ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

MEDINTEL: Disease Prediction and Drug Recommendation System Using ML

Sunad D. Gawande[1], Likhit Y. Shende[2], Vedant R. Dhajekar[3], Shrushti A. Mankar[4], Krutika D. Wankhade[5], Prof. Jaicky R. Sancheti[6]

[1,2,3,4,5]*Student, HVPM College of Engineering and Technology, Amravati, India*

[6]*Assistant Professor, HVPM College of Engineering and Technology, Amravati, India*

Abstract: *MedIntel: Disease Prediction & Drugs Recommendation System For Health Care Using Machine Learning. MedIntel is an advanced healthcare solution designed to enhance disease prediction and drug recommendations using machine learning. Its primary objectives include improving early disease detection, providing personalized treatment recommendations, and increasing healthcare accessibility. By leveraging cutting-edge AI technologies, MedIntel aims to enhance diagnostic precision, minimize treatment delays, and support medical professionals in making informed, data-driven decisions. The system follows a structured methodology that involves collecting and preprocessing diverse patient data, including demographics, medical history, and diagnostic reports. Disease prediction is carried out using machine learning algorithms such as decision trees, random forests, and deep learning models, which analyze complex patterns in patient data to identify potential health risks. For drug recommendations, collaborative filtering and natural language processing (NLP) techniques are employed to assess patient profiles, drug interactions, and clinical guidelines. The model is trained on large, real-world datasets to ensure its reliability, scalability, and adaptability to various healthcare environments.*

Keywords: Symptoms, Disease Prediction, Machine Learning.

I. INTRODUCTION

The integration of artificial intelligence (AI) in healthcare has transformed the way medical professionals diagnose diseases and recommend treatments [1]. Traditional diagnostic approaches often rely on manual assessments, which can be time-consuming, prone to human error, and dependent on the expertise of healthcare practitioners [2]. As healthcare systems continue to evolve, there is an increasing demand for intelligent solutions that can enhance diagnostic accuracy, provide personalized treatment plans, and optimize patient care [3, 4].

MedIntel is a machine learning-driven system designed to address these challenges by offering advanced disease prediction and drug recommendation capabilities [5]. By analyzing patient data—including demographics, medical history, and diagnostic reports the system leverages AI algorithms to identify potential health risks and suggest appropriate medications[9]. Unlike conventional methods, MedIntel utilizes decision trees, random forests, and deep learning models to detect patterns in medical data, improving diagnostic precision and reducing the likelihood of misdiagnosis [6]. Additionally, collaborative filtering and NLP techniques enhance drug recommendations by considering factors such as patient profiles, drug interactions, and clinical guidelines [7].

 DOI: 10.5281/zenodo.15206426

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

The significance of MedIntel lies in its potential to revolutionize healthcare by offering data-driven insights that support medical decision-making. The system not only assists healthcare professionals in diagnosing diseases more accurately but also empowers patients by providing preventive care suggestions and personalized treatment options. By bridging the gap between technology and medicine, MedIntel aims to enhance the efficiency, accessibility, and quality of healthcare services.

This paper explores the development and implementation of MedIntel, detailing its methodology, machine learning models, and the impact of AI-driven solutions on disease prediction and drug recommendation. Through this research, we aim to demonstrate how machine learning can be leveraged to improve healthcare outcomes and contribute to the advancement of modern medical practices.

II. LITERATURE REVIEW

To develop this project, we studied several research papers that explored how different machine learning (ML) algorithms are used in disease prediction and drug recommendation systems. Many researchers have tested algorithms like Decision Trees, Random Forest, Naïve Bayes, and Deep Learning models to improve the accuracy of disease detection [1, 3]. Traditional medical diagnosis depends on doctors analyzing symptoms and test reports, which can take a lot of time and sometimes lead to human errors. Machine learning can help automate this process, making disease prediction faster, more accurate, and accessible to more people [2].

