e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:03/March-2022 Impact Factor- 6.752 www.irjmets.com

DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS

Prof. Swati Dhabarde*1, Rohit Mahajan*2, Satyam Mishra*3, Sanjiv Chaudhari*4, Satish Manelu*5, Prof. Dr. N S Shelke*6

*2,3,4,5Student, Dept. of Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, India.

*1,6Professor, Dept. of Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, India.

ABSTRACT

With big data growth in biomedical and healthcare communities, accurate analysis of medical data is devisee for early complaint discovery, patient care, and community services. One similar perpetration of machine literacy algorithms is in the field of healthcare. Medical installations need to be advanced so that better opinions for patient opinion and treatment options can be made. Accurate and on- time analysis of any health-related problem is important for the forestallment and treatment of the illness. The traditional way of opinion may not be sufficient in the case of a serious disease.

Still, supervised machine literacy (ML) algorithms have showcased significant eventuality in surpassing standard systems for complaint opinion and abetting medical experts in the early discovery of high- threat conditions. In this literature, the end is to fete trends across colorful types of supervised ML models in complaint discovery through the examination of performance criteria.

Developing a medical opinion system grounded on machine literacy (ML) algorithms for vaticination of any complaint can help in a more accurate opinion than the conventional system. We've designed a complaint vaticination system using multiple ML algorithms. The data set used had further than 230 conditions for processing. Grounded on the symptoms, age, and gender of an individual, the opinion system gives the affair as the complaint that the existent might be suffering from. By relating significant patterns and detecting correlations and connections among numerous variables in huge databases, the use of colorful data mining tools and machine literacy approaches has changed healthcare associations. It serves as an important instrument in the medical sector, furnishing and comparing being data for the unborn course of action. This technology combines multiple logical methodologies with ultramodern and complex algorithms, allowing for the disquisition of massive quantities of data. Our opinion model can act as a croaker for the early opinion of a complaint to insure the treatment can take place on time and lives can be saved.

Keywords: Decision Tree, Logistic Regression, Naive Bayes, Random Forest, SVM.

I. INTRODUCTION

Medicine and healthcare are some of the most pivotal corridor of the frugality and mortal life. There's a tremendous quantum of change in the world we're living in now and the world that was a many weeks back. Everything has turned horrible and divergent. There are also some remote townlets which warrant medical installations. Virtual croakers are board- certified croakers who choose to exercise online via videotape and phone movables, rather than in-person movables but this isn't possible in the case of exigency. Machines are always considered better than humans as, without any mortal error, they can perform tasks more efficiently and with a harmonious position of delicacy. A complaint predictor can be called a virtual croaker, which can prognosticate the complaint of any case without any mortal error. Machine literacy offers veritably accurate styles for determining conditions which have large and well collected databases. Machine literacy in the field of drug is a veritably active area of exploration. Despite that, there's no good mobile app available in the request for complaint vaticination which can help croakers in their day to day life. So we've tried to develop a mobile app which can help them. Still, mobile operations present some challenges sort of a stoner can not input numerous figures of input fields. For, eg UCI heart complaint dataset has roughly 75 features. So the main challenge before us was to reduce the number of features to an applicable position. We've presented our approach during this paper of point birth by using multi algorithm.

In this study, we probe studies that use further than one supervised ML model for each complaint recognition problem. This approach renders further comprehensiveness and perfection because the evaluation of the performance of a single algorithm over colorful study settings induces bias which generates squishy results. The analysis of ML models are going to be conducted on many conditions located at bottom, order, bone, and brain. The stylish performing ML models in respect of each complaint will be concluded.

II. EXISTING SYSTEM

Vaticination using traditional complaint threat model generally involves a machine literacy and supervised learning algorithm which uses training data with the markers for the training of the models. High threat and Low threat case bracket is done in groups test sets. But these models are only precious in clinical situations and are extensively studied. A system for sustainable health monitoring using smart apparel by Chenet.al. He completely studied miscellaneous systems and was suitable to achieve the stylish results for cost minimization on the tree and simple path cases for miscellaneous systems.

