MULTIPLE DISEASE PREDICTION SYSTEM USING MACHINE LEARNING

Harshvardhan Gupta[1] , Mayank Singhal[1] , Vaishnav Yadav[1] , Vaibhav Panjiyar[1]

Anjali Chauhan[1]

Computer Science And Engineering (AI/AIML), KIET Groups of Institutions, Ghaziabad, Uttar Pradesh, India

Abstract **—** The urgent rise in chronic disease incidences demands state-of-the-art medical instruments for the timely detection and treatment of such diseases. This investigation presents a Multiple Diseases Prediction System (MDPS), which implements machine learning (ML) algorithms for predicting diabetes, cardiac disease, and Parkinson's disease. The system transmits important health parameters into the logistic regression and support vector machine models: blood pressure, pulse rate, blood cholesterol, or heart rate. An elegant online application designed using the Streamlit framework will be accessible to patients and doctors. Different from other applications that focus on single disease predictions, MDPS stands out by taking into account and assessing the risk of multiple diseases all at once. The comfort of the models is again increased by the preprocessing methods, which consist of normalization and feature selection. Risk factors, information, and disease prediction are fed back to the health professionals for a more informed decision. AI in healthcare uses large-scale data resources to aid in making evidence-based decisions, thereby enhancing patient outcomes and eventually relieving the burden of chronic diseases. Further enhancement will embrace deep learning, real-time patient monitoring through IoT-enabled devices, and cloud deployment for scalability. Bridging AI technology to healthcare issues should render MDPS more customized, effective, and accessible.

Keywords – Machine Learning, Multiple Disease Prediction System (MDPS), Logistic Regression, Heart Disease, Disease Prediction, Diabetes Prediction, Support Vector Machine (SVM), Parkinson's Disease Detection.

I. INTRODUCTION

Due to the rapid advances in machine learning (ML), almost every other industry has transformed by keeping pace with this trend. The most prominent transformation has been in the field of medicine. Early and accurate prognosis of a disease is vital to improving patient outcomes and reducing mortality rates. By contrast, repeated tests and consultations are usually necessary under classical diagnostic methods, delaying timely intervention. Intelligent systems that are capable of making correct predictions from clinical and lifestyle data are also desired because of the prevalence of chronic diseases like diabetes, Parkinson's, and heart disease.

This paper presents a Multiple Disease Prediction System (MDPS) which, through machine learning algorithms, predicts diabetes, heart diseases, and Parkinson's disease risk. The current models, designed primarily for single-disease prediction, restrict the analysis of multiple diseases. The proposed solution fills this gap and employs many machine learning techniques for extensive predictive analysis, thereby rendering great help to the patient and the healthcare provider in decision-making.

MDPS leverages state-of-the-art ML algorithms, such as logistic regression and support vector machine (SVM), to assess essential health parameters, including vital signs, blood pressure, cholesterol, and heart rate. The system was created using the Streamlit framework, building a web-based application that ensures usability and easy access.

Excellent preprocessing techniques, such as feature selection and normalization, increase the robustness and efficiency of the model and reduce misinterpretation of inconsistent data.

One of the main advantages of MDPS is its ability to evaluate different kinds of datasets and the patterns across different diseases; hence, this may provide more extensive knowledge of where diseases interrelate. It generates a risk assessment report from risk evaluation against the patient's data from several sources to inform the users of possible health concerns.

Besides facilitating predictive healthcare and early detection of diseases, this also relieves long-term strains on healthcare infrastructures. The proposed technique bridges the divide between machine-learning technology and real-world healthcare applications while accommodating the increasing demand for AI-based diagnostic solutions. MDPS will use predictive analytics to mitigate health risks and promote rapid medical intervention. Further improvements may include introducing deep learning techniques, real-time monitoring via Internet of Things devices, and cloud-based implementation for accessibility purposes.

Finally, the Multiple Disease Prediction System becomes a realistic approach in disease diagnosis at the levels of patients and medical professionals by providing reliable, data-supported insights. Through machine learning techniques, this system envisions the advancement of innovative healthcare, thus ultimately paving the way toward a proactive and efficient medical ecosystem.

