ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

FEATURE OPTIMIZATION-BASED HEART DISEASE PREDICTION USING DEEP LEARNING ALGORITHM

Vikash Kumar Singh[1] and Dinesh Kumar Sahu[2]
[1]Department of Computer Science & Engg., SRK University, Bhopal, India
[2]Department of Computer Science & Engg., SRK University, Bhopal, India
[1]vikashpiet@gmail.com and [2]drdineshksahu@gmail.com

ABSTRACT
The multitude of features present in heart disease datasets often hampers accurate prediction outcomes. This decline in prediction performance is attributed to the inclusion of redundant features and suboptimal feature selection methods. In response, this paper introduces a hybrid learning approach that combines a convolutional neural network (CNN) with the firefly optimization algorithm. The hybrid learning approach aims to mitigate the adverse effects of redundant features and enhance the feature space mapping for the CNN model. To evaluate the performance of the proposed algorithms, a 10-fold cross-validation testing method is employed. Additionally, hyperparameter tuning is conducted, specifically focusing on the number of fully connected layers, set at 10 in the feature vector. Results indicate that the firefly optimization algorithm achieves an accuracy of 95.148% when utilizing the full set of attributes. However, the highest accuracy of 96.46% is attained when employing the optimal attribute set obtained through optimized attribute evaluation. Moreover, employing the deep learning classifier bagging with firefly optimization yields the highest Receiver Operating Characteristic (ROC) area of 0.91, utilizing both the full and optimal attribute sets derived from the firefly optimization process. Overall, the hybrid learning classifier emerges as the most effective prediction method compared to alternative techniques, showcasing a 3.25% accuracy improvement through the tuning of hyperparameter, particularly the fully connected layer set at 10, in conjunction with the optimized feature set. This analysis of state-of-the-art results demonstrates superior performance over conventional CNN, LSTM, and RNN models.

Keywords: - Heart Disease, Feature Optimization, FA, Deep Learning, CNN, Hybrid Learning.

INTRODUCTION
According to the World Health Organization (WHO), heart diseases rank among the leading causes of death globally. Identifying cardiovascular diseases (CVD) poses significant challenges due to various contributing factors, including high blood pressure, cholesterol levels, diabetes, abnormal pulse rate, and others [1,2]. Moreover, CVD symptoms may manifest differently based on gender; for instance, males may experience chest pain, while females might exhibit additional symptoms like chest discomfort, nausea, extreme fatigue, and shortness of breath [3,4]. Although researchers have explored numerous techniques for predicting heart diseases, early-stage prediction remains inefficient due to complexities, execution time constraints, and challenges related to accuracy. Recognizing the importance of timely diagnosis and treatment, advancements in predictive methodologies hold the potential to save many lives [5]. Diagnosing heart health issues poses challenges due to various factors like blood pressure, cholesterol levels, and creatine levels. Authors in [6] conducted an analysis to identify contributing factors to heart disease, highlighting controllable elements such as alcohol usage, smoking, diabetes, high cholesterol, and limited physical activity. In the modern healthcare landscape, electronic health records (EHRs) play a pivotal role in both clinical practice and research endeavors[7]. However, reliance solely on physical examinations may introduce errors, which could have grave consequences in the context of heart disease diagnosis. Leveraging machine learning-based expert systems has proven effective in diagnosing cardiovascular diseases, ultimately contributing to a reduction in mortality rates. Automatic detection of heart disease presents several challenges, primarily stemming from the diverse array of factors influencing the accuracy of detection devices[8,9]. The efficacy of these devices heavily relies on the efficiency of algorithms, particularly those employed in machine learning and deep learning. Machine learning encompasses various algorithms, including unsupervised and supervised learning methods. Within supervised learning, regression algorithms like

Support Vector Regression (SVR) and other statistical regression functions are commonly utilized for heart disease detection [11,12]. Alternatively, classification algorithms such as Support Vector Machine (SVM), Random Forest, and ensemble-based approaches offer effective means of classification. However, machine learning's limitations are being addressed through the adoption of deep learning algorithms, which significantly enhance detection capacity. Deep learning algorithms excel in feature selection and optimization, crucial components of effective detection [13]. Feature selection methods, including rank-based and random-based algorithms, play a vital role in enhancing detection accuracy by identifying relevant features. Additionally, feature optimization processes contribute significantly to heart disease detection. Recent advancements have introduced algorithms for feature optimization, such as swarm intelligence-based approaches like Ant Colony Optimization and Particle Swarm Optimization, as well as meta-heuristic-based functions. These algorithms further refine feature selection and optimization processes, thereby improving the accuracy and efficiency of heart disease detection systems. The paper introduces a novel approach to optimizing features in medical data, specifically targeting the reduction of redundant features in the context of heart disease detection. The proposed method utilizes the firefly algorithm, a meta-heuristic function inspired by swarm intelligence, to enhance the convergence of medical data and identify optimal features relevant to heart disease detection [14,15]. By leveraging the firefly algorithm, the proposed technique aims to improve the accuracy and efficiency of disease detection by refining the feature selection process. An interesting aspect of this approach is its integration with convolutional neural networks (CNNs), a powerful deep learning technique commonly used in image recognition and classification tasks. By incorporating the firefly algorithm into the pooling process, the sample size is increased, facilitating more comprehensive analysis and potentially improving the detection accuracy of heart disease. This paper presents an innovative method that combines meta-heuristic optimization with deep learning techniques to address feature redundancy in medical data, with a specific focus on enhancing heart disease detection. This approach has the potential to contribute significantly to the advancement of medical diagnostics by improving the accuracy and efficiency of disease detection algorithms. The main contributions of this paper is:

- Accelerates the detection ratio of heart disease on early stage of disease.
- Finding a novel methodology for classification of heart disease based on feature optimization
- Compare proposed algorithm with existing algorithm as detection ratio on different parameters.
- Improves the capacity of convolutional neural network and reduces complexity of CNN networks

The rest of the article explores, as in Section II, related work in the area of heart disease detection; in Section III, proposed methodology based on firefly algorithms and CNN; in Section IV, experimental analysis and datasets; in Section V, results and discussion; and in Section VI, a paper with future direction

II. RELATED WORK

Heart disease remains a prevalent and serious health concern globally, often linked to lifestyle factors. Its impact on mortality rates is significant, with approximately one in three individuals affected by heart disease-related issues. Accurately predicting heart disease poses a major challenge due to the multitude of factors involved in the prediction process. However, recent efforts from various researchers have focused on developing methodologies aimed at enhancing the accuracy of heart disease prediction through feature selection and optimization techniques. This section delves into the methodologies proposed by these authors, aiming to shed light on the advancements made in this field and their potential implications for improving healthcare outcomes. In [1], utilized a decision tree classifier in conjunction with backward feature selection to achieve a maximum classification accuracy of 88.52%, with precision at 91.30%, sensitivity at 80.76%, and an f-measure of 85.71%. In contrast to healthcare informatics [2], the suggested approach extends its applicability to diverse fields such as sports analytics, bioinformatics, and financial analysis. The Naive Bayes model and XG Boost significantly differ at a 95% confidence interval when two characteristics are combined. In [3], to address class imbalance, the Synthetic Minority Oversampling Method (SMOTE) significantly improved predictive performance. The combination of

