DiagniQ: An AI-Based Multi-Disease Prediction System Using Hybrid Machine Learning and Deep Learning Models

Saara Goyal[1], *Aastha Maheshwari[1], Varun Srivastava[1] and Rajiv Kumar Mishra[1]

[1] Jaypee Institute of Information Technology, Noida, India
*Aasthamaheshwari.am@gmail.com, mishrarajiv99@gmail.com

Abstract. The timely diagnosis of chronic and life-threatening diseases is crucial to enhancing patient outcomes and lowering the rate of burden in the healthcare systems. The paper introduces DiagniQ, which is an artificial intelligence (AI) multi-disease prediction algorithm that helps to identify breast cancer, lung cancer, diabetes and heart disease at the early stages. The proposed system combines both machine learning and deep learning to work with structured clinical data as well as medical images. Image-based disease prediction uses Convolutional Neural Networks (CNN) and Residual Neural Networks (ResNet), whereas structured datasets are analyzed with the help of Support Vector Machines (SVM) and Decision Tree classifiers. The suggested system is a multi-disease prediction system that is integrated and unified. DiagniQ accepts two types of input channels that can be entered in by a user manually or by submitting a structured medical report. Experimental analysis shows the potential of DiagniQ though with all diseases which means that DiagniQ can be a useful early check and decision-support system that facilitates preventive health and medical timely consultation.

Keywords: Convolutional Neural Networks (CNN), Residual Neural Networks (ResNet), Support Vector Machines (SVM) and Decision Tree classifiers.

Introduction

The high rate of chronic and non-communicable diseases has become one of the greatest challenges affecting health care systems in the global arena. Breast cancer, lung cancer, diabetes, and heart disease are some of the major causes of morbidity and mortality in the world on average per year. It is always clinically proven that early diagnosis and prompt treatment is very important in enhancing the survival rates, minimizing the complexity of treatment and decreasing the overall healthcare costs. [1] Although medical technology has advanced, early diagnosis is an area that is inaccessible to a big section of the population. Conventional methods of diagnosis usually involve specialised medical equipment, trained medical practitioners and expensive laboratory tests. The delays in the detection of the disease and the ineffectiveness of treatments in such situations are common in remote and low-resource

A. Agnihotri et al. (eds.), *Proceedings of the Conference on Bridging Engineering Disciplines with AI and Machine Learning (BEDAIML 2026)*, Advances in Intelligent Systems Research 209,
https://doi.org/10.2991/978-94-6239-697-5_20

regions (Machiavelli, 2009). As a result, diseases are often diagnosed at a later stage when treatment methods are not many and effective.
The recent advances in the field of artificial intelligence (AI), especially machine learning (ML) and deep learning (DL), have allowed analyzing complex medical data with the help of the automated method. The AIs have proven to be able to work with large datasets, detect latent patterns, and also help in clinical decision-making with high precision. Deep learning models have demonstrated good results in the analysis of medical images, whereas classical machine learning algorithms are still useful in the analysis of structured clinical data.
Nevertheless, generally, existing AI diagnostic instruments have been designed to address a single disease, or employ a single type of data. This is a small scale method that is difficult to expand and restricts their practical application in practice in a real healthcare facility. DiagniQ is suggested to fill this gap and bring together heterogeneous AI models into a single framework in order to predict multiple diseases jointly.

