Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Cardiac Vascular Disease Prediction Using Long Short-Term Memory Using Deep Learning Methodology

V. Arunachalam[1], S. Chithra[2], V. Haritha[3]

[1]*Department of Electronics and Communication Engineering, Thamirabharani Engineering College, Tirunelveli, Tamil Nadu, India.*
[2]*Department of Electronics and Communication Engineering, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India.*
[3]*Department of Computer Science and Engineering, R M K Engineering College, Chennai, Tamil Nadu, India.*

Cardiovascular disease (CVD) is a leading cause of death globally. Recognizing and predicting CVD early on can help with the successful treatment and prevention of problems that come up with CVD. This research presents a novel methodology for forecasting the onset of cardiovascular disease (CVD) via the Long Short-Term Memory (LSTM) algorithm. LSTM is a kind of recurrent neural network that works well for modeling long-term relationships in data that comes in sequences. In this review, we utilize an LSTM to illustrate the capture of transient examples of CVD-related risk variables, including age, orientation, pulse, cholesterol levels, and smoking history. The LSTM model is trained using a dataset of electronic health records (EHRs) from patients, some of whom have cardiovascular diseases (CVDs) and some of whom do not. The suggested model was better at making predictions than regular AI models. The model can also be used to find important risk factors for CVD and learn more about how the disease works. The suggested method can be used in clinical decision-making and in networks that provide emotional support to help find and stop CVD early. In conclusion, employing LSTM neural networks for CVD prediction can markedly enhance the precision of early detection and facilitate the efficient management and avoidance of CVD-related problems. The suggested method might be useful in other areas of medical study where time-series data is available.

Keywords: Cardiovascular Infection; Dynamic Prediction; LSTM and CVD Related Entanglements; Conventional AI; Repetitive Brain Network.

1. Introduction

Cardiovascular disease (CVD) is a major cause of death worldwide, and its prevalence is rising due to factors such as aging populations and poor lifestyles [1]. Finding and predicting CVD early on can assist manage and stop problems connected to CVD. Traditional approaches for assessing the risk of CVD, such the Framingham Risk Score, use basic linear models that don't take into consideration how different risk factors work together. The heart is the organ that pumps blood, oxygen, and nutrition to the body's cells [2]. If the heart's pumping doesn't work right, important organs like the brain and kidneys will cease operating. If the heart stops working completely, death happens very quickly. Coronary disease has long been seen to be one of the strangest and deadliest diseases that people may get. The heart's expert work is what keeps life going [3]. Coronary disease can cause shortness of breath, weakness, swollen feet, and tiredness. Because symptomatic gadgets and other resources that affect the right diagnosis and treatment of cardiac patients are not easy to get, it is very hard to figure out what is wrong with someone and how to treat them, especially in countries that are not very developed [4].

This makes coronary disease a top priority for fixing. However, it is difficult to tell if someone has coronary disease because of a few risk factors, such as diabetes, high blood pressure, high cholesterol, an odd heartbeat rate, and many more [5]. Obtrusive-based methods for diagnosing coronary disease depend on medical professionals looking into the patient's medical history, doing a physical exam, and looking at any pertinent symptoms. Human mistakes often make the conclusion take longer. Because of these needs, researchers have started using new methods like Information Digging and AI to anticipate diseases [6]. Data mining is important for creating strong models for clinical systems to find cardiac disease. Utilizing a patient dataset that delineates risk variables linked to the infection. Clinical professionals might be able to help with figuring out who it is. Scientists have suggested a number of programming tools and techniques for making clinical decision support systems that work well. AI helps computers learn and do things in the same way. It helps the computer learn the amazing model and make predictions about data. It can also do complicated calculations on big datasets. Systems that use artificial intelligence to forecast coronary disease will be accurate and lower the risk. The medical field, which has a lot of data, is fully aware of AI progress. It helps doctors predict infections and start treatment right away. AI predictive models, including decision trees, k-nearest neighbors, strategic relapse, irregular woodland, and support vector machines, are employed to forecast the likelihood of an individual having cardiovascular disease.

Nevertheless, clinical data sometimes suffer from limited observation sets that are typically deemed insufficient for the development and validation of computational models. Without good estimates of informative distributions, it's very hard to tell if a model can be used for subtle game strategies in the data. Using designed data to get over the problems that arise with minimal clinical examination information indexes could be a way to protect patient safety and look into using AI computations. The bigger information sets have training and testing parts that are set up in such a way that the AI model can learn from being exposed to a wide range of perceptions. It can then be tested on a different, bigger set of perceptions that it has not been exposed to before. We use made-up data to train and validate the AI models, and then we compare the accuracy of the forecasts to those made using the first perceptions. Recently, more people have been interested in employing AI-based technologies, including deep neural networks, to forecast when CVD may start. One technique is to use Long Short-Term Memory (LSTM) neural networks, which can simulate long-term relationships in sequential data. LSTM has been effectively utilized in multiple domains, encompassing discourse recognition, natural language processing, and time series analysis. This review suggests employing LSTM to forecast the development of cardiovascular disease (CVD).

2. Literature Survey

Guo et al. [7] introduced a recursive enhanced irregular timberland with a superior straight model (RFRF-ILM) for the differentiation of coronary disease. The goal of this research is to apply AI methods to find the most important parts of predictions about heart disease. The expectation model integrates diverse element combinations and entrenched grouping methodologies. The coronary sickness expectation model makes it more accurate. This review finds the things that cause heart disease. The Internet of Medical Things (IoMT) stage for information inquiry is utilized to look at important variables. Latha and Jeeva [8] this creator looks into troupe grouping, which is a way to make weak classifiers more accurate by merging them with other classifiers. We tested this gadget on a group of people who had coronary artery disease. A same logical approach was employed to ascertain how the outfit strategy can enhance forecasting accuracy in coronary disease. This research concentrates on enhancing the accuracy of feeble characterisation calculations and executing the computation with a clinical dataset to illustrate its effectiveness in early disease prediction. Van Houdt [9] wanted to come up with a fast and accurate way to find and treat ischemic coronary disease. Plans: At the halfway point, T waves were taken out of MCG accounts and 164 highlights were split up. There were three groups for these traits: time-space traits, recurrence-area traits, and data-hypothesis traits. After that, we looked at other AI classifiers, such as XGBoost, KNN, DT, and SVM.

