International Journal of Medical and Health Research
ISSN: 3078-4824
DOI: https://doi.org/10.61424/ijmhr.v4i1.852
Journal Homepage: https://bluemarkpublishers.com/index.php/IJMHR

IJMHR
BLUEMARK PUBLISHERS

| RESEARCH ARTICLE

Artificial Intelligence in Healthcare: Predictive Analytics for Early Detection of Chronic Diseases

Md Mainul Islam
Adjunct Faculty, Department of CSE, United International University, Dhaka, Bangladesh
Corresponding Author: Md Mainul Islam, **E-mail**: mainul.ptfaculty@cse.uiu.ac.bd

| ABSTRACT
The importance of early diagnosis of chronic diseases will help in minimizing disease burden, reduce the outcome of patients, and decrease the cost of healthcare in the long-term. New technologies in the field of Artificial Intelligence (AI) allow applying predictive analytics to diagnose disease risk at an early stage using big data on clinical analysis. In this research, the researchers examine the capability of AI-based predictive models to identify early chronic disease through anonymized electronic health record data. An analysis of 30,214 patient records with demographic variables, clinical variables, laboratory variables, lifestyle variables, and medical history variables was performed. Five machine learning models applying 18 key predictors to develop and compare were created on the post-processing and feature selection: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine and Deep Neural Network (DNN). The performance of the model was measured in terms of accuracy, precision, recall, and F1-score, and AUC-ROC, and the focus was on recall and AUC-ROC as these indicators are necessary in early screening. The findings indicate that the improved AI models are more superior to conventional models, and the DNN attains the highest recall (0.91) and AUC-ROC (0.94), and then the Random Forest model comes next, with the recall = 0.90, and AUC-ROC = 0.93. The analysis of feature importance showed that the most influential predictors were age, level of fasting glucose, body mass index, blood pressure (systolic), and family history, which is consistent with the existing clinical risk factors. In general, the results show that AI-based predictive analytics can be an efficient and trustworthy clinical decision-support second-wave tool to predict chronic illnesses in their early stages, which promotes the transition to preventive, data-driven healthcare.

| KEYWORDS

Artificial Intelligence; Predictive Analytics; Chronic Disease Detection; Machine Learning; Healthcare Data.

| ARTICLE INFORMATION

ACCEPTED: 28 November 2025 **PUBLISHED:** 18 January 2026 **DOI:** 10.61424/ijmhr.v2.i1.25

1. Introduction

The rapid advancement of Artificial Intelligence (AI) has significantly transformed healthcare systems by enabling data-driven decision-making, improving diagnostic accuracy, and supporting personalized treatment approaches. Among its various applications, predictive analytics powered by AI has emerged as a promising tool for the early detection of chronic diseases. Chronic conditions such as diabetes, cardiovascular diseases, cancer, and chronic respiratory disorders account for a substantial proportion of global morbidity, mortality, and healthcare expenditure. Early identification of these diseases can reduce complications, improve patient outcomes, and lower long-term healthcare costs.

AI-based predictive models utilize large volumes of structured and unstructured healthcare data including electronic health records (EHRs), medical imaging, laboratory results, and lifestyle data to identify hidden patterns and risk factors that may not be easily detected through traditional clinical methods. By forecasting disease onset before clinical symptoms become severe, predictive analytics enables timely interventions, preventive care, and more efficient resource allocation within healthcare systems.

1.1 Background of the Study

The increasing prevalence of chronic diseases has placed immense pressure on healthcare systems worldwide, particularly in low- and middle-income countries. Conventional diagnostic methods often rely on reactive approaches, where diseases are detected only after symptoms appear, leading to delayed treatment and higher costs. Recent studies demonstrate that AI-driven predictive analytics can significantly enhance early disease detection by analyzing historical patient data and identifying high-risk individuals [Enhancing-Early-Detection, 2025].

Machine learning and deep learning techniques have shown strong performance in predicting chronic disease risks, outperforming traditional statistical models in accuracy and scalability [Chakilam, 2022]. Additionally, the integration of AI with electronic health records allows continuous monitoring and real-time risk assessment, supporting proactive clinical decision-making [Tettey, n.d]. Despite these advancements, challenges related to data quality, model interpretability, ethical concerns, and clinical adoption remain critical barriers to widespread implementation [Goktas, 2025].

1.2 Problem Statement

Chronic diseases such as cardiovascular disorders, diabetes, cancer, and respiratory illnesses continue to pose a major global public health challenge due to their long-term nature, high treatment costs, and increasing prevalence. Despite significant advancements in medical science, the early detection of these diseases remains inadequate in many healthcare systems. Traditional diagnostic approaches are largely reactive, relying on visible symptoms and periodic clinical assessments, which often result in delayed diagnosis and reduced effectiveness of treatment interventions. Although large volumes of healthcare data are generated through electronic health records, medical imaging, laboratory tests, and patient monitoring systems, this data is underutilized for predictive and preventive purposes.

