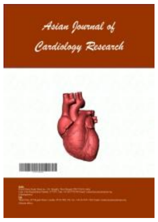

Asian Journal of Cardiology Research

3(1): 8-15, 2020; Article no.AJCR.57729

Early Detection of Heart Disease Using Gated Recurrent Neural Network

Shawni Dutta[1] and Samir Kumar Bandyopadhyay[2*]

[1]Department of Computer Science, The Bhawanipur Education Society College, Kolkata, India.
[2]The Bhawanipur Education Society College, Kolkata, India.

Authors' contributions

This work was carried out in collaboration between both authors. Author SD designed the proposed method, coding, statistical work and wrote the first draft. Author SKB initiates the work and edited all processes before finalizing the manuscript. Both author read and approved the final manuscript.

Article Information

Editor(s):
(1) Dr. Hugo R. Ramos, Hospital de Urgencias, Argentina.
Reviewers:
(1) N. R. Ananthanarayanan, SCSVMV, India.
(2) Mehmet Akif Ozdemir, Izmir Katip Celebi University, Turkey.
Complete Peer review History: http://www.sdiarticle4.com/review-history/57729

Original Research Article

Received 02 April 2020
Accepted 06 June 2020
Published 07 June 2020

ABSTRACT

Cardiovascular disease (CVD) is an important factor in life since it may cause the death of human by effecting the heart and blood vessels of the body. Early detection of this disease is necessary for securing patients life. For this purpose, an automated tool is proposed in this paper for detecting patients with CVD and assisting health care systems also. A stacked-GRU based Recurrent Neural Network model, abbreviated as, stk-G, is proposed in this paper that considers interfering factors from past health records while detecting patients with cardiac problems. This proposed model is compared with two benchmark classifiers known as Support Vector Machine (SVM) and K-Nearest Neighbour (K-NN). The comparative analysis concludes that the proposed model offers enhanced efficiency for heart disease prediction. A promising result is given by the proposed method with an accuracy of 84.37%, F1-Score of 0.84 and MSE of 0.16.

Keywords: Heart disease prediction; machine learning; deep learning; GRU; SVM; K-NN.

**Corresponding author: Email: 1954samir@gmail.com;*

1. INTRODUCTION

Heart, an important organ of the human body, pumps signals throughout the body for functioning different parts of the body. The signal is created by a node, called the Sinus Node. The heart is divided into two halves. One half is atria and another part is Ventricle. There are two atria-left and right as well as two ventricle-left and right. The atria collect blood and the ventricles push blood out of the heart. The right half of the heart blood sent oxygen to the lungs so that blood cells can obtain more oxygen. The newly oxygenated blood travels from the lungs into the left atrium and the left ventricle. It then pumps it into the organs and tissues of the body. This oxygen provides the body with energy and is essential to keep the body healthy [1]. Hence to detect troubles related to heart is quite necessary for the medical field.

During heart disease diagnostic process, data mining and knowledge discovery approaches are explored. The system is proposed in this paper automatically captures previous health records of the patient and detects the tendency of heart disease. This paper applies Machine learning (ML) algorithms to identify patients with cardiac severity. Given a set of messages, ML methods can obtain information and later can use the acquired information to classify unknown new messages. Early heart disease can be predicted by utilizing supervised machine learning approaches those takes patient’s record as input. To address the problem of heart disease detection, classification techniques are implemented that maps input variable to target classes by considering training data. The input variables include several parameters such as patient’s age, gender, chest pain type, resting electrocardiographic results, maximum heart rate achieved, fasting blood sugar, serum cholesterol, the slope of the peak exercise ST segment, exercise-induced angina, old peak (ST depression induced by exercise relative to rest), number of major vessels (0-3) coloured by fluoroscopy, thalassemia, resting blood pressure. All these data turn out to be good predictors while identifying patients having cardiac problem symptoms. The predictive models can act as a tool to analyze the information of patients about their past health history records and predict their chances of being effected in cardiac trouble. This prediction will support doctors to make informed decisions and prescribe medicines and surgeries accordingly.