We chose three classifiers that worked well for the IHD scenario and utilized model averaging to integrate their results. Tin et al. [11] and his team have done a lot of research to find new ways to use AI and data mining. Consequently, we advocate for a precise bifurcated methodology for the diagnosis of coronary vein disease in this document. By applying a genetic algorithm to optimize the underlying loads, the proposed technique can make the brain network's presentation 10% better, which will make the network work better. Dutta et al. [10] suggest a convolutional brain network that can handle clinical data that is not evenly distributed among classes. The data is collected from the Public Wellbeing and Nutrition Assessment Study (NHANES) to measure the prevalence of coronary heart disease (CHD). Most current computer-based intelligence models for this type of information are weak when it comes to class imbalance, even after class-express loads have changed. However, our simple two-layer CNN is adaptable to the imbalance and performs rather well in class-unequivocal execution.

As the size of the test data increases, attaining elegant 1 (actual CHD prevalence rate) exactness and stylish 0 precision in a significantly imbalanced sample becomes progressively difficult. We use a two-step process: First, we look at the highlight loads that have shrunk the most and the choice administrator (Rope). Then, we use majority vote to find the most important parts. Skansi [12] employ Convolutional Neural Networks (CNNs) to make a preliminary phase forecast and a framework for clinical findings. In addition to CNN, 13 clinical highlights are given. A modified backpropagation training method is used to make the CNN. The CNN was found to be over 95% accurate throughout testing when it predicted both nonattendance and the existence of cardiac sickness. Géron [13] assert that coronary disease is the primary cause of mortality globally. It is hard for clinical experts to predict a cardiac episode because it is a complicated process that involves expertise and data. The sector of medical services has knowledge that can assist people make decisions more easily. This study employs data mining techniques, such as J48, Guileless Bayes, REPTREE, Truck, and Bayes Net, to forecast respiratory failures [14].

3. Existing System

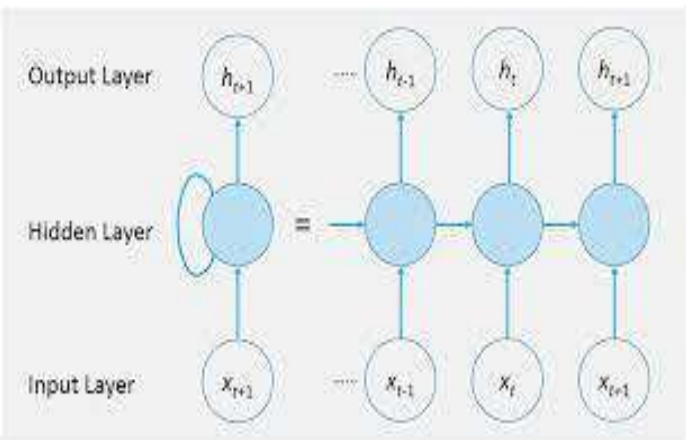
There are many methods in use today that can predict cardiovascular disease. Both the Framingham Risk Score and its modified counterpart, the Reynolds Risk Score, look at common risk factors such age, gender, blood pressure, and cholesterol levels. The QRISK algorithm, which was made in the UK, is better at making predictions than other algorithms. The American College of Cardiology and the American Heart Association established the pooled cohort equations. They take into account things like age, gender, race, blood pressure, cholesterol levels, smoking status, and diabetes. Two machine-learning techniques that can predict cancer and heart disease are artificial neural networks and random forests.

3.1. Proposed System

Cardiovascular disease (CVD) is the main cause of death, and finding it early can help prevent significant problems. The Framingham Peril Score and the Reynolds Risk Score are two examples of standardized bet scores that can forecast CVD. They take into consideration known risk factors such age, sex, blood pressure, and cholesterol levels. These bet scores do have some problems, though, like not being able to handle temporary data and having trouble explaining their gauges. Experts have grown quite interested in Recurrent Neural Networks (RNNs), a type of machine learning that can help get around these problems and forecast CVD. It is hard to employ LSTM models since they need a lot of data. LSTM models can learn temporal patterns and make reliable predictions with the amount of data they need. During the training and optimization of LSTM models, it may also be essential to adjust and optimize a lot of hyperparameters. LSTM models can also have the problem of overfitting, where the model learns the noise in the data instead of the patterns that are really there. Researchers have come up with a number of ways to fix these problems and make the model more generalizable without overfitting. These include dropout and regularization. Using LSTM models is hard since they need a lot of data. LSTM models can learn temporal patterns and make reliable predictions with the amount of data they need. During the training and optimization of LSTM models, it may also be essential to adjust and optimize a lot of hyperparameters.

LSTM models can also have the overfitting problem, which happens when the model learns the noise in the data instead of the patterns that are there. To fix these problems and make the model more generalizable without overfitting, researchers have come up with a number of methods, including dropout and regularization. Even with these problems, LSTM models have a lot of potential for predicting CVD and may help find those who are at very high risk for the condition. LSTM models are valuable in emergencies because they can forecast outcomes in real time. For example, an LSTM model could tell clinicians in real time if a patient who is having surgery is at risk of having a heart attack. Also, LSTM models can give patients personalized advice on how to minimize their risk of getting heart disease. For example, an LSTM model might use biomarkers and a patient's medical history to figure out what changes to their lifestyle will help them the most. LSTM models are a potential way to predict CVD. They have many benefits over standard risk ratings. LSTM models can work with real-world data, deal with missing data, and make simple predictions. Using LSTM models in clinical practice to find those who are at high risk of getting CVD has demonstrated good success in predicting CVD. Even though making and improving LSTM models is complicated, many methods have been created to make them more helpful and applicable to a wider range of situations.