1. Literature Survey

The use of artificial intelligence in healthcare has received considerable interest because it can process complicated medical information and help to identify disease before it occurs. Both machine learning and deep learning approaches have been studied in the prediction of various diseases, and the techniques have proven to be more accurate and also more efficient than the conventional methods of diagnosis.
Image based disease prediction has broadly been done using deep learning models. Convolutional Neural Networks (CNNs) have demonstrated excellent performance in detection of breast cancer when using histopathological images since the technique is capable of extracting spatial and texture-based features automatically. These models are based less on manually devised characteristics and are more dependable in various forms of imaging scenarios[1], [2].
Research has suggested that CNN-based methods show high classification accuracy when they are contrasted with traditional image-processing methods. On the same note, RNs have successfully been implemented in detecting lung cancer with CT scan images [3], [4].
ResNet has deep architecture and residual connections, which allow learning features effectively and enhancing convergence, thus making it appropriate to learn high-resolution medical images.
Classical machine learning algorithms have been shown to be useful in the case of diseases that are dependent on structured clinical information. The diabetes prediction is commonly performed by Support Vector machines (SVMs) because they allow manipulating high dimensional data and non-linear boundaries of the decisions. [5], [6].
Studies have demonstrated that SVM-based models can have a high degree of reliability and accuracy with regard to patient attributes information like the level of glucose, body mass index and insulin values on the dataset. Moreover, Decision Tree classifiers have found wide application in prediction of heart diseases. They are also useful in medical

decision-support systems due to their capacity to model non-linear relationships and give interpretable decision rules. [7], [8]. The existing systems can identify only one disease and can act autonomously which restricts their application in the real world. This explains why a single AI based tool such as DiagniQ could integrate image processing and organized data to assist in the early detection of various diseases.

2. Proposed Method

The offered solution DiagniQ should be an AI-based multi-disease prediction system combining machine learning and deep learning models in the framework of a single web based architecture. The overall goal of the model is to aid the early detection of breast cancer, lung cancer, diabetes, and heart disease through the analysis of heterogeneous medical data in an effective and convenient way. [5], [7].

System Architecture: DiagniQ has a modular three-layer structure with a frontend layer, a backend layer and a model execution layer. The frontend is created based on the React.js framework and offers an easy to use interface in data entry, report posting, and visualisation of results. It uses the Node.js and Express libraries to implement the backend that will perform user authentication, routing of requests, and communication with predictive models.

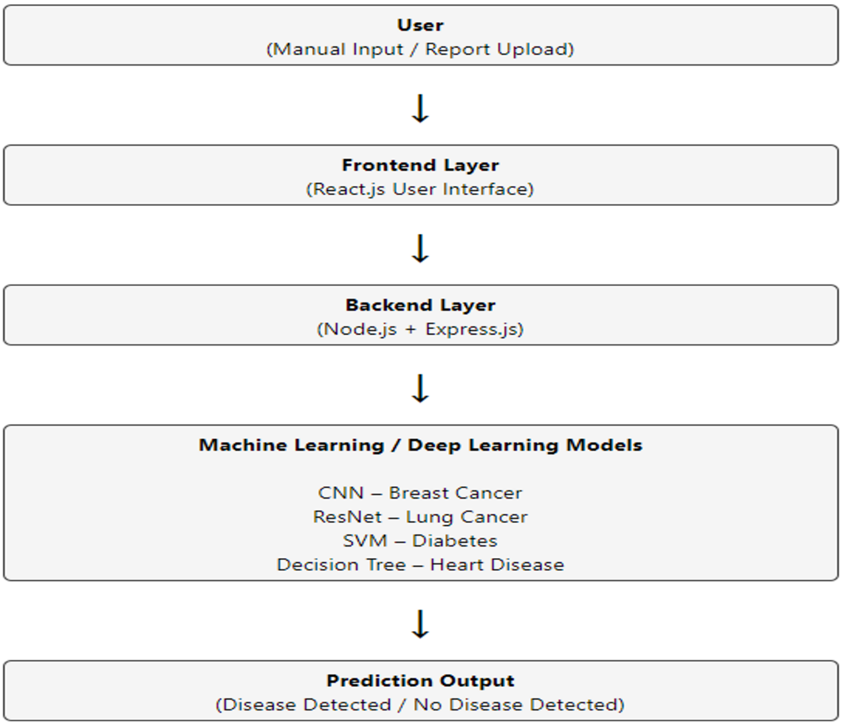


Fig. 1. Block diagram of the proposed DiagniQ system.

