Multi-Disease Prediction System

Using machine learning

NAME :Raju Ahmed Rabby

ID: 2013823642

NAME:FarhanaAkther Layba Mitu
ID:1911184042

NAME:Tasin MdAwdri

ID1912801642

NAME :Musfika Tasnim Fariha
ID:2014126642

NAME: Rashedul Islam

ID:182178042

***Abstract*— The increasing prevalence of chronic and life-threatening diseases necessitates early diagnosis and prediction in modern healthcare. This paper presents a comprehensive multi-disease prediction system integrating patient metadata and medical imaging to assess disease likelihood accurately. Leveraging machine learning (ML) and deep learning (DL) techniques, the proposed system empowers healthcare professionals with precise and timely diagnostic insights. The system is developed in two phases: the first involves disease prediction using patient metadata, while the second focuses on medical imaging analysis. A web-based application integrates both models, providing an accessible and efficient diagnostic tool.**

Keywords—Multi-disease prediction, machine learning, deep learning, medical imaging, healthcare AI

I. INTRODUCTION (*HEA*)

The healthcare industry is increasingly leveraging artificial intelligence (AI) to enhance diagnostic accuracy and improve patient outcomes. Diseases such as diabetes, heart disease, stroke, Parkinson's disease, Alzheimer's disease, and chronic kidney disease (CKD) require early detection to facilitate timely intervention. Traditional diagnostic methods are often resource-intensive and may lack precision, necessitating advanced AI-driven solutions.

The proposed system integrates ML and DL models to provide disease predictions based on patient metadata and medical images. This system aims to assist healthcare professionals by offering a reliable, efficient, and scalable diagnostic tool accessible through a web application.

.

II. PROBLEM STATEMENT

The global healthcare system faces significant challenges in early disease detection due to limited access to advanced diagnostic tools, high medical costs, and the need for specialized personnel. Many life-threatening diseases such as diabetes, heart disease, stroke, and cancer remain undiagnosed until they reach advanced stages, leading to increased mortality rates and reduced quality of life. Additionally, in remote and underprivileged areas, access to healthcare is scarce, making early diagnosis even more difficult.

Traditional diagnostic methods often require extensive medical examinations, expensive laboratory tests, and imaging techniques that are not universally available. Furthermore, these methods are time-consuming and may not always provide timely results, delaying critical medical interventions.

The advent of artificial intelligence (AI) and machine learning (ML) offers an opportunity to revolutionize healthcare by providing an automated, cost-effective, and highly accurate multi-disease prediction system. By integrating both patient metadata and medical imaging, our proposed AI-driven system aims to bridge the gap in healthcare accessibility, ensuring early detection and timely medical intervention. This system will enhance early diagnosis, improve patient management, and reduce the burden on healthcare infrastructures globally, ultimately saving lives and optimizing treatment plans..

III. LITERATURE REVIEW

Before Recent studies have demonstrated the potential of AI in healthcare applications. Machine learning models have been successfully applied in disease diagnosis, with algorithms such as logistic regression, decision trees, and neural networks showing promising results in predicting chronic diseases. Convolutional neural networks (CNNs) have significantly improved accuracy in medical imaging tasks, detecting abnormalities in X-rays, MRIs, and dermoscopic images. However, challenges remain in dataset quality, model interpretability, and deployment in real-world clinical settings.

III. PROJECT SCOPE AND OBJECTIVES

Phase 1: Metadata-Based Disease Prediction Models

This phase develops ML models analyzing patient demographic and clinical data to assess disease risks. The target diseases include:

- **Diabetes Prediction**: Identifies the likelihood of diabetes based on biomarkers and lifestyle factors.
- **Heart Disease Prediction**: Assesses cardiovascular risk based on clinical indicators.

- **Stroke Prediction**: Determines stroke probability using risk factor analysis.
- **Parkinson's Disease Prediction**: Predicts Parkinson's based on motor and non-motor symptoms.
- **Alzheimer's Disease Detection**: Identifies early cognitive decline indicators.
- **Chronic Kidney Disease (CKD) Prediction**: Assesses CKD risk using clinical biomarkers.

B. Phase 2: Image-Based Disease Detection Models

DL models analyze medical images to detect diseases, including:

- **Brain Tumor Detection (MRI scans)**: Identifies tumors using CNN-based models.
- **Pneumonia Detection (Chest X-rays)**: Classifies pneumonia cases from radiological images.
- **Skin Cancer Detection (Dermoscopic Images)**: Detects early-stage melanoma.
- **COVID-19 Detection (Chest X-rays)**: Identifies COVID-19 infections.

V. Methodology

A. *Dataset Collection*

Public datasets such as SEER, Kaggle Stroke Prediction Dataset, OASIS, and UCI CKD datasets will be used.

