

An intelligent framework of Swine flu status prediction by rainforest algorithm

V. Kakulapati^{1[0000-0002-1753-3298]}, V. Kranthi Kumar¹, V. Sai Srikar¹³, Y. Swapnika Rao⁴,

Thierry Oscar Edoh

^{1,2,3,4} Sreenidhi Institute of Science and Technology

Yamnampet, Ghatkesar, Hyderabad, Telangana-501301

¹vldms@yahoo.com, ²kranthivankadaru@gmail.com

Abstract: The modernization has influenced people in these days in lifestyle and food habits are trying to defray from the healthier food which we have from an ancient culture, but are falling into indigent practices, and we have to find the reasons for the main causes of disease and how people are prone to illness. A variety of artificial neural network prototypes are examining in terms of their categorization effectiveness in a swine flu infection. The implementation results exhibit enhanced accuracy than conventional classification methods and are a reasonable and earlier diagnosis of swine flu. This method provides a suitable alternative of medical features for the identification of swine flu status. It also provides insight into the classification of individuals regarded as an “unidentified” phenotype on the origin of standard diagnosis methods. Scalability is the key challenge for large volumes of data, and we applied the Rainforest algorithm for improving the quality of classification.

Keywords: swine, flu, rainforest, artificial neural network, status, classification, diagnosis.

1. Introduction

H1N1 is also known as Swine flu, is a breathing illness which is the origin of infection viruses. This virus will spread to humans that regularly cause outbreaks of influenza in pigs. Those who suffer from the H1N1 virus have symptoms like severe cough, hungry less, running nose, and lethargic actions. This disease is contagious and rapidly spread among people through air and water. Predicting influenza is complex that includes determining whether the person is affected by the virus or not, intended for diagnosis. ANN classification gives prediction results that have to determine with a decision of ‘Yes’ or ‘No.’

The prevention of swine flu is very important as sources of serious health problems by recurrent effects throughout the world [1] because now a day’s Swine flu is increasing and also increasing death cases. The virus symptoms can fluctuate which requires hospitalization, whereas severe impediments like a high fever and breathing problems lead to death. The death caused by Swine flu is between 151,700 and 575,400 people during the year (2009) the virus transmitted. This number is increasing yearly. In 2019, to date, 493 people are dead in Telangana state only. Thousands of people are dying due to this virus, and the main reason for death is that they are facing breathing difficulty. In this work, we implemented a machine learning technique. An artificial neural network model is utilizing to the swine flu prediction that reduces the diagnosis cost and time. The main benefit of an artificial neural network is relying upon the stimulating function which can execute non-linear categorization provinces due to analogous nature and can be proficient even if the network fails. This technique is frequently preferring for its capability of outcomes from hidden records.

Then we apply the Random Forest algorithm, which a supervised learning method, utilized as a predictor of data for categorization and regression. In the categorization procedure algorithm develop many decision trees at the time of model training and build the category to facilitate the method of the classes yield by utilizing every single tree [2].

In this work, utilize an ensemble learning technique, i.e., Random Forest, for the prediction, which is a massive amount of decision trees whose result is the method of the results from each tree. This algorithm congregates the features above, is addressing, which can achieve by decisive repeatedly the amendment parameter, which is having many base classifiers that invent the collection and influences its

achievement. The implementation part of random forest builds an ensemble with the finest accuracy. The features as mentioned earlier allow to be completely incorporated into any investigative or healing because it enhances RF algorithm, which given that a high-performance categorization along with time, and efficient computational cost that efforts autonomously on the therapeutic issues and this can hold piercing or missing data, a general feature of health care datasets[18]. Random Forest model exhibits extensively elevated accuracy than the ANN classifier. These models are training by high dimensional input exhibits, comparable performance; however, the random forest was more efficient in terms of computational and time costs.

