ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Predictive Analytics for Multi-Disease Detection Based on User Symptoms Using ML Algorithms

[1] Tatavarthi Sivanwita, [2] Karri Lakshmi Ganapathi, [3] Pandiripalli Chandu, [4] Devalacheruvu Sreehari, [5] Dr. A. Rama Murthy,

[1,2,3,4,] Students, Dept.of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

[5] Professor, Dept. of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

Abstract—

A wide variety of diseases affect many people daily. Treatment decisions are based on the prognosis of the condition. Accurate medical data analysis, made possible by the tremendous growth in healthcare and medical data, helps in the early detection of illnesses and in patient treatment prior to their onset. In order to forecast the most likely disease based on the symptoms and to directly predict the likelihood that the individual may be suffering from that specific illness, this study employs a number of supervised classification algorithms, such as Decision Tree, Support Vector Machine, K-Nearest Neighbor, Logistic Regression, Naive Bayes, and Random Forest, to conduct research on the massive amounts of medical data. The model takes the output of the supervised classification algorithms and applies them to the symptoms in order to provide a final validation that suggests the possible illness that the entity may be experiencing. By integrating the predictions from each of the following classification algorithms, the model is able to provide a more precise verification than would be possible with the predictions from each model alone. Research like this may cut down on the number of false negatives and speed up the decision-making process. As a result, healthcare organizations are able to make more informed choices on the course of early patient treatment. Additionally, it helps medical practitioners come up with more efficient methods of patient treatment.

Decision Tree, Naïve Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Random Forest

INTRODUCTION

Worldwide, some 55 million people lose their lives each year, with different illnesses accounting for about 55 percent of those casualties. And according to the numbers, 74% of all fatalities are attributable to non-communicable diseases. Every now and again, a lot of people become sick. An enormous quantity of complex patient data, comprising clinical variables, diagnostic information, patient records, medical research, and ancillary materials, is generated by the advent of the digital era and technical breakthroughs. The most important aspect of therapy is illness prediction. Making a correct diagnosis from a patient's symptoms alone is sometimes the hardest thing for clinicians to perform [11]. As a result, forecasting relies heavily on machine learning to accomplish the crucial and challenging job of sickness prediction. The model finds patterns based on symptoms using medical data. By inputting their symptoms, users may get a diagnosis of the most likely illness they might be experiencing. Predicting the illness before it worsens a patient's health via realistic medical data analysis, it enables early patient therapy.

RELATED WORKS

Using existing medical data, this article offers early illness prediction and medicinal advantages. By feeding users' symptoms into supervised machine learning algorithms, this research can determine the most likely illness they may be suffering from. The algorithms that were used were Decision Tree, SVM, Naïve Bayes, RF, and KNN. The system is designed in a way that uses Naïve Bayes for illness prediction, KNN for classification, Logistic Regression for feature extraction, and Decision tree for dataset partitioning. Using metrics such as recall, precision, and F1, we found that KNN achieved an accuracy of 71.28%, Decision tree 84.50%, Random Forest 99.995%, Naive Bayes 89.50%, and SVM 96.49%. Using the grails architecture, the scientists constructed a disease predictor that achieved

a system accuracy of 98.3%. When it comes to illness predictions, this system employs the holy grail paradigm. I When physicians aren't present, patients may still access their medical records using a web or android app, which is the main topic of this article. The software can fix little issues, but for a more comprehensive checkup, users need go to a hospital. Using user-supplied symptoms and the anticipated degree of disease, the system aims to provide users with fast and accurate illness prediction. In order to diagnose the user, the system checks the supplied symptoms against the data in the database. It either gives the correct diagnosis or flags it as a "wrong symptom" depending on whether the symptoms are in the database. After that, a window would appear asking if the user wants to preserve the symptoms in the data. The system used a number of algorithms to provide reliable predictions, including RF, Decision Tree, Naive Bayes, and KNN. After that, the system is trained in a web app using Django and Python so that users may get solutions without having to travel to the doctor.

