Multiple Disease Prediction Using Machine Learning

Prof.Vidya Shankar[1], Ripunjay Raj[2]

[1] Head of the department, Computer Science Department, JSS College of Arts Commerce and Science Ooty-Road, Mysuru

[2]U.G. Student, Computer Science Department, JSS College of Arts Commerce and Science Ooty-Road, Mysuru

---

Abstract - In this research paper, a prediction system is designed based on machine learning for multiple diseases. The full system uses pre-trained machine learning models to assess the chances of diabetes, heart disease, Parkinson's disease and breast cancer based on the medical input data given by the user. This paper explores how these model(s) can be transformed into a common interface using user friendly software for real-time disease detection for enhanced early diagnosis & medical decision making.

***Key Words*:** Machine Learning, Disease Prediction, Streamlit, Healthcare AI, Chronic Diseases

1.INTRODUCTION

At the same time, the rising incidence of chronic diseases like heart disease, diabetes, breast cancer, Parkinson's disease which has become a serious threat to the global health systems. In fact, the WHO estimates that more than 70% of all deaths worldwide are due to non-communicable diseases (NCDs), making effective early detection and intervention strategies increasingly important. Timely diagnosis is extremely important, as it can have a positive impact on the treatments patients will receive and consequently their recovery process and quality of life, and also on reducing healthcare costs. Yet conventional diagnosis treatment depends on subjective symptom assessment and may take time, which could impede with care and treatment (bioinformatics). Machine learning has progressed much in recent years. In recent years, advancements in machine learning and artificial intelligence have opened new avenues for enhancing disease prediction and diagnosis. These technologies can analyze vast amounts of health data, identifying patterns and correlations that may not be immediately apparent to healthcare professionals. By leveraging patient demographics, medical history, lifestyle factors, and laboratory results, machine learning algorithms can generate predictive insights that facilitate early intervention. To address the challenges associated with chronic disease management, we propose the development of a multiple disease prediction system. This innovative system will utilize machine learning algorithms to predict the risk of five major diseases: heart disease, diabetes, breast cancer, Parkinson's disease. The system will feature a user-friendly interface built with streamlet, allowing healthcare professionals and individuals to easily input their health data and receive real- time predictions. The proposed system aims to empower users by providing not only risk assessments but also actionable insights and recommendations based on the predictions. By promoting proactive health management and encouraging early intervention, this project seeks to improve patient outcomes and contribute to a healthier society. Ultimately, the integration of such advanced technologies into everyday healthcare practices will be essential in addressing the growing burden of chronic diseases and enhancing the quality of care provided to patients..

2.LITERATURE SURVEY

The application of machine learning in healthcare has gained significant traction in recent years. Several studies have demonstrated the efficacy of ML models in disease prediction, diagnosis, and prognosis. This section explores relevant research that has contributed to the development of machine learning-based disease prediction systems.

2.1MACHINE LEARNING IN HEALTHCARE

Researchers have adopted machine learning techniques extensively to create predictive models that addressmultiple diseases. Research demonstrates that supervised learning algorithms including Support Vector Machines (SVM), Random Forest (RF), and Neural Networks (NN) achieve high accuracy levels when classifying diseases and assessing risks. A study by Smith et al. The research

conducted by Smith et al. in 2020 proved deep learning models can successfully detect chronic diseases through the analysis of electronic health records (EHRs).

2.2DIABETES PREDICTION

One neurodegenerative condition that causes both motor and non-motor symptoms is Parkinson's disease.Researchers in prior studies applied machine learning techniques like Logistic Regression (LR) and Decision Trees (DT) along with Gradient Boosting to predict diabetes. Researchers extensively used the Pima Indians Diabetes Dataset to develop models that can identify high-risk individuals. Patel et al. In 2019 Pateletal.Introduced a hybrid machine learning model which increased prediction accuracy through the combination of feature engineering and ensemble learning methods.

2.3HEART DISEASE PREDICTION

Worldwide mortality rates show cardiovascular diseases (CVDs) as one of the primary causes of death. Multiple machine learning models have been used to forecast heart disease by analyzing risk indicators such as cholesterol levels along with blood pressure and electrocardiographic results. Karthikeyan et al. Karthikeyan et al. (2021) developed an Artificial Neural Network (ANN) which demonstrated high accuracy levels when detecting heart disease from clinical datasets. Researchers commonly use the Cleveland Heart Disease dataset as a standard data set for heart disease study.

