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

Artificial Intelligence-Based Multi-Disease Prognosis from User Symptoms

[1] Mrs. Neha Shireen, [2] MOHAMMED ABDUL GHAFOOR ANSARI, [3]ISLAM KHAN, [4]ABRAR AHMED KHAN, [5] AMAAN SHAREEF

[1] Assistant Professor, Department of AIML, Lords Institute of Engineering & Technology.

[2345] Student, Department of AIML, Lords Institute of Engineering & Technology.

Abstract—

Many people deal with various diseases on a daily basis. Treatment decisions are based on the prognosis of the condition. The exponential growth of healthcare IT has made it possible to analyze medical data with pinpoint accuracy, which in turn helps in the early detection of diseases and the planning of patient treatment. 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. The model provides a more precise verification when compared to the predictions given by individual models after integrating the predictions from all the classification algorithms listed below. Medical data is used by the 3rd Deekshitha Karingula Department of Computer Science and Engineering model to discover patterns based on symptoms. The users may input their symptoms and get a diagnosis of the most likely illness based on those findings. Predicting the illness before it worsens a patient's health via realistic medical data analysis, it enables early patient therapy.Research like this may cut down on the number of false negatives and speed up the decision-making process. In terms of early patient care, it aids healthcare associations in making more informed judgments. Additionally, it helps medical practitioners come up with more efficient methods of patient treatment.

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

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 first In this study, we use existing medical data to provide early illness prediction and other medicinal advantages. By feeding users' symptoms into supervised machine learning algorithms, this research can determine the most likely illness they may be suffering from. Decision Tree, SVM, Naïve Bayes, RF, and KNN were the algorithms that were used. The system is designed in a way that incorporates 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%. The holy grail framework for illness predictors is used by this system. [2] in 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. While the software may help with little issues, a more comprehensive checkup is still required in a hospital. Rapid and accurate disease prediction based on symptoms and expected severity is the system's main

emphasis. After receiving the symptoms, the system checks them against the data in the dataset. If the symptoms are in line with what is in the database, it will either provide the correct diagnosis or flag it as an incorrect symptom. The next question that would pop up is whether or not the user wants to keep the symptoms in their records. The system used RF, SVM, Decision Tree, Naive Bayes, and KNN algorithms to provide precise predictions. Users may get solutions without going to the doctor's office since the system is then taught via a web application utilizing Django and Python.

[3] The primary goal 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. Therefore, in order to make it shorter, the study utilizes the dataset to speculate on the person's possible pain. Approximately 230 illnesses are available for processing in the dataset. 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. At 93.5%, the Weighted KNN method outperformed the second-best algorithm, Fine KNN, which achieved 80.3% accuracy. [4] The application of supervised learning algorithms for the prediction of cardiac illnesses is the main subject of the article. Age, cholesterol, chest discomfort, and blood pressure are some of the risk factors for cardiovascular disease, and this study tries to help improve the accuracy of these predictions. The patient's risk level was classified using a number of algorithms and data mining approaches, including Random Forest Classifier, Support Vector Classifier [13], KNN Classifier, Decision Tree, Naive Bayes Classifier, and Logistic Regression. The best accuracy, 87%, was achieved using the KNN classifier. 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. There was evidence that these models were more accurate in their estimations. the eleventh Predicting the user's risk of developing chronic illnesses such as diabetes, cancer, arthritis, and others from his reported symptoms is the primary focus of this research. Its only purpose was to serve the needs of end consumers. To provide the best possible illness prediction, the scientists used structured and textual data collected from actual hospitals. A number of machine learning methods are used by the model. These include Naive Bayes for illness prediction, KNN for clustering, and Logistic Regression for the final output, which will be a series of ones and zeros. Accuracy [19], Recall, and F1-Measure are some of the evaluation measures used to determine performance. [14] Two main categories of analysis are covered in this paper: organized and unstructured. The author employs structured analysis to zero down on a specific area and community's chronic illness burden. 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. The models with the greatest accuracy rates were the random forest model (95.7%), the K-Nearest Neighbor classifier (95.6%), the Naïve Bayes classifiers (94.5%), and the decision tree (92.4%). The first step in any pattern analysis or formulation is data collection, which entails gathering the appropriate information from a variety of sources and loading it into an analytical or pattern-making program. According to [5], the first step is to get the necessary datasets from the Kaggle repository. Then, divide the dataset into two parts: the training dataset and the testing dataset. Following the instructions in [5][14], the datasets include 132 features and 4920 records for training and 41 records for testing. Section B: Preprocessing Data In business, it's a way to go from raw data to data that's useful for making choices. A more accurate output is the result of data preprocessing, which entails 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 (as in [18]) to obtain relevant data (as in [5][12]). To enhance the model's performance and provide more accurate predictions, the datasets have been preprocessed by adding records in both the training and testing datasets, as indicated in [5][14]. This was done in accordance with the method outlined in [10].

