



Medical Disease Prediction Using X-Ray Images

S Mohanappriya, S Kavin, GD Boopalan, M Ajay Yaswanth

Kamaraj College of Engineering & Technology, Virudhunagar, Tamil Nadu, India

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03.2026; Published: 28.03.2026.

ABSTRACT: This project presents a machine learning-based disease prediction system designed to identify heart disease and pneumonia at an early stage. The system integrates two different types of data, including structured clinical data and medical image data, to provide accurate predictions. For heart disease prediction, clinical parameters such as age, blood pressure, and cholesterol levels are analyzed using the Random Forest algorithm, which improves accuracy by combining multiple decision trees. For pneumonia detection, chest X-ray images are processed using a Convolutional Neural Network (CNN), which is effective in identifying patterns and features in medical images. The system includes a data preprocessing stage where clinical data is cleaned, normalized, and encoded, while image data is resized, normalized, and augmented to enhance model performance. A user-friendly interface is developed using a Flask web application, allowing users to easily input data or upload images and receive results. The system runs locally without the need for external deployment tools, making it simple and efficient to use. The prediction results are displayed clearly, indicating the risk level of heart disease and the presence of pneumonia. Additionally, the system supports data storage for future reference and improvement. This project reduces the time required for diagnosis and minimizes human error, thereby assisting healthcare professionals in decision-making. It also makes healthcare technology more accessible, especially in areas with limited medical resources. The modular design allows future enhancements, such as adding more diseases and improving model accuracy. Overall, this project demonstrates the practical application of artificial intelligence in healthcare by providing a fast, reliable, and easy-to-use disease prediction system.

KEYWORDS: Random Forest Algorithm, Convolutional Neural Network (CNN), Chest X-ray Pneumonia Dataset (Chest X-ray Images for Pneumonia – Kaggle).

I. INTRODUCTION

In recent years, the use of artificial intelligence in healthcare has increased significantly, helping improve the accuracy and speed of disease diagnosis. Early detection of serious diseases like heart disease and pneumonia is very important to reduce health risks and save lives. Traditional diagnostic methods often depend on manual analysis and expert knowledge, which can be time-consuming and sometimes inaccurate. With the growth of machine learning, automated systems can now assist doctors in analyzing medical data efficiently. This project focuses on developing a smart system that supports early and reliable disease prediction.

The proposed system combines clinical data analysis and medical image processing to provide better diagnostic results. It uses the Random Forest algorithm for predicting heart disease based on patient data and a Convolutional Neural Network (CNN) for detecting pneumonia from X-ray images. The system is implemented as a web application, making it easy for users to interact and obtain results. Proper data preprocessing techniques are applied to improve the accuracy of predictions. Overall, the project aims to create a simple, efficient, and accessible tool that enhances healthcare decision-making.

II. LITERATURE REVIEW

Many researchers have worked on disease prediction systems using machine learning techniques to improve early diagnosis. Previous studies show that algorithms like Decision Trees, Support Vector Machines, and Random Forest are widely used for predicting heart disease. These models analyze clinical data such as age, blood pressure, and cholesterol levels to identify risk factors. Among them, Random Forest provides better accuracy due to its ensemble learning approach. This makes it a suitable choice for reliable heart disease prediction.

In the field of medical image analysis, deep learning models, especially Convolutional Neural Networks (CNN), have shown significant improvements. Researchers have applied CNN models to detect pneumonia from chest X-ray images



with high accuracy. These models automatically extract important features such as patterns and textures from images. Compared to traditional image processing methods, CNN reduces the need for manual feature extraction. This has made deep learning a powerful tool in medical imaging applications.

Several studies have also focused on combining different types of data, such as clinical records and medical images, for better prediction results. Hybrid systems that use both machine learning and deep learning techniques have shown improved performance. Data preprocessing plays a crucial role in these systems by cleaning, normalizing, and preparing data for model training. Proper preprocessing helps in reducing noise and improving model accuracy. These approaches highlight the importance of integrating multiple techniques in healthcare systems.

Recent research emphasizes the development of user-friendly web-based applications for healthcare solutions. Many systems are implemented using frameworks like Flask to provide easy interaction between users and prediction models. These applications allow users to input data, upload images, and receive instant results. However, some systems require complex deployment tools, which can limit accessibility. This project addresses that limitation by creating a simple, locally executable system that is efficient and easy to use.

III. RESEARCH METHODOLOGY

The methodology of this project focuses on using the Random Forest algorithm to predict heart disease based on clinical data. Initially, a suitable dataset containing patient information such as age, blood pressure, cholesterol level, and other medical attributes is collected. This data serves as the input for the prediction model. Since raw data may contain errors or inconsistencies, it cannot be used directly for training. Therefore, proper data preparation is required before applying the algorithm.

