International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Cardiovascular Disease (CVD) Prediction using Deep Learning Algorithm

Sarita Rathi[1], Vijay Gadicha[2]

[1]PG Scholar, Department of Computer Science & Engineering, G H Raisoni University, Amravati, Maharashtra, India

[2]Professor, Department of Computer Science & Engineering, G H Raisoni University, Amravati, Maharashtra, India

Abstract: ***Brain and Heart both organs keep our body working with co-ordination comparing to brain heart is very delicate and need more care for well-functioning. So heart is essential inevery living body. Pumping of oxygenated blood and supplies it to all organs in proper flow makes our body well operated. So keeping our heart healthy and disease free is our priority. For this propose prediction of genesis of cardiovascular diseases will be very notable work in healthcare sector. Processing on raw data with useful tools and technique to make it significant for prediction of disease at early stage will be useful for healthcare centre to cure disease at its born stage. Enormous disease related data of patients is stored every day in that can be pre-process and can be used to train any machine learning or deep learning model which will help to predict disease at early stage. In this study deep learning algorithms like RNN(Recurrent Neural Network) with its type LSTM(Long short Term Memory) is used to train and predict the cause of CVD so patient can start medication very soon. The proposed work, come up with comparative study of pervious used algorithm and proposed working algorithms with accurate prediction without time delay. The results of the diagnosis model are generated by using classification method, accuracy and severity of causing parameters. This review paper purpose is to develop the model to predict whether patient is having a cardiovascular disease or not and make patient more aware on CVD. The perfection is achieved by using categorization technique, RNN (Recurrent Neural Network) with LSTM to achieve accuracy in prediction.***

Keywords: Cardiovascular Diseases: RNN (Recurrent Neural Network), Long Short-Term Memory (LSTM) networks, CVD (Cardiovascular Disease)

1. Introduction

Cardiovascular diseases (CVDs) consider as big umbrellas of disorders related to heart and blood vessels, including list of diseases which cannot be ignorable. Mark D et, al[9], observed in their survey that nearly 17.9 million people died due to CVDs in 2021, from 31% of all death toll and 85% of death arises due to heart related disorders. As stress is main cause of cardiovascular disease (CVD). Every human is having their own stress reasons, it may be job, family stress, health stress and so many, which leads to rapid. **Cardiovascular Disease (**The big umbrella). As mention above CVD diseases that affect the heart or blood vessels, comprise clogged arteries, leads to heart attacks, **stroke, failure, and peripheral artery disease.**

Researches tried tosspot dominant factor that causes heart disease as well as summaries the risk causing factors. Cardiovascular Disease is even nominated asasilentgunwhichcausethedeathofthepatientwithoutclearsymptoms.The before time prediction of cardiovascular disease plays a vital role in adoption of healthful lifestyle can change high-risk of causing disease in patients and may bring down the complications.

The proposed work aims to predict future Cardiovascular Disease by studying patients' data and based on training model will classify whether they have cardiovascular disease or not using Deep-learning algorithms.

1.1 Problem Definition

The major thing in Cardiovascular disease is its detection at early stage , as eating habit, healthy habits and vegan or non-vegan, major risk factor is smoking, which drastically vary from person to person. There are many devices presently used for prediction of cardiovascular disease, but they are not cost effective and accuracy in results. Early detection of cardiovascular diseases can decrease the mortality rate or death rate and overall problematic conditions. The patients suffering from CVD requires real time consultation and treatment, however monitoring of patient 24*7 is not possible. Hence, a time bounded effective and efficient model is required. To serve medical practitioner and patients, said gap is analyzed using existing literature review. To bridge this, model is proposed is required.

1.2 Objectives

The main objective of paper:

1) Develop a deep learning model to forecast possibility cardiovascular disease by implementation Recurrent Neural Network with LSTM.
2) To resolve significant risk factors based trained dataset which may lead to Cardiac arrest or any severity leads to CVD.
3) To examine feature selection methods and know their working principle as to predict severity of heart failure and cause of CVD.

