

An Intelligent REMORA Optimized RNN With LSTM Mechanism for Detection of Atherosclerosis

Paulraj Ranjith Kumar*, P Govindamoorthi

PSR Engineering College, Sivakasi, Tamilnadu, India

Received: 11 Oct. 2025; Accepted: 14 Jun. 2026

Abstract- Atherosclerosis is a medical condition where plaque accumulates in the arteries. It is important to detect and intervene early to minimize the negative effects of atherosclerosis on a person's health. The World Health Organization (WHO) states that 17.9 million deaths globally are caused by cardiovascular diseases (CVDs), the primary cause of death in the country. Several medical innovations and advancements were being explored in the field of atherosclerosis. The proposed research aims to build a system based on artificial intelligence (AI) automatic framework for an accurate detection and classification of Atherosclerosis using a hybrid REMORA Optimized Recurrent Neural Network with Long Short. Term Memory (RO-RNN-LSTM) method. In this work, various techniques for feature selection, preprocessing, and classification are implemented to achieve more accuracy. The min-max normalization followed by the relevant features from the preprocessed data are extracted using the innovative REMORA Optimization method. To maximize accuracy and minimize classifier time, the dimensionality of features was reduced in this work. The publicly available datasets such as Cleveland, Kaggle (CVD) and Stulong are used to evaluate the proposed algorithms. The prediction findings indicate that utilizing the RO-RNN-LSTM algorithms improves accuracy up to 98.5%, sensitivity to 98.7%, specificity to 95.4%, recall to 98.7% and f1-score to 98.7%. The RO-RNN-LSTM exhibits superior performance outcomes while processing large dimensional medical datasets, according to the overall comparison investigation and analysis.

© 2026 Tehran University of Medical Sciences. All rights reserved.

Acta Med Iran 2026;64(8):531-544.

<https://doi.org/10.18502/acta.v64i8.22850>

Keywords: Atherosclerosis; Cardiovascular diseases; Artificial intelligence; RO-RNN-LSTM; Min-max normalization

Introduction

Atherosclerosis refers that a condition that described by the amassing of plaque in the arteries, making them constrict and limit blood flow. This plaque is made up of blood-derived substances such as cholesterol, fat, and calcium. Plaque narrows and hardens the arteries over time, reducing the quantity of blood in oxygen that reaches the body's vital organs. The major cause of death worldwide, accounting for around 17.9 million fatalities annually, is cardiovascular diseases (CVDs), according to the World Health Organization (WHO). These conditions include heart disease, cerebrovascular disease, peripheral arterial disease, and other disorders that affect the blood

vessels and heart. Tobacco use, excessive alcohol consumption, bad eating choices, and physical inactivity are the main behavioral risk factors for heart disease and stroke. High blood pressure, high blood sugar, raised blood lipids, and other disorders can be brought on by these risk factors. In order to prevent premature deaths, it is essential to detect and treat individuals at high risk of cardiovascular diseases (CVDs) at an early stage. The World Health Organization (WHO) has also recognized the importance of addressing non-communicable diseases, including CVDs, and formulated a worldwide action plan with the goal of reducing premature mortality by 25% by 2025. Medication and counseling are required for at least 50% of eligible patients in order to avoid heart

Corresponding Author: P.R. Kumar

PSR Engineering College, Sivakasi, Tamilnadu, India

Tel: +91 8072823175, E-mail address: vijiranjith2011@gmail.com

Copyright © 2026 Tehran University of Medical Sciences. Published by Tehran University of Medical Sciences

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited

attacks and strokes. Other optional global targets included in this strategy include lowering the prevalence of high blood pressure globally. By achieving these goals, significant investments need to be made in healthcare systems. Early detection of atherosclerosis, a condition that contributes to CVDs, is crucial for implementing preventive measures and providing appropriate treatment. When determining the existence and severity of atherosclerosis, diagnostic techniques are essential and allowing healthcare providers to develop personalized management strategies for individuals at risk.

Various imaging modalities, such as single-photon emission computed tomography (SPECT), computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), and magnetic resonance angiography (MRA), are employed to measure the fibrous membrane thickness and lesion size in atherosclerosis (1). Atherosclerotic plaques identified and described using ultrasound techniques such contrast-enhanced ultrasound imaging (CEUS), Doppler mode, B-mode ultrasound, and elastography (2). These imaging modalities play a crucial role in assessing the severity and progression of atherosclerosis, aiding in diagnosis and treatment planning. The degree of severity in cardiac disease in humans has been determined by using different data mining and neural network techniques. The accuracy of cardiovascular risk prediction is successfully improved by machine learning algorithms; however more patient records are needed to get greater results. Several machine learning and deep learning algorithms are used by various researchers to accurately identify and categorize CAD from the patient's medical data. However, the majority of methods have issues with processing speed, accuracy, and efficiency when dealing with massive amounts of data. As a result, the goal of the ongoing research is to create a novel, automated system for reliably detecting the presence of atherosclerotic disease in medical records. A new method based on an advanced deep learning algorithm has been presented to efficiently assess coronary computed tomography angiography (CCTA) data. With the use of this cutting-edge technology, coronary artery atherosclerosis can be accurately detected or not (3). The creation and integration of a Medical Diagnostic Support System (MDSS) for cardiovascular conditions align with the increasing focus on utilizing technology to improve medical diagnosis and decision-making processes. By incorporating advanced algorithms and machine learning techniques, the MDSS potentially improves the accuracy and precision of cardiovascular disease diagnosis. The proposed research is focused on the development of Remora Optimized–Recurrent Neural

Network–Long Short-Term Memory (RO-RNN-LSTM) classification technique for an efficient diagnosis of Atherosclerosis. The main contribution of this work is to develop an Artificial Intelligence (AI) based automatic framework for an accurate disease diagnosis. In order to reduce the dimensionality of data for an accurate disease prediction, this work aims to develop a novel optimization algorithm which incorporated with deep learning algorithm for classification of atherosclerosis disease.

