iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42639

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

A Review of Common Ailments Possibility Using Machine Learning

Mohammed Sameer[1], Omjee[2], Raghav Gupta[3], Ritik Tyagi[4], Swapna Singh[5]
[1, 2, 3, 4, 5]*Inderprastha Engineering College*

Abstract: ***Machine learning in healthcare helps humans to process large and complex medical datasets and then analyze them into clinical insights which can help physicians in providing better medical care. Therefore, machine learning, when implemented in the medical field can lead to increased patient satisfaction. In this research, we will try to implement the functionalities of machine learning in healthcare in a single system. Health care can be made smart with the help of machine learning. Many cases can occur when the early diagnosis of an ailment is not within reach, So, their ailment prediction cannot be effectively implemented. As widely said "Prevention is better than cure", prediction of diseases would lead to early prevention of occurrence of disease. Medical Staff are often overworked in the medical field and hence the diagnosis becomes prone to human errors and negligence. Patients should be given treatment and diagnosis that are accurate and precise. Mistreatment may result in worsening the condition of the patient and hence the need for precise diagnosis. Therefore, the application of machine learning in disease prediction is considered in this paper as the best practice to facilitate a better healthcare system and provide better treatment to a patient as soon as possible. This paper majorly focuses on the development of a web app that would work on symptoms collected from the user and medical data and store it in the system. This data then will be analyzed using different machine learning algorithms to deliver results with maximum accuracy.***

I. INTRODUCTION

Medicine and healthcare are some of the most crucial parts of the economy and human life. There is a tremendous amount of change in the world we are living in now. Nowadays, hospitals are well equipped with monitoring and other data collection devices resulting in enormous data which are collected continuously through health examination and medical treatment.

In this situation, where everything has changed so much, the doctors and nurses are putting up maximum efforts to save people's lives even if they have to risk their own. In remote areas where there is a lack of medical facilities, some virtual assistance in case of emergency can play a major role. Machines are always considered better than humans as, without any human error, they can perform tasks more efficiently and with a consistent level of accuracy. A disease predictor can be called a virtual doctor, which can predict the disease of any patient without any human error. Also, in situations like COVID-19, a disease predictor can be a blessing as it can identify a person's disease without any physical contact. Disease Prediction using Machine learning is a system that predicts the disease based on the symptoms provided by the user. It is a system that also provides the user with a specialized doctor for the disease predicted. It is a system that provides the user the tips and tricks to maintain the health system of the user.

The purpose of making this project is to predict the accurate disease using the symptoms provided by the user. Using this information, we will compare with our previous datasets of the patients and predict the disease of the patient he/she has been through. If this prediction is done at the early stages of the disease can be cured and in general, this prediction system can also be very useful in the health industry. If the health industry adopts this project, then the work of the medical staff can be reduced and they can easily diagnose the disease of the patient. The general purpose of this research is to provide a prediction for the various and generally occurring diseases that when unchecked and sometimes ignored can turn into dangerous diseases and cause a lot of problems to the patient and as well as their family members. This system will predict the most possible ailment based on the symptoms provided. So, with the help of all these algorithms, techniques, and methodologies we have done this project with hopes to help the people who are in need.

II. ANALYSIS OF VARIOUS MODELS

Iwendi C [et.al] proposed a fine-tuned Random Forest model boosted by the AdaBoost algorithm in 2020. The model uses the COVID-19 patient's data to predict the severity of the case and the possible outcome, recovery, or death. The model has an accuracy of 94% and an F1 score of 0.86 on the dataset used. The data analysis reveals a positive correlation between patients' gender and deaths and indicates that most patients are aged between 20 and 70 years.[1]

©IJRASET: All Rights are Reserved

Rinkal Keniya [et.al] in their research done in 2020 used different machine learning models to examine the prediction of disease for available input datasets. Total of 11 different ML models were used . Out of the 11 models, they managed to get 50 % or above accuracy for 6 models. As shown in the table, The Highest accuracy of a few of the models are SVM, Random Forest, and naïve Bayes.[2]