The information of case's statistics, test results, and complaint history is recorded in EHR which enables to identify implicit data-centric results which reduce the cost of medical case studies. Bates etal. propose six operations of big data in the healthcare field. Being systems can prognosticate the conditions but not the subtype of conditions. It fails to prognosticate the condition of people. The prognostications of conditions have beennon-specific and indefinite.

RELATED WORK

A structural model and a collection of tentative chances are used by Bayesian classifiers. They make the supposition that the benefactions of all factors are independent. It first calculates the previous probability for each class, and also applies the circumstance of each variable value to an unknown script. A Bayes network classifier is erected on a Bayesian network, which reflects a common probability distribution over a set of order characteristics. The SVM system and the Nave Bayes fashion were used to prognosticate order complaint. The authors tried to classify colorful stages of order complaint using the suggested ANFIS algorithm. The study's purpose was to design an effective categorization algorithm using several assessment criteria similar as delicacy and prosecution time. While the SVM Algorithm handed advanced bracket delicacy, the Nave Bayes fared better since it produced results in lower time. The results show that SVM outperforms the Nave Bayes Approach in prognosticating renal illness.

The fuzzy fashion with a class function was used to read cardiac complaint [2]. Using the Fuzzy KNN Classifier, the authors tried to exclude nebulosity and query from data. The 550- record dataset was separated into 25 classes, with each class having 22 particulars. The dataset was separated into two equal corridor training and testing. The fuzzy KNN methodology was enforced after pre-processing ways were used. This fashion was examined using several assessment criteria similar as delicacy, perfection, and recall, among others. Grounded on the data, it was discovered that the fuzzy KNN classifier outperformed the KNN classifier in terms of rigor. For the vaticination of cardiac complaint, a new fashion grounded on the ANN algorithm was cooked [3]. The experimenters created an interactive vaticination system grounded on categorization using an artificial neural network algorithm and taking into account the thirteen most important clinical parameters. The suggested system proved effective for prognosticating heart complaint with an accuracy of 80 and can be veritably useful for healthcare interpreters.

Authors in [4] presented an automated approach for answering delicate inquiries for heart complaint vaticination. The Naive Bayes methodology was used to produce this intelligent system in order to give quick, better, and more accurate issues. It might prop croakers in making clinical judgments about heart attacks. This system may be enhanced by including SMS functionality, erecting Android and IOS mobile operations, and including a trendsetter in the order.

Diabetes and breast cancer were diagnosed by incorporating the adaptivity specific into support vector machines [5]. The thing was to offer a rapid-fire, automated, and adaptable individual system using adaptive SVM. To achieve better results, the bias value in conventional SVM was changed. The suggested classifier produced affair in the form 'if-then' rules. The proposed system was used to diagnose diabetes and bone cancer, and it handed 100 right bracket rates for both conditions. Unborn exploration should concentrate on developing more effective ways for changing the bias value in conventional SVM.

For the vaticination of type 2 diabetes, a mongrel model grounded on clustering followed by bracket was proposed [6]. For prediction, the suggested model uses K- means clustering and the C4.5 bracket system withk-foldcross-validation. The model generated encouraging results with a bracket delicacy of 88.38 percent using the mongrel fashion, which might be largely useful for clinicians in making applicable clinical choices related to diabetes..

III. PROPOSED SYSTEM

In this paper, we've combined the structure and unshaped data in healthcare fields that let us assess the threat of complaint [7]. By using statistical knowledge, we could determine the major habitual conditions in a particular region. In the case of unshaped textbook data, we elect the features automatically with the help of multi algorithm which is Decision Tree, Naïve Bayes, Logistic Regression, Random Forest, SVM. We propose a multi algorithm for both structured and unshaped data.