The remaining parts of this paper are divided into the following categories: The thorough literature analysis in Section 2 examines various learning-based classification algorithms that are employed for the automatic detection of CAD. The goals of this research work are listed in Section 3. The full justification for the suggested optimization integrated classification algorithms for atherosclerotic disease diagnosis is provided in Section 4. Additionally, Section 5 assesses the effectiveness and comparative results of the proposed approaches using widely available public datasets and evaluation measures. In Section 6, the overall process is summarized together with the main conclusions, research contributions, and future work.

Background of the work

The transition from traditional clinical diagnostics to AI-driven methodologies marks a pivotal era in cardiovascular medicine. Poznyak *et al.*, (1) provide a comprehensive baseline of these existing methods, underscoring a critical need for high-precision screening tools. This gap is increasingly filled by Artificial Intelligence. Pudjihartono *et al.*, (2) and Hasnat *et al.*, (3) emphasize that the efficacy of these AI models is not solely dependent on the algorithms themselves but on the quality of feature selection and input data. Their research suggests that the "garbage in, garbage out" principle is particularly high stakes in heart disease prediction, where model reliability dictate clinical outcomes. Furthermore, in specialized fields like vascular surgery, Fischer *et al.*, (4) identify AI as a transformative force for real-time intraoperative decision-making, moving the technology from theoretical prediction to practical surgical assistance. A significant hurdle in the adoption of AI is the lack of transparency in "black-box" models. Chen *et al.*, (5) address this by integrating data-mined knowledge into Support Vector Machines (SVM), thereby enhancing interpretability and fostering clinical trust. The robustness of these models is further explored by Shukur and Mijwil (6), who demonstrate how dataset dimensionality impacts performance, and Vafaei *et al.*, (8), who advocate for

rigorous normalization to maintain consistency across disparate medical data sources. To maximize accuracy without sacrificing stability, Gowthamani *et al.*, (7) introduced "hard voting" ensemble techniques, proving that a consensus among multiple models often outperforms a single algorithmic approach. A major trend observed in the shift toward metaheuristic optimization—using nature-inspired algorithms to fine-tune neural network parameters. This approach addresses the limitations of standard backpropagation by exploring a wider solution space. Binarized Butterfly Optimization was proposed by Amutha and Sekar (9) for enhanced CVD prediction. Likelihood Swarm Whale Optimization: Specifically tailored by Govindamoorthi and Ranjith Kumar (15) to optimize the LeNet classifier for atherosclerosis detection. Specialized Algorithms: The application of Remora Optimization (12), Dwarf Mongoose Optimization (13), and Aquila-optimized feature selection in Quantum Neural Networks (14) showcases the diversity of these optimization strategies in improving diagnostic classification.

As medical data becomes more diverse, the focus has shifted toward specific architectures suited for different data types. Convolutional Neural Networks (CNNs) have become the gold standard for image-based diagnostics, such as the multi-class plaque detection framework developed by Zhao *et al.*, (16). For sequential and time-series data—such as ECG rhythms or health factor trends—RNNs and LSTMs have proven superior. Studies by Sunil *et al.*, (18) and Elkomy *et al.*, (19) utilize RNN-LSTM combinations to process heartbeats and multimodal data, while Dhawale *et al.*, (20) found that attention-based Deep Neural Networks (DNNs) offer the highest accuracy by focusing on the most relevant temporal segments. Finally, the reach of these models is expanding. Recent work by Shoaib *et al.*, (17) uses Deep Learning to analyze plasma cytokine data, while extend these applications to accelerometer-based hypertension classification. Collectively, these studies indicate that the future of atherosclerosis diagnostics lies in a hybrid approach: combining the raw power of deep learning with the strategic refinement of nature-inspired optimization and the transparency of interpretable AI.

Materials and Methods

Block diagram of proposed RO-RNN-LSTM classification technique

This section gives the complete description for proposed Remora Optimized Recurrent Neural Network with Long Short-Term Memory (RO-RNN-LSTM)

classification technique for an efficient diagnosis of Atherosclerosis. The main role of this work is to develop an Artificial Intelligence (AI) based automatic framework for an accurate disease diagnosis. In our previous work, we have implemented an effective automated detection tool for the identification and classification of atherosclerosis disease using Real-Time Invariant Atherosclerosis Feature Selection and Classification Model (RIFSACM). In order to reduce the dimensionality of data for an accurate disease prediction, this work aims to develop a novel optimization algorithm which incorporated with deep learning algorithm for classification of atherosclerosis disease. The flow of the proposed RO-RNN-LSTM system is shown in Figure 1 which comprises the following modules:

- Dataset collection
- Data preprocessing using min-max normalization technique
- Feature selection using Remora Optimization (RO) algorithm
- Disease classification using Recurrent Neural Network with Long Short Term Memory (RNN-LSTM) model
- Performance evaluation

The freely available Cleveland, Alizadeh and Hungarian datasets have been used to evaluate the proposed system. The typical normalization operation is performed to get preprocessed data to improve accuracy of the detection system. In order to remove unnecessary information from the heart disease dataset, we utilize the min-max normalization technique along with data imputation. Following this, we employ the innovative RO algorithm to extract the necessary features from the preprocessed data. The optimized feature selection process, which utilizes the best optimum function, aids in reducing data dimensionality, thereby increasing classifier accuracy and decreasing time consumption. Moreover, the optimum features are passed to the RNN-LSTM classification approach for an effective prognosis of atherosclerosis disease. A hybrid deep learning algorithm developed by integrating the operations of RNN and LSTM techniques. By using this technique, the disease is accurately detected from the given data.

Min-max normalization

The data preprocessing is performed with the use of min-max normalization technique, which maximizes the accuracy and efficiency of disease detection. One of the popular data transformation methods for maintaining the sensitive specific characteristics in a set of data is a min-

max normalization. By considering the lowest and highest values in a dataset, the raw dataset values are normalized using the min-max normalization function. The original dataset is transformed linearly using this method. The primary goal of this method is to standardize the initial dataset into a retained form, which satisfies the standards for resilient private data with little information loss. This method transforms the variables linearly, with s being the set of observed values and min-max representing the minimum and maximum values of s , and the function is calculated as follow.

$$m = \frac{s - s_{min}}{s_{max} - s_{min}} \quad (1)$$

This is a mapping or scaling of the data, mapping the minimum value of s to 0 and the largest value of s to 1.