Anant Agrawal [et.al] in 2018 proposed a hybrid machine learning model consisting of a genetic algorithm and support vector machine. They tested their model on different datasets like liver, diabetes, and heart. By reducing the number of features, they were able to get good enough accuracies for all three datasets.[3]

Shahadat Uddin [et.al] in 2019 attempted to study comparative performances of different supervised machine learning algorithms in disease prediction. To compare the algorithms, a common benchmark is to be established on dataset and scope since clinical data and scope varies between disease prediction studies. Regardless of the variations on frequency and performances, the results show the potential of these families of algorithms in the disease prediction.[4]

Jyoti Soni [et.al] have found that Decision Tree is better when comparing the performance of predictive data mining technique on the same dataset and sometimes Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering is not performing well.[5]

K.M. Al-Aidaroos [et.al] conducted research for the medical diagnosis using mining techniques in 2018. For this, authors compared Naïve Bayes with five other classifiers .For this, 15 real-world medical problems from the UCI machine learning repository were selected for evaluating the performance of all algorithms. It was found that Naive Bayes gave better performance in 8 out of 15 data sets so it was concluded that the predictive accuracy results in Naïve Bayes is better than other techniques.[6]

M. Marimuthu [et.al] In conclusion, as identified through the literature survey, believe only a marginal success is achieved in the creation of predictive model for heart disease patients and hence there is a need for combinational and more complex models to increase the accuracy of the predicting the early onset of heart disease. With the more amount of data being fed into the database the system will be very intelligent.[7]

Shadab Adam [et.al] The system uses a historical heart disease database for information**.** This model could answer complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy. HDPS can be further enhanced and expanded. It can also include other data mining techniques. Instead of categorical data, continuous data can be used . HDPS can be further enhanced and expanded.[8]

S. Jadhav [et.al] In the proposed device, it affords system getting to know algorithms for effective prediction of numerous disorder occurrences in disease-common societies and predicts the waiting time for each remedy task for each affected person in addition to a Hospital Queuing Recommendations (HQR) gadget is evolved for recommending remedy mission series with appreciate to anticipated waiting time. .[9]

Purushottam [et.al] In this research paper, they have presented a system that can help medical practitioners in efficient decision making based on the given parameter. They have trained and tested the system using a 10 fold method. The result shows the accuracy of 86.3 % in the testing phase and 87.3 % in the training phase. This model demonstrates better results and helps the area specialists and even individual related with the field to get ready for a superior determine and give the patient to have early determination results as it performs sensibly well even without retraining.[10]

N. Skyttberg [et.al] studied that standardization of the workflow is an important concept when discussing vital sign data quality in Swedish emergency departments. An effective method to reduce individual variation and increase quality is using measurement and documentation in a well defined workflow. However, to make sure that the documentation is digitalized, information technology has to provide adequate documentation support, otherwise paper-based documentation will be favored.[11]

A. Wright [et.al] CDSS malfunctions are common and often go undetected.The alerts can sometimes fail to fire and that is difficult to detect. A range of causes, including changes in codes and fields, software upgrades, inadvertent disabling or editing of rules, and malfunctions of external systems commonly contribute to these malfunctions, and current approaches for preventing and detecting CDSS malfunctions are inadequate. As CDSSs becomes more complex and widespread and clinicians increase their reliance on them, improved processes and tools for preventing and detecting CDSS malfunctions are essential.[12]

C. Y. Tsai [et.al] Physicians are not likely to adopt recommendations provided by false positive alerts in patient-safety-related CDSS. The system will lead to an underestimation of system effectiveness if adoption rate of CDSS is reported without differentiating between TP and FP alerts.[13]

