. Blume, Sebastian, Tim Benedens, and Dieter Schramm. "Hyperparameter optimization techniques for designing software sensors based on artificial neural networks." Sensors 21, no. 24 (2021): 8435.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12]. Ahmed, Kanwal, Muhammad Imran Nadeem, Dun Li, Zhiyun Zheng, Yazeed Yasin Ghadi, Muhammad Assam, and Heba G. Mohamed. "Exploiting Stacked Autoencoders for Improved Sentiment Analysis." Applied Sciences 12, no. 23 (2022): 12380.

[13]. Houssein, Essam H., Ibrahim E. Ibrahim, Nabil Neggaz, Mahmoud Hassaballah, and Yaser M. Wazery. "An efficient ECG arrhythmia classification method based on Manta ray foraging optimization." Expert systems with applications 181 (2021): 115131.

[14]. Wasimuddin, Muhammad, Khaled Elleithy, Abdelshakour Abuzneid, Miad Faezipour, and Omar Abuzaghleh. "Multiclass ECG signal analysis using global average-based 2-D convolutional neural network modeling." Electronics 10, no. 2 (2021): 170.

[15]. Kim, Tae-Young, and Sung-Bae Cho. "Optimizing CNN-LSTM neural networks with PSO for anomalous query access control." Neurocomputing 456 (2021): 666-677.

[16]. Han, Hyojoon, Hyukho Kim, and Yangwoo Kim. "An efficient hyperparameter control method for a network intrusion detection system based on proximal policy optimization." Symmetry 14, no. 1 (2022): 161.

[17]. Wang, Hang, Min-jun Peng, Abiodun Ayodeji, Hong Xia, Xiao-kun Wang, and Zi-kang Li. "Advanced fault diagnosis method for nuclear power plant based on convolutional gated recurrent network and enhanced particle swarm optimization." Annals of Nuclear Energy 151 (2021): 107934.

[18]. Weicheng Xie, Wenting Chen, Linlin Shen, Jinming Duan, Meng Yang, Surrogate Network-based Sparseness Hyper-parameter Optimization for Deep Expression Recognition, Pattern Recognition (2020), doi: https://doi.org/10.1016/j.patcog.2020.107701.

[19]. Jalaeian Zaferani, Effat, Mohammad Teshnehlab, Amirreza Khodadadian, Clemens Heitzinger, Mansour Vali, Nima Noii, and Thomas Wick. "Hyper-Parameter Optimization of Stacked Asymmetric Auto-Encoders for Automatic Personality Traits Perception." Sensors 22, no. 16 (2022): 6206.

[20]. Awotunde, Joseph Bamidele, Agbotiname Lucky Imoize, Oluwafisayo Babatope Ayoade, Moses Kazeem Abiodun, Dinh-Thuan Do, Adão Silva, and Samarendra Nath Sur. "An Enhanced Hyper-Parameter Optimization of a Convolutional Neural Network Model for Leukemia Cancer Diagnosis in a Smart Healthcare System." Sensors 22, no. 24 (2022): 9689.

[21]. Li, Yaru, Yulai Zhang, and Yongping Cai. "A new hyper-parameter optimization method for power load forecast based on recurrent neural networks." Algorithms 14, no. 6 (2021): 163.

[22]. Kumari, Kirti, Jyoti Prakash Singh, Yogesh K. Dwivedi, and Nripendra P. Rana. "Multi-modal aggression identification using convolutional neural network and binary particle swarm optimization." Future Generation Computer Systems 118 (2021): 187-197.

[23]. Bhatia, Surbhi, Saroj Kumar Pandey, Ankit Kumar, and Asma Alshuhail. "Classification of Electrocardiogram Signals Based on Hybrid Deep Learning Models." Sustainability 14, no. 24 (2022): 16572.

[24]. Zulfiqar, M., Kelum AA Gamage, M. Kamran, and M. B. Rasheed. "Hyperparameter Optimization of Bayesian Neural Network Using Bayesian Optimization and Intelligent Feature Engineering for Load Forecasting." Sensors 22, no. 12 (2022): 4446.

[25]. Ismail, Ali Rida, Slavisa Jovanovic, Naeem Ramzan, and Hassan Rabah. "ECG Classification Using an Optimal Temporal Convolutional Network for Remote Health Monitoring." Sensors 23, no. 3 (2023): 1697.

[26]. Habtemariam, Ejigu Tefera, Kula Kekeba, María Martínez-Ballesteros, and Francisco Martínez-Álvarez. "A Bayesian Optimization-Based LSTM Model for Wind Power Forecasting in the Adama District, Ethiopia." Energies 16, no. 5 (2023): 2317.