ISSN 2277-2685
IJESR/April-May 2025/ Vol-15/Issue-2s/1083-1086
G. Shanker *et. al.*, / International Journal of Engineering & Science Research

Diseases Prediction Using Machine Learning For Health Care

[1]Dr.N.Dinesh Kumar, [2]G. Shanker, [3]K. Nanditha Reddy, [4]M.Niharika, [5]P. Durga Prasad
[1]Dean R&D, Electronics And Communication Engineering, Vignan Institute of Technology and Science, India.
[2,3,4,5]Electronics And Communication Engineering, Vignan Institute of Technology and Science, India.
shivashanker99456@gmail.com, nandithareddiee18@gmail.com, malganiharikam@gmail.com,
dpdurgaprasad771@gmail.com

Abstract*—Disease Prediction using Machine Learning is the system that is used to predict the diseases from the symptoms which are given by the patients or any user. The system processes the symptoms provided by the user as input and gives the output as the probability of the disease. Naïve Bayes classifier is used in the prediction of the disease which is a supervised machine learning algorithm. The probability of the disease is calculated by the Naïve Bayes algorithm. With an increase in biomedical and healthcare data, accurate analysis of medical data benefits early disease detection and patient care. By using linear regression and decision tree we are predicting diseases like Diabetes, Malaria, Jaundice, Dengue, and Tuberculosis.

Keywords: Disease Prediction, Machine learning, Naive bayes algorithm

1. INTRODUCTION

Machine Learning is the domain that uses past data for predicting. Machine Learning is the understanding of computer system under which the Machine Learning model learn from data and experience. The machinelearning algorithm has two phases: 1) Training & 2) Testing. To predict the disease from a patient's symptoms and from the history of the patient, machine learning technology is struggling from past decades.

Healthcare issues can be solved efficiently by using Machine Learning Technology. We are applying complete machine learning concepts to keep the track of patient's health. ML model allows us to build models to get quickly cleaned and processed data and deliver results faster. By using this system doctors will make good decisions related to patient diagnoses and according to that, good treatment will be given to the patient, which increases improvement in patient healthcare services. To introduce machine learning in the medical field, healthcare is the prime example. To improve the accuracy of large data, the existing work will be done on unstructured or textual data. For the prediction of diseases, the existing will be done on linear, KNN, Decision Tree algorithm.

2. EXISTING SYSTEM

The existing system predicts the chronic diseases which are for a particular region and for the particular community. Only particular diseases are predicted by this system. In this System, Big Data & CNN Algorithm is used for Disease risk prediction. For S type data, the system is using Machine Learning algorithm i.e K-nearest Neighbors, Decision Tree, Naïve Bayesian. The accuracy of the existing System is up to 94.8%. In the existing paper, they streamline machine learning algorithms for the effective prediction of chronic disease outbreak in disease-frequent communities. They experiment with the modified prediction models over reallife hospital data collected from central China. They propose a convolutional neural network-based multimodal disease risk prediction(CNN-MDRP) algorithm using structured and unstructured data from the hospital.

3. PROPOSED SYSTEM

Most of the chronic diseases are predicted by our system. It accepts the structured type of data as input to the machine learning model. This system is used by end-users i.e. patients/any user. In this system, the user will enter all the symptoms from which he or she is suffering. These symptoms then will be given to the machine learning model to predict the disease. Algorithms are then applied to which gives the best accuracy. Then System will predict disease on the basis of symptoms.

This system uses Machine Learning Technology. Naïve Bayes algorithm is used for predicting the disease by using symptoms, for classification KNN algorithm is used, Logistic regression is used for extracting features which are having most impact value, the Decision tree is used to divide the big dataset into smaller parts. The final output of this system will be the disease predicted by the model

4. METHODOLOGY

To calculate performance evaluation in the experiment, first, we denote TP, TN, Fp and FNias true positive(the number of results correctly predicted as required), true negative (the number of results not required), false positive (the number of results incorrectly predicted as required), false negative(the number of results incorrectly predicted as not required)respectively. We can obtain four measurements: recall, precision, accuracy, and F1 measures as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

5. REVIEW OF LITERATURE

Machine learning (ML) has become a transformative tool in healthcare, particularly in the prediction of diseases. By analyzing large datasets, including patient demographics, medical histories, lab results, and even genetic information, ML algorithms can predict the onset of various diseases such as cardiovascular diseases, diabetes, cancer, and infectious diseases. Supervised learning techniques like decision trees, support vector machines (SVM), and logistic regression have been extensively used in this domain, where models are trained on labeled data to predict disease outcomes. Additionally, deep learning, especially convolutional neural networks (CNNs), has shown remarkable performance in analyzing medical images for cancer detection, including breast and lung cancer. Furthermore, ML has played a crucial role in infectious disease prediction, with notable applications during the COVID-19 pandemic, where ML models helped predict infection rates, patient outcomes, and potential outbreaks based on demographic and symptom data. Despite the success of ML in disease prediction, challenges remain, such as the quality and availability of medical data, the interpretability of complex models, and the generalization of predictions across diverse populations. Moreover, ethical concerns around patient privacy and algorithmic bias also need to be addressed to ensure fairness and accuracy in healthcare decision-making. Looking ahead, advancements in techniques like federated learning—allowing for decentralized data analysis while maintaining patient privacy—and the development of explainable AI (XAI) hold promise for improving model transparency and application in clinical practice. The continued integration of ML into healthcare will likely revolutionize disease prediction, diagnosis, and personalized treatment, ultimately improving patient outcomes.