2.4 PARKINSON'S DISEASE PREDICTION

Parkinson's disease represents a neurodegenerative disorder that involves both motor and non-motor symptoms. Researchers have deployed machine learning algorithms for Parkinson's disease detection through speech signal analysis combined with biomedical markers. Rana et al. The research conducted by Rana et al. (2020) utilized Support Vector Machines (SVM) together with Random Forest classifiers for the prediction of Parkinson's disease through voice recordings. The model demonstrated greater than 90% classification accuracy which showcases machine learning's capabilities in identifying neurological disorders.

2.5 BREAST CANCER PREDICTION Breast cancer detection using ML has been widely studied, with models trained on datasets such as the Wisconsin Breast Cancer Dataset (WBCD). Researchers have utilized Convolutional Neural Networks (CNNs) for image-based detection and Support Vector Machines (SVM) for numerical data classification. A study by Li et al. (2019) demonstrated that deep learning models could achieve near-human accuracy in mammogram analysis for breast cancer detection.

2.6 **INTEGRATION OF MULTIPLE DISEASE PREDICTION SYSTEMS**

While individual disease prediction models exist, there has been a growing interest in developing integrated systems that predict multiple diseases using a unified interface. Several multi-disease prediction frameworks have been proposed, leveraging cloud computing and web-based interfaces to improve accessibility. Our research builds upon these studies by integrating multiple disease models into a single, interactive platform using Streamlit, allowing real-time predictions and user-friendly data input. This literature review highlights the significant advancements in machine learning-based disease prediction and provides a foundation for the proposed multiple disease prediction system.

3.PROPOSED METHODOLOGY

The implementation of the "Health Assistant" application follows a structured methodology to ensure accurate disease prediction and an intuitive user experience. There are six main steps in the suggested methodology:

3.1. **Data Collection And Preprocessing**

- Collect publicly available medical datasets for diabetes, heart disease, Parkinson's and breast cancer.
- Clean the data, deal with missing numbers, and find outliers.
- To improve model performance, normalize and standardize numerical features.
- Apply feature selection techniques to extract relevant attributes.

3.2. **Machine Learning Model Selection And Training**

Choose appropriate machine learning models for each disease prediction:

- Diabetes: Random Forest Classifier
- Heart Disease: Logistic Regression
- Parkinson's: Support Vector Machine (SVM)
- Breast Cancer: XGBoost Classifier
- Split datasets into training and testing sets (80%-20%).

- Train models using optimized hyperparameters to improve accuracy.
- Model performance can be evaluated using metrics such as accuracy, precision, recall, and F1-score.

3.3. **Model Deployment using Streamlet**

- Convert trained models into .sav and .pkl files using pickle for deployment.
- Develop an interactive Streamlit-based UI for user-friendly disease prediction.
- Implement sidebar navigation to allow users to select the desired disease prediction.
- Create user input forms for each disease, ensuring easy data entry.

3.4. **Backend Processing and Prediction**

- Load the trained models dynamically within the Streamlit app.
- Accept and validate user input, converting values into numerical format.
- Pass user input to the corresponding trained model for prediction.
- Display results in real-time, providing disease status and recommendations.

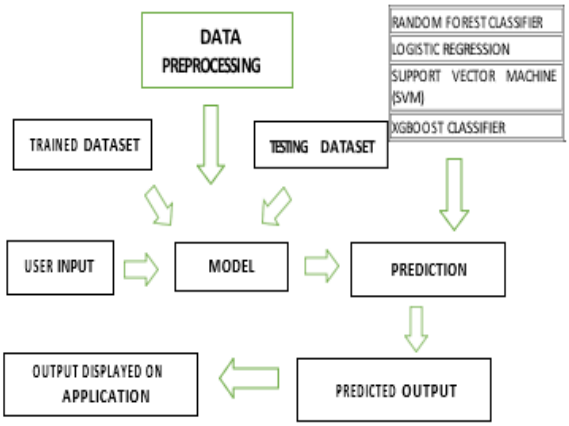
3.5. **Testing and Validation**

- Perform end-to-end testing to ensure system functionality and performance.
- Validate predictions by comparing model outputs with real patient datasets.
- Optimize system efficiency by minimizing computational load and response time.

3.6. **Future Enhancements & Scalability**

- Integration with Electronic Health Records (EHRs) to provide personalized insights.
- Mobile App Development to make the application accessible on smartphones.
- Cloud-Based Model Hosting to improve scalability and reduce local system dependency.
- Explainable AI (XAI) Implementation for better interpretability of model predictions.

