



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DISEASE PREDICTION USING ML

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

^{12,3,4,5} Student, Dept. of Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, India.

^{1,6} Professor, 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 devised for early complaint discovery, patient care, and community services. One analogous performance of machine knowledge 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-affiliated problem is important for the prevention and treatment of the illness. The traditional way of opinion may not be sufficient in the case of a serious disease.

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

Developing a medical opinion system predicated on machine knowledge (ML) algorithms for prophecy of any complaint can help in a more accurate opinion than the conventional system. We have designed a complaint prophecy system using multiple ML algorithms. The data set used had further than 230 conditions for processing. Predicated 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 multitudinous variables in huge databases, the use of various data mining tools and machine knowledge approaches has changed healthcare associations. It serves as an important instrument in the medical sector, furnishing and comparing being data for the future course of action. This technology combines multiple logical methodologies with modern and complex algorithms, allowing for the exploration of massive amounts of data. Our opinion model can act as a croaker for the early opinion of a complaint to ensure the treatment can take place on time and lives can be saved. place on time and lives can be saved.

KEYWORDS- Logistic Regression, Naive Bayes, SVM, Decision Tree, Random Forest.

a. INTRODUCTION

Medicine and healthcare are some of the most vital corridors of the economy and mortal life. There is a tremendous amount of change in the world we are living in now and the world that was a numerous weeks back. Everything has turned horrible and divergent. There are also some remote villages which warrant medical installations. Virtual croakers are board-certified croakers who choose to exercise online via video and phone movables, rather than in-person movables but this is not possible in the case of emergency. Machines are always accounted better than humans as, without any deadly error, they can accomplish tasks more efficiently and with a harmonious position of delicacy. A complaint predictor can be called a virtual complainer, which can predict the complaint of any case without any mortal error. Machine knowledge offers truly accurate styles for deciding conditions which have bulky and well-collected databases. Machine knowledge in the field of medicine is a truly active area of disquisition. Despite that, there is no good mobile app available in the request for complaint prophecy which can help croakers in their day-to-day life. Hence we have tried to develop a portable app which can help them. Still, mobile operations present some challenges sort of a user can't input multitudinous numbers of input fields. For, eg UCI heart complaint dataset has approximately 75 features. So the main challenge before us was to reduce the number of features to an working position. We have offered our way during this paper of point birth by applying multi algorithm.

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

b. EXESTING SYSTEM

Forecasting using traditional complaint trouble model generally involves a machine knowledge and supervised literacy algorithm which uses training data with the labels for the training of the models. High trouble and Low trouble case type is done in groups test sets. But these models are only important in clinical situations and are considerably studied. A system for sustainable health monitoring using smart vesture byChenet.al. He Fully studied eclectic systems and was suitable to achieve the swish results for cost minimization on the tree and simple path cases for eclectic systems.

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

c. RELATED WORK

A structural model and a collection of conditional chances are used by Bayesian classifiers. They make the supposition that the contributions of all factors are independent. It first calculates the former probability for each class, and also applies the circumstance of each adjustable value to an unknown script. A Bayes network classifier is raised on a Bayesian network, which reflects a frequent probability distribution over a set of order characteristics. The SVM system and the Nave Bayes fashion were used to predict order complaint. The authors tried to classify various stages of order complaint using the suggested ANFIS algorithm. The study's purpose was to design an effective categorization algorithm using several assessment criteria analogous as delicacy and execution time. While the SVM Algorithm handed advanced type delicacy, the Nave Bayes did better since it delivered results in lower time. The outcomes show that SVM outperforms the Nave Bayes Approach in predicting renal illness. The fuzzy rage with a class function was used to read cardiac complaint. applying the Fuzzy KNN Classifier, the authors tried to count ambiguity and query from data. The 550- record dataset was decoupled into 25 classes, with each class having 22 particulars. The dataset was decoupled into two equal corridor training and testing. The fuzzy KNN methodology was executed afterpre-processing ways were used. This fashion was examined using several assessment criteria analogous as delicacy, perfection, and recall, among others. Predicated on the data, it was found that the fuzzy KNN classifier outperformed the KNN classifier in terms of rigor. For the prophecy of cardiac complaint, a new fashion predicated on the ANN algorithm was cooked. The researchers created an interactive prophecy system predicated on categorization using an artificial neural network algorithm and taking into account the thirteen most meaningful clinical parameters. The suggested system proved effective for predicting heart complaint with an delicacy of 80 and can be truly useful for healthcare practitioners.

