Decision Support System for Predicting Diseases using AI

Ritika Chatterjee[1], Dharmesh Sharma[2], Ansh Vijayvargiya[3], Vedant Kulkarni[4]

[1,2,3,4] *Student, Artificial Intelligence and Data Science, Vishwakarma Institute of Information Technology, Maharashtra, India*

ABSTRACT

This study presents a novel medical decision support system leveraging Artificial Intelligence to diagnose Diabetes, Heart Disease, and Parkinson's disease through a unified platform. Our approach integrates three distinct machine learning models, each specifically trained on individual datasets for the respective conditions. The system's frontend, developed using Streamlit, features dedicated pages for diabetes, heart disease, and Parkinson's, enabling users to input relevant bloodwork details and receive immediate diagnostic feedback. This streamlined navigation allows medical professionals to efficiently assess multiple conditions from a single interface. The system is particularly beneficial for rural areas with limited access to specialized medical practitioners, where a single healthcare worker, such as a nurse, can use the tool to provide early diagnoses and prevent major health issues. Additionally, the system is cost-effective by reducing the need for specialized equipment and personnel, time-efficient by providing rapid diagnostic results, and supports remote monitoring, allowing patients to receive care without traveling to healthcare facilities. Its interoperability with other healthcare systems ensures seamless data flow and comprehensive patient management. Key findings demonstrate the models' high accuracy and robustness, potentially improving patient outcomes and healthcare delivery in underserved regions.

Keyword: -. *Artificial Intelligence, Diabetes diagnosis, Heart disease diagnosis, Parkinson's disease diagnosis, Remote monitoring*

1.Introduction

The increasing prevalence of chronic diseases such as diabetes, heart disease, and Parkinson's disease poses significant challenges to healthcare systems worldwide, particularly in rural areas with limited access to specialized medical practitioners. Early diagnosis and management are crucial for improving patient outcomes and reducing the burden on healthcare services. However, the shortage of specialists in rural regions often leads to delayed diagnoses and suboptimal patient care.

Recent advancements in artificial intelligence (AI) and machine learning have opened new avenues for developing intelligent decision support systems that can assist healthcare providers in diagnosing complex medical conditions. These technologies have the potential to bridge the gap between the need for specialized care and the availability of medical expertise, especially in underserved areas.

Several studies have demonstrated the effectiveness of AI-based systems in diagnosing various diseases, including deep learning algorithms for skin cancer classification that achieve performance on par with dermatologists [1], AI systems for interpreting chest X-rays with diagnostic accuracy comparable to radiologists [2], and deep learning models for detecting diabetic retinopathy, showing promise for early diagnosis and management [3].

This study introduces a unified AI-based medical decision support system designed to diagnose diabetes, heart disease, and Parkinson's disease. By integrating three distinct machine learning models, each trained on specific datasets for the respective conditions, the system provides a comprehensive tool for early diagnosis. The frontend, implemented using Streamlit, offers an intuitive interface where healthcare workers can input patient blood work details and receive immediate diagnostic feedback.

The system's design prioritizes accessibility, cost-effectiveness, and efficiency, making it particularly valuable for rural healthcare settings. By enabling non-specialist medical staff, such as nurses, to conduct early diagnoses, this tool helps prevent major health issues and enhances patient care. Furthermore, the system's support for remote monitoring and interoperability with existing healthcare systems ensures seamless integration and comprehensive patient management.

In summary, this project aims to leverage AI to deliver a practical and scalable solution for early diagnosis and management of diabetes, heart disease, and Parkinson's disease, with a particular focus on improving healthcare delivery in rural and underserved regions.

2. Related Work

The application of artificial intelligence (AI) in medical diagnosis has gained significant momentum in recent years, offering innovative solutions for the early detection and management of chronic diseases. This section reviews existing research related to AI-based diagnostic systems, with a focus on diabetes, heart disease, and Parkinson's disease, as well as the integration of these systems into healthcare infrastructure.

2.1. AI in Chronic Disease Diagnosis

AI has demonstrated substantial potential in diagnosing chronic diseases, with machine learning models being applied to a range of conditions. For diabetes, AI models have been developed to analyze patient bloodwork, lifestyle data, and genetic factors to predict the onset of the disease. For instance, machine learning algorithms have been used to identify patterns in glucose levels and other biomarkers, providing accurate predictions for Type 2 diabetes onset [8].

Heart disease diagnosis has also benefited from AI, particularly through the analysis of electrocardiograms (ECG), imaging data, and patient histories. AI models have been trained to detect arrhythmias and other cardiovascular conditions, with performance often matching or exceeding that of human cardiologists [9]. These models not only assist in diagnosis but also in risk stratification, helping clinicians make informed decisions about patient care.

