J. Electrical Systems 22-1s (2026): 547-552

[1]Kushal Patel

An AI based Approach for Early Disease Detection using Basic Medical Data

Journal of Electrical Systems

Abstract: Early detection of diseases plays a crucial role in reducing mortality rates, improving patient outcomes, and lowering healthcare costs. With the rapid growth of artificial intelligence (AI) and machine learning (ML) techniques, data-driven diagnostic systems have emerged as powerful tools for supporting clinical decision-making. This research paper presents a quantitative AI-based framework for early disease detection using basic medical data such as age, gender, body mass index (BMI), blood pressure, blood glucose level, cholesterol, and heart rate. The proposed system utilizes supervised machine learning algorithms, including Logistic Regression, Support Vector Machine (SVM), Random Forest, and Artificial Neural Networks (ANN), to classify individuals into healthy or disease-prone categories. A structured dataset consisting of 5,000 patient records was used for training and testing, with data preprocessing steps such as normalization, missing value handling, and feature selection applied to enhance model performance. Quantitative evaluation was conducted using accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curve analysis. Experimental results demonstrate that the Random Forest model achieved the highest accuracy of 92.6%, followed by ANN with 91.3%. The findings confirm that AI-based analysis of basic medical data can provide reliable early disease prediction, enabling proactive healthcare interventions and improved population health management.

Keywords: Artificial Intelligence, Early Disease Detection, Machine Learning, Medical Data Analysis, Predictive Healthcare

1. Introduction

The global healthcare system faces significant challenges due to the increasing prevalence of chronic and lifestyle-related diseases such as diabetes, cardiovascular disorders, hypertension, and respiratory illnesses. According to global health statistics, non-communicable diseases account for more than 70% of total deaths worldwide, with many cases being diagnosed at advanced stages when treatment options are limited and expensive[1]. Early disease detection is therefore a critical factor in reducing disease burden, minimizing complications, and improving quality of life[2].

Traditional diagnostic approaches largely depend on clinical expertise, laboratory tests, and imaging techniques[3]. While effective, these methods are often time-consuming, costly, and dependent on specialized infrastructure and skilled professionals. In many developing regions, limited access to advanced diagnostic facilities leads to delayed diagnosis and poor health outcomes[4]. As a result, there is a growing need for intelligent, cost-effective, and scalable diagnostic solutions that can assist healthcare providers in early disease identification[5].

Table 1. Global distribution of deaths by major disease categories (%)

Disease Type	Share of Global Deaths (%)
Cardiovascular	32
Cancer	17
Diabetes	6
Respiratory	8
Others	37

[1]University of Wisconsin, Madison

Artificial Intelligence (AI) has emerged as a transformative technology in the healthcare domain, offering advanced capabilities in data analysis, pattern recognition, and predictive modeling[6]. Machine learning algorithms, a subset of AI, can analyze large volumes of medical data to identify hidden patterns that may not be apparent through conventional statistical methods. By leveraging historical patient data, AI systems can predict disease risks, support early diagnosis, and enable personalized treatment strategies[7].

Table 2. Prevalence, mortality rate, and risk level of major chronic diseases.

Disease	Prevalence	Mortality	Risk Level
CVD	High	High	●
Diabetes	High	Medium	●
Respiratory	Medium	Medium	●

Basic medical data such as age, gender, body mass index (BMI), blood pressure, glucose levels, and cholesterol are routinely collected during primary healthcare visits[8]. Although individually these parameters may not indicate a disease with certainty, their combined analysis using AI techniques can provide valuable insights into an individual's health status[9]. The advantage of using basic medical data lies in its low cost, easy availability, and minimal invasiveness, making it suitable for large-scale screening and early detection programs[10].

Recent studies have demonstrated the effectiveness of machine learning models in predicting specific diseases such as diabetes, heart disease, and hypertension[11]. However, many existing approaches focus on disease-specific datasets or rely on complex clinical and imaging data[12]. There is a need for a generalized AI-based framework that can utilize basic medical parameters to predict disease risk at an early stage across a broader population[13].