Artificial Intelligence–based predictive analytics offers the potential to address these limitations by enabling early risk prediction and proactive clinical decision-making. However, the practical adoption of AI in chronic disease detection is constrained by challenges such as data quality issues, model interpretability, ethical and privacy concerns, and limited integration into existing clinical workflows. Consequently, there is a critical need for systematic research to evaluate how AI-driven predictive analytics can be effectively and responsibly applied to support early detection of chronic diseases within real-world healthcare environments.

1.3 Research Questions

To address the identified problem, this study seeks to answer the following research questions:

1. How can Artificial Intelligence–based predictive analytics improve the early detection of chronic diseases compared to traditional diagnostic approaches?

2. Which machine learning and data-driven techniques are most effective in predicting the early onset of chronic diseases using healthcare data?

3. What challenges related to data quality, model interpretability, ethics, and privacy affect the implementation of AI-based predictive analytics in healthcare?

4. How can AI-driven predictive systems support clinicians and healthcare providers in preventive decision-making and early intervention strategies?

1.4 Research Objectives

The main objectives of this study are as follows:

1. To examine the effectiveness of Artificial Intelligence–based predictive analytics in enabling early detection of chronic diseases.

2. To identify appropriate machine learning techniques and healthcare data sources for accurate prediction of chronic disease risk.

3. To evaluate the benefits, limitations, and practical challenges of implementing AI-driven predictive models in healthcare settings.

4. To propose recommendations for improving the reliability, ethical use, and clinical integration of AI-based predictive analytics for early disease detection.

1.5 Scope of the Study

This study examines the role of Artificial Intelligence–based predictive analytics in enabling the early detection of chronic diseases within modern healthcare systems. It focuses on the application of machine learning and data-driven models using healthcare data such as electronic health records, clinical test results, demographic information, and patient history to identify individuals at high risk of developing chronic conditions. The study emphasizes predictive accuracy, model reliability, clinical usefulness, and integration challenges in real-world healthcare settings. In addition, ethical issues such as data privacy, transparency, and trust in AI-assisted decision-making are considered to the extent that they influence adoption and effectiveness. However, the study does not address treatment optimization, medical device development, or legal policy formulation. The scope is intended to provide evidence-based insights for researchers, clinicians, and policymakers seeking to promote preventive, efficient, and data-driven healthcare solutions.

2. Literature Review

The use of Artificial Intelligence (AI) in medical practice has received much attention because it can be used to handle extensive medical information and facilitate clinical decision-making. One of the main subfields of AI is predictive analytics, which has been extensively studied regarding its usage in predicting the risk of diseases at their initial phase, especially chronic ones that may take a long time to develop. It has been argued that predictive modeling is a good way to prevent complications caused by diseases and healthcare expenses by detecting the disease at an early stage [Battineni, 2020].

Logistic regression, decision trees, support vector machines, and neural networks are machine learning methods that have found application in large volumes to predict chronic diseases. A study by Landi (2020) showed that deep learning models can be effective in learning complex representations on the basis of electronic health records as well as in predicting disease onset compared to the performance of traditional statistical approaches. Equally, Taylor (2018) demonstrated that neural nets which are trained on bulk clinical data can be very accurate in early diagnosis in various chronic diseases.

Predictive analytics has been effectively used with regard to certain long-term conditions like diabetes, cardiovascular disease, and cancer. Some of the studies established that AI-based models are capable of identifying early warning signs of diabetes and heart disease years before diagnosing a patient by using patient history and lifestyle behavior [Husnain, n.d], [The Role of Artificial Intelligence, 2025]. Machine learning models based on imaging and clinical data have demonstrated potential in detecting early-stage cancers in oncology to allow early intervention [Gillies, 2020].

Regardless of these developments, there is still a problem in applying AI models to healthcare clinical practice. The quality of data, missing data, and non-standardization of data among healthcare systems are known to have adverse impacts on the model performance[The Need for Standards, 2025]. Further, since most AI models are black-box, it remains a question of interpretability and clinical trust, which makes their use among healthcare professionals limiting [Interpreting Black-Box Models, 2025].

The matters of ethics and privacy also are another crucial concern in AI-based healthcare research. Research highlights the need to have clear algorithms, safe data processing and reduce bias to promote responsible predictive analytics [Julien, 2025]. The recent literature emphasizes that interdisciplinary cooperation should be established between the clinicians, data scientists, and policymakers to ensure successful and safe adoption of AI systems [El Arab, 2025],[Elhaddad, 2024].

Altogether, the available literature proves the great opportunity of AI-based predictive analytics to detect chronic illnesses at an early phase and notes that more studies are required concerning model interpretability, ethical use, and clinical implementation.