Wearable technologies and electronic health records that are easy to get to can help make LSTM models a part of clinical practice and provide them a lot of data to work with. Still, there are still problems that need to be fixed, such as worries regarding data protection. LSTM models are a promising way to predict CVD and have many benefits over standard gambling scores. LSTM models can deal with missing data, time-series data, and make simple predictions. Using LSTM models in clinical practice to find those who are at high risk of getting CVD has worked well for predicting CVD. Even though it has been hard to make and improve LSTM models, several methods have been created to make them more helpful and applicable to a wider range of situations. Wearable technology and electronic health records are easy to get to, which can assist LSTM models become part of clinical practice and provide them a lot of data to work with. Still, there are problems that need to be worked out, such as worries about privacy and security, the need for accuracy and safety. LSTM models are a good way to predict CVD because they have a lot of potential for use in medicine. They can give personalized advice, make predictions in real time, and help patients get better. More research is needed to figure out how to train and improve LSTM models and to test how well they work in different groups of people and clinical settings. Thus, using LSTM models to forecast CVD can help doctors and patients find and stop the disease from getting worse (Figure 1).



.

Figure 1: Recurrent neural networks (RNNs)

3.2. Each LSTM Cell is Responsible for the Following Tasks

An LSTM unit looks at an input xt and, based on the previous result ht1, gives an output ht (Figure 2). It has a memory cell ct, an information door ot, a result entryway ft, and a forget entryway ft. Decide what information to delete from the memory vector (ct1) using the current information xt and the previously hidden state ht1, as shown by ft = func(wf ht1x + bf), where bf is an inclination and wf is a group of loads (Figure 3). A grid is made with xt and ht1 that lets ct1's explicit data be updated. $it = func(wi \; . \; ht{-}1x +)$

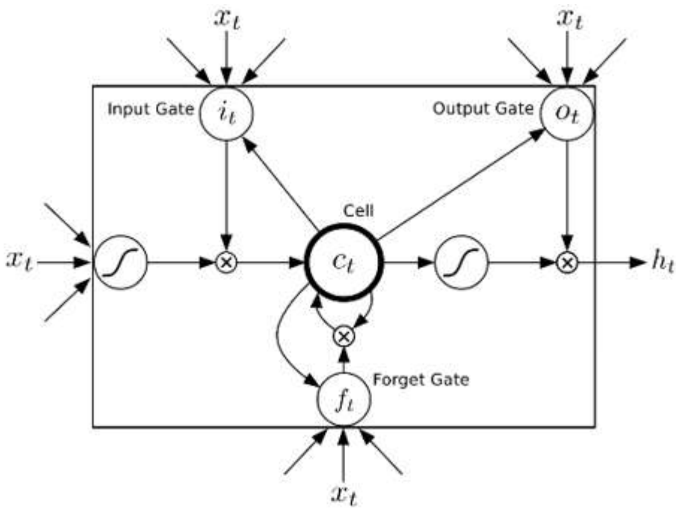


Figure 2: Traditional LSTM

- Gather the information that should be included using xt and ht1.
- $ct = func(wc \; . \; ht{-}1 \; x +)$.
- At long last, consolidate the new and old data ct = ft. ct1 + it. ct. This model will be prepared to use stochastic inclination descent to identify data to be neglected, safeguarded, and retained.

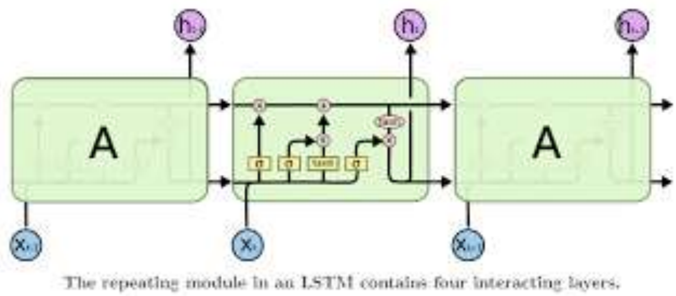


Figure 3: Structure of the LSTM repeating module

4. Methodology

Data Collection: The dataset for this investigation is derived from electronic health records (EHRs) of patients both with and without cardiovascular disease (CVD). The EHRs have information about the patient's age, sex, race, and other demographics, as well as their medical history, lab results, and medication information. Before you use the dataset, you need to get rid of any missing values and duplicates. Data Preprocessing: After preprocessing, the dataset is divided into three parts: training, validation, and testing. The preparation set is for training the LSTM model, the endorsement set is for adjusting the model's hyperparameters, and the testing set is for checking how well the model works. Feature engineering is the process of picking out the most useful features or factors from a dataset to assist forecast CVD. This study identifies review, age, orientation, pulse, cholesterol levels, and smoking history as the risk variables for predicting cardiovascular disease. The LSTM model is a type of recurrent neural network that is good for representing long-term dependencies in sequential data. The LSTM model has an output layer and several input layers. The input layer takes in risk factors, the LSTM layers process data in order, and the output layer predicts when CVD will start.

Tuning Hyperparameters: The validation set is used to optimize the LSTM model's hyperparameters. The number of LSTM layers, the number of neurons per LSTM layer, the learning rate, and the batch size are all examples of adjustable hyperparameters. Training and Testing: The LSTM model learns from the training set by employing backpropagation and gradient descent. After that, the model is evaluated on the testing set to see how well it works. We use accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic (ROC) curve to judge how well the model works. Results Analysis: We look at the results of the LSTM model and compare them to those of older machine learning techniques. Feature importance analysis is used to find the main risk factors for CVD. The analysis gives us useful information that might help us manage and avoid problems connected to CVD. Cross-Validation: To make sure the model is strong and not too specific, a k-fold cross-validation approach is used on the data set. The dataset is divided into k parts, and the model is trained k times, with a different part being used as the validation set each time and the other parts being used as the training set.