1.3 Scope

Deep learning algorithms played crucial role in defining the symptomatic pattern to predict cardiovascular diseases (CVDs). Deep learning models have the ability to learn complex patterns and relationships from large datasets, which can be used to identify risk factors and predict the likelihood of developing CVDs.

Some areas of cardiovascular disease prediction that have

been explored using deep learning include:

- **Diagnosis:** Deep learning models have been trained to accurately diagnose CVDs from medical images such as echocardiograms, electrocardiograms (ECGs), and angiograms. This can lead to early detection and treatment, which can improve patient outcomes.
- **Risk prediction:** Neural Network is capable to diagnose patient data such as medical history, healthy habits, stress, heritable information to anticipate the risk of developing CVDs. This can help identify high-risk individuals who may benefit from preventive interventions.
- **Prognosis:** Deep learning models can be used to predict the progression of CVDs and the likelihood of adverse outcomes such as heart failure, stroke, or death. This can help clinicians make informed decisions about treatment and monitoring.
- **Personalized medicine:** Deep learning models can analyze patient data to develop personalized treatment plans that are modify as per the need of patients and risk factors. This can be to more effective treatments and healthier outcomes.

Overall, the scope of cardiovascular disease prediction using deep learning is vast and holds great potential for improving patient outcomes and reducing the burden of CVDs.

2. Literature Review

2.1 Reviews

Vast research study done on various types of Heart related diseases and its prediction techniques using different data mining techniques and machine learning algorithms at infirmary.

1) **Muktevi Srivenkatesh [2]** this research work is based on training different models with same instance of dataset but outcomes from each trained model is different but clearly classification accuracy of Regression algorithm was better compared to other algorithms.
2) **Ngure Ngare,** [4] developed heart disease prediction model using Bayes Net and Multilayer to get optimum performance. The exactitude performances achieved by those algorithms was not satisfactory, so the accuracy's performance is improved more to give better decision to diagnosis disease.
3) **Rohit Bharti ,**et al,[3] used technique that predict chronic disease by mining the data containing in historical health records using different algorithms and Neural Network. A comparative study for better performance using all algorithms were observed ,out of all SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.
4) **Surenthiran Krishnan**,et al, [7] ,based on hybrid deep learning model of RNN and GRU for heart disease prediction, this model has successfully shown increased the prediction accuracy for Heart Disease and gave appropriate bifurcation using categorization technique.

2.2 Summary report

1) **Sai Yeshwanth Chagant** , et al [8] , study said that, SVM is very accurate for small amount of dataset to achieve an accuracy of 93% and if tripled the size of dataset and applied SVM again, perfection reduces to 82%. Unsatisfied with the results, but using deep learning techniques like CNN, the accuracy of a staggering 93.57% for tripled dataset. So without compromising with accuracy of CVD prediction the RNN as well CNN would generate more accurate result compared to other machine learning algorithms.
2) **Migran N**, et al [11] showed the comparison of different machine learning libraries, which shows the accuracy in percentages of different libraries with respect to time in seconds for each machine learning libraries. We reach to conclusion that nearly the accuracy of any library is same but if we compare learning time for each ,like Py-Torch takes longer learning time compare to others.
3) **Vassili Kovalev,** et, al[12], displayed very deep comparison of all libraries as well programming language compatibility, complexity in which implementation of Tensor flow with Keras using python programming language reduce complexity and code lines.

Table 1: Comparison of Algorithms (Source -Surenthiran Krishnan et, al December 2021)

Algorithmic Model	Accuracy (%)	Specificity (%)	Sensitivity (%)
RNN	98.23	97.6	99.1
Back Propagation RNN	98.2	89.65	87.64
RNN	92	---	----
Prognosis prediction using RNN	98	98.8	98.47
Advanced RNN using GRU	98.4	98.4	99.1
Proposed model	98.69	96.31	1.000

Table 1, shows the comparison of performance based on Accuracy, Specificity and Sensitivity using RNN with GRU and other algorithms. Which concludes that the RNN is best finer to predict heart disease?