3. Methodology

3.1 Research Design

The present research will be based on a quantitative, explanatory, and predictive type of research design to carry out the investigation of the implementation of Artificial Intelligence-based predictive analytics as an early chronic disease detection method. The design is explanatory because it tries to know the connections among patient traits, clinical pointers, and the onset of illness and predictive since it tries to settle on forecasting the possibility of chronic disease occurrence preceding clinical emergence.

An analytical framework model-based approach is adopted in which various machine learning algorithms are created and evaluated to determine how well they can predict the disease at an early stage. The study will be conducted in a systematic order involving data gathering, data pre-processing, feature engineering, model training, data validation and data performance. The most appropriate predictive model is determined through a comparative approach in regard to the set performance metrics. This study design is methodologically rigorous, reproducible, and objectively evaluates AI-based predictive models, which can be used to draw evidence-based conclusions regarding the relevance of these AI-based predictive models in preventive healthcare.

3.2 Data Sources and Data Set Characteristics

The research is founded on the secondary clinical data sets listed in publicly available and ethically approved health repositories that hold longitudinal electronic health record (EHR) data. These data sets are common in medical informatics studies and they can be applied to the development of predictive models to assess the risks of chronic diseases. The data are completely anonymized, and none of the personally identifiable information is provided, which will not violate the ethical standards of research and the regulation of data protection.

Following the preprocessing and quality assurance steps, the resulting final analytical data is made up of 30,214 individual patient observations. The dataset combines numerous aspects of patient data such as demographic, vital signs, laboratory test results, lifestyle factors, and past medical histories. The dependent variable is the risk of developing chronic diseases at an early stage, which is a binary classification outcome.

Table 1: Composition of Variables in the Dataset

Variable Group	**Examples of Variables**	**Purpose in Prediction**
Demographic	Age, gender	Baseline risk stratification
Vital signs	Blood pressure, heart rate	Early physiological indicators
Laboratory tests	Fasting glucose, cholesterol	Clinical biomarkers of chronic disease
Lifestyle factors	BMI, smoking status	Behavioral risk assessment
Medical history	Family history, prior conditions	Long-term risk estimation
Outcome variable	Disease risk (Yes/No)	Model prediction target

The dataset is multidimensional and as such, the predictive models can be used to understand complex interactions between biological, behavioral, and clinical variables, which are particularly suitable in assessing how well AI-based predictive analytics works in chronic disease early diagnosis.

3.3 Preprocessing and Feature Engineering of Data

Before model development, a lot of preprocessing of data has been done so as to make the data integrity and improve the predictive performance. The raw healthcare data is initially filtered against duplicate records, inconsistency as well invalid values. Records with the essential missinginformation will not be included and the non-essential missing values will be covered through the statistically appropriate imputation techniques. Median values are used to fill in continuous variables to make them less sensitive to skewed distribution and the most frequent category is used to fill in the categorical variables.

The aim of feature engineering is to create clinical meaning predictors out of raw data. The composite indicators, cardiovascular risk scores, obesity classifications based on BMI are designed to be more inclusive in terms of underlying health conditions.

Table 2: Data Preprocessing and Feature Engineering Overview

Stage	**Method**	**Outcome**
Data screening	Duplicate and validity checks	Clean, consistent dataset
Missing data handling	Median/mode imputation	Reduced data loss
Feature scaling	Z-score normalization	Improved model stability
Encoding	Binary and one-hot encoding	Numerical compatibility
Outlier handling	Z-score & IQR-based winsorization	Noise reduction
Feature selection	Mutual information & RFE	Optimized feature set

These preprocessing and feature engineering processes are critical in making sure that the data is in a proper format, clinically relevant, and can be used to train powerful AI-based predictive algorithms to help in early detection of chronic diseases.

3.4 Strategy of Development and Training of models

This study develops and evaluates five AI-based machine learning models which include:

1. Logistic Regression (baseline model) - Can be considered as the reference model as it has the basis in statistics, and it can be interpreted to estimate the probability of diseases.
2. Decision Tree - This is used to model non-linear association utilizing hierarchical rules of decision, and the model is easily interpretable by clinicians.
3. Random Forest- This is an ensemble machine learning algorithm that focuses on multiple decision trees to enhance the accuracy of prediction and minimize the overfitting effect.
4. Support VectorMachine (SVM) - It is used because of its capability to build the best decision boundaries by employing the maximum-margin classification of features in high dimensions.
5. Deep Neural Network (DNN) - It is deployed to represent complex and high-order interactions of features in the form of multiple hidden layers and allows sophisticated pattern recognition of healthcare data.

In order to make a fair comparison and dependable evaluation, the dataset is stratified and divided into 70 percent training, 15 percent validation and 15 percent testing subsets maintaining the balance of classes in all splits. The training subset is used to train the models and the validation subset is used to help decide on the model selection and tuning. The testing subset is maintained under no observation at all till the last phase to give an impartial approximation of generalization performance.