In the next step, data preprocessing is performed to clean and organize the dataset. This includes handling missing values, removing duplicate entries, and converting categorical data into numerical form using encoding techniques. The dataset is also normalized to ensure that all values are within a similar range. These preprocessing steps improve the quality of the data and help the model learn more effectively. Feature selection is also carried out to identify the most important attributes that influence heart disease prediction.

After preprocessing, the Random Forest algorithm is applied to train the model. Random Forest works by creating multiple decision trees using different subsets of the data and features. Each tree makes its own prediction, and the final output is determined by majority voting. This approach reduces overfitting and increases accuracy compared to a single decision tree. The model is trained using a portion of the dataset and then tested on unseen data to evaluate its performance.

Finally, the trained model is integrated into a Flask-based web application for user interaction. Users can enter their clinical details through the interface, and the system processes the input to generate predictions. The results are displayed clearly, indicating the likelihood of heart disease. This methodology ensures an efficient, accurate, and user-friendly system for disease prediction. It also allows future improvements by incorporating more data and advanced techniques.

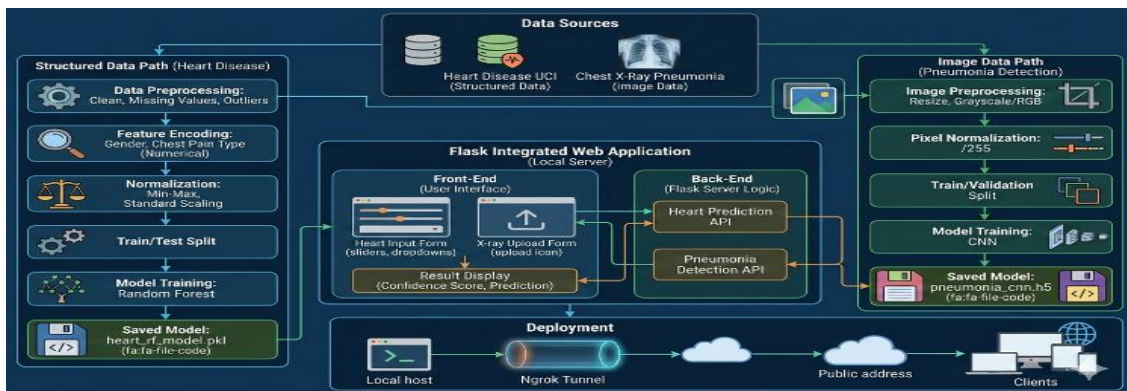


FIG:1



Dataset Collection – Heart disease clinical data and chest X-ray image datasets are collected from reliable medical sources such as Kaggle and public healthcare repositories. These datasets contain patient health parameters and labeled X-ray images used for training the machine learning models.

Data Preprocessing – The collected dataset is cleaned by removing missing, duplicate, or incorrect values to improve data quality. Normalization and scaling techniques are applied to ensure that the data is suitable for machine learning algorithms.

Feature Encoding – Categorical attributes such as gender, chest pain type, and other non-numerical data are converted into numerical format. This allows machine learning models to process and analyze the data effectively.

Image Preprocessing – Chest X-ray images are resized to a fixed dimension and normalized to improve model performance. Image preprocessing also helps in reducing noise and preparing the images for CNN training.

Model Selection – The Random Forest algorithm is selected for heart disease prediction because of its high accuracy and ability to handle structured data. A Convolutional Neural Network (CNN) is chosen for pneumonia detection due to its effectiveness in analyzing image patterns.

Model Training – The Random Forest model is trained using the processed clinical dataset to learn patterns related to heart disease. Similarly, the CNN model is trained with chest X-ray images to recognize pneumonia infection patterns.

Model Testing – After training, both models are tested using separate testing datasets. This step evaluates the accuracy, precision, and performance of the models in predicting diseases.

Model Integration – The trained machine learning models are integrated into a web application using the Flask framework. This integration allows the models to interact with the user interface and process user inputs.

User Input – Users can enter patient health details such as age, blood pressure, and cholesterol levels for heart disease prediction. They can also upload chest X-ray images for pneumonia detection.

Prediction Process – The system processes the user input and sends it to the appropriate machine learning model. The model analyzes the data and generates a prediction result.

Result Display – The prediction results are displayed on the web interface in an easy-to-understand format. The system may also show the probability or confidence level of the prediction.