M. El-Bardini [et.al] proposed direct AIT2-FLC to deal with the multivariable anesthesia system. After testing the proposed controller by using three simulation tasks including the inter and intra-individual variability of the patient's parameters, the results for the proposed controller were compared with T1-FLC and IT2-FLC that were implemented for controlling the multivariable

anesthesia system and published previously. Results show that the proposed controller is able to respond to the uncertainty that is introduced by large inter and intra-individual variability of the patient's parameters. The proposed controller is better than IT2-FLC because there is no deviation from set-point for both muscle relaxation and blood pressure. So, the proposed controller is superior to IT2-FLC and T1-FLC that were published previously for controlling the multivariable anesthesia system.[14]

Tarigoppula V.S Sriram [et.al] concluded that analysis of voice data is important in the present decade to understand and diagnostic methods for human diseases. The present method provides the diagnosis of PD using voice dataset through machine learning algorithms. [15]

Jaymin Patel [et.al] By analyzing the experimental results, it is concluded that J48 tree technique turned out to be the best classifier for heart disease prediction because it contains more accuracy and less total time to build. We can clearly see that highest accuracy belongs to J48 algorithm with reduced error pruning followed by LMT and Random Forest algorithm respectively. Also Observed that applying reduced error pruning to J48 results in higher performance while without pruning, it results in lower Performance. The best algorithm J48 based on UCI data has the highest accuracy i.e. 56.76% and the total time to build model is 0.04 seconds while LMT algorithm has the lowest accuracy i.e 55.77% and the total time to build model is 0.39seconds**.** [16]

Sellappan Palaniappan et.al] analyzed the classification tree techniques in data mining. Decision Stump, Random Forest and LMT Tree algorithm are used and tested. The objective of this research was to compare the outcomes of the performance of different classification techniques for a heart disease dataset. [17]

Vinitha S [et.al] bid a Machine Learning Decision tree map algorithm by using structured and unstructured data from hospitals.For partitioning the data, Map reduce algorithm is used To the highest of gen, none of the current work attentive on together data types in the zone of remedial big data analytics. Compared to several typical calculating algorithms, the scheming accuracy of our proposed algorithm reaches 94.8% with a regular speed which is quicker than that of the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm and produces reports. The report consists of possibility of occurrences of diseases.[18]

Kedar Pingale [et.al] aims to predict the disease on the basis of the symptoms. The system takes symptoms from the user as input, processes the data and predicts the disease as output. In conclusion, for disease risk modeling, the accuracy of risk prediction depends on the diversity feature of the hospital data.[19]

Nishant Yede [et.al] proposed a framework using big data and K - means clustering method to select drugs for patients. Big Data Analytics combined with machine learning has brought a new era for biomedical engineering research. From biomedical classification, machine learning plays a key role in today's world. The proposed framework shows the significance of incorporating machine learning and big data in healthcare research. SVM Algorithm used for classification in big healthcare data. Currently, the proposed framework shows performance based on K-means clustering. In future, this framework will be used to estimate the accuracy of the data and this will help the healthcare professionals to make clinical decisions.[20]

III. CONCLUSION

Disease prediction using patient symptoms by applying data mining and machine learning techniques is ongoing struggle for the past decades.The recent success of deep learning in disparate areas of machine learning has driven a shift towards machine learning models that can learn rich, hierarchical representations of raw data with little preprocessing and produce more accurate results. Numbers of papers have been published on several data mining techniques for diagnosis of heart disease such as Decision Tree, Naive Bayes, neural network, svm, kernel density, automatically defined groups, bagging algorithm and support vector machine showing different levels of accuracies in disease prediction. In this paper, we have reviewed various ML techniques that have been presented in different published and available literature.On this basis, it can safely conclude that the Support Vector Machine and Random Forest Technique has higher accuracy and shows promising results.

REFERENCES

[1] Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, Mishra R, Pillai S and Jo O, "COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm". (2020) Front. Public Health 8:357. DOI: 10.3389/fpubh.2020.00357

[2] Keniya Rinkal and Aman Khakharia and Vruddhi Shah and Vrushabh Gada and Ruchi Manjalkar and Tirth Thaker and Mahesh Warang and Ninad Mehendale "Disease Prediction from Various Symptoms Using Machine Learning". (2020) SSRN: 3661426.

