International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI-Powered Medical Diagnosis System Using Machine Learning Techniques

Dr. P.Chiranjeevi[1], K. Sai Nikith[2], K. Ganesh[3], G. Varun[4], S. Shiva Prasad[5], G. Vinay[6]

Hod-CSE(DS) III B.Tech, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India

ABSTRACT :

In the healthcare industry, the accuracy and speed of diagnosis are critical. However, conventional methods often rely on manual analysis by medical professionals, which may lead to errors and diagnostic delays. This paper presents an AI-powered medical diagnosis system that employs machine learning algorithms to assist in detecting various diseases, including diabetes, heart disease, Parkinson's, lung cancer, and hypothyroidism. The system leverages patient health records and medical inputs to predict disease likelihood through a web-based interface built with Streamlit. Extensive model training was conducted using real-world datasets, including PIMA, Cleveland, and Parkinson's voice data, sourced from Kaggle and UCI repositories. Results show that the system enhances diagnostic accuracy while reducing the cognitive load on healthcare professionals.

Keywords: Medical diagnosis, machine learning, AI in healthcare, disease prediction, Streamlit, patient data analysis, EHRs.

INTRODUCTION

The growing complexity of healthcare data and the increasing patient load challenge the efficiency and accuracy of medical diagnostics. In high-pressure settings like emergency rooms or rural clinics, clinicians face time constraints and limited resources. Artificial Intelligence (AI) has shown immense promise in augmenting diagnostic decision-making by processing large volumes of electronic health records (EHRs), lab results, and patient symptoms. This paper proposes an AI-based multi-disease prediction system that integrates machine learning models for five prevalent health conditions. The goal is to develop a lightweight yet effective decision support tool for preliminary disease detection.

2.EXISTING SYSTEM

The current medical diagnosis process primarily relies on human expertise, where doctors analyze patient symptoms, medical history, lab reports, and imaging to determine a diagnosis. While experienced medical professionals provide accurate diagnoses, several challenges persist:

1. High Risk of Human Error: Due to the vast amount of patient data and complex disease patterns, misdiagnoses or overlooked symptoms are common.
2. Time-Consuming Analysis: Manually reviewing electronic health records (EHRs) and test results takes significant time, leading to diagnostic delays.
3. Overburdened Healthcare Professionals: Doctors, especially in emergency rooms or resource-limited settings, face heavy workloads, increasing stress and the likelihood of errors.
4. Limited Decision Support: The existing system lacks advanced tools to assist doctors in quickly identifying critical conditions, such as strokes or heart attacks.
5. Inconsistent Accuracy: Diagnosis quality may vary depending on the experience and specialization of the healthcare professional, potentially leading to disparities in patient care.

3.PROPOSED SYSTEM

AI can streamline medical diagnosis by analyzing large volumes of patient data, such as symptoms, history, labs, and imaging-quickly and accurately. In high-pressure settings like emergency rooms, AI can help detect critical conditions, such as heart attacks, reducing the risk of misdiagnosis. By processing EHRs efficiently, Al ensures no crucial data is overlooked, supporting doctors in making faster, more informed decisions, even in complex or resource-limited scenarios. This leads to improved accuracy, reduced diagnostic delays, and better patient outcomes.

ADVANTAGES OF THE PROPOSED SYSTEM

1. **Enhanced Accuracy:** AI algorithms can analyze vast amounts of patient data with high precision, reducing diagnostic errors.
2. **Faster Decision-Making:** AI processes EHRs in real time, helping doctors make informed decisions quickly, particularly in emergencies.
3. **Reduced Workload:** By automating data analysis, AI reduces the burden on healthcare professionals, allowing them to focus on patient care.
4. **Early Detection of Critical Conditions:** AI can identify life-threatening conditions like heart attacks and strokes more efficiently, ensuring timely intervention.
5. **Scalability and Accessibility:** AI-driven diagnostic tools can be deployed in remote or resource-limited areas, improving healthcare accessibility.
6. **Data-Driven Insights:** AI continuously learns from vast medical datasets, enhancing predictive capabilities for better treatment outcomes.

