**Prediction of cardiopulmonary arrest**

**Bikash Chandra Sahoo1 and Dr. Sandeep Kumar Satapathy2,**

1School of Computer Science and Engineering, VIT Chennai

2 School of Computer Science and Engineering, VIT Chennai

**E-mail address:** bikashchandra.sahoo2022@vitstudent.ac.in

**Abstract**

The cardiopulmonary arrest is a major issue in any country. Gone are the days when it used to happen to those who are aged but now it is a major concern emerging among adolescents as well. According to the World Health Organization (WHO), cardiac arrest and stroke is still a major concern and remains a public health crisis. In past years India has witnessed many cases of heart related issues which used to occur predominantly among people having high cholesterol. But now the scenario has changed, and cases were observed in people having normal cholesterol level. There are several factors involved in heart stroke such as age, sex, blood pressure etc. which are used by doctors to monitor and diagnose the same. The factors contributing to heart issues can be used as a beacon to predict the stroke and help an individual to further consult a doctor beforehand. This paper focuses on different predictive models and ways to improve the accuracy of prediction by analyzing datasets on how they affect the accuracy of certain algorithm. The idea is to target the datasets and the prediction algorithms of machine learning and deep learning including advanced ones in order to improvise it and attain a better result. Furthermore, it discusses about a new attribute called “gamma prime fibrinogen” which could be used in future for boosting prediction performance.

***Keywords:*** *heart stroke; adolescent; dataset; predictive models; fibrinogen*

1. **Introduction**

Health plays a vital role in any person’s life. It is very important for everyone to keep health in check and be free from ailment. But there are some health issues which are uncertain and could become a barrier to smooth running lifestyle. Such uncertain issues involve heart stoke, cancer, diabetes etc. and they need to be diagnosed and controlled beforehand. Cardiopulmonary arrest or heart stroke is one of the health conditions that can happen to any individual. The probability of happening a stroke depends on diet and the lifestyle that a person is having. It is evident from the past observations recognized by reputed organizations like World Health Organization (WHO) and Centre for Disease Control (CDC) that heart stroke is a frequently occurred issue and the pattern of occurrence is changing day by day. Heart stroke is contributed by many factors like age, blood pressure, smoking status, cholesterol level etc. and consist of pattern. Hence it can be predicted and diagnosed with the help predicting models of machine learning and deep learning which is a boon of technology.

With the evolution of technology, researchers and scientist can now depend more on the same in order to analyze and predict health issue like heart stroke. Machine learning and deep learning is one of the emerging media for foretelling an issue. Techniques like Support Vector Machine (SVM), Random Forest, Naïve Bayes, Recurring Neural Network (RNN) and many more are used as predicting models. In all these algorithms the ultimate goal is to fetch relation between attributes and perform some operations to calculate prediction accuracy. So far there is no consistency in prediction accuracy as it varies with the dataset and the type of algorithm. Therefore, there is one such attribute called “gamma prime fibrinogen” which could be gamechanger in predicting heart stroke in a precise manner. Fibrinogens are the soluble protein which plays a vital role in blood coagulation but sometimes play major role in cardiac related issues. On the other hand, improvising algorithms like using different activation function could result in better outcome. Overall, the prediction models should cope up with datasets which may vary overtime.

1. **Literature Survey**

Chethan Malode [1] made use of fuzzy rules enabled SVM (Support Vector Machine) to predict likelihood of heart stroke. The proposed model involved four phases namely creation of soft set, generation of fuzzy rules, SVM classifier and decision making. The necessary data was first obtained through device and fed into the soft set creator generating binary output. The binary output was then fed into fuzzy generator to classify data using SVM. Finally, based on SVM classifier the result was displayed as low or high risk. The proposed framework is good compared to conventional SVM with respect to classification accuracy and efficiency.

