Calculation of Heart Disease Analysis using Machine Learning Method

[1]Prof.Geetika Dikshit, [2]Dr. Ravindra Tiwari
[1]Ph. D. Scholar, [2]Associate Professor
[1]Department of Computer Science Engineering, LNCTU, Bhopal, India
[1]geetudikshit@gmail.com, [2]ravindratiwari.s@gmail.com

Abstract:- **The enhancing the number of persons through heart diseases along with deaths connected by these diseases which is the top reason behind the record of causes of death. Hopefully early finding as well as action can decrease the death risk of persons with heart disease along with people among heart disease. In recent years, regularly growth in the health field through the artificial intelligence technology development, the model of artificial intelligence along with classification and prediction ability that will put effort positively toward health workers and patients are being developed. The death status through heart disease is classified as per the record of patient's life information and the clinical data records of the patients which are included in the heart disease data set at heart failure rate. The objective of this learning is to estimate the death which is related with heart disease base on the patient's life information as well as patient's clinical data and to show to doctors and patients to near the beginning diagnosis or else near the beginning treatment processes. Various machine learning methods performs classification processes and success rates of these methods were shown. Unlike various machine learning algorithms have been experienced to get success rates among 73% to 78%. Amongst the applied methods or algorithms, the mainly victorious classification method is providing through the SVM (Support Vector Machine) algorithm with adaptive tree algorithm.**
Keywords: Disease; Machine learning; support vector machine; Dataset

1. Introduction

Health is considered as an approval which is depends on various factors along with proper care. Special civilizations build up for an only some centuries as well as died away. Decay and disease overtook them. In our country peoples are not liberated from health disease. In this present daily life has a crash on health disease like heart failure which is growing day by day. Amongst the entire health diseases, one of them heart disease is currently a key concern as numerous peoples are lose their own lives who are the facing the problem of heart disease [1].

This death rate increasing, this can be lessened through heart disease early detection. A lot of peoples are not aware of the heart disease's early detection, although numerous healthcare industries are regularly try to identify heart disease at an near the beginning stage. Though, it is not found yet at any earlier stage. While an outcome, it becomes approximately unfeasible to bear the treatment cost. It is currently a key concern amongst the various researchers toward the detection of heart disease at near the beginning stage [4, 3].

Generally heart disease is all through the world as well as it is the top disease that gives reason as high risk toward individual life. As per the Statistics report of Health of the health Ministry of the India, the 46.5 % cause of deaths in the world other than several contagious diseases in 2012

was declared as cardiovascular diseases. According to this report this rate is increasing. So we can say that, deaths due to heart diseases are predictable as 22.4 million till 2030 [2].

2. Literature Review

There are some papers studied and found some things about heart disease. Also we have to find heart disease is the one of the major issues for death of civilizations. In recent years, due to the successful classification and prediction outcomes, machine learning algorithms are frequently used in educational studies and they have to illustrate on data sets which is ready to analysis. Generally machine learning arise in various field of learning such as cryptology [15], time series [16, 5], and health [6, 17].

In this literature, there are numerous studies on the heart disease's classification as well as diagnosis. Out of these learning were carried out by frequently used machine learning algorithms [7 - 11]. Mohan et. al. preferred a machine learning algorithm which is combination of more two methods so we called as hybrid method where linear regression and random forest algorithms were merge to use heart disease prediction. There are used the data set which have 303 data lines and they used in their learning as well as the model success rate as good [12].

Haq et. al. as well used a machine learning algorithms which is also hybrid method for the heart diseases classification. They have to use the classifiers like as K-Nearest Neighbor algorithm, Logistic Regression algorithm, and Support Vector Machine algorithm (SVM), inside the combined model. As a success rate of model, they reached as well compare to others [13].

Kukar et. al. also used machine learning algorithms for the ischemic heart diseases diagnosis classifications. In their learning, they use sequential ECG test, and electrocardiogram (ECG), during myocardial scintigraphy, controlled exercise, and lastly coronary angiography imagery as used as data sets. This model has been build up in their learning provide accuracy as much better from previous one [14].