Extra Trees Classifier (ETC) with SMOTE achieved an impressive accuracy of 92.62% in predicting the survival of heart patients. In [4], employing a machine-learning-based method, the author successfully identified highly correlated features in patient electronic clinical data. Both datasets achieved remarkable accuracy ratings, with the first dataset reaching 100% and the second dataset at 97.68%, along with perfect precision, recall, and F1 values. In [5], using various well-known machine learning algorithms, including logistic regression, support vector machines, and decision trees, the author demonstrated that fewer features can lead to a higher prediction accuracy of approximately 87%. In [6], comparing different methods, the SMO classifier emerged as the most accurate prediction method, while setting the hyperparameter "k" to 9 in IBk improved performance by 8.25% using the chi-squared attribute set. In[7], Employing data collection, pre-processing, and transformation techniques, the suggested model achieved superior outcomes through wrapper feature selection methods compared to other filter feature selection methods. In [8], utilising the LASSO technique for feature selection and Gaussian Naive Bayes for classification, the proposed system showcased impressive results with an accuracy score of 94.92% after utilizing feature selection techniques like Lasso and Ridge regression. In[9] Based on the CDF-DI monitoring system, a machine learning system was developed for wearable IoT-enabled smart healthcare, achieving a prediction accuracy of 96% for the survival status of heart failure patients using the winning RF method. In [10], suggested approaches were verified on 23 medical datasets, with LBMFO-R1 outperforming other techniques on 39% of the datasets and LBMFO-R2 outperforming others on three datasets. [11] Techniques proposed in the literature can be applied to classify heart diseases using patient datasets from various sources, with model-assembling techniques aiming to improve classification accuracy. In [12], using an embedded method for analytical attribute selection, the suggested SVM HFWE-FS algorithm achieved accuracy results of 93.33%, surpassing other algorithms by significant margins. In [13], a fuzzy logic strategy employing greedy hill climbing feature selection methods was developed for diabetes classification, addressing issues with diagnosis. In [14], using the Improved Squirrel search algorithm and rank aggregation, a hybrid feature selection model for biological datasets was proposed, achieving notable information-based measure values across attributes. In [15], employing six feature selection approaches and four prediction algorithms on glycaemic-related biological characteristics, Random Forest emerged as the best-performing algorithm, providing the best average performance over multiple prediction horizons. In [16], a novel approach utilizing heart rate variability features in machine learning algorithms achieved high accuracy rates for predicting cardiac health, with CNN, SVM, and overall performance reaching notable levels. In [17], including demographic, plasma biomarkers, and echocardiography data, Ada Boost properly classified patients with LVH and DHF, demonstrating high accuracy rates with limited data inputs. In [18], the suggested model showcased increased accuracy and reduced computational load compared to other techniques, indicating its superiority in predicting cardiac diseases. In [19], employing various machine learning classifiers, the proposed technique achieved high accuracy rates in categorizing animal status into distinct groups, highlighting its effectiveness in classification tasks. In [20], offering a machine-learning tool for precise estimation of particulate matter concentration, the suggested technique, particularly using random forest and GBM, achieved exceptional accuracy and AUC scores in classification. In [21], using SMOTE and ensemble classifiers, the proposed model achieved high accuracy in the early detection of heart illness, showcasing the effectiveness of the approach in healthcare applications. In [22], by combining aggregation methods and filter feature selection strategies, the proposed ensembles, evaluated with various classifiers, demonstrated improved performance in dataset classification. In [23] Utilizing existing datasets in healthcare, the study aimed to assist clinicians in making accurate decisions for patient management, with 90% of the dataset used for training and 10% for testing. In [24], employing MSU as a selection criterion with algorithms like rapid correlation and relief, followed by penalty-based regression, the proposed method achieved satisfactory RMSE for blood pressure prediction. Utilising ensemble models and machine learning algorithms, the study predicted mortality rates in specific patient cohorts, providing valuable insights for clinical decision-making. In[26] Demonstrating high accuracy in predicting heart illness, the proposed technique utilizing data mining algorithms offers potential for effective healthcare management. In [27], building on advanced algorithms, the proposed OBEFS technique offers promising advancements in deep learning for engineering and medicine, with future work aimed at further extensions. In [28], using the Cleveland UCI dataset, the study highlighted the utility of Choose Best in enhancing heart disease prediction, with hybridized classifiers achieving high accuracy rates. In [29], for the leukaemia dataset, SVM and KNN models achieved remarkable accuracy with a limited number of features, indicating their efficacy in disease classification. In [30], various metrics were used to evaluate machine learning models' performance, with the proposed model achieving notable accuracy, recall, cross-validation mean, and AUC scores. In [31], employing a strategy to remove non-informative characteristics from the ResNet50 model improved pneumonia diagnosis accuracy, with implementation carried out using PyTorch. In [32] The study described a model for finding depression and used ensembles and feature selection algorithms for classification. It also tested performance across several splits of training data. In [33], developing a risk stratification tool using machine learning algorithms, the study accurately predicted mortality risk levels, demonstrating clinical utility for patient management. In [34], employing machine learning classifiers, the study effectively differentiated individuals using original and SMOTE data, achieving high performance metrics including accuracy, sensitivity, specificity, and AUC scores. In [35], utilising nonnegative matrix factorization in feature-based rejection of PCG signals, the study compared feature-based and DL classifiers' performance, highlighting challenges and costs associated with each approach. In [36], employing RF, SVC, KNN, LGBM, Bagging, and AdaBoost classifiers, the study accurately predicted the death rate of heart failure patients, demonstrating the model's effectiveness and capability.

III. METHODOLOGY

This section describes the proposed methodology for heart disease prediction based on a hybrid learning approach. Hybrid learning approach based on feature optimization and deep learning algorithms. The process of feature optimization employed the Firefly algorithm. The Firefly algorithm reduces the redundancy of features in heart disease data. The process of optimization improves the feature mapping of CNN networks. The employed CNN networks are 10-layer networks along with ReLU activation and apply max-poling feature selection to the input feature matrix. The processing of the algorithm is presented in Figure 1. The methodology section is described in three sections: the first section describes the firefly algorithm (FA), the second section describes the CNN algorithm, and the third section describes the proposed methodology for heart disease prediction.

1st SECTION

FIREFLY ALGORITHM (FA)

The firefly algorithm (FA), proposed by Yang, belongs to the family of swarm intelligence techniques. It operates as a stochastic, meta-heuristic algorithm designed to locate global optimal solutions within search spaces, addressing optimization problems including those categorized as NP-hard. Inspired by the natural phenomenon of fireflies emitting light, the FA mirrors the physical formula governing light intensity observed in these insects. The fundamental concept of the firefly algorithm lies in interpreting the characteristics of light intensity, where: All fireflies are considered unisex, with any two fireflies potentially exhibiting attraction towards each other. The level of attraction between fireflies is directly proportional to their respective light intensities. Fireflies with lower light intensity tend to move towards those emitting higher light intensity. In the absence of a firefly with higher light intensity nearby, a firefly will randomly explore the search space. The light intensity emitted by a firefly is determined by a fitness function, which serves as a measure of its effectiveness in solving the optimization problem at hand. Within the firefly algorithm, three key aspects are addressed[37,38]:

The core component of firefly algorithm is attraction function and movement of fireflies.

Attraction function

The attractiveness function β can be monotonically decreasing function such as

$$\beta = \beta_0 X\, e^{-\gamma r2_{ij}} \ldots\ldots\ldots\ldots\ldots\ldots. (1)$$

Here r is distance between two fireflies I and j respectively. Euclidean distance is applied in this work. β_0 is an attraction parameter that attractiveness at r=0 and γ is a light absorption coefficient.

Movement of fireflies

The movement of fireflies depends on lower light intensity to higher light intensity as equation (2)

$$X_i(t+1) = X_{i\,(t)} + \beta\left(X_{i\,(t)} - X_{j\,(t)}\right) + \alpha\left(rand - \frac{1}{2}\right) \dots\dots\dots (2)$$

where Xi(t) and Xj(t) are positions of firefly i with lower light intensity and firefly j with higher intensity at time t respectively, a is a random parameter which determines randomly behavior of movement of fireflies, rand is a random number generator uniformly distributed in [0,1].

2nd SECTION

Deep learning algorithm

Convolutional Neural Networks (CNNs) are a class of deep neural networks that are primarily used for analyzing visual imagery. They have been highly successful in tasks such as image recognition, object detection, and image segmentation. Here's an overview of CNNs: In a convolutional layer, a set of learnable filters (also known as kernels) is applied to the input image. These filters detect patterns such as edges, textures, or more complex features. The filters slide (or convolve) across the input image, computing dot products between the filter weights and the input at each position. This operation produces feature maps that highlight the presence of different features in the input. After convolutional layers, pooling layers are often used to reduce the spatial dimensions of the feature maps while retaining important information. Common pooling operations include max-pooling, which selects the maximum value within a region of the feature map, and average pooling, which computes the average value. Pooling helps make the representations learned by the network more robust to variations in the input, while also reducing computational complexity. Activation functions like Rectified Linear Units (ReLU) are typically applied after convolutional and pooling operations to introduce non-linearity into the network. CNNs often include one or more fully connected layers. These layers connect every neuron in one layer to every neuron in the next layer, enabling the network to learn high-level features and make predictions based on the learned features. Fully connected layers are typically used in the final stages of the network to map the learned features to specific output classes. CNNs are trained using supervised learning, where labeled training data is used to adjust the network's parameters (weights and biases) through backpropagation and gradient descent. During training, the network learns to minimize a loss function that measures the difference between its predictions and the true labels.