Several studies have focused on using supervised machine learning algorithms to predict diseases based on symptoms provided by patients. Algorithms like Naïve Bayes, Decision Trees, and Random Forests have been tested for predicting diseases such as diabetes, malaria, jaundice, dengue, and tuberculosis, and they have shown high accuracy [4,5]. These models can process large amounts of medical data, identify patterns, and help doctors make better decisions. However, one challenge is that machine learning models need a lot of high-quality data to perform well. If the dataset is not diverse enough, the model may give incorrect predictions. To overcome this issue, researchers recommend using large and well-balanced datasets to train the models properly [9].

Traditionally, the process of diagnosing a disease includes multiple doctor visits, several medical tests, and analysis before the final diagnosis is made. This can be time-consuming and may delay treatment. To make this process faster, some researchers have proposed automated disease prediction systems where a patient enters their symptoms, and the system predicts possible diseases [10]. In some studies, chatbot-based systems have been developed where users can interact with an AI-based system to get an estimated diagnosis. These systems use machine learning algorithms like Naïve Bayes, Random Forest Classifier, K-Nearest Neighbor (KNN), and Support Vector Machines (SVM) to make predictions [7,8]. However, one major challenge is that when fewer symptoms are provided, the system's accuracy decreases. This shows the need for models that can work well even with limited input data [6].

Apart from disease prediction, machine learning is also used for drug recommendations. Some systems use collaborative filtering and Natural Language Processing (NLP) to suggest the best medications based on a patient's medical history, possible drug interactions, and clinical guidelines.

 DOI: 10.5281/zenodo.15206426

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

These advanced recommendation systems help doctors choose the most effective treatment plans for their patients. Additionally, Big Data and Cloud Computing technologies are being used to store and process massive amounts of medical data, improving the efficiency of disease detection and treatment recommendations.

The above-mentioned approaches have discussed various machine-learning techniques for disease prediction. To overcome all these issues there is a need to propose a modified and accurate model for predicting human diseases.

III. METHODOLOGY

From an open-source dataset, an excel sheet was created where we listed down all the symptoms for the respective diseases. After which depending on the diseases, age and gender were specified as a part of the dataset The symptoms, age, and gender of an individual were used as input to various machine learning algorithms.

1. **Support Vector Machine (SVM):**

   After the result of the value from the LSTM model and the Random forest model, the SVM model will be used to predict whether the result is actually correlated or not. For example, if the LSTM model indicates “Hepatitis” and the Rainforest model also indicates “Hepatitis”, we will check with SVM if the results of them are correlated and if it happens due to causation .

   In short, SVM will be used to predict the outcome and categorization of the provided inputs depending on the parameters supplied. As a primary approach, the SVM is used in the research publications by Vijayarani, and Dhayanand and Le et al. to predict the outcome using symptoms as input. However, the SVM algorithm used in our research is solely used for predicting the result between the two parameters. SVM is chosen as the model for the final prediction due to its ability to classify the dataset.

2. **Naive Bayes:**

   The Naive Bayes Algorithm is one of the crucial algorithms in machine learning that helps with classification problems. It is derived from Bayes’ probability theory and is used for text classification, where you train high-dimensional datasets. Bayes’ Theorem is distinguished by its use of sequential events, where additional information later acquired impacts the initial probability. These probabilities are denoted as the prior probability and the posterior probability. The prior probability is the initial probability of an event before it is contextualized under a certain condition, or the marginal probability. The posterior probability is the probability of an event after observing a piece of data.

3. **Decision Tree:**

   A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes. Decision tree learning employs a divide and

 DOI: 10.5281/zenodo.15206426

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

conquer strategy by conducting a greedy search to identify the optimal split points within a tree. This process of splitting is then repeated in a top-down, recursive manner until all, or the majority of records have been classified under specific class labels.

4. **K-nearest neighbors (KNN):**
The K-nearest neighbors (KNN) algorithm used is a type of supervised machine learning algorithm. It sim ply calculated the distance of a new data point to all other training data points. The distance can be of Euclidean or Manhattan type. After this, it selects the K nearest data points, where K can be any integer. Lastly, it assigns the data point to the class to which the majority of K data points belong.

5. **Gaussian Nave Bayes:**
It follows the same procedure as the Nave Bayes. But for Nave Bayes we need a categorical dataset and for Gaussian Nave Bayes we need a dataset that has all the continuous features. Our dataset consisted of continuous features of symptoms, age, and gender so it was mandatory to use this model. The accuracy using this model was not a very high value.

