

A Deep Learning Framework for Prediction of Cardiopulmonary Arrest

Sirisha Potluri^{1,*}, Bikash Chandra Sahoo², Sandeep Kumar Satapathy³, Shruti Mishra⁴, Janjhyam Venkata Naga Ramesh⁵ and Sachi Nandan Mohanty⁶

¹ Department of Artificial Intelligence and Data Science, Faculty of Science and Technology (IcfaiTech), ICFAI Foundation for Higher Education, Hyderabad, India-501203

² School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India- 600127

³ Department of Computer Science, Yonsei University, 50 Yonsei-ro, Sudaemoon-gu, Seoul 03722, South Korea

⁴ Centre for Advanced Data Science, Vellore Institute of Technology, Chennai, Tamil Nadu, India- 6001273

⁵ Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., Andhra Pradesh - 522302, India

⁶ School of Computer Science & Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India-522237

Abstract

INTRODUCTION: 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 have been observed in people having normal cholesterol levels. 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.

OBJECTIVES: 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 algorithms.

METHODS: 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. The idea is to target the datasets and the prediction algorithms of deep learning including advanced ones to improvise it and attain a better result.

RESULTS: This paper brings out the comparative analysis among neural network techniques like ANN, Transfer Learning, MAML and LRP in which ANN showed the best result by giving the highest accuracy of 94%.

CONCLUSION: Furthermore, it discusses a new attribute called “gamma prime fibrinogen” which could be used in the future to boost prediction performance.

Keywords: Heart Stroke; Adolescent; Neural Network; Predictive Models; Fibrinogen

Received on 03 December 2024, accepted on 08 March 2024, published on 14 March 2024

Copyright © 2024 S. Potluri et al., licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.10.5420

*Corresponding author. Email: sirisha.vegunta@gmail.com

1. Introduction

Health plays a vital role in any person's life. Everyone needs to keep their health in check and be free from ailment. However, some health issues are uncertain and could become a barrier to smooth smooth-running lifestyle. Such uncertain issues involve heart stroke, 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 a stroke depends on diet and the lifestyle that a person has. It is evident from past observations recognized by reputed organizations like the World Health Organization (WHO) and Centre for Disease Control (CDC) that heart stroke is a frequently occurring issue and the pattern of occurrence is changing day by day. Heart stroke is caused

by many factors like age, blood pressure, smoking status, cholesterol level, etc., and consists of pattern. Hence it can be predicted and diagnosed with the help of predicting models of machine learning and deep learning which is a boon of technology.

With the evolution of technology, researchers and scientists can now depend more on the same to analyze and predict health issues 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 the 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 a game changer in predicting heart stroke in a precise manner. Fibrinogens are soluble protein that plays a vital role in blood coagulation but sometimes play a major role in cardiac-related issues. On the other hand, improvising algorithms like using different activation functions could result in better outcomes. Overall, the prediction models should come up with datasets that may vary over time.

2. Literature Survey

Malode et al. [1] made use of fuzzy rules that enabled SVM (Support Vector Machine) to predict the likelihood of heart stroke. The proposed model involved four phases namely creation of a soft set, generation of fuzzy rules, SVM classifier, and decision making. The necessary data was first obtained through the device and fed into the soft set creator generating binary output. The binary output was then fed into a fuzzy generator to classify data using SVM. Finally, based on the SVM classifier the result was displayed as low or high risk. The proposed framework is good compared to conventional SVM for classification accuracy and efficiency. One of the works involved a bagging classifier to predict and detect heart attacks carried out by Bulut [2]. The process indulged in asking questions to patients who have already suffered heart attacks. The answers to the questionnaire were used as a 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 the final result and the cross-validation process showed high performance in regression.