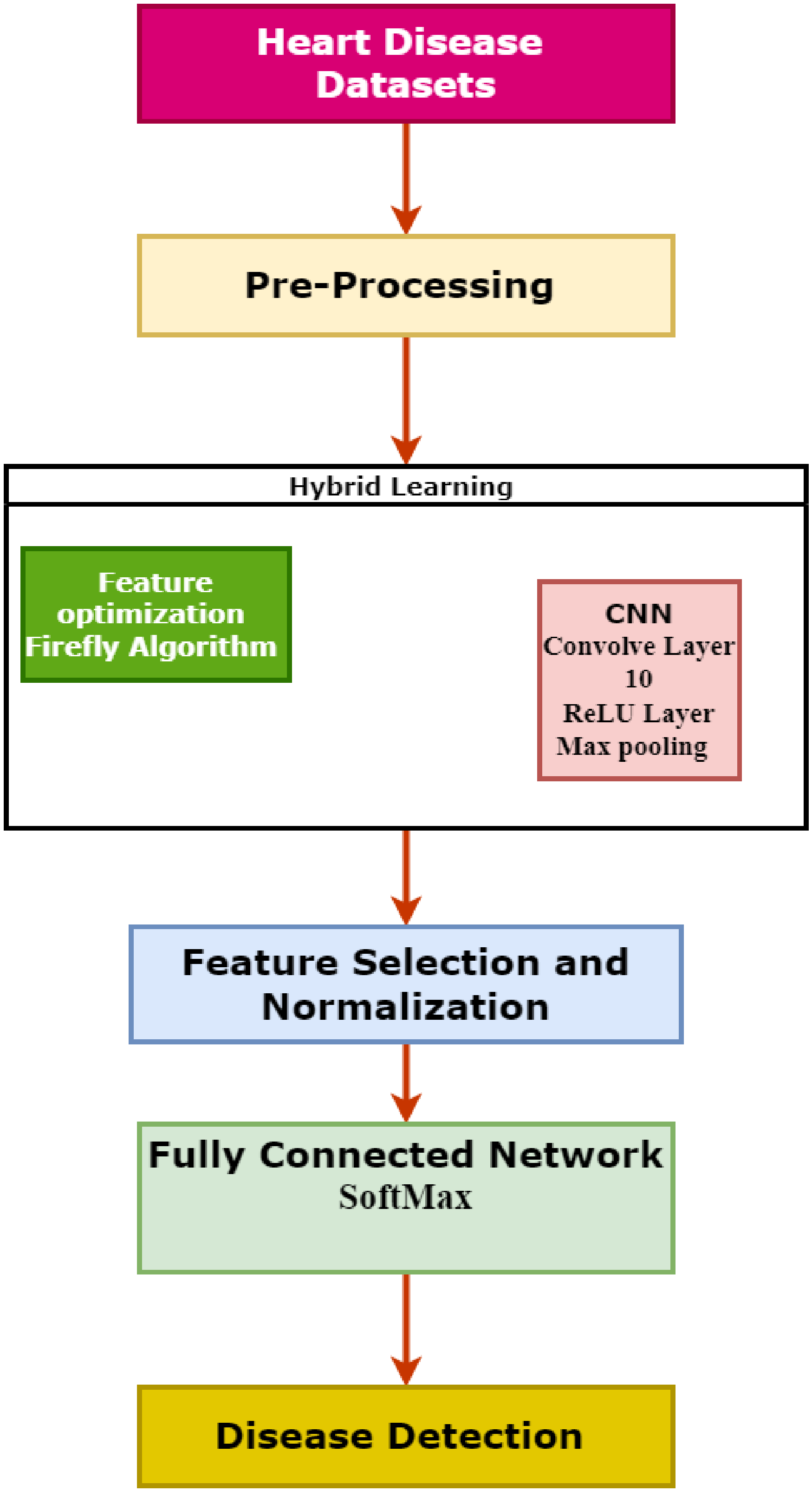

Figure 1 proposed model of hybrid learning algorithm for heart disease predication

3rd SECTION

Proposed Methodology

The proposed hybrid learning is resolving issue of feature mapping. the processing of model mapped feature of dataset of heart disease, and search all features approach as non-linear. The proposed model employed rectified linear active function for the training of the networks. The ReLU function describes as

$$f_{ReLU\,(x)} = max(0, x)\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Here x defines the argument of the function. Also consider O and xin to be the output of the network and the input of the learning , the expression can write as

$$O = f(x_{in}, w) = f^{(n-1)}(f^{(n-2)}(\ldots f^{1}(x_{in})))\ldots\ldots\ldots\ldots(4)$$

Where n and W are defined as the number of layers of the network and the weight of the network respectively.

Processing of CNN

1. The input of network is featuring vector F_m and precoding matrix P_m
2. After the processing of feature vector and other parameters employe firefly algorithm. the employed firefly algorithm removes local interference and improve the process of feature mapping.
3. The layer of convolutional network 64 different 4 X 4X1 filters and 1 stride produces feature map follows ReLU.
4. Employed the feature constraints for next layer for the selection of distant features. The fitness of selection is $D\sum_{m=1}^{M}|p_m| \leq F_m$
5. The layer of fully connected to process of feature factors. The total numbers of neurons is 256.
6. The output of FC layer is proceeded in 3X3X1 filter of convolutional layers filter and stride 1
7. The maximum pooling of network filter is 3X3X1
8. The total number of convolutional layers is 10
9. Finally estimates optimal preceded matrix P_m
10. Predict features as normal and disease.
11. Exit
12. Table 1 parameters value of CNN network

Algorithm	Parameter	Values
CNN	Layer	10
	Activation function	ReLU
	Optimizer	FA
	Loss	Binary cross entropy
	Epochs	200
	Validation _split	2
	Batch size	500

IV. Experimental Analysis

To evaluate the performance of proposed algorithms using MATLAB software version R2018a, the system configuration comprises an Intel Core i7 processor, 16GB RAM, and the Windows 10 operating system. While

MATLAB offers built-in support for certain classification algorithms recurrent neural network (RNN), other classifier functions are defined and programmed using function files, which are compiled with library files. For the evaluation process, UCI machine learning datasets are employed. The data processing involves utilizing 10-fold cross-validation for prediction and assessing parameters like accuracy, sensitivity, and specificity. The prediction of disease categorized into two class, class-1 and class-2. The representation of class-1 is abnormal and class-2 is normal. The description of parameters as [28,29,30,31]

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \dots\dots\dots\dots\dots (4)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \dots\dots\dots\dots\dots (5)$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \dots\dots\dots\dots\dots (6)$$

$$F1 = 2x\frac{Pricision\ X\ Recall}{Pricision+Recall} \times 100 \dots\dots\dots\dots\dots (7)$$

TP: True Positive

TN: True Negative

FP: False Positive

FN: False Negative

Dataset Descriptions:

Hungarian: This dataset, collected at the Hungarian Institute of Cardiology in Budapest by Andras Janosi, consists of ten features. Initially comprising 294 samples, 34 were discarded due to missing values, leaving 262 records. These records are divided into 62.21% representing healthy subjects and 37.78% with heart disease.

Cleveland: With a total of 303 instances, the Cleveland dataset contains 76 attributes, of which only 14 are considered. It serves as another dataset for heart disease prediction analysis.

Z-Alizadeh Sani: This dataset includes 270 instances and 13 attributes. Patients are classified into two categories: CAD or Normal. A patient is categorized as CAD if their diameter narrowing is greater than or equal to 50%, otherwise labeled as Normal.

Statlog: This dataset, comprising 270 instances and 13 attributes, features no missing attribute values.

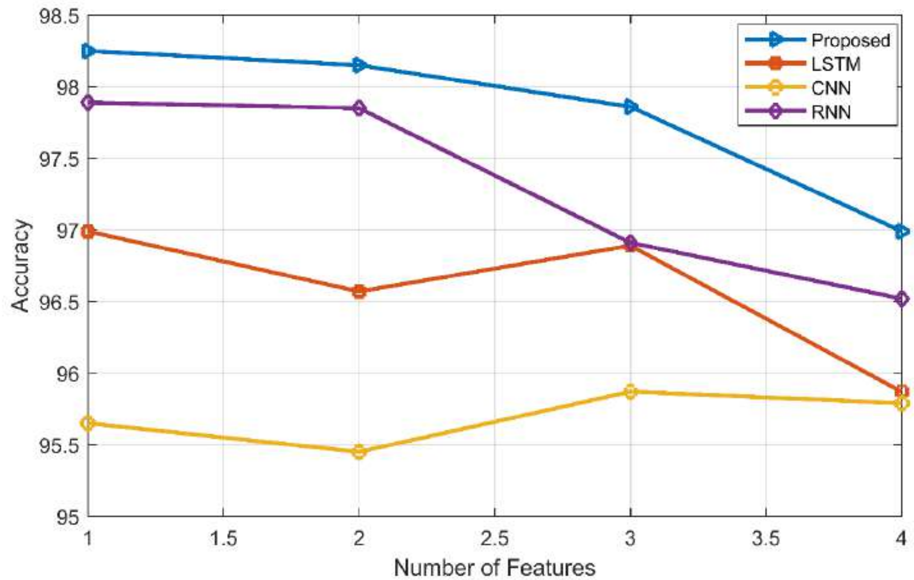

Figure: 2 Performance analysis of accuracy for Cleveland dataset.