The model execution layer is developed in Python and it incorporates both the machine learning and deep learning models that are applied in the prediction of diseases.
The backend executes prediction models with worker threads or other processes so as not to slow down the main server to ensure that the system runs smoothly. This method enables live predictions as well as maintaining the system scalability and responsiveness.Fig. 1 shows the general architecture of the suggested DiagniQ system and demonstrates the interactions between the user interface, back-end server, and machine learning models.
Depending on the character of the medical data, various predictive models are used:
1. Breast Cancer Prediction: A Convolutional Neural Network that is trained on histopathological images of breast cancer is used to detect cancer in the breast. Preprocessing of the images is applied by using resizing, normalisation and data augmentation to enhance generalisation. The CNN model will categorize the input images as either benign or malignant according to acquired spatial and texture characteristics.
2. Lung Cancer Prediction: This approach to predict lung cancer uses a deep learning model, which will be the Residual Neural Network (ResNet) architecture. The images of CT scan of the lungs are processed following the common pre-processing procedures like resizing and normalisation. The residual relationships in ResNet can extract more features and better classify cancerous and non-cancerous cases.
3. Diabetes Prediction: A Support Vector Machine model is used to make predictions of diabetes using structured clinical data to predict the disease. The input parameters will contain the level of glucose in the body mass index, insulin value, age among other pertinent medical factors. The SVM model is quite effective in dealing with high dimensional data and non-linear decision boundaries and thus gives high reliability in the prediction outcomes.

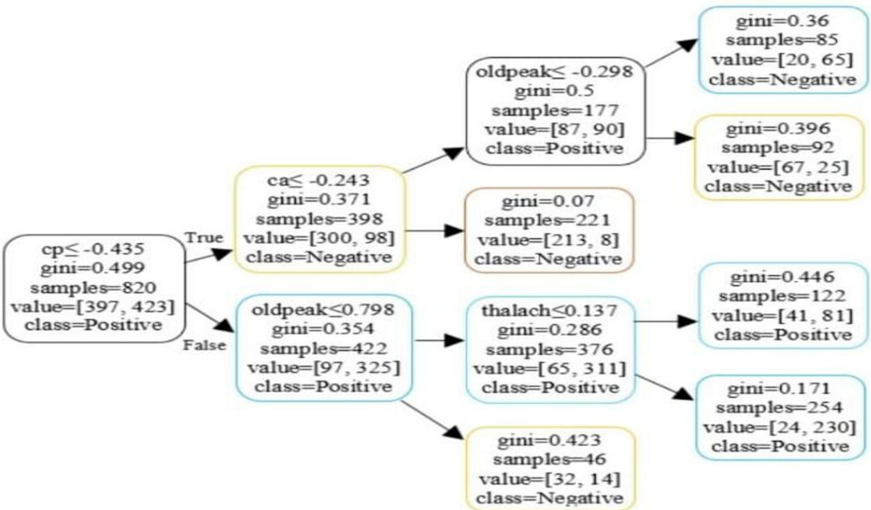
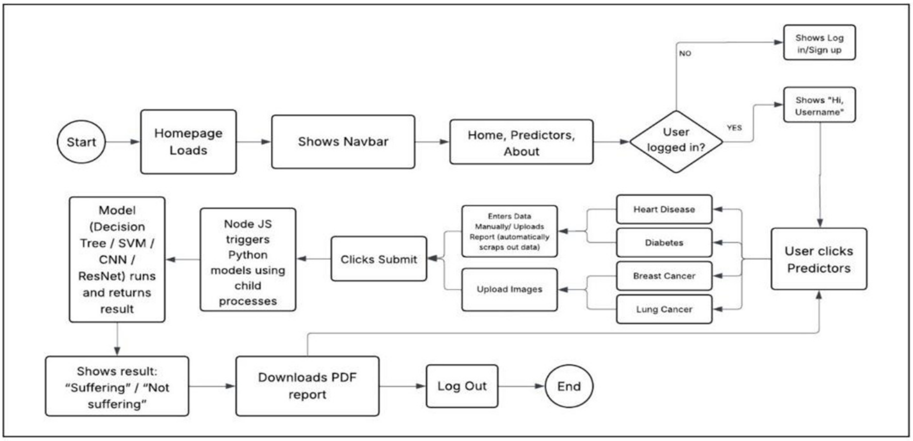


Fig. 2. Decision Tree model used for heart disease prediction (depth = 3).