Ravish et al. [3] proposed a neural network-based model which used ECG data as a prime attribute to predict heart stroke. The features of ECG such as QRS duration, R-R interval etc. and statistical features like blood pressure, and cholesterol were 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 soon. One of the works carried out by 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 for implementing a bagging classifier using a 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], a comparative analysis of heart stroke prediction implemented by Rakshit made use of different machine learning algorithms in which the dataset was passed through phases like data description, pre-processing, visualization, and splitting. In this work, the decision tree resulted in the highest accuracy. The AI based prediction proposed by Yu et al. [6] used two more attributes namely ECG and PPG apart from readily available datasets. The photo plethysmography 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 biosignals like EEG and EMG. The data was then normalized using the 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 metafile. The prognostic symptoms of stroke patients can be accurately predicted by more than 90% based solely on ECG and PPG. The author further proposed a depth analysis of biosignals for future scope.

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

The work proposed by Gavhane et al. [8] made use of multilayer perceptron (MLP) [9] to predict heart disease. The deep learning method proves to be effective when compared to machine learning as it decreases the variance to a great extent. The author prepared a dataset with the help of different sensors like AliveKor, Health Gear, fitbit, etc., and used it as input weights in the 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 Rajamhoana et al. [10] on the analysis of heart stroke prediction using a neural network. The analysis involved feature selection and classification techniques in which multilayer perceptron with back propagation learning resulted in the highest accuracy of 94%. The work also showed that a hybrid

system that uses a combination of artificial intelligence methods gives the highest accuracy.

Paranthaman et al. [11] made use of a deep learning algorithm to predict cardiovascular disease. The proposed work used a neural network having multiple layers. The problem of limiting and outlining various data mining techniques 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 the work proposed by Kumar et al. [12], ECG data was used for prediction using neural networks. Two methods namely artificial neural network (ANN) and convolutional neural network (CNN) were implemented where CNN outperformed ANN by 4%. The author proposed the addition of more new data to check the accuracy of predictive models shortly. Digumarthi et al. [13] proposed a bio-inspired algorithm to predict cardiac arrhythmia. It incorporates one of the techniques called Modified Salp Swarm Optimization (MSSO) and Adaptive neurofuzzy Inference (ANFIS) gaining the highest accuracy of 99.4%. The problem that arose in this work was to apply algorithms [14] to multiple data and create a unique framework to automate the process. The filters and fuzzy logic could increase the performance of the algorithm in the proposed work.

3. Methodology

The project focuses on the prediction of heart stroke and it is very necessary to predict vital issues for diagnostic purposes [15]. Such steps could help in taking measures 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 databases which consist of IEEE, ACM, etc. For the implementation phase dataset was extracted from hospitals, kaggle, and UCI repositories. 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 algorithm was used to deduce comparative analysis and finally, the highest accuracy algorithm is highlighted.

3.1. Dataset Acquisition

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

3.2. Block Diagram

Figure 1 below outlines the basic steps to be followed while implementing different algorithms. The dataset needs to be selected which is publicly available. Well-defined data extracted from the preprocessing phase is the key to better analysis with the help of an algorithm chosen based on the user’s objective. The decision phase should not be solely based on accuracy. Certain other performance metrics like F1 score, mean square error, etc. should also be considered to justify the prediction.

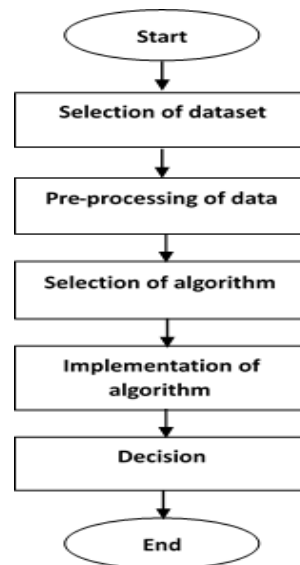


Figure 1. Generalized workflow model for implementing different algorithms

3.3. Neural Techniques Used

Transfer learning: Transfer learning is a machine learning technique that enhances the performance of a related task by influencing knowledge and training obtained from another task. Unlike starting from scratch, transfer learning utilizes pre-trained models that have already been learned from extensive data. In this case, the numerical data is transformed into a heatmap representation resembling an image. Consequently, the heatmap can be treated as an image and utilized in subsequent training processes. The heatmap equivalent images were compared for the training process.