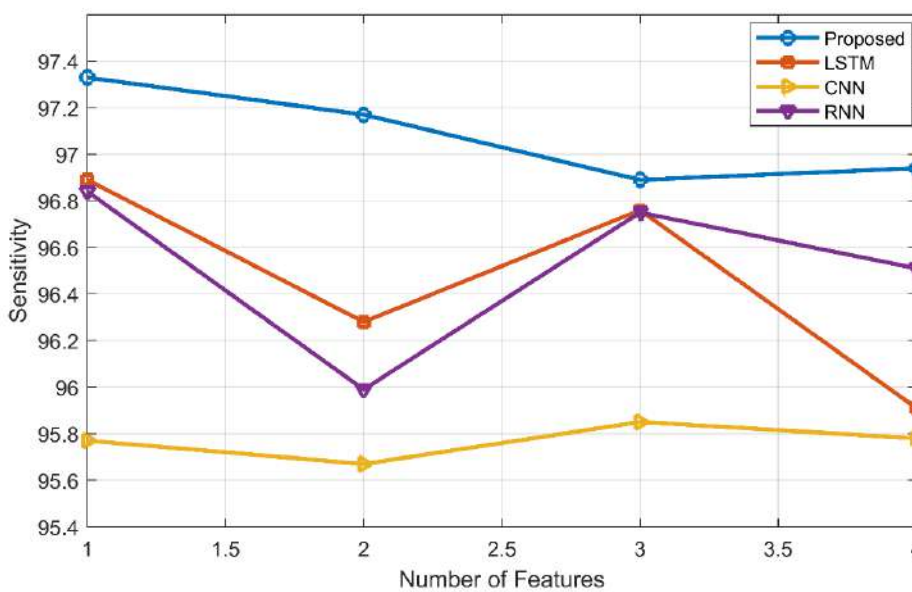

Figure: 3 Performance analysis of sensitivity for Cleveland dataset

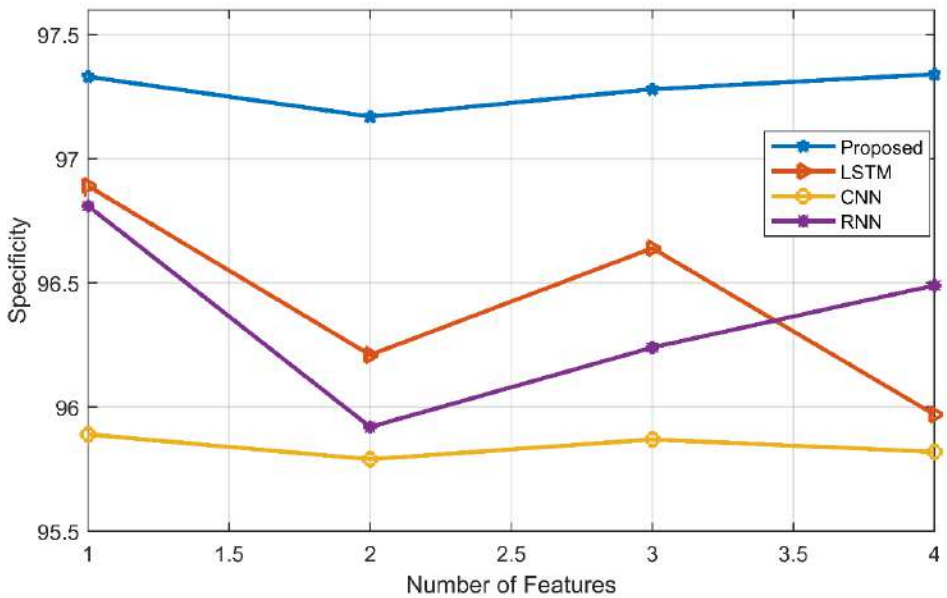

Figure: 4 Performance analysis of Specificity for Cleveland dataset.

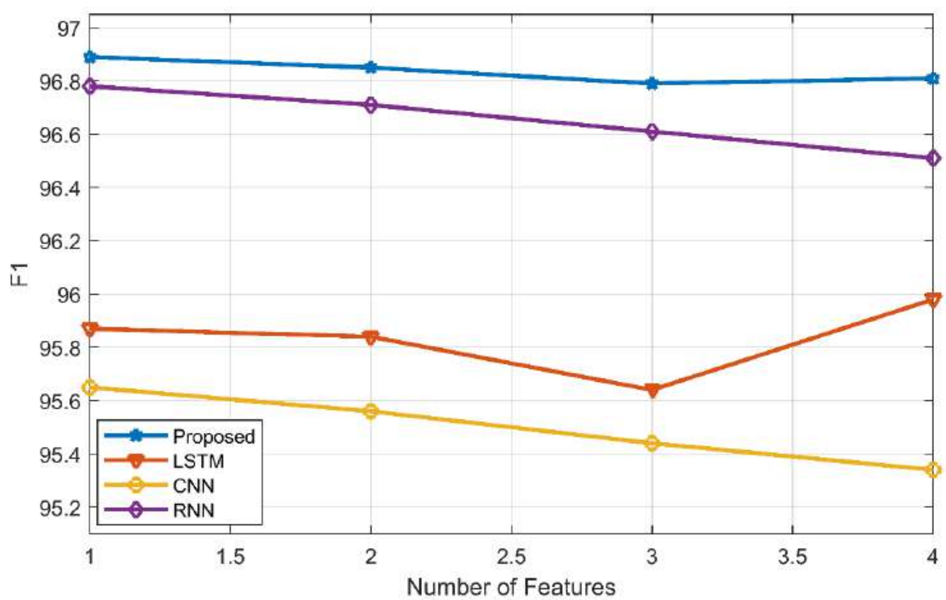

Figure: 5 Performance of analysis F1 for Cleveland dataset.

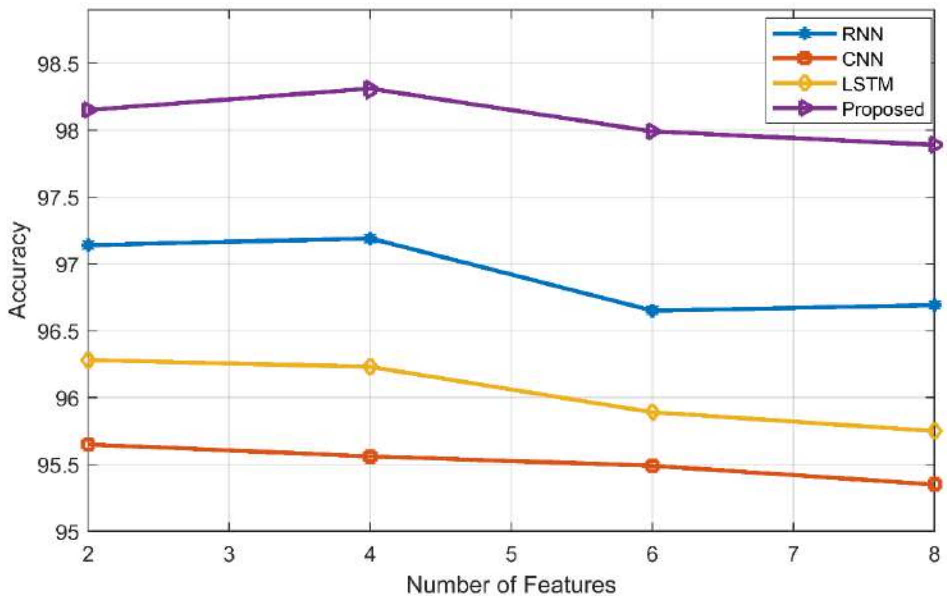

Figure: 6 Performance of analysis Accuracy for Hungarian dataset.

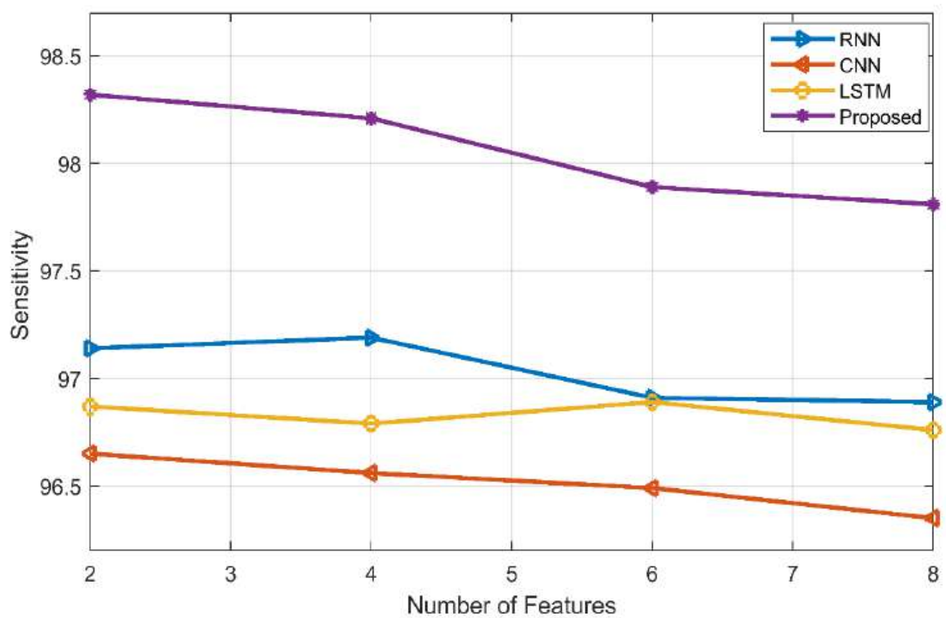

Figure: 7 Performance analysis of Sensitivity for Hungarian dataset.