4. Heart Disease Prediction: A Decision Tree classifier is applied to predict heart diseases based on clinical variables of age, cholesterol level, blood pressure, ECG outcomes, and the type of chest pain. Decision Tree model is interpretable and thus it is applicable in medical decision support applications

Fig.3.Workflow Diagram



Input and Output Mechanism

DiagniQ can take two modes of input, one being manually typing the health parameters and the other being the uploading of structured medical report in supported formats. After the input is sent the backend examines the data and routes them to the relevant predictive model. The result of the prediction is then delivered to the frontend where it will be presented in a lucid and understandable form. The system gives a risk assessment but not a final diagnosis, and there is a clear presentation of DiagniQ as a screening and decision-support tool but not a replacement of medical professionals.

5. Results

The functionality of the proposed DiagniQ system was tested with publicly accessible benchmark datasets that were associated with each disease. All the prediction models were individually trained and tested and finally incorporated into the web based platform. The assessment criterion was on the accuracy of prediction, response time and the usability of the system.

The Convolutional Neural Network that was trained on the histopathological images data reached a maximum accuracy of about 89% in predicting breast cancer. The model provided good discrimination of benign and malignant tissue samples which means that this model can be used in the screening activities at the early stages. Image processing and data augmenting helped in better generalisation performance.

The model of prediction of lung cancer, which is based on the architecture of Residual Neural Network (ResNet) reached an accuracy of about 83 percent in case of CT scan images. The remaining connections allowed extracting more features and made the learning more stable, leading to the reliable classification of cancerous and non-cancerous lung images.

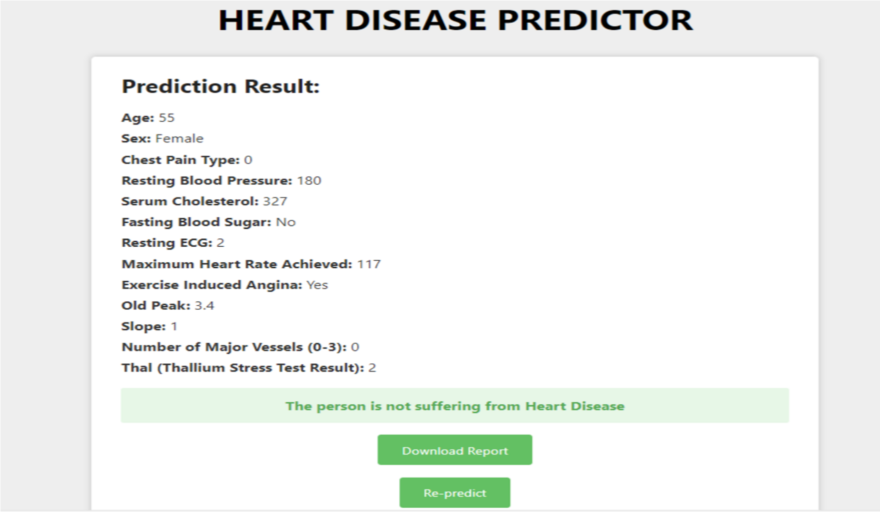


Fig. 4. Sample output generated by the DiagniQ system for heart disease prediction.

In case of structured clinical data, the Support Vector Machine model that was applied to predict diabetes had an accuracy in the range of about 85%. The model was useful in managing non-linear correlations between clinical attributes, including glucose level, body mass index, and insulin levels. Likewise, the Decision Tree classifier that predicts heart disease had the ability to achieve the accuracy of about 87 percent which gives the system reliable predictions as well as offers the interpretation of the decision regulation.

Along the lines of prediction performance, system performance was also considered on the aspects of response time and user interaction. The system produced prediction output in several seconds with respect to manual input and report upload mode. The interface was reported to be user friendly, and the prediction results were easy to read and understand as well as the interface was found to be intuitive and easy to navigate through as indicated by user testing. These findings indicate that DiagniQ is an efficient real-time and user-friendly early disease screening and decision-support system.