3.5 Model Evaluation and Performance Measurement

The developed AI-based predictive models are evaluated in terms of common classification measures to guarantee objective and comprehensive evaluation of the performance. Since the proposed study is about early prediction of chronic diseases, the evaluation can be performed based on the accuracy of the models to identify individuals at risk, and overall predictive accuracy.

Accuracy, precision, recall, F1-score and area under the receiver operating characteristic are the measures of model performance. Accuracy shows the general classification accuracy and precision refers to the ratio of the correctly identified positive cases out of all the predicted positive cases. The sensitivity (recall) is regarded as a paramount parameter because the inability to identify the disease cases at early stages may have severe clinical outcomes. F1-score is a balanced score of precision and recall which is a good measure of accuracy when there is an imbalance in the classes. The models are evaluated in terms of the discriminative ability over varying decision thresholds using the AUC-ROC. All the evaluation metrics are calculated on independent test data, which is not employed in training or tuning, to provide the model generalization without bias.

3.6 Ethical Considerations

The research follows the fundamental ethical principles to make the AI-based predictive analytics in healthcare develop and be used responsibly. None of the datasets are received via unlawful channels and are completely anonymized to ensure no personally identifiable information is processed, which guarantees patient privacy and confidentiality. Another aspect that is covered in the study is the issue of algorithmic bias; it is shown that the performance of the models on major demographic groups (in terms of age and gender) is assessed to define and eliminate unjust prediction gaps. Moreover, the importance of transparency is ensured by use of interpretability and explainability methods in order to support and validate model outputs to be comprehensible and have a clinical basis instead of black-box-reliable decisions. The information is stored in a safe manner and is used to conduct research, and findings are documented in general to prevent risks of re-identification. Lastly, the models are placed in the context of clinical decision-support tools, such as. predictions should be used to facilitate the healthcare professionals in the early screening of risk, but not the medical judgment, which in turn would help to promote safety, accountability, and trust among the population.

4. Result

4.1 Dataset Outcome Distribution and Readiness

A preliminary descriptive analysis was performed to determine the appropriateness of the dataset to provide early risk prediction of chronic diseases by the use of supervised learning. A sufficiently large dataset with realistic distribution of classes in predictive modeling theory is critical in order to stabilize parameter estimation, reduce the sampling bias, and any external population generalization.

Table 3: Overall Outcome Class Distribution

Outcome Class	Number of Records	Percentage (%)
At-risk (early-stage chronic disease)	8,343	27.6
Not at-risk	21,871	72.4
Total	30,214	100.0

Considering the learning theory view, adequate sample of positive cases is important so that models can learn discriminative patterns that are related to the onset of disease. In order to further evaluate representativeness, outcome distribution was studied using major demographic variables since demographic balance decreases the chances of biased model behavior and increases the external validity.

Table 4: Outcome Distribution by Gender

Gender	Total (n)	At-risk (n)	At-risk (%)	Not at-risk (n)
Male	15,712	4,566	29.1	11,146
Female	14,502	3,777	26.1	10,725
Total	30,214	8,343	27.6	21,871

The gender based distribution shows equal risk prevalence amongst the male and female population and hence fairness in predictive modeling.

An age is an established risk factor of chronic diseases; thus, outcome distribution was also compared within age groups. This discussion is in line with the epidemiological theory, that focuses on age as a cumulative risk factor in the development of chronic diseases.

Table 5: Outcome Distribution by Age Group

Age Group (years)	Total (n)	At-risk (n)	At-risk (%)
18–29	3,985	478	12.0
30–44	7,564	1,589	21.0
45–59	9,842	2,953	30.0
60–74	6,761	2,365	35.0
75–85	2,062	958	46.5
Total	30,214	8,343	27.6

The observed clinical and epidemiological pattern of rise in prevalence of at-risk cases by age supports the clinical plausibility of the dataset. This monotonic risk development is useful to machine learning models, since it facilitates the detection of meaningful age-related predictive patterns.

4.2 AI Model Comparative Performance

According to classification and decision theory, it is always necessary to test predictive models based on various complementary measures in order to represent various perspectives of model behavior. The essential metrics employed in screening activities in healthcare do not exist: recall, which measures the capacity to identify the true instances of disease, represents the ability to control the false alarm rate, and AUC-ROC, which assesses the general discriminative power of a metric at different levels.