2. Related work

Intelligent prototype predicting the swine flu is established [3-5], and in this prototype utilized Naïve Bayes for categorizing the Swine flu patients of swine flu instead of the KNN algorithm and attain the classification accuracy is 63.3 %. One of the popular classification methods is SVM, which generally utilized a machine learning algorithm for the classification of swine flu narrated ailments [6-10]. A model [8] proposed, which contains three stages of categorization, utilizing the Support Vector Machine for the enhanced dissimilarity between the genuine tweets regarding flu and fake flu tweets. In this model, the first level is utilized to classify flu tweets data set to positive and negative tweets, and the second level is utilized to retrieve the flu-related tweets, whereas the last level utilized for the classification of infection. ANN is utilizing for classification for the new attribute selection technique for heart illness, categorization and different attribute selection techniques for particular state people [11]. To determine the difficulties in persistent and non-persistent based heart disease diagnosis, many research works are developed on machine learning techniques such as SVM, K-NN, Artificial neural network, decision tree, logistic regression, etc., [12-13] on non-persistent coronary illness diagnosis system, and because of this therapeutic decision method, the death rate of coronary illness disease is reduced [14]. Many research studies related to heart illness diagnosis based on machine-learning methods.

A three-phase ANN model [15] proposed to analyze heart disease in angina and attained a precision of classification is 88.89%, and this model is installing in healthcare providers. A collection-based predictive prototype of an artificial neural network [16] proposed to facilitate identifies coronary and utilized the analysis of statistical methods with the categorization system and attained 89.01% of accuracy, sensitivity is 80.09%, and specificity is 95.91% — an investigative method [17] designed coronary illness based on ML classifier multilayer perceptron. An artificial neural network determined the backpropagation learning algorithm and attribute selection algorithm.

In machine learning techniques, the most popular method is RF, which is based on aggregating bootstrap or bagging ([19], whereas bagging is a technique for creating several communicated apart from unmoving dissimilar training sets from training data, through enhancing accurateness. According to [20], it reduces Overfitting, fitting the prototype to the noise in the data set training data in its place of the essential association of characteristics, which consequential in an excessively multifaceted prototype and the significance of the poorer quality of attainment of prediction. RF utilizing the Classifiers or regressor is utilizing an arbitrary choice of features when creating decision trees, which is proposed by [21] called as random decision forests with an improvement of generalized more than one decision trees, which utilized bagging within the RF algorithms [22].

3. Methodology

In this work, Machine Learning classification algorithms and their theoretical background are discussed to classify the swine flu patients. We discussed Artificial Neural Network (ANN), and Random Forest classification algorithms of machine learning techniques are utilized for classification accuracy of patients.

3.1. Artificial Neural networks

These are analogous computational models involved in closely interrelated processing elements to the capability to take action to participation stimuli and to be trained to become accustomed to the surroundings. It is estimating the prototype based on the formation and tasks of biological neural networks. Data flows during the association influences the formation of Artificial neural network as a neural network transforms - or be trained, in logic - established on that participation and amount produced. These are reflecting on nonlinear statistical data representation devices where the compound associations among contributions and yields are prototypes are created.

3.2. Random Forest

Produce a forest of numerous trees and every tree on an autonomous model from the training data set.

At every node,

1. Choose k variables at random out of all K probable variables.
2. Finding the best split from the chosen m variables
3. Produce maximum classification of the tree
4. Estimating the average of the trees to acquire prediction.

Subsequent measures can estimate the proposed system classification performance. We calculate the sensitivity, specificity, precision, and accuracy of each algorithm.

Sensitivity: It described as the number of positive records is accurately classified.

Specificity: It described as the number of negative records is accurately classified

Precision: This is the percentage calculation of relevant records with the overall records.

Accuracy: It is the test data percentage that is classified by any algorithm.

4. Experimental analysis

For implementation purposes, we take the UCI machine repository data set, which contains symptoms like a high fever, running nose, headache, chill, etc. First, we implement the neural network on the swine flu data set by using two-node and five node layers. Then we apply the random forest algorithm for better classification accuracy.