A recommended system is proposed in this paper that utilizes the past health record of a patient and provides early prediction of this disease. Timely detection and screening play a leading role in the prevention of heart attacks. A deep learning-based method is implemented in this paper as a recommended system for improving the efficiency in heart trouble prediction using medical data. A Recurrent Neural Network (RNN) with a feedback loop structure is often helpful in forecasting purpose. Gated Recurrent Unit (GRU) [2] based model is proposed in this paper while detecting the cardiac problem. This neural network classifier receives all interfering factors as features for heart disease prediction. This paper focuses on applying deep learning technique based stk-G model and traditional machine learning classifier SVM and K-NN to identify patients with CVD. Finally, in the experimental results section, it has been concluded that deep learning-based stk-G model performs well in compared with other benchmark classifiers such as SVM [3] and K-NN Classifier [4].

1.1 Related Work

By implementing data mining rules, data related to coronary illness is extracted from a large database. For this purpose, the weighted association implemented in [5]. Using rule mining algorithms on patients dataset, heart disease is predicted. Prediction results achieved 61% training accuracy and 53% testing accuracy.

In [6], historical medical data is utilised to Predict Coronary Heart Disease (CHD) using three supervised MLalgorithms such as Naïve Bayes (NB), SVM and Decision Tree (DT).South African Heart Disease dataset of 462 instances is used for prediction purpose. All these algorithms were performed using 10-fold cross-validation method. It is concluded in [6] that probabilistic NB classifier achieves better performance over other classifiers.

In [7], patients with heart failure (HF) are classified into three categories such as HF with preserved ejection fraction (HFPEF) and HF with reduced ejection fraction(HFREF).Several classification methods such as classification trees, bagged classification trees, random forests, boosted classification trees, and SVMs and for prediction, logistic regression, regression trees, bagged regression trees, random forests, and boosted regression trees are utilised for detecting patients with aforementioned three categories of heart failure. The empirical study concluded that the use of tree-based methods indicates superior performance over conventional classification and regression trees for predicting and classifying HF subtypes. However, these methods do not offer substantial improvements over logistic regression for predicting the presence of HFPEF.

K. Gomathi et al. in [8] predicted heart disease using NB Classifier and J48 classifier. They have concluded that NB classifier reaches an accuracy of 79% where J48 classifier reaches an accuracy of 77%. P.Sai Chandrasekhar Reddy et al. in [9] used ANN while predicting Heart disease. This ANN method is used to predict the condition of the patient considering various parameters like heartbeat rate, blood pressure, cholesterol etc.Different classification techniques such as J48, DT, KNN, SMO and NBwere implemented by Boshra Brahmi et al. In [10] while evaluating the prediction and diagnosis of heart disease. After comparing these classifiers concerning evaluation metrics J48 and decision tree have shown the best result for heart disease prediction. J48 classifier reached the best accuracy of 83.732%. A Hybrid Random Forest with Linear Model (HRFLM) is proposed in [11] for detecting patients with cardiac disease. This model uses all features without any restrictions on feature selection. Another study considered Arrhythmia which is irregular changes of normal heart rhythm as a prediction field. A CNN based model is proposed in this paper that accepts images of Electrocardiogram (ECG) signals to predict patients with Arrhythmia [12].

2. PROPOSED METHODOLOGY

The proposed methodology aims in predicting cardiac disease of patients with cardiac problem. The exact process of prediction using stk-G model is illustrated through a series of steps as follows.

2.1 Data Collection and Preprocessing

In this framework, *Heart disease dataset* from UCI machine learning repository [13] is utilized for predicting cardiac trouble tendency of a patient. The dataset can be formulated as a collection of attributes that include several criteria for detecting heart disease tendency such as patient's age, gender, chest pain type, resting electrocardiographic results, maximum heart rate achieved, fasting blood sugar, serum cholesterol, the slope of the peak exercise ST segment, exercise-induced angina, old peak (ST depression induced by exercise relative to rest), number of major vessels (0-3) coloured by fluoroscopy, thalassemia, resting blood pressure, target(probability of being heart patient). However, the attribute 'target' is utilized as the output class of the prediction. Following the diagram, Fig. 1 shows an overall understanding of the dataset. For obtaining a balanced dataset, preprocessing techniques such as missing value handling, scaling some attributes are performed. Presence of NaN values, irrelevant non-numeric symbols hamper the process of classification. Hence they are taken into consideration for pre-processing. Scaling of attributes means the attribute values should rely on a pre-defined range. In this case, attribute values are scaled within a range between 0 to 1. Performing these techniques will yield a transformed dataset that can be fitted to the classifier. The transformed dataset is partitioned into a training set and testing dataset. The training and testing dataset are obtained by partitioning the transformed dataset with the ratio of 7:3.