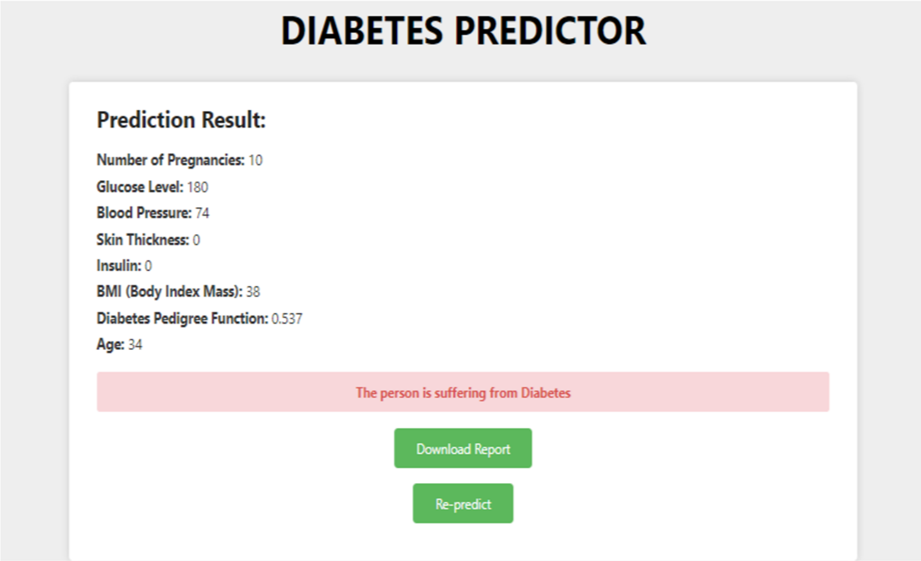


Fig. 5. Sample output generated by the DiagniQ system for diabetes disease prediction

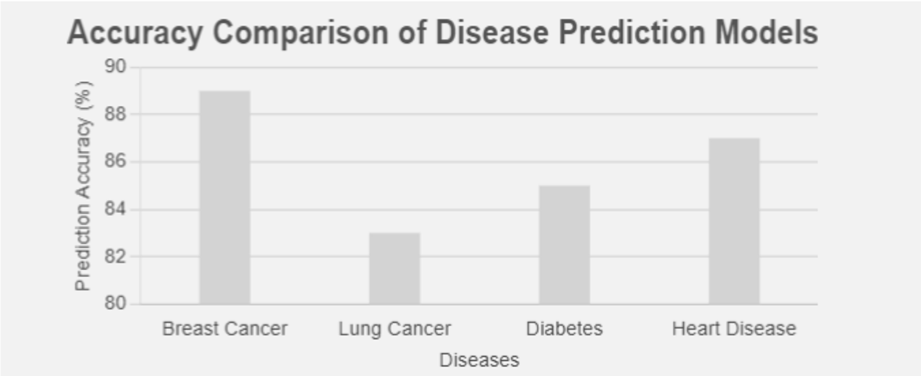


Fig. 6. Comparison of prediction accuracy for different disease prediction models.

Table I presents a concise comparison of the performance of the disease-specific prediction models used in the DiagniQ system. This comparison allows for an easy evaluation of model accuracy across different diseases within the system.

Table I: Quantitative Performance Comparison of Disease Prediction Models

Disease	Model Used	Data Type	Accuracy (%)
Breast Cancer	CNN	Histopathological Images	89
Lung Cancer	ResNet	CT Scan Images	83
Diabetes	SVM	Structured Clinical Data	85
Heart Disease	Decision Tree	Structured Clinical Data	87

6. Conclusion

In this paper, a case of AI-based multi disease prediction system, DiagniQ, has been introduced that is capable of supporting early screenings of breast cancer, lung cancer, diabetes, and heart disease. The system incorporates machine learning and deep learning algorithms into a single, web based framework that has the ability to analyse structured clinical data as well as medical images. Convolutional Neural Networks, Residual Neural Networks, Support Vector Machines, and Decision Tree classifiers are the main classifiers DiagniQ combines to cater to the wide range of data needs of various diseases.