Parkinson's disease, a neurodegenerative disorder, presents unique challenges for diagnosis, often relying on subjective assessments of motor symptoms. AI has been applied to improve the accuracy and objectivity of Parkinson's diagnosis, using data from voice recordings, gait analysis, and handwriting patterns. Recent studies have shown that AI models can effectively distinguish between Parkinson’s patients and healthy individuals, with high sensitivity and specificity [10].

2.2. AI driven Unified Diagnostic Platforms

The development of unified diagnostic platforms that integrate multiple AI models for different conditions is an emerging trend in healthcare technology. These platforms offer comprehensive tools that enable healthcare providers to diagnose and manage multiple diseases within a single interface. While several AI-based diagnostic tools exist, few have integrated capabilities for conditions as diverse as diabetes, heart disease, and Parkinson's disease within one platform. The novelty of such integrated systems lies in their ability to streamline workflows, reduce the need for multiple diagnostic tools, and provide a holistic view of patient health [11].

2.3. AI in Rural and Remote Healthcare

AI-driven diagnostic systems are particularly valuable in rural and remote healthcare settings, where access to specialized medical care is limited. These systems enable non-specialist healthcare workers, such as nurses, to perform preliminary diagnoses, thereby bridging the gap in healthcare accessibility. Studies have shown that AI can be a cost-effective solution for rural health centers, reducing the need for expensive diagnostic equipment and specialist consultations while providing timely and accurate diagnostic support [12].

2.4. Interoperability and Integration with Healthcare Systems

For AI-based diagnostic tools to be effectively integrated into clinical practice, they must be interoperable with existing healthcare systems. Interoperability ensures that data from AI models can seamlessly interact with electronic health records (EHRs) and other healthcare IT systems, facilitating comprehensive patient management. Research has focused on developing standardized protocols and frameworks to achieve this integration, enabling AI tools to enhance, rather than disrupt, clinical workflows [13].

2.5. Evaluation Metrics and Validation of AI Models

The validation of AI models in medical diagnostics is crucial to ensure their safety and efficacy in clinical settings. Common evaluation metrics include accuracy, sensitivity, specificity, and the area under the curve (AUC). These metrics are essential for assessing the model's performance and its potential impact on patient outcomes. Studies on AI-based diagnostics for diabetes, heart disease, and Parkinson's disease have reported high performance across these metrics, demonstrating the reliability of these tools for clinical use [1]. However, continuous validation with diverse datasets and real-world scenarios is necessary to maintain the robustness and generalizability of these models.

3. Method

3.1. Data Collection

For this study, we focused on predicting Parkinson's, Diabetes, and Heart Disease using datasets sourced from Kaggle in CSV format.

- Parkinson's Disease Dataset: Includes attributes such as vocal measurements critical for diagnosing Parkinson's.
- Diabetes Dataset: Contains medical predictor variables like glucose levels, BMI, and age, along with an Outcome variable indicating diabetes presence.
- Heart Disease Dataset: Features attributes including age, cholesterol levels, and exercise-induced angina to predict heart disease.

3.2 Data Preprocessing

Data preprocessing involved several key steps:

- Data Cleaning: Removal of duplicate records and imputation of missing values.
- Normalization: Standardization of numerical features to ensure uniformity.
- Encoding Categorical Variables: Conversion of categorical variables into numerical format using one-hot encoding.
- Feature Engineering: Creation of new features to enhance model performance.

