A Machine Learning-Based Dual Diagnostic System for Medical Imaging and Symptom-Based Disease Prediction

Dr. Mohammed Arshad [1], Omer Ahmed Quadri[2], Mohammed Furqan Ur Rehman[3], Mirza Shareef Baig[4]
[1]Principal & Head, Department of Computer Science & Engineering, Deccan College of Engineering and Technology, Hyderabad, India
[2,3,4]UG Students, Department of Computer Science & Engineering, Deccan College of Engineering and Technology, Hyderabad, India

***Abstract*—Diagnostic accessibility and accuracy have greatly increased as a result of the application of machine learning (ML) and artificial intelligence (AI) in healthcare. A dual-input diagnostic system that combines symptom-based disease prediction and medical image analysis is presented in this project. While machine learning classifiers analyse patient-reported symptoms to suggest possible diseases, convolutional neural networks (CNNs) are used to classify medical images like brain scans and chest X-rays. For increased prediction accuracy, the symptom analysis pipeline is improved by Natural Language Processing (NLP) techniques. The system is trained on publicly accessible benchmark datasets to guarantee scalability, dependability, and efficiency. It is implemented in Python using frameworks such as TensorFlow, Keras, OpenCV, and Scikit-learn. The goal of this integrated solution is to assist medical professionals with rapid, initial evaluations, particularly in settings with limited resources.**

Index Terms— Convolutional Neural Networks, Artificial Intelligence, Machine Learning, Medical Image Analysis, Natural Language Processing, Symptom-Based Diagnosis

I. INTRODUCTION

Machine Learning (ML) is playing a vital role in modern healthcare by improving the accuracy and efficiency of disease diagnosis. Techniques like Deep Learning (DL) and supervised ML algorithms are increasingly applied for analyzing medical data and assisting clinical decision-making.

This project introduces a dual-module diagnostic system designed to predict diseases using two independent approaches. The first module uses Convolutional Neural Networks (CNNs) to classify medical images such as chest X-rays and brain scans, identifying conditions like tumors, pneumonia, and lung infections. The second module processes patient-reported symptoms and predicts possible diseases using supervised ML models.

Developed using Python and libraries such as TensorFlow, Keras, OpenCV, Scikit-learn, and Flask, the system is trained on public datasets for both medical imaging and symptoms. It aims to provide a scalable, accessible, and efficient diagnostic tool that can assist healthcare institutions and practitioners in making quicker preliminary assessments.

- Key features of this system include:
- CNN-based medical image analysis
- ML-based symptom prediction
- Independent dual-module diagnostic pipelines
- Scalable and cost-effective implementation

II. LITERATURE REVIEW

Recent advancements in Machine Learning (ML) have improved disease diagnosis through medical imaging and symptom analysis. Studies show CNN models perform well in classifying medical images, detecting conditions like tumors and pneumonia. Separately, ML classifiers such as Random Forest and SVM are used for symptom-based disease prediction. Limited research exists on combining both approaches into a unified diagnostic system, highlighting a gap this work addresses. Recent research highlights the potential of combining imaging data with clinical symptom analysis for better diagnostic accuracy. However, most existing works focus on either image-based diagnosis or symptom-based prediction in isolation. Hybrid diagnostic systems integrating both have been limited, often constrained by complexity, data privacy concerns, and lack of accessible, scalable solutions.

This project addresses these gaps by implementing a dual-input ML-based diagnostic tool, integrating image analysis and symptom interpretation within a unified, efficient, and user-friendly framework.

III. EXISTING SYSTEM

Current diagnostic support systems typically rely on single-modality analysis. Some tools use deep learning models like CNNs for medical image classification, while others employ ML classifiers such as Random Forest, SVM, and Naive Bayes for analyzing symptom-based datasets.
A few advanced systems have attempted integrating these approaches but remain either too complex for widespread use or dependent on expensive hardware and proprietary software.

A. LIMITATIONS

- Use Most systems handle medical imaging and symptom analysis separately.
- Lack of integrated platforms combining both data sources.
- High dependence on high-end computational resources.
- Limited accessibility in under-resourced regions.
- Existing tools often have closed-source constraints and high licensing costs.
- Insufficient focus on user-friendly interfaces for medical and non-expert users.

IV. PROPOSED SYSTEM

The proposed system introduces a dual-input ML-based diagnostic solution that processes both medical images and patient-reported symptoms. It employs CNN models for classifying diseases from images such as chest X-rays, brain scans, and lung images. In parallel, an NLP pipeline combined with ML algorithms interprets symptoms input by users to suggest potential diagnoses.