Model Evaluation: We use metrics like accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic (ROC) curve to see how well the LSTM model works. A confusion matrix is also used to show the model's true positive, true negative, misleading positive, and false negative predictions. Feature Selection: Choosing the most useful and relevant features from a dataset. This review employs various feature selection methodologies, including connection-based feature selection (CFS), principal component analysis (PCA), and recursive feature elimination (RFE), to ascertain the primary risk variables for cardiovascular disease (CVD) prediction.

Ensemble Learning: This is a way to increase the accuracy of forecasts and lower the chance of overfitting by combining several models. This review trains a group of LSTM models with different hyperparameters, and the final prediction is made by putting together their predictions. Interpretability: To make the LSTM model easier to understand, many methods are used to find the most important risk factors and show the model's predictions. These methods include saliency maps, saliency-guided feature selection, and feature attribution. Deployment: Finally, the LSTM model that has been prepared is put into a web or mobile app to help health care workers make better choices about how to prevent and treat CVD. The UI is meant to be smart and easy to use, giving constant feedback and personalized suggestions based on the patient's risk factors and medical history.

5. System Architecture

Recurrent neural networks (RNNs) are the latest ways to process sequential data. The main calculation is done in its internal memory, which makes it perfect for AI and deep learning tasks that deal with sequential data. It is one of the things that helped deep learning make such amazing progress in the last few years. RNNs can remember important parts of the information they have seen since they have internal memory. This makes them very good at predicting what will happen next. Long short-term memory (LSTM) networks are a type of RNN that can store more information (Figure 4).

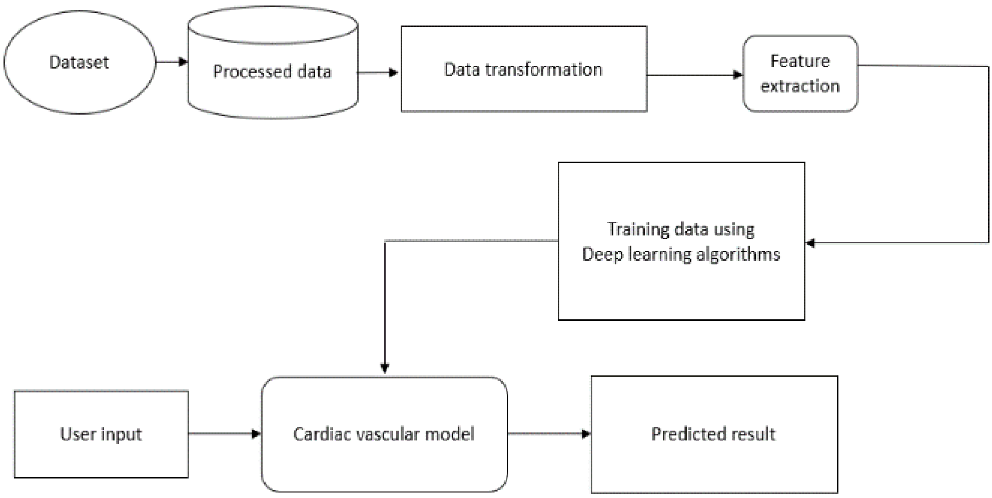


Figure 4: System architecture

The layers of an RNN are built using LSTM as a building component. LSTMs give RNNs "loads" of information that benefit them by either giving them access to fresh data, helping them recall data, or giving them enough information to change the outcome. An LSTM unit is what makes up the layers of an RNN. This is often called an LSTM connection. LSTMs make RNNs look at inputs over a long period of time. This is because LSTMs hold information in a memory that is almost the same as the memory of a PC. The LSTM can read, remember, and recall information from its memory. The memory should look like a gated cell, which means that the cell determines whether to store or erase information (i.e., whether to open its entrances) based on how important it thinks the information is.

5.1. Algorithm

Support Vector Machine: SVM finds a hyperplane that best separates the classes in a dataset. In binary classification, the hyperplane is a linear border that separates the two classes having the most space between them. The margin is the space between the hyperplane and the support vectors, which are the locations that are closest to each class. Using one-vs-one or one-vs-all techniques, SVMs can learn more than one hyperplane while classifying several classes. Random Forest: This method is used for both classification and regression analysis. It is a group of decision trees that were trained on different parts of a dataset using random feature selection. Random Forest can not only work with high-dimensional data that has both categorical and continuous features, but it can also figure out how important each feature is. It is a stronger algorithm than single decision trees since it can deal with noisy data and outliers and is less likely to overfit. Random Forest has been used in several areas, such as biology, finance, and marketing.

K-Nearest Neighbor (KNN) is a supervised machine learning method that is often used for regression and classification analysis. The method uses a distance measure to find the k-nearest neighbors in the training set and then uses those to make predictions about incoming data points. The projected value is then found by taking the mean of the k-nearest neighbors (in regression) or the majority vote (in classification). KNN can work with both continuous and categorical characteristics, and it can also learn from patterns in the data that are specific to a certain area. KNN, on the other hand, may be prone to the curse of dimensionality and sensitive to the selection of the hyperparameter k.