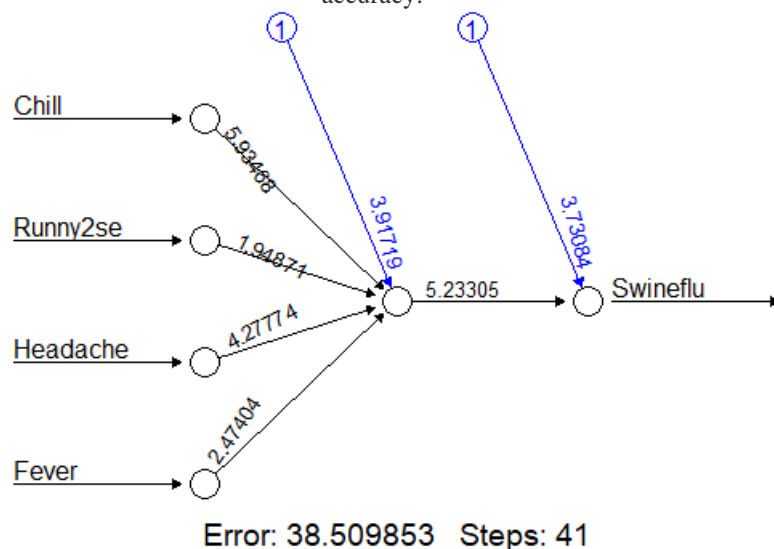


Fig 1: Neural n/w with 2nodes and 2layers of swine flu

Confusion matrix

```
Pred 2 1 2
      1 57 36
```

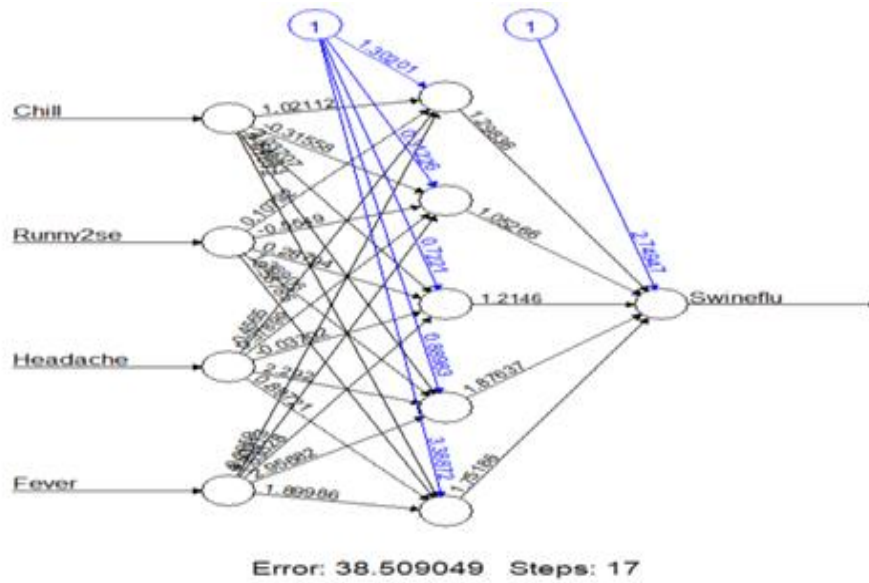


Fig 2 Swine flu neural n/w with 5 nodes in a layer

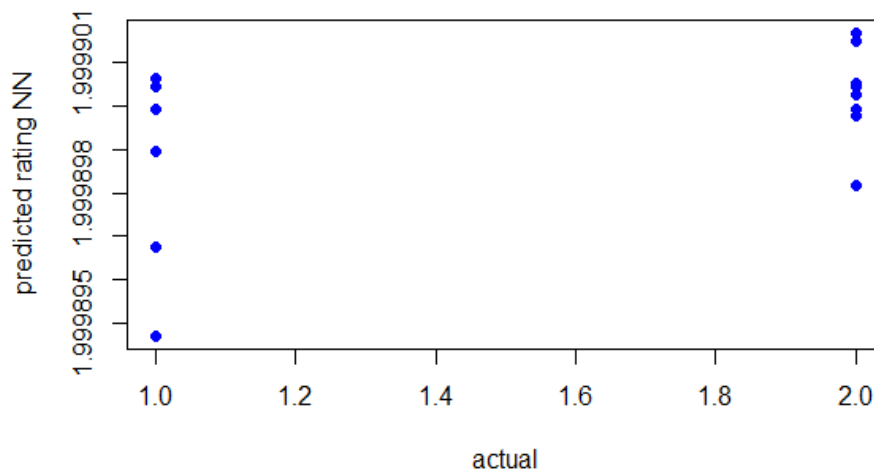


Fig 3. Swine flu prediction by using the neural network

Type of random forest: classification
 Number of trees: 500
 No. of variables tried at each split: 2
 OOB estimate of error rate: 1.39%

Table 1: Confusion matrix:

| S.no. | 1 class | 2 class | error |
|-------|---------|---------|-------------|
| 1 | 133 | 1 | 0.007462687 |
| 2 | 2 | 80 | 0.024390244 |

Confusion Matrix and Statistics

| | | |
|-------------------|-----------|---|
| | Reference | |
| Prediction | 1 | 2 |
| 1 | 133 | 2 |