Model-agnostic meta-learning: Model-agnostic meta-learning (MAML) is a meta-learning strategy designed to enhance the machine learning model’s ability to rapidly adapt to new tasks. The core concept of MAML involves acquiring an initial set of model parameters that can be easily fine-tuned for diverse tasks. The approach consists of two stages: an inner loop and an outer loop. During the inner loop, the model is trained on a small amount of task-specific data, and the gradients are calculated to update the model parameters. In the outer loop, the updated model parameters are assessed using an independent validation set. This iterative process facilitates the acquisition of transferable knowledge, enabling the model to effectively tackle various tasks.

Layer-wise relevance propagation: Layer-wise relevance propagation (LRP) is an interpretability method employed in deep learning models for comprehending the impact of individual input features or neurons on the model's output. Its objective is to gain insights into the model's decision-making process by assigning relevance scores to each feature. The core concept of LRP involves propagating the relevance scores from the model's output layer to the input layer. During this propagation, the scores are redistributed based on the contribution of each neuron or feature. Additionally, the technique includes a step called "rule application," which is carried out at each layer to determine the appropriate distribution of relevance among the neurons within that layer.

3.4. Implementation

The selection of a proper dataset based on objective is very necessary. After the selection of the dataset, it needs to be analyzed manually regarding missing values, NaN, etc. The dataset then needs 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 needs 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 about which algorithm best fits the dataset. The decision is made based on various

performance metrics like accuracy, F-Test [16], precision, recall [17-20], etc. The performance metrics also vary based on the algorithm implemented.

3.5. Algorithm

- a) Import the necessary packages and libraries.
- b) Perform the feature selection and gather all attributes in variable X and target features in variable Y which will be used to predict the heart stroke.
- c) Preprocess the dataset by imputing missing values or by normalizing it with statistical parameters. Also based on the algorithm convert attributes to categorical or numerical values using one hot encoding wherever necessary.
- d) Split the dataset into training and testing data in an 80-20 or 70-30 ratio.
- e) Select algorithms and create a neural network architecture using suitable functions.
- f) Train the model and track status based on epochs value.
- g) Use hyperparameter tuning if required and skip if the dataset is too large.
- h) Evaluate the model based on accuracy and visualize it using any suitable chart.
- i) The stroke attribute in the Cleveland dataset and Framingham stroke in 10 years is chosen as the target variable for prediction.
- j) To get a 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.
- k) The implementation uses epochs to track accuracy and loss.
- i) The cross-validation process is also implemented to check if the algorithm overfit.

4. Results and Discussions

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

Table 1. An overall comparative analysis with various algorithms and dataset

Table 1 above showcases the major deep learning algorithms executed on two different datasets. It can be inferred that almost all algorithms are performing exceptionally well having an accuracy of more than 90% in the Cleveland dataset and more than 80% in the Framingham dataset. ANN scores the maximum accuracy and therefore should be chosen for the prediction of cardiac arrest.