3. Proposed System

The purpose of CVD prediction is to identify individuals who are at high risk of developing cardiovascular diseases (CVDs) and to take relevant preventive measures to decrease the risk or delay the onset of the disease. CVDs are a major cause of morbidity and mortality widely, and early detection of risk factors can help to take care of causes of CVDs.

CVD prediction can be consider few important factors like risk assessment tools, medical history of family, exercise habits, eating habits and genetic testing. By predicting CVDs, physicians and wellness programmers can develop

personalized prevention instruments so they can modify to the individual's unique needs and risk factors. This can include lifestyle modifications such as diet and exercise, medication therapy, and other interventions to reduce the risk of CVDs.

The purpose of CVD prediction is to identify the risk of CVDs causing factors and to take early preventative steps to prevent from the disease, eventually main aim to improved patient consequences and reduced healthcare costs.

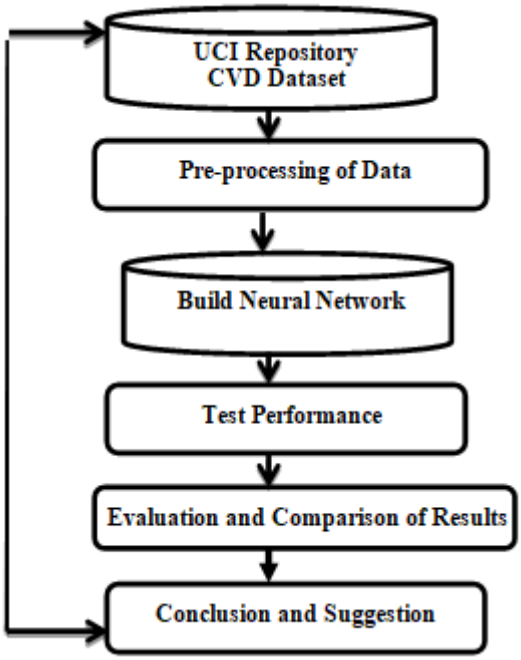

Figure 1: Proposed work

The above (figure : Proposed work) shows the actual flow of this dissertation work, as CVD prediction can be achieved through various means, including risk assessment tools, medical history, lifestyle factors, and genetic testing.

In order to predict and diagnose CVD, here using algorithms, like RNNs (Recurrent Neural Networks), this is type of neural network that can process sequential data and time series data. Whenever we reach classification based on input, it will conclude whether the patient is suffering from CVD or not.

3.1 System Requirements:

- Python Programming Language (version 3.6 or later)
- Tensor Flow (an open-source framework)
- Keras (a high-level API)
- Kaggle Notebook or other development environment for writing and executing code
- Pandas, NumPy, and Scikit-learn and other libraries for data pre-processing and analysis

3.2 Methodology

- **Data Collection:** Collect relevant data related to cardiovascular diseases from various sources such as medical journals, research papers, and public datasets.
- **Data Pre-processing:** Clean the collected data by removing any missing or erroneous values, normalizing the data, and converting it into a format suitable for training the RNN model.
- **Model Selection:** Choose an appropriate RNN architecture to predict cardiovascular disease based on performance metrics such as accuracy and precision.
- **Model Training:** collected data is used to train our RNN model called pre-processed of data using relevant parameters and tuning techniques for optimization.
- **Model Evaluation:** Judge the trained model's outcome on which data get discrete in pre-processing step by evaluating different parameters like accuracy, sensitivity.
- **Prediction:** Deploy the trained model to make predictions on new data and generate insights that can help in preventing cardiovascular diseases