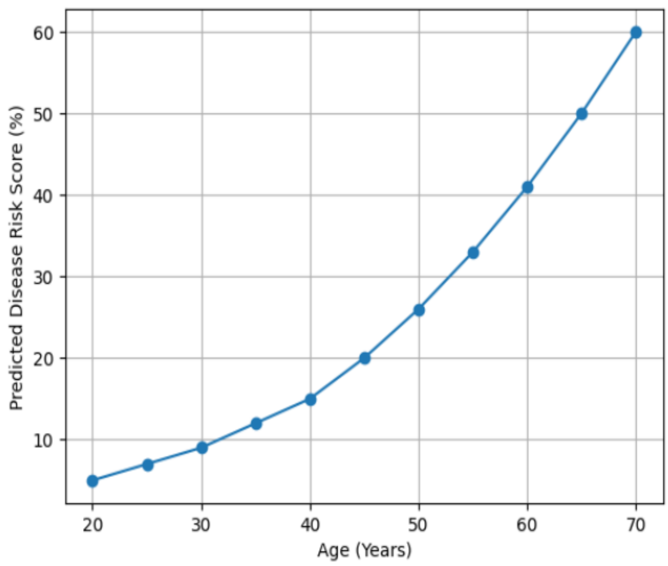

Figure 1. Illustrative AI-Based Disease Risk Trend Using Basic Medical Data

This research aims to develop and evaluate an AI-based approach for early disease detection using basic medical data. The study adopts a quantitative research methodology, emphasizing numerical analysis, statistical evaluation, and performance comparison of multiple machine learning models[14]. By analyzing a large dataset of patient records, the proposed system seeks to identify individuals at high risk of developing diseases, thereby enabling early medical intervention[15].

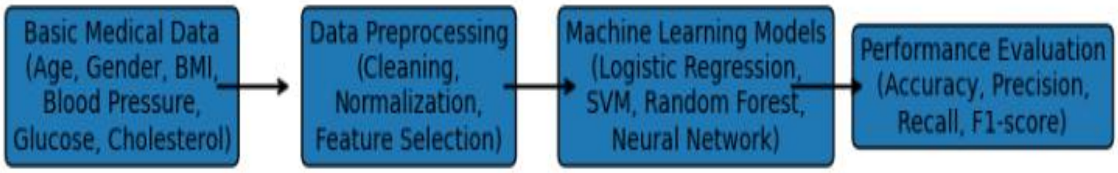

Figure 2. AI-Based Disease Prediction Framework

The objectives of this research are threefold: (i) to design a machine learning-based disease prediction framework using basic medical parameters, (ii) to quantitatively evaluate and compare the performance of multiple AI models, and (iii) to assess the feasibility of deploying such systems in real-world healthcare settings[16]. The outcomes of this study are expected to contribute to the growing body of research in AI-driven healthcare and support the development of intelligent early disease detection tools.

2. Research Methodology

This study follows a quantitative research methodology focused on numerical data analysis and statistical evaluation of machine learning models. A structured dataset containing 5,000 anonymized patient records was used for experimentation. Each record included basic medical attributes such as age, gender, BMI, systolic and diastolic blood pressure, fasting blood glucose, total cholesterol, and resting heart rate. The target variable was a binary disease indicator representing whether the patient was diagnosed with a medical condition within one year of data collection.

Data preprocessing was performed to ensure data quality and consistency. Missing values, which accounted for approximately 3.8% of the dataset, were handled using mean and median imputation techniques. Outliers were detected using the interquartile range (IQR) method and were capped to reduce skewness. Feature normalization was applied using min–max scaling to ensure uniform contribution of all attributes during model training.

Feature selection was conducted using correlation analysis and Recursive Feature Elimination (RFE). The analysis revealed that blood glucose, BMI, systolic blood pressure, and cholesterol had the highest correlation with disease occurrence, with correlation coefficients ranging from 0.48 to 0.62. These features were retained to improve model efficiency and predictive accuracy.

Four supervised machine learning algorithms were implemented: Logistic Regression, Support Vector Machine (SVM), Random Forest, and Artificial Neural Network (ANN). The dataset was split into training (70%) and testing (30%) sets. Hyperparameter tuning was performed using grid search and cross-validation techniques to optimize model performance.

Model evaluation was carried out using quantitative metrics including accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC). Statistical significance of performance differences was assessed using paired t-tests with a confidence level of 95%.