2 1 80

Accuracy: 0.9861

Sensitivity: 0.9925

Specificity: 0.9756

Positive PredValue : 0.9852

Negative PredValue : 0.9877

Prevalence : 0.6204

Detection Rate : 0.6157

Detection Prevalence : 0.6250

Balanced Accuracy: 0.9841

'Positive' Class: 1

Reference

Prediction 1 2

1 51 1

2 1 30

Accuracy: 0.9759

95% CI: (0.9157, 0.9971)

No Information Rate: 0.6265

P-Value [Acc > NIR]: 1.758e-14

Kappa: 0.9485

Sensitivity: 0.9808

Specificity: 0.9677

Pos PredValue : 0.9808

Neg PredValue : 0.9677

Prevalence : 0.6265

Detection Rate : 0.6145

Detection Prevalence : 0.6265

Balanced Accuracy : 0.9743

'Positive' Class: 1

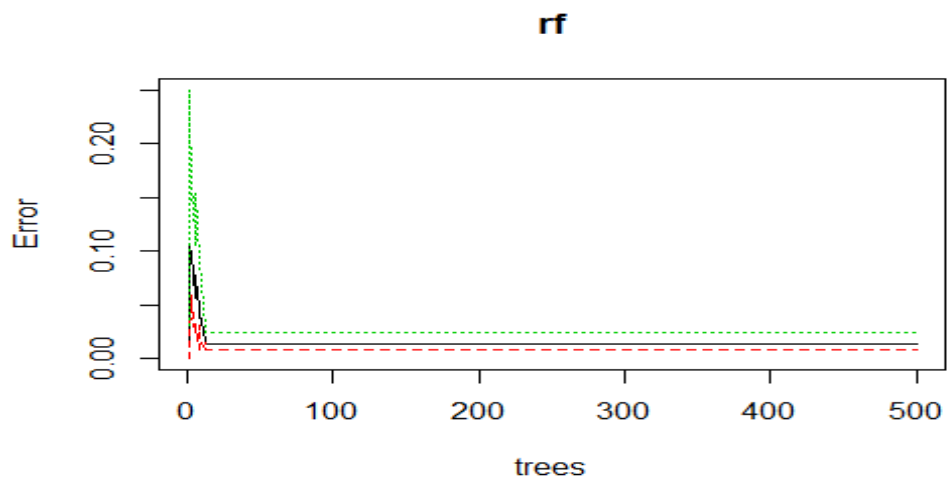


Fig 4. Predicting swine flu classification by random forest algorithm

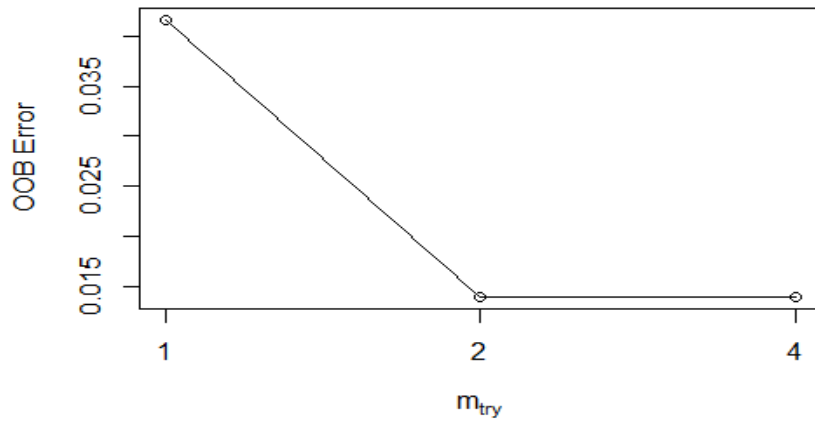


Fig 5: an increasing number of nodes by using a random forest algorithm

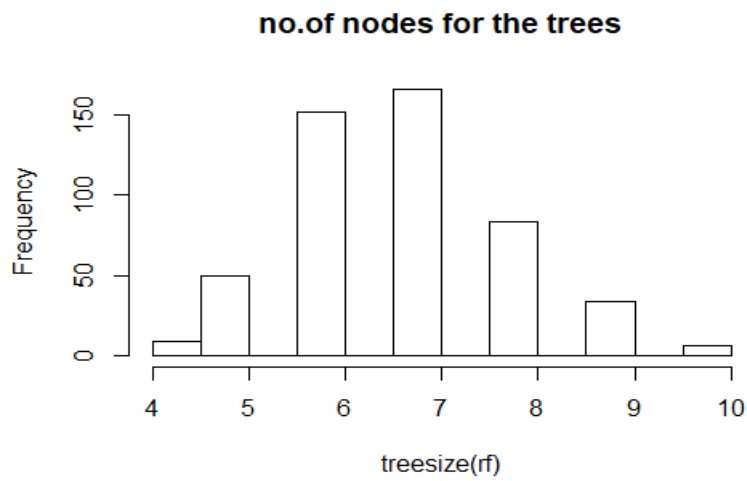


Fig 6: Histogram representation of RF algorithm

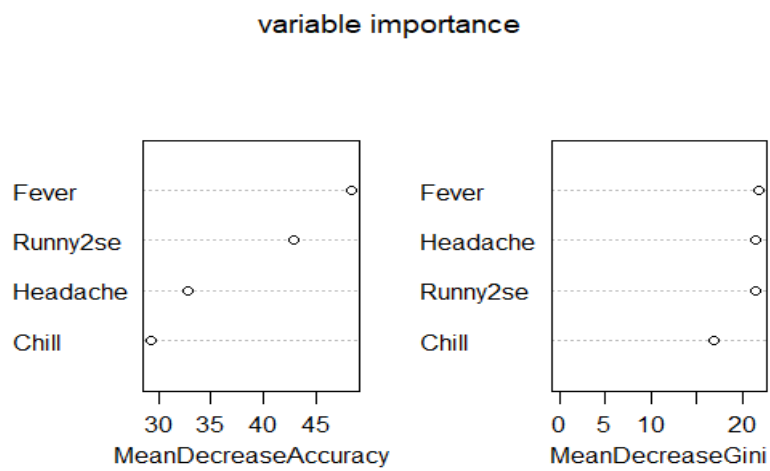


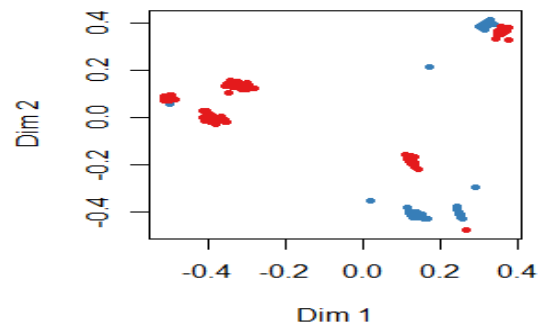
Fig 7: mean accuracy and Gini values of swine flu symptoms.

Table 2. Swine flu accuracy and Gini values symptom wise

| symptom | 1 | 2 | MeanDecreaseAccuracy | MeanDecreaseGini |
|-----------|----------|----------|----------------------|------------------|
| chill | 26.3183 | 24.3157 | 29.3527 | 16.90823 |
| runnynose | 41.32695 | 34.041 | 42.85651 | 21.48069 |
| headache | 26.61973 | 29.02683 | 32.72877 | 21.51662 |
| fever | 39.11247 | 37.53227 | 48.25632 | 21.86046 |

Table 3. Swine flu spitting and status prediction