3. Proposed Work

There are data set hold 14 medical characteristics. Of these characteristics, the initially twelve characteristics are measured the same as independent variables as well as the very last characteristic like dependent variables. Here, our proposed model is subjected to forecast the dependent variable fraction as well as that is subjected to training toward independent variable. While the prediction of value is 1 or 0, the action execute will be classification.

There are a 303 total data lines inside the heart disease data set. In this data set out of total 303 data lines, 203 consider as training data and 100 data lines were arbitrarily owed as test data. Unlike classification approaches have been effort on the heart disease data set during order on the way to get the greatest outcome of the training. There are each algorithm's mathematical models; the classification achievement of values of confusion matrix which is obtained from the training.

3.1 Support Vector Machine Model

The term "h" this is called as hypothesis function. Here, dimension of classifications are represented by X and y parameters. The class +1, classified by point which is on or above the hyperplane, and another side class -1, classified by the point which is below the hyperplane. This is represented by mathematical equation or model which shown in below:

$$h(x_i) = \begin{cases} -1, \text{ if w.x} + b < 0 \\ +1, \text{ if w.x} + bx >= 0 \end{cases}$$

Where, the number of features had represented by n parameter and w is represented the point which is on the hyperplane. If this used lying on the soft-margin classifier while selecting a adequately little value intended for lambda given in the way the hard-margin classifier intended for input data which is linearly-classifiable.

$$\frac{1}{n}\sum_{i=1}^{n} \max(0,1 - yi(w.xi - b)) + Y||w||^2$$

This is also represented by mathematical equation or model that permits these mentions action to be performed which is shown above:
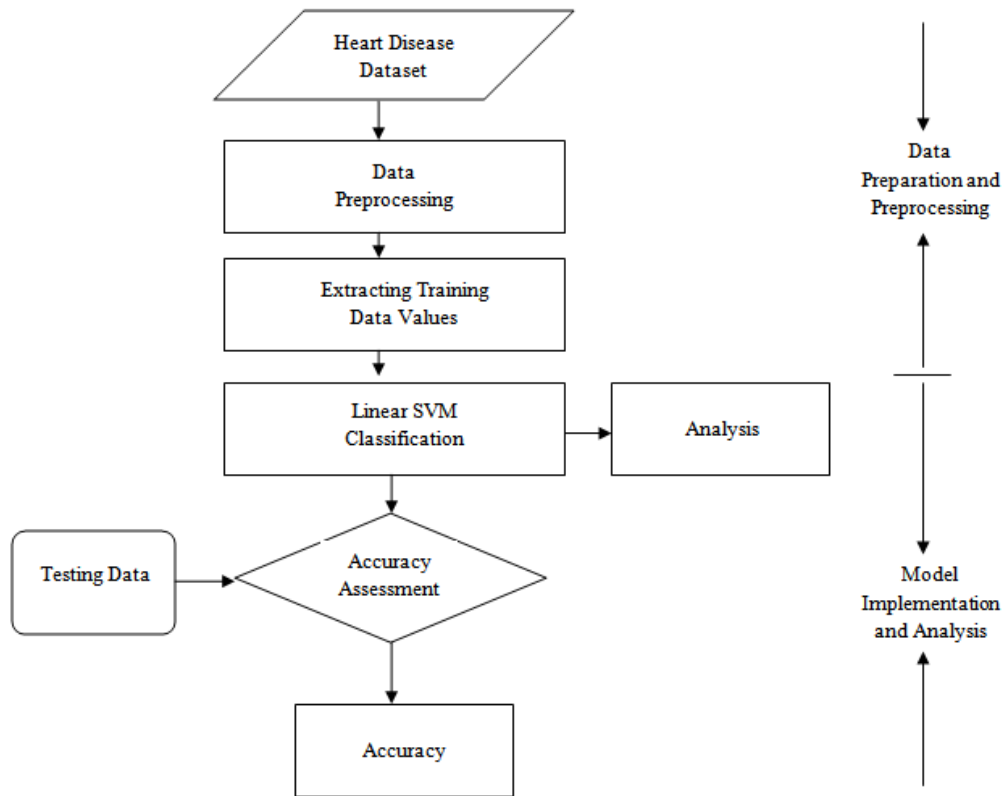

Figure 1: Flow Chart of Proposed Model

Our model is Tree Partitioning Adaptive Support Vector Machine was applied toward the learned data furthermore obtained an accuracy value was 0.83. This model is effective or gets more accurate result as accuracy from various other models. Figure 1 show the flow chart of our proposed model.