The core idea of the article is to provide a virtual, accurate, and early assessment of any health issues. Going to the doctor and getting a diagnosis via the old methods takes a long time. As a result, the study spends much of its time analyzing the dataset to determine what the person may be going through. There are around 230 illnesses included in the collection. By analyzing the user's age, gender, and symptoms, the system is able to make a diagnosis and display the condition accordingly. Next, a number of machine learning models were applied to the dataset, including RUS Boosted trees, Fine, Medium, and Coarse Decision trees; Kernel and Gaussian Naive Bayes; and Weighted and Subspace KNN. Different models have different levels of accuracy; the project was built using the most accurate model. Of the algorithms tested, Weighted KNN achieved a remarkable 93.5% accuracy, while Fine KNN came in second with an impressive 80.3%. 2 Cardiology illness prediction using supervised learning algorithms is the main topic of the article. With the use of characteristics including age, cholesterol, chest discomfort, and blood pressure, the report intends to provide a more accurate means of predicting the likelihood of cardiac problems. We used a number of data mining methods and algorithms to determine the patient's risk level, including Random Forest Classifier, Support Vector Classifier [13], KNN Classifier, Decision Tree, Naive Bayes Classifier, and Logistic Regression. With an accuracy of 87%, the KNN classifier has achieved the best results.
In [5], a disease prediction system was built using Decision Tree, K-Nearest Neighbour, Random Forest, and Naive Bayes algorithms. Each technique was treated as a separate model, and the results of each model's predictions were shown. When it came to estimating the outcome, these models demonstrated considerable improvements in accuracy. the eleventh This study focuses on the idea of using the user's symptoms to potentially diagnose chronic illnesses such as diabetes, cancer, arthritis, and others. It was made specifically for those using it. In order to make the most precise illness prediction possible, the scientists used structured and textual data collected from actual hospitals. For illness prediction, the model employs the Naive Bayes method; for clustering, it employs the KNN algorithm; and for the final output, which will be a set of integers, it employs Logistic Regression. Accuracy [19], Recall, and F1-Measure are some of the evaluation measures used to determine performance.
[14] Structured and unstructured analysis are the main topics of this study. Using structured analysis, the author zeroes down on a specific area and community to find cases of chronic illnesses. To forecast potential illnesses based on input symptoms, the author used Random Forest, Naïve Bayes Classifiers, Decision tree, and K-Nearest Neighbor algorithms in unstructured analysis. With accuracies of 95.7%, 94.5%, and 92.4% respectively, the Random Forest model was the most accurate, followed by K-Nearest Neighbor, Naïve Bayes Classifiers, and Decision Tree.

THE PROPOSED METHOD

Data Collection

It is a procedure that involves gathering data from several sources and then loading it into an analysis program so that patterns may be formed. Following the instructions in [5], this stage entails obtaining the necessary datasets from the Kaggle repository and splitting them into a training dataset and a testing dataset. Following the instructions in [5][14], the datasets include 132 features and 4920 records for training and 41 records for testing. Preparing Data for Use It is a method for making better business choices based on more insightful data that is transformed from raw data. Obtaining the dataset, importing it and any necessary libraries, identifying and handling missing values (as in [10][18]), encoding categorical data (as in [12][18]), splitting the dataset, and feature scaling to extract relevant data (as in [18]) are all parts of data preprocessing, which leads to more accurate results (as cited in [5][12]). The datasets have been preprocessed by adding records in both the training and testing datasets to enhance the model's performance and for more accurate prediction. This was done in addition to the current datasets, as indicated in

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

[5][14]. There are 54 records for testing and 5000 records for training in the proposed data set. Additionally, a new categorization has been introduced to the goal feature, which tells the user they are healthy if they don't have any symptoms. Model functionality Importing libraries that enable machine learning techniques [13] into the proposed model is the first stage. These libraries include NumPy, pandas, sklearn, statistics, and tkinter. Reading the training and testing datasets is the next stage, as described in [5]. Following the preprocessing phase, the training and testing datasets are divided into feature and target attribute sets. The next step is to fit the training data into the models by importing different machine learning classifiers using the sklearn module. On one hand, we have machine learning classifiers like Decision Tree, Random Forest, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes Classifier, and Logistic Regression. On the other hand, we have Support Vector Machine, K-Nearest Neighbor, Random Forest, Support Vector Machine, Naïve Bayes Classifier, and Logistic Regression. The most likely illness the patient may be experiencing is determined by using the mode function from the statistics library on a global list that contains the predictions from the aforementioned models. This list may have numerous values that are similar or different. To get the final forecast, the suggested model makes use of statistical mode, as shown in Fig. 1. A global list is created by combining the predictions acquired from several machine learning algorithms. The statistics library's mode() method returns the most commonly predicted ailment.