Remora optimization (RO) based feature selection

The RO approach is used to choose the necessary characteristics from the preprocessed dataset in order to correctly forecast the disease from the medical dataset after preprocessing. It is a recently developed meta-heuristic technique for handling global optimization issues. It has the major benefits like increased convergence rates, accurate solution rates, and minimum local optima. Similar to other meta-heuristic models, the ROA leverages the parasitic habit of remora, which is one of their biological traits, to finish the optimization process. Remora may stick to whales, swordfish, and other creatures, and it readily obtains food with the host's assistance. Moreover, it bears a striking resemblance to the sponging behavior of remora fish. Known for their exceptional intelligence, remoras attach themselves to whales, sailfish, and other marine creatures to feed off their host's resources while conserving their own energy and protecting themselves from potential threats. Consequently, in an effort to update their positions, the Remora Optimization Algorithm (ROA) incorporates specific formulas derived from the Sailed Fish Optimizer (SFO) and Whale Optimization Algorithm (WOA). Additionally, similar to how remoras navigate within a limited range around their host or fend off attacks, the algorithm evaluates whether it is necessary to switch hosts in order to optimize its performance. Moreover, this optimization technique comprises the following phases of operations:

- Initialization
- Exploration
- Exploitation
- Identification of Best solution

During initialization, the population updating is

performed as shown in the following equation:

Recurrent neural network–long short-term memory (RNN-LSTM) based disease classification

The RNN-LSTM classification algorithm is used to classify the disease class with high accuracy and detection efficiency after feature optimization. Several deep learning techniques were proposed by various researches. However, they are having trouble with characteristics including longer training and testing times, less accuracy in predicting the type of sickness, and higher error rates. Thus, the goal of this research is to apply the cutting-edge and potent RNN-LSTM deep learning algorithm for the detection and categorization of diseases. A feed forward neural network with an internal memory is generalized into an RNN. The outcome of the current input in an RNN depends on the previous computations. Once the output is obtained, it is replicated and redirected into the recurrent network. RNNs function inputs related to one another using their internal memory to make decisions. Back-propagation is used by LSTM for training. The architecture is made up of memory chunks or linked sub-networks. Long-term memory retention is memory blocks' default function. Every memory block has many multiplicative units, including input, forget, and output gates, as well as a minimum of one self-connected accumulator cell. By setting a counter, such as 0 or 1, these three gates enable users to store and retrieve data. Since the RNNs increase classification performance and produce better results than conventional feed-forward networks, they are the ideal choice for classifying time-series data. Because the output contains traces of variables retained in the memory and the network is given back with feedback with current value. As shown in Figure 2, there are three gates in an LSTM network such as input, forget, and output. The input gate updates the memory by choosing the parameters from the input array to activate using a function called sigmoid. Certain details from past states are removed from the block via the forget gate and the output is managed by the output gate. The RNN-LSTM prediction operations are performed as represented in the following equations:

$$\mathbb{i}_k = \delta(\omega_{gi}g_k + \omega_{qi}h_{k-1} + \omega_{ci}p + \mathbb{b}_i) \quad (12)$$

$$\mathbb{f}_k = \varphi(\omega_{gf}g_k + \omega_{qf}h_{k-1} + \omega_{cf}p + \mathbb{b}_f) \quad (13)$$

$$\mathbb{o}_k = \varphi(\omega_{go}g_k + \omega_{qo}h_{k-1} + \omega_{co}p + \mathbb{b}_o) \quad (14)$$

$$\mathbb{g}_k = \varrho(\omega_{gs}g_k + \omega_{qs}h_{k-1} + \omega_{cs}p + \mathbb{b}_s) \quad (15)$$

$$\zeta_k = \mathbb{f}_k \odot \zeta_k - 1 + \mathbb{i}_k \odot \mathbb{g}_k \quad (16)$$

$$h_k = \mathbb{o}_k \odot \varrho(\zeta_k) \quad (17)$$

Where, $\mathbb{i}_k$ is the input gate, $\mathbb{f}_k$ represents the forget gate, $\mathbb{o}_k$ denotes the output gate, $\mathbb{g}_k$ indicates the input

modulation gate, g is the input vector, h represents the hidden state, p is the visual feature, ζ represents the memory cell, ω is the weight value, and b represents the bias value. Moreover, the softmax function is used to predict the classified label during the training process.

Results

The performance of the proposed RO-RNN-LSTM model is evaluated and compared with the publically available datasets such as Cleveland, Kaggle (CVD) and Stulong. The accuracy and overall effectiveness of the disease diagnosis system are generally reduced by the lower training data, which typically leads to poor approximation. Hence, it is necessary to apply for a suitable dataset having the best collection of features for a perfect classification. Moreover, the following performance measures are considered for evaluation such as accuracy, precision, recall, f1-score, sensitivity, and specificity. The mathematical equations for calculating these parameters are described in the previous chapter. As a result, a confusion matrix is a tabular representation of the dataset's real values and the estimated model. It is used to assess the models' performance in classification-based machine learning challenges. The four different arrangements of the actual and predicted numbers that make up this composition are False Positive, False Negative, True Positive, and True Negative.

Figure 3 illustrates that Confusion matrix of Cleveland Dataset. Cleveland datasets contain 303 records, the model predicted 295 times "Yes" and 8 times

"No". True Positive is the model that has predicted 231 times that the patient is having the disease accurately. True Negative is the model has predicted 64 times that the patient is not having heart disease precisely. False Negative is thrice the model has predicted that the patient is not having the disease when the patient is having the disease. False Positive is the model that has forecasted 5 times that the patient is having disease while the patient is not having the disease. Figure 4 shows accuracy, sensitivity, specificity, precision, Recall and F1-Score of the existing and proposed disease prediction models using Cleveland dataset, and the corresponding values are available in Table 1. The proposed RO-RNN-LSTM has produced higher accuracy compared to other approaches. It is observed that accuracy is increased from 82.9% to 97.5%. Similarly, the sensitivity is increased from 88.2% to 97.4%, specificity is increased from 89.6% to 95.4%, precision is increased from 83.5% to 98.7%, recall is increased from 87.4% to 97.4%, f1-scores increased from 86.4% to 98.1%. Figure 5 shows that Confusion matrix of Kaggle (Cardio Vascular Disease) Dataset. Kaggle (CVD) dataset contains 70000 records, the model has forecasted 68950 times "Yes" and 1050 times "No". True Positive is the model that has predicted 54000 times that the patient is having the disease perfectly. True Negative is the model has predicted 14950 times that the patient is not having heart disease accurately. False Negative is the model has predicted 400 times that the patient is not having the disease when the patient is having the disease. False Positive is the model has predicted 650 times that the patient is having disease while the patient is not having the disease.