4.METHODOLOGY

4.1 Data Collection

Gather data related to the diseases you want to predict. This data should include both features (such as symptoms, patient demographics, and medical history) and labels (whether or not the patient has each disease). Datasets are collected from Kaggle and the UCI Machine Learning repository.
o Diabetes - PIMA Dataset.
o Heart - Cleveland, Statlog.
o Parkinson's, Hepatitis - Kaggle

4.2 Data Preprocessing

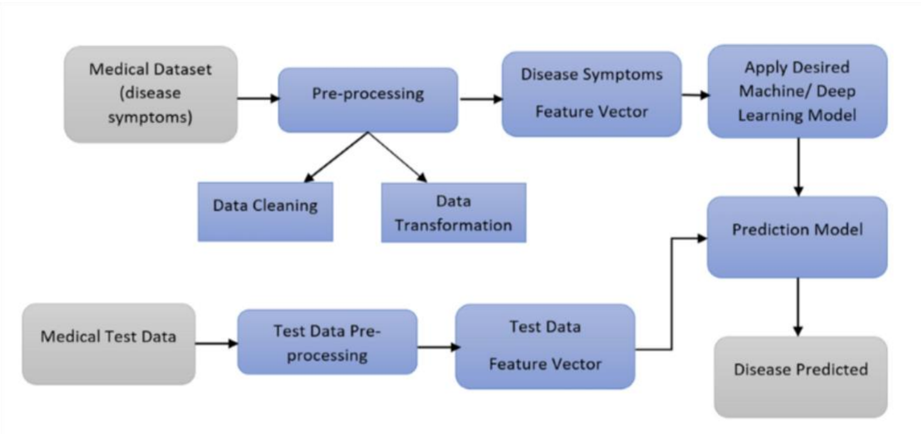
Clean the data to handle missing values, outliers, and inconsistencies. Normalize or standardize the features if necessary to ensure that they're on the same scale. Encode categorical variables into numerical representations if needed (e.g., one-hot encoding). Datasets are imbalanced. So, they are balanced using SMOTE, Over sampling technique.
There are missing values and null values. So, a few attributes with missing values and null values are replaced by their mean value. A few other attributes play a major role in predicting disease, so their missing values are replaced using Regression Imputation

4.3 Model Training

This step involves choosing the appropriate algorithm and representation of data in the form of the model. The cleaned data is split into two parts – train and test28 (proportion depending on the prerequisites); the first part (training data) is used for developing the model. The second part (test data), is used as a reference.

5.SYSTEM ARCHITECTURE

System architecture is a comprehensive blueprint that defines the structure, behavior, and interactions of various components within a system—whether it's a software application, a computer system, or a complex network of systems. It provides a high-level view of how the system is organized and how different parts such as hardware, software, data storage, processing units, communication protocols, and user interfaces interact to perform specific functions. In software systems, architecture describes how modules or services are divided, how they communicate (e.g., via APIs or message queues), and how data flows through the system.



In hardware systems, it includes the design of processors, memory units, input/output devices, and how they are connected. System architecture also includes considerations for scalability (handling growth in users or data), security (protecting data and operations), maintainability (ease of updates and debugging), and performance (speed and efficiency).