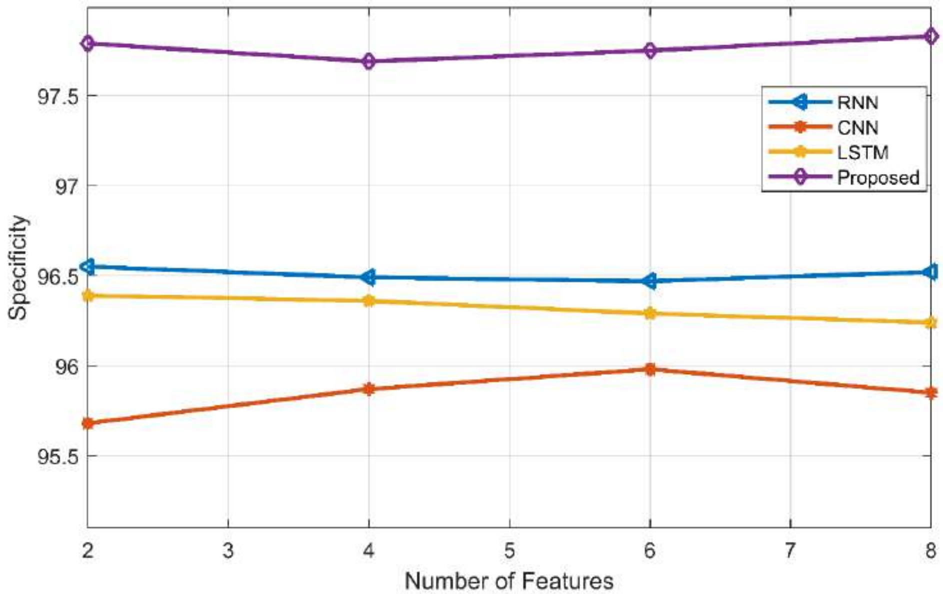

Figure: 8 Performance analysis of Specificity for Hungarian dataset.

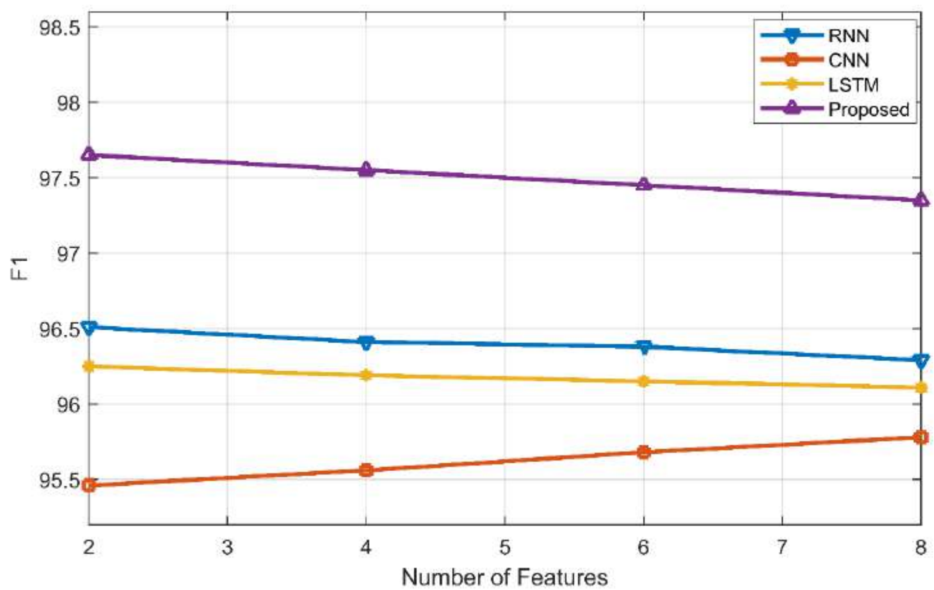

Figure: 9 Performance analysis F1 for Hungarian dataset.

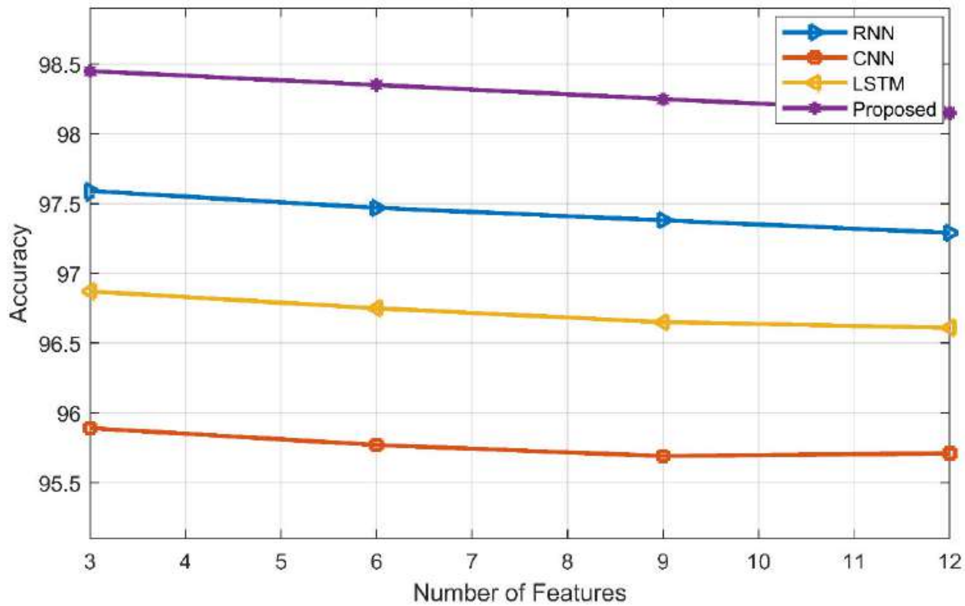

Figure: 10 Performance analysis of accuracy for Sani dataset.

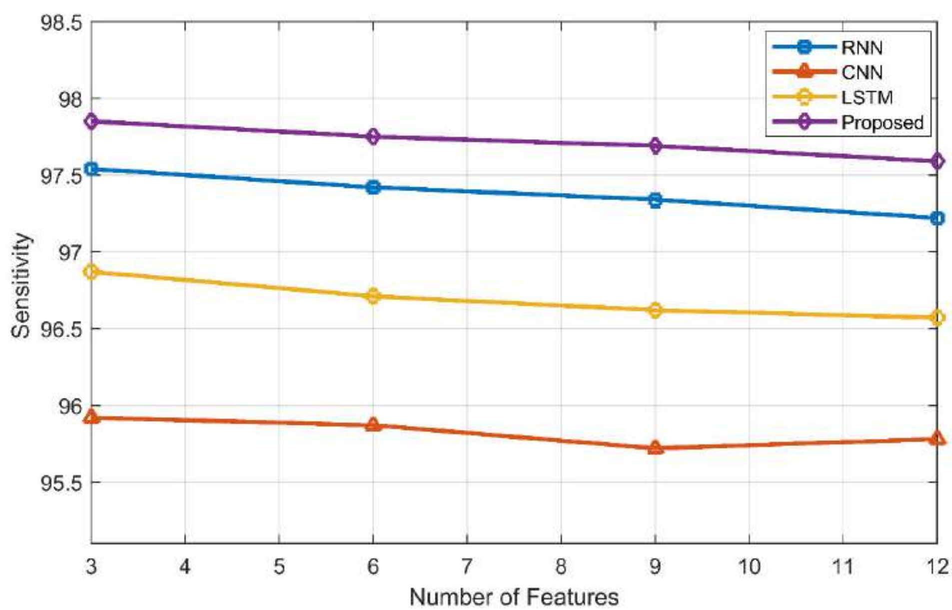

Figure: 11 performance analysis of sensitivity for Sani dataset.