Table 1. Analysis on various Performance Metrics Using Cleveland Dataset

Methods	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-Score
RF	82.9	88.7	93.8	90.1	91.3	91.4
SVM (RBF)	85.3	90.4	93.8	90.3	91.2	92.4
SVM (Lin)	85.6	90.7	93.9	92.1	91.4	90.4
DT	87.8	88.3	89.6	89.2	89.5	87.6
Adaboost	87.8	88.2	89.7	83.5	87.4	86.4
RIFSACM	96.1	96.2	95.5	98.7	96.2	97.4
RO-RNN-LSTM	97.5	97.4	95.4	98.7	97.4	98.1

Table 2. Analysis on various Performance Metrics Using Kaggle (CVD) Dataset

Methods	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-Score
SVM (RBF)	78.3	95.7	68.4	91.1	90.2	90.6
Adaboost	80.1	82.8	76.9	86.8	85.7	86.2
DT	82.3	87.7	67.9	87.9	86.4	87.1
SVM (Lin)	86.4	94.3	64.5	91.7	91.7	91.7
RF	93.2	95.2	91.4	92.6	91.7	92.1
RIFSACM	97.2	97.4	95.6	98.7	97.4	98.1
RO-RNN-LSTM	98.5	98.7	95.5	98.7	98.6	98.7

Table 3. Analysis on various Performance Metrics Using Stulong Dataset

Methods	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-Score
RF	82.9	88.7	93.8	90.1	91.3	91.4
SVM (RBF)	85.3	90.4	93.8	90.3	91.2	92.4
SVM (Lin)	85.6	90.7	93.9	92.1	91.4	90.4
DT	87.8	88.3	89.6	89.2	89.5	87.6
Adaboost	87.8	88.2	89.7	83.5	87.4	86.4
RIFSACM	98.6	98.7	95.2	98.7	98.7	98.7
RO-RNN-LSTM	98.1	98.7	95.2	98.7	98.7	98.7

Table 4. Comparative analysis based on accuracy with respect to varying training data for Cleveland dataset

Training Data (%)	BF-PSO	Bi-LSTM-CRF	XGB	RLNNC	DMOA-Without FS	DMOA	RO-RNN-LSTM
50	0.79	0.8	0.81	0.815	0.82	0.835	0.9
60	0.8	0.81	0.81	0.83	0.83	0.85	0.95
70	0.82	0.83	0.84	0.84	0.86	0.88	0.97
80	0.86	0.87	0.87	0.87	0.87	0.90	0.978
90	0.89	0.892	0.9	0.91	0.91	0.92	0.99

Table 5. Comparative analysis based on sensitivity with respect to varying training data for Cleveland dataset

Training Data (%)	BF-PSO	Bi-LSTM-CRF	XGB	RLNNC	DMOA-Without FS	DMOA	RO-RNN-LSTM
50	0.81	0.83	0.85	0.88	0.89	0.91	0.91
60	0.81	0.84	0.85	0.88	0.89	0.91	0.96
70	0.82	0.84	0.85	0.89	0.90	0.91	0.97
80	0.82	0.84	0.86	0.89	0.90	0.91	0.98
90	0.83	0.84	0.87	0.89	0.90	0.92	0.99

Table 6. Comparative analysis based on specificity with respect to varying training data for Cleveland dataset

Training Data (%)	BF-PSO	Bi-LSTM-CRF	XGB	RLNNC	DMOA-Without FS	DMOA	RO-RNN-LSTM
50	0.78	0.78	0.80	0.82	0.82	0.84	0.91
60	0.79	0.82	0.83	0.83	0.83	0.84	0.96
70	0.80	0.82	0.86	0.83	0.83	0.85	0.97
80	0.84	0.86	0.88	0.90	0.90	0.90	0.98
90	0.86	0.88	0.9	0.91	0.91	0.91	0.98

Table 7 Comparative analysis with optimization integrated classification models using Cleveland dataset

Methods	Accuracy	Sensitivity	Specificity
BF-PSO	0.88	0.82	0.87
Bi-LSTM-CRF	0.89	0.84	0.88
XGB	0.89	0.87	0.89
RLNNC	0.90	0.89	0.90
DMOA-SN without FS	0.90	0.90	0.90
DMOA-SN	0.91	0.91	0.90
Ro-RNN-LSTM	0.98	0.98	0.98

Table 8. Comparative analysis with standard machine learning classification models using Z-Alizadeh dataset

Methods	Precision (%)	Recall (%)	F1-Score (%)
MNB	0.91	0.99	0.96
SVM	0.98	0.98	0.98
LR	0.91	0.91	0.91
CART	0.85	0.93	0.89
LDA	0.98	0.95	0.96
AB	0.93	0.98	0.95
RF	0.91	0.95	0.93
ET	0.95	0.98	0.97
XGB	0.90	0.99	0.95
Proposed	98.8	99.3	98.7

Table 9. Comparative analysis with other state of the art model techniques using Cleveland dataset

Methods	Recall (%)	Specificity (%)
Ensemble model	89.68	89.31
Hybrid neural network	93.32	78.51
PPCA+SVM	75.21	90.57
Fuzzy rule	84.70	92.90
Gaussian NB	87.80	97.95
DBN	96.03	93.15
Hybrid NN-GA	88.35	91.49
DE + BPNN	82.35	92.31
DT + FIS	92.44	96.18
Proposed	98.73	98.80

Table 10. Comparative analysis with other state of the art model techniques using Hungarian dataset