2.2 Classification Method

A classifier model maps the input variable to target classes after learning from training data. The objective of using classifier is to predict whether a patient has cardiac problem or not. A deep learning-based classification technique is proposed in this paper. Deep learning model focuses on combining multiple layers with linear or non-linear activation functions and trained together for achieving complex problem-solving approach. A RNN follows deep neural network architecture that processes both sequential and parallel information which in turn makes RNN superior to other traditional neural networks. Similar operations like the human brain can be simulated by incorporating memory cells to the neural network [14].

Following Fig. 2 shows a general structure of simple RNN. The following section discusses the description and implementation of the proposed model.

2.3 Description

In this framework, a stk-G model is proposed to detect heart disease at an early stage. GRU is an improved version of the standard recurrent unit with two types of gates- Update gate and Reset gate. How much the previous data needs to be passed to the future is determined by update gate and use of reset gate decides how much amount of past data to forget [2].

Stk-G model is constructed by incorporating GRU layers, dense layers as well asdropout layers.Use of dropout layers prevents this model from overfitting.Between hidden layers and output layers, a function called as activation function is used to perform diverse computations. The sigmoid activation function is a non-linear activation function that allows nodes to identify and learn critical structures from the data. Sigmoid activation function [15] transforms input data in the range of 0 to 1 and it is shown in equation(1).

$$f(x) = 1/(1 + exp^{-x}) \tag{1}$$

After configuring this stk-G-RNN model, the training process is executed. The training process goes through one cycle known as an epoch where the dataset is partitioned into smaller sections. An iterative process is executed through a couple of batch size that considers subsections of training dataset for completing epoch execution.