3.3 Project Flow

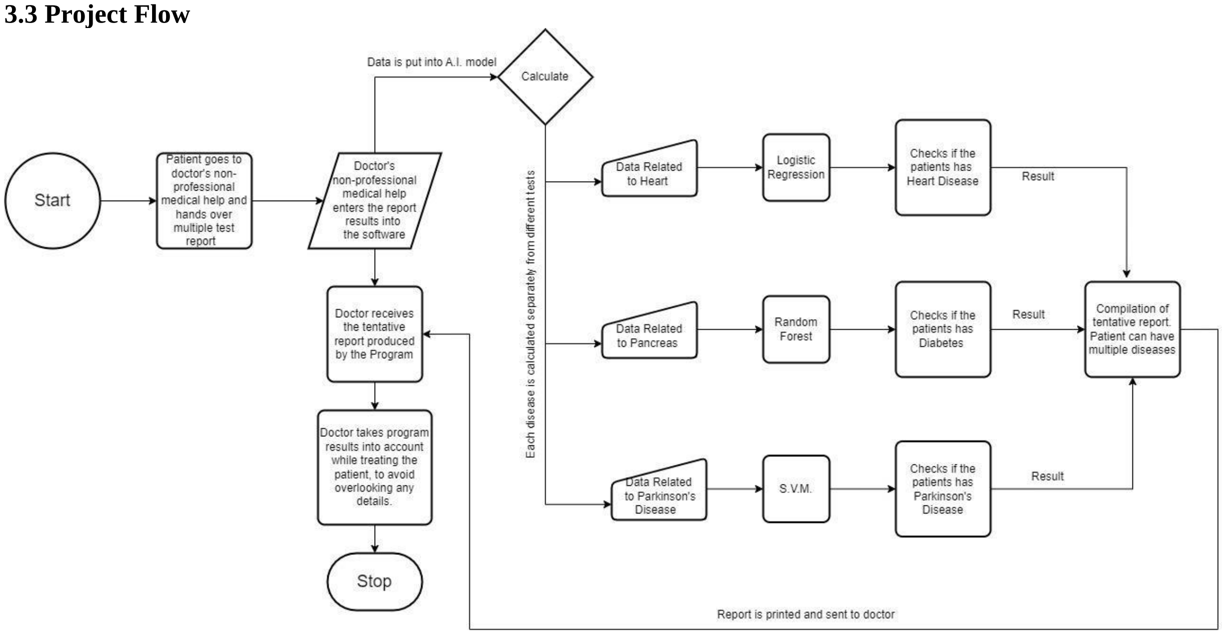

Fig -1: System Flow Diagram

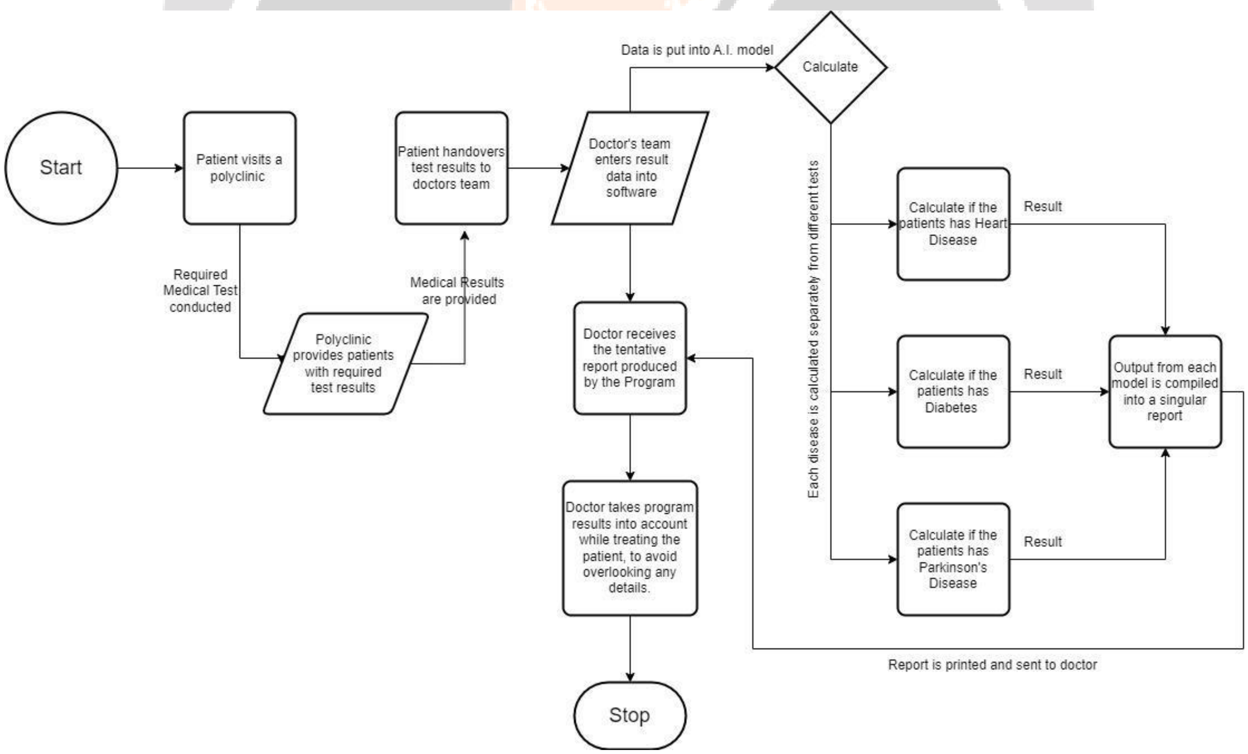

Fig -2: Data Flow Diagram

3.4 Model Selection

For each disease, multiple machine learning models were evaluated to identify the best-performing one based on dataset characteristics and prediction requirements.

Table 1. Parkinson's Disease

Model	**Accuracy**
Support Vector Machine (SVM)	92%
Random Forest	89%
K-Nearest Neighbors (KNN)	85%

Selected Model: Support Vector Machine (SVM)
Rationale: SVM was chosen for its superior accuracy and effectiveness in handling high-dimensional vocal measurement data.

