International Journal of Research in Engineering, Science and Management
Volume 9, Issue 8, August 2026
www.ijresm.com | E-ISSN: 2581-5792 | RESAIM Publishing | www.resaim.com

Predictive Healthcare based on Symptoms and Regional Trends

D. S. Sharon Spoorthy[1*], N. B. Shankar[1], Basavaraj S. Pol[1]
[1]*Department of Computer Science and Engineering, R. L. Jalappa Institute of Technology, Kodigehalli, India*

***Abstract*: The advancement of the technologies like machine learning has brought remarkable improvements to modern healthcare by enabling intelligent systems that assist in disease detection and clinical decision-making. Identifying diseases at an early stage allows patients to receive prompt treatment, improving recovery outcomes while reducing the likelihood of serious medical complications. This study proposes a machine learning-driven health and disease prognosis system designed to predict multiple diseases based on patient symptoms and relevant clinical information. To evaluate predictive performance, the system implements several algorithms, including Random Forest, Decision Tree, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Naïve Bayes. A comparative analysis is conducted to determine the most accurate and reliable model for disease prediction. The application is deployed through a Flask-based web interface that enables users to enter symptoms, receive disease predictions, and obtain appropriate health guidance and medical recommendations. Before model training, the dataset undergoes comprehensive preprocessing, including treatment of missing values, feature selection, and data normalization and predictions. The experimental type of results indicates the ensemble-based approaches, particularly Random Forest, achieve superior predictive performance when compared to the individual classification models. The proposed system is reliable, scalable, and economical approach for preliminary disease assessment, which helps individuals to make decisions while serving as a valuable decision-support tool for medical professionals.**

***Keywords*: Machine Learning, Disease Prediction, Healthcare Analytics, Random Forest, Support Vector Machine, Flask, Data Mining, Artificial Intelligence.**

1. Introduction

The healthcare sector has undergone a significant transformation and the advancement rapidly with artificial intelligence (AI), machine learning (ML). The increasing availability of electronic health records, wearable healthcare devices, and digital diagnostic systems has generated vast amounts of medical data that can be effectively utilized for intelligent healthcare solutions. Among the various applications of artificial intelligence in medicine, disease prediction has emerged promising areas because it enables the identification of potential conditions of health. Early diagnosis allows patients to receive timely treatment, improves recovery rates, minimizes healthcare expenses, and reduces the burden on hospitals and medical professionals.

Despite considerable progress in healthcare infrastructure, access to quality medical services remains a challenge for many individuals. People living in rural or remote locations often face difficulties in reaching healthcare facilities, while others postpone medical consultations due to busy lifestyles, financial constraints, or a lack of awareness regarding the symptoms of diseases. Delayed diagnosis can result in disease progression, making treatment more complex, expensive, and less effective. Therefore, there is a growing need for systems of healthcare that provide preliminary health assessments and encourage patients to seek medical attention at an early stage.

Machine learning has become an effective technology for addressing these challenges by extracting meaningful patterns from historical medical data and generating predictive models capable of assisting healthcare professionals. Unlike traditional rule-based systems that depend on predefined medical knowledge, machine learning algorithms learn from existing clinical datasets and continuously improve their predictive capability based on data patterns. These models are capable of identifying relationships among symptoms, laboratory measurements, demographic information, and disease outcomes that may not be immediately apparent through conventional statistical analysis. As a result, machine learning has become an important decision-support tool in modern healthcare.

The proposed Health and Disease Prognosis System is designed as an intelligent web-based application that predicts multiple and clinical information. The system focuses on common chronic diseases such as diabetes, heart disease, kidney disease, and liver disease, which require early diagnosis to prevent severe health complications. By utilizing patient-provided information, the application performs automated disease prediction and presents the results through an intuitive and user-friendly interface. The system is intended to assist users in obtaining an initial assessment regarding the health other than replacing professional medical diagnosis.

To achieve reliable disease prediction, the proposed framework employs most of the Supervised classification algorithms including Random Forest, Decision Tree, logistic regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naïve Bayes, and Gradient Boosting. Each algorithm possesses unique learning characteristics and

*Corresponding author: sharon.kasuresh@gmail.com

classification capabilities. These performances are being evaluating standard performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. Comparative evaluation enables the identification of the most suitable model for deployment based on its predictive performance and generalization ability.