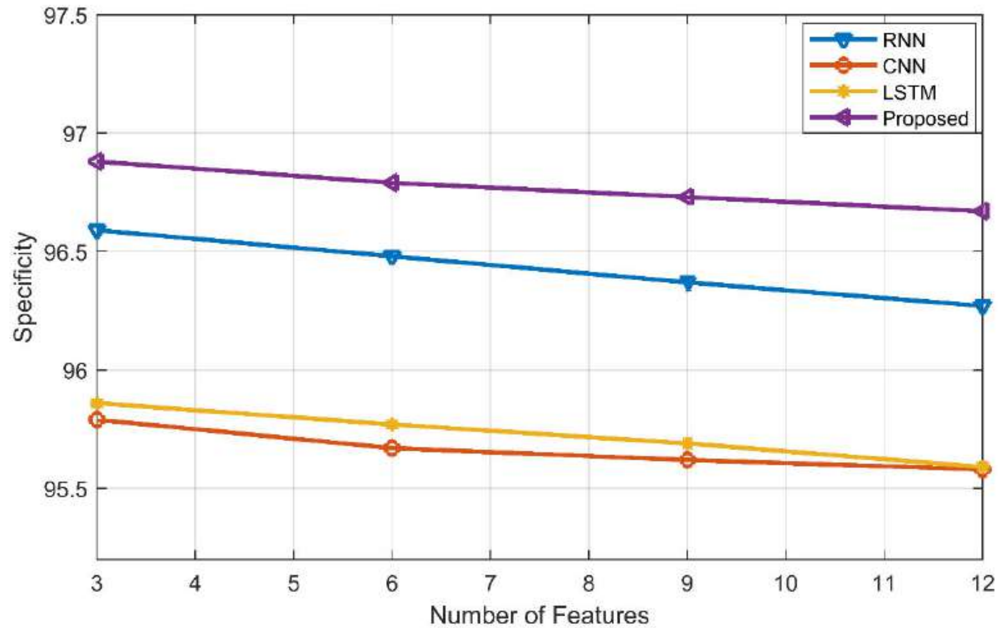

Figure: 12 Performance analysis specificity for Sani dataset.

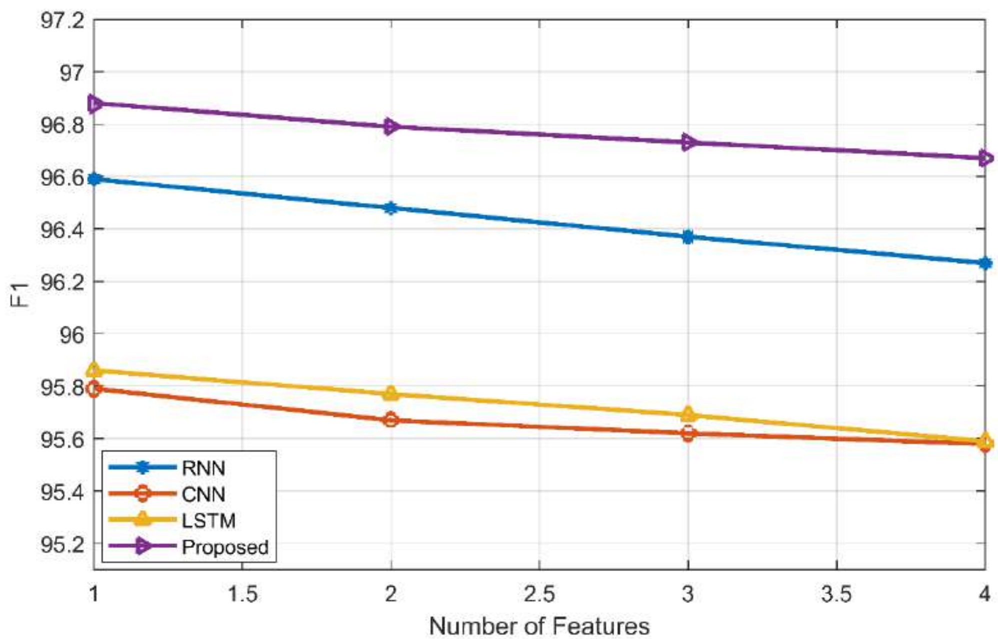

Figure: 13 Performance analysis F1 for Sani dataset.

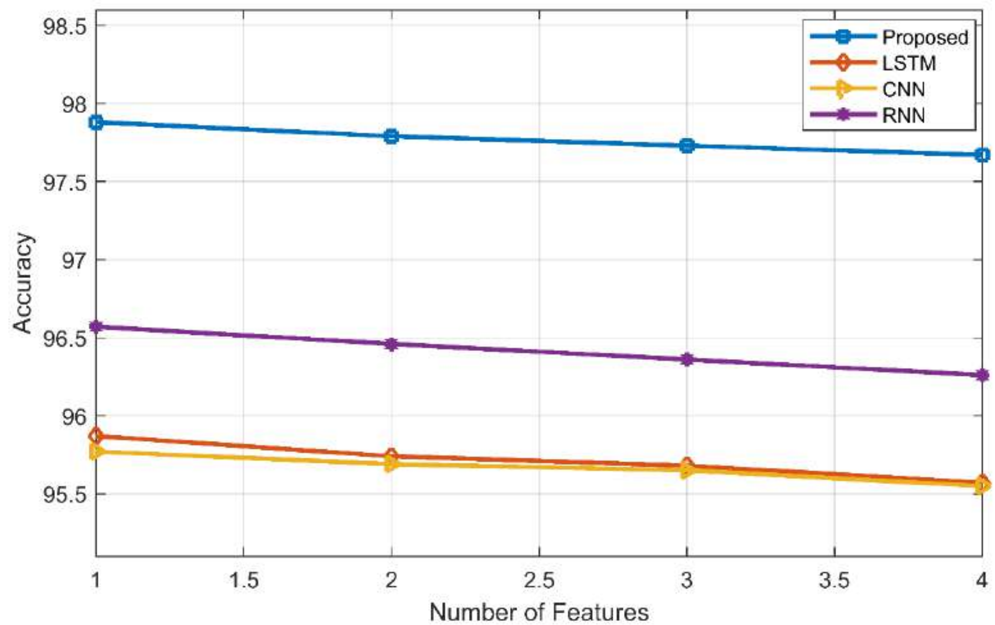

Figure: 14 Performance analysis of accuracy for Statlog dataset.

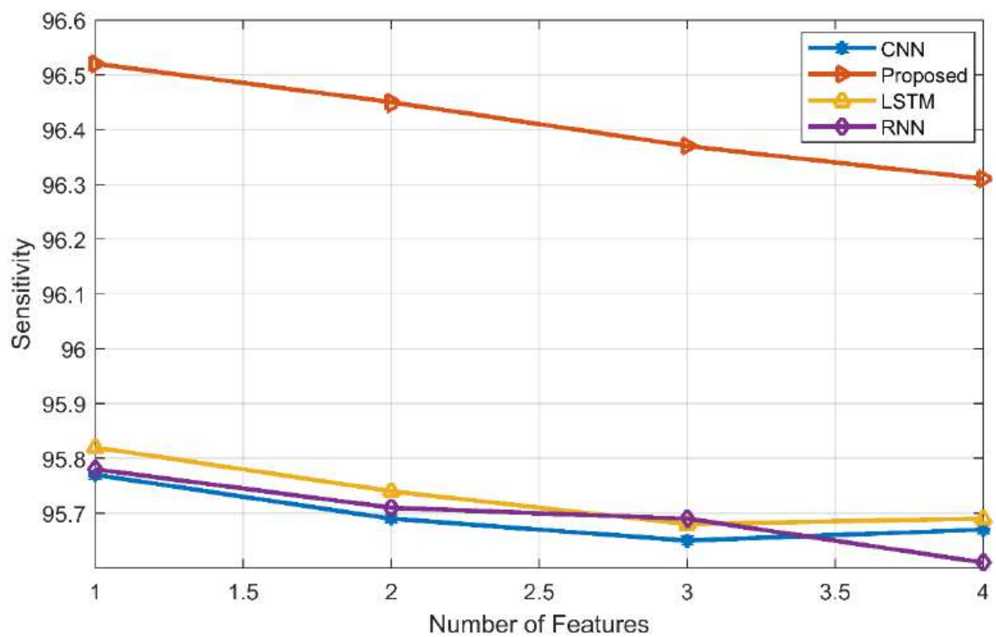

Figure: 15 Performance analysis for sensitivity for Statlog dataset.

V. RESULTS AND DISCUSSION

This section scrabble into an examination of the performance of a proposed algorithm in comparison to existing deep learning algorithms like RNN (Recurrent Neural Networks), LSTM (Long Short-Term Memory), and CNN (Convolutional Neural Networks). The evaluation encompasses four distinct datasets: Cleveland, Hungarian, Statlog, and Sani datasets. Performance evaluation metrics include accuracy, sensitivity, specificity, and the F1 score. The results of the performance evaluation are presented graphically, likely in figures 2 to 17. Figures may include plots or charts illustrating the performance metrics for each algorithm across the different datasets. Comparative analysis of the algorithms' performance across various metrics and datasets may be provided to identify strengths and weaknesses.

Table 2: Result analysis of CNN, RNN, LSTM, and Proposed for Cleveland dataset.