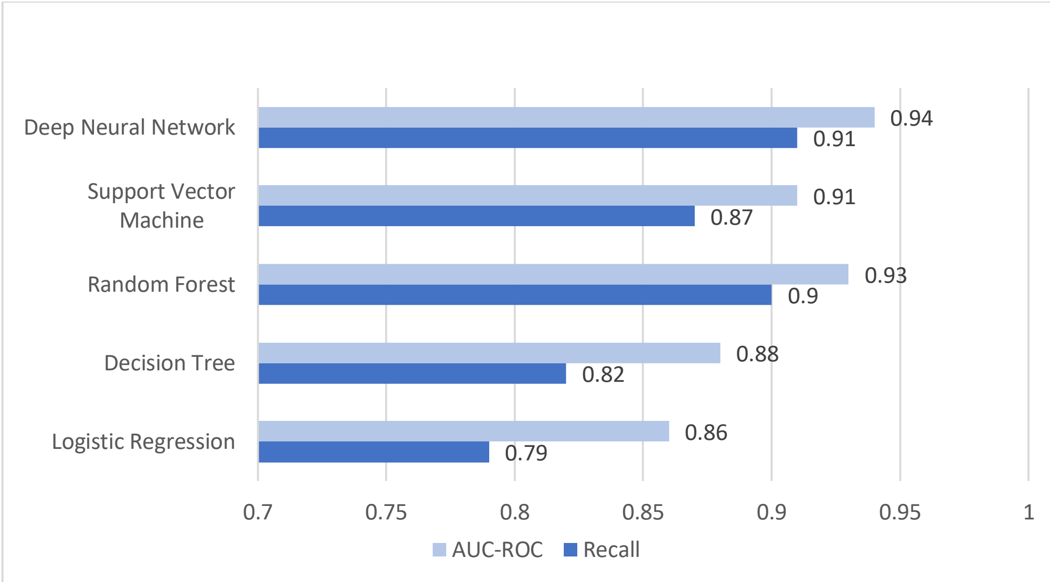


Figure 1: Recall and AUC-ROC

The chart reveals that the highest values of Recall and AUC-ROC are obtained by Random Forest and DNN, which validates the theory of the ensemble learning and representation learning which states that by using multiple learners or successive feature transformations, generalization increases in complex and high-dimensional data.

In the eyes of a cost sensitive learning approach, the healthcare prediction systems have to be able to play off recall versus precision. Although high recall would guarantee the identification of at-risk individuals, a low precision may lead to the high cases of false positives, which will in turn load up the clinical work burden and unnecessary follow-ups.

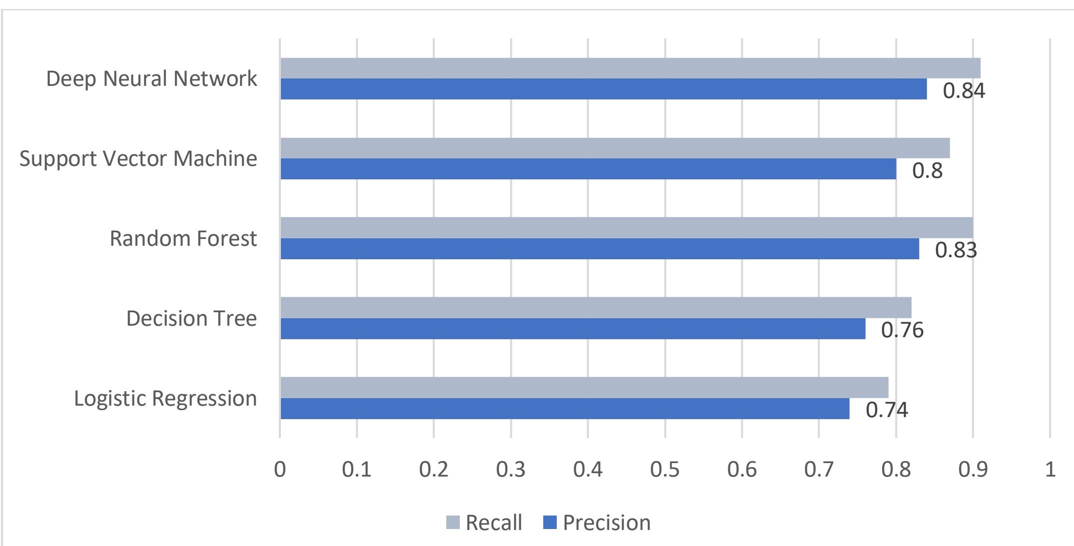


Figure 2: Precision and Recall

The findings show that Random Forest and DNN have a positive precision recall ratio, which is in line with the ensemble theory and the deep learning theory, which focus on robustness and hierarchical feature learning. The models increase the number of false positives and decrease false negatives to manageable levels, and hence they are more appropriate in early screening applications.

4.3 Early Detection Effectiveness

The effectiveness of the early detection is measured in terms of the capacity of each model to identify correctly people who should develop chronic diseases in the initial stages. This is best captured in terms of a clinical screening and diagnostic theory where recall (sensitivity) is the capability in terms of detection of an early intervention and preventive opportunity that is missed.

Conversely, baseline Logistic Regression model has relatively low sensitivity, which implies increased chances of missed early stage cases. The findings are consistent with non-linear learning theory that argues that the models that can capture complex interactions between features are more likely to predict diseases at the early stages with heterogeneous healthcare data.