Methods	Recall (%)	Specificity (%)
Ensemble model	82.39	92.02
PPCA+SVM	80.43	88.42
Hybrid NN + GA	85.87	88.32
Fuzzy classification	82.98	90.57
DT + FIS	90.42	79.24
Proposed	98.41	98.57

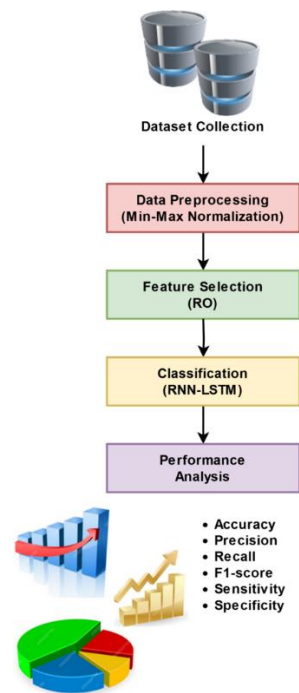


Figure 1. Flow of the proposed RO-RNN-LSTM model

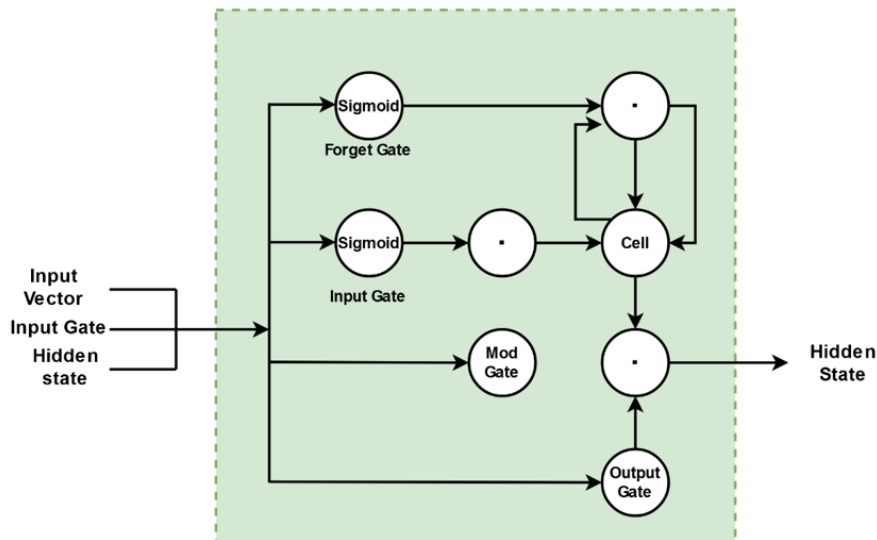


Figure 2. RNN-LSTM architecture model

Training Set + Testing			
TARGET \ OUTPUT	True	False	SUM
True	231 76.24%	3 0.99%	234 98.72% 1.28%
False	5 1.65%	64 21.12%	69 92.75% 7.25%
SUM	236 97.88% 2.12%	67 95.52% 4.48%	295 / 303 97.36% 2.64%

Figure 3. Confusion Matrixes for Cleveland Dataset

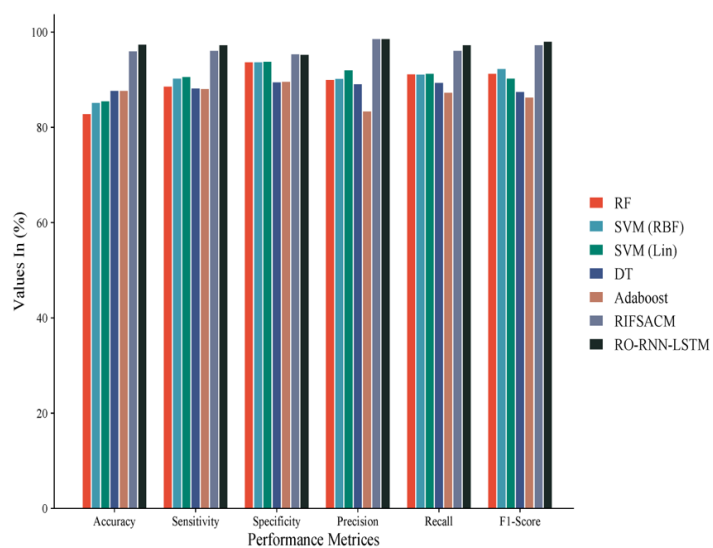


Figure 4. Comparative analysis using cleveland dataset

Training Set + Testing			
TARGET \ OUTPUT	True	False	SUM
True	54000 77.14%	400 0.57%	54400 99.26% 0.74%
False	650 0.93%	14950 21.36%	15600 95.83% 4.17%
SUM	54650 98.81% 1.19%	15350 97.39% 2.61%	68950 / 70000 98.50% 1.50%

Figure 5. Confusion matrix for kaggle (CVD) dataset

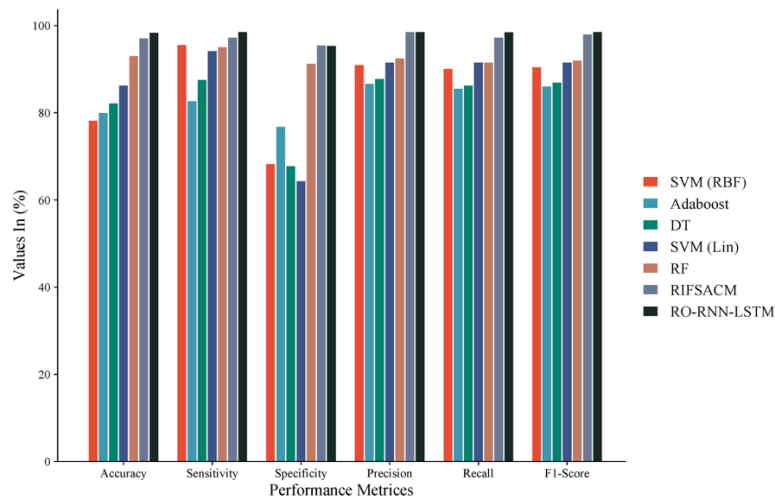


Figure 6. Comparative analysis using kaggle (cardio vascular disease) dataset

Training Set + Testing			
TARGET \ OUTPUT	True	False	SUM
True	1107 78.01%	14 0.99%	1121 98.75% 1.25%
False	13 0.92%	285 20.08%	298 95.64% 4.36%
SUM	1120 98.84% 1.16%	299 95.32% 4.68%	1392 / 1419 98.10% 1.90%