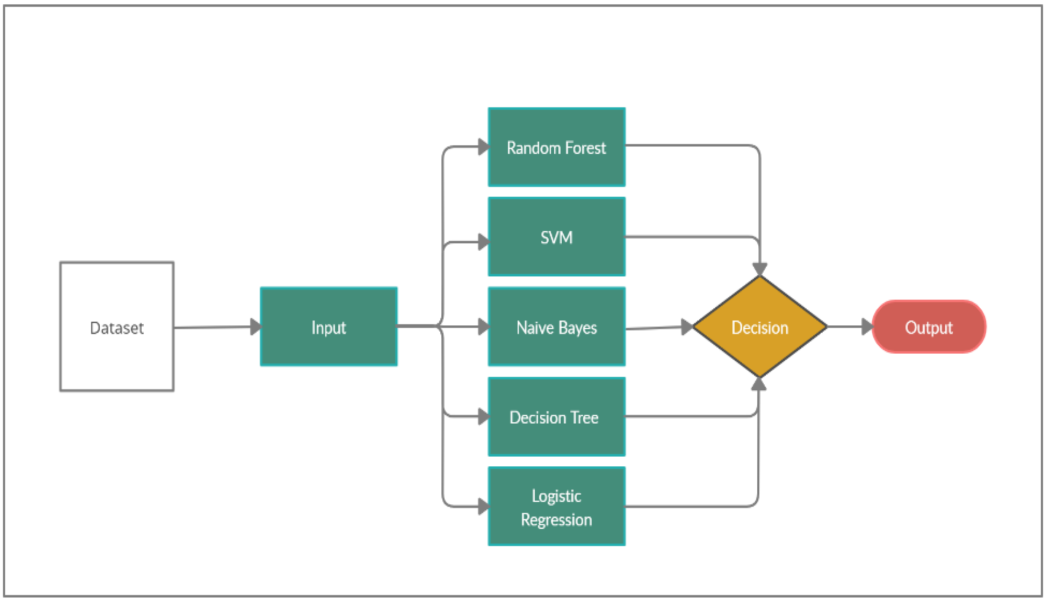

Figure 1

The decision taken from the multi algorithm stylish 3 results are shown in the affair with its delicacy and is stored in the database if for once references. The delicacy we reached by using multiple algorithm is 93.24%.

Logistic Regression

Logistic regression is another important supervised ML algorithm used for binary classification problems. The stylish way to suppose about logistic regression is that it's a linear regression but for classification problems. Logistic regression basically uses a logistic function defined below to model a binary output variable. The primary difference between direct retrogression and logistic regression is that logistic regression's range is bounded between 0 and 1. In addition, as opposed to direct regression, logistic regression doesn't bear a linear relationship between inputs and output variables. This is due to applying a nonlinear log metamorphosis to the odds rate.

$$Logistic\ function = \frac{1}{1+e^{-y}} \quad \text{--(1.1)}$$

In the logistic function equation, y is the input variable. Let's feed in values – 20 to 20 into the logistic function. As illustrated in Fig.5.17, the inputs have been transferred to between 0 and 1.

As opposed to direct regression where MSE or RMSE is used as the loss function, logistic regression uses a loss function appertained to as “maximum liability estimation (MLE)” which is a tentative probability. If the probability is lesser than0.5, the vaticinations will be classified as class 0. Else, class 1 will be assigned. Before going through logistic regression derivate, let's first define the logit function. Logit function is defined as the natural log of the odds. A probability of0.5 corresponds to a logit of 0, chances lower than 0.5 correspond to negative logit values, and chances lesser than 0.5 correspond to positive logit values. It's important to note that as illustrated inFig.5.17, logistic function ranges between 0 and 1 (W ∈ [0,1]) while logit function can be any real number from minus perpetuity to positive perpetuity (W ∈ (− ∞, ∞)).

$$odds = \frac{W}{1-W} \rightarrow \text{logit}\,(W) = \ln\left(\frac{W}{1-W}\right) \qquad \text{-- (1.2)}$$

Let's set logit of W to be equal to my + b, therefore:

$$\text{logit}\,(W) = my + b \rightarrow my + b = \ln\left(\frac{W}{1-W}\right)$$

$$\left(\frac{W}{1-W}\right) = e^{(my+b)} \rightarrow P = \frac{e^{(my+b)}}{1+e^{(my+b)}} \rightarrow P(y) = \frac{1}{1+e^{-(my+b)}} \qquad \text{--(1.3)}$$

Random Forest

As the name suggests, "Random Forest" is a classifier that contains a number of decision trees on colorful subsets of the given dataset and takes the average to ameliorate the prophetic delicacy of that dataset." Rather of counting on one decision tree, the arbitrary forest takes the vaticination from each tree and grounded on the maturity votes of vaticinations, and it predicts the final affair.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \text{--(2)}$$

Naive Bayes

Naive Bayes classifiers are a collection of classification algorithms based on Bayes Theorem. It isn't a single algorithm but a family of algorithms where all of them partake a common principle, i.e. every pair of features being classified is independent of each other.

