International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Disease Prediction System using Machine Learning

Aditya Singh[1], Ankur Pandey[2], Ankur Kumar Tiwari[3] , Chandan Nishad[4]

[1]Aditya Singh CSE department ITM Gida Gorakhpur
[2]Ankur Pandey CSE department ITM Gida Gorakhpur
[3]Ankur Kumar Tiwari CSE department ITM Gida Gorakhpur
[4]Chandan Nishad CSE department ITM Gida Gorakhpur

Abstract - This paper presents an approach of employing Machine Learning for disease prediction by utilizing input user symptoms. Supervised Machine Learning strategies are implemented through Decision Tree, Random Forest, Naïve Bayes, and KNN algorithms to achieve more accurate and timely disease identification; thus, providing optimal patient care. The results ascertained the feasibility and practicality of the system to promptly diagnose diseases.

***Key Words*:** Machine Learning, Disease Prediction, Decision Tree, Random Forest.

1.INTRODUCTION *(Size 11, Times New roman)*

The Earth is currently experiencing a period of rapid technological expansion, where the need for intelligence and accuracy is continually increasing [11]. People in the modern age have become heavily reliant on the internet, often neglecting their physical health. All too often, they ignore small problems until they become much more serious diseases, as a result of never having visited a hospital or physician [11]. Taking advantage of this growing technology, our primary objective is to develop an efficient system which can be used to predict multiple diseases just by inputting the relevant symptoms without ever having to visit a hospital.

Machine Learning is a subset of artificial intelligence (AI) that mainly deals with algorithms which improve and optimize with the use of data and experience. Generally speaking, Machine Learning consists of two phases: training and testing [17]. It provides an efficient platform when it comes to solving healthcare issues at an expedited rate. Machine Learning can be divided into two distinct categories: supervised learning and unsupervised learning. With supervised learning, a model is built using previously labeled data while with unsupervised learning model learns from unlabeled data.

The aim here is to come up with a suitable Machine Learning algorithm that yields accurate results when predicting disease. To achieve this goal, we will employ algorithms such as Decision Tree, Random Forest, Naïve Bayes and KNN as these have been proven to successfully predict diseases accurately while improving patient care [11].

We have tools that can give us or suggest relevant information within our reach, and the internet is one of those tools. Today billions of searches are made every day and sometimes the results provided are relevant and sometimes not.

2. RELATED WORK

Numerous works related to the Disease Prediction System utilizing different Machine Learning algorithms have been done and achieved various results in the medical field.
The paper [1] “Disease Prediction System” employed Decision Tree, Random Forest, and Naïve Bayes algorithms to forecast an ailment based on systems and optimize synchronized and knowledgeable medical systems ensuring maximum patient satisfaction.

The paper [2] “Heart Disease Prediction with Machine Learning Approaches” used LR, NB, KNN, SVM, DT and RF algorithms for heart disease prediction with competent data processing and implementation of Machine Learning algorithms with distinct parameters; KNN scored the highest accuracy of 87%.

The paper [3] “Heart Attack Prediction by Using Machine Learning Techniques” made a comparison of numerous Machine Learning models via performance metrics to detect heart-related problems with an accuracy of 89.34% achieved by SVM.

Moreover, the paper [4] “Disease Prediction using Machine Learning over Big Data” proposed a CNN-MDRP algorithm which combined structured and unstructured data and proved that CNN-MDRP is more precise than earlier prediction algorithms.

Additionally, the paper [5] “A Review of Heart Disease Prediction Using Machine Learning and Data Analytics Approach” utilized diverse Data Mining (DM) and Machine Learning methods to predict heart disease occurrences and applied the proposed system where needed.

Furthermore, the paper [6] “Application of Machine Learning Predictive Models in Chronic Diseases” concentrated on SVM and LR algorithms to assess study models related to chronic disease diagnosis. These models showed high applicability in classification as well as diagnosis of Chronic Diseases.

Additionally, the paper [7] "COVID-19 Outbreak Prediction with Machine Learning" utilized MLP as well as ANBEIS and presented a comparative analysis between ML models and soft ones for forecasting COVID-19 outbreak while providing initial benchmarking to illustrate ML potential for future use.

Moreover, the paper [8] "Heart Disease Prediction System using Machine Learning" constructed a heart disease prediction system using NB algorithm that attained an accuracy of 88.163% among others.

The paper [9] "Heart Disease Prediction Using Machine learning" suggested a robust model for predicting heart diseases which ranked Logistic Regression algorithm as having most efficiency with 82.89% accuracy followed by DT (80.40%) & NB (80.40%) & SVM (81.75%).