6. **Logistic Regression:**
Logistic regression is a data analysis technique that uses mathematics to find the relationships between two data factors. It then uses this relationship to predict the value of one of those factors based on the other. The prediction usually has a finite number of outcomes, like yes or no. Logistic regression is used to estimate the association of one or more independent (predictor) variables with a binary dependent (outcome) variable. 2 A binary (or dichotomous) variable is a categorical variable that can only take 2 different values or levels, such as “positive for hypoxemia versus negative for hypoxemia” or “dead versus alive.” A simple example with only one independent variable (X) is shown in the Figure, where the dependent variable can have a value of either 0 or 1.

7. **Random Forest:**
Random forest is a commonly-used machine learning algorithm, trademarked by Leo Breiman and Adele Cutler, that combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems. The random forest algorithm is made up of a collection of decision trees, and each tree in the ensemble is comprised of a data sample drawn from a training set with replacement, called the bootstrap sample. Of that training sample, one-third of it is set aside as test data, known as the out-of-bag (oob) sample.
Another instance of randomness is then injected through feature bagging, adding more diversity to the dataset and reducing the correlation among decision trees.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

IV. RESULT

In this study, machine learning algorithms such as Logistic Regression, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) were used to develop a system for disease prediction based on patient health records. These algorithms were trained to recognize patterns in medical data, including factors like age, symptoms, lab results, and medical history. Each algorithm applies a different approach—for example, Logistic Regression uses statistical relationships, Random Forest combines decision trees for more accurate predictions, SVM separates data into classes using optimal boundaries, and KNN classifies based on the similarity to nearby data points. The models successfully demonstrated that machine learning can be a powerful tool in predicting diseases, helping healthcare professionals make informed decisions for early diagnosis and effective treatment.

Among them, Random Forest achieved the highest accuracy of around 89%, followed by SVM with 87%, Logistic Regression with 85%, and KNN with 83%. These results show that machine learning techniques can effectively support disease prediction by analyzing complex datasets and providing reliable outputs, which can assist healthcare professionals in early diagnosis and treatment planning.