One of the works involved bagging classifier to predict and detect heart attack carried out by Faruk Bulut [2]. The process indulged in asking questions to patients who have already suffered heart attack. The answers to the questionnaire were used as dataset to feed into the bagging classifier where it was sampled into multiple subsets and processed using different machine learning algorithms. The average of the prediction accuracy was considered as final result and the cross-validation process showed high performance in regression.

D. K. Ravish [3] proposed neural network-based model which used ECG data as prime attribute to predict heart stroke. The features of ECG such as QRS duration, R-R interval etc. and statistical features like blood pressure, cholesterol was processed. The training stage involved MATLAB to establish a neural network involving 100 layers for computational work. After passing through successive layers the errors were reduced and the result was displayed in the output stage. The proposed model not only helped in predicting heart stroke but was also capable of predicting myocardial infarction in near future.

One of the works carried out by D. R. Krithika [4] was based on ensemble-based prediction where a large dataset was fed to multiple machine learning algorithms. The dataset was also sampled into multiple subsets to for implementing bagging classifier using decision tree. Bootstrapping aggregation was performed to decrease the variance for accurate prediction. Hyper Parameter Tuned Random Forest (HPTRF) resulted in very high accuracy and hematocrit proved to be an important attribute for stroke prediction.

Similarly, in [5], comparative analysis of heart stroke prediction implemented by Tanisha Rakshit made use of different machine learning algorithms in which dataset was passed through phases like data description, pre-processing, visualization and splitting. In this work decision tree resulted in highest accuracy.

The AI based prediction proposed by Jaehak Yu [6] used two more attributes namely ECG and PPG apart from readily available datasets. The photoplethysmography or PPG data which tells regarding the volumetric change in blood played a vital role in analyzing and predicting heart stroke better. In this work, first the data is collected along with other bio signals like EEG and EMG. The data was then normalized using Z score method. The prediction process can be executed either in offline or online mode. After the feature selection the resulting data is fed to machine learning and deep learning models. The model was then saved as meta file. The prognostic symptoms of stroke patients can be accurately predicted by more than 90% based solely on ECG and PPG. The author further proposed in depth analysis of bio signals for future scope.

D. Sharath Chandra [7] proposed a way to detect heart disease and diabetes. The implementation undergoes similar steps like data preprocessing, fitting ML algorithms, training and prediction. The author used dataset from UCI repository and data world webserver. The work was carried out specifically using logistic regression (LR) and support vector machine (SVM) which resulted in accuracy of 85% and 78% respectively. The important aspect of the work was its interactive feature to display health status.

The work proposed by Aditi Gavhane [8] made use of multilayer perceptron (MLP) to predict heart disease. Deep learning method proves to be effective when compared to machine learning as it decreases the variance to a great extent. The author prepared dataset with the help of different sensors like AliveKor, Health Gear, fitbit etc. and used it as input weights in MLP mechanism. The neural network consisted of one input layer, one hidden layer and one output layer. The output layer resulted in binary health status as “yes” or “no”. The proposed work can be implemented on other diseases as well like diabetes, cancer etc.

A study-oriented work was carried out by SP Rajamhoana [9] on the analysis of heart stroke prediction using neural network. The analysis involved feature selection and classification technique in which multilayer perceptron with back propagation learning resulted in the highest accuracy of 94%. The work also showed that hybrid system that use the combination of artificial intelligence methods gives the highest accuracy.

Paranthaman M [10] made use of deep learning algorithm to predict cardiovascular disease. The proposed work used neural network having multiple layers. The problem of limiting and outlining various data mining technique used in the field of prediction was inspected in this work. The classification technique is delicate to noisy data. If there is any noisy data present it could establish major issues in classification.

In one of the previous work ECG data was used and processed using machine learning but in one of the work proposed by Ashish Kumar [11] ECG data was used for prediction using neural networks. Two methods namely artificial neural network (ANN) and convolutional neural network (CNN) was implemented where CNN outperformed ANN by 4%. The author proposed addition of more and new data to check the accuracy of predictive models in near future.