| S.no. | Left node | Right node | Split var | Split point | status | prediction |
|-------|-----------|------------|-----------|-------------|--------|------------|
| 1 | 2 | 3 | Runnynose | 1.5 | 1 | <NA> |
| 2 | 4 | 5 | headache | 3.5 | 1 | <NA> |
| 3 | 6 | 7 | chill | 1.5 | 1 | <NA> |
| 4 | 8 | 9 | Fever | 1.5 | 1 | <NA> |
| 5 | 0 | 0 | <NA> | 0 | -1 | 1 |
| 6 | 10 | 11 | headache | 3.5 | 1 | <NA> |
| 7 | 12 | 13 | headache | 3 | 1 | <NA> |
| 8 | 0 | 0 | <NA> | 0 | -1 | 1 |
| 9 | 14 | 15 | chill | 1.5 | 1 | <NA> |
| 10 | 16 | 17 | headache | 2.5 | 1 | <NA> |

**Fig 7:** random forest two-dimensional representation of classification.

5. Conclusion

In this work, the swine flu prediction system is implemented utilizing the ensemble technique called random forest algorithm, which retrieves unseen data e from the swine flu data set. The classification prototypes are training and authenticated in contrast to an investigation dataset. The proposed prototype is intelligent to retrieve prototypes in terms of the status of prediction. Our proposed system accuracy obtained in the random forest algorithm is enhanced when compared with the accuracy of ANN. Our proposed approach attained an accuracy of 98.6 for the swine flu data. Our method exhibits enhanced accurateness in the swine flu prediction. Our proposed method does better than conventional categorization algorithms for the efficient classification of swine flu. It can effectively utilize in predicting the risk factors of swine flu, and to facilitate healthcare providers for the swine flu prediction.