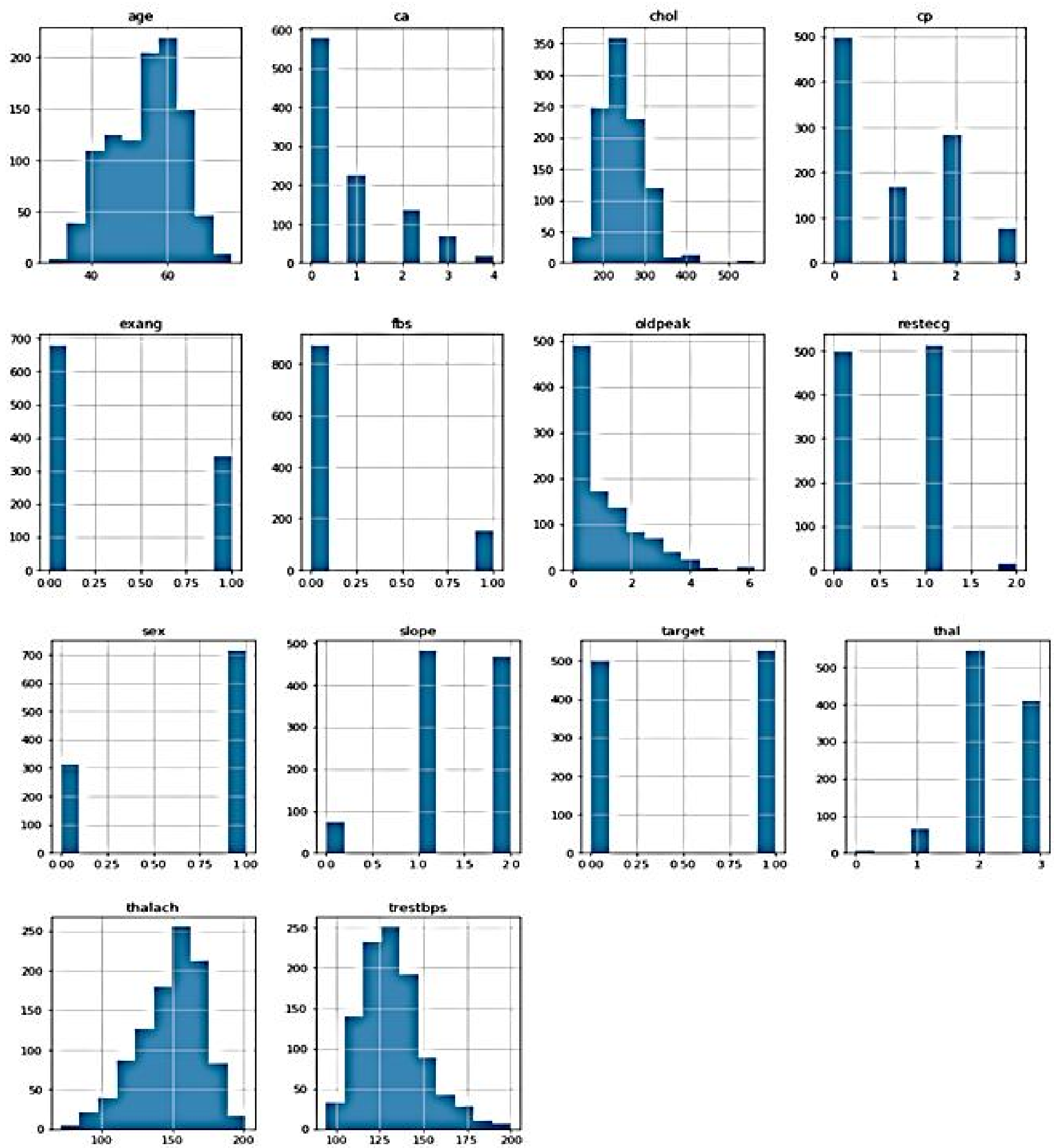

Fig. 1. Histogram interpretation of the collected dataset

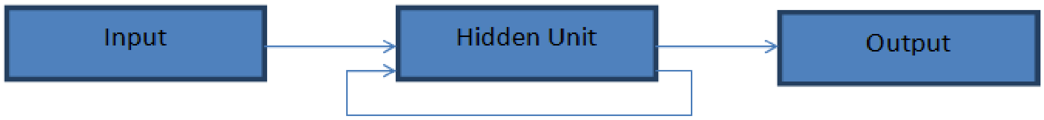

Fig. 2. General structure of simple RNN [14]

2.4 Implementation

While designing this model it is necessary to tune hyper-parameters to achieve maximized efficiency. This section describes the specification of the model along with its hyperparameters. The proposed model consists of having four GRU layers with 128,64,32,16 number of units respectively. Each of these layers is followed by layer having a drop out rate of 20%. Next, four dense layers are stacked in this model with 8,4,2,1 number of nodes respectively. The sigmoid function is applied to GRU layers as well as Dense layers. It is to be noted that, Dropout layers do not receive Sigmoid functions. Finally, these aforementioned layers are compiled using adam solver through 50 epochs and with a batch size of 32. Adjustment of the hyperparameters assists the model to obtain the best predictive result. All aforementionedinterferingattributes as features are given as input to this classifier model. The neural network receives a total of 98,825 number of parameters and trains those parameters to obtain a prediction. The architecture of the proposed framework is shown in Table 1.

After implementing the classifier model, the training set is fitted to the classifier model and later prediction is obtained for the test set. One prediction is obtained it is compared concerning performance evaluating metrics that are discussed in the following section.

2.5 Performance Measure Metrics

While evaluating the performance of a model, performance measure metrics are used. They are needed because it is necessary to identify the similarity measure of actual and predicted result of a classifier model. Following are the metrics that are required to justify the performance of the given model.

Accuracy [16] is a metric that ascertains the ratio of true predictions over the total number of instances considered. However, evaluating a model in terms of accuracy may not be enough since it does not consider wrong predicted cases. For addressing the above-mentioned problem, we yield two more metrics known as, Recall and Precision. *Precision* [16] identifies the ratio of correct positive results over the number of positive results predicted by the classifier. *Recall* [16] denotes the number of correct positive results divided by the number of all relevant samples. *F1-Score* or *F-measure* [16] is a parameter that is concerned for both recall and precision and it is calculated as the harmonic mean of precision and recall. Since F1-Score reflects the results of both precision and recall, in this paper F1-Score calculation is shown. *Mean Squared Error* (MSE) [17] is another evaluating metric which is used for measuring absolute differences between the prediction and actual observation of the test samples. A model having higher values of accuracy, F1-Score and lower MSE value indicates a better performing model.