Additionally, the paper [10]"Implementation Of machine learning model to predict Heart Failure Disease" explored, recommended & applied a machine learning model in which Rapid Miner tool is used calculating high degree of correctness than MATLAB & Weka tool.

Moreover, the Paper [11]"Multiple Disease prediction using Different machine learning Algorithms Comparatively" proposed multiple disease prediction system providing Medicine & Drug consultation of disease predicted.

Besides, the Paper [12]"Heart Disease prediction Using Data Mining Techniques" used KNN, Naive Bayes and SVM Algorithms and collaborated with respect To Accuracy Using Heart Disease Dataset and Achieved Highest Accuracy Of 86.6 %Using Naive Bayes.

Additionally, The Paper [13]"Heart Disease Prediction Using machine learning Techniques" proposed method For Heart Disease Prediction using machine learning and results showed great accuracy standards for better estimated results.

Also, The Paper [14]" Diseased Prediction Using Machin Leaning " Used KNN, Naive Bayes, Logistic Regression And Decision Tree Algorithms To Make Disease Prediction System Which Can Predict Disease On Basis Of Symptoms Implemented Using Grails Framework.

Lastly, The Paper [15]"Diseased Prediction using Machine Learning " Used Naive Bayes, Decision Tree And Random Forest Algorithms To Create Disease Prediction System With Better Accuracy which also provides motivational thoughts and images.

3. PROPOSED MODEL

This proposed methodology entails the following steps:
i. Gather records regarding signs and associated functional issues in the body.
ii. Collect data correlating signs to potential medical conditions.
iii. Input acquired symptoms from patient via multilinear regression for processing.
iv. Utilize multilinear regression to forecast possible medical conditions from given symptoms.
v. System will reveal diagnosis by displaying Highest Possible Disease and Lowest Possible Disease options.
The flowchart of the methodology is presented below:

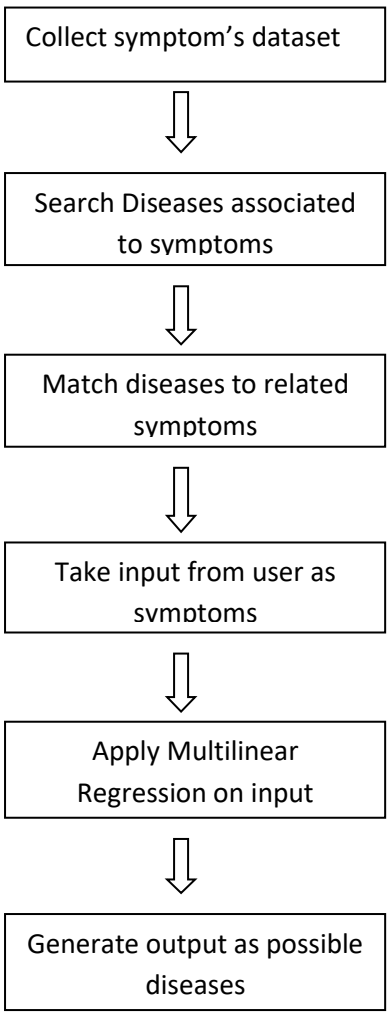

Figure 1: Flowchart of proposed model

ALGORITHM USED

Our disease prediction system leverages Support Vector Machine (SVM) for classification and Multilinear Regression (MLR) for predicting the result. MLR is a form of regression algorithm which takes into account two or more predictor variables to determine response variable(Y).
This differs from Simple Linear Regression, in which only one independent/predictor variable (X) is used. MLR is ideal for cases in which the response variable is influenced by multiple predictor variables.

4. RESULTS ANALYSIS

Result analysis in our proposed system is an essential part of this research paper. By the analysis of results, we can compare that how much better this proposed system is performing. In result analysis we will see accuracy of different diseases that are predicted using our proposed system. We have taken datasets of 100 cases for result analysis. disease prediction

system leverages Support Vector Machine (SVM) for classification and Multilinear Regression (MLR) for predicting the result. MLR is a form of regression algorithm which takes into account two or more predictor variables to determine response variable(Y).