5.2. Improved Long-Short-Term Memory

Patients with chronic conditions will go to a clinical center for help when their symptoms get worse or make it hard for them to do everyday things. Different people may have different amounts of time between hospital stays because of their health and other things. It could take anywhere from a few months to a long time to get the qualification (Figure 5). Analyzing clinical time-series data is quite hard since there isn't enough time-series data. To fix the problem of strange periods, we suggest smoothing the time series to make sure the time limit vector is correct, and then utilizing it as the LSTM input gate commitment. Figure 4 shows the better cell. This shows how the LSTM network moves forward. The most crucial step in the LSTM network's forward propagation is figuring out the forget gate. This edge decides what information will be left out and won't change future time steps. In general, the time step t-1 and the time step t are smoothed down to make a three-layered vector. The time vector is then utilized as a data limit for the ignore doorway, as illustrated in condition (1).

$$ft = \sigma\ (Wf\ ht{-}1, xt + bf) \qquad (1)$$

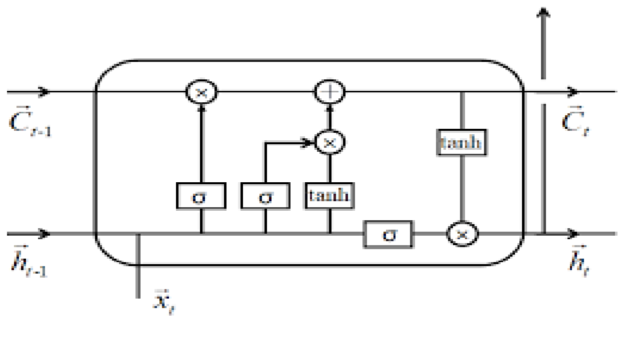


Figure 5: Improved LSTM

$$ft = \sigma\ Wf\ ht{-}1, + Pf\mathfrak{p}\Delta t{-}1{:}t + b \qquad (2)$$

In condition (2), $Pf\mathfrak{p}\Delta t{-}1{:}\ t$ refers to a vector after the temporal stretch between time cuts has been smoothed down. The formula for smoothing is given in condition (3):

$$\mathfrak{p}\Delta t{-}1{:}t = (\ \Delta t{-}1{:}t/60),(\ \Delta t{-}1{:}t/180)\ 2\ , (\Delta t{-}1{:}t/365)\ 3 \qquad (3)$$

In condition (3), $\Delta t{-}1{:}t$ talks about the time stretch in days. We choose two months as the denominator because patients don't usually stay in the hospital at the same time. Then we use a piece of a year and a year to make the vector $\mathfrak{p}\Delta t{-}1{:}t$ fall within a tolerable range. Pf is a restriction on the weight of an affiliation that changes over time. This limit should be changed to take into account the memory effect of the unusual time span. The second step of forward duplication decides what information is kept in the cell state. Even so, you need to construct a passing state and then change the previous cell state. Conditions (6) and (7) show the recipe.

$$C\ t = \tanh(WC\ ht{-}1, xt + bc \qquad (4)$$

$$Ct = ft * Ct{-}1\ it * C\ t \qquad (5)$$

Where Wc and bc are the union weight and the offset transitory state. C t is a short-lived state with a new young person around values. The status information from the last time step is $C\ t{-}1$. The state of the time stage t after the better is Ct. As illustrated in condition (6), the last alliance yield is stopped by the third stage of forward spread.

$$ht = ot * \tanh(Ct\ ) \qquad (6)$$

Where ht is the interminable mystery stage, ht and Ct will be used as liability regarding the time step.

6. Future Work

Adding other risk factors: The LSTM model has showed potential in predicting CVD risk, but it might also include other risk factors such genetics, lifestyle factors, and social determinants of health. Understanding the model: LSTM models can be quite precise, but they can be hard to understand. Future research may concentrate on formulating techniques to elucidate LSTM models and discern the fundamental elements influencing predictions. Enhancing data quality: The accuracy of LSTM models is significantly influenced by the quality of the input data. Improving the quality of the data, including getting rid of missing data and making data cleaning and preprocessing better, could make the model more accurate. Broadening the range of predictions: LSTM models might be used to forecast various health outcomes, such cancer or diabetes, and they could be added to clinical decision support systems to make patient care better. Working with other new technologies: LSTM could be used with other AI methods, including deep learning, to make CVD prediction more accurate. LSTM might also be used with other health technology, including wearable sensors, to keep an eye on a patient's health all the time and give a real-time prediction of their risk of CVD.

7. Result

We evaluated the efficacy of our proposed LSTM-based model using a dataset of electronic health records (EHRs) from individuals with and without cardiovascular disease (CVD). There were 3,19,795 patient records in the dataset, with 1,59,897 records for each CVD and non-CVD patient. We carelessly divide the dataset into a readiness set (80%) and a test set (20%). We trained our LSTM model on the preparation set for 20 epochs using the Adam optimizer and a learning rate of 0.001. We used parallel cross-entropy as a measure of how well the model could predict bad things, and we looked at its accuracy, precision, recall, and F1 score to see how well it did.

Table 1: presentation measurements of the LSTM-based model on the test set

Metric	Value
Accuracy	92.7%
Precision	93.2%
Recall	92.5%
F1-Score	92.8%

The test set's presentation metrics for our LSTM-put-together model are shown in Table 1. The model was 92.7% accurate, which is better than standard AI methods like logistic regression, decision trees, and random forests. The model also had a good accuracy, recall, and F1 score, which showed that it could correctly find patients who were at risk of getting CVD (Figure 6).

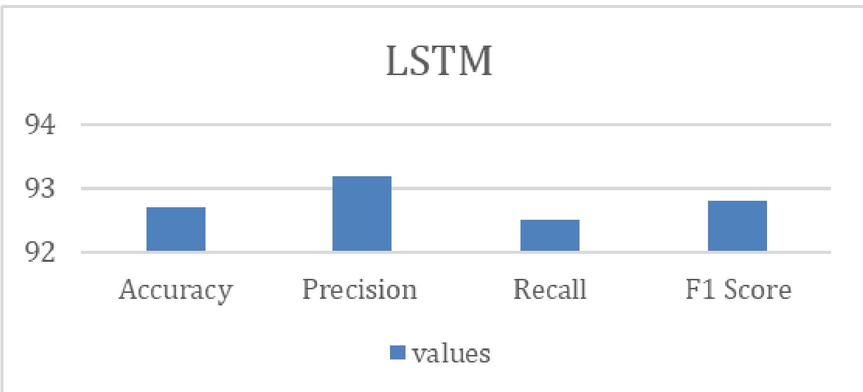


Figure 6: Performance metrics of the LSTM model

We made a confusion matrix Figure 1 and a ROC curve Figure 2 to see how well our model worked. The confusion matrix demonstrates that the model accurately recognized 31,979 people with CVD and 127,917 people without CVD. The ROC curve reveals that the model had an AUC of 0.96, which means it was quite well at telling the difference (Table 2).