II. LITERATURE REVIEW

2.1 Diabetes Prediction

Diabetes is a chronic condition affecting many millions of people worldwide, with early detection being a prerequisite for treatment. The PIMA Indian Diabetes Dataset (PIDD) has been used by researchers in developing machine learning approaches for diabetes prediction. [11]

Commonly applied for binary classification, decision trees, and logistic regression constitute the base of the methods used in diabetes prediction models. Another form, support vector machines (SVMs), together with Random Forest classifiers, is shown to be much more effective in management and works more easily with complicated data.[2]

Neural networks and other deep-learning techniques improve effectiveness. Deep learning further predicts effective incorporation of the complex patterns in health data. Hence, by consulting these machine learning tools, healthcare providers can evaluate risk variables for early intervention and personal treatment regimens.

Patient conditions improve through cutting-edge algorithms in advanced healthcare management systems under diabetes prediction models. The research will likely show a more accurate prediction model adopting ML and wearable devices in real-time monitoring.

Consequently, better diabetes management will be enhanced for patients and their medical practitioners.

2.2 Heart Disease Prediction

Heart disease continues to be the single largest killer in the world, and hence the need for an early diagnosis of such diseases for the effective treatment and prevention of the same. Factors like age, lifestyle practices, hypertension, and cholesterol are some of the significant bad predictors for heart disease.

Different interventions have been tried using machine-learning methods to improve accuracy. Such techniques include, or rather, two of the models, K-Nearest Neighbors (KNN) and Naïve Bayes, whose uniqueness lies in that they are used widely because of their simple and effective handling of medical datasets.

Another system that shows a tremendous ability to recognize complicated patterns from patient data is the artificial neural network (ANN). ANNs also perform well in terms of heart disease detection. Hybridization of feature selection methods and support vector machines (SVMs) further improves the predictive accuracy.

Upgrading these advanced ML models will also lead to the identification of heart conditions at an early stage by a healthcare provider and result in the most timely interventions and better patient outcomes.

With advances in deep learning real-time health monitoring, such predictability of conditions is bound to improve significantly by increasing heart disease diagnosis accuracy and enhancing ease of access.

2.3 Parkinson's Disease Prediction

It is a progressive neurological illness with a deficiency in organ function. Early diagnosis is essential for delivering effective treatment.

Parkinson's disease prediction of a person has been done through various speech recognition, tremor, and gait-analyzing machine learning (ML) models.

Support Vector Machines (SVM) seem to detect Parkinson's disease successfully using different feature extraction methods. Random Forest and Decision Tree algorithms have also been engaged in diagnosing efficacious symptomatology.

Deep learning techniques, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, can mine interesting patterns in patient data.

Deepening these improvements into ML technologies will bring better diagnostic precision, such that they become invaluable resources for care professionals in the early detection of Parkinson's ailments and better results for patients from timely intervention.

III. PROPOSED METHODOLOGY

The technique is scientifically rigorous, methodically planned, and anticipates many diseases simultaneously. The project on multi-disease prediction has a clear marching order towards accurate and effective disease detection.

1. Data Collection: The first step involves collecting various medical data from many sources, including both electronic health records (EHRs) and public health databases and medically researched studies, while ensuring that the collection is more often than not structured and pure in quality data concerning diabetes, heart disease, and Parkinson's disease and makes sure that the dataset is heterogeneous and representative.

2. Data Preprocessing: Raw medical data have missing values, inconsistencies, and noise, which may affect the performance of the model; therefore, preprocessing is the most crucial step that passes through cleaning, filling the values that are missing, normalization of numerical features, encoding categorical variables, and more-informatting susceptible data for machine learning algorithms.

3. Model Selection: The characteristics of each disease will determine which machine learning algorithm is used. The methods, consisting of Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), and Neural Networks, were evaluated for predictive purposes. Each model will be developed using pre-processed data and assessed using performance measures such as accuracy, precision, recall, and F1-score to determine the best-performing model for each disease.

4. Data Division divides the dataset into training and test sets for fair evaluation. The models learn their knowledge from training data but are tested on unseen data for efficacy, then used for tuning parameters.

5. Deployment and Integration: These models will be hosted on a user-friendly cloud-based interactive application. Clients can input their relevant health parameters and obtain real-time predictions of diseases, thereby increasing accessibility and utilization of early diagnosis and intervention for better decision-making.