One hundred forty-four test records and five thousand training records make up the suggested dataset. 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. C. The Model's Operation 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. The next step is to read the training and testing datasets in the same way as described in [5]. Partitioning the training and test datasets according to feature and

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

target properties follows preprocessing. The next step is to fit the training data into the models by importing different machine learning classifiers using the sklearn module. Various machine learning classifiers such as Decision Tree, Random Forest, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes Classifier, and Logistic Regression are used for training and prediction purposes. A global list is used to hold various similar or different values, which are predicted by the aforementioned models. Then, the mode function from the statistics library is applied to this list in order to determine the most likely ailment that the individual may be suffering from. The suggested model's usage of statistical mode, as seen in Figure 1, aids in obtaining the final forecast. 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. Projection Outcome = Mode (p1, p2, p3, p4, p5, p6) pi - algorithm i's prediction (where i = 1, 2,..., 6); its output is consequently regarded as the ultimate forecast, as seen in Figure 1. 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.

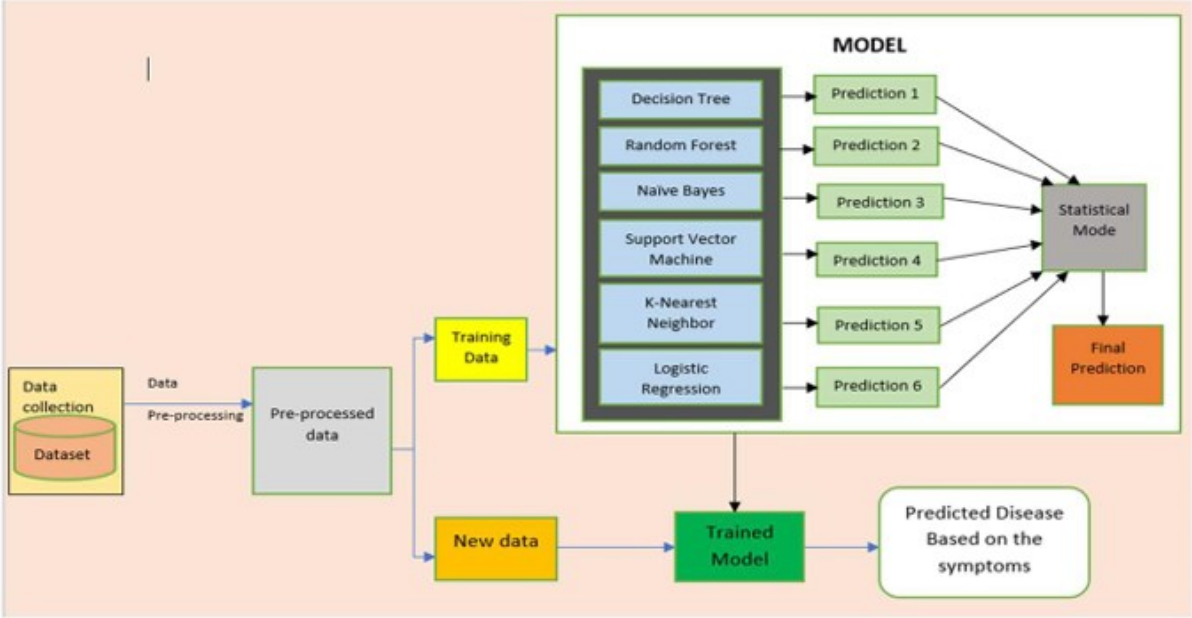

Fig. 1. System Architecture

Section D. Upcoming 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. By adding this fresh information to the database, we can fortify the dataset and make more accurate predictions going forward. E. Steps to Take Before Starting Include NumPy, pandas, statistics, tkinter, and sklearn in the model you're proposing to build. • Use a variety of pre-processing methods on the datasets imported from Kaggle. To begin, divide the two sets of data into one for testing and one for features and target characteristics. • Incorporate the training data into prediction and training models such as Decision Tree [8], Logistic Regression [9], Random Forest [1][2][6][14], Naive Bayes Classifier, K-Nearest Neighbors, and Support Vector Machine [8][16]. After feeding the models the right data for training, you may see which illness is most likely afflicting the patient in the statistical mode of the models' predictions. • New data is being entered on the website; in this case, the patient's symptoms. • The user's symptoms are used to determine the most likely illness, and the most likely diagnosis is then shown by merging the model predictions.

RESULTS