Authors in [5] offered an automated approach for responding delicate inquiries for heart complaint prophecy. 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 findings about heart attacks. This system may be perfected by comprehending SMS functionality, setting up Android and IOS mobile operations, and including a trendsetter in the order. Diabetes and bone cancer were diagnosed by incorporating the adaptivity specific into support vector machines. The thing was to offer a rapid-fire- fire, automated, and adaptable individual system using adaptive SVM. To attain better outcomes, the bias value in conventional SVM was altered. The suggested classifier produced affair in the form'if- also' rules. The proposed system was used to diagnose diabetes and bone cancer, and it handed 100 right type rates for both conditions. Future disquisition should concentrate on developing further effective ways for changing the bias value in conventional SVM.

d. PROPOSED SYSTEM

In this paper, we have combined the structure and unstructured data in healthcare fields that let us assess the trouble of complaint. By applying statistical knowledge, we could judge the major habitual conditions in a particular region. In the case of unstructured text data, we handpick the features automatically with the help of multi algorithm which is Decision Tree, Naïve Bayes, Logistic Regression, Random Forest, SVM. We offer a multi algorithm for both structured and unstructured data. The decision taken from the multi algorithm swish 3 results are shown in the affair with its delicacy and is stored in the database if for formerly references. The delicacy we cossed by using numerous algorithm is 93.24%

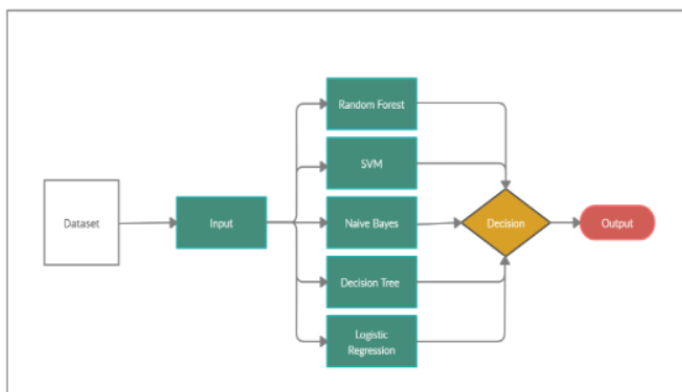


Figure 1

I. Random Forest

Random Forest is a classifier that contains a multiple of decision trees on various subsets of the given dataset and takes the average to meliorate the predictive delicacy of that dataset. Rather of counting on one decision tree, the arbitrary timber takes the prophecy from each tree and predicated on the maturity votes of vaticinations, and it predicts the concluding affair. direct relationship between inputs and affair variables.

b. Logistic Regression

Logistic regression is another major supervised ML algorithm used for binary classification problems. The swish way to suppose about logistic regression is that it's a direct regression but for classification problems. Logistic regression principally uses a logistic function defined below to model a double affair variable. The primary difference between direct regression and logistic regression is that logistic regression's range is bounded between 0 and 1. In addition, as opposed to direct retrogression, logistic retrogression does not bear a direct relationship between inputs and affair variables.

c. Decision Tree

Decision tree is the most influential and popular tool for classification and forecasting. A Decision tree is a flowchart like tree structure, where each inner knot denotes a test on an trait, each branch represents an outgrowth of the test, and each terminal knot holds a class marker.