6. RESULTS AND OUTPUT

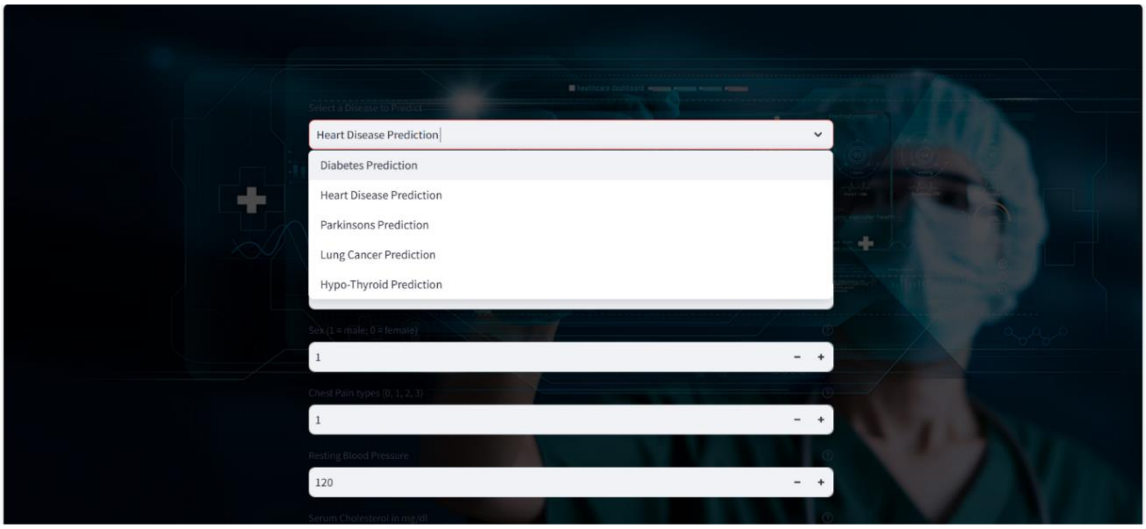


Fig: User Interface

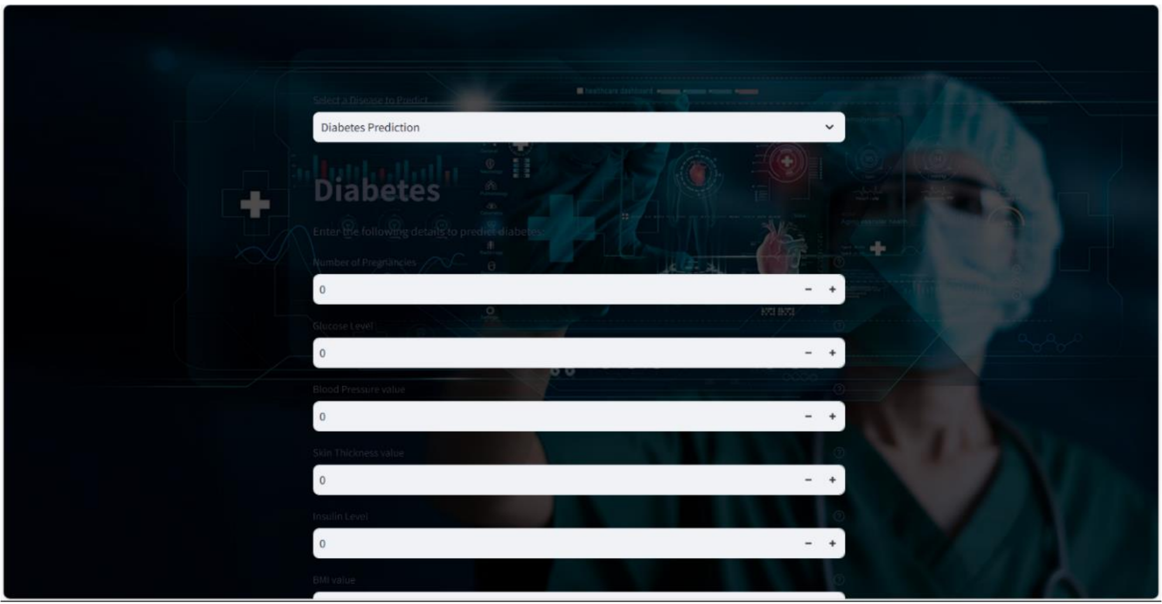


Fig: Enter input values

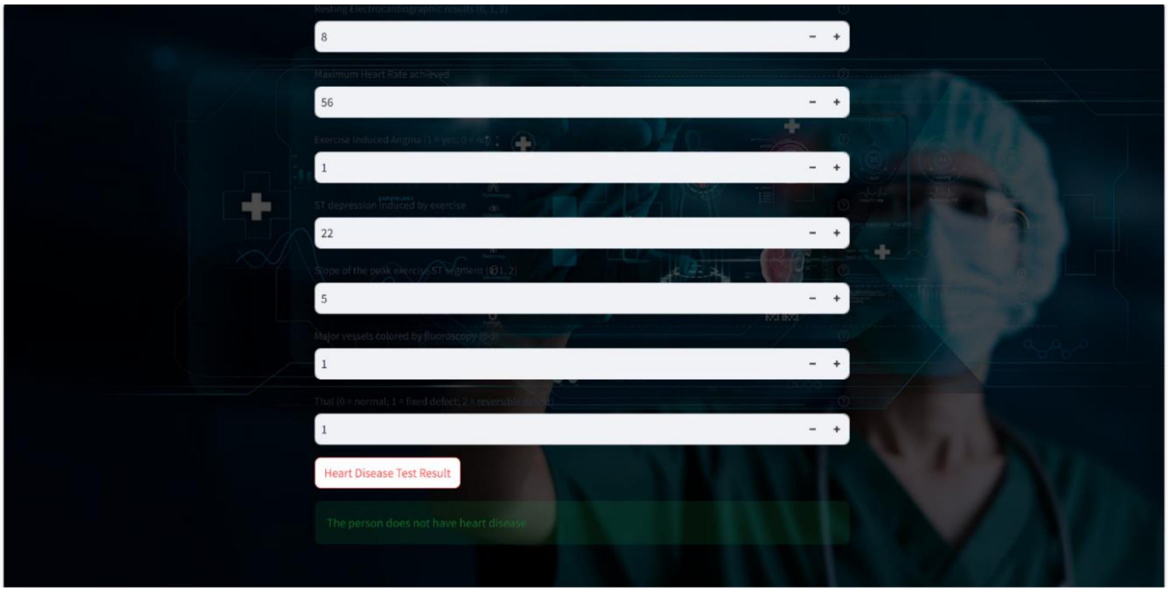


Fig: Result for the given inputs

7. CONCLUSION

Predicting multiple diseases using machine learning algorithms can be a promising approach with significant potential benefits in healthcare. Through the utilization of various machine learning techniques, such as logistic regression, support vector machines, algorithms, and so on, it is possible to analyze complex datasets containing a multitude of health-related features and predict the likelihood of multiple diseases in patients. One key advantage of machine learning in disease prediction is its ability to learn patterns and relationships from labeled data, enabling the model to make predictions on unseen data. By training the model on historical patient data with known disease outcomes, it can learn to identify subtle patterns and associations indicative of different diseases. Additionally, machine learning algorithms allow for the integration of diverse data sources, including demographic information, medical history, genetic markers, and diagnostic test results. This comprehensive approach can enhance the accuracy and robustness of disease prediction models, enabling healthcare providers to make more informed decisions.

However, there are several challenges and considerations associated with the implementation of machine learning for disease prediction. These include the need for large and high-quality datasets for training, potential biases in the data that may affect model performance, and the interpretability of complex models, especially in clinical settings where transparency and understanding are crucial.

8.FUTURE SCOPE

Integration of Multi-Modal Data:

Incorporating diverse data sources beyond traditional electronic health records, such as wearable devices, genomic data, environmental factors, and social determinants of health, can provide a more comprehensive understanding of disease risk factors. Integrating these multi-modal data streams into predictive models can enhance their accuracy and predictive power.