6. Future enhancement

In the future, automation of swine flu prediction by using a large data set and also develops similarity measures of swine flu symptoms. We can utilize the clustering techniques based on location wise and predicting the swine flu and provide recommendations to prevent the disease which can be built using deep learning techniques. In the future, we will develop new prediction techniques to enhance classification accuracy and efficiency of swine flu patient privacy and provide a personalized recommendation to the patient for better diagnosis by applying nature-inspired optimization techniques.

REFERENCES

1. Rajoura, O. P., Roy, R., Agarwal, P., & Kannan, A. T. (2011). A study of the swine flu (H1N1) epidemic among health care providers of a medical college hospital of Delhi. *Indian Journal of Community Medicine*, July-September, 36(3), 187-190.
2. <http://www.datasciencecentral.com/profiles/blogs/random-forests-algorithm>.
3. Thakkar, Hasan and Desai, 2010,—Health care decision support system for swine flu prediction using Naïve bayes classifier, International Conference on Advances in Recent Technologies in Communication and Computing.,IEEE.
4. www.tb.evidence.org/documents/dxres/models/tb_diagnosics.pdf World Health Organization
5. http://www.ijarcse.com/docs/papers/Volume_4/6_June2014/V4I6-0235.pdf.
6. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Comput Biol*. 2015;11(10):1004513.
7. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: detecting influenza epidemics using twitter. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics*; 2011. p. 1568–76.
8. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS ONE*. 2013;8(12):83672.
9. Cui X, Yang N, Wang Z, Hu C, Zhu W, Li H, Ji Y, Liu C. Chinese social media analysis for disease surveillance. *Pers Ubiquit Comput*. 2015;19(7):1125–32.
10. Santos JC, Matos S. Analysing twitter and web queries for flu trend prediction. *Theor Biol Med Modell*. 2014;11(1):6.
11. M. Akhil Jabbar et al., “Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection” *Global Journal of Computer Science and Technology Neural & Artificial Intelligence* Volume 13 Issue 3 Version 1.0 Year 2013, Online ISSN: 0975-4172 & Print ISSN: 0975-4350.
12. S. Nazir, S. Shahzad, S. Mahfooz, and M. Nazir, “Fuzzy logic based decision support system for component security evaluation,” *International Arab Journal of Information Technology*, vol. 15, pp. 1–9, 2015.
13. S. Nazir, S. Shahzad, and L. Septem Riza, “Birthmark-based software classification using rough sets,” *Arabian Journal for Science and Engineering*, vol. 42, no. 2, pp. 859–871, 2017.
14. Methaila, P. Kansal, H. Arya, and P. Kumar, “Early heart disease prediction using data mining techniques,” in *Proceedings of Computer Science & Information Technology (CCSIT-2014)*, vol. 24, pp. 53–59, Sydney, NSW, Australia, 2014.
15. E. O. Olaniyi and O. K. Oyedotun, “Heart diseases diagnosis using neural networks arbitration,” *International Journal of Intelligent Systems and Applications*, vol. 7, no. 12, pp. 75–82, 2015.
16. R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675–7680, 2009.
17. M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Classification of heart disease using artificial neural network and feature subset selection,” *Global Journal of Computer Science and Technology Neural & Artificial Intelligence*, vol. 13, no. 11, 2013.
18. Evanthia E. Tripoliti, "Automated Diagnosis of Diseases Based on Classification: Dynamic Determination of the Number of Trees in Random Forests Algorithm", IEEE, 2012.
19. Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*.
20. Bramer, M.A. (2013). Principles of data mining. Springer.
21. Ho, T. K. (1995). Random decision forests. *International Conference on Document Analysis and Recognition*. 1, 278.
22. Breiman, L. (2001). Random Forests. *Machine Learning*. 45, 5-32.