d. Naive Bayes

Naive Bayes classifiers are a collection of classification algorithms grounded on Bayes Theorem. It is not a single algorithm but a lineage of algorithms where all of them participate a common principle, i.e. every brace of features being classified is independent of each other.

e. Support Vector Machine

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

IV. SYSTEM ARCHITECTURE

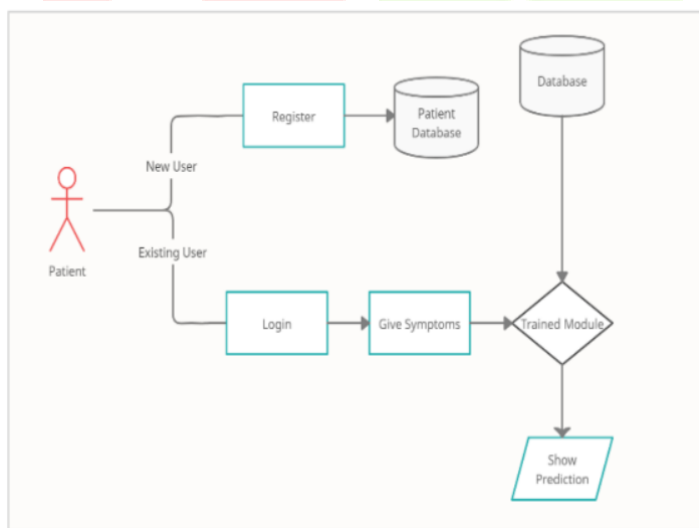


Figure 2

V. CONCLUSION

With the suggested system, improved delicacy can be attained. We not only apply structured data, but also the text data of the case predicated on the suggested multi algorithm which is Decision Tree, Random Forest, Naive Bayes, SVM, Logistic Regression. To determine that out, we join both data, and the delicacy rate can be reached up to 92.67% None of the being system and work is concentrated on applying both the data types in the field of medical big data analytics. We suggest multi clustering algorithm which is Naïve Bayes, Decision Tree, Random Forest, Logistic Regression and SVM Algorithms for both structured and unstructured data. The complaint trouble model is attained by combining both structured and unstructured features.

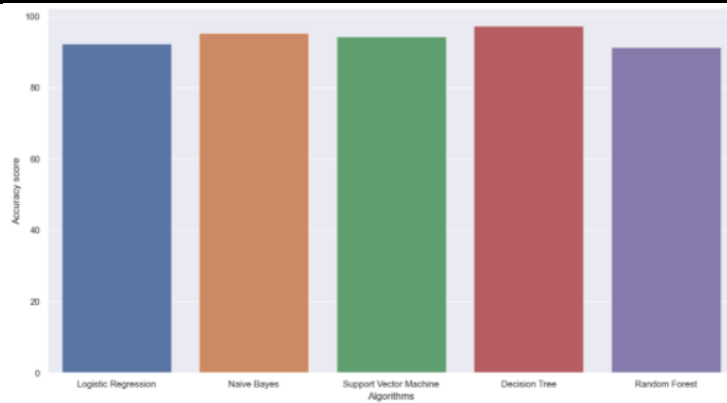


Figure 3

REFERENCES

- [1] 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.
- [2] 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.
- [3] 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
- [4] 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.
- [5] 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>.
- [6] 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>.
- [7] 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>
- [8] MIN CHEN , (Senior Member, IEEE), YIXUE HAO, KAI HWANG , (Life Fellow, IEEE), LU WANG , AND LINWANG, Disease Prediction by Machine Learning Over Big Data From Healthcare Communities, n:2017 IEEE Intellectual Property Rights, 2017 pp. 2169-3536
- [9] 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.
- [10] Akash C. Jangade, Prof. S. D. Zade, Disease Prediction Using Machine Learning, in:2019 International Research Journal of Engineering and Technology (IRJET), 2019,pp. 2395-0072.