Multiple Disease Prediction Using ML

Alok Katiyar, Sajid Ali, Sameer Ray

***Abstract:** Accurate and on-time analysis of any health-related drawback is vital for the interference and treatment of the sickness. The standard method of diagnosis may not be sufficient in cases of a significant illness. Developing a medical diagnosis system supported by machine learning (ML) algorithms for predicting illnesses will facilitate more accurate diagnoses than the standard methodology. We've designed a disease prediction system using ML. A Disease Prediction System using Machine Learning could be a system that predicts sickness based on data or symptoms entered into the system and provides accurate results based on that data. This predictive disease model, utilising machine learning, is completed entirely with the assistance of Learning Machines and the Python programming language, leveraging its Flask Interface and utilising previously offered databases from hospitals to predict the illness.*

***Keywords:** Machine Learning, Disease Prediction, Decision Tree, Random Forest, Symptoms.*

I. INTRODUCTION

Disease Prediction using Machine Learning is a system that predicts the likelihood of a disease based on the information provided by the user. It also predicts a patient's illness or the user's condition based on the information or symbols that enter the system, and provides accurate results based on that information. If the patient is not so bad and the user just wants to know the type of disease, go through it. It's a system that provides users with tips and tricks to maintain a healthy lifestyle, and also offers a way to diagnose diseases using this prediction. This Disease Prediction Using Machine Learning is fully developed with the help of Python Programming Learning and Language Equipment with its Flask Interface and using the previously available database by Hospitals that use that will predict the disease (Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., & Lo, B. et al. (2017). Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics, 21(1), 4-21.) [1].

Today, doctors utilise various scientific technologies to identify and diagnose not only common diseases, but also many deadly ones. Effective treatment is permanent; it results from an accurate diagnosis. Doctors can sometimes fail to make the right decision while diagnosing a patient's disease, so disease planning systems that use machine learning skills help in such situations to obtain accurate results.

II. LITERATURE REVIEW

A Bayesian Inference Naive Bayes classifiers are a type of simple probabilistic classifier based on applying Bayes' theorem to characteristics with strong (naive) independence assumptions. They're one of the most basic Bayesian network models, but when combined with kernel density estimation, they can achieve higher levels of accuracy. The number of parameters required by Naive Bayes classifiers is linear in the number of variables (features/predictors) in a learning problem, instead of expensive iterative approximation, which is used for many other types of classifiers, maximum likelihood training can be done simply by evaluating a closed-form expression, which requires linear time.

A. Logistic regression:

Logistic regression is a type of analytical modelling. It's used to examine a dataset in which one or more independent variables influence the outcome. With a random state of 0, Logistic Regression was imported. The training model was then fitted. The accuracy of the test was 87.09 per cent.

B. Random forest:

Random forest classifier may be a powerful supervised classification tool. RF generates a forest of classification trees from a given dataset, instead of a single classification tree. Every one of those trees produces a classification for a given set of attributes. From 'sklearn ensemble', 'Random Forest Classifier' was imported (Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32) [2]. The estimators are unbroken at ten and random at 0. Then the coaching model was fitted. The testing accuracy was 90.32%. Call Tree. The testing accuracy was 90.32%. A decision tree is a tree-like diagram in which the internal nodes represent a check on an attribute, each branch denotes the outcome of the test, and each leaf node denotes a category label. Call Tree was foreign wherever the random state was unbroken as zero, and so the coaching model was fitted. The testing accuracy was 83.87%. 6. Results Amongst all classification techniques, testing accuracy was best within the case of the random forest and SVM approach, with an accuracy of 90.32%.