Advanced Feature Engineering Techniques:

Future research could explore more sophisticated feature engineering methods, including deep feature synthesis, autoencoders, and generative adversarial networks (GANs). These techniques can automatically learn hierarchical representations of complex data, capturing intricate relationships between variables and improving the discriminative power of predictive models.

Ensemble Learning Approaches:

Ensemble learning methods, such as stacking, boosting, and bagging, can combine multiple base models to create a more robust and accurate predictive model. Future enhancements could focus on developing novel ensemble strategies tailored specifically for multiple disease prediction tasks, leveraging the strengths of different algorithms and data representations.

Personalized Medicine Approaches:

Moving towards personalized medicine requires tailoring disease prediction models to individual patient characteristics, preferences, and health trajectories. Future enhancements could explore the integration of patient-specific data, such as longitudinal health records, lifestyle factors, and treatment histories.

9. REFERENCES

[1] Ankush Singh, Ashish Yadav, Saloni Shah, and Prof. Renuka Nagpure,” Multiple Disease Prediction System”, Dept. of Information Technology Engineering, Atharva College of Engineering. (2022).

[2] Yaganteeswarudu, A. (2020). Multi Disease Prediction Model by using Machine Learning and Flask API. 2020 5th International Conference on Communication and Electronics Systems (ICCES).

[3] Marouane Fethi Ferjani Computing Department Bournemouth University Bournemouth, England s5319941@bournemouth.ac.uk,” Multi Disease Prediction Model by using Machine Learning (2020)”

[4] Sonar, P., & Jaya Malini, K. (2019). Diabetes Prediction Using Different Machine Learning Approaches. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). doi:10.1109/iccmc.2019.8819841

[5] American Heart Association, http://www.heart.org/HEARTORG/, (2014)