Deployment – The web application is deployed using Flask and for real-time accessibility. This allows users to access the system through a browser and obtain predictions instantly.

IV. RESULTS AND DISCUSSION

The system successfully predicts the risk of heart disease using the Random Forest algorithm based on clinical parameters. It also detects pneumonia from chest X-ray images using a Convolutional Neural Network (CNN). The prediction results are displayed through a user-friendly web interface. The models show reliable performance with good accuracy in disease prediction. This helps in faster diagnosis and supports healthcare professionals in making better medical decisions.

The developed system demonstrates how machine learning can assist in the early prediction of diseases such as heart disease and pneumonia. The Random Forest model effectively analyzes clinical parameters and provides reliable predictions for heart disease risk. The CNN model successfully identifies pneumonia patterns from chest X-ray images. Proper data preprocessing and feature extraction improve the performance of both models. Integrating these models into a Flask-based web application makes the system easy to use. The web interface allows users to input data and obtain predictions quickly. The system reduces the time required for manual diagnosis and supports medical professionals in decision-making. Overall, the project highlights the potential of artificial intelligence in improving healthcare diagnosis and efficiency.



Heart Disease Prediction

Age: Gender (1: Male, 0: Female):

Chest Pain Type (1-4): Resting Blood Pressure:

Cholesterol (mg/dl): Fasting Blood Sugar (1: True, 0: False):

Resting ECG (0, 1, 2): Max Heart Rate:

Exercise Induced Angina (1: Yes, 0: No): ST Depression:

Slope of Peak (1-3): Major Vessels (0-3):

Thal (3: normal; 6: fixed defect; 7: reversible defect):

FIG:2

Prediction Result

There is a low risk of heart disease.

Factors Contributing to Prediction

- High fasting blood sugar is linked to heart disease.

Advice

- Control blood sugar with a healthy diet and regular exercise.

Pneumonia Detection

Patient Name:

Age:

Sex:

Weight (kg):

Upload Chest X-ray Image: No file chosen

FIG:4

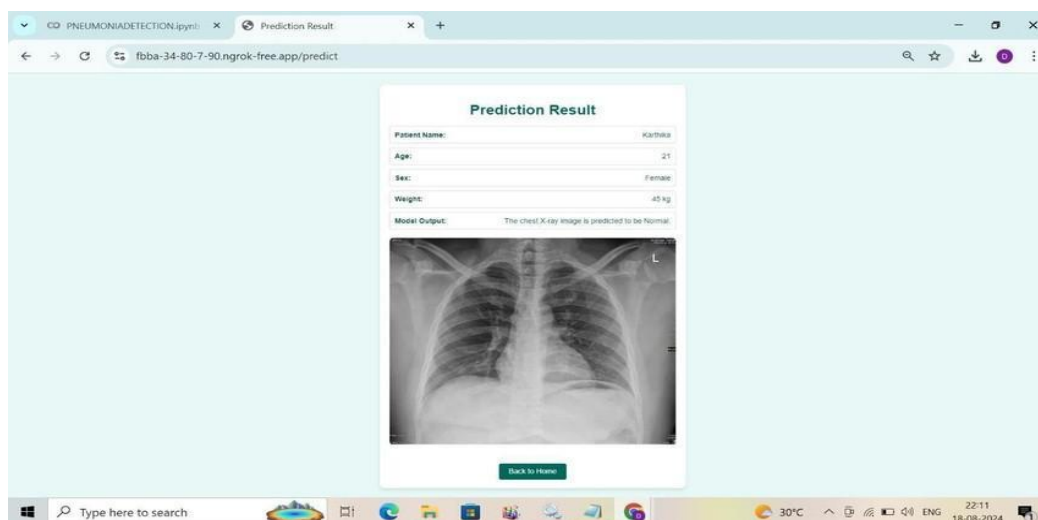


FIG:5

V. CONCLUSION

The project successfully develops a machine learning–based system for predicting heart disease and pneumonia in an efficient and user-friendly manner. It combines the Random Forest algorithm for analyzing clinical data and a Convolutional Neural Network (CNN) for processing chest X-ray images to ensure accurate predictions. The system reduces the time required for diagnosis and minimizes the chances of human error. It is implemented as a Flask web application and runs locally, making it simple and accessible without the need for complex deployment tools. Proper data preprocessing techniques are applied to improve the quality and reliability of the results. The system provides clear outputs, helping users understand the prediction easily. It can assist healthcare professionals in making faster and better decisions. The design of the system is flexible and can be extended to include more diseases in the future. This project demonstrates the practical use of artificial intelligence in real-world healthcare applications. Overall, it highlights how technology can make disease prediction faster, more accurate, and more accessible.