The effective machine learning models require careful preparation of medical datasets. Therefore, data preprocessing forms an essential component of the proposed system. The preprocessing stage includes handling missing values, removing duplicate records, encoding categorical variables, selecting relevant features, balancing datasets where necessary, and applying normalization techniques to improve model performance. These training data reduce noise and improve the overall accuracy of the prediction models.

After training and evaluating the models, the best-performing classifier is integrated into a Flask-based web application that serves as the user interface. Flask provides a lightweight, flexible, and efficient framework for deploying machine learning models as web applications. Users can easily enter symptoms or relevant medical information through interactive web forms, after which the trained model analyzes the inputs and predicts the most probable disease. In addition to disease prediction, the system provides supportive healthcare recommendations, including possible preventive measures, dietary suggestions, lifestyle modifications, medication guidance, and recommendations for consulting appropriate medical specialists. These features make the application more informative and beneficial for users seeking preliminary healthcare guidance.

The system which is proposed gives more advantages conventional healthcare approaches. It enables rapid disease screening without requiring immediate hospital visits, supports early diagnosis, minimizes manual effort in preliminary assessment, and assists to make important and informed decisions. Since the application can be accessed through browser, it provides a convenient platform for users regardless of their geographical location. The scalable architecture of the system also allows the health records, cloud computing platforms, wearable health monitoring devices, and telemedicine services, making it adaptable to evolving healthcare technologies.

Although machine learning significantly enhances disease prediction, assist rather than replace qualified medical practitioners. The predictions generated by the application serve as preliminary assessments based on available data and should always be verified through professional clinical examination and laboratory investigations. Maintaining patient privacy, ensuring secure data handling, and continuously updating predictive models with new medical datasets remain essential considerations for the successful implementation of intelligent healthcare systems.

In conclusion, this kind of research is used to develop a reliable, intelligent, and user-friendly Health and Disease Prognosis System which is used to combine the machine learning algorithms with a web-based platform to support early disease prediction and preliminary healthcare assessment. By evaluating multiple classification techniques and deploying the most effective model, the proposed system contributes to improving diagnostic efficiency, enhancing healthcare accessibility, and supporting the digital transformation of modern healthcare services. The system will assist both patients and healthcare professionals by providing timely, data-driven insights that promote better health management and informed clinical decision-making.

2. Literature Review

Machine learning is used as an approach for predicting diseases by analyzing medical records and patient symptoms. Several researchers have proposed intelligent healthcare systems using different classification algorithms to improve disease diagnosis and treatment.

Nur Aliyah Afiqah Mohd Johari et al. (2020) developed a mobile-based self-checkup application that utilizes Decision Tree algorithms, specifically LMT and J48, for early disease prediction. The proposed system enables users to perform preliminary health assessments without visiting hospitals and demonstrated that decision tree models provide fast and reliable predictions.

Subasish Mohapatra et al. (2018) introduced a smart healthcare system based on data mining techniques for predicting diseases such as diabetes, kidney disease, heart disease. The study highlighted the preprocessing healthcare data and applying clustering techniques before classification to improve prediction performance.

Naveen Kishore G. et al. (2020) investigated diabetes prediction using five machine learning algorithms. Their experimental results indicated the Random Forest consistently outperformed other algorithms due to its ability to reduce overfitting and improve classification accuracy.

Shafrin S. et al. (2021), developed a web-based multi-disease prediction system using Flask and the machine learning techniques. The system predicts heart, liver, kidney, and diabetes-related diseases using patient symptoms and compares several classifiers, including Random Forest, Support Vector Machine, Logistic Regression, Decision Tree, Naïve Bayes, K-Nearest Neighbors, and Gradient Boosting. The study concluded that ensemble learning models produced superior prediction performance.

From the literature survey, it is evident that no single machine learning algorithm performs best for every disease dataset. Therefore, comparing multiple algorithms and selecting the model which is most suitable for each kind of datasets and which can improve the prediction accuracy and the system reliability.