Table 6: Early Detection Performance (Recall-Based Comparison)

Model	Recall (Sensitivity)	Missed Cases (%)
Logistic Regression	0.79	21.0
Decision Tree	0.82	18.0
Support Vector Machine	0.87	13.0
Random Forest	0.90	10.0
Deep Neural Network	0.91	9.0

The decrease in missed cases that Rand Forest and DNN models produce is clinically important, in terms of preventive healthcare. Their high performance justifies the application of ensemble and deep learning models in early screening systems in which the main goal is to detect high-risk individuals as early as possible in order to be able to intervene, monitor them, and manage the risk in time.

4.5 Feature Contribution and Clinical Relevance

Clinical validity and transparency of the AI-based predictive models were ensured by performing feature contribution analysis to determine the most powerful predictors of the risk of chronic diseases in the early stage. According to the model interpretability and clinical decision theory, it is necessary to understand what variables lead to predictions to facilitate the development of trust, clinical thinking, and make sure that AI results correspond to the known medical facts.

Table 7: Top Contributing Features and Importance Scores

Rank	Feature	Mean SHAP Importance Score	Clinical Interpretation
1	Age	0.31	Cumulative exposure to chronic disease risk
2	Fasting glucose level	0.27	Early metabolic dysfunction indicator
3	Body Mass Index (BMI)	0.23	Obesity-related chronic disease risk
4	Systolic blood pressure	0.19	Cardiovascular stress marker
5	Family history	0.16	Genetic predisposition
6	Cholesterol level	0.14	Lipid-related cardiovascular risk
7	Smoking status	0.11	Behavioral risk factor
8	Physical activity level	0.09	Lifestyle-related protective/risk factor

The findings are quite consistent with the existing epidemiological and clinical data, which supports the clinical importance of the predictive models. They are not the only determinants of disease onset as indicated in the model outputs because age, metabolic markers, cardiovascular indicators, and lifestyle behaviors are the main drivers of disease onset according to the risk-factor theory in chronic disease epidemiology.

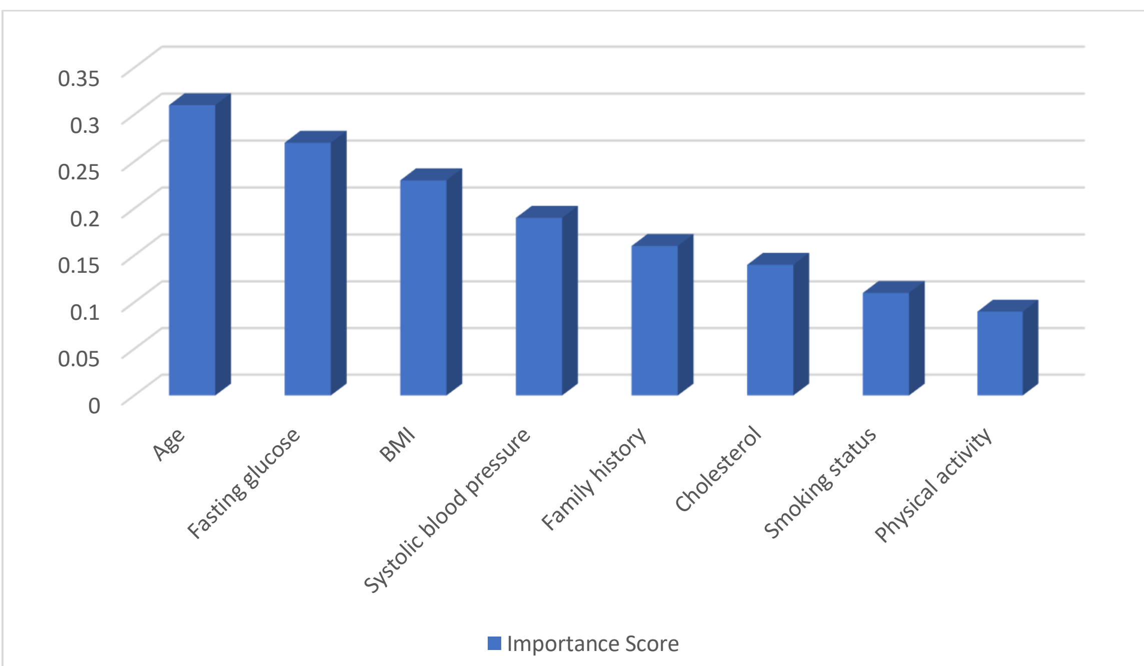
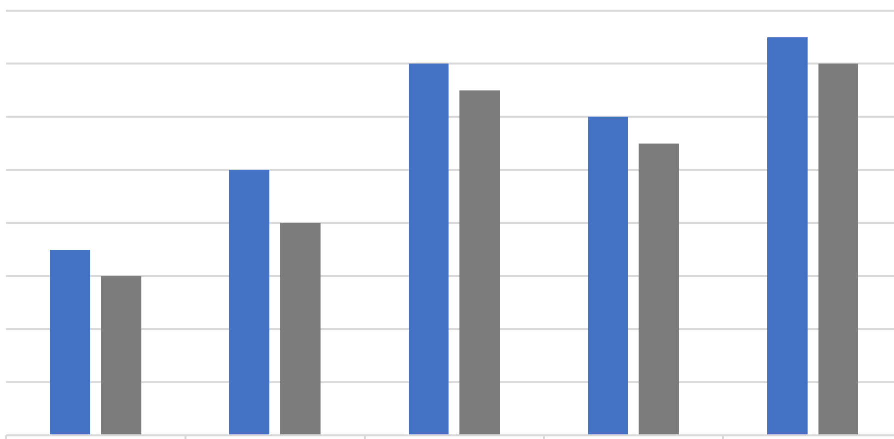