$$\text{Final Prediction} = \text{Mode}(p1, p2, p3, p4, p5, p6)$$

$$pi \text{ - prediction of algorithm } i \text{ (for } i = 1, 2, \ldots 6)$$

This outcome, as illustrated in Figure 1, is thus taken as the ultimate forecast. The combined predictions of the classifiers provide a final prediction that strengthens the model significantly [11]. It improves upon the results achieved by the separate models mentioned in [5] and makes the prediction even more precise.

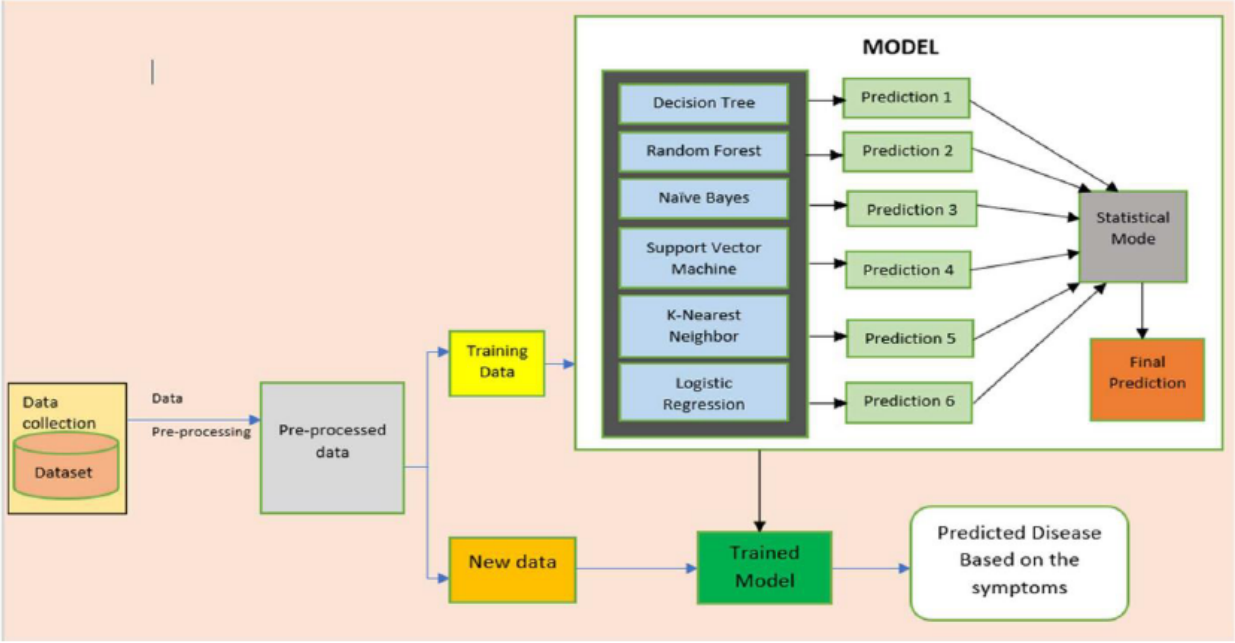

Fig. 1. System Architecture

Fresh Information

By inputting more data in the form of symptoms, the user may get a forecast back that shows which condition is most likely affecting them. The addition of this fresh data to the database strengthens the dataset, which in turn allows for more accurate predictions in the future. Method for Execution The suggested model requires the import of many libraries, including NumPy, pandas, statistics, tkinter, and sklearn. Apply several pre-processing methods to the datasets imported from Kaggle. Divide the datasets used for training and testing into those with features and those with targets. Naive Bayes Classifier, K-Nearest Neighbor, Decision Tree, Logistic Regression, Random

Forest, and others should be fitted using the training data. When the models have been trained with the necessary data, the statistical mode of their predictions is shown, showing which illness is most likely to As fresh information is being added, please enter the patient's symptoms on the website. The Community and Mutual Aid For both training and prediction, we use a Vector Machine [8][16]. the individual could be suffering from. By merging the predictions of the various diagnostic tools, the most likely illness that the user is experiencing is shown.