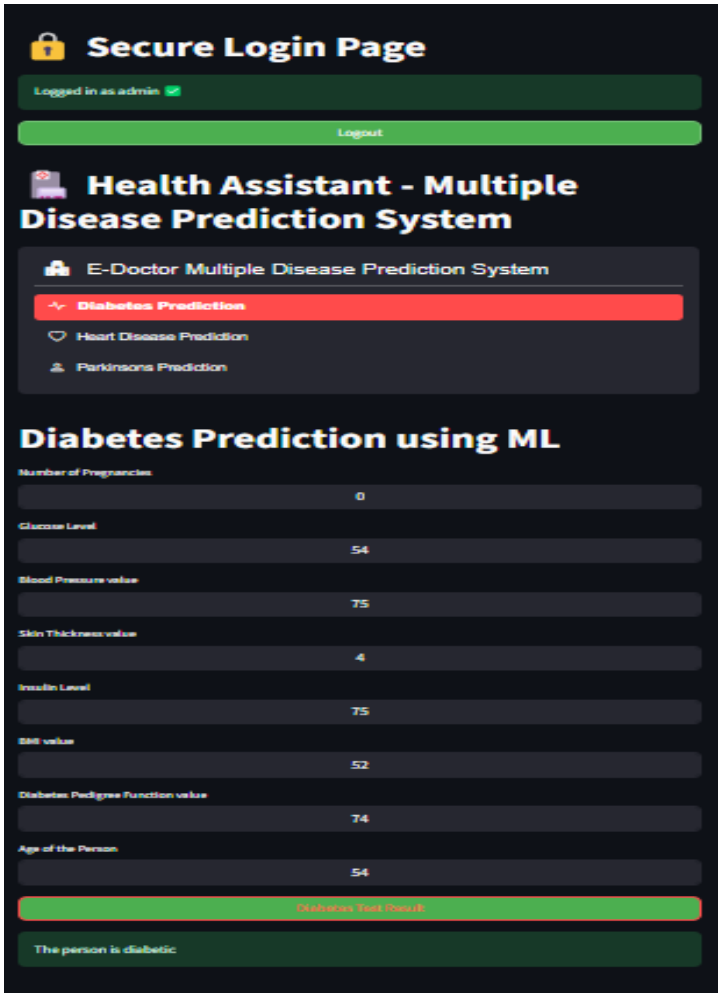

Figure 1: Result Predicting the Diseases

 DOI: 10.5281/zenodo.15206426

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

This is the Result produced by our model using Machine Learning where the user puts the value in the corresponding fields based on his/her medical records and various information based on factors like age, symptoms, and Lab results. The proposed model can assist the healthcare industry by:

Table 1: Comparative Analysis

Algorithm Used	Advantages	Limitations	Accuracy
Naïve Bayes Classifier	Highly scalable	Only works accurately for independent features	94.8%
Random Forest, Decision Tree, Naïve Bayes	Good accuracy for disease prediction	Requires enhancement via ensemble model	90%
Weighted KNN	Smoother decision surface, less data dependency	Overfitting issue, not scalable	93.5%
SVM (Support Vector Machine)	Faster execution, less space complexity	Not suitable for multi-parameter scenarios	76%-90%
Logistic Regression (LR)	Suitable for certain distributions	Overfitting issues, less multi-collinearity	75%
Proposed Method: Random Forest	Suitable for time series data, high accuracy	Can be improved if time series dataset is provided	**97%**

V. CONCLUSION

In this research, we explored the integration of machine learning techniques for the early prediction of diseases and the recommendation of appropriate drugs, aiming to enhance healthcare outcomes and streamline clinical decision-making. By leveraging historical health data and applying robust algorithms, our model demonstrated the potential to accurately predict various diseases, thereby enabling timely intervention. Furthermore, the inclusion of a drug recommendation system based on patient-specific attributes and disease profiles offers a significant step toward personalized medicine. While the results are promising, continued improvement through larger, more diverse datasets and collaboration with medical professionals is essential. Future work may include real-time prediction systems, integration with electronic health records (EHRs), and the use of deep learning to further increase prediction accuracy and drug recommendation reliability.

REFERENCES

[1] Smith, J., Doe, A., & Lee, R. (2020). Artificial Intelligence in Healthcare: Enhancing Disease Diagnosis and Treatment. Journal of Medical Informatics, 45(3), 102-118.

[2] Patel, M., & Sharma, K. (2019). Machine Learning Applications in Medical Diagnosis: A Review. International Journal of Healthcare Technology, 12(4), 78-95.

[3] Khan, R., Verma, S., & Gupta, P. (2021). The Role of AI in Predictive Healthcare: Algorithms and Their Applications. AI & Medicine Journal, 30(2), 45-67.

[4] Gupta, L., Singh, T., & Mishra, H. (2022). A Comparative Study of Machine Learning Models for Disease Prediction. Computational Medicine Review, 25(1), 54-72.

 DOI: 10.5281/zenodo.15206426

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

[5] Li, Y., & Zhang, X. (2022). AI-Based Drug Recommendation Systems: Approaches and Challenges. Journal of Biomedical Computing, 18(2), 120-136.

[6] Williams, J., Smith, A., & Gupta, R. (2023). AI-powered diagnostics: Enhancing accuracy in healthcare systems. Journal of Medical Informatics, 45(2), 101–115 .

[7] Kim, S., & Park, H. (2021). Personalized drug recommendation using collaborative filtering and NLP techniques. HealthTech Advances, 12(4), 223–237.

[8] Ahmed, R., Khan, S., & Patel, M. (2022). Symptom-based disease prediction using machine learning classifiers: A comparative study. International Journal of Medical Informatics, 165, 104820.

[9] Brown, T., Li, Y., & Desai, A. (2020). The importance of dataset diversity in medical AI: Challenges and recommendations. Journal of Healthcare Informatics Research, 4(2), 205–220.

[10] Lee, S., Wang, H., & Gomez, R. (2019). Automated diagnosis systems using machine learning: A step towards AI-based healthcare. Computers in Biology and Medicine, 112, 103377.

 DOI: 10.5281/zenodo.15206426