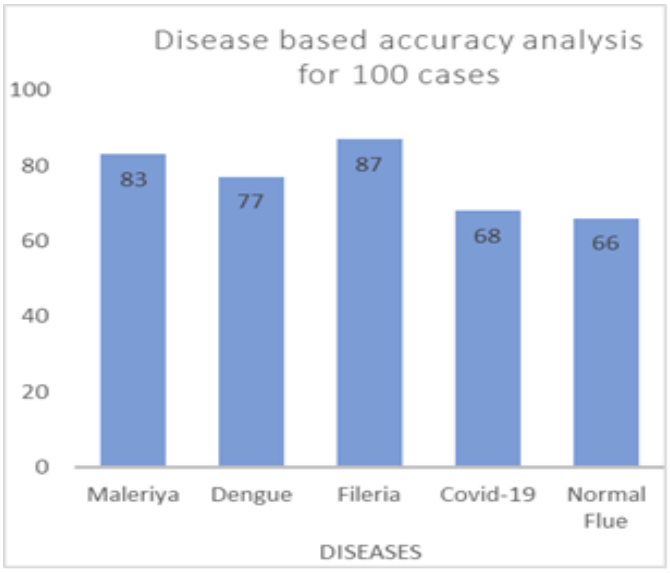

Fig -1: Disease based accuracy analysis on 100 cases

Above diagram shows the accuracy of 5 diseases that are malaria, dengue, filaria, covid-19 and normal flu. Disease based accuracy analysis for 100 cases using SVM and CNN:

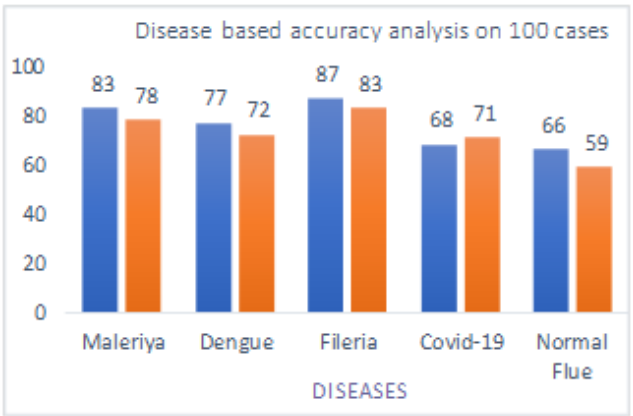

Fig -2: Disease analysis 100 cases comparison

In the above chart we can see that five diseases are given and for these 5 diseases their accuracies are also given. These five diseases are processed using two different algorithms for each consecutive bars. The blue bar shows accuracy for the diseases processed using SVM. The Orange bar shows the accuracy of diseases processed using CNN.

The results of this study have important implications for the field of disease prediction. The Disease Prediction system has the potential to improve the accuracy and speed of disease diagnosis, leading to better treatment outcomes for patients. However, there are some limitations to the system, such as the reliance on accurate and complete patient records, which may not always be available.

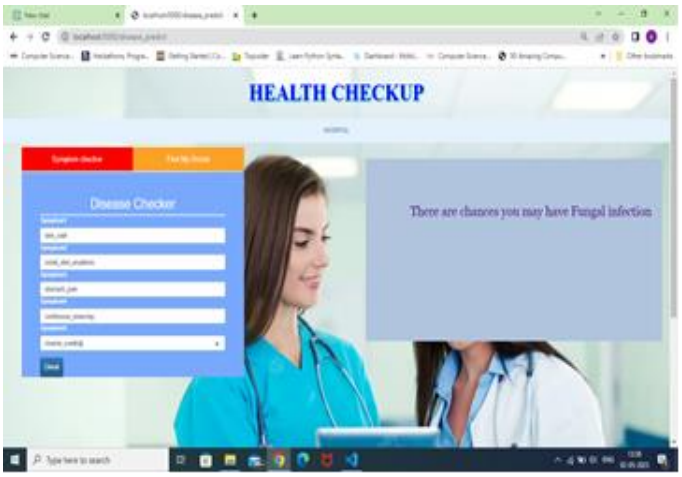

Fig -3: The final/output page

5. CONCLUSIONS

In our research, we employed a Support Vector Machine (SVM) and Multilinear Regression Algorithm to predict disease. Moreover, we tested other algorithms, such as k-Nearest Neighbor, Convolution Neural Network, and Decision Tree. Out of these algorithms, the combination of SVM and Multilinear Regression gave a higher accuracy rate than the others. We have achieved a potential disease prediction rate of up to 87%, for certain diseases, and a maximum of 68% accuracy for some diseases with a minimal number of datasets. Nevertheless, if our system is provided with an extensive amount of datasets, it has demonstrated the capacity to reach up to 95% accuracy in its disease predictions.

The collection process for such a large number of datasets on diseases and symptoms is highly time consuming; it may take several years to obtain them all and train the system with these data sets. This predictive system could be used by doctoral students for their research activities. By utilizing this disease prediction system we are now able to diagnose people based on their symptoms. Even though this system only provides possible results which do not ensure that the disease will be correctly predicted; it is still considerably accurate in predicting probable diseases. During our research sessions, we tested this model's accuracy for five different conditions and attained a maximum prediction rate of 87%.