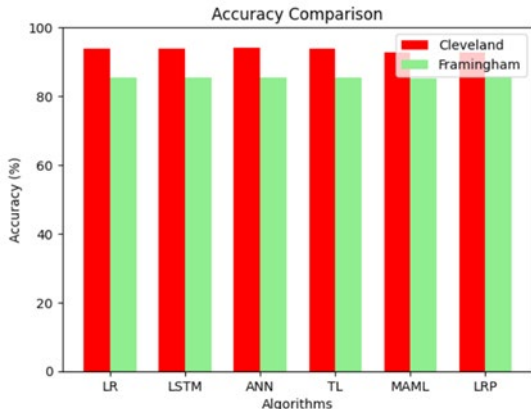


Figure 2. Accuracy comparison for different algorithm on two different datasets

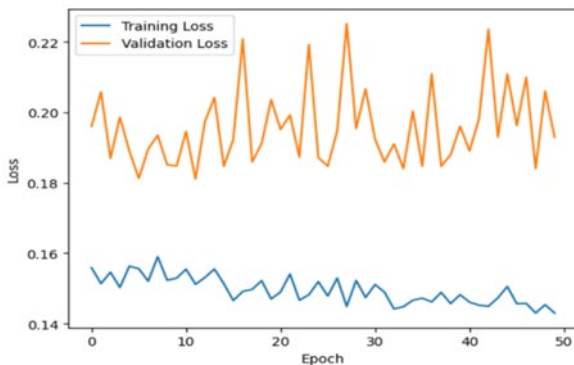


Figure 3. Training and validation loss for best working algorithm

Figure 2 above helps us visualize side by side comparison of different algorithms concerning two different datasets. The Cleveland dataset achieves more accuracy than the Framingham dataset when trained with different models. This shows that the features of the Cleveland dataset are more vital when compared to Framingham. Figure 3 shows the training and validation loss for best performing algorithm i.e., artificial neural network (ANN) [21-24]. Since the training loss and validation loss are deviating i.e., their incrementing nature is opposite to each other shows that the algorithm is not overfitting which is a good sign.

Figure 4 shows the ROC curve for ANN and its value to be 0.82 which indicates that the model has good overall discriminatory power in distinguishing between positive

and negative cases. The specificity and NPV value also came out to be 1 and 94% respectively which shows the promising performance measure for an algorithm.

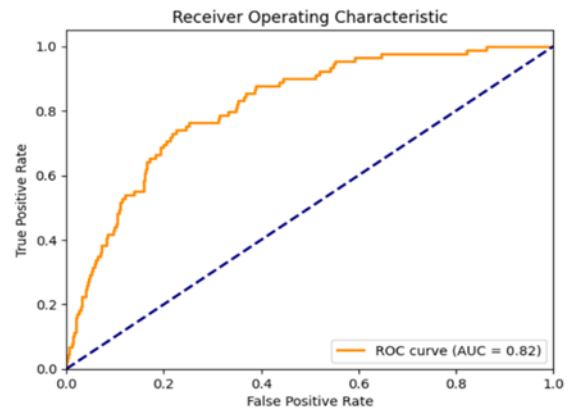


Figure 4. ROC curve for the best working model

Figure 4 shows the ROC curve for ANN and its value to be 0.82 which indicates that the model has good overall discriminatory power in distinguishing between positive and negative cases. The specificity and NPV values also came out to be 1 and 94% respectively which shows the promising performance measure for an algorithm.

4.1. Gamma prime fibrinogen

Gamma prime fibrinogen is a form of fibrinogen, a protein that plays a role in blood clotting. The normal quantity in the human body measures 22.5-23.7 mg/dL. Studies indicate that higher levels of gamma prime fibrinogen are linked to a greater likelihood of cardiovascular diseases, such as cardiopulmonary arrest. Machine learning techniques can be utilized to analyze data related to gamma prime fibrinogen and create models that predict the risk of cardiac arrest.

5. Conclusions

ANN resulted in the highest accuracy of 94% with hyperparameter tuning 95% on the Cleveland dataset and 85.4% on the Framingham dataset. In our specific case, artificial neural networks (ANN) demonstrated superior performance as the most suitable model. This is attributed to their adaptability in handling various types of data, which aligns with the characteristics of our dataset. ANNs excel in capturing non-linear relationships between features and the target variable, enabling them to effectively handle complex patterns. Unlike transfer learning, which relies on pre-existing models, ANNs can start from scratch and learn directly from the data, making them well-suited for our scenario where there is no predefined model specifically trained on cardiac arrest data. Additionally, our dataset does not involve a diverse set of related tasks, making model-agnostic meta-learning