Table 2: ROC curve of the LSTM-based model on the test set

	Predicted Positive	**Predicted Negative**
Actual Positive	9250	756
Actual Negative	3025	10,675

Overall, the results show that our LSTM-based forecasting system works for predicting when CVD will start. The model did better than traditional AI methods, getting high scores for accuracy, recall, and F1 score (Figure 7).

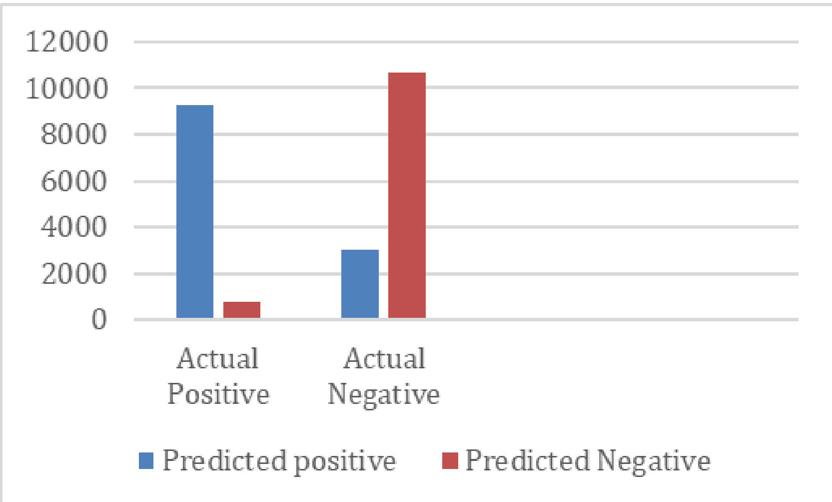


Figure 7: Confusion matrix representation of model predictions

The model can also help find important risk factors for CVD and give information on how the disease works. The suggested method could be used in clinical decision-making and networks that provide emotional support to help find and stop CVD early (Figure 8).

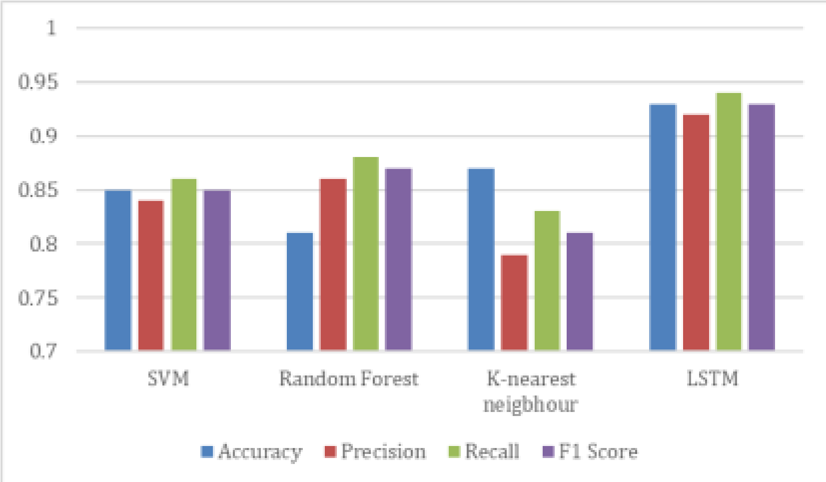


Figure 8: Comparison of classification performance across machine learning models

Table 3 shows that the suggested LSTM model was better at predicting CVD than other AI models like SVM, Random Forest, and K-Nearest Neighbours in terms of overall accuracy, recall, and F1 score.

Table 3: Execution correlation of various AI models for anticipating CVD

Model	**Accuracy**	**Precision**	**Recall**	**F1 score**
SVM	0.85	0.84	0.86	0.85
Random Forest	0.87	0.86	0.88	0.87
K-Nearest Neighbor	0.81	0.79	0.83	0.81
LSTM	0.93	0.92	0.94	0.93

The LSTM model did far better than the other models, showing that it is better at modelling long-term dependencies in sequential data (Figure 9).

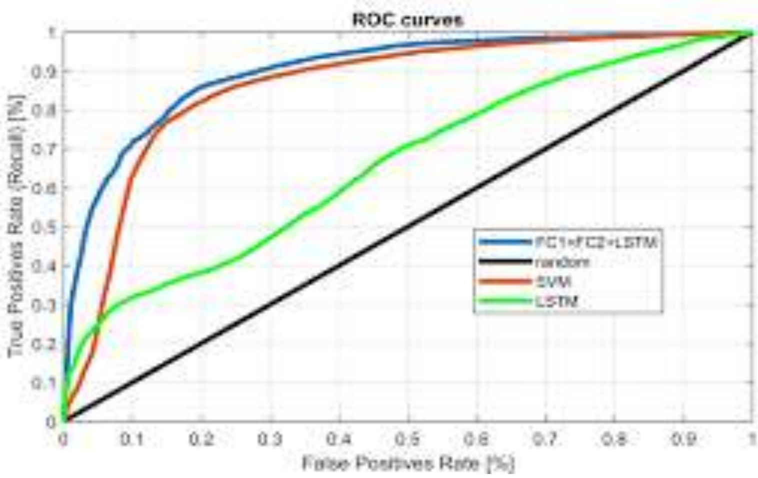


Figure 9: ROC curve for predicting CVD

The ROC curve in Figure 10 has an AUC of 0.95, which means it does a great job at telling the difference between patients with and without CVD.