4. Result Discussion

Here, we have to use the Support Vector Machine and an algorithm which is based on tree partitioning with filtering approach that's called as TPASVM (Tree Partitioning Adaptive Support Vector machine). In this method, we have to complete our process in two steps; firstly we have to train our data as build a binary decision tree, secondly we have done testing phase. We have to use the heart disease data set with these attribute such as age, sex, cp, trestbpl, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, and thal, with 303 rows. We have to use the model SVM with linear kernel function type and linear kernel scale is automatic then we have to get accuracy 82.8% with training time 4.974 sec as well as prediction speed is -3300 obs/sec. This model will be executed on MATLAB 2015a version. This result as much better than from our base Tree model which has accuracy 78.2% in three phase as shown in figure below-
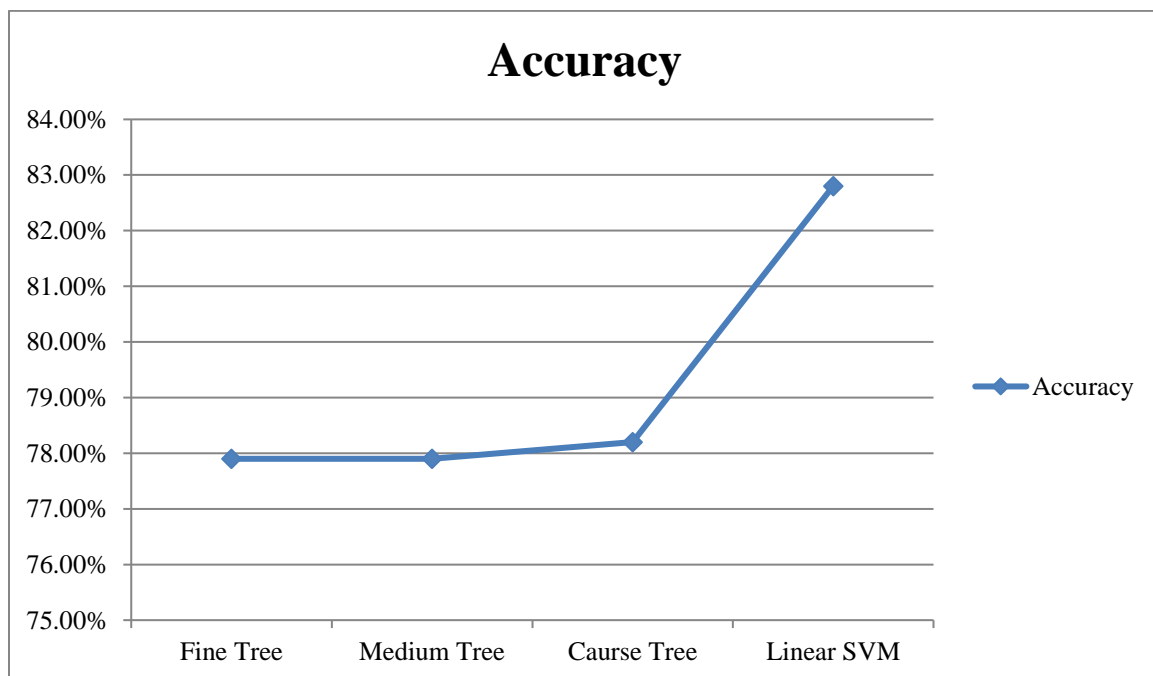

Figure 2: Accuracy values by type of algorithms

This developed model measured as success rate with 82.8%, which cannot be measured as bad, in another hand the algorithms considered and applied in this learning, the developed model will provide higher successes rate. The reason behind of this success rate towards stay on at this stage is measured to be the various numbers of data within the heart disease data set. Here, 203 data lines were accessible towards the models like training data. Furthermore, as the quantity of data

within the data set enhance, the model learning ability will obviously get better as well as the value of accuracy will be improve accordingly as shown in figure 2.