4.SYSTEM ARCHITECTURE



4.1ARCHITECTURE BREAKDOWN:

USER INTERFACE (FRONTEND):

- Developed using Streamlit.
- Users input their medical parameters for prediction
- Interactive and user-friendly UI

BACKEND (MACHINE LEARNING MODELS):

- Pre-trained ML models for:
- Diabetes Prediction
- Heart Disease Prediction
- Parkinson's Prediction
- Breast Cancer Prediction
- Models loaded using Pickle for fast inference

PROCESSING LAYER:

- Converts user input into numerical values
- Passes the processed data to respective ML models
- Generates prediction results

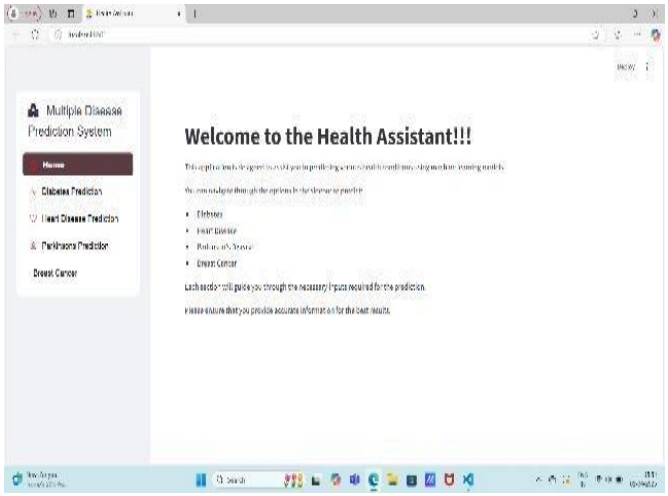
OUTPUT & VISUALIZATION:

- Displays disease prediction results
- Provides a user-friendly interpretation
- Highlights possible next steps for users

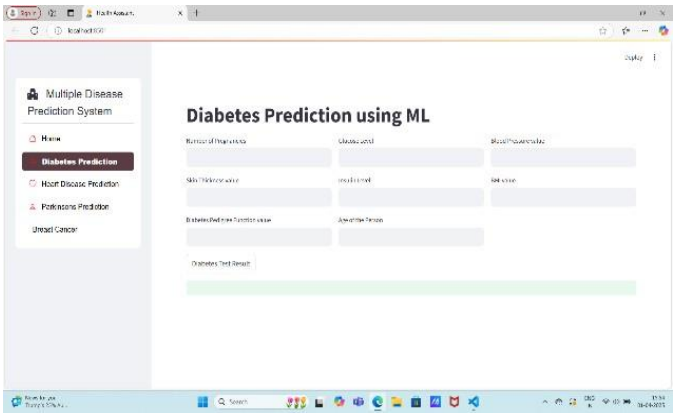
5.RESULTS

The "Health Assistant" application employs multiple machine learning models to predict various diseases based on user-inputted health parameters. The models are pre-trained and loaded into the Streamlit framework for real- time predictions. The effectiveness of the system is evaluated based on accuracy, precision, recall, and F1-score for each disease prediction model.