Decision Tree

Decision tree is the most influential and popular tool for classification and vaticination. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each terminal node holds a class label.

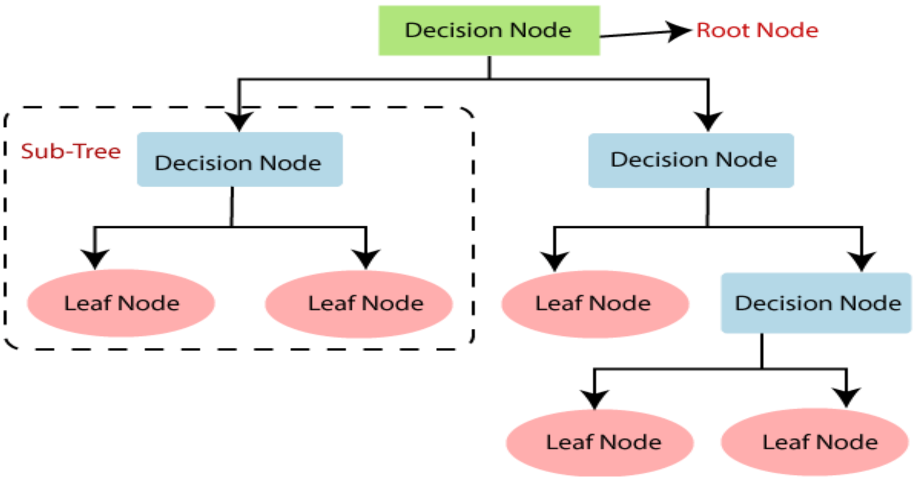

Figure 2

Support Vector Machine

Support Vector Machine (SVM) is a fairly simple Supervised Machine Learning Algorithm used for classification and/ or regression. It's further preferred for classification but is occasionally veritably useful for regression as well. Principally, SVM finds a hyperactive- plane that creates a boundary between the types of data. In 2-dimensional space, this hyperactive- plane is nothing but a line.