5. Conclusion

In this learning, we have to investigational studies about death conditions arise during heart failure which is conducted as classification using various machine learning approaches according to the life information and measurement values which is achieved from individuals. The heart_diesease_info data set is used in the learning. In this data set, 14 features are definite as input data along with one considered as target value. The probable outcome as result from the classification is to be capable to categorize the person's death conditions due to the very big reason which is heart disease as an algorithm's outcome subjected toward input values. Out of the 303 data rows in the given dataset is divided into two parts first one is training data which is 67% of total as well as test data which is 33% of the total. The maximum accuracy value get from various approaches has been improved to 82.8%. In future learning, it is expected to augment the data amount in the given data set and boost the closing success rate toward above 90%.

References

[1] Davide Chicco, Giuseppe Jurman: "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone". BMC Medical Informatics and Decision Making 20, 16 (2020).

[2] Türkiye Cumhuriyeti Sağlık Bakanlığı Sağlık İstatistikleri Yıllığı, 2015

[3] Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC medical informatics and decision making*, *20*(1), 16.

[4] Ahmad, T., Munir, A., Bhatti, S. H., Aftab, M., & Raza, M. A. (2017). Survival analysis of heart failure patients: A case study. PloS one, 12(7), e0181001.

[5] He, L., Chen, S., Liang, Y., Hou, M., & Chen, J. (2020). Infilling the missing values of groundwater level using time and space series: case of Nantong City, east coast of China. Earth Science Informatics, 1-15.

[6] Zemmal, N., Azizi, N., Sellami, M., Cheriguene, S., Ziani, A., AlDwairi, M., & Dendani, N. (2020). Particle Swarm Optimization Based Swarm Intelligence for Active Learning Improvement: Application on Medical Data Classification. Cognitive Computation, 1-20.

[7] Yin, H. (2020). Smart Healthcare Via Efficient Machine Learning (Doctoral dissertation, Princeton University).

[8] Anand, H., Anand, A., Das, I., Rautaray, S. S., & Pandey, M. (2020, July). Hridaya Kalp: A Prototype for Second Generation Chronic Heart Disease Detection and Classification. In International Conference on Innovative Computing and Communications (pp. 321-329). Springer, Singapore.

[9] Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H., & Gutierrez, J. (2017, July). A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease. In 2017 IEEE Symposium on Computers and Communications (ISCC) (pp. 204-207). IEEE.

[10] Buettner, R., & Schunter, M. (2019, October). Efficient machine learning based detection of heart disease. In 2019 IEEE International Conference on E-health Networking, Application & Services (HealthCom) (pp. 1-6). IEEE.

[11] Parthiban, G., & Srivatsa, S. K. (2012). Applying machine learning methods in diagnosing heart disease for diabetic patients. International Journal of Applied Information Systems (IJAIS), 3(7), 25-30.

[12] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542-81554.

[13] Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018

[14] Kukar, M., Kononenko, I., Grošelj, C., Kralj, K., & Fettich, J. (1999). Analysing and improving the diagnosis of ischaemic heart disease with machine learning. *Artificial intelligence in medicine*, *16*(1), 25-50.

[15] Al Shibli, M. (2020). Hybrid Artificially Intelligent Multi-Layer Blockchain and Bitcoin Cryptology (AI-MLBBC): Anti-Crime-Theft Smart Wall Defense. In Encyclopedia of Criminal Activities and the Deep Web (pp. 1089-1111). IGI Global.

[16] Jiang, Y., Bao, X., Hao, S., Zhao, H., Li, X., & Wu, X. (2020). Monthly Streamflow Forecasting Using ELM-IPSO Based on Phase Space Reconstruction. Water Resources Management, 1-17.

[17] Hsu, C. N., Liu, C. L., Tain, Y. L., Kuo, C. Y., & Lin, Y. C. (2020). Machine Learning Model for Risk Prediction of Community-Acquired Acute Kidney Injury Hospitalization from Electronic Health Records: Development and Validation Study. Journal of Medical Internet Research, 22(8), e16903.