RESULTS AND DISCUSSION

Remarkable Outcomes Decisions are made faster as a result of the project. It offers a straightforward graphical user interface (GUI) that allows the user to input symptoms from a pre-populated list. Using a website built with the Django framework, users have access to the same functionality as the GUI. It The user may access the login page by clicking on "I have already account," as seen in Figure 3. On Then, based on the symptoms they provided, get a diagnosis of the most likely illness they might be experiencing. include a registration page that allows users to register, as seen in Figure 2. and sign up for an account from scratch if they don't already have one. After the user logs in, the system takes them to the main screen (Fig. 4), where they may input their symptoms and get a forecast.

Outcomes

Previous work has been compared with the output of the suggested model. In earlier work, there existed a method for illness prediction in which each algorithm was shown as a separate model and the results of those predictions were shown. Each of these models was accurate in its own way and produced its own set of findings. In comparison to other efforts, the suggested model provides a straightforward and easy-to-understand user interface that accurately predicts the most likely illness a person may be experiencing. The combined strength of six supervised ML algorithms is used in this forecast. As a consequence, the forecast is improved upon and becomes even more accurate than before.

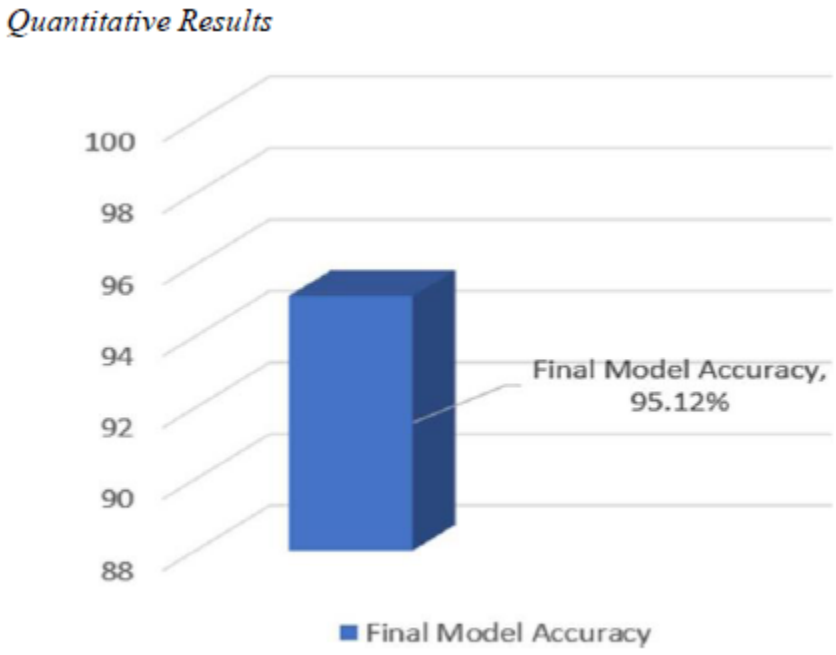

Fig. 6. Accuracy of proposed model

Accuracy of the final model developed by combining the six supervised machine learning algorithms is stated in Fig. 6.

FUTURE SCOPE

In order to improve the accuracy of illness prediction using image data sets and algorithms such as CNN. For the purpose of expanding the model to include additional categorization methods or algorithms beyond the already With the goal of reducing death rates via more precise prediction. current methods to enhance precision. More illnesses may be added to the current API in the future.

CONCLUSION

Using a model that integrates the results of several Supervised Machine Learning Classification Algorithms, the research hopes to diagnose the patient's illness based on their chosen symptoms. With a mean accuracy of 95.1%, this study's final prediction was obtained using six Machine Learning methods. Compared to earlier studies, this demonstrates a considerable improvement in refinement, accuracy, and dependability. Along with the patient's illness, the model saves the user-entered data. These might be referred to as historical records that can assist in future forecast. In order to enhance the model's operability and user experience, this work incorporates a graphical user interface (GUI) for more seamless interaction with the model. It helps healthcare organizations and hospitals make more informed choices on patient care. In order to facilitate early patient treatment, the model analyses practical medical data and forecasts the illness before the patient's health deteriorates.