3.3 Data Set Used

The data set is publicly available on the Kaggle.com or any UCI repository. It provides patient time series or sequential dataset with different attributes including age, gender, sysBP, DiaBP, Total Cholesterol , HDL, LDL cholesterol, fasting glucose, smoking status ,alcohol status, exercise etc.Thedatasetisincsv(CommaSeparatedValue)formatwhichi sfurtherprepared to data frame as supported by library in python.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Patient ID	Age	Gender	Systolic BP	Diastolic BP	Total Cholesterol	HDL Cholesterol	LDL Cholesterol	Triglycerides	Fasting Blood Glucose	Smoking Status	Family History of CVD	Alcohol intake	Exercise	CVD Diagnosi
2	1	45	Male	130	80	220	40	150	200	90	Non-smoker	No	No	No	No
3	2	60	Female	150	90	180	50	120	150	100	Smoker	Yes	Yes	No	Yes
4	3	50	Male	140	85	200	45	130	180	95	Non-smoker	No	Yes	Yes	No
5	4	65	Male	160	95	240	35	160	250	110	Non-smoker	Yes	No	No	Yes
6	5	55	Female	135	80	190	60	100	120	80	Non-smoker	No	No	No	No
7	1	45	Male	130	80	220	40	150	200	90	Non-smoker	No	No	No	No
8	2	60	Female	150	90	180	50	120	150	100	Smoker	Yes	Yes	No	Yes
9	3	50	Male	140	85	200	45	130	180	95	Non-smoker	No	Yes	Yes	No
10	4	65	Male	160	95	240	35	160	250	110	Non-smoker	Yes	No	No	Yes
11	5	55	Female	135	80	190	60	100	120	80	Non-smoker	No	No	No	No
12	1	45	Male	130	80	220	40	150	200	90	Non-smoker	No	No	No	No
13	2	60	Female	150	90	180	50	120	150	100	Smoker	Yes	Yes	No	Yes
14	3	50	Male	140	85	200	45	130	180	95	Non-smoker	No	Yes	Yes	No
15	4	65	Male	160	95	240	35	160	250	110	Non-smoker	Yes	No	No	Yes
16	5	55	Female	135	80	190	60	100	120	80	Non-smoker	No	No	No	No
17	1	45	Male	130	80	220	40	150	200	90	Non-smoker	No	No	No	No
18	2	60	Female	150	90	180	50	120	150	100	Smoker	Yes	Yes	No	Yes
19	3	50	Male	140	85	200	45	130	180	95	Non-smoker	No	Yes	Yes	No
20	4	65	Male	160	95	240	35	160	250	110	Non-smoker	Yes	No	No	Yes
21	5	55	Female	135	80	190	60	100	120	80	Non-smoker	No	No	No	No
22															

Figure 2: Sample Data set

3.4 Methodology and Algorithm Used

The main purpose of designing this system is to predict the causes of cardiovascular disease causes at early stage and to get proper medication as well treatment for future betterment. Here, I used Recurrent Neural Network(RNN) as a deep-learning algorithm to train my model and LSTM (Long-Short Term Memory) to get best time series significant variable selection. The figure 3: given below explain the detailed architecture of CVD prediction.