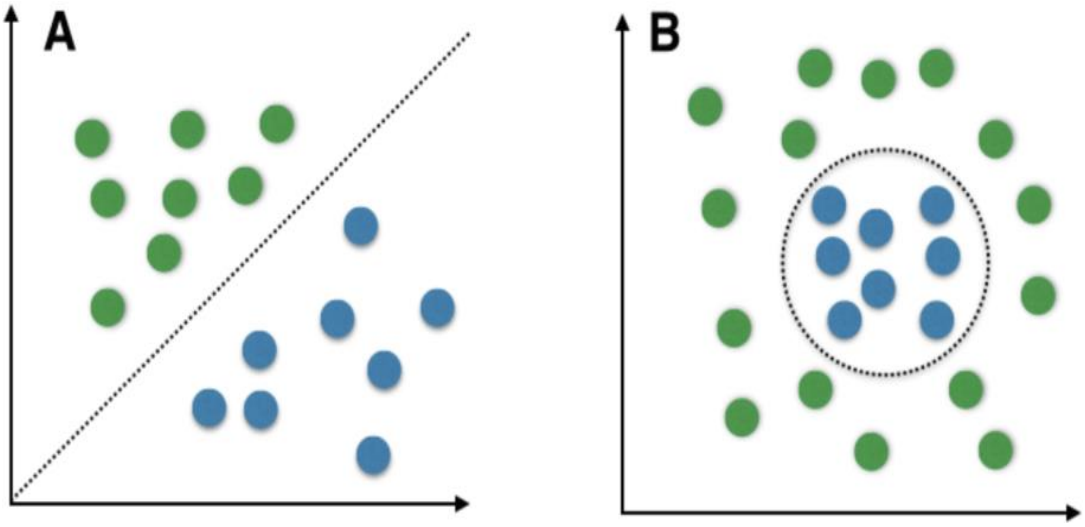

Figure 3

IV. SYSTEM ARCHITECTURE

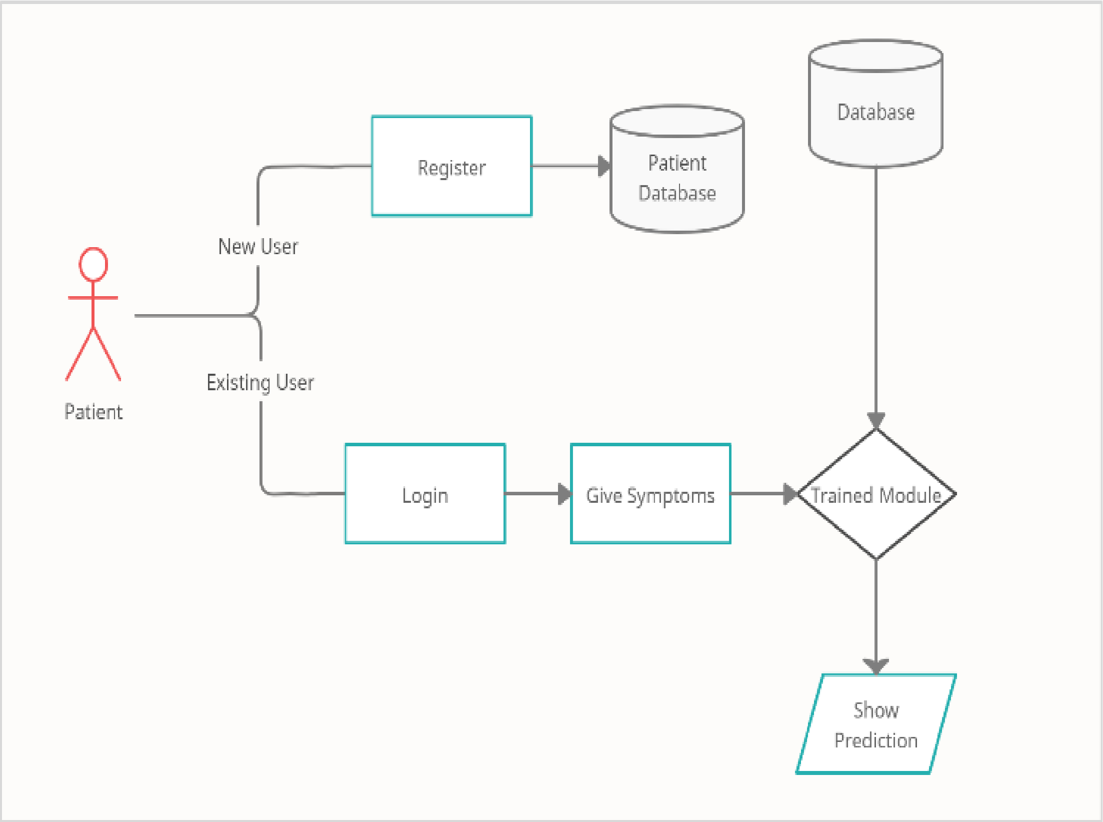

Figure 4

V. CONCLUSION

With the proposed system, advanced delicacy can be achieved. We not only use structured data, but also the textbook data of the case grounded on the proposed multi algorithm which is Decision Tree, Naïve Bayes, Logistic Regression, Random Forest, SVM. To find that out, we combine both data, and the delicacy rate can be reached up to 93.24%. None of the being system and work is concentrated on using both the data types in the field of medical big data analytics. We propose multi clustering algorithm which is Decision Tree, Naïve Bayes, Logistic Regression, Random Forest, SVM Algorithms for both structured and unshaped data. The complaint threat model is attained by combining both structured and unshaped features.