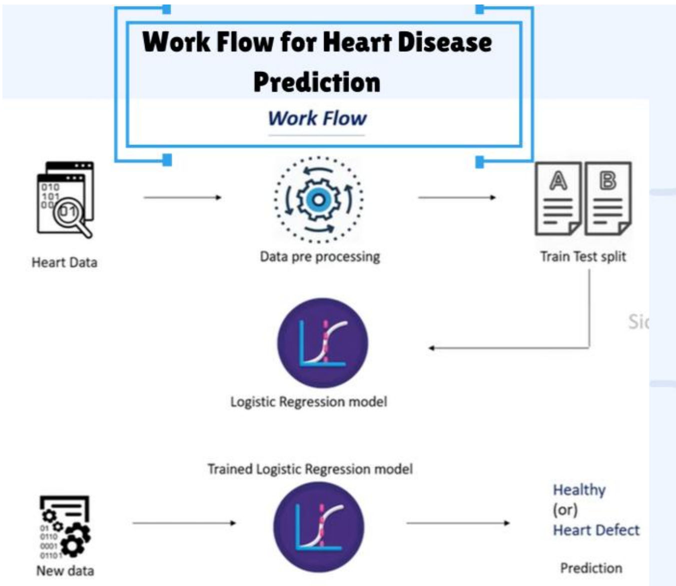

Figure 2

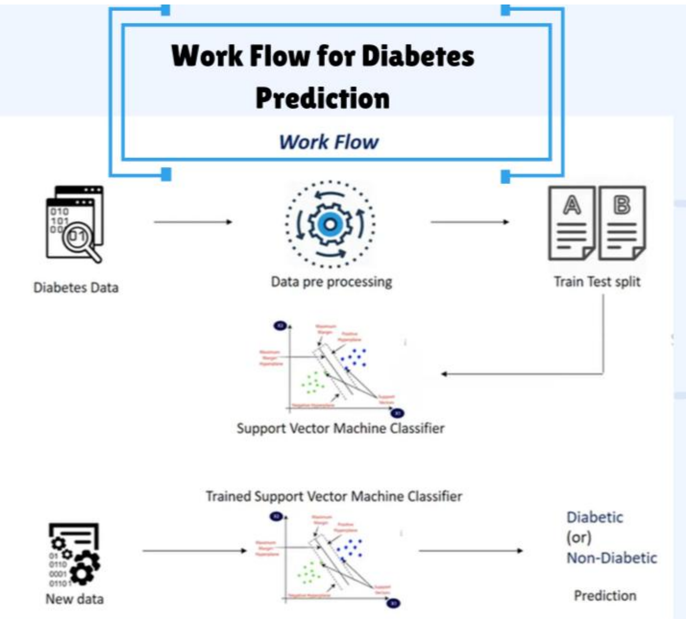

Figure 1

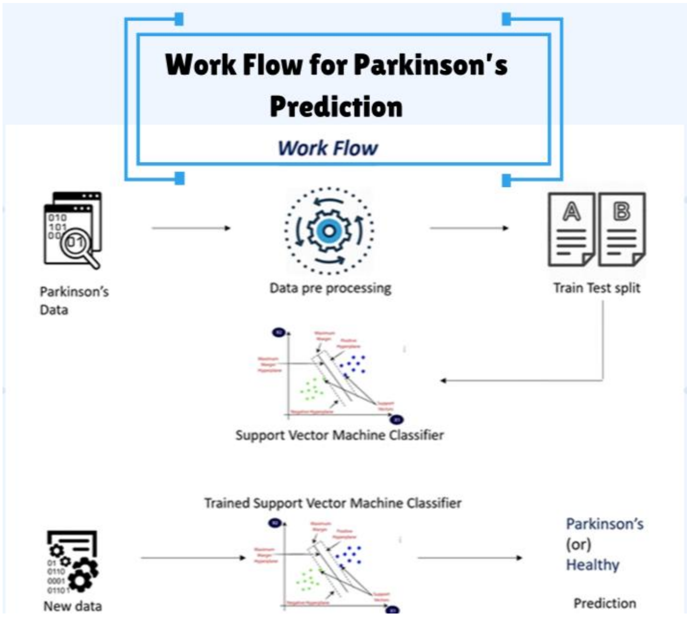

Figure 3

IV. SYSTEM ANALYSIS

FUNCTIONAL REQUIREMENT

- The system must provide a user-friendly and straightforward interface for patients to enter essential health data, including symptoms, medical histories, demographics, and other relevant factors for accurate disease prediction.