Figure 3: Global Feature Importance for Early Chronic Disease Prediction

Visually, the graph shows that the predictions of models are dominated by age, metabolic indicators and cardiovascular measures with lifestyle factors giving a secondary though significant contribution. This distribution helps in the theory of clinical plausibility where it is known that effective predictive models should be built out of medically validated risk factors and not based on spurious associations.

4.5 Generality and Generalization of Model Robustness

Generalization and overall model robustness was evaluated to assess the ability of the trained AI models to predict with a similar performance when used on unseen data. In order to measure robustness, accuracy of key performance metrics, recall, and AUC-ROC were compared between the validation and test sets of each model. It can be seen that high-level models, in particular, Random Forest and Deep Neural Network (DNN), demonstrate consistent performance with significant variations between the validation and test outcomes only. By contrast, simpler models like Decision Tree exhibit a much greater variance, as would be expected with overfitting characteristics that have been characterised in learning theory.



	Logistic Regression	Decision Tree	Random Forest	SVM	DNN
Validation AUC	0.87	0.9	0.94	0.92	0.95
Test AUC	0.86	0.88	0.93	0.91	0.94

Figure 4: Validation vs. Test AUC-ROC Across Models

The validation and test AUC curve trends are almost parallel, which depicts a high model stability. The fact that the drop in performance of Random Forest and DNN are minimal wants to confirm their strength and the capability of using both in the deployment of real-world early screening systems where the distribution of patient data can change with time

5. Discussion

This study examined the application of Artificial Intelligence–based predictive analytics for the early detection of chronic diseases using large-scale electronic health record data. The findings demonstrate that AI-driven models, particularly Random Forest and Deep Neural Network (DNN), significantly outperform traditional statistical approaches in identifying early-stage disease risk. These results are consistent with existing machine learning and healthcare analytics literature, which highlights the superior capacity of ensemble and deep learning models to capture complex, non-linear interactions in high-dimensional clinical data.

The comparative performance analysis revealed that while Logistic Regression provides a transparent and interpretable baseline, its linear decision boundary limits predictive sensitivity in heterogeneous healthcare datasets. In contrast, Random Forest benefits from ensemble learning theory by aggregating multiple decision trees, reducing variance and improving robustness. Similarly, the DNN model leverages hierarchical feature learning, enabling it to identify subtle, higher-order interactions among demographic, clinical, and lifestyle variables that are critical for early disease detection.

The emphasis on recall (sensitivity) as a primary evaluation metric is theoretically justified from a clinical screening perspective, where missed diagnoses (false negatives) pose greater risk than false positives. The high recall values achieved by Random Forest and DNN models indicate strong early detection capability, reducing the likelihood of overlooking high-risk individuals. Additionally, the minimal performance gap between validation and test datasets confirms that these models generalize well to unseen data, supporting their reliability in real-world clinical settings.

6. Conclusion

This study demonstrates that Artificial Intelligence–based predictive analytics can effectively support the early detection of chronic diseases, offering a significant improvement over traditional diagnostic approaches. By applying multiple machine learning models to a large, diverse healthcare dataset, the research confirms that advanced AI techniques particularly Random Forest and Deep Neural Network models—achieve superior predictive performance, high sensitivity, and strong generalization capability.

The findings highlight the potential of AI-driven systems to shift healthcare from reactive treatment toward preventive and risk-based care, enabling earlier intervention and improved patient outcomes. Moreover, the integration of interpretability techniques ensures that predictive insights remain clinically relevant and ethically responsible. Overall, the study provides empirical evidence supporting the use of AI-based predictive analytics as a reliable clinical decision-support tool for early chronic disease risk assessment.