Jyothirmai Digumarthi [12] proposed bio-inspired algorithm to predict cardiac arrhythmia. It incorporates one of the techniques called Modified Salp Swarm Optimization (MSSO) and Adaptive Neuro fuzzy Inference (ANFIS) gaining the highest accuracy of 99.4%. The problem which arose in this work was to apply algorithms to multiple data and creating a unique framework to automate the process. The filters and fuzzy logic could increase the performance of algorithm in the proposed work.

1. **Methodology**

The project focuses on prediction of heart stroke and it is very necessary to predict vital issues beforehand for diagnostic purpose. Such steps could help in taking measures in order to counteract health problems and will help in the betterment of society. A search for the scope was executed in google scholar and digital library database which consist of IEEE, ACM, ICSCC, ICCNT journals. For the implementation phase dataset was extracted from hospital, kaggle and UCI repository. In the design phase different machine learning algorithms like Logistic Regression, ANN, Transfer Learning etc. were used on two different datasets namely Cleveland and Framingham. Both datasets differ based on certain features. The accuracy obtained from executing each machine learning algorithms were used to deduce comparative analysis and finally the highest accuracy algorithm is highlighted.

* 1. **Data Acquisition**

There are two datasets that were used for our objective. These are “Cleveland” and “Framingham”. The Cleveland dataset is a mixed dataset and consist of both categorical and numerical value while Framingham consist of only numerical data. The common attributes in both the datasets are age, gender, diabetes, cholesterol, BMI, glucose, smoking status. The other attributes are resident type, work type, marital status, stroke in ten years. The Cleveland dataset consist of 5110 data while Framingham dataset consist of 4240 data.

**3.2. Block Diagram**

**Selection of dataset**

**Processing of data**

**Selection of algorithm**

**Implementation of algorithm**

**Decision**

* 1. **Implementation**

Selection of proper dataset based on objective is very necessary. After the selection of dataset, it needs to be analyzed manually regarding missing values, NaN etc. The dataset is then need to be preprocessed to make it suitable for processing using particular machine learning algorithms. The preprocessing involves dropping of row values, imputing missing values with mean or median, scaling, normalization and feature selection. After this an algorithm need to be selected to which the processed data will be fed. Multiple suitable algorithms need to be tested to achieve the best possible accuracy. Based on the implementation using various algorithms, the decision needs to be made which algorithm best fits the dataset. The decision is made on the basis of various performance metrics like accuracy, F-Test, precision, recall etc. The performance metrics also varies based on the algorithm implemented.