Dataset	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)
Cleveland	CNN	95.65	95.89	95.89	95.65
	RNN	97.89	96.84	96.81	96.78
	LSTM	96.99	96.89	96.89	95.87
	Proposed	**98.25**	**97.33**	**97.33**	**96.89**
	CNN	95.45	95.67	95.79	95.56
	RNN	97.85	95.99	95.92	96.71
	LSTM	96.57	96.28	96.21	95.84
	Proposed	**98.15**	**97.17**	**97.17**	**96.85**
	CNN	94.87	95.85	95.87	95.44
	RNN	96.91	96.75	96.24	96.61
	LSTM	96.89	96.76	96.64	95.64
	Proposed	**97.86**	**96.89**	**97.28**	**96.79**
	CNN	95.79	95.78	95.82	95.34
	RNN	96.52	96.51	96.49	96.51
	LSTM	95.87	95.91	95.97	95.98
	Proposed	**96.99**	**96.94**	**97.34**	**96.81**

Table 3: Result analysis of CNN, RNN, LSTM, and Proposed for Hungarian dataset.

Dataset	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)
Hungarian	CNN	95.65	96.65	95.68	95.46
	RNN	97.14	97.14	96.55	96.51
	LSTM	96.28	96.87	96.39	96.25
	Proposed	**98.15**	**98.32**	**97.79**	**97.65**
	CNN	95.56	96.56	95.87	95.56
	RNN	97.19	97.19	96.49	96.41
	LSTM	96.23	96.79	97.69	96.19
	Proposed	**98.31**	**98.32**	**97.69**	**97.55**
	CNN	95.49	96.49	95.98	95.68
	RNN	96.65	96.91	96.47	96.38
	LSTM	95.89	96.89	96.29	96.15
	Proposed	**97.99**	**98.21**	**97.75**	**97.45**
	CNN	95.35	96.35	98.85	95.78
	RNN	96.69	96.89	96.52	96.29
	LSTM	95.75	96.76	96.24	96.11
	Proposed	**97.89**	**97.81**	**97.83**	**97.35**

Table 4: Result analysis of CNN, RNN, LSTM, and Proposed for Sani datasets.

Dataset	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)
	CNN	95.89	95.92	95.79	95.78

Sani	RNN	97.59	97.54	96.59	96.55
	LSTM	96.87	96.87	95.89	95.85
	Proposed	**98.45**	**97.85**	**96.88**	**96.87**
	CNN	95.77	95.87	95.67	95.66
	RNN	97.47	97.42	96.48	96.49
	LSTM	96.75	96.71	95.77	95.76
	Proposed	**98.35**	**97.75**	**96.79**	**96.78**
	CNN	95.96	95.72	95.62	95.61
	RNN	97.38	97.34	96.37	96.38
	LSTM	96.65	96.62	96.69	95.68
	Proposed	**98.25**	**97.69**	**96.73**	**96.74**
	CNN	95.71	95.78	95.58	95.59
	RNN	97.29	97.22	96.27	95.26
	LSTM	96.61	96.57	95.59	95.58
	Proposed	**98.15**	**97.59**	**96.67**	**96.66**

Table 5: Result analysis of CNN, RNN, LSTM, and Proposed for Statlog datasets.

Dataset	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)
Statlog	CNN	95.77	95.78	95.72	95.89
	RNN	96.57	95.78	96.27	96.78
	LSTM	95.87	95.82	95.81	96.56
	Proposed	**97.88**	**96.52**	**97.28**	**97.59**
	CNN	95.69	97.69	95.66	95.82
	RNN	96.46	95.71	96.23	96.69
	LSTM	95.74	95.74	95.73	96.47
	Proposed	**97.79**	**96.45**	**97.24**	**97.68**
	CNN	95.65	95.65	95.59	95.78
	RNN	96.36	95.69	96.15	96.72
	LSTM	95.68	95.68	95.67	96.38
	Proposed	**97.73**	**96.37**	**97.21**	**97.75**
	CNN	95.55	95.67	95.67	95.73
	RNN	96.26	95.61	95.98	96.65
	LSTM	95.57	95.69	95.67	96.29
	Proposed	**97.67**	**96.37**	**96.99**	**97.71**

Tables 2 to 5 present a detailed analysis of the results obtained from applying CNN, RNN, LSTM, and a proposed algorithm to four different datasets: Cleveland, Hungarian, Sani, and Statlog. Each method (CNN, RNN, LSTM, and Proposed) is evaluated for accuracy, sensitivity, specificity, and F1 score across multiple runs on the Cleveland dataset. For example, under the CNN method, the accuracy ranges from 94.87% to 95.79%, sensitivity from 95.78% to 95.89%, specificity from 95.79% to 95.87%, and F1 score from 95.34% to 95.65%. Similar to Table 2, this table provides performance metrics for each method across various runs on the Hungarian dataset. For instance, under the RNN method, accuracy ranges from 96.23% to 97.19%, sensitivity from 96.56% to 97.19%, specificity from 95.87% to 97.69%, and F1 score from 96.19% to 96.71%. This table offers a comprehensive analysis of each method's performance on the Sani dataset. For example, under the LSTM method, accuracy ranges from 95.75% to 96.87%, sensitivity from 95.89% to 96.87%, specificity from 95.62% to 95.89%, and F1 score from 95.61% to 96.87%. Similarly, this table presents the performance metrics for each method on the Statlog dataset. For instance, under the Proposed method, accuracy ranges from 97.67% to 97.88%, sensitivity from 96.37% to 96.52%, specificity from 95.59% to 97.28%, and F1 score from 97.59% to 97.75%. Overall, this section provides a comprehensive analysis of the performance of the proposed algorithm in comparison to established deep learning techniques across multiple datasets, offering insights into their effectiveness in addressing the task at hand.

VI. CONCLUSION & FUTURE SCOPE

The paper aims to address the critical issue of early prediction of heart disease with the goal of saving lives globally. It proposes a novel hybrid learning approach that combines a convolutional neural network (CNN) with the firefly algorithm. This hybrid approach is designed to enhance the accuracy of heart disease prediction. The proposed approach combines two distinct techniques: a convolutional neural network (CNN) and the firefly algorithm. This combination aims to leverage the strengths of both approaches to improve the accuracy of heart disease prediction. The researchers utilized a dataset from the UCI repository, which is a widely used source for machine learning research. The dataset initially consisted of 76 instances, with 14 instances used for prediction after preprocessing. Preprocessing likely involved tasks such as data cleaning, normalization, and feature selection to prepare the data for training. The hybrid learning approach achieved impressive performance metrics, with accuracy reported at 91.71%, sensitivity at 98.88%, specificity at 92.75%, and F1 score at 95.70%. These metrics indicate the effectiveness of the proposed approach in accurately predicting heart disease. The research also compared the performance of the hybrid learning approach with other neural network architectures, such as recurrent neural networks (RNN) and long short-term memory (LSTM) networks. The results suggest that the hybrid approach outperformed these other models in terms of prediction accuracy. The high-performance metrics achieved demonstrate the potential of this approach to significantly improve the accuracy of heart disease prediction, ultimately contributing to saving lives worldwide.

REFERENCES

[1]. Dissanayake, Kaushalya, and Md Gapar Md Johar. "Comparative study on heart disease prediction using feature selection techniques on classification algorithms." *Applied Computational Intelligence and Soft Computing* 2021 (2021): 1-17.

[2]. Hasan, Najmul, and Yukun Bao. "Comparing different feature selection algorithms for cardiovascular disease prediction." *Health and Technology* 11 (2021): 49-62.

[3]. Ishaq, Abid, Saima Sadiq, Muhammad Umer, Saleem Ullah, Seyedali Mirjalili, Vaibhav Rupapara, and Michele Nappi. "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques." *IEEE access* 9 (2021): 39707-39716.

[4]. Senan, Ebrahim Mohammed, Ibrahim Abunadi, Mukti E. Jadhav, and Suliman Mohamed Fati. "Score and Correlation Coefficient-Based Feature Selection for Predicting Heart Failure Diagnosis by Using Machine Learning Algorithms." *Computational and Mathematical Methods in Medicine* 2021 (2021).

[5]. Le, Minh Tuan, Minh Thanh Vo, Nhat Tan Pham, and Son VT Dao. "Predicting heart failure using a wrapper-based feature selection." *Indones. J. Electr. Eng. Comput. Sci* 21, no. 3 (2021): 1530-1539.

[6]. Reddy, Karna Vishnu Vardhana, Irraivan Elamvazuthi, Azrina Abd Aziz, Sivajothi Paramasivam, Hui Na Chua, and S. Pranavanand. "Heart disease risk prediction using machine learning classifiers with attribute evaluators." *Applied Sciences* 11, no. 18 (2021): 8352.

[7]. USHA, S., and S. KANCHANA. "Predicting Heart Disease Using Feature Selection Techniques Based On Data Driven Approach." *Webology (ISSN: 1735-188X)* 18, no. 4 (2021).

[8]. Narayanan, Velmurugan Sathya, Kavin N. Raj, Kishore Kumar, and Manoj Kumar. *Survival Prediction of Heart Failure Patients Using Lasso Algorithm and Gaussian Naive Bayes Classifier*. No. 6912. EasyChair, 2021.