Manuscript received on 27 March 2023 | Revised Manuscript received on 04 April 2023 | Manuscript Accepted on 15 May 2023 | Manuscript published on 30 May 2023.
*Correspondence Author(s)

Dr. Alok Katiyar, Professor, Department of CSE, Galgotias University, Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India ORCID ID: https://orcid.org/0000-0002-1645-6585

Sajid Ali, Student, Department of CSE, Galgotias University, Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. ORCID ID: https://orcid.org/0009-0007-5961-3070

Sameer Ray*, Student, Department of CSE, Galgotias University, Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. E-mail: sameer.ray.official@gmail.com, ORCID ID: https://orcid.org/0009-0006-7319-2332

III. PROPOSED SYSTEM

We have blended structured and unstructured facts within the healthcare field to determine disease risk in this project—the Use of a latent aspect model to recreate missing statistics in scientific Information acquired from online sources. We may also verify the most critical persistent illnesses in a particular area and population using statistical data. We consult with specialists in sanatorium management to learn about functional capabilities while working with dependent data. Inside the Case of unstructured textual content files, we use the random forest Algorithm to pick out capabilities automatically.

A. Data collection

Data collection has been conducted online to identify the disease. Here, the essential symptoms of the sickness are collected, i.e., no dummy values are entered. The symptoms of the illness are collected from various health-related websites.

Data Preprocessing

Before feeding the information into the Prediction model, the following data cleaning and preprocessing steps are performed.

● Checking null values and filling in the missing values using the forward fill technique

● changing information into completely different cases

● Standardizing the information, mean, and variance

● Splitting the dataset into coaching and testing sets

B. Building a model

Many methods are used to perform data mining. Machine learning is one such approach. Random forest Machine learning strategies include grouping, clustering, summarization, and many others. Since classification techniques are used in this project, classification is one of the data mining processes in the Phase of categorical data classification.

C. Prediction

Prediction using random forest:

Prediction done by a Random Forest model using the Flask framework, trained on a chronic disease dataset.

IV. THE DATABASE AND THE METHODS

A. Set of data

The dataset we utilise is the Kaggle dataset, which is freely available on Kaggle. (n.d.). Chronic Kidney Disease Dataset Retrieved from https://www.kaggle.com/mansoordaku/ckdisease [3]. The dataset was retrieved from its repository. It comprises 402 samples—numerous types of classes. Eleven of the twenty-five qualities are numerical, and thirteen are non-numerical; one is a class attribute, while the other is a nominal. The data collection includes the total number of missing values. The dataset's information is available here, including data from the patient, such as age, blood pressure, and specific symptoms, as well as albumin, sugar, red blood cells, and other relevant details.

B. Data Gathering

We use a dataset from the Kaggle Machine Learning Repository in this experiment (Mehta, N., Pandey, S., & Verma, S., 2017). Disease Prediction System using Machine Learning over Cloud. 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 331-336.) [4]. Furthermore, the initial dataset was gathered. There are 583 liver patients in this dataset, with 75.64 per cent of male patients and 24.36 per cent of female patients. There were 11 distinct parameters in this sample, whereas

Select ten parameters for further investigation and one parameter as a target class. For example,

•Patient's age (in years)

•Gender: The Patients' Gender

•TB stands for total bilirubin.

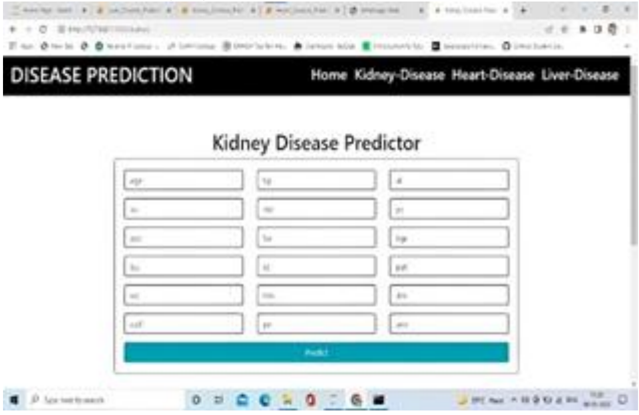
•DB: Bilirubin Direct

• Alkaline Phosphatase

• Alanine Aminotransferase (SGPT)

• Aspartate Aminotransferase (SGOT)

•Total Proteins (TP)