Figure 7. Confusion matrix for stulong dataset

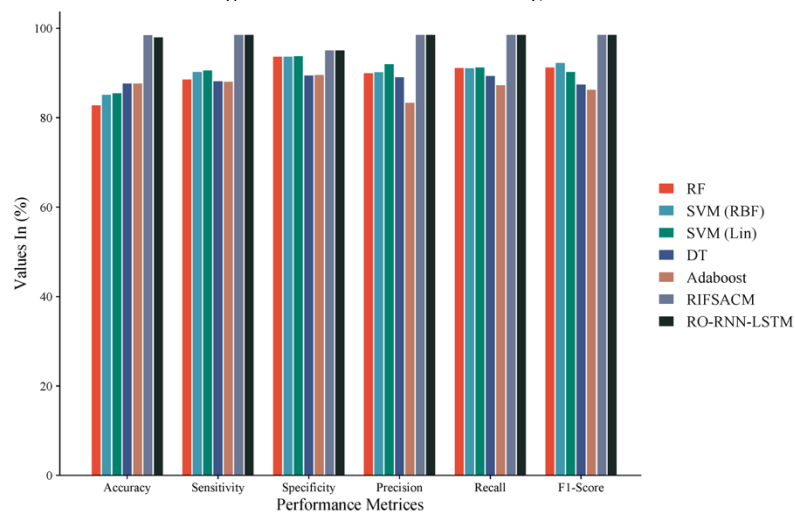


Figure 8. Comparative analysis using stulong dataset

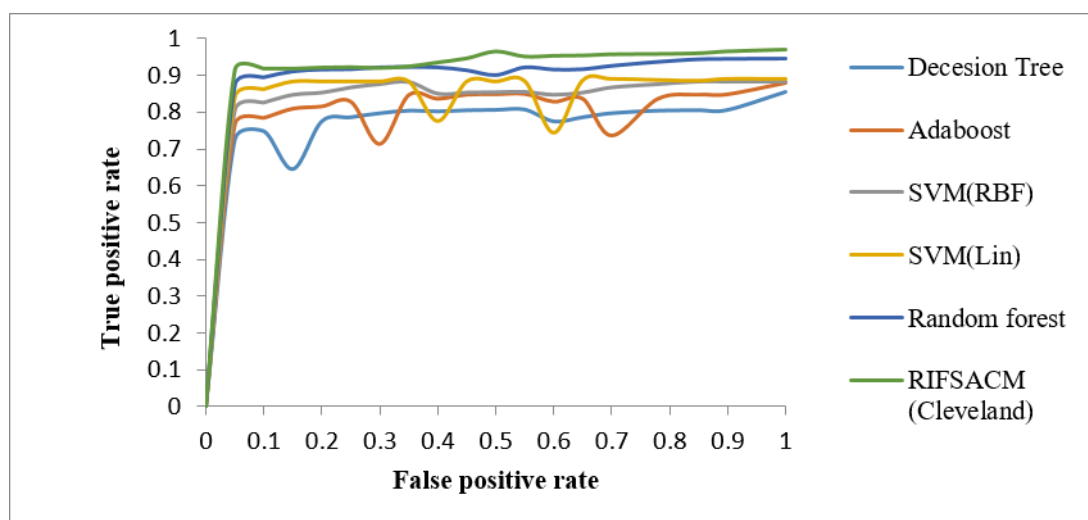


Figure 9. ROC analysis for cleveland dataset

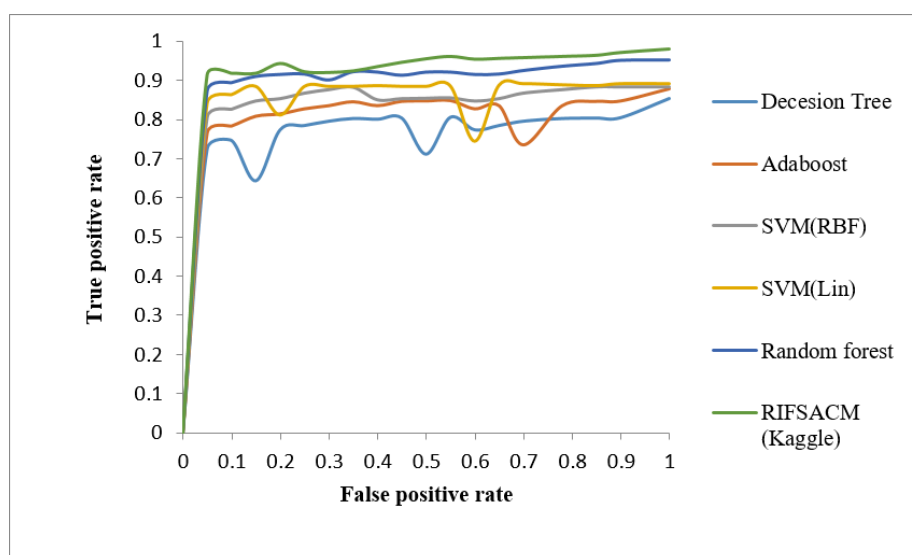


Figure 10. ROC analysis for Kaggle dataset

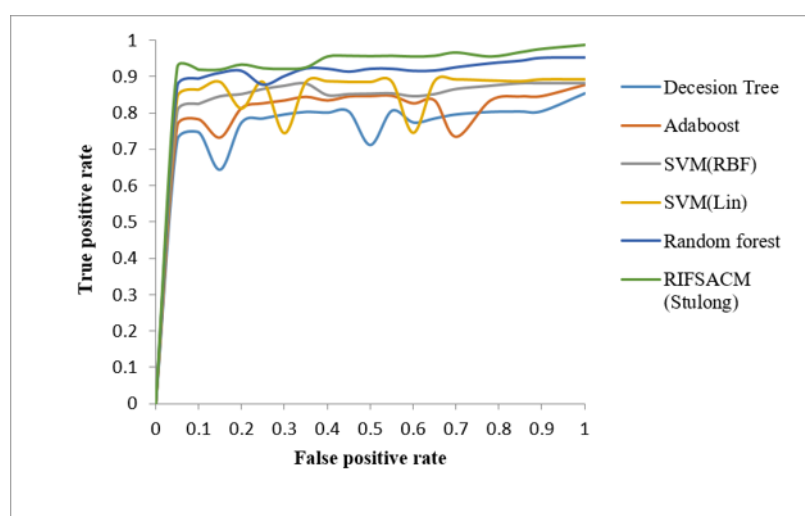


Figure 11. ROC analysis for Stulong dataset

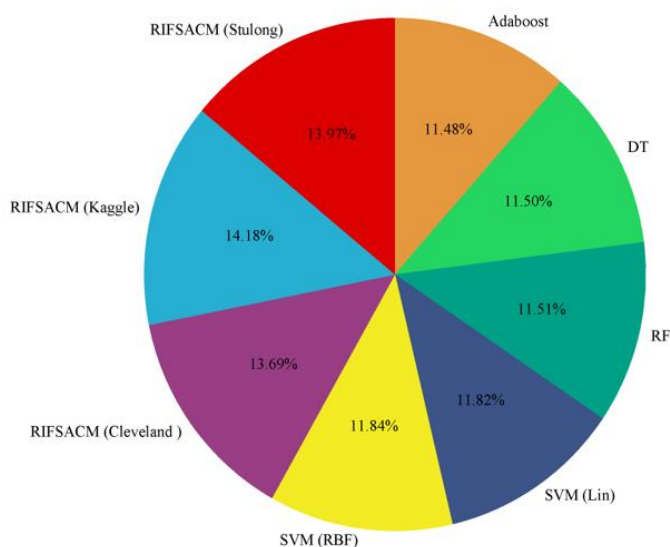


Figure 12. Comparison based on MCC using different dataset

Discussion

Figure 6 shows an accuracy, sensitivity, specificity, precision, Recall and F1-Score of the existing and proposed disease prediction models using Kaggle (CVD) dataset, and the corresponding values are available in Table 2. It is observed that the accuracy is increased from 78.3% to 98.5%. Similarly, the sensitivity is increased from 82.8% to 98.7%, specificity is increased from 64.5% to 95.5%, precision is increased from 86.8% to 98.7%, recall is increased from 85.7% to 98.6%, f1-scores increased from 86.2% to 98.7%. Figure 7 shows that Confusion matrix of Stulong Dataset. Stulong dataset contains 1419 records; the proposed model has predicted 1392 times “Yes” and 27 times “No”. True Positive is the model has predicted 1107 times that the patient is having the disease truthfully. True Negative is the model has predicted 285 times that the patient is not having heart disease accurately. False Negative is the model has forecasted 14 times that the patient is not having the disease when the patient is having the disease. False Positive is the model that has predicted 13 times that the patient is having disease while the patient is not having the disease. Figure 8 shows an accuracy, sensitivity, specificity, precision, Recall and F1-Score of the existing and proposed disease prediction models using Stulong dataset, and the corresponding values are available in Table 3. It is observed that the accuracy is increased from 82.9% to 98.1%. Similarly, the sensitivity is increased

from 88.2% to 98.7%, specificity is increased from 89.6% to 95.2%, precision is increased from 83.5% to 98.7%, recall is increased from 87.4% to 98.7%, f1-scores increased from 86.4% to 98.7%. The true positive rate (TPR) and false positive rate (FPR) are two metrics that are plotted on this curve at various categorization criteria. TPR is shown on the Y-axis and FPR is shown on the X-axis of