6. BLOCK DIAGRAM

KNN K Nearest Neighbour (KNN) could be terribly easy, simple to grasp, versatile and one amongst the uppermost machine learning algorithms. In the Healthcare System, the user will predict the disease. In this system, the user can predict whether the disease will detect or not. In the proposed system, classifying disease in various classes that shows which disease will happen on the basis of symptoms. KNN rule used for each classification and regression issue. KNN algorithm is based on feature similarity approach. It is the best choice for addressing some of the classification related tasks. K-nearest neighbor classifier algorithm is to predict the target label of a new instance by defining the nearest neighbor class. The closest class will be identified using distance measures like Euclidean distance. If K = 1, then the case is just assigned to the category of its nearest neighbor.

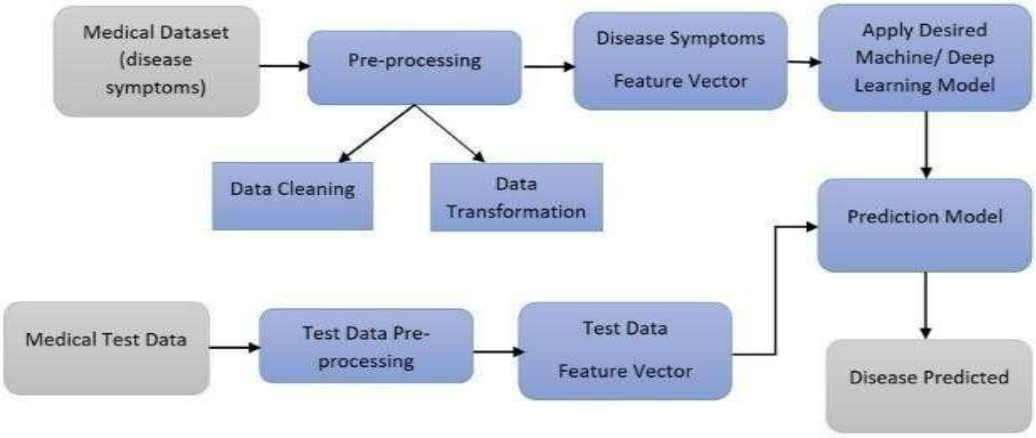
The value of 'k' has to be specified by the user and the best choice depends on the data. The larger value of 'k' reduces the noise on the classification. If the new feature

i.e in our case symptom has to classify, then the distance is calculated and then the class of feature is selected which is nearest to the newer instance. In the instance of categorical variables, the Hamming distance must be used. It conjointly brings up the difficulty of standardization of the numerical variables between zero and one once there's a combination of numerical and categorical variables within the dataset

NAIVE BAYES:

Naive Bayes is an easy however amazingly powerful rule for prognosticative modeling. The independence assumption that allows decomposing joint likelihood into a product of marginal likelihoods is called as 'naive'. This simplified Bayesian classifier is called as naive Bayes. The Naive Bayes classifier assumes the presence of a particular feature in a class is unrelated to the presence of any other feature. It is very easy to build and useful for large datasets. Naive Bayes is a supervised learning model. Bayes theoremmprovides some way of calculative posterior chance P(b|a) from P(b), P(a) and P(a|b). Look atithe equation below:

P(b v a)= P(a v b)P(b)/P(a)
Above,
• P(b|a) is that the posterior chance of class (b,target) given predictor (a, attributes). • P(b) is the prioriprobability of class.
• P(a|c)iis that chance that is that the chance of predictor given class.



• P(a) is the prioriprobability ofipredictor. In our system, Naïve Bayes decides which symptom is to put in classifier and which is not. 8.3 LOGISTIC REGRESSION Logistic regression could be a supervised learning classification algorithm accustomed to predict the chance of a target variable that is Disease.

7. *Advantages*

Early Detection and Prevention

Machine learning enables the identification of disease patterns in their early stages, often before the patient even shows symptoms. This early detection is key in preventing diseases from progressing to more severe stages, which often require more complex and expensive treatments.

Personalized Treatment Plans

By predicting an individual's susceptibility to certain diseases, healthcare providers can tailor prevention and treatment plans specific to that patient's risk factors,

lifestyle, and genetic makeup, resulting in more effective and personalized care.