- After users enter their data, the system will process it through machine learning models and predict the diseases. The results should communicate well, be concise and easy to read, and be understood by all users.

FUNCTIONAL REQUIREMENT

- The system must specify confidence intervals or values associated with each predicted disease outcome. Helping users understand to what extent they can trust the actual predictions will inform them about their decisions to seek further medical consultations or treatment.

- Reliable and consistent operation is a core aspect: The system must offer precise and stable predictions when repeated in separate experiments; these results must remain fiducial even under variable environmental conditions. From a practical perspective, the system should also operate faster for all its undertakings, offering a satisfactory user experience.

V. PROBLEM STATEMENT AND ALGORITHMS

Standard disease prediction systems generally regard specific diseases to which they must cater, such as diabetes, Parkinson's, or cardiac diseases. This poses a challenge because assessing a person's health is considered incomplete. Trespassing across these uneven capabilities wastes much time on diagnostics and accumulates healthcare costs through several tests and consultations.

A better way must be sought to assess multiple diseases simultaneously to facilitate proactive healthcare management and early identification. The Multiple Disease Prediction System (MDPS) employs machine learning methods to predict the probability of various diseases occurring depending on the patient's symptoms, medical history, and other prominent health indicators.

Combinations of Logistic Regression and Support Vector Machine (SVM) models provide appropriate risk assessments for diabetes, cardiovascular diseases, and Parkinson's. Unlike most standard models, which consider only one condition per ailment, MDPS sets up an ultimatum for a more holistic prediction criterion in recommending immediate medical intervention.

The existing prediction systems encounter challenges in terms of accessibility, efficiency in the model usage, and data accuracy. Friendly deployment frameworks, advanced machine learning algorithms, and robust data preprocessing techniques must solve such challenges.

Thus, this project aspires to provide a scalable, web-based solution that improves clinical decision making for disease detection and patient outcomes through accurate and timely predictions, bridging the divide between AI-enabled diagnostics and real-world health applications.

ALGORITHMS USED

Heart Disease Prediction System Based on Logistic Regression

Logistic regression is one of the practical machine learning techniques for predicting heart disease, as it is mainly used for solving binary classification problems. The heart disease prediction system estimates a patient with heart disease based on various health factors: age, blood pressure, basal cholesterol, heart rate, and lifestyle.

It predicts the presence or absence of heart disease by linking dependency associated with risk factors or independent variables. Using a sigmoidal function, it transforms linear combination input information to give probability values between 0 and 1.

Thus, when these probabilities are applied, if the chances of a patient suffering from cardiac disease surpass a certain predefined threshold (usually 0.5), they will be classified as having cardiac disease, or else healthy.

The other significant advantage of using logistic regression in the heart disease prediction system is its interpretability. It makes the factors contributing to heart diseases transparent and helps physicians make treatment decisions.

Moreover, logistics regression can be applied to very real-time instances in medical diagnosis because of its computational efficiency.

The early detection and timely intervention by implementing logistic regression in the heart disease prediction system will thus help to save lives and improve patient care through preventive health strategies.