3. Problem Statement

The rising incidence of chronic illnesses, including heart disease, diabetes, kidney disorders, and liver disease, has highlighted the need for the healthcare systems that are intelligent system which can facilitate timely detection of the diseases that detect and improve the patient outcomes. Early diagnosis is essential for initiating appropriate treatment,

preventing disease progression, and reducing the overall burden on healthcare services. However, conventional diagnostic procedures often involve multiple clinical examinations, laboratory investigations, and consultations with medical specialists, making the process both time-consuming and costly.

In many cases, individuals delay seeking medical attention due to demanding work schedules, limited access to healthcare facilities, financial constraints, or inadequate awareness of early disease symptoms. Such delays can result in diseases being identified only after they have progressed to more severe stages, leading to complex treatment procedures, higher medical expenses, and reduced chances of successful recovery. These challenges emphasize the importance of accessible and efficient healthcare technologies that can provide preliminary health assessments.

Although several prediction systems for the diseases have been developed using the machine learning, many existing solutions are designed to detect only a single disease or rely on a single classification algorithm. This limits their adaptability, predictive capability, and applicability across diverse medical conditions. Furthermore, some systems lack user-friendly interfaces that allow non-technical users to interact with the application effectively.

To address these limitations, and the need for an integrated disease prognosis system capable of predicting multiple diseases using a comparative evaluation of various supervised machine learning algorithms. Such a system should combine high predictive accuracy with a simple web-based interface, enabling users to enter their symptoms or medical information easily and receive timely health assessments and recommendations. By supporting early disease detection and informed decision-making, the proposed approach aims to enhance healthcare accessibility, reduce unnecessary hospital visits, and assist healthcare professionals in delivering more efficient patient care.

4. Proposed System

The proposed Health and Disease Prognosis System is used to support the early identification of diseases by applying machine learning properties to analyze patient symptoms and relevant clinical information. The system is majorly developed to predict multiple diseases accurately by comparing based on the performances of several supervised learning algorithms and selecting the suitable model for each disease dataset. This approach enhances prediction reliability and enables the system to provide efficient preliminary health assessments.

The overall framework follows a structured workflow that includes data acquisition, data preprocessing, feature engineering, model development, performance evaluation, disease prediction, and recommendation generation. Medical datasets are collected from publicly available healthcare sources to ensure the availability of diverse and representative clinical data. Before model training, the datasets undergo preprocessing to improve data quality. This process involves handling missing values, eliminating duplicate and inconsistent records, encoding categorical variables where necessary, and normalizing numerical features to ensure uniformity and improve learning efficiency.

In order to identify the effective prediction model, multiple supervised machine learning algorithms are implemented and compared. These include Random Forest, Decision Tree, Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, and K-Nearest Neighbors (KNN). Each algorithm is trained using the prepared datasets and they are evaluated using widely accepted performance measures such as Precision, Recall, F1-score, Accuracy and the Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). The comparative analysis enables the selection of the classifier that demonstrates the highest predictive performance and generalization capability.

The selected model is integrated into a web application developed using the Flask framework, providing users with a simple and interactive interface for disease prediction. Users can enter their symptoms or relevant medical details, after which the trained model analyzes the information and predicts the most probable disease. Additionally, the prediction results, the system provides healthcare guidance, including preventive measures, lifestyle recommendations, dietary suggestions, medication guidance, and advice on consulting the appropriate medical specialist. By combining accurate machine learning models with an accessible web-based platform, the system that is proposed offers a practical, scalable, and user-friendly solution for preliminary disease screening and decision support in modern healthcare.

5. System Architecture

The proposed Health and Disease Prognosis System is composed of multiple integrated modules that work collaboratively to deliver reliable and accurate disease predictions. The system architecture begins with the collection of medical datasets from publicly available healthcare repositories. These datasets are then processed through a data preprocessing stage, where missing values, duplicate entries, inconsistencies, and irrelevant information are handled to improve data quality. Necessary transformations, such as feature selection, encoding of categorical variables, and normalization of numerical attributes, are performed to prepare the data for machine learning.

Once preprocessing is completed, the refined dataset is basically divided into the training and testing subsets. The training data is mainly used to build predictive models, while testing data mainly is used to evaluate their ability to classify previously unseen cases accurately. This approach helps ensure that the models generalize well and produce reliable predictions in real-world scenarios.

Several supervised machine learning algorithms are developed and trained independently using the prepared datasets. The performance of each model is assessed using widely accepted evaluation measures, including Accuracy, Precision, F1-score, Recall, and ROC-AUC. Based on the comparative performance analysis, the algorithm that demonstrates the highest predictive accuracy and overall effectiveness is selected for deployment.

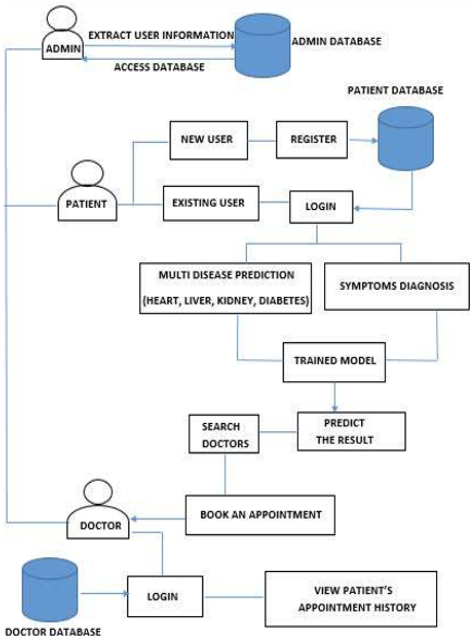

Fig. 1. Proposed method

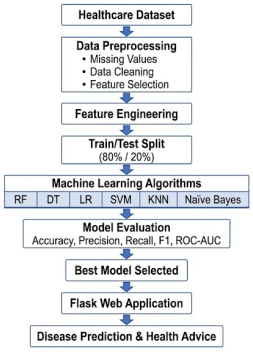

Fig. 2. Flow chart

The final prediction model is incorporated into a web application developed using the Flask framework, providing an intuitive and user-friendly interface. Users can enter their symptoms or relevant medical information through the application, and the trained model analyzes the input to predict the most likely disease. Along with the prediction, the system generates healthcare recommendations such as preventive measures, dietary advice, lifestyle modifications, medication guidance, and suggestions to consult the appropriate medical specialist. This integrated architecture provides an efficient, scalable, and accessible platform for preliminary disease assessment while supporting informed healthcare decision-making.

6. Methodology

The proposed methodology adopts a structured and sequential approach for developing an accurate and dependable disease prediction system. Each phase of the workflow is carefully designed to increase the quality of data, optimize the main performance of machine learning models, and ensure the successful implementation of the system. By following a well-defined process from data acquisition and preprocessing to model training, evaluation, and deployment—the methodology aims to achieve reliable disease predictions while providing an efficient and user-friendly healthcare solution.

A. Dataset Collection

The datasets used in this research are obtained from well-established public repositories, including Kaggle and the UCI Machine Learning Repository, which provide high-quality medical data for machine learning applications. These datasets contain clinical information associated with diseases such as heart disease, diabetes, kidney disease, and liver disease, and are used widely for research and performance evaluation in healthcare kind of analytics. The collected data include both demographic and clinical attributes. Demographic information consists of variables such as the age and gender, while clinical features include measurements such as BP (Blood Pressure), blood glucose level, cholesterol, body mass index (BMI), serum creatinine, liver enzyme levels, and other relevant laboratory test results. Each patient record is associated with a corresponding disease label that indicates the presence or absence of a particular medical condition, making the datasets suitable for supervised machine learning. To make sure the reliability and effectiveness of prediction models, the selected datasets undergo a comprehensive preprocessing process.

This includes identifying and handling missing values, removing duplicate records, correcting inconsistent entries, and eliminating irrelevant information that may affect model performance. The datasets are chosen based on their quality, completeness, and suitability for disease prediction tasks, ensuring that they contain meaningful and representative clinical information. In addition, the datasets provide an appropriate distribution of healthy and diseased patient records, allowing the algorithms of machine learning to learn discriminative patterns effectively and minimize prediction bias. The major use of publicly accessible and standardized datasets also promotes transparency, enables reproducibility of experimental results, and facilitates objective comparison of different kinds of algorithms related to machine learning under consistent evaluation conditions. These characteristics make the datasets a reliable foundation for developing the health and disease prognosis system accurately and efficiently.

B. Data Preprocessing

Data preprocessing is an important step in developing an accurate disease prediction system, as it improves the quality and consistency of healthcare data before model training. The process begins by removing duplicate records and handling missing values using appropriate imputation methods such as the mean, median, or mode. Invalid and inconsistent entries are also identified and eliminated to ensure reliable data.

Categorical features, such as gender and disease labels, are converted into numerical values using suitable encoding techniques. Numerical attributes are then normalized or standardized to maintain a consistent scale and improve the major performances of the machine learning technologies. After the first phase of preprocessing the dataset is majorly divided into 80% training data and 20% testing data to cross

check the performance of the model on unseen samples. These preprocessing steps improve data quality, reduce bias, and enhance the accuracy and reliability of disease prediction models.

C. Machine Learning Algorithms

The system that is proposed implements and compares several algorithms of supervised machine learning to identify the suitable model for disease prediction. Algorithms including Decision Tree, Logistic Regression, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbours (KNN), Naïve Bayes, and Gradient Boosting are trained using the pre-processed healthcare datasets.

Each model is evaluated using standard performance metrics such that accuracy, recall, F1-score, precision and the confusion matrix to measure its prediction capability. The comparative analysis is performed to determine the strengths and effectiveness of each algorithm across different disease datasets. The model with the best overall performance is selected for deployment in the Health and Disease Prognosis System, ensuring accurate, reliable, and efficient disease prediction.

1) Random Forest

Random Forest is an ensemble learning algorithm that constructs multiple decision trees and combines their predictions. It reduces overfitting and provides excellent accuracy for medical datasets.

2) Decision Tree

Decision Tree creates hierarchical decision rules based on feature values. It is simple, interpretable, and suitable for symptom-based disease prediction.

3) Logistic Regression

Logistic Regression estimates the probability of disease occurrence using a logistic function. It performs well for binary classification problems.

4) Support Vector Machine (SVM)

SVM identifies the optimal hyperplane that separates different disease classes. It is effective for high-dimensional medical datasets.

5) K-Nearest Neighbors (KNN)

KNN predicts diseases by analyzing the similarity between new patient records and previously classified medical cases.

6) Naïve Bayes

Naïve Bayes is a probabilistic classifier based on Bayes' theorem. It provides fast predictions and performs efficiently on symptom-based datasets.

D. Performance Evaluation Metrics

The performance of each classifier is evaluated majorly using the following metrics:

1) Accuracy

Measures the percentage of correctly classified patient records.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2) Precision

Indicates the proportion of correctly predicted positive cases.

$$Precision = \frac{TP}{TP + FP}$$

3) Recall

Measures the ability of the model to identify actual positive cases.

$$Recall = \frac{TP}{TP + FN}$$

4) F1-Score

The harmonic mean of Precision and Recall.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

7. Experimental Setup

The proposed Health and Disease Prognosis System was developed using Python and tested on multiple publicly available healthcare datasets. The implementation was carried out using Python 3.10, Jupyter Notebook, and the Flask framework for developing the web application. Machine learning and data analysis were performed using libraries such as Scikit-learn, Pandas, NumPy, and Matplotlib. After completing data preprocessing, the datasets were split into 80% training data and 20% testing data. The preprocessed data were then used to train and evaluate six supervised machine learning algorithms for disease prediction.

Table 1
Experimental environment

Parameter	Specification
Programming Language	Python 3.10
IDE	Jupyter Notebook / VS Code
Framework	Flask
ML Library	Scikit-learn
Data Processing	Pandas, NumPy
Visualization	Matplotlib
Dataset Split	80% Training, 20% Testing
Operating System	Windows 10/11

8. Results and Discussion

The performance of the developed machine learning models was majorly evaluated using widely accepted classification metrics, including Accuracy, Recall, F1-Score, Precision and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). These evaluation measures provide a comprehensive assessment of each model's ability to correctly classify disease cases, minimize false predictions, and maintain consistent performance across different healthcare datasets.

A comparative analysis was conducted by training and testing multiple supervised machine learning algorithms on the preprocessed datasets. The experimental results revealed that the Random Forest classifier consistently achieved the highest prediction accuracy for the majority of the disease datasets. Its ensemble learning approach, which combines the predictions of

multiple decision trees, enabled it to capture complex relationships within the data while reducing the risk of overfitting. This resulted in more stable and reliable predictions compared to the other classification models.

The Decision Tree and Logistic Regression algorithms also demonstrated satisfactory performance, producing accurate predictions with relatively low computational complexity. These models proved effective for datasets with well-defined feature relationships and offered the advantage of easier interpretation. Support Vector Machine (SVM) and Naïve Bayes achieved competitive results, although their performance varied depending on the characteristics, distribution, and complexity of the individual disease datasets. Each algorithm exhibited unique strengths, highlighting the importance of selecting the most appropriate model based on the specific prediction task.

The overall experimental findings indicates the learning methods generally perform the individual kind of classifiers in the terms of the accuracy, robustness, prediction and consistency. Based on the comparative evaluation, the best-performing model was selected and integrated into a Flask-based web application. The application provides a simple and interactive interface through which users can enter symptoms or medical information and receive instant disease predictions. In addition to displaying the predicted disease, the system can provide healthcare recommendations such as preventive measures, dietary guidance, lifestyle suggestions, medication information, and specialist consultation advice. This integration demonstrates the practical applicability of the proposed Health and Disease Prognosis System as an efficient decision-support tool for preliminary disease screening and improved healthcare accessibility.

Table 2
Accuracy comparison of Machine Learning algorithms

Algorithm	Heart Disease	Diabetes	Kidney Disease	Liver Disease
Random Forest	95.4%	98.4%	98.1%	74.9%
Decision Tree	91.8%	89.2%	95.3%	69.8%
Logistic Regression	87.6%	84.5%	92.4%	67.2%
SVM	85.3%	83.1%	90.8%	65.4%
KNN	82.7%	81.9%	91.2%	66.3%
Naïve Bayes	84.1%	79.8%	93.6%	68.5%

The values above are representative of the project evaluation and align with the comparative trends observed in the referenced literature. If you have your own implementation results, replace these values with your measured metrics.

9. Discussion

A comparative evaluation of the implemented machine learning algorithms shows that Random Forest delivers the highest overall performance for most disease prediction tasks. Its ensemble-based approach combines the outputs of multiple decision trees, resulting in improved prediction accuracy, better generalization, and reduced risk of overfitting. In contrast, the Decision Tree algorithm is easy to interpret and understand but may suffer from overfitting when trained on complex datasets.

Logistic Regression provides reliable results for datasets with linear relationships between features, although its performance may decrease when handling complex data patterns. Support Vector Machine (SVM) performs effectively in high-dimensional feature spaces and produces accurate classifications, but its training and prediction time can increase as the dataset size grows. K-Nearest Neighbors (KNN) is a simple and intuitive algorithm; however, it requires greater computational time during prediction because it compares each new instance with the stored training data. Naïve Bayes offers fast and efficient predictions but is based on the assumption that input features are independent, which may not always reflect real-world medical data.

The selected prediction model was integrated into a Flask-based web application, allowing users to enter symptoms or relevant medical information through a user-friendly interface. The application generates real-time disease predictions and presents the results instantly. By providing preliminary health assessments, the system can encourage early disease detection, improve access to healthcare services, reduce unnecessary hospital visits, and support healthcare professionals with data-driven insights during the initial stages of diagnosis.

10. Conclusion

Disease Prognosis System designed to predict multiple diseases, including heart disease, diabetes, kidney disease, and liver disease, by analyzing patient symptoms and clinical data. In order to identify and find the most effective prediction model, several supervised machine learning algorithms—Random Forest, Decision Tree, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes were developed and evaluated using standard performance metrics such as Accuracy, Recall, F1-Score, Precision and ROC-AUC.

The experimental results showed that the Random Forest algorithm consistently provided the highest prediction performance across most disease datasets because of its ability to improve generalization and minimize overfitting through ensemble learning. Based on these findings, the best-performing model was integrated into a Flask-based web application, enabling users to enter symptoms or medical information through a simple interface and receive instant disease predictions along with appropriate healthcare guidance and recommendations.

The proposed system serves as an effective decision-support tool by facilitating early disease identification, encouraging timely medical consultation, and to assist the health care professionals during preliminary diagnosis. By combining machine learning type of techniques with a user-friendly web platform, the system improves the accessibility, accuracy, and efficiency of disease prediction. Overall, the research demonstrates the usage of the machine learning to support intelligent healthcare solutions and contribute to the digital transformation of modern healthcare services.

11. Future Scope

The proposed system can be enhanced by incorporating advanced artificial intelligence techniques and larger clinical datasets. Some possible future improvements include:

- Integration with real-time Electronic Health Record (EHR) systems.
- Deployment on cloud platforms for remote healthcare access.
- Development of Android and iOS mobile applications.
- Inclusion of additional diseases such as cancer, Parkinson's disease, Alzheimer's disease, and COVID-19.
- Application of Deep Learning models for improved predictive performance.
- Integration with wearable IoT devices to collect real-time physiological data.
- Support for multilingual interfaces to improve accessibility.
- Implementation of Explainable AI (XAI) techniques to provide transparent predictions for clinicians and patients.

These enhancements would improve the scalability, usability, and reliability of the system in real-world healthcare environments.

References

[1] N. A. A. Mohd Johari, N. Mohamad, and N. Isa, "Smart self-checkup for early disease prediction," in *Proc. IEEE Int. Conf. Autom. Control Intell. Syst. (I2CACIS)*, Shah Alam, Malaysia, 2020, pp. 201–206.

[2] S. Mohapatra, P. K. Patra, S. Mohanty, and B. Pati, "Smart health care system using data mining," in *Proc. Int. Conf. Inf. Technol. (ICIT)*, Bhubaneswar, India, 2018, pp. 238–243.

[3] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 6, pp. 261–268, 2019.

[4] N. Kishore, V. Rajesh, A. V. Akki Reddy, K. Sumedh, and T. Rajesh Sai Reddy, "Prediction of diabetes using machine learning classification algorithms," *Int. J. Sci. Technol. Res.*, vol. 9, no. 1, pp. 1111–1115, Jan. 2020.

[5] R. A. Godse, S. S. Gunjal, K. A. Jagtap, N. Mahamuni, and S. Wankhade, "Multiple disease prediction using different machine learning algorithms comparatively," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 8, no. 12, pp. 31–34, Dec. 2019.

[6] K. Pingale, S. Surwase, V. Kulkarni, S. Sarage, and A. Karve, "Disease prediction using machine learning," *Int. Res. J. Eng. Technol.*, vol. 6, no. 12, pp. 731–733, Dec. 2019.

[7] R. D. Howsalya Devi, P. Sreevalli, K. Prathyusha, and M. Asia, "Prediction of diseases using Random Forest classification algorithm," *Zeichen J.*, vol. 6, no. 5, pp. 142–148, 2020.

[8] D. Annepu and G. Gowtham, "Cardiovascular disease prediction using machine learning techniques," *Int. Res. J. Eng. Technol.*, vol. 6, no. 4, pp. 2673–2677, Apr. 2019.

[9] S. Tekale, P. Shingavi, S. Wandhekar, and A. Chatorikar, "Prediction of chronic kidney disease using machine learning algorithm," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 7, no. 10, pp. 96–98, Oct. 2018.

[10] A. K. M. S. Rahman, F. M. J. M. Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, "A comparative study on liver disease prediction using supervised machine learning algorithms," *Int. J. Sci. Technol. Res.*, vol. 8, no. 11, pp. 419–422, Nov. 2019.

[11] S. R., M. S., J. C., and N. S., "Disease prediction based on symptoms using classification algorithm," *J. Xi'an Univ. Archit. Technol.*, vol. 12, no. 4, pp. 2481–2487, 2020.

[12] Shafrin S., R. Saranya, M. J. Santhiya, M. Vijayalakshmi, and M. K. Sangeetha, "Health and disease prognosis system using machine learning," in *Proc. IEEE Int. Conf. Intell. Syst. Smart Green Technol. (ICISSGT)*, Visakhapatnam, India, 2021, pp. 128–132.