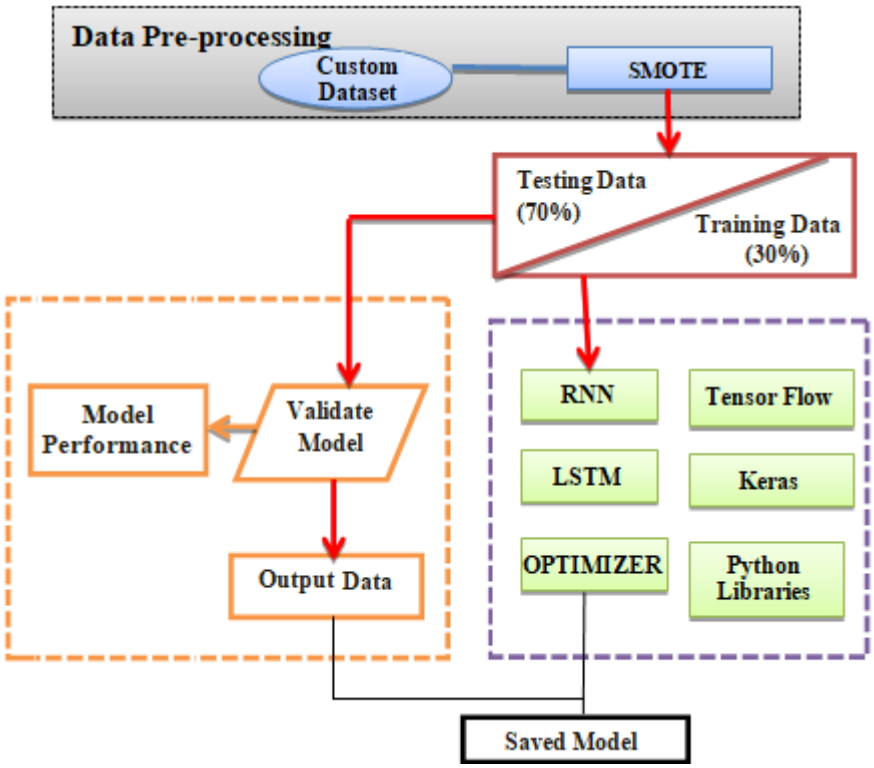

Figure 3: Detailed Working Architecture of CVD Prediction

3.5 RNN (Recurrent Neural Network)

Recurrent Neural Networks (RNNs) is deep learning model which work on time series type of dataset. RNN can be applied to cardiovascular disease prediction using time based data. Time-series data refers to data that is collected over time and includes a sequence of observations. In the case of cardiovascular disease prediction, time-series data can include patient data such as blood pressure, heart rate, and cholesterol levels, which are collected at different points in time.

RNNs can capture the temporal dependencies in sequential data and are well-suited for CVD prediction. Recent studies have demonstrated the potential of RNNs in predicting CVD using time-series data such as electrocardiograms (ECGs) and medical records. One study applied RNNs to predict the risk of CVD using ECG data. The study used a Long Short-

Term Memory (LSTM) network, a type of RNN, to predict the risk of CVD in patients with hypertension. The model achieved an accuracy of 80% in predicting the risk of CVD, which was significantly better than traditional statistical methods.