Improved Accuracy

Machine learning models can analyze large datasets much more effectively than human doctors alone, leading to more accurate diagnoses and predictions. They can detect subtle patterns and relationships in data that might be missed by traditional clinical methods. **Cost Savings**

Early disease prediction helps reduce the need for expensive treatments in later stages of disease progression. Preventive care can lower healthcare costs by avoiding hospitalizations, surgeries, and long-term care for chronic conditions.

Real-Time Monitoring

Machine learning models can analyze real-time data (e.g., from wearable devices) to provide continuous monitoring of patients' health. This real-time analysis helps in providing immediate feedback to patients and doctors, which is crucial for managing chronic diseases or detecting sudden health changes.

8. CONCLUSION

The problems faced by the medical industry with the unaffordability of the patients to seek dictators and the unavailability of the medical staff can be diminished. machine learning has shown immense potential in revolutionizing the prediction and early detection of various diseases, significantly improving healthcare outcomes. By harnessing the power of large datasets and advanced algorithms, ML models can provide accurate and timely predictions that assist in early diagnosis, personalized treatment, and better management of diseases like cardiovascular issues, diabetes, cancer, and infectious diseases. Despite challenges related to data quality, model interpretability, and ethical concerns, ongoing advancements in ML techniques—such as federated learning and explainable AI—offer promising solutions to these issues. As the field continues to evolve, the integration of machine learning into healthcare systems is poised to enhance decision-making, optimize treatment strategies, and ultimately contribute to more efficient, precise, and equitable healthcare for patients worldwide.

This can happen by automating the channelization of the patients to a specialist instead of a generalist. This can happen via the use of a disease prediction system. This system will input the patient's symptoms and produce possible disease as an output with 97% accuracy as compare to earlier models. The proposed

model can assist the healthcare industry by:

9. FUTURE SCOPE

In the future, the model can be used in various sectors and can enhance efficiency by considering more symptoms to predict disease. The model can be used for providing an enhanced, more accurate framework that would lead to a better human disease prediction model.

REFERENCES

1. Zhou, S.-M., Fernandez-Gutierrez, F., Kennedy, J., Cooksey, R., Atkinson, M., Denaxas, S., Siebert, S., Dixon, W.G., O'Neill, T.W. and Choy, E., "Defining disease phenotypes in primary care electronic health records by a machine learning approach: A case study in identifying rheumatoid arthritis", PloS One, Vol. 11, No. 5, (2016), e0154515. https://doi.org/10.1371/journal.pone.0154515
2. Littell, C.L., "Innovation in medical technology: Reading the indicators", Health Affairs, Vol. 13, No. 3, (1994), 226-235. https://doi.org/10.1377/hlthaff.13.3.226
Milella, F., Minelli, E.A., Strozzi, F. and Croce, D., "Change and innovation in healthcare: Findings from literature", ClinicoEconomics and Outcomes Research, (2021), 395-408. doi: 10.2147/CEOR.S301169.
3. Rathi, M. and Pareek, V., "Disease prediction tool: An integrated hybrid data mining approach for healthcare", IRACSTInternational Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN, (2016), 2249-9555.
4. Kelly, C.J. and Young, A.J., "Promoting innovation in healthcare", Future Healthcare Journal, Vol. 4, No. 2, (2017), 121. doi: 10.7861/futurehosp.4-2-121.
5. Mobeen, A., Shafiq, M., Aziz, M.H. and Mohsin, M.J., "Impact of workflow interruptions on baseline activities of the doctors working in the emergency department", BMJ Open Quality, Vol. 11, No. 3, (2022), e001813. doi: 10.1136/bmjoq-2022- 001813.
6. Ahmed, S., Szabo, S. and Nilsen, K., "Catastrophic healthcare expenditure and impoverishment in tropical deltas: Evidence from the mekong delta region", International Journal for Equity in Health, Vol. 17, No. 1, (2018), 1-13. doi: 10.1186/s12939-018-0757-5.
7. Roberts, M.A. and Abery, B.H., "A person-centered approach to home and community-based services outcome measurement", Frontiers in rehabilitation Sciences, Vol. 4, (2023). doi: 10.3389/fresc.2023.1056530
8. Farooqui, M. and Ahmad, D., "Disease prediction system using support vector machine and multilinear regression", International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN, (2020), 2347-5552. https://doi.org/10.21276/ijircst.2020.8.4.15
9. Olatunji, O.O., Adedeji, P.A., Akinlabi, S., Madushele, N., Ishola, F. and Aworinde, A.K., "Improving classification performance of skewed biomass data", in IOP Conference Series: Materials Science and Engineering, IOP Publishing. Vol. 1107, (2021), 012191.
10. Cao, J., Wang, M., Li, Y. and Zhang, Q., "Improved support vector machine classification algorithm based on adaptive feature weight updating in the hadoop cluster environment", PloS One, Vol. 14, No. 4, (2019), e0215136. https://doi.org/10.1371/journal.pone.0215136
11. Hamidi, H. and Daraee, A., "Analysis of pre-processing and post-processing methods and using data mining to diagnose heart diseases", International Journal of Engineering, Transactions B: Applications, Vol. 29, No. 7, (2016), 921-930.