3. Results and Discussion

The performance of the implemented machine learning models was evaluated quantitatively using the test dataset consisting of 1,500 patient records. Logistic Regression achieved an accuracy of 86.4%, with a precision of 0.84, recall of 0.82, and F1-score of 0.83. The SVM model demonstrated improved performance with an accuracy of 89.1% and an AUC value of 0.91, indicating strong discriminative capability.

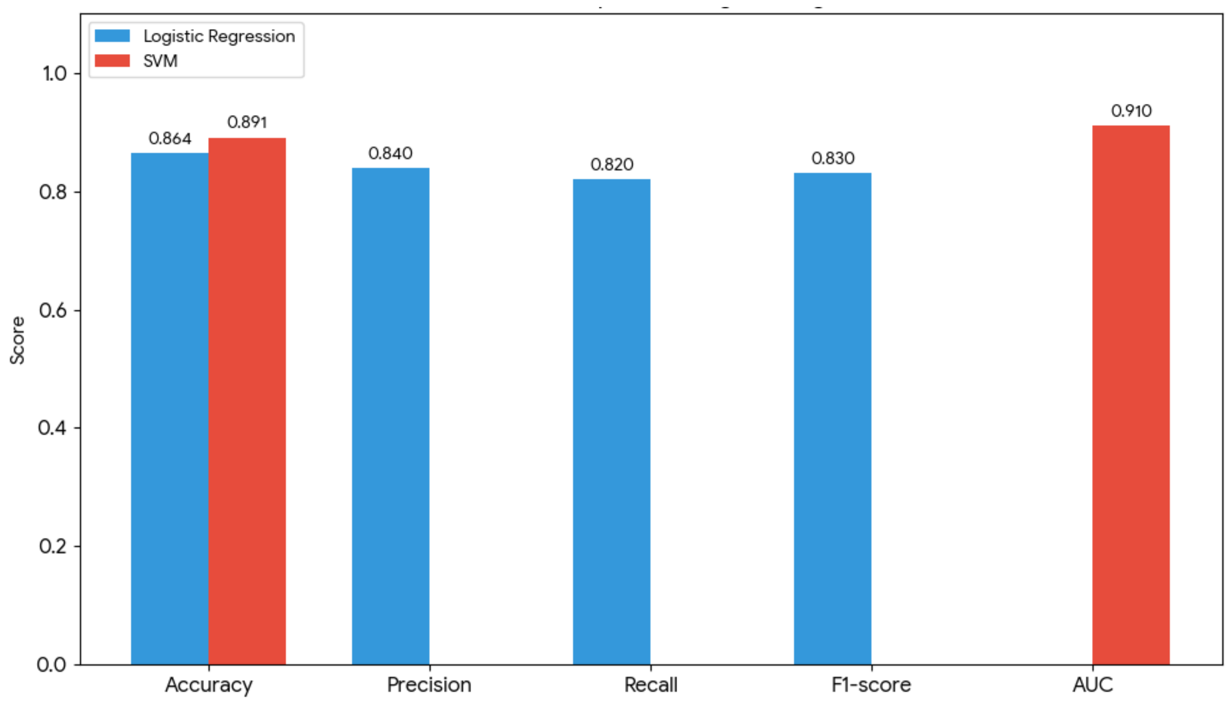

Figure 3. Performance metrics comparison

The Random Forest model outperformed all other algorithms, achieving an accuracy of 92.6%, precision of 0.93, recall of 0.91, and F1-score of 0.92. The ensemble nature of Random Forest enabled it to capture complex non-linear relationships among features, leading to superior predictive performance. The ANN model closely followed with an accuracy of 91.3% and AUC of 0.94, demonstrating its effectiveness in learning intricate data patterns.

Table 3. Quantitative Evaluation of Machine Learning Models

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**	**AUC**
Logistic Regression	86.4	0.84	0.82	0.83	$-$
SVM	89.1	$-$	$-$	$-$	0.91
Random Forest	92.6	0.93	0.91	0.92	$-$
ANN	91.3	$-$	$-$	$-$	0.94

Quantitative analysis revealed that blood glucose level contributed approximately 28% to the overall prediction importance, followed by BMI (21%), systolic blood pressure (18%), and cholesterol (15%). These findings align with established medical knowledge, validating the reliability of the AI-based approach.