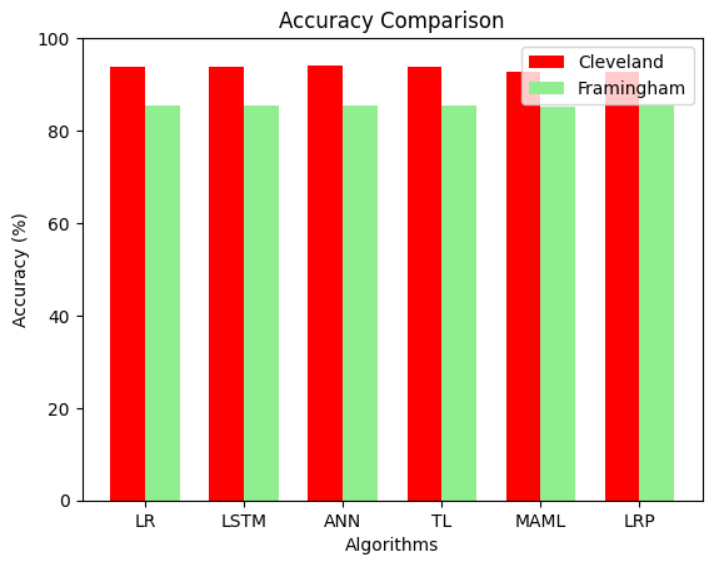
* 1. **Algorithm**

1. Import the necessary packages and libraries.
2. Perform the feature selection and gather all attributes in variable X and target feature in variable y which will be used to predict the heart stroke.
3. Preprocess the dataset by imputing missing values or by normalizing it with statistical parameters. Also based on algorithm convert attribute to categorical or numerical value using one hot encoding wherever necessary.
4. Split the dataset into training and testing data in 80-20 or 70-30 ratio.
5. Select algorithms and create a neural network architecture using suitable function.
6. Train the model and track status based on epochs value.
7. Use hyperparameter tuning if required and skip if the dataset is too large.
8. Evaluate the model based on accuracy and visualize it using any suitable chart.
9. The stroke attribute in Cleveland dataset and in Framingham stroke in 10 years is chosen as target variable for prediction.
10. To get the better analysis the algorithms are applied both on Cleveland and Framingham so that the accuracy obtained in each dataset can be compared and chosen for future generic use.
11. The implementation uses epochs to track accuracy and loss.
12. The cross-validation process is also implemented to check if the algorithm overfit.
13. **Results and Discussions**

**Table 1:** Comparative analysis

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Dataset Accuracy (in %)** | |
| **Cleveland** | **Framingham** |
| Logistic Regression | 93.9 | 85.4 |
| Long Short-Term Memory | 93.9 | 85.4 |
| Artificial Neural Network (ANN) | 94.1 | 85.4 |
| Transfer Learning | 93.9 | 85.4 |
| Model Agnostic Meta Learning (MAML) | 92.6 | 85.1 |
| Layer wise Relevance Propagation (LRP) | 92.6 | 85.5 |

Almost all algorithm is performing exceptionally good having accuracy more than 90% in Cleveland dataset and more than 80% in Framingham dataset. ANN having the maximum accuracy is therefore should be chosen for prediction of cardiac arrest. Furthermore, Cleveland dataset is availing more accuracy than Framingham dataset. This shows that the features of Cleveland dataset are more vital when compared to Framingham.



**Figure 1**: Accuracy comparison for different algorithm on two different datasets

**A picture containing text, screenshot, font, plot

Description automatically generated**

**Figure 2**: Training and validation loss for best working algorithm

Figure 2 shows the training and validation loss for best performing algorithm i.e., artificial neural network (ANN). Since the training loss and validation loss are deviating i.e., their incrementing nature are opposite to each other shows that the algorithm is not overfitting which is a good sign.

**4. Conclusions**

The deep learning algorithm resulted in the accuracy of 93% on Cleveland dataset and 85.4% on Framingham dataset. Other useful algorithm like transfer learning also proved itself as a good choice as it gave the same accuracy on Cleveland dataset. Testing algorithm on two different datasets also showed that the features in a dataset might impact the accuracy of algorithm.

For future scope by improvising the algorithms like bringing change in activation function in neural networks or by introducing new attribute like gamma prime fibrinogen could help predictions perform even further.

**References:**

[1] Harshita Puri, Jahnavi Chaudhary, Kulkarni Rakshit Raghvendra, Rhea Mantri, Kiahore Bingi, “Prediction of stroke using SVM algorithm”, 2021 8th International conference on smart computing and communications (ICSCC).

[2] “Heart attack risk detection using bagging classifier”, IEEE Signal processing and communication applications.

[3] Jaehak Yu, Sejin Park, Soon Hyun Kwan, Kang-Lee Cho, Hansung Lee, “”AI-Based stroke disease prediction system using ECG and PPG bio-signals”.

[4] CM Chethan Malode, K Bhargavi, B G gunasheela, G kavana, R Sushmitha, “Soft set and fuzzy rules enabled SVM approach for heart attack risk classification among adolescents”, 2018 fourth international conference on computing communication control and automation (ICCUBEA).