The tests indicate that the system can provide high prediction accuracy rates of all diseases it supports and remain fast and convenient to use. The system is more approachable and practical with users being able to manually input their health data or load structured health reports. DiagniQ is not a final diagnosis but rather a screening and decision-support tool and assists the user in seeking timely medical advice and facilitating preventive healthcare. In general, DiagniQ opens up the opportunities of hybrid AI in healthcare use, and it shows how the use of integrated multi-disease platforms can contribute to early awareness of disease. The system can be improved in future by adding optical character recognition so that it will analyse reports automatically, integration of more diseases and running it on scalable cloud computing infrastructure to ensure a broader adoption.

References

1. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using convolutional neural networks,” Proc. Int. Joint Conf. Neural Networks (IJCNN), Vancouver, BC, Canada, 2016, pp. 2560–2567, doi: 10.1109/IJCNN.2016.7727519.
2. Forderation, “BreakHis: Breast cancer histopathological image dataset (400×),” Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/forderation/breakhis-400x .
3. A. Srivastava, A. Rai, and A. Khan, “Early-stage lung cancer detection using ResNet,” in Emerging Trends in Computer Science and Its Applications. Boca Raton, FL, USA: CRC Press, 2025, pp. 330–335.
4. S. Vaishnavii and A. Senthilselvi, “Early detection of lung cancer with deep learning: An optimized ResNet-50 model,” in Proc. 8th Int. Conf. Circuit, Power and Computing Technologies (ICCPCT), IEEE, Aug. 2025, pp. 1728–1733.
5. M. Nickparvar, “Lung cancer CT scan image dataset,” Kaggle, 2021. [Online]. Available: https://www.kaggle.com/ datasets/hamdallak/the-iqothnccd-lung-cancer-dataset
6. A.Srivastava, A. Tripathi, and A. Singh, “A review of diabetes prediction systems using SVM machine learning algorithms,” Int. Res. J. Modernization Eng., Technol. Sci., vol. 6, no. 5, pp. 8422–8428, May 2024.
7. UCI Machine Learning Repository, “Pima Indians diabetes database,” Kaggle. [Online]. Available: https:// www.kaggle.com/datasets/uciml/pima-indians-diabetesdatabase.
8. R. Vijaya Saraswathi, K. Gajavelly, A. K. Nikath, R. Vasavi, and R. R. Anumasula, “Heart disease prediction using decision tree and SVM,” in Proc. 2nd Int. Conf. Advances in

Computer Engineering and Communication Systems (ICACECS). Singapore: Springer, 2022, pp. 69–78.

9. H. Puri, J. Chaudhary, K. R. Raghavendra, R. Mantri, and K. Bingi, “Prediction of heart stroke using support vector machine algorithm,” in Proc. 8th Int. Conf. Smart Computing and Communications (ICSCC), IEEE, 2021, pp. 21–26.
10. Kaggle Community, “Heart disease dataset,” Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/ johnsmith88/heart-disease-dataset
11. R. Islam, A. Sultana, and M. R. Islam, “A comprehensive review of chronic disease prediction using machine learning algorithms,” J. Electrical Systems and Information Technology, vol. 11, Art. no. 27, Jul. 2024, doi: 10.1186/s43067-024-00150-4.
12. W. DeGroat, H. Abdelhalim, E. Peker, et al., “Multimodal AI/ML for discovering novel biomarkers and predicting disease using multi-omics profiles of patients with cardiovascular diseases,” Scientific Reports, vol. 14, Art. no. 26503, Nov. 2024, doi: 10.1038/ s41598-024-78553-6.
13. R. Islam, A. Sultana, and M. R. Islam, “Machine learning approaches for chronic disease prediction: A systematic review,” Scientific Reports, vol. 14, Art. no. 11245, 2024, doi: 10.1038/s41598-024-61234-9.