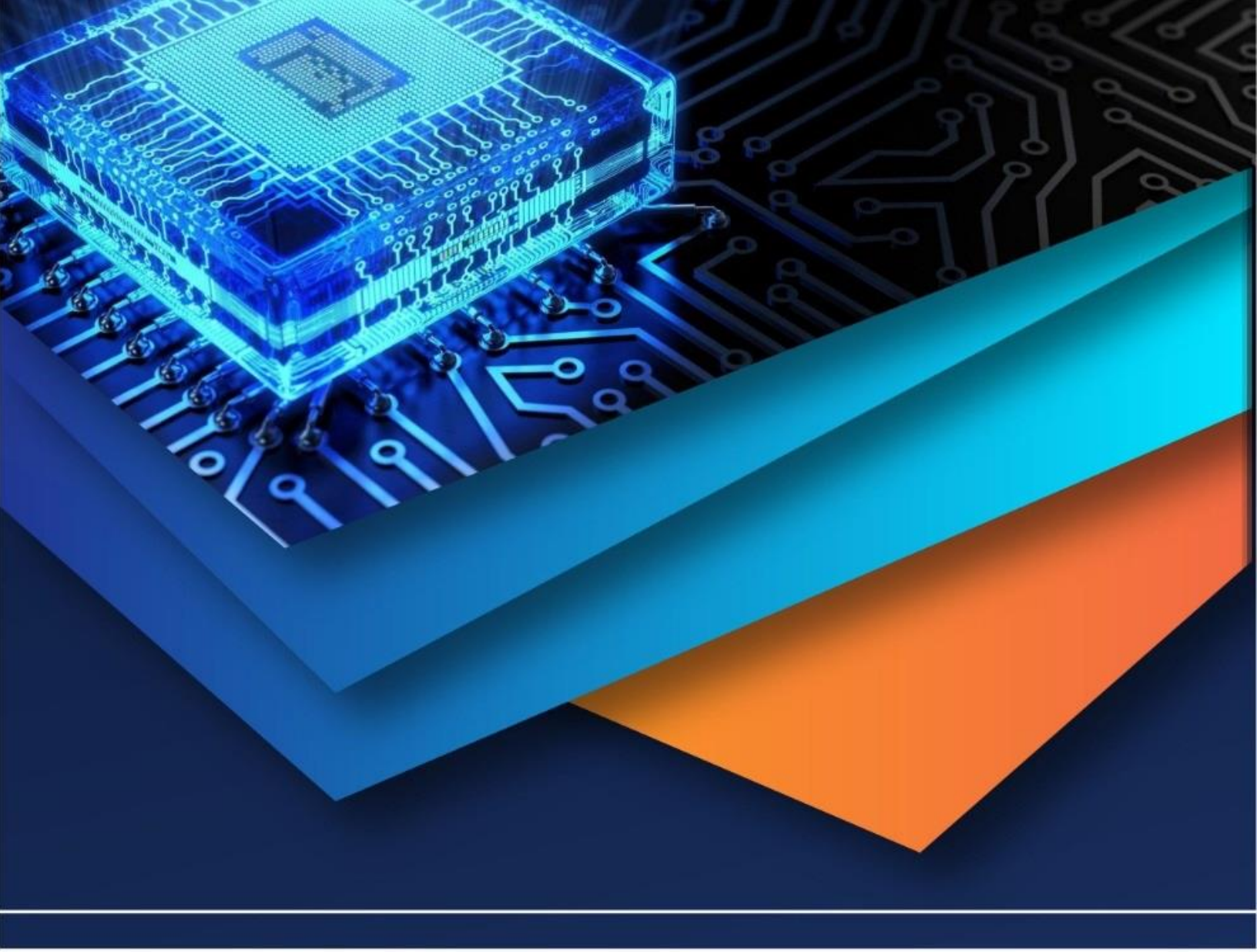
[3] Anant Agrawal, Harshit Agrawal, Shivam Mittal, Mradula Sharma, "Disease Prediction Using Machine Learning". 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT), ISSN: 1556-5068. 2018.