Key highlights of the system:

- CNN-based deep learning models for image classification.
- ML models (Random Forest, SVM, Naive Bayes) for symptom analysis.
- A modular, scalable architecture developed in Python using open-source libraries like TensorFlow, Keras, OpenCV, Scikit-learn, and Flask.
- Trained on benchmark open-source datasets for robustness and performance.
- Designed for deployment across hospitals, clinics, and remote healthcare centers.
- Cost-effective, easily maintainable, and privacy-compliant, ensuring data security.

This integrated approach enhances diagnostic reliability, optimizes resource usage, and improves healthcare accessibility, particularly in areas lacking advanced diagnostic infrastructure.

VI. IMPLEMENTATION

In implementation firstly in the ML based symptom Based disease Prediction user has to enter his/her details such as name, age, gender, and contact information. This helps maintain a record of user profiles for future reference and tracking purposes. Once the details are submitted, the admin reviews and verifies the user profile. After approval, the user is granted access to the prediction module. The user can then enter details of their symptoms through a simple form-based interface. Based on the entered symptoms, the trained machine learning model processes the input and predicts the most likely disease along with its probability score.

Similarly, in the Medical Image Processing module, the user is required to register by providing basic personal information. The admin again validates the profile, ensuring proper user management and system security. Once permission is granted, the user can upload their medical image, such as a chest X-ray or brain scan. The deep learning model — trained using Convolutional Neural Networks (CNNs) — analyzes the uploaded image and predicts the possible medical condition (e.g., pneumonia, lung infection, or tumor) with a corresponding confidence level.

This structured workflow not only ensures proper record-keeping and access control but also streamlines the diagnostic process by integrating machine learning-based symptom analysis with medical image interpretation under a controlled and user-friendly environment.