a ROC curve graph. When the classification threshold is decreased, more observations are classified as positive, which increases TPR and FPR. Figure 9 to Figure 11 validate the Receiver Operating Characteristics (ROC) of the conventional and proposed classification techniques used for atherosclerosis disease detection using Cleveland, Kaggle (Cardiovascular Disease) and Stulong datasets respectively. The ROC is specifically used to assess the binary classifier's ability to diagnose abnormalities. The analysis ROC curve is frequently employed in machine learning and medical decision-making research applications. In order to draw the ROC curve. The effectiveness of the classifier is assessed using the ROC evaluation metric with an effort to raise the disease prediction rate.

The total findings show that the proposed RIFSACM model's ROC value has been increased to 0.971 for the Cleveland dataset, 0.981 for the Kaggle dataset and 0.987 for the Stulong dataset. It is observed that the ROC is significantly increased when compared to the traditional approaches. Consequently, the Matthews Correlation Coefficient (MCC) is also validated and compared as

shown in Figure 12. The existing techniques considered in this study are Random Forest (RF), Decision Tree (ET), Adaboost, Support Vector Machine (SVM) – Radial Basis Function (RBF), and SVM – Linear Kernel (Lin). The total findings show that the proposed RO-RNN-LSTM model's MCC value has been increased to 0.924 for the Cleveland dataset, 0.957 for the Kaggle dataset and 0.943 for the Stulong dataset. In this analysis, 80% of data has been used for training and 20% of data has been used for testing or validation. This comparison analysis shows that the enhanced accuracy of 98.5%, sensitivity of 98.7%, specificity of 95.4% and precision of 98.7% of the suggested RO-RNN-LSTM model outperforms the traditional classification methodologies. The combination of the proposed RO-RNN-LSTM model surpasses all other conventional classification approaches with great performance outcomes, according to the comparative assessment. The proposed approach significantly increases the disease detection accuracy of RO-RNN-LSTM by utilizing an efficient feature optimization mechanism. The RO-RNN-LSTM exhibits superior performance outcomes while processing large dimensional medical datasets, according to the overall comparison investigation and analysis.