(MAML) less effective in the long run. Lastly, while layer-wise relevance propagation (LRP) does not contribute significantly to improving predictive performance, ANN models are capable of achieving better results in this regard. Testing algorithms on two different datasets also showed that the features in a dataset might impact the accuracy of the 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] C Malode, K Bhargavi, B G Gunasheela, G kavana and R Sushmitha. Soft set and fuzzy rules enabled SVM approach for heart attack risk classification among adolescents. ICCUBE, 2018.
- [2] F Bulut. Heart attack risk detection using Bagging Classifier. SIU. 2016.
- [3] D K Ravish, K J Shanthi, Nayana R Shenoy and S Nisargh. Heart function monitoring, prediction and prevention of heart attacks: using artificial neural networks. ICCCI. 2014.
- [4] D Krithika and K Rohini. Ensemble Based Prediction of Cardiovascular Disease Using Bigdata analytics. IEEE. 2021. pp.42-46.
- [5] T Rakshit and A Shreshtha. Comparative Analysis and Implementation of Heart Stroke Prediction using Various Machine Learning Techniques. IJETR. 2021. 10(6), pp. 6-9.
- [6] J Yu, S Park, S H Kwan, K Cho and H Lee. AI-based stroke disease prediction system using ECG and PPG bio-signals. IEEE. 2022. 10(3), pp.23-38.
- [7] D Sharathchandra and M Ram. ML-based interactive disease prediction model. IEEE. 2022.
- [8] A. Gavhane, G. Kokkula, I. Pandya and K. Devadkar. Prediction of Heart Disease Using Machine Learning. Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). 2018.
- [9] S. K. Satapathy, S. Mishra, P.K. Mallick, G.S. Chae. ADASYN and ABC Optimized RBF Convergence Network for Classification of Electroencephalograph Signal. Personal and Ubiquitous Computing, Springer. 2022.
- [10] S Rajamhoana, C Devi, K Umamaheswari, R Kiruba, K Karunya and R Deepika. Analysis of Neural Networks Based Heart Disease Prediction System. 11th International Conference on Human System Interaction (ICHSI). 2018. pp. 18-25.
- [11] Paranthaman, Yaathash and Sanjairam. Cardiovascular Disease Prediction using Deep Learning, 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022.
- [12] A Kumar, K Rathor, S Vaddi, D Patel, P Vanjarapu and M Maddi. ECG Based Early Heart Attack Prediction Using Neural Networks. IEEE. 2022.
- [13] J Digumarthi, V Gayathri and R Pitchai. Early Prediction of Cardiac Arrhythmia using Novel Bio-inspired Algorithms. 8th International Conference on Smart Structures and Systems (ICSSS). 2022. pp.1-4.
- [14] S. Mishra, S.K. Satapathy, S.N. Mohanty and C. R. Patnaik. A DM-ELM based classifier for EEG brain signal classification for epileptic seizure detection. Communicative and Integrative Biology. 2023. 16:1.
- [15] H Puri, J Chaudhary, K R Raghvendra, R Mantri and K Bingi. Prediction of stroke using SVM algorithm. 8th International conference on smart computing and communications (ICSCC). 2021. 5(9), pp.7-18.
- [16] M Salman, N Shila, K Hasan, P Ahmed, S Khushbu and S Noori. Data Mining Technique for Prediction System of Heart Disease Using Associative Classifications. 12th International Conference on Computing Communication and Networking Technologies (ICCNT). 2021. pp. 1-5.
- [17] S. K. Satapathy, S. Saravanan, S. Mishra, S. N. Mohanty, A Comparative Analysis of Multidimensional COVID-19 Poverty Determinants: An Observational Machine Learning Approach, New Generation Computing, Springer. 2023.
- [18] S Krishnan, P Magalingam and R Ibrahim. Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction. IJECE. 2021.
- [19] N. Panda, S. K. Satapathy, S. Mishra, P. K. Mallik. Empirical Study on Different Feature Selection and Classification Algorithms for Prediction of Hepatitis Disease. Technical Advancements of Machine Learning in Healthcare 936, 75. Springer. 2021.
- [20] H Tiwari. Early Prediction of heart disease using deep learning approach. CIIRC. 2022.
- [21] S Priscila and M Hemalatha. Improving the Performance of Entropy Ensembles of Neural Networks on Classification of Heart Disease Prediction. IJPAM. 2017. 117(7), pp.14-19.
- [22] D Swain, S Pani and D Swain. An efficient system for the prediction of coronary artery disease using a dense neural network with hyperparameter tuning. IJITEE. 2019. 8(6), pp. 54-67.
- [23] Y Xie, H Yang, X Yuan and Q He. Stroke prediction from electrocardiograms by deep neural network. Multimed Tools Application (MTA), 2021.
- [24] M Chae, S Han, H Gil, N Cho and H Lee. Prediction of In-Hospital Cardiac Arrest Using Shallow and Deep Learning. 2021. 11(7), pp. 25-35.