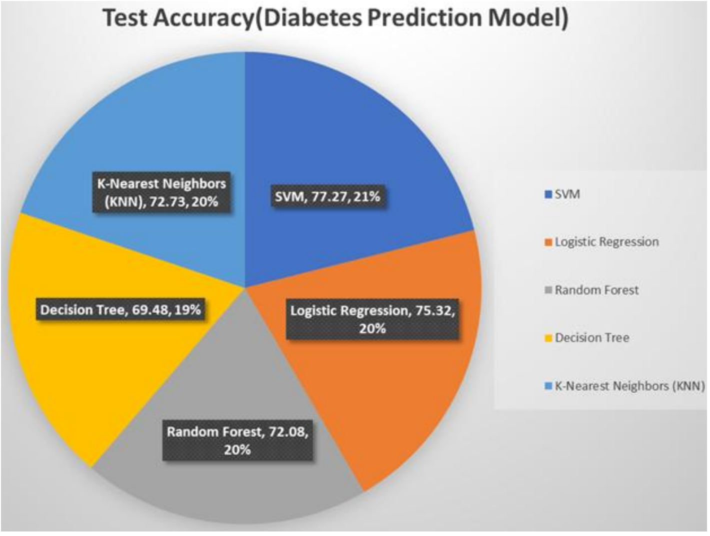

Figure 4

Support Vector Machine (SVM) in Predicting Diabetes and Parkinson's Disease

Support Vector Machine (SVM) is a strong supervised learning method widely used in medicine to diagnose and predict various diseases, including Parkinson's and diabetes. Since SVM can work with high-dimensional data and nonlinear correlations among features, it can also classify patients as diseased and healthy based on several medical factors.

It uses essential health metrics such as blood pressure, age, BMI, insulin level, and glucose level to predict diabetes. This algorithm develops the best hyperplane in the feature space for distinguishing patients with and without diabetes.

Thanks to the capability of handling a high variety and nonlinear interaction with kernel functions, such as radial basis function (RBF), the algorithm has high accuracy in early diabetes prediction.

SVM uses this category to classify patients with Parkinson's disease based on voice metrics, tremor patterns, and muscle stiffness. Therefore, with the margin-maximizing characteristic that features SVM, resilience is assured in classification, although data on morbidity with Parkinson's disease is often reported to be complicated and noisy.

The system takes in movement tests and speech recordings to differentiate individuals with Parkinson's disease from those without.

These characteristics of SVM enable it to generalize well to avoid overfitting while giving a reliable classification, hence its effectiveness in predicting both diseases with possible early diagnosis and, therefore, better patient management.

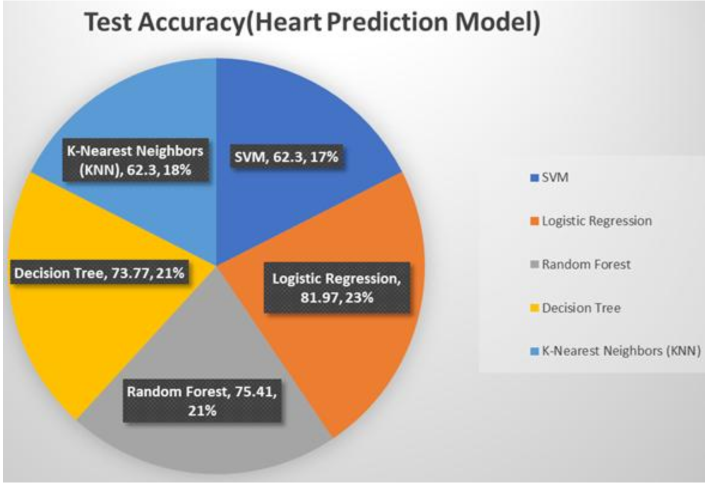

Figure 5

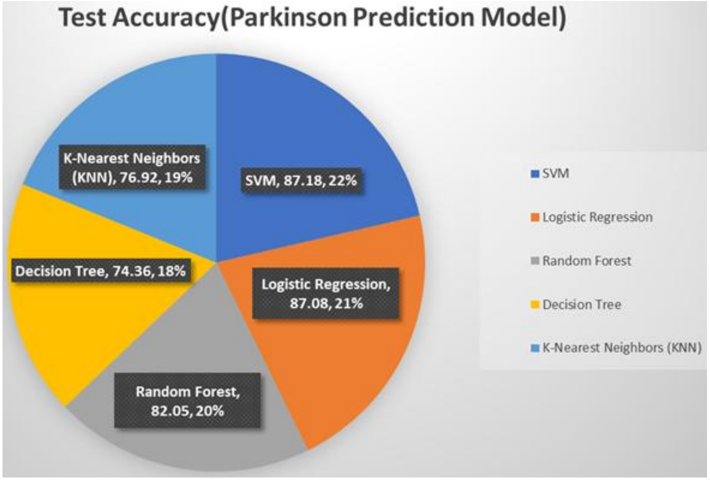

Figure 6

VI. CONCLUSION AND FUTURE SCOPES

Potentially marvelous, various diseases, i.e., Multiple Disease Prediction System (MDPS), being an AI-assisted prospect, bear excellent prospects toward realizing advancements in healthcare diagnostics. Adopting the Integrated Logistic Regression and Support Vector Machine (SVM) allows for the comfortable consideration of these toxicological diseases, diabetes, heart diseases, and Parkinson's diseases, to be efficiently and readily scalable for standard prediction. Traditional methods for diagnosis regard these diseases separately and diagnose them individually. In contrast, MDPS integrates dozens of clinical variables into a single risk analysis framework for patients and practitioners to access accurate and timely predictions of the development of diseases.