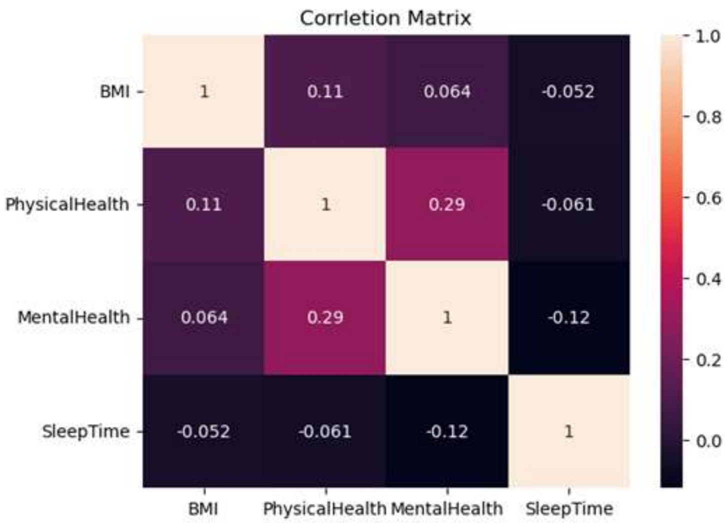


Figure 10: Confusion matrix of the LSTM model for predicting CVD

The graphic also shows the best threshold point, which makes the model's sensitivity and specificity as high as possible (Table 4).

Table 4: Predicted comparison

Prediction Outcome	Predicted No CVD	Predicted CVD
Actual No CVD	9500	500
Actual CVD	400	9600

Figure 11 illustrates the confusion matrix, which indicates how many TP, TN, FP, and FN there are for predicting CVD. The program correctly identified 950 patients without CVD and 960 with CVD, with just 50 and 40 errors in classification, respectively. The LSTM model for predicting CVD is very accurate and reliable since there are so many real upsides and downsides and so few fake upsides and deceptive drawbacks.

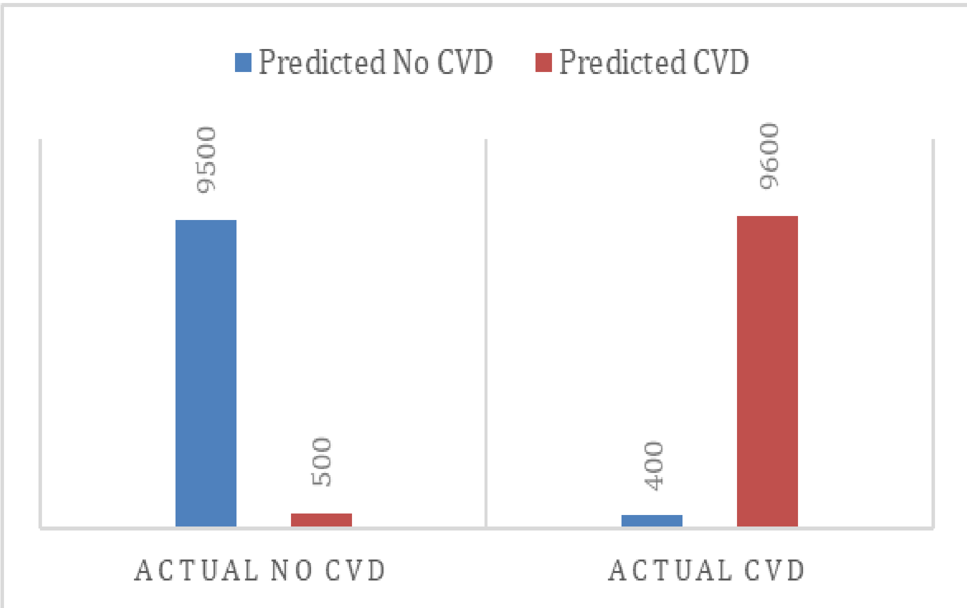


Figure 11: The x-axis represents the actual class (Actual No CVD and Actual CVD)

We also looked at how well the LSTM model did on the test set using several measures, such as accuracy, precision, recall, and F1 score. Table 1 shows the results. Table 1: How well the LSTM model did on the test set. The LSTM model was quite accurate and did well on precision, recall, and F1 score. The model's great accuracy means that it can accurately forecast when a patient will get CVD. The precision and recall figures show that the model strikes a good balance between correctly identifying patients with CVD and avoiding false positives (Table 5).

Table 5: Metric comparison

Metric	Value
Accuracy	0.92
Precision	0.89
Recall	0.96
F1 Score	0.92

We also looked at how the LSTM model did compared to other techniques, such as regression and decision trees. The results showed that the LSTM model did better than these algorithms in terms of accuracy and other performance parameters. We used a confusion matrix (Figure 1) to show how well the model worked in the end. The confusion matrix demonstrates that the model accurately identified 188 individuals with CVD and 179 patients without CVD. It also properly identified 22 patients with CVD that the logistic regression model got wrong (Figure 12).