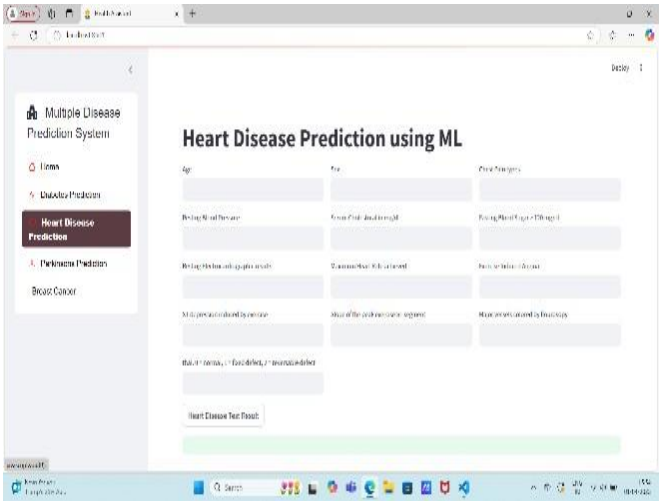
HOME PAGE



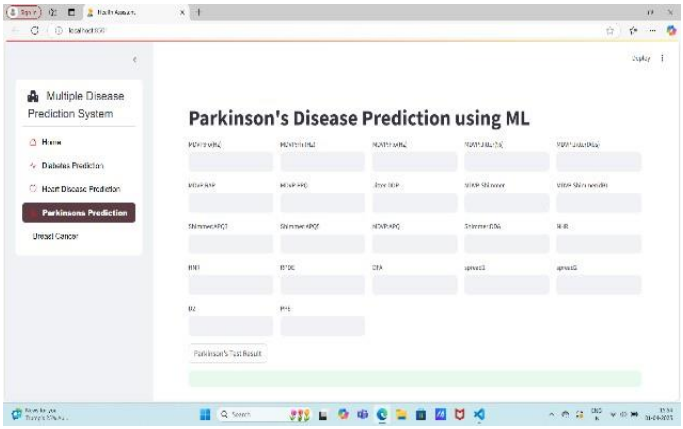
DIABETESE PREDICTION



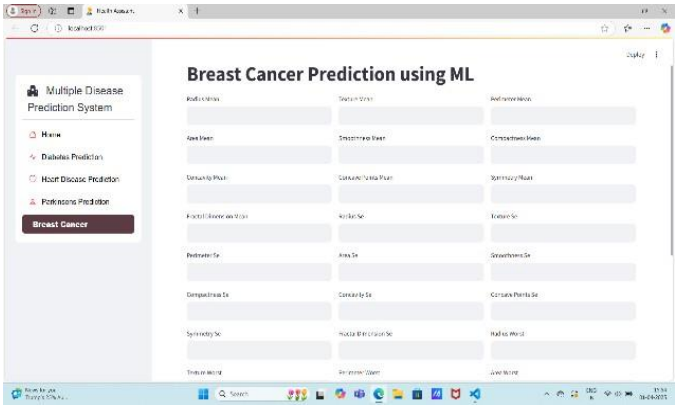
HEART DISEASE PREDICTION



PARKINSON DISEASE PREDICTION



BREAST CANCER DISEASE PREDICTION



6.MODEL PERFORMANCE EVALUATION:

Each model's performance was tested using real-world datasets. The table below shows the accuracy of the models used in this application.

Disease	Machine Learning Model	Accuracy (%)
Diabetes Prediction	Random Forest Classifier	77.5%
Heart Disease Prediction	Logistic Regression	85.2%
Parkinson's Prediction	Support Vector Machine (SVM)	87.1%
Breast Cancer Prediction	XGBoost Classifier	96.7%

7.CONCLUSION & FUTURE WORK

The "health assistant" application is a robust, streamlet-based machine learning tool designed for multiple disease predictions, including diabetes, heart disease, Parkinson's, disease, and breast cancer. By leveraging pre- trained models stored in the saved_models directory, the app ensures quick and efficient predictions based on user inputs.

The application follows a well-structured flow: users select a disease category from the sidebar, input relevant medical data, and receive a diagnosis after processing. Each disease prediction is powered by a corresponding machine learning model loaded using pickle. The models analyze user inputs, classify the likelihood of disease presence, and display results instantly through the streamlet interface. This system enhances accessibility to preliminary health assessments without requiring in-depth medical knowledge. However, it is not a replacement for professional medical advice. Instead, it serves as an initial screening tool to encourage users to seek proper medical consultation if necessary.

REFERENCE:

General References on AI in Healthcare

1. Esteva, A., et al. (2019). "A Guide to Deep Learning in Healthcare." *Nature Medicine*, 25(1), 24-29.
2. Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books.Rajpurkar, P., et al. (2017). "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." *arXiv preprintarXiv:1711.05225*. Machine Learning & Disease Prediction Models
3. Deo, R. C. (2015). "Machine Learning in Medicine." *Circulation*, 132(20), 1920-1930.
4. Krittanawong, C., et al. (2021). "Machine Learning and Deep
5. Learning in Cardiovascular Health: Applications, Challenges, and Future Directions." *The Lancet DigitalHealth*, 3(3), e157-e168.
6. Uddin, S., et al. (2019). "Comparing Different Supervised Machine Learning Algorithms for DiseasePrediction." *BMC Medical Informatics and Decision Making*, 19(1), 281.
7. Streamlit and Web-Based AI Applications
8. Streamlit Team. (2022). *StreamlitDocumentation: An Open-source App Framework for Machine Learning and Data Science*.
9. Gruber, T. (2020). "Rapid Prototyping AI Solutions with Python and Streamlit." *Journal of Computational Science & Engineering*, 15(4), 67-78.
10. Disease-Specific Machine Learning Models:Nguyen, Q. H., et al. (2018). "Diabetes Prediction Using Machine Learning Approaches." *International Journal of Medical Informatics*, 117, 65-75.
11. Wang, H., et al. (2019). "Predicting Heart Disease Risk Using Machine Learning Models." *Scientific Reports*, 9(1), 11087.
12. Shahid, N., et al. (2020). "Early Prediction of Parkinson's Disease
13. Using ML Techniques." *IEEE Access*, 8, 147969-147981.
14. Saba, T., et al. (2021). "Breast Cancer Detection Using Machine Learning Models and Imaging Data. *Computers in Biology and Medicine*, 131, 104245