[4] Shahadat Uddin, Arif Khan, Md Ekramul Hossain, and Mohammad Ali Moni, "Comparing different supervised machine learning algorithms for disease prediction", BMC Medical Informatics and Decision Making (2019) 19:281.

[5] Jyoti Soni, Ujma Ansari, Dipesh Sharma, and Sunita Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction". (2011) International Journal of Computer Applications. DOI: 10.5120/2237-2860.

[6] K.M. Al-Aidaroos, A.A. Bakar, and Z. Othman, "Medical Data Classification With Naive Bayes Approach". (2012) International Technology Journal 11. DOI: 10.3923/itj.2012.1166.1174.
[7] M. Marimuthu, M. Abinaya, K. S. Hariesh, K. Madhankumar, V. Pavithra,"A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach", International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 18, September 2018.
[8] Shadab Adam et al., "Prediction system for Heart Disease using Naïve Bayes", International Journal of Advanced Computer and Mathematical Sciences, vol. 3, no. 3, pp. 290-294, ISSN 2230-9624 (2012).
[9] S. Jadhav, R. Kasar, N. Lade, M. Patil, and S. Kolte, "Disease Prediction by Machine Learning from Healthcare Communities," InternationalJournal of Scientific Research in Science and Technology, pp. 29–35,2019.
[10] Purushottam, Kanak Saxena, Richa Sharma, "Efficient Heart Disease Prediction System", Procedia Computer Science Volume 85, 2016, Pages 962-969.
[11] N. Skyttberg, J. Vicente, R. Chen, H. Blomqvist, and S. Koch, "How to improve vital sign data quality for use in clinical decision support systems? A qualitative study in nine Swedish emergency departments," BMC medical informatics and decision making, vol. 16, p. 1, 2016
[12] A. Wright, T.-T. T. Hickman, D. McEvoy, S. Aaron, A. Ai, J. M. Andersen, et al., "Analysis of clinical decision support system malfunctions: a case series and survey," Journal of the American Medical Informatics Association, p. ocw005, 2016.
[13] C. Y. Tsai, S.H. Wang, M.H. Hsu, and Y.C. J. Li, "Do false positive alerts in naïve clinical decision support systems lead to false adoption by physicians? A randomized controlled trial," Computer Methods and Programs in Biomedicine, vol. 132, pp. 83-91, 2016.
[14] M. El-Bardini and A. M. El-Nagar, "Direct adaptive interval type-2 fuzzy logic controller for the multivariable anesthesia system," Ain Shams Engineering Journal, 2011.
[15] Tarigoppula V.S Sriram , M. Venkateswara Rao , G V Satya Narayana , DSVGK Kaladhar , T Pandu Ranga Vital "Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms", International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 3, September 2013.
[16] Jaymin Patel, Prof. Tejal Upadhyay and Dr. Samir Patel, "Heart Disease Prediction Using Machine learning and Data Mining Technique" IJCSC, Volume 7, Number 1, September 2015
[17] Sellappan Palaniappan and Rafiah Awang, "Intelligent heart disease prediction system using data mining techniques" 2008 IEEE/ACS International Conference on Computer Systems and Applications, 31 March-4 April 2008.
[18] Vinitha S, Sweetlin S, Vinusha H and Sajini S, "Disease Prediction using Machine Learning over Big Data"Computer Science & Engineering: An International Journal (CSEIJ), Vol.8, No.1, February 2018
[19] Kedar Pingale, Sushant Surwase, Vaibhav Kulkarni, Saurabh Sarage and Prof. Abhijeet Karve, "Disease Prediction using Machine Learning"International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume 06, Issue 12, Dec 2019
[20] Nishant Yede , Ritik Koul , Chetan Harde , Kumar Gaurav and Prof. C. S. Pagar, "Disease Prediction by Machine Learning over Big data from Healthcare Communities" International Journal of Advance Scientific Research and Engineering Trends, Volume 5, Issue 11, November 2020

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)