[9]. Kumar, Deepak, Chaman Verma, Sanjay Dahiya, Pradeep Kumar Singh, Maria Simona Raboaca, Zoltán Illés, and Brijesh Bakariya. "Cardiac diagnostic feature and demographic identification (CDF-DI): an IoT enabled healthcare framework using machine learning." *Sensors* 21, no. 19 (2021): 6584.

[10]. Khurma, Ruba Abu, Ibrahim Aljarah, and Ahmad Sharieh. "A simultaneous moth flame optimizer feature selection approach based on levy flight and selection operators for medical diagnosis." *Arabian Journal for Science and Engineering* 46 (2021): 8415-8440.

[11]. Rath, Adyasha, Debahuti Mishra, Ganapati Panda, and Suresh Chandra Satapathy. "Heart disease detection using deep learning methods from imbalanced ECG samples." *Biomedical Signal Processing and Control* 68 (2021): 102820.

[12]. Parthiban, R., S. Usharani, D. Saravanan, D. Jayakumar, Dr U. Palani, Dr D. StalinDavid, and D. Raghuraman. "Prognosis of chronic kidney disease (CKD) using hybrid filter wrapper embedded feature selection method." *European Journal of Molecular & Clinical Medicine* 7, no. 9 (2021): 2511-2530.

[13]. Al-Behadili, Hayder Naser Khraibet, and Ku Ruhana Ku-Mahamud. "Fuzzy unordered rule using greedy hill climbing feature selection method: An application to diabetes classification." *Journal of Information and Communication Technology* 20, no. 3 (2021): 391-422.

[14]. Nagarajan, Gayathri, and L. D. Dhinesh Babu. "A hybrid feature selection model based on improved squirrel search algorithm and rank aggregation using fuzzy techniques for biomedical data classification." *Network Modeling Analysis in Health Informatics and Bioinformatics* 10, no. 1 (2021): 39.

[15]. Rodríguez-Rodríguez, Ignacio, José-Víctor Rodríguez, Wai Lok Woo, Bo Wei, and Domingo-Javier Pardo-Quiles. "A comparison of feature selection and forecasting machine learning algorithms for predicting glycaemia in type 1 diabetes mellitus." *Applied Sciences* 11, no. 4 (2021): 1742.

[16]. Alkhodari, Mohanad, Mamunur Rashid, Mohammad Abdul Mukit, Khawza I. Ahmed, Raqibul Mostafa, Sharmin Parveen, and Ahsan H. Khandoker. "Screening cardiovascular autonomic neuropathy in diabetic patients with microvascular complications using machine learning: a 24-hour heart rate variability study." *IEEE Access* 9 (2021): 119171-119187.

[17]. Ward, Michael. "Diagnosis of Myocardial Hypertrophic Disease States through Machine Learning and Mechanistic Modeling." PhD diss., Clemson University, 2021.

[18]. Sarra, Raniya R., Ahmed M. Dinar, Mazin Abed Mohammed, and Karrar Hameed Abdulkareem. "Enhanced heart disease prediction based on machine learning and χ2 statistical optimal feature selection model." *Designs* 6, no. 5 (2022): 87.

[19]. Hevia-Montiel, Nidiyare, Jorge Perez-Gonzalez, Antonio Neme, and Paulina Haro. "Machine Learning-Based Feature Selection and Classification for the Experimental Diagnosis of Trypanosoma cruzi." *Electronics* 11, no. 5 (2022): 785.

[20]. Mengash, Hanan Abdullah, Lal Hussain, Hany Mahgoub, A. Al-Qarafi, Mohamed K. Nour, Radwa Marzouk, Shahzad Ahmad Qureshi, and Anwer Mustafa Hilal. "Smart cities-based improving atmospheric particulate matters prediction using chi-square feature selection methods by employing machine learning techniques." *Applied Artificial Intelligence* 36, no. 1 (2022): 2067647.

[21]. Mahesh, T. R., V. Dhilip Kumar, V. Vinoth Kumar, Junaid Asghar, Oana Geman, G. Arulkumaran, and N. Arun. "AdaBoost ensemble methods using K-fold cross validation for survivability with the early detection of heart disease." *Computational Intelligence and Neuroscience* 2022 (2022).

[22]. Benhar, Houda, Ali Idri, and Mohamed Hosni. "Ensemble Feature Selection for Heart Disease Classification." In *HEALTHINF*, pp. 369-376. 2022.

[23]. Al Gahtani, Abdullah, and Hoda El-Sayed. "Building a Novel Ensemble Learning–Based Prediction Framework for Diagnosis of Coronary Heart Disease." *International Journal of Intelligent Systems and Applications in Engineering* 10, no. 3 (2022): 265-273.

[24]. Khan Mamun, Mohammad Mahbubur Rahman, and Ali T. Alouani. "Cuffless blood pressure measurement using linear and nonlinear optimized feature selection." *Diagnostics* 12, no. 2 (2022): 408.

[25]. Huang, Yen-Chun, Kuan-Yu Chen, Shao-Jung Li, Chih-Kuang Liu, Yang-Chao Lin, and Mingchih Chen. "Implementing an Ensemble Learning Model with Feature Selection to Predict Mortality among Patients Who Underwent Three-Vessel Percutaneous Coronary Intervention." *Applied Sciences* 12, no. 16 (2022): 8135.

[26]. Vasudev, R. Aravind, B. Anitha, G. Manikandan, B. Karthikeyan, Logesh Ravi, and V. Subramaniyaswamy. "Heart disease prediction using stacked ensemble technique." *Journal of Intelligent & Fuzzy Systems* 39, no. 6 (2020): 8249-8257.

[27]. Sabeena, B., S. Sivakumari, and Dawit Mamru Teressa. "Optimization-Based Ensemble Feature Selection Algorithm and Deep Learning Classifier for Parkinson's Disease." *Journal of Healthcare Engineering* 2022 (2022).

[28]. Sada, Wasilah, Celinus Kiyea, Baththama Bello Alhassan, Aminat Bolatito Yusuf, Fatsuma Jauro, and Maria Abur. "Enhancing Heart Disease Prediction Using Ensemble Techniques." *SLU Journal of Science and Technology* 4, no. 1&2 (2022): 130-146.

[29]. Elemam, Tarneem, and Mohamed Elshrkawey. "A Highly Discriminative Hybrid Feature Selection Algorithm for Cancer Diagnosis." *The Scientific World Journal* 2022 (2022).

[30]. Desai, Forum, Deepraj Chowdhury, Rupinder Kaur, Marloes Peeters, Rajesh Chand Arya, Gurpreet Singh Wander, Sukhpal Singh Gill, and Rajkumar Buyya. "HealthCloud: A system for monitoring health status of heart patients using machine learning and cloud computing." *Internet of Things* 17 (2022): 100485.

[31]. Pramanik, Rishav, Sourodip Sarkar, and Ram Sarkar. "An adaptive and altruistic PSO-based deep feature selection method for Pneumonia detection from Chest X-rays." *Applied Soft Computing* 128 (2022): 109464.

[32]. Cohn, F., and Gordon Parker. "Interpretation of Depression Detection Models via Feature Selection Methods."

[33]. Luo, Cida, Yi Zhu, Zhou Zhu, Ranxi Li, Guoqin Chen, and Zhang Wang. "A machine learning-based risk stratification tool for in-hospital mortality of intensive care unit patients with heart failure." *Journal of Translational Medicine* 20, no. 1 (2022): 136.

[34]. Hussain, Lal, Kashif Javed Lone, Imtiaz Ahmed Awan, Adeel Ahmed Abbasi, and Jawad-ur-Rehman Pirzada. "Detecting congestive heart failure by extracting multimodal features with synthetic minority oversampling technique (SMOTE) for imbalanced data using robust machine learning techniques." *Waves in Random and Complex Media* 32, no. 3 (2022): 1079-1102.

[35]. Imran, Zaria, Ethan Grooby, Vinayaka Vivekananda Malgi, Chiranjibi Sitaula, Sunil Aryal, and Faezeh Marzbanrad. "A Fusion of Handcrafted Feature-Based and Deep Learning Classifiers for Heart Murmur Detection."

[36]. Chiu, Chih-Chou, Chung-Min Wu, Te-Nien Chien, Ling-Jing Kao, Chengcheng Li, and Han-Ling Jiang. "Applying an Improved Stacking Ensemble Model to Predict the Mortality of ICU Patients with Heart Failure." *Journal of Clinical Medicine* 11, no. 21 (2022): 6460.

[37]. Nayak, Janmenjoy, Bighnaraj Naik, Paidi Dinesh, Kanithi Vakula, and Pandit Byomakesha Dash. "Firefly algorithm in biomedical and health care: advances, issues and challenges." *SN Computer Science* 1, no. 6 (2020): 311.

[38]. Long, Nguyen Cong, Phayung Meesad, and Herwig Unger. "A highly accurate firefly based algorithm for heart disease prediction." *Expert Systems with Applications* 42, no. 21 (2015): 8221-8231.