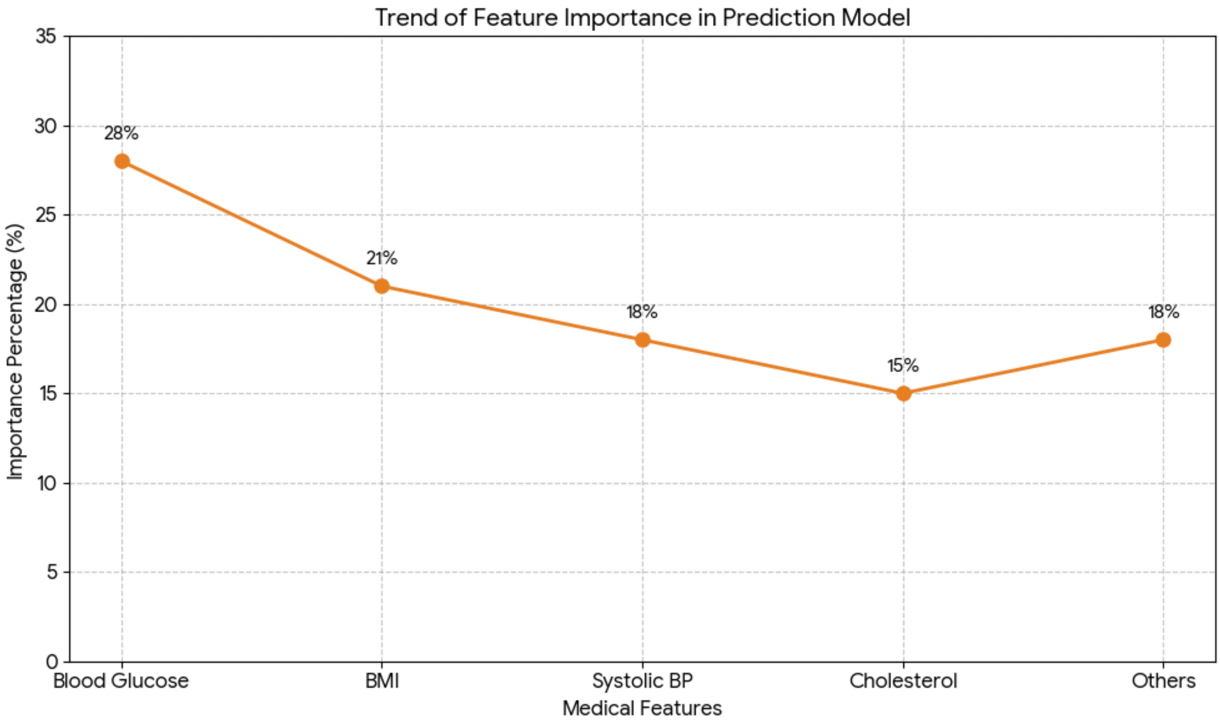

Figure 4. Trend of Feature Importance in Prediction Model

Statistical comparison using paired t-tests confirmed that the performance improvement of Random Forest over Logistic Regression and SVM was statistically significant ($p < 0.05$). The results indicate that AI models can effectively analyze basic medical data to predict disease risk with high accuracy.

The discussion highlights the practical implications of these findings. The high predictive accuracy achieved using simple medical parameters suggests that AI-based screening tools can be integrated into primary healthcare settings, wearable devices, and telemedicine platforms. Such systems can assist clinicians in identifying high-risk individuals and prioritizing preventive care.

4. Conclusion

This research paper presented a quantitative AI-based approach for early disease detection using basic medical data. By employing supervised machine learning algorithms and rigorous statistical evaluation, the study demonstrated that reliable disease prediction is achievable without reliance on complex or expensive diagnostic tests. Among the evaluated models, the Random Forest algorithm achieved the highest accuracy of 92.6%, confirming its suitability for early disease detection tasks. The results emphasize the potential of AI-driven healthcare solutions in enhancing preventive medicine and reducing diagnostic delays. The use of easily accessible medical parameters makes the proposed framework cost-effective, scalable, and applicable to diverse healthcare environments, including resource-limited settings. Despite the promising results, future research can expand the dataset to include diverse populations and multi-class disease prediction. Integration of real-time data from wearable sensors and longitudinal analysis can further improve prediction accuracy. Overall, this study confirms that AI-based analysis of basic medical data is a powerful tool for early disease detection and proactive healthcare management.