DISEASE PREDICTION

Kidney Disease Predictor

Fig. 1 Web applications interface

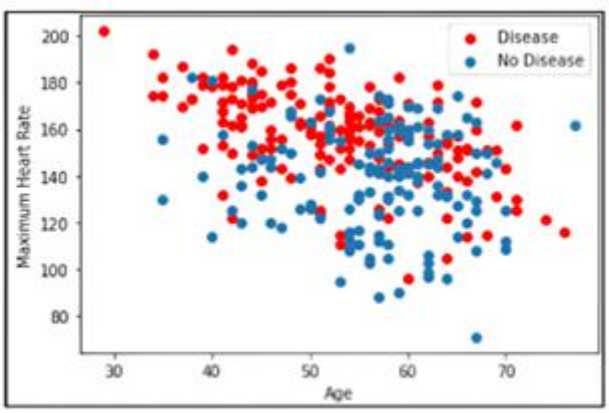


Fig. 2 Positive and negative cases

V. RELATED WORK

Many researchers have used machine learning techniques, such as KNN, Naïve Bayes, and Decision trees, to develop disease Prediction strategies.

Retrieval Number: 100.1/ijrte.A75680512123
DOI: 10.35940/ijrte.A7568.0512123
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Satyabhama Balasubramanian and Balaji Subramani discussed a system to reduce the multiple diseases that exhibit similar symptoms, thereby increasing the accuracy of such diagnoses (Balasubramanian, S., & Subramani, B., 2019). A system for reducing multiple diseases showing similar symptoms using machine learning techniques. Journal of Medical Systems, 43(7), 157.) [5]. It has received 71.53% accuracy. Aditya Arya, Sudhanshu, Rohan Agarwal, attempted to show and visualized the result of our study and this project By comparing with other techniques, it scores accuracy of 68.5% (Arya, A., Sudhanshu, & Agarwal, R. (2018). Disease prediction using machine learning algorithms. International Journal of Advanced Research in Computer Science, 9(5), 415-419.) [6]. Iqra Anjum, Mohammed Afreed, Mohammed Kalam has developed a system which predicts the disease based on the information or the symptoms he/she enter into the system and provides the accurate results based on that information (Anjum, I., Afreed, M., & Kalam, M. (2020). Disease prediction using machine learning. International Journal of Advanced Science and Technology, 29(5), 1592-1601.) [7]. Raj H. Chauhan, Daksh N. Naik, Rinal A. Halpati, Sagarkumar J. Patel, Mr. A.D. Prajapati developed a system analyzes the symptoms provided by the user as input and gives the probability of the disease as an output Disease Prediction is done by implementing the Decision tree Classifier (Chauhan, R. H., Naik, D. N., Halpati, R. A., Patel, S. J., & Prajapati, A. D. (2021). Disease prediction using a decision tree classifier. In Proceedings of the 2021 3rd International Conference on Advances in Electronics, Computers and Communications (pp. 148-152). IEEE) [8]. Decision tree Classifier calculates the probability of the disease. With the growth of big data in biomedical and healthcare communities, it also provides probability estimation of the system, showing the various probabilities of how the system behaves when a certain number of predictions are made. It also provides recommendations for patients based on their final results and symptoms, indicating which treatments to use and which to avoid, ultimately leading to the desired outcome. Here, we've combined the structured and unstructured forms of data for the general risk analysis required for predicting the disease. Using the structured analysis, we will identify the chronic diseases in a particular region and community. In unstructured analysis, we automatically select features with the help of algorithms and techniques. This technique takes symptoms from the user and predicts the disease accordingly supported the symptoms that it takes and also from the previous datasets, it also helps in continuous evaluation of viral diseases, heart rate, blood pressure, sugar level and far more which is within the system and along with other external symptoms its predicts the acceptable and accurate disease and it gives the prescription details for that disease. Additionally, the data entered by the user is stored within the database that has been created.