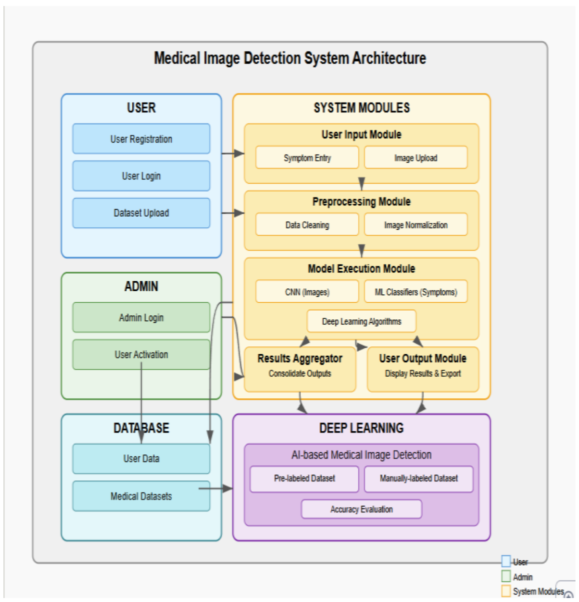

Fig 1. System Architecture

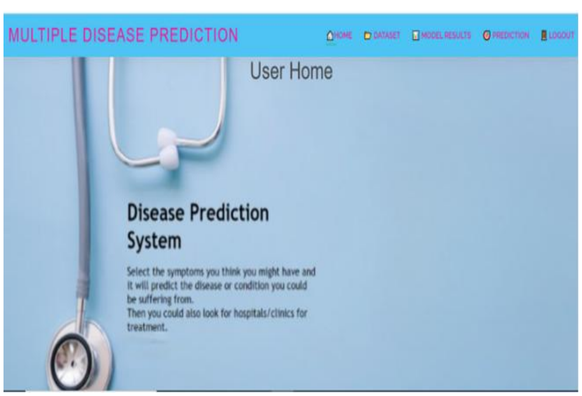

Fig 2 . Home page of symptom Based disease prediction

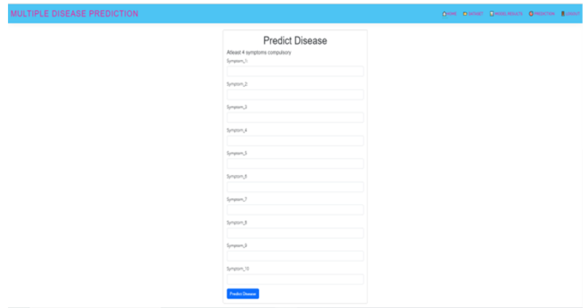

Fig 3. . Prediction of symptom Based disease prediction

Fig 4 . Prediction result page

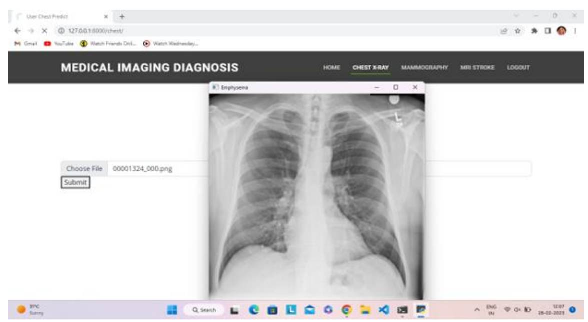

Fig 5. Home page of medical image analysis

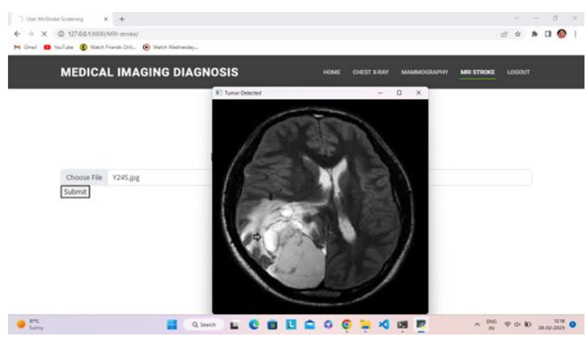

Fig 5. Image selection page

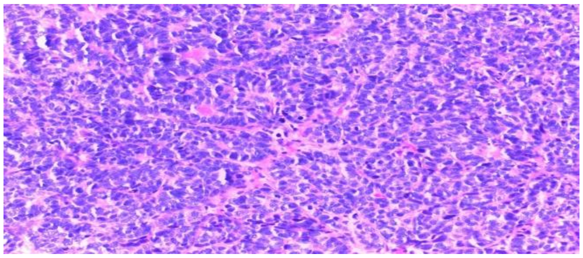

Fig 6. Expected output

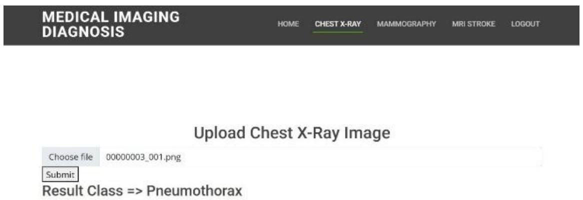

Fig 7. Prediction result page.

VII. CONCLUSION

This project successfully demonstrates a dual-module diagnostic system that combines machine learning-based symptom prediction with medical image analysis for efficient disease detection. By integrating these two approaches, the system offers a practical, scalable, and accessible tool for preliminary medical assessment. Developed using open-source technologies and trained on publicly available datasets, it provides faster, cost-effective, and reliable diagnostic support. This system can assist healthcare professionals in early diagnosis and decision-making, especially in remote or under-resourced areas. Future improvements may include expanding disease categories, enhancing model accuracy, and integrating real-time data for broader clinical applications.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Dr. P.Vishwapathi , Dr. Mohammed Arshad for his valuable guidance, support, and encouragement throughout the development of this project. His insights and expertise played a significant role in shaping the direction and execution of this work. The authors also thank the faculty and staff of Deccan College of Engineering and Technology for providing the necessary resources and a productive research environment.

REFERENCE

[1] S. Rajpurkar, J. Irvin, K. Zhu, et al., “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2097–2105, Jul. 2017.

[2] P. Mooney, “NIH Chest X-Ray Dataset Documentation,” *NIH Clinical Center*, 2017. [Online]. Available: https://nihcc.app.box.com/v/ChestXray-NIHCC.

[3] A. Cohen, P. Morrison, and L. Dao, “COVID-19 Image Data Collection,” *arXiv Preprint*, Apr. 2020.

[4] B. Menze, A. Jakab, et al., “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[5] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[6] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[7] X. Zhang, Y. Zhang, and Y. Huang, “Naive Bayes based disease classification using symptoms,” *J. Biomed. Inform.*, vol. 55, pp. 90–100, Mar. 2015.

[8] F. Chollet, *Deep Learning with Python*, 1st ed., Greenwich, CT: Manning, 2018, ch. 4, pp. 100–124.

[9] A. Paszke et al., “Automatic differentiation in PyTorch,” *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 8024–8035, Dec. 2019.

[10] M. Abadi et al., “TensorFlow: Large-scale Machine Learning on Heterogeneous Distributed Systems,” *Proc. USENIX Symp. Oper. Syst. Des. Implement.*, pp. 265–283, Feb. 2016.

[11] A. Gupta, “OpenCV: Computer Vision with Python,” *Packt Publishing*, 2017.

[12] P. Vishwapathi, “Flask-based Web Deployment for ML Models,” *Proc. Int. Conf. Eng. Res.*, pp. 45–49, Jan. 2021.

[13] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. 22nd ACM SIGKDD*, pp. 785–794, Aug. 2016.

[14] J. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine,” *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[15] C. Deng et al., “UCI Machine Learning Repository: Diabetes Data,” Univ. of California, Irvine. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/diabetes.

[16] B. Van Rossum and F. Drake, *Python Reference Manual*, Amsterdam: Python Software Foundation, 2001.
[17] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May 2007.