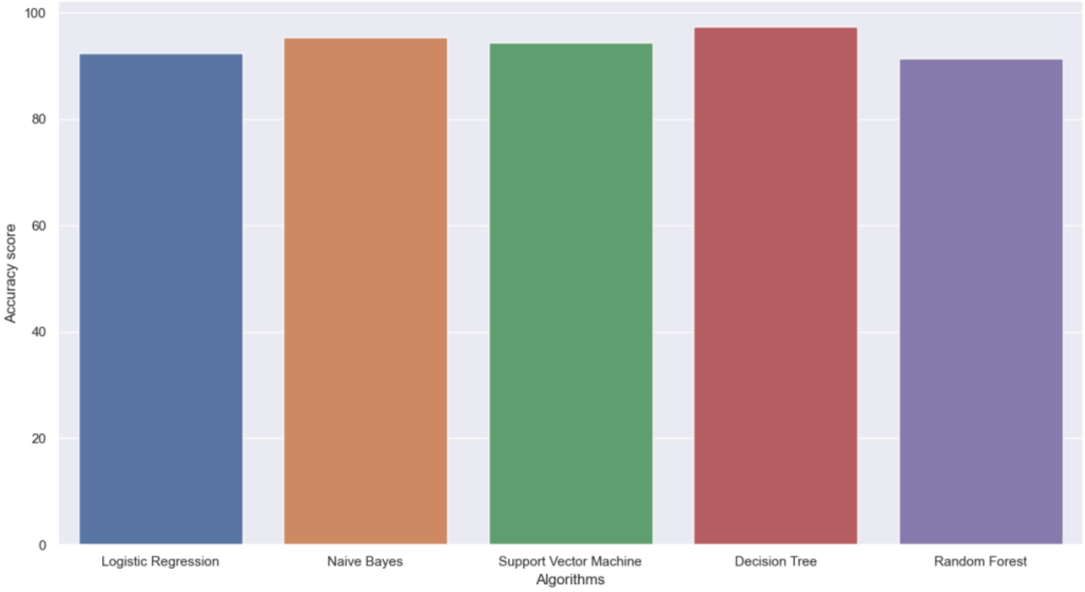

Figure 5

VI. REFERENCES

[1] H. Barakat, P. Andrew, Bradley, H. Mohammed Nabil Barakat, Intelligible support vector machines for diagnosis of diabetes mellitus, IEEE Trans. Inf. Technol. Bio Med. J. 14 (4) (2009) 1–7.

[2] R. Tina Patil, S.S. Sherekar, Performance analysis of Naive bayes and J48 classification algorithm for data classification, Int. J. Comput. Sci. Appl. 6 (2) (2013) 256–261.

[3] Shruti Ratnakar, K. Rajeswari, Rose Jacob, Prediction of heart disease using genetic algorithm for selection of optimal reduced set of attributes, Int. J. Adv. Comput. Eng. Netw. 1 (2) (2013) 51–55.

[4] S. Grampurohit, C. Sagarnal, Disease prediction using machine learning algorithms, 2020 Int. Conf. Emerg. Technol. (INCET) (2020) 1–7, https://doi. org/10.1109/INCET49848.2020.9154130.

[5] R.J.P. Princy, S. Parthasarathy, P.S. Hency Jose, A. Raj Lakshminarayanan, S. Jeganathan, Prediction of Cardiac Disease using Supervised Machine Learning Algorithms, in: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 570–575, https://doi.org/10.1109/ ICICCS48265.2020.9121169.

[6] P. Deepika, S. Sasikala. Enhanced Model for Prediction and Classification of Cardiovascular Disease using Decision Tree with Particle Swarm Optimization, 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 1068-1072, doi: 10.1109/ ICECA49313.2020.9297398

[7] MIN CHEN , (Senior Member, IEEE), YIXUE HAO, KAI HWANG , (Life Fellow, IEEE), LU WANG , AND LIN WANG, Disease Prediction by Machine Learning Over Big Data From Healthcare Communities, in:2017 IEEE Intellectual Property Rights, 2017 pp. 2169-3536

[8] Akash C. Jamgade, Prof. S. D. Zade, Disease Prediction Using Machine Learning, in:2019 International Research Journal of Engineering and Technology (IRJET), 2019,pp. 2395-0072.

[9] Shahadat Uddin , Arif Khan, Md Ekramul Hossain and Mohammad Ali Moni, Comparing different supervised machine learning algorithms for disease prediction, 2019 Uddin et al. BMC Medical Informatics and Decision Making doi: https://doi.org/10.1186/s12911-019-1004-8

[10] A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain, A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms, in: 2020 International Journal Of Scientific & Technology Research , 2020 , pp. 2277-8616.