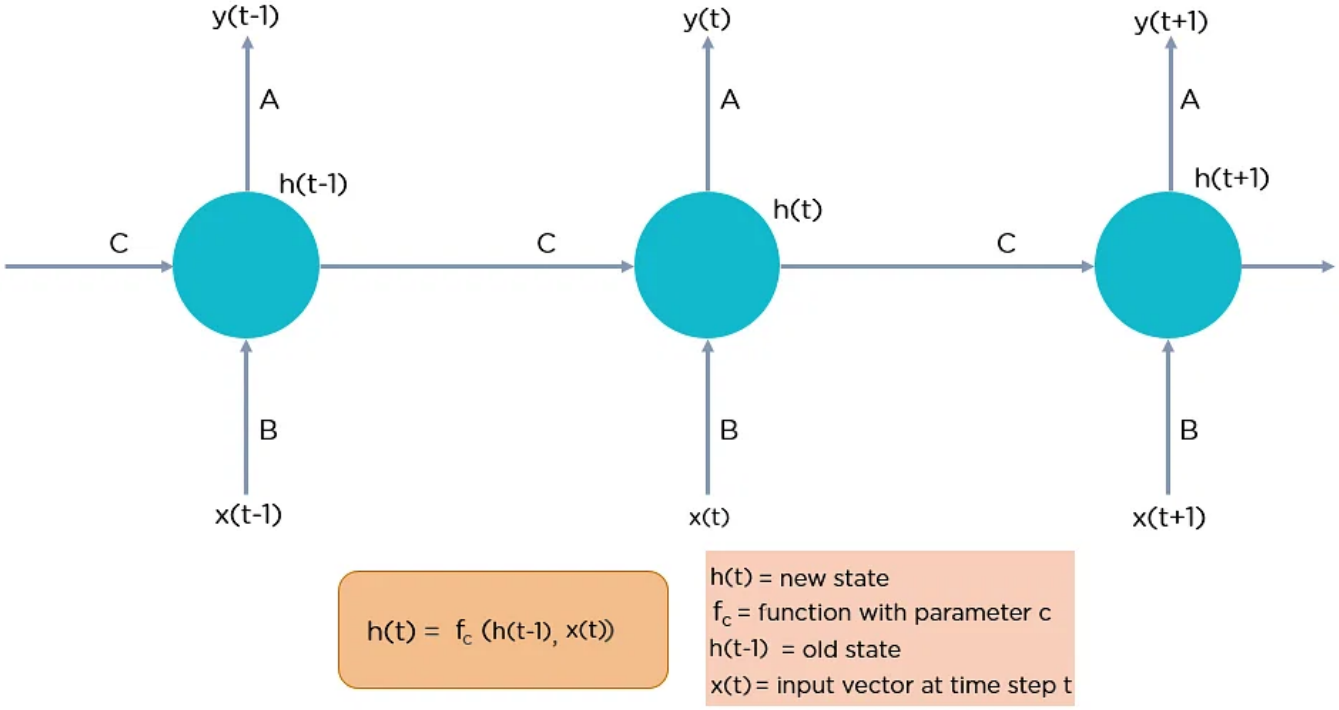

Figure 4: Working of Recurrent Neural Network

In figure 4, the input layer 'x' takes in the input to the neural network which goes through hidden layer which is also refer as middle layer. The hidden layer 'h' can consist of number hidden layers, which has theri own activation functions and weights and biases. In neural network which has different parameters for different hidden layers not affected by the previous layer, because it does not have memory, for that purpose, we can use a recurrent neural network. The RNN will standardize the different activation functions and weights and biases so that every hidden layer has the same parameters. instead of creating multiple hidden layers, it will create one layer and loop over to many times as required.

3.6 LSTM (Long Short- Term Memory)

LSTM is a type of recurrent neural network (RNN) which specially design to resolve vanishing gradient problem in RNNs. LSTMs are effective to study sequential data as well time series data. LSTMs work by using a set of gates to control the flow of information through the network. The gates are composed of sigmoid activation functions that produce outputs between 0 and 1. The gates can either allow information to pass through, or block it. The four main gates used in LSTMs are:

1) **Input gate:** This gate make note of how much new information should get pass to cell state and all input are taken at current time step and this generated output consider as input for next time step.
2) **Forget gate:** This gate decides which information should get omitted from cell sate and it receives input at current time step and product from previous time step take as input.
3) **Output gate:** As it's a output gate it decide which information should send as output from cell state. It takes current time step as input and earlier time step as output.
4) **Cell state:** Cell state is known as Memory of LSTM, has capacity to store data for longer time. Here input and forget gate are used to change the cell state and output gate monitor and control what will be output from cell state.

All above gates in LSTM design to select remember and forget the data from early time stage. LSTM is design to solve vanishing gradient from RNN, so LSTM is powerful to get accurate prediction also LSTM is capable to learn long term dependencies in sequential and time series data.