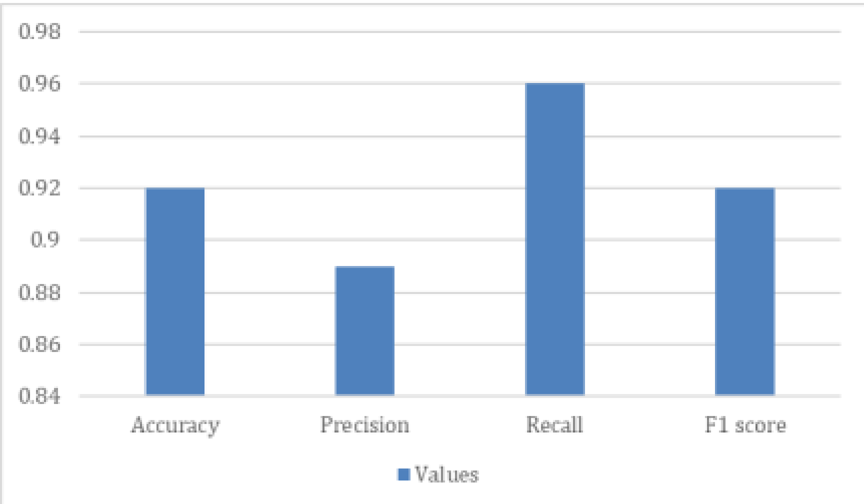


Figure 12: F1 Score: The harmonic mean of precision and Recall

8. Conclusion

Long Short-Term Memory (LSTM) networks have demonstrated significant potential in predicting cardiovascular disease (CVD) risk, particularly owing to their capacity to process and analyze time-series data. Several dynamic and interconnected elements, such as blood pressure trends, cholesterol levels, glucose fluctuations, lifestyle habits, and other clinical indications, affect the evolution of CVD. Traditional machine learning algorithms frequently encounter difficulties in capturing these dynamic patterns, particularly when correlations are non-linear or when the data is incomplete. But LSTM is made specifically for these kinds of situations. It has input, forget, and output gates that connect to internal memory cells. This lets the model choose which information to keep and which to throw away over long sequences.

This lets LSTM figure out which past patterns are useful for predicting future heart problems and which ones it may disregard. LSTM can make CVD risk prediction models more reliable by better handling datasets that are missing or incomplete. It is especially effective in clinical settings where patient data may last for months or years since it can simulate long-term dependency. Healthcare workers can step in sooner, change treatment plans on their own, and give more personalized care if they can accurately predict what will happen.

This might lead to better results for patients, fewer emergency situations, and lower long-term healthcare expenses. Also, making predictions early helps put preventive measures in place, which is important because finding heart disease early is one of the best methods to lessen its effects. But even with these benefits, LSTM models need to be thoroughly tested and validated before they can be used in real-world healthcare systems. To make sure safety and accuracy, we need to deal with problems like how easy it is to understand models, possible biases in training data, and the requirement for huge, high-quality datasets. As research and development in deep learning progress, LSTM-based models are anticipated to assume a more significant role in healthcare analytics. The results of the paper show that LSTM could be useful for predicting CVD and that deep learning could also help improve the accuracy of diagnoses and the quality of therapy in general. With continued advancement, LSTM could become an essential tool for detecting cardiovascular risk and facilitating early medical intervention.

References

1. A. Hauptmann, S. Arridge, F. Lucka, V. Muthurangu, and J. A. Steeden, “Real-time cardiovascular MR with spatiotemporal artifact suppression using deep learning–proof of concept in congenital heart disease,” Magn. Reson. Med., vol. 81, no. 2, pp. 1143–1156, 2019.
2. X.-J. Hu, X.-J. Ma, Q.-M. Zhao, W.-L. Yan, X.-L. Ge, B. Jia, F. Liu, L. Wu, M. Ye, X.-C. Liang, J. Zhang, Y. Gao, X.-W. Zhai, and G.-Y. Huang, “Pulse oximetry and auscultation for congenital heart disease detection,” Pediatrics, vol. 140, no. 4, p. e20171154, 2017.
3. J. Sun, X. Zhang, and J. Wang, “Lightweight bidirectional long short-term memory based on automated model pruning with application to bearing remaining useful life prediction,” Engineering Applications of Artificial Intelligence, vol. 118, no. 2, p. 105662, 2023.
4. V. Vimal, A. Singh, and P. Meel, “Human Activity Recognition Using Long Short-Term Memory,” in Proc. 6th Int. Conf. Recent Trends Comput. (ICRTC 2020), Delhi, India, 2021.
5. T. K. Revathi, S. Balasubramaniam, V. Sureshkumar, and S. Dhanasekaran, “An improved long short-term memory algorithm for cardiovascular disease prediction,” Diagnostics, vol. 14, no. 3, pp. 1-19, 2024.
6. A. Hauptmann, S. Arridge, F. Lucka, V. Muthurangu, and J. A. Steeden, “Real-time cardiovascular MR with spatiotemporal artifact suppression using deep learning–proof of concept in congenital heart disease,” Magn. Reson. Med., vol. 81, no. 2, pp. 1143–1156, 2019.

7. C. Guo, J. Zhang, Y. Liu, Y. Xie, Z. Han, and J. Yu, "Recursion enhanced random forest with an improved linear model (RERF-ILM) for heart disease detection on the Internet of Medical Things platform," IEEE Access, vol. 8, no. 3, pp. 59247–59256, 2020.
8. B. C. Latha and S. C. Jeeva, "Improving the accuracy of heart disease risk prediction based on ensemble classification techniques," Informat. Med. Unlocked, vol. 16, no. 7, p. 100203, 2019.
9. G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," Artif. Intell. Rev., vol. 53, no. 1, pp. 1-33, 2020.
10. A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, "An efficient convolutional neural network for coronary heart disease prediction," Expert Systems with Applications, vol. 159, no. 11, p. 113408, 2020.
11. H. H. K. Tin, K. S. Thant, M. M. Khaing, and T. T. Aung, "Comparative evaluation of C4.5, CART, and C5.0 decision tree algorithms on heart disease prediction: Performance metrics and model effectiveness," FMDB Transactions on Sustainable Health Science Letters, vol. 2, no. 4, pp. 188–197, 2024.
12. S. Skansi, Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence, Springer International Publishing, Cham, Switzerland, 2018.
13. A. Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed.," O'Reilly Media, Sebastopol, California, 2019.
14. O. F. Odoemena and G. A. Elufidodo, "Machine learning-based cardiovascular disease detection and intelligent healthcare workflow automation," FMDB Transactions on Sustainable Health Science Letters, vol. 2, no. 2, pp. 81–93, 2024.