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Section A. Analytical Findings • The project gives a user-friendly graphical user interface (GUI), which improves the speed of decision-making. The user may input symptoms from a list and get a prediction of the most likely ailment they could be suffering from based on those symptoms. • Users may access the same features offered by the GUI via a website built using the Django framework. In case the user does not already have an account, they may create one by going to the registration page, as seen in Figure 2. Figure 3 shows the login screen that the user may access by clicking the "I have already account" button. As seen in Figure 4, the user is sent to the home page upon entering their login credentials. From there, they may input their symptoms and get a prognosis.

The health of the patient is improved, which enables prompt medical attention. There was a comparison between the prior work and the suggested model's output. In the earlier study, there existed a disease prediction system that presented the predictions produced by each individual model. Each algorithm was treated as an independent model. The outcomes produced by each of these models were distinct and contingent upon their own accuracy. In contrast to other efforts, the suggested model provides a clear and straightforward interface for determining the patient's most likely illness. Six supervised machine learning algorithms work together to make this prediction. As a consequence, the forecast is improved upon and becomes even more accurate than before. Findings Based on Numbers

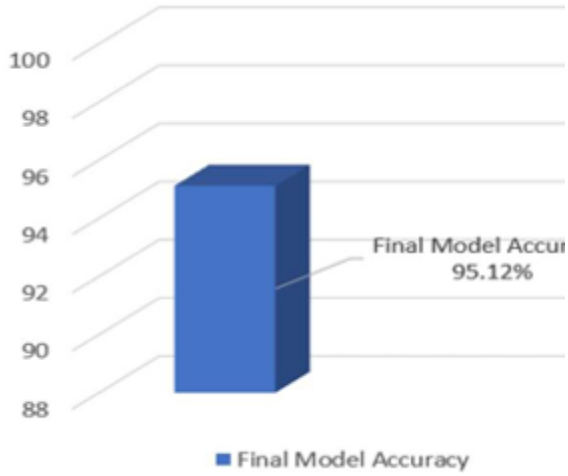

Fig. 6. Accuracy of proposed model

Figure 6(V) displays the accuracy of the combined model from all six supervised machine learning techniques. Perspectives for the Future • To improve the accuracy of illness prediction by adding picture data sets and using algorithms such as CNN. • To increase the model's accuracy, expand it to include other categorization approaches or algorithms other than the ones now in use. Additional illnesses may be included into the current API down the road. • In order to reduce death rates, we need to make better predictions.

Conclusion

By using a model that integrates the predictions from several Supervised Machine Learning Classification Algorithms, the research intends to predict the illness based on the symptoms specified by the patient. The final forecast in this research was obtained using six Machine Learning methods, which had a mean accuracy of 95.1%. Superior to earlier efforts, this demonstrates a new level of precision and dependability. The user-entered data and the patient's illness are stored in the model. For the sake of future forecast, they may serve as historical records. In order to make the model more user-friendly and easier to operate, this project makes use of a graphical user interface (GUI) to facilitate better interaction with the model. Healthcare facilities and organizations may use it to improve their decision-making when it comes to patient care. Prior to the disease's progression, the model analyses real-world medical data and makes a prediction about it.

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)