REFERENCES

1. Mr. A. Rohith Naidu, Dr C K Gomathy: “The Prediction of Disease Using Machine Learning”, International Journal of Scientific Research in Engineering and Management (IJSREM)
2. Prof. Suchita Wankhade, Rudra A. Godse, Karan A. Jagtap, Smita S. Gunjal, Mahamuni Neha S: “Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 9, Issue 4, April 2020
3. Keniya Rinkal, Ninad Mehendale, Aman, Khakharia, Mahesh Warang, Vruddhi Shah, Vrushabh Gada, Tirth Thaker, Manjalkar Ruchi: “Disease prediction from various symptoms using machine learning”
4. Megha Kamboj: “Heart Disease Prediction with Machine Learning Approaches”, International Journal of Science and Research (IJSR) (2018)
5. Mr. Analp Pathak, Anuj Kumar: “A Machine Learning Model for Early Prediction of Multiple Diseases to Cure Lives”, Turkish Journal of Computer and Mathematics Education, 2021
6. B. Prajna, Divya Mandem: “Multiple Disease Prediction System”, International Journal of Innovative Research in Technology (IJIRT) (2021)
7. S.MD. Jabeer, G. Chakravarthi: “Heart Disease Prediction Based on Machine Learning Techniques”, International Journal of Innovative Research in Technology (IJIRT)
8. M.Lakshmi Narayana, O. Rama Praneeth Kumar, N.Sai Prasad, T. Naga Sampath, N Md Jubair Basha: “Chronic Disease Prediction Using gradient Boosting and KNN Classifier”, International Journal of Innovative Research in Technology (IJIRT) October 2021
9. Mursal Furqan, Kanwal Awan, Hiba Rajput, Sanam Narejo, Adnan Ashraf: “Heart Disease Prediction Using Machine Learning Algorithms”, International Conference on Computational Sciences and Technology December (2020)
10. Khurana Sarthak, Gupta Dr. Akhilesh Das, Jain Atishay, Bhasin Kunal, Kataria Shikhar, Aror Sunny: "Disease Prediction System", International Research Journal of Engineering and Technology, 6(5), 5178-5184.
11. Sarage Saurabh, Pingale Kedar, Kulkarni Vaibhav, Surwase Sushant, Karve Prof. Abhijeet: "Disease Prediction using Machine Learning", International Research Journal of Engineering and Technology, 6 (12), 2810-2813.
12. Rinal A, Chauhan Raj H, Naik Daksh N, Sagarkumar J, Halpati Patel, Prajapati Mr. A.D: "Disease Prediction using Machine Learning", International Research Journal of Engineering and Technology, 7(5), 2000-2002.
13. Gopi Battineni, Chinatalapudi Nalini, Francesco Amenta, Sagaro Getu Gamo: “Application Of Machine Learning Predictive Models in the Chronic Disease”, International of PersonalisedMedicine, 10(21), 1- 11.
14. Narendran. G, Pramoth Krishnan. T, Nivethitha. A: “Smart Disease Prediction Using Machine Learning”, International Journal of Innovative Science and Research Technology, Volume 6, Issue 6, June 2021
15. H. Patel, S. Patel, "Survey of data mining techniques used in the healthcare domain," Int. J. of Inform. Sci. and Tech., Vol. 6, pp. 53-60, March 2016.

16. K. Arumugam, Tatiana Gonzales-Yanac, Mohd Naved, Orlando Leiva- Chauca, Priyanka P. Shinde, Antonio Huaman-Osorio, Materials Today: Proceeding, August 2021
17. Computer Science, University of Benin: "Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction", International Journal of Scientific and Research Publications (IJSRP) 2021
18. Priya S. S., and Kumar, S. (2018). Chronic Kidney Disease Prediction Using Machine Learning. International Journal of Computer Science and Information Security (IJCSIS), 16(4)
19. T.V. Sriram, M.V. Rao, G.S. Narayana, D. Kaladhar, T.P.R. Vital, Intelligent parkinson disease prediction using machine learning algorithms, International Journal of Engineering and Innovative Technology (IJEIT) 3(3), 1568 (2013)