VI. FUTURE WORK

1. **Addition of More Diseases** -The system can be extended to include prediction of other diseases such as diabetes and cancer..This will make the application more comprehensive and useful.It will help in providing a multi-disease diagnosis platform
2. **Improving Model Accuracy** - Advanced algorithms and larger datasets can be used to improve prediction accuracy.Fine-tuning models and using hybrid techniques can enhance performance.This will lead to more reliable and precise results.
3. **Cloud-Based Deployment** - The system can be deployed on cloud platforms for global accessibility.Users can access the application from anywhere using the internet.This removes the limitation of running only on a local system.
4. **Real-Time Data Integration** - Integration with real-time health monitoring devices can be added.Data from wearable devices can be directly used for prediction.This helps in continuous health monitoring and early detection.
5. **Mobile Application Development** - A mobile app version of the system can be developed for easy access.Users can input data and get results directly from their smartphones.This increases usability and makes the system more user-friendly.

REFERENCES

1. Al-Alshaikh, H. A., et al., “Comprehensive Evaluation of Machine Learning in Heart Disease Prediction,” *Scientific Reports*, 2024.
2. Rehman, M. U., et al., “Predicting Coronary Heart Disease with Advanced Machine Learning Classifiers,” *Scientific Reports*, 2025.
3. Mohammad, F., & Al-Ahmadi, S., “WT-CNN: Hybrid Model for Heart Disease Prediction,” *Mathematics Journal*, 2023.



4. Shin, H. J., et al., "AI-Based Chest Radiograph Model for Pneumonia Prediction," *Scientific Reports*, 2024.
5. Bikku, T., et al., "Deep Learning-Based Diagnosis of Respiratory Diseases using CNN-RNN," *Scientific Reports*, 2025.
6. Karandikar, A., et al., "Hybrid CNN Model for Heart Disease Prediction," *Journal of Electrical Systems*, 2024.
7. Heliyon Journal, "Early Heart Disease Prediction using Machine Learning," *ScienceDirect*, 2024.
8. C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, *Electric Power Components and Systems*, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
9. C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - *Journal of Electrical Engineering*, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2
10. C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, *Electrical Engineering*, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
11. S.Tamilselvi, R.Prakash, C.Nagarajan, "Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller" *Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering*, DOI10.1007/s40998-025-00917-z,2025
12. S.Tamilselvi, R.Prakash, C.Nagarajan, "Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance" *Electric Power Systems Research* 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428
13. S.Thirunavukkarasu, C. Nagarajan, 2024, "Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," *Journal of Electrical Engineering And Technology*, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
14. C. Nagarajan, M.Madheswaran and D.Ramasubramanian- 'Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model'- *Acta Electrotechnica et Informatica Journal* , Vol.13 (2), pp.18-31, April-June.2013, DOI: 10.2478/aei-2013-0025.
15. C.Nagarajan and M.Madheswaran - 'DSP Based Fuzzy Controller for Series Parallel Resonant converter'- Springer, *Frontiers of Electrical and Electronic Engineering*, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
16. C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- *Iranian Journal of Electrical & Electronic Engineering*, Vol.8 (3), pp.259-267, September 2012.
17. C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
18. Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", *Revista Materia (Rio J.)* Vol. 30, 2025, DOI: <https://doi.org/10.1590/1517-7076-RMAT-2024-0721>
19. M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", *Journal of Environmental Protection and Ecology*, Volume 23, Issue 2, pp: 520-530,2022
20. *Expert Systems with Applications*, "Quantum CNN for Heart Disease Prediction," *ScienceDirect*, 2024.
21. Mabrouk, A., et al., "Pneumonia Detection using Ensemble CNN Models," *arXiv*, 2023.
22. Hao, R., et al., "Hybrid CNN-Transformer Model for Heart Disease Prediction," *arXiv*, 2025.
23. Rajasekar, M., (2024). Secure Digital Banking with Federated AI: An AWS Cloud-Based Predictive Analytics Architecture for Financial Risk Intelligence. *International Journal of Research and Applied Innovations*, 7(3), 10735-10740.
24. Anand, L., Tyagi, R., & Mehta, V. (2024, January). Food recognition using deep learning for recipe and restaurant recommendation. In *Proceedings of Eighth International Conference on Information System Design and Intelligent Applications* (pp. 269-279). Singapore: Springer Nature Singapore.
25. Kiran, A., Rubini, P., & Kumar, S. S. (2025). Comprehensive review of privacy, utility and fairness offered by synthetic data. *IEEE Access*.