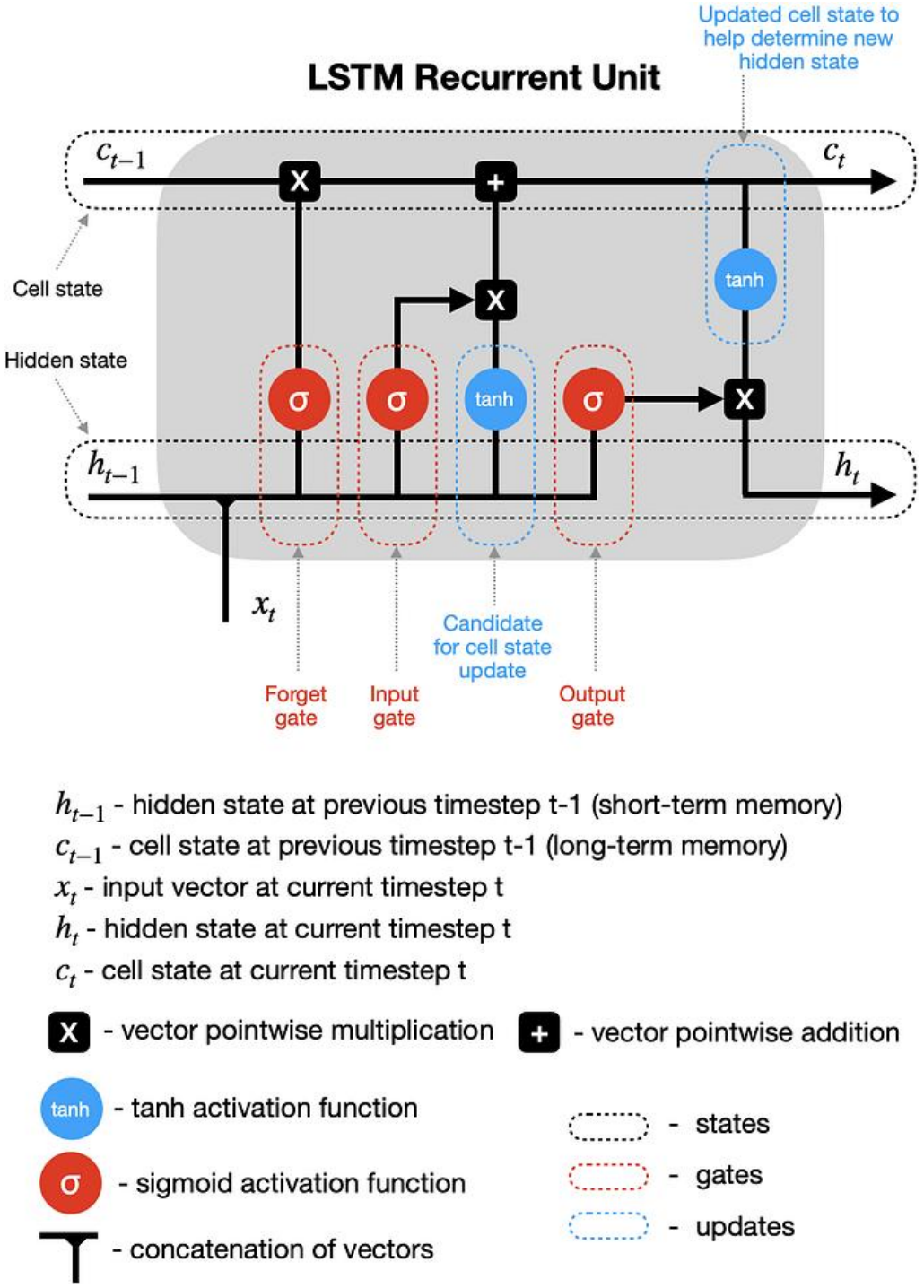

Figure 5: LSTM Unit (source: *https://solclover.com)*

4. Implications

The detection of CVD (Cardiovascular Disease) using deep learning algorithms has several implications, including:

- **Early detection:** Deep learning algorithms can help identify patterns in large amounts of data that may not be possible for us, allowing for the early detection of CVD. This can lead to pre- interventions and medications, which will help to reducing the risk and complications for patients.
- **Personalized medicine:** Trained model include information about individual patient characteristics, such as age, sex, and medical history, exercise habits, eating habits that can help the developed a personalized and proper treatment designing for patient.
- **Improved accuracy:** Deep learning algorithms can analyze large amounts of data with high accuracy, potentially improving the accuracy of CVD detection compared to traditional diagnostic methods. This can reduce the risk of misdiagnosis and improve patient outcomes.
- **Reduced healthcare costs:** As CVD is not having certain symptoms so preventative measures at early stage may reduce the medication cost and will help to avoid expensive medical procedures. Neural Network can recognise the difficult pattern to identify risk factors which can cause CVD, so proposed work will be good as

preventive measure that will help to decrease patients' burden by cost as well health.