Limitations

The proposed RO-RNN-LSTM model, while achieving a high accuracy of 98.5% on the Kaggle dataset, faces several critical limitations regarding its clinical application and technical execution. The efficacy of cardiovascular risk prediction is fundamentally dependent on the availability of massive patient records, as lower volumes of training data typically lead to poor approximation and reduced diagnostic effectiveness. Furthermore, the model contends with the inherent challenges of deep learning, such as significant training and testing times and the potential for higher error rates when processing large-dimensional medical datasets. A major barrier to adoption remains the "black-box" nature of architecture, which lacks the transparency necessary to foster clinical trust; consequently, the authors acknowledge the need to integrate explainable artificial intelligence (XAI) in future iterations to improve interpretability. Finally, while the model is optimized for time-series data, its performance and processing speed can fluctuate significantly based on the dimensionality and consistency of the input data.

This paper presents a novel and automated detection method for the recognition and diagnosis of atherosclerosis using three different heart disease datasets

such as Cleveland, Kaggle (CVD) and Stulong. In this work, various techniques for feature selection, preprocessing, and classification have been implemented to achieve more accuracy. The min-max normalization procedure was carried out to obtain the preprocessed data, which enhances the detection system's accuracy and efficiency. Followed by the relevant features from the preprocessed data were extracted using the innovative RO method. In order to maximize accuracy and minimize classifier time, the dimensionality of features was reduced. Additionally, the RNN-LSTM classification system received the best attributes for a reliable prognosis of atherosclerotic disease. The performance validation parameters such as f1-score, MCC, sensitivity, specificity, accuracy, precision, recall, and sensitivity have been calculated to assess the effectiveness of the proposed method. The prediction findings indicate that utilizing the RO-RNN-LSTM algorithms can improve accuracy up to 98.5%, sensitivity to 98.7%, specificity to 95.4%, recall to 98.7% and f1-score to 98.7%. In future, the proposed work will be implemented by using explainable artificial intelligence techniques.

References

1. Amutha S, Sekar JR. Deep neural network based cardio cardiovascular disease prediction using binarized butterfly optimization. *Intell Autom Soft Comput*. 2023;36:1863-1880.
2. Balasubramaniam S, Kumar KS, Kavitha V, Prasanth A, Sivakumar T. Feature Selection and Dwarf Mongoose Optimization Enabled Deep Learning for Heart Disease Detection. *Comput Intell Neurosci* 2022;2022:1-15.
3. Chen S, Gao C, Zhang P. Incorporation of Data-Mined Knowledge into Black-Box SVM for Interpretability. *ACM Trans Intell Syst Technol* 2023;14:1-22.
4. Darolia A, Chhillar RS, Alhussein MA, Dalal S, Aurangzeb K, Lilhore U. Enhanced cardiovascular disease prediction through self-improved Aquila optimized feature selection in quantum neural network & LSTM model. *Front Med (Lausanne)* 2024;11:1364319.
5. Dhawale H, Thakare D, Dhule C, Agrawal R, Morris N. Coronary Heart Disease Prediction Using CNN, RNN, LSTM and Attention DNN Model. In: 2025 International Conference on Computational, Communication and Information Technology (ICCCIT); 2025.
6. Elkomy OM, Rushdy E, Nasser S, Khashaba MM. Exploring differential privacy in CNNs, LSTMs, GRUs, and RNNs for heartbeat detection from multimodal data. *J Big Data* 2025;12.
7. Govindamoorthi P, Ranjith Kumar P. A likelihood swarm

- whale optimization based LeNet classifier approach for the prediction and diagnosis of patients with atherosclerosis disease. *Comput Methods Biomech Biomed Engin* 2023;26:326-337.
8. Hasnat R, Al Mamun A, Musha A, Tahabilder A. A Review on Heart Diseases Prediction Using Artificial Intelligence. In: *International Conference on Machine Intelligence and Emerging Technologies*. Springer; 2022.
9. Majeed MA, Phichaisawat S, Asghar F, Hussan U. Data-driven optimized load forecasting: an LSTM-based RNN approach for smart grids. *IEEE Access*; 2025.
10. Malibari AA, Hassine SBH, Motwakel A, Hamza MA. Metaheuristics with deep learning empowered biomedical atherosclerosis disease diagnosis and classification. *Comput Mater Contin* 2022;72:2859-75.
11. Ouellet M, Wac K, Siebra CA. Long Short-Term Memory Network for accelerometer-based hypertension classification. *Stud Health Technol Inform* 2025.
12. Poznyak AV, Sukhorukov VN, Eremin II, Nadelyaeva II, Orekhov AN. Diagnostics of atherosclerosis: Overview of the existing methods. *Front Cardiovasc Med* 2023;10:1134097.
13. Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Front Bioinform* 2022;2:927312.
14. Saboor A, Usman M, Ali S, Samad A, Abrar MF, Ullah N. A method for improving prediction of human heart disease using machine learning algorithms. *Mob Inf Syst* 2022;2022:1-12.
15. Shahzadi K, Abid MK, Naeem A, Fuzail M, Aslam N, Sajjad R. Predicting the Risk of Cardiovascular Diseases Using Deep Learning Models. *Kashf J Multidiscip Res* 2025.
16. Shoaib M, Junaid A, Husnain G, Qadir M, Ghadi Y, Askar SS, et al. Advanced detection of coronary artery disease via deep learning analysis of plasma cytokine data. *Front Cardiovasc Med* 2024;11:1374790.
17. Shukur B, Mijwil M. Involving machine learning techniques in heart disease diagnosis: a performance analysis. *Int J Electr Comput Eng* 2023;13:2177-85.
18. Soni T, Pasha SN. Deep Learning for Early Detection of Cardiac Diseases: RNN Analysis of PTB XL+ Data. In: *2025 3rd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*; 2025.
19. Sunil N, Tejaswini K, Lakshmi BB, Nani C, Gurudatta P, Ashok Kumar, et al. RNN-LSTM based regular health factor analysis in medical environment. *Int J Eng Appl Sci Technol* 2025;9:84-87.
20. Vafaei N, Ribeiro RA, Camarinha-Matos LM. Assessing normalization techniques for simple additive weighting method. *Procedia Comput Sci* 2022;199:1229-36.
21. Wang S, Hussien AG, Jia H, Abualigah L, Zheng R. Enhanced remora optimization algorithm for solving constrained engineering optimization problems. *Mathematics* 2022;10:1696.