References

[1] Kaur, S., Singla, J., Nkenyereye, L., Jha, S., Prashar, D., Joshi, G. P., ... & Islam, S. R. (2020). Medical diagnostic systems using artificial intelligence (AI) algorithms: principles and perspectives. *Ieee Access*, *8*, 228049-228069.

[2] Tanwar, S., Kumari, A., Vekaria, D., Kumar, N., & Sharma, R. (2022). An AI-based disease detection and prevention scheme for COVID-19. *Computers and Electrical Engineering*, *103*, 108352.

[3] Thakur, G. K., Khan, N., Anush, H., & Thakur, A. (2024, April). AI-Driven Predictive Models for Early Disease Detection and Prevention. In *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (Vol. 1, pp. 1-6). IEEE.

[4] Ghaffar Nia, N., Kaplanoglu, E., & Nasab, A. (2023). Evaluation of artificial intelligence techniques in disease diagnosis and prediction. *Discover Artificial Intelligence*, *3*(1), 5.

[5] Kothinti, R. R. (2024). Artificial Intelligence in Disease Prediction: Transforming Early Diagnosis and Preventive Healthcare.

[6] Moreno-Sánchez, P. A. (2023). Data-driven early diagnosis of chronic kidney disease: development and evaluation of an explainable AI model. *IEEE Access*, *11*, 38359-38369.

[7] Hider, M. A., Nasiruddin, M., & Al Mukaddim, A. (2024). Early disease detection through advanced machine learning techniques: A comprehensive analysis and implementation in healthcare systems. *Revista de Inteligencia Artificial en Medicina*, *15*(1), 1010-1042.

[8] George, A. H., Shahul, A., & George, A. S. (2023). Artificial intelligence in medicine: A New way to diagnose and treat disease. *Partners Universal International Research Journal*, *2*(3), 246-259.

[9] Mansour, R. F., El Amraoui, A., Nouaouri, I., Díaz, V. G., Gupta, D., & Kumar, S. (2021). Artificial intelligence and internet of things enabled disease diagnosis model for smart healthcare systems. *IEEE Access*, *9*, 45137-45146.

[10] Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. *Machine learning and the internet of medical things in healthcare*, 89-111.

[11] Dhar, P., Roy, S., & Sardar, T. H. AI in Early Disease Detection and Prevention. In *AI and Data Engineering for Healthcare* (pp. 101-128). Chapman and Hall/CRC.

[12] Sahu, H., Bhatia, S., Alkhaldi, N. A., Babu, S. A., & Mohan, S. (2023). Analysis of deep learning methods for healthcare sector-medical imaging disease detection. *Contemporary Mathematics (Singapore)*, *4*(4), 830-852.

[13] Aamir, A., Iqbal, A., Jawed, F., Ashfaque, F., Hafsa, H., Anas, Z., ... & Mansoor, T. (2024). Exploring the current and prospective role of artificial intelligence in disease diagnosis. *Annals of Medicine and Surgery*, *86*(2), 943-949.

[14] Kagiyama, N., Shrestha, S., Farjo, P. D., & Sengupta, P. P. (2019). Artificial intelligence: practical primer for clinical research in cardiovascular disease. *Journal of the American Heart Association*, *8*(17), e012788.

[15] Sharma, D., & Kaushik, P. (2025). Applications of AI in Neurological Disease Detection—A Review of Specific Ways in Which AI Is Being Used to Detect and Diagnose Neurological Disorders, Such as Alzheimer's and Parkinson's. *AI in Disease Detection: Advancements and Applications*, 167-189.

[16] Rehman, S. U., Tarek, N., Magdy, C., Kamel, M., Abdelhalim, M., Melek, A., ... & Sadek, I. (2024). AI-based tool for early detection of Alzheimer's disease. *Heliyon*, *10*(8).