REFERENCES

1. Khurana, Sarthak, Jain, Atishay., Kataria, Shikhar, Bhasin, Kunal, Arora, Sunny, & Gupta, Dr. Akhilesh . Das. (2019). Disease Prediction System.International Research Journal of Engineering and Technology, 6(5), 5178-5184.
2. Kamboj, Mgha. (2020). Heart Disease Prediction with Machine Learning Approaches. International Journal of Science and Research, 9(7), 1454-1458.
3. Ware, Miss.Sangy, Rakesh,Mrs.Shanu. K., &Choudhary, Mr. Bharat (2020). Heart Attack Prediction by Using Machine Learning Techniques. International Journal of Recent Technology and Engineering, 8(5), 1577-1580.

4. Shirsath, Shraddha.Subhash., & Patil, Prof. Shubhangi. (2018). Disease Prediction Using Machine Learning over Big Data. International Journal of Innovative Research in Science and Technology, 7(6), 6752-6757.
5. Marimuthu, M. , Abinaya, M. ,Hariesh,K.S., Madhan,K.,& Pavithra, Kumar. V. (2018). A Review of Heart Disease Prediction Using Machine Learning and Data Analytics Approach International Journal of Computer Application, 181(18), 20-25.
6. Battineni, Gopi., Sagaro, Getu.Gamo., Chinatalapudi, Nalini., &Amenta, Francesco. (2020). Application Of Machine Learning Predictive Models in the Chronic Disease. International of PersonalisedMedicine, 10(21), 1-11.
7. Ardabili, Sina. F., Mosavi, Amir., Khamosi, Pedram., Ferdinand, Filip., Varkonyi-Koczy. Annamaria.R. Reuter, Uwe., Rabczuk, Timon., & Atkinson, Peter M. (2020). COVID -19 Outbreak Prediction with Machine Learning.Journal of Algorithms,13(249), 1-36.
8. Shrestha, Ranjit., & Chatterjee, Jyotir. Moy. (2019). Heart Disease Prediction System Using Machine Learning. LBEF Research Journal of Science Technology and Management, 1(2), 115-132.
9. Magar, Rishabh., Memane, Rohan., Raut, Sura., &Rupnar, Prof. V.S. (2020). Heart Disease Prediction Using Machine Learning. Journal of Emerging Technologies and Innovative Research, 7(6), 2081-2085.
10. Alotaibi, Fahd. Saleh. (2019). Implementation of Machine Learning Model to Predict Heart Failure Disease. International Journal of Advanced Computer Science and Application, 10(6), 261-268.
11. Godse, Rudra A., Gunjal, Smita S., JagtapKaran A., Mahamuni, Neha S., &Wankhade, Prof. Suchita. (2019). Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively. International Journal of Advance Research in Computer and Communication Engineering, 8(12), 50-52.
12. Anitha, Dr.S., & Sridevi, Dr.N. (2019). Heart Disease Prediction Using Data Mining Techniques.Journal of analysis and Computation ,13(2), 48-55.
13. Bindhika, Galla Siva Sai., Meghana, Munaga., ReddyManchuriSathvika., &Rajalakshmi. (2020). Heart Disease Prediction Using Machine Learning Techniques. International Research Journal of Engineering and Technology, 7(4), 5272-5276.
14. Pingale, Kedar., Surwase, Sushant., Kulkarni, Vaibhav. Sarage, Saurabh., &Karve, Prof. Abhijeet. (2019). Disease Prediction using Machine Learning.International Research Journal of Engineering and Technology, 6(12), 2810-2813.
15. Chauhan Raj H., NaikDaksh N., Halpati, Rinal A., Patel, Sagarkumar J. , &PrajapatiMr. A.D. (2020). Disease Prediction using Machine Learning.International Research Journal of Engineering and Technology, 7(5) ,2000-2002.

BIOGRAPHIES

Aditya Singh is an undergraduate and his areas of interests are data structures, cloud computing, machine learning, natural language processing, and artificial intelligence. He has done training in Python with Machine Learning from Coding Ninja and have the knowledge of Java, SQL, Web Development.

Ankur Pandey is an undergraduate and his areas of interests are machine learning, natural language processing, and artificial intelligence. He has done training in Java.

Ankur Kumar Tiwari is an undergraduate and his areas of interests are machine learning and Web Development. He has a knowledge of Java, Python and MySQL.

Chandan Nishad is an undergraduate and his areas of interests are machine learning, ethical hacking. He has done training in Ethical hacking and have a good knowledge of Java.