[5] D K Ravish, K J Shanthi, Nayana R Shenoy, S Nisargh, “Heart function monitoring, prediction and prevention of heart attacks: using artificial neural networks”, 2014 International conference on contemporary computing and informatics (IC3I.)

[6] D Sharathchandra, M Raghu Ram, “ML based interactive disease prediction model”, 2022 IEEE Delhi section conference (DELCON).

[7] Dr. Kritika K Rohini, “Ensemble based prediction of cardiovascular disease using bigdata analytics”, 2021ICCS.

[8] M. S. Salman et al., "Data Mining Technique for Prediction System of Heart Disease Using Associative Classifications," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 1-5, doi: 10.1109/ICCCNT51525.2021.9579825.

[9] S. P. Rajamhoana, C. A. Devi, K. Umamaheswari, R. Kiruba, K. Karunya and R. Deepika, "Analysis of Neural Networks Based Heart Disease Prediction System," 2018 11th International Conference on Human System Interaction (HSI), 2018, pp. 233-239, doi: 10.1109/HSI.2018.8431153.

[10] Paranthaman. M, Yaathash. B and Sanjairam. M, "Cardiovascular Disease Prediction using Deep Learning”, 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1399-1404, doi: 10.1109/ICOEI53556.2022.9777135.

[11] A. Kumar, K. Rathor, S. Vaddi, D. Patel, P. Vanjarapu and M. Maddi, "ECG Based Early Heart Attack Prediction Using Neural Networks”, 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), 2022, pp. 1080-1083, doi: 10.1109/ICESC54411.2022.9885448.

[12] J. Digumarthi, V. M. Gayathri and R. Pitchai, "Early Prediction of Cardiac Arrhythmia using Novel Bio-inspired Algorithms”, 2022 8th International Conference on Smart Structures and Systems (ICSSS), 2022, pp. 01-04, doi: 11109/ICSSS54381.2022.9782178.

[13] Krishnan et al., “Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction”, 2021 International Journal of Electrical and Computer Engineering (IJECE), 11. 5467-5476. 10.11591/ijece. v11i6.pp5467-5476.

[14] Harshvardhan Tiwari, “Early Prediction of heart disease using deep learning approach”, 2022 Centre for Incubation, Innovation, Research, and Consultancy (CIIRC), https://doi.org/10.1016/B978-0-12-824145-5.00014-9.

[15] S. Priscila, M. Hemalatha, “Improving the Performance of Entropy Ensembles of Neural Networks (EENNS) on Classification of Heart Disease Prediction”, 2017 International Journal of Pure and Applied Mathematics (IJPAM), Volume 117 No. 7 2017, 371-386 ISSN: 1314-3395.

[16] Mohammad Amin Morid, Olivia R. Liu Sheng, and Joseph Dunbar. 2023. Time Series Prediction Using Deep Learning Methods in Healthcare. ACM Trans. Manage. Inf. Syst. 14, 1, Article 2 (March 2023). https://doi.org/10.1145/3531326.

[17] Debabrata Swain, Santosh Pani and Debabala Swain, "An efficient system for the prediction of Coronary artery disease using dense neural network with hyper parameter tuning", International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 6S, pp. 689-695, April. 2019.

[18] N. Sharma, M. K. Mishra, J. S. Chadha and P. Lalwani, "Heart Stroke Risk Analysis: A Deep Learning Approach," 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2021, pp. 543-598, doi: 10.1109/CSNT51715.2021.9509665.

[19] Xie, Y., Yang, H., Yuan, X. et al. Stroke prediction from electrocardiograms by deep neural network. Multimed Tools Appl 80, 17291–17297 (2021). https://doi.org/10.1007/s11042-020-10043-z.

[20] Chae M, Han S, Gil H, Cho N, Lee H. Prediction of In-Hospital Cardiac Arrest Using Shallow and Deep Learning. Diagnostics. 2021; 11(7):1255. https://doi.org/10.3390/diagnostics11071255.