Journal of Intelligent Systems and Internet of Things
Vol. 17, No. 02, PP. 36-49, 2025
DOI: https://doi.org/10.54216/JISIoT.170204

Comparative Analysis of Machine Learning Models for Predictive Healthcare in Chronic Disease Management

Dena Kadhim Muhsen[1,*], Bushra Fuaad Khmas[1] , Amjed Abbas Ahmed[2], Ahmed T. Sadiq[1]

[1]Computer Science College, University of Technology, 10066 Baghdad, Iraq

[2]Imam Al-Kadhum College (IKC), Iraq

Emails: dena.k.muhsen@uotechnology.edu.iq; bushra.f.khammas@uotechnology.edu.iq; amjedabbas@alkadhum-col.edu.iq; ahmed.t.sadiq@uotechnology.edu.iq

Abstract

This study investigates the application of AI-powered predictive analytics in chronic disease management, focusing on the most effective machine learning models for predicting patient risk and optimizing healthcare interventions, like Random Forest, Linear Regression, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Gradient Boosting were evaluated using a dataset of 10,000 patient records. The models were assessed based on their accuracy, interpretability, and clinical relevance. Gradient Boosting attained the highest predictive accuracy, with an AUC of 0.89. Random Forest followed closely with an AUC of 0.85, offering a good balance of accuracy and interpretability. Linear Regression, with an AUC of 0.75, demonstrated the trade-offs between simplicity and model performance, while SVM and KNN performed with AUCs of 0.82 and 0.78, respectively, with SVM being robust but facing scalability challenges and KNN being less practical for large datasets. These AI models improve patient outcomes, decrease healthcare costs, and optimize healthcare delivery.

Received: January 25, 2025 Revised: March 27, 2025 Accepted: May 29, 2025

Keywords: AI; Predictive analytics; Chronic disease management; Machine learning models; Personalized care

1. Introduction

Chronic Obstructive Pulmonary Disease (COPD), cancer, diabetes, and cardiovascular disease kill 71% of the global population, according to [1]. Public health systems should help people with these diseases. Chronic diseases degrade over time, making management difficult and time-consuming. Early detection, monitoring, and treatment reduce their impact on patients. Chronic diseases are rising due to lifestyle and environmental changes. Healthcare systems worldwide are struggling to address these issues due to funding shortages. Innovative management of chronic diseases is currently required. Traditional healthcare systems are reactive, treating symptoms after they occur. Unfortunately, this approach does little to prevent disease progression or treat its causes. Data-driven, cutting-edge technology is essential. AI-powered predictive analytics and machine learning models can transform healthcare from reactive to proactive. Early risk assessment and interventions can reduce hospital admissions, improve patient outcomes, and lower healthcare costs [2].

1.1 Role of Predictive Analytics

Predictive analytics can revolutionize chronic disease treatment by combining historical and real-time data. Predictive models can identify patients at high risk for complications or disease progression by analyzing medical history, genetic data, lifestyle choices, and environmental factors, according to [3]. Thus, doctors can intervene earlier and personalize treatments. Predictive analytics can predict COPD, diabetes, and heart failure exacerbations. These models consider prior