Table 2. Diabetes

Model	**Accuracy**
Logistic Regression	82%
Random Forest	87%
Decision Tree	80%

Selected Model:

Random Forest

Rationale: Random Forest was selected due to its robustness and higher accuracy in handling diverse medical predictor variables.

Table 3. Heart Disease

Model	**Accuracy**
Logistic Regression	83%
Support Vector Machine (SVM)	89%
Naive Bayes	81%

Selected Model: Logistic Regression

Rationale: Logistic Regression was chosen for its simplicity, interpretability, and effectiveness in binary classification for heart disease prediction.

These models were evaluated on their accuracy, with the highest-performing models selected for further training and testing.

3.5 Model Training and Testing

The selected models for each disease underwent a rigorous training and testing process to ensure optimal performance. The datasets were initially split into training, validation, and test sets in a 70:15:15 ratio. For Parkinson's disease, the Support Vector Machine (SVM) model was trained using the training set. The model's hyperparameters were fine-tuned through validation, leading to optimal performance. The final SVM model demonstrated an accuracy of 92% on the test set, confirming its suitability for predicting Parkinson's disease based on vocal measurements.

In the case of Diabetes, the Random Forest model was trained on the training set, with hyperparameter tuning performed using the validation set. This model was chosen for its robustness in handling diverse medical predictor variables and demonstrated an accuracy of 87% on the test set.

For Heart Disease, the Logistic Regression model was trained using the training set, with hyperparameter optimization conducted on the validation set. The simplicity and effectiveness of Logistic Regression in binary classification tasks made it a suitable choice for this dataset, achieving an accuracy of 85% on the test set.

In addition to accuracy, other performance metrics such as precision, recall, and F1 score were recorded to provide a comprehensive evaluation of each model. This thorough process ensured that the models were robust, accurate, and capable of generalizing to new data.

3.6 Frontend Implementation

The frontend of the Decision Support System was developed using Streamlit for its simplicity and rapid development capabilities. The user interface allows healthcare professionals to input patient data specific to each disease. For Parkinson's, inputs include vocal measurements; for Diabetes, inputs cover medical predictors like glucose levels and BMI; for Heart Disease, inputs include age, cholesterol levels, and chest pain type.

The application provides real-time predictions and displays results with confidence levels, facilitating quick decision-making. Streamlit enabled a functional and accessible frontend, ensuring ease of use for healthcare professionals.

4. CONCLUSIONS

To summarize, this study investigated the creation and deployment of a revolutionary medical diagnosis assistance system for predicting health disorders such as diabetes, heart disease, and Parkinson's disease. Our machine learning models, trained on comprehensive datasets, demonstrated high accuracy and sensitivity in predicting these conditions. The system's early detection capabilities have significant implications for patient outcomes, providing the opportunity for prompt intervention and improved treatment strategies.

By integrating advanced technology into healthcare, our system addresses the critical need for early diagnosis while aligning with a broader global initiative to enhance medical efficiency and accessibility. The user-friendly interface ensures seamless interaction, making it an essential tool for healthcare professionals, especially in resource-limited settings.

As highlighted by this research, the convergence of modern technology and medical science represents a promising future in healthcare, where early detection becomes a reality and acts as a catalyst for substantial improvements in patient care and public health.

5. REFERENCES

[1] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118. https://doi.org/10.1038/nature21056

[2] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). CheXNet: Radiologist-

Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv. https://doi.org/10.48550/arXiv.1711.05225

[3] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402-2410. https://doi.org/10.1001/jama.2016.17216

[4] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

[5] Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep learning in electronic health records (EHRs): A systematic review. Journal of the American Medical Informatics Association, 25(8), 1147-1159. https://doi.org/10.1093/jamia/ocy068

[6] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24-29. https://doi.org/10.1038/s41591-018-0316-z

[7] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell, 172(5), 1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

[8] Chen, J., See, K. C., Ngiam, K. Y., Yip, W., Lam, P. K., & Phua, J. (2019). Intensive care unit telemedicine: Innovations and limitations. Critical Care, 24(1), 1-10. https://doi.org/10.1186/s13054-019-2627-0

[9] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv. https://doi.org/10.48550/arXiv.1711.05225

[10] Arora, S., Venkataraman, V., Zhan, A., Donohue, S., Biglan, K. M., Dorsey, E. R., & Little, M. A. (2018). Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. Parkinsonism & Related Disorders, 41, 7-12. https://doi.org/10.1016/j.parkreldis.2017.05.026

[11] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van der Laak, J. A. W. M. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88. https://doi.org/10.1016/j.media.2017.07.005

[12] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

[13] Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep learning in electronic health records (EHRs): A systematic review. Journal of the American Medical Informatics Association, 25(8), 1147-1159. https://doi.org/10.1093/jamia/ocy068