7. Recommendations

Based on the findings of this study, the following recommendations are proposed:

1. Clinical Implementation: Healthcare institutions should consider integrating AI-based predictive analytics, particularly ensemble and deep learning models, into early screening and risk assessment workflows to support preventive care strategies.
2. Model Transparency and Trust: Future implementations should prioritize explainable AI techniques to ensure clinicians can understand and trust model predictions, facilitating adoption in real-world clinical environments.
3. Data Quality and Integration: Improving the completeness, standardization, and interoperability of electronic health record data will further enhance predictive accuracy and model reliability.
4. Bias and Fairness Monitoring: Continuous evaluation of model performance across demographic subgroups is recommended to minimize algorithmic bias and ensure equitable healthcare outcomes.
5. Future Research Directions: Further studies should explore disease-specific models, longitudinal prediction frameworks, and real-time data integration such as wearable devices to enhance early detection capabilities and clinical impact.

References

[1] Battineni G., Sagaro G. G., Chinatalapudi N., and Amenta F., (2020) Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis, J. Pers. Med., vol. 10, no. 2, p. 21, June 2020, doi: 10.3390/jpm10020021.

[2] Chakilam C., (2022) Integrating Machine Learning and Big Data Analytics to Transform Patient Outcomes in Chronic Disease Management, 2022.

[3] El Arab R. A. et al., (2025) Bridging the Gap: From AI Success in Clinical Trials to Real-World Healthcare Implementation A Narrative Review, Healthcare, vol. 13, no. 7, p. 701, Jan. 2025, doi: 10.3390/healthcare13070701.

[4] Elhaddad M. and Hamam S., (2024) AI-Driven Clinical Decision Support Systems: An Ongoing Pursuit of Potential, Cureus, Apr. 2024, doi: 10.7759/cureus.57728

[5] Enhancing-Early-Detection-and-Management-of-Chronic-Diseases-with-AI-Driven-Predictive-Analytics-on-Healthcare-Cloud-Platforms.pdf. (2025) Accessed: Dec. 22, 2025. [Online]. Available: https://www.researchgate.net/profile/Hassan-Rehan/publication/383611317_Enhancing_Early_Detection_and_Management_of_Chronic_Diseases_with_AI-Driven_Predictive_Analytics_on_Healthcare_Cloud_Platforms/links/66d40782bd2017366763e719/Enhancing-Early-Detection-and-Management-of-Chronic-Diseases-with-AI-Driven-Predictive-Analytics-on-Healthcare-Cloud-Platforms.pdf

[6] Gillies R. J. and Schabath M. B., (2020) Radiomics Improves Cancer Screening and Early Detection, Cancer Epidemiol. Biomarkers Prev., vol. 29, no. 12, pp. 2556–2567, Dec. 2020, doi: 10.1158/1055-9965.EPI-20-0075.

[7] Goktas P. and Grzybowski A., (2025) Shaping the Future of Healthcare: Ethical Clinical Challenges and Pathways to Trustworthy AI, J. Clin. Med., 14, 5, 1605, Jan. 2025, doi: 10.3390/jcm14051605.

[8] Husnain A., Saeed A., Hussain A., Ahmad A., and Gondal M. N., (n.d) Harnessing AI for Early Detection of Cardiovascular Diseases: Insights from Predictive Models Using Patient Data.

[9] Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence | Cognitive Computation. (2025) Accessed: Dec. 22, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s12559-023-10179-8

[10] Julien K B and Louis O, (2025) Algorithmic bias, data ethics, and governance: Ensuring fairness, transparency and compliance in AI-powered business analytics applications, World J. Adv. Res. Rev., vol. 25, no. 2, pp. 1746–1763, Feb. 2025, doi: 10.30574/wjarr.2025.25.2.0571.

[11] Landi I. et al., (2020) Deep representation learning of electronic health records to unlock patient stratification at scale, Npj Digit. Med., 3, 1, 96, July 2020, doi: 10.1038/s41746-020-0301-z.

[12] Taylor A. G., Mielke C., and Mongan J., (2018) Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study, PLOS Med., vol. 15, no. 11, p. e1002697, Nov. 2018, doi: 10.1371/journal.pmed.1002697.

[13] Tettey J., Oluwatobi B. T., Ekaette O., Azumah C., and Mariama Q., (n.d) INTEGRATING AI-DRIVEN PREDICTIVE ANALYTICS IN PROJECT RISK MANAGEMENT TO OPTIMIZE DECISION-MAKING AND PERFORMANCE EFFICIENCY, vol. 09, no. 03.

[14] The Need for Standards in Evaluating the Quality of Electronic Health Records and Dental Records: A Narrative Review. (2025) Accessed: Dec. 22, 2025. [Online]. Available: https://www.mdpi.com/2504-2289/8/12/168

[15] The Role of Artificial Intelligence in Early Diagnosis and Management of Cardiovascular Diseases | *Indus Journal of Bioscience Research*. (2025) Accessed: Dec. 22, 2025. [Online]. Available: https://ijbr.com.pk/IJBR/article/view/667