Table 1. Architecture of Stk-G based model.

Layer (type)	Output Shape	Param #
gru_1 (GRU)	(None, 14,128)	49920
dropout_1(Dropout)	(None, 14,128)	0
gru_2 (GRU)	(None, 14,64)	37056
dropout_2(Dropout)	(None, 14,64)	0
gru_3 (GRU)	(None, 14,32)	9312
dropout_3(Dropout)	(None, 14,32)	0
gru_4(GRU)	(None, 16)	2352
dropout_4(Dropout)	(None, 16)	0
dense_1 (Dense)	(None, 8)	136
dense_2 (Dense)	(None, 4)	36
dense_3 (Dense)	(None, 2)	10
dense_4 (Dense)	(None, 1)	3

3. RESULTS AND DISCUSSION

The proposed model is implemented and evaluated in terms of the above-mentioned metrics. This model is later compared with the other two benchmark classifiers known as SVM [3] and K-NN [4]. The comparative study is shown in Table 2. From the comparative study, it is clear that the proposed model indicates much promising result over other traditionalclassifiers.

3.1 Analysis

Fig. 3 depicts the training process of the best model. During training, while fitting the training data into the classifier, the training process is evaluated in terms of accuracy as well as loss. For each epoch, data loss and accuracy is calculated. The best performing model will show accuracy to be increased as the number of epochs are increased. Similarly, the best model will show loss to be decreased when the number of epochs is increased. The training loss declines rapidly within 10 epochs and later decreases gradually as the number of epochs increases. After 50th epoch, it approaches a loss of 0.394. During training, this model starts from obtaining a lower value accuracy which is increased till 0.8542 after certain epochs. Once the training process is completed, the model is evaluated in terms of accuracy and loss during prediction. Prediction is obtained for the testing dataset and the loss, accuracy observed during prediction is summarized in Table 3.

Fig. 4 depicts the confusion matrix obtained for stk-G model. All True predictions, as well as wrong predictions, are shown in the confusion matrix. Low false prediction rates will signify the best problem-solving approach. In this case, it is observed that false prediction rates are quite lower than true prediction rates which signify the superiority of this model in heart disease prediction purpose.

Table 2. Comparative study among proposed model, SVM, and K-NN

Performance evaluating metrics	SVM	K-NN	Proposed stk-Gmodel
Accuracy	74.63%	75.12%	84.37%
F1-Score	0.75	0.75	0.84
MSE	0.25	0.25	0.16