- **Data privacy concerns:** Any Neural Network algorithms needed large dataset to train and achieve accurate and effective results, but the main concern is data privacy and its security. So need to ensure every patient that data is protected and safe to avoid unauthorized use of patients' record.

5. Conclusion

From the literature review and comparative study of all machine learning algorithm , reached to conclusion that accuracy is very important factor when we consider health and wellness of human beings. We should not compromise our health at any cost. So RNN with LSTM would give better and accurate results compare to other machine learning algorithm

References

[1] Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms, Kaushalya Dissanayake and Md Gapar Md JoharApplied Computational Intelligence and So Computing Volume 2021, Article ID 5581806, ISBN 1155/2021/5

[2] Prediction of Cardiovascular disease using ML, MuktéviSrivenkatesh ,International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

[3] Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning, RohitBharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande ,and Parneet Singh, Computational Intelligence and Neuroscience Volume 2021, Article ID 8387680

[4] Heart Disease Prediction System, Kennedy NgureNgare, St. Paul's University

[5] Machine learning prediction in cardiovascular diseases: a meta-analysis, Chayakrit Krittanawong1,9*, HafeezUl Hassan Virk, Sripal Bangalore, ZhenWang, Kipp W. Johnson6, Rachel Pinotti7, HongJu Zhang8, Scott Kaplin9, Bharat Narasimhan,Takeshi Kitai10, Usman Baber, Jonathan L. Halperin & W. H. Wilson Tang10

[6] A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach , M. Marimuthu, M. Abinaya, K. S. Hariesh , K. Madhan kumar , V. Pavithra , International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 18, September 2018

[7] Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction,Surenthiran Krishnan, Pritheega Magalingam, Roslina Ibrahim Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia, International Journal of Electrical and Computer Engineering (IJECE) Vol. 11, No. 6, December 2021, pp. 5467~5476 ISSN: 2088-8708, DOI: 10.11591/ijece.v11i6.pp5467-5476

[8] Image Classification using SVM and CNN, Sai Yeshwanth Chagant, Test Engineer Irdeto Technologies Pvt Ltd New Delhi, Ipseeta Nanda Electronics and Communication Koneru Lakshmaiah Education Foundation Vaddeswaram, India, University of Southern Queensland, July 2014 UTC from IEEE.

[9] Incidence if Cardiovascular Risk Factor in an Indian Urban Cohort, Mark D, Huffman, Europe PCM Funders Author Manuscripts Am CollCardiol. 2011 April 26; 57(17): 1765–1774. doi:10.1016/j.jacc.2010.09.083.

[10] B. S. Babu, A. Suneetha, G. C. Babu, Y. J. N. Kumar, and G. Karuna, "Medical disease prediction using grey wolf optimization and auto encoder based recurrent neural network," Periodicals of Engineering and Natural Sciences, vol. 6, no. 1, pp. 229-240, 2018, doi: 10.21533/pen.v6i1.286.

[11] Migran N. Gevorkyan, Anastasia V. Demidova,Tatiana S. Demidova, Anton A. Sobolev, "Review and comparative analysis of machine learning libraries for machine learning", UDC 519.21;51-76DOI: 10.22363/2658-4670-2019-27-4-305-315

[12] Vassili Kovalev, Alexander Kalinovsky, Sergey Kovalev, " Deep Learning with Theano, Torch, Caffe, Tensor flow, andDeeplearning4J: Which One is the Best in Speed and Accuracy?" United Institute of Informatics Problems, Belarus National Academy of Sciences Surganova St., 6, 220012 Minsk, Belarus