VI. EVALUATING THE MODEL % RESULT

The results obtained from our model are summarised in the following table:

Model	Accuracy
Liver disease	77.97%
Kidney disease	90%
Heart disease	85%

VII. CONCLUSION

Ultimately, I conclude that this project, Disease Prediction Using Machine Learning, is beneficial to everyone's day-to-day life. It is mainly more Important for the healthcare sector, because they are the ones who use these systems daily to predict diseases in patients based on their general information and symptoms they have experienced. Today, the health industry plays a significant role in curing patients' diseases. Hence, this also provides some assistance to the health industry in informing users. Additionally, it is helpful for the user in case they do not want to visit the hospital or any other clinic. By entering their symptoms and other relevant information, the user can identify the disease they are suffering from. The healthcare industry can also benefit from this system by simply asking the user about their symptoms and entering them into the system. With just a few seconds, they can identify the exact disease, and to some extent, the accuracy is also ensured. If the health industry adopts this project, doctors' workloads can be reduced, and they can more easily predict patients' diseases. The Disease prediction aims to provide a forecast for various and generally occurring diseases that, when unchecked and sometimes ignored, can turn into fatal diseases and cause numerous problems for the patient and their family members.

DECLARATION

Funding/ Grants/ Financial Support	No, we did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence
.Availability of Data and Material/ Data Access Statement	Data sets are taken from the Kaggle community.
Authors Contributions	The entire project was completed under the guidance of Dr Alok Katiyar. All documentation work and somewhat research has been done by (Author) Sajid Ali. Complete Web applications and Research have been done by (Author) Sameer Ray.

REFRENCES

1. Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., & Lo, B. et al. (2017). Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics, 21(1), 4-21. [CrossRef]
2. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. [CrossRef]
3. Kaggle. (n.d.). Chronic Kidney Disease Dataset. Retrieved from https://www.kaggle.com/mansoordaku/ckdisease
4. Mehta, N., Pandey, S., & Verma, S. (2017). Disease Prediction System using Machine Learning over Cloud. 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 331-336.

5. Balasubramanian, S., & Subramani, B. (2019). A system for reducing multiple diseases showing similar symptoms using machine learning techniques—Journal of Medical Systems, 43(7), 157.
6. Arya, A., Sudhanshu, & Agarwal, R. (2018). Disease prediction using machine learning algorithms. International Journal of Advanced Research in Computer Science, 9(5), 415-419.
7. Anjum, I., Afreed, M., & Kalam, M. (2020). Disease prediction using machine learning. International Journal of Advanced Science and Technology, 29(5), 1592-1601.
8. Chauhan, R. H., Naik, D. N., Halpati, R. A., Patel, S. J., & Prajapati, A. D. (2021). Disease prediction using a decision tree classifier. In Proceedings of the 2021 3rd International Conference on Advances in Electronics, Computers and Communications (pp. 148-152). IEEE.

AUTHORS PROFILE

Dr. Alok Katiyar, Professor, Department of CSE (Galgotias University, Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India **About**: Dr. Alok Katiyar is a highly respected professor in the Department of Computer Science and Engineering at Galgotias University, located in Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. With years of experience in academia and industry, Dr. Katiyar is known for his expertise in areas such as data analytics, machine learning, and artificial intelligence. He is an accomplished researcher and has published numerous papers in top-tier conferences and journals. Dr. Katiyar is also a sought-after speaker, having delivered talks at various national and international conferences. His dedication to teaching and research has earned him the respect and admiration of his students, colleagues, and peers in the academic community.

Sajid Ali, Student, Department of CSE (Galgotias University, Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. **About:** Sajid Ali is a dedicated student pursuing his education in the Department of Computer Science and Engineering at Galgotias University in Greater Noida, Uttar Pradesh, India. With a passion for technology, he continually explores new areas within the field of computer science, working to expand his knowledge and skills. Sajid is a hardworking individual with a keen eye for detail and a commitment to excellence. He is actively involved in various extracurricular activities, such as coding contests and hackathons, which allow him to apply his skills and gain practical experience. Sajid's enthusiasm for computer science and his drive to succeed make him a valuable member of the Galgotias University community.

Sameer Ray, Student, Department of CSE (Galgotias University, Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. **About:** Sameer Ray is a dedicated and enthusiastic student currently pursuing his studies in the Department of Computer Science and Engineering at Galgotias University, located in Greater Noida, Uttar Pradesh, India. With a passion for technology and a desire to innovate, Sameer has already begun to establish himself as a promising young talent within the field. He is recognised for his strong work ethic, commitment to learning, and ability to collaborate effectively within a team. With his drive and determination, Sameer is sure to achieve great things both academically and professionally in the years to come.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.