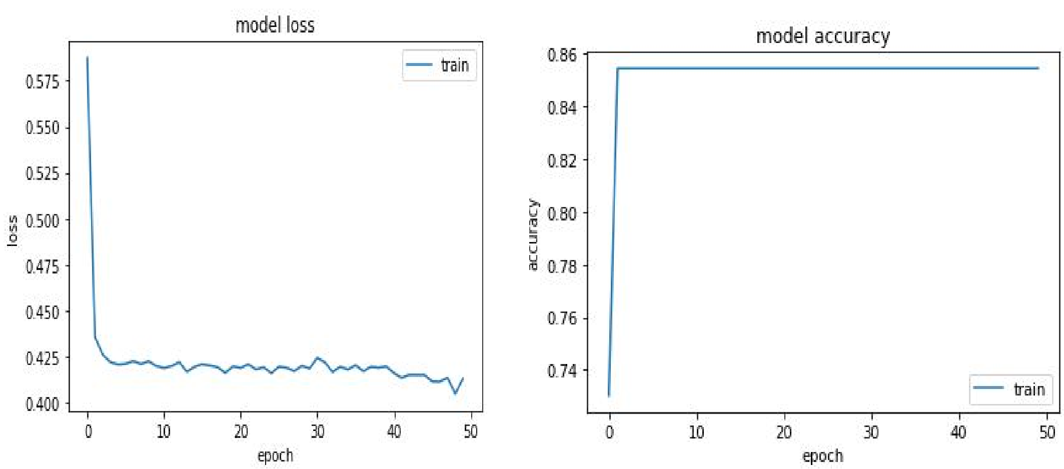

Fig. 3. Loss and accuracy obtained during each training epoch of stk-G model

True Negative 128	False Positive 28
False Negative 24	True positive 159

Fig. 4. Confusion matrix for stk-G model prediction

Table 3. Accuracy and loss for the testing dataset of Stk-G model

Test Loss	0.39405598243077594	Test Accuracy	84.37%

4. CONCLUSION

Heart disease is an impactful disease that needs to be handled carefully. Detecting the disease at an early stage is quite helpful in saving patients life. The objective of this study is to detect the feasibility of utilising previous medical records and determine the probability of being affected by cardiac arrest. Using deep learning methods, a stacked GRU layer-based model is proposed and implemented in this paper. Interfering attributes those have an impact on heart disease are considered while designing the model with necessary parameter tuning. The proposed method achieves promising result with an accuracy of 84.37%, F1-Score of 0.84 and MSE of 0.16.

CONSENT

As per international standard informed and written participant consent has been collected and preserved by the authors.

ETHICAL APPROVAL

As per international standard written ethical permission has been collected and preserved by the author(s).

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Sylva M, Van den Hoff MJB, Moorman AFM. Development of the human heart," Am. J. Med. Genet. Part A. 2014;164(6): 1347–1371.
DOI: 10.1002/ajmg.a.35896.
2. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. 2014;1–9.
3. Evgeniou T, Pontil M. Workshop on support vector machines: Theory and applications. Mach. Learn. Its Appl. Adv. Lect. 2001;249–257.
DOI: 10.1007/3-540-44673-7
4. Cunningham P, Delany SJ. K-Nearest neighbour classifiers. Mult. Classif. Syst. 2007;1–17.
DOI: 10.1016/S0031-3203(00)00099-6
5. Chauhan A, Jain A, Sharma P, Deep V. Heart disease prediction using evolutionary rule learning. Int. Conf. Computational Intell. Commun. Technol. CICT. Cict. 2018; 1–4.
DOI: 10.1109/CIACT.2018.8480271
6. Gonsalves AH, Thabtah F, Mohammad RMA, Singh G. Prediction of coronary heart disease using machine learning: An experimental analysis. ACM Int. Conf. Proceeding Ser. 2019;51–56.
DOI: 10.1145/3342999.3343015
7. Austin PC, Tu JV, Ho JE, Levy D, Lee DS. Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes. J. Clin. Epidemiol. 2013;66(4): 398–407.
DOI: 10.1016/j.jclinepi.2012.11.008
8. Kirmani M. Cardiovascular disease prediction using data mining techniques. Orient. J. Comput. Sci. Technol. 2017; 10(2):520–528.
DOI: 10.13005/ojcst/10.02.38
9. Sai PP, Reddy C. Heart disease prediction using ann algorithm in data mining. International Journal of Computer Science and Mobile Computing. 2017;6(4):168-172.
10. Bahrami B, Hosseini Shirvani M. Prediction and diagnosis of heart disease by data mining techniques. J. Multidiscip. Eng. Sci. Technol. 2015;2(2):3159–40.
11. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019;7:81542–81554.
DOI: 10.1109/ACCESS.2019.2923707
12. Izci E, Ozdemir MA, Degirmenci M, Akan A. Cardiac arrhythmia detection from 2d ecg images by using deep learning technique. TIPTEKNO 2019 - Tip Teknol. Kongresi. 2019;1–4.
DOI: 10.1109/TIPTEKNO.2019.8895011
13. Blake C, Keogh E, Merz CJ. UCI repository of machine learning databases. Department of Information and Computer Science, University of California, Irvine, CA; 1998.
Available:http://www.ics.uci.edu/~mlearn/MLRepository.html

14. Lipton ZC, Berkowitz J, Elkan C. A critical review of recurrent neural networks for sequence learning. 2015;1–38.
15. Nwankpa C, Ijomah W, Gachagan A, Marshall S. Activation functions: Comparison of trends in practice and research for deep learning. 2018;1–20.
16. Canbek G, Temizel TT, Sagiroglu S, Baykal N. Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights," 2nd Int. Conf. Comput. Sci. Eng. UBMK. 2017;821–826. DOI: 10.1109/UBMK.2017.8093539
17. HM, SMN. a review on evaluation metrics for data classification evaluations. Int. J. Data Min. Knowl. Manag. Process. 2015;5(2):01–11. DOI: 10.5121/ijdkp.2015.5201

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/57729