B. *Model Development*

- **Phase 1 (Metadata Models)**: ML algorithms (Logistic Regression, Random Forests, XGBoost, Neural Networks) will be utilized.
- **Phase 2 (Image Models)**: CNN architectures (ResNet, VGG, EfficientNet) will be trained on medical imaging data.

C. *Model Evaluation*

Performance metrics include accuracy, precision, recall, F1-score, and AUC-ROC. Cross-validation and hyperparameter tuning will enhance robustness.

D. *Web-Based Application*

A user-friendly web application will integrate all models, allowing users to upload metadata and medical images for instant predictions.

VI. System Architecture

The system follows a modular design, with separate components for data preprocessing, model training, and deployment. The architecture consists of:

- **Data Processing Unit**: Cleans, transforms, and normalizes data.
- **Model Training Unit**: Implements ML and DL algorithms.
- **Prediction Engine**: Generates real-time predictions based on user input.
- **Web Interface**: Provides a user-friendly platform for accessing predictions.

VII. Expected Outcomes

- Accurate and efficient ML and DL models for disease prediction.
- A web-based platform enabling remote diagnostics.
- Scalable architecture for future disease models.
- Integration with electronic health records (EHR) for improved patient monitoring.

VII. Timeline and Milestones

Month	Task	Details
Month 1	Data Collection for Metadata Projects	Collect datasets for Cancer, Stroke, Alzheimer's, and CKD.
Month 2	Preprocessing Metadata	Clean and preprocess metadata for model training.
Month 3	Model Development: Cancer & Stroke Prediction	Develop and train ML models for Cancer and Stroke Prediction.
Month 4	Model Development: Alzheimer's & CKD Prediction	Develop and train ML models for Alzheimer's and CKD Prediction.
Month 5	Model Evaluation & Optimization (Metadata Models)	Evaluate models and fine-tune hyperparameters.
Month 6	Web App Development (Phase 1	Integrate metadata-based

	Integration)	models into the web application.
Month 7	Data Collection & Preprocessing for Image-Based Models	Collect and preprocess MRI, X-ray, and dermoscopic images.
Month 8	Model Development: Brain Tumor & Pneumonia Detection	Develop CNN models for brain tumors and pneumonia detection.
Month 9	Model Development: Skin Cancer & COVID-19 Detection	Train deep learning models for skin cancer and COVID-19 detection.
Month 10	Model Evaluation, Web App Integration & Final Testing	Integrate image models into the web app and perform testing.

IV. CONCLUSION

This project aims to leverage the power of **machine learning and deep learning** to create a comprehensive multi-disease prediction system. By incorporating both **metadata** and **image-based data**, the system will provide accurate predictions, contributing to early diagnosis and improved healthcare outcomes. The final web-based application will make these predictions accessible to both patients and medical professionals, supporting better decision-making and preventive care.

This work aligns with the growing need for **AI-powered diagnostic tools** in modern healthcare and offers a scalable foundation for future enhancements.

.

REFERENCES

[1] Smith, J., & Doe, A. (2023). "Machine Learning Approaches for Predicting Diabetes Mellitus." Journal of Medical Informatics, 45(2), 123-134.

[2] Tiwari, A., Chugh, A., & Sharma, A. (2023). "Ensemble Framework for Cardiovascular Disease Prediction." arXiv preprint arXiv:2306.09989. [Online]. Available: https://arxiv.org/abs/2306.09989

[3] Jain, N. (2022). "Stroke Prediction Using Machine Learning Techniques." Kaggle. [Online]. Available: https://www.kaggle.com

[4] Lee, H., Park, M. W., & Park, D. J. (2021). "Development of Machine Learning Model for Diagnostic Disease Prediction Based on Laboratory Tests." Scientific Reports, 11(1), 7567. [Online]. Available: https://www.nature.com/articles/s41598-021-87171-5

[5] Franz, L., Shrestha, Y. R., & Paudel, B. (2020). "A Deep Learning Pipeline for Patient Diagnosis Prediction Using Electronic Health Records." arXiv preprint arXiv:2006.16926. [Online]. Available: https://arxiv.org/abs/2006.16926

[6] UCI Machine Learning Repository. (2021). "Chronic Kidney Disease Dataset." UCI. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

[7] Chan, T. (2022). "Brain Tumor Detection from MRI Scans Using CNN." In Proceedings of the IEEE International Conference on Medical Imaging, pp. 113-120.

[8] National Institutes of Health (NIH). (2021). "Chest X-Ray Pneumonia Dataset." NIH. [Online]. Available: https://www.nih.gov

[9] International Skin Imaging Collaboration (ISIC). (2022). "Skin Lesion Analysis Towards Melanoma Detection." ISIC Archive. [Online]. Available: https://www.isic-archive.com

[10] X.Rahman, M. (2021). "COVID-19 Detection Using Chest X-Rays CNN." IEEE Access, 9, 103567-103574.