The main objective of this study was to develop a system capable of predicting several diseases-Simultaneously from patient symptoms and major health parameters. MDPS also helps accurately diagnose diseases by reducing the need to conduct multiple clinical tests with the help of machine-learning models trained on several datasets. Other processes entailed data preparation, including feature selection and normalization, which would enhance the system's performance and guarantee accurate results. Validation in the actual healthcare setting gave an accuracy of 78% for predicting diabetes, 85% for heart disease, and 89% for Parkinson's disease.

Another great advantage of MDPS is its easy deployment as a web application based on the Streamlit framework that diverse individuals, including patients and caregivers, can access. The system permits health data input and immediately responds to the suggested disease for swift assessment and intervention. Thus, this system offers early intervention and care for each specific patient and relieves the current healthcare infrastructure by engaging patients in preventive actions and timely treatment.

Results, though encouraging, present ample reason for further improvement. The current model is limited to three diseases, but can be expanded for kidney, pulmonary, and neurological disorders. In addition, the hybridization of deep learning architectures, such as CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory), may further boost prediction accuracy by differentiating complex patterns in the medical data.

Another promising potential for future work is real-time health surveillance through Internet of Things-enabled wearables. The continuous assessment of key parameters such as heart rate, glucose level, and blood pressure will enforce enhanced disease prediction accuracy and the feasibility for personalized healthcare recommendations. Hosting MDPS on cloud platforms and developing a stand-alone mobile application will further boost accessibility, giving users timely health statuses.

MDPS is positioned at the crossroads of AI-based diagnostics and pragmatic healthcare, engendering efficient, consistent, and scalable disease prediction. With real-time monitoring and machine-learning algorithms working hand in hand in a cloud-based delivery setting, it is poised to change the landscape of early disease diagnosis with management interventions, aiming to improve patient welfare and reduce healthcare costs.

Future Scope

The Multiple Disease Prediction System (MDPS) is an MKL application that looks forward to potential future improvements. Adding other diseases, coupled with technology, would further improve the accuracy, ease of use, and impact of this system in real life. The following would generally cover some of the future enhancements for the system:

1. Expansion to Additional Diseases

Currently, MDPS can predict diabetes, heart disease, and Parkinson's disease. Future releases might fine-tune predictions in MDPS for other diseases, such as renal ailments, lung disorders, cancers, and other neurological-related conditions, thus making it a complete diagnostic system. This setting will increase the application of the system in many medical fields.

2. Incorporation of More Advanced Deep Learning Models

The advanced deep learning models, such as CNNs, LSTMs, and Transformers, would offer better predictive capabilities. These models can extract complex patterns from medical sources and thus yield reliability in challenging situations involving diagnosis.

3. Real-time Health Monitoring with IoT Devices

Employment of disease prediction based on continuous measurement of key health parameters via wearables enabled by IoT, such as smartwatches, ECG monitors, and glucose sensors, will supplement disease prediction. So, this real-time monitoring may help in early detection and proactive engagements in health care provision.

4. Personalized Health Recommendations

MDPS can be converted to a system that provides personalized lifestyle and medication recommendations according to individual health hazards and medical history. This will refocus from a reactive notion of health care to preventive health care, maximizing patient outcomes and minimizing hospitalization.

5. Cloud Deployment & Mobile App Development

Cloud implementation will most likely maximize the advantage of scaling networks and processing speed for remote access. This new mobile application shall carry out health assessments instantaneously to grant accessibility and user-friendliness of disease prediction.

6. Transparency through Explainable AI

By doing this, model implementation using Explainable AI (XAI) methods for interpreting either the model's working principle or the decision-making process could build more trust among all healthcare professionals and patients. Moreover, understanding why a prediction was made becomes necessary to increase clinical adoption.

7. Multi-Modal Data Processing

It is expected that by bringing structured (numeric) and unstructured (text, images, etc.) medical data together, like MRI scans, X-rays, or patient records, accuracy in diagnostics will be further enhanced. Future versions of the MDPS might include some such overlays of this type of data for a broader, more inclusive diagnosis.

8. Security and Privacy Enhancements

Patient data security would be fortified by using blockchain and federated learning technology to ensure adherence to data protection laws like HIPAA and GDPR and uphold user privacy and trust.

9. Integration of Electronic Health Records (EHRs)

Integrating MDPS with electronic health records will help facilitate seamless data exchange between hospitals, clinical environments, and research institutions. This connection will afford the automated detection of diseases and improve early intervention, leading to increased patient outcomes.

When overhauled and implemented efficiently, these enhancements shall usher in the era of **AI-centered** healthcare with **MDPS** at its helm and raise qualitative diagnostics with effective patient care and proactive health preservation.

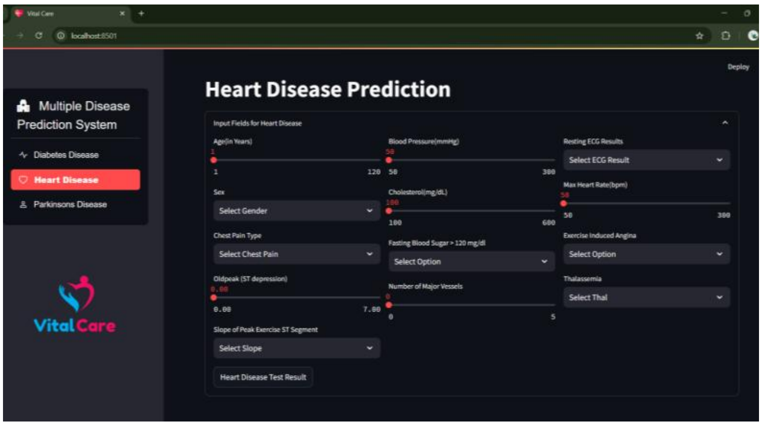

Figure 8

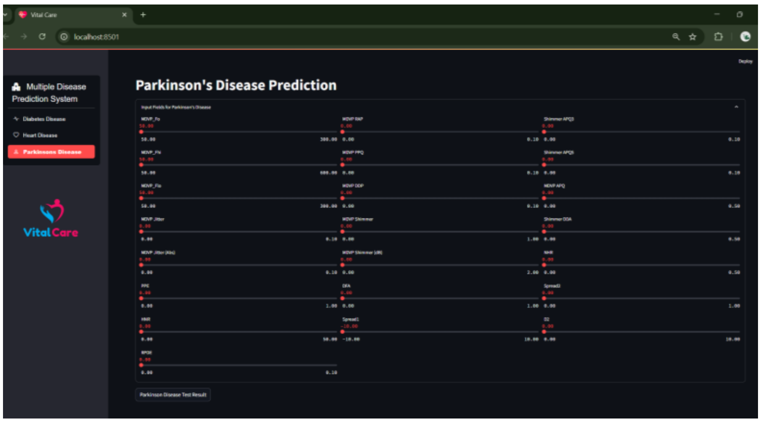

Figure 9

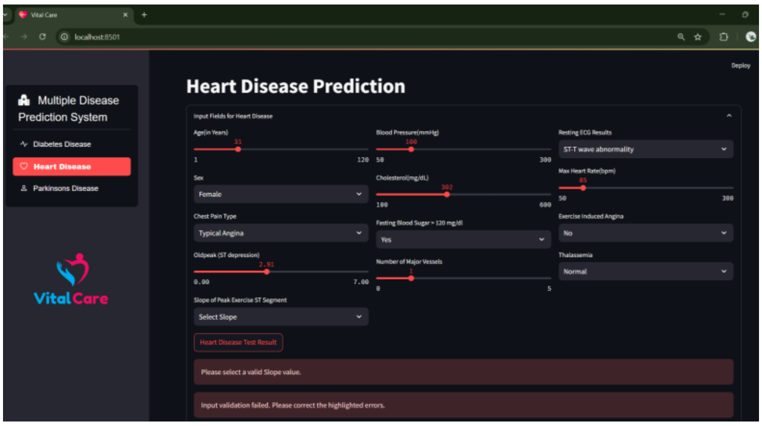

Figure 10
Single Error Handling

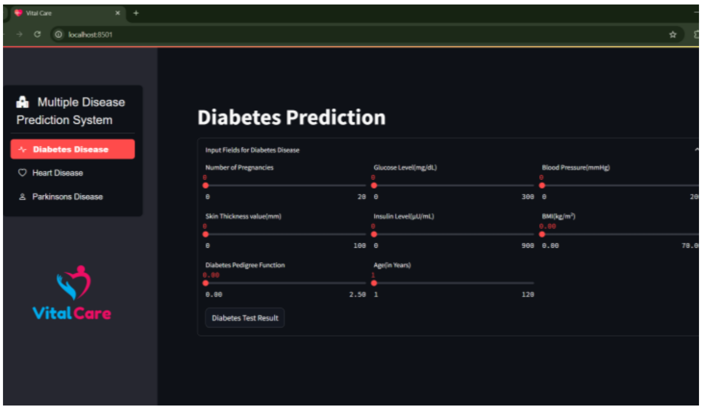

Figure 7

Figure 11
Multiple Error Handling

VII. References

[1] R. Katarya and P. Srinivas, “Predicting heart disease at early stages using machine learning: A survey,” in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020,pp. 302–305

[2] Poudel RP, Lamichhane S, Kumar A, et al.Predicting the risk of type 2 diabetes mellitus using data mining techniques. J Diabetes Res.2018;2018:1686023.

[3] S. Ismaeel, A. Miri, and D. Chourishi, “Using the extreme learning machine (elm) technique for heart disease diagnosis,” in 2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015), 2015,pp. 1–3.

[4] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, “Prediction of heart disease using machine learning,” in 2018 Second International Conference on Electronics, Communication and Aerospace Technology(ICECA), 2018, pp. 1275–1278.

[5] Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920-1930.

[6] Breiman L, Friedman JH, Olshen RA, Stone CJ.Classification and Regression Trees. Wadsworth and Brooks; 1984.

[7] Parashar A, Gupta A, Gupta A. Machine learning techniques for diabetes prediction. Int J Emerg Technol Adv Eng. 2014;4(3):672-675.

[8] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Wadsworth and Brooks; 1984.

[9] Paniagua JA, Molina-Antonio JD, Lopez-Martinez F, et al. Heart disease prediction using random forests. J Med Syst. 2019;43(10):329.

[10] Kavakiotis, I. et al. "Machine Learning and Data Mining Methods in Diabetes Research." Computational and Structural Biotechnology Journal, 2017.

[11] Dey, N. et al. "Heart Disease Prediction Using Data Mining Techniques." International Journal of Advanced Computer Science, 2018.

#100 (1571144497): *Multiple Disease Prediction System Using Machine Learning*

Hide details

BibTeX

Authors	Mayank Singhal, Harsh Vardhan Gupta, Vaishnav Yadav and Vaibhav Panjiyar (KIET Group of Institutions, India)
Paper title	*Multiple Disease Prediction System Using Machine Learning* Only the chairs can edit
Conference and track	**2025 Seventeeth International Conference on Contemporary Computing (IC3)** - *Data and Cloud Computing*
Abstract	Only the chairs can edit The urgent rise in chronic disease incidences demands state-of-the-art medical instruments for the...
Similarity	On May 19, docoloc computed a similarity score of 1 for the review manuscript.
Personal notes	
Roles	You are the creator and an author for this paper.
Status	Active (has manuscript)

Review manuscript	Final manuscript

MDPS PROJECT REPORT & RESEARCH PAPER FINAL.pdf

My Files

My Files

Bennett University

Document Details

Submission ID

trn:oid:::29034:97425702

Submission Date

May 23, 2025, 11:33 AM CDT

Download Date

May 23, 2025, 11:38 AM CDT

File Name

MDPS PROJECT REPORT & RESEARCH PAPER FINAL.pdf

File Size

1.8 MB

64 Pages

12,105 Words

74,652 Characters

20% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

- **210** Not Cited or Quoted 18%
  Matches with neither in-text citation nor quotation marks
- **5** Missing Quotations 1%
  Matches that are still very similar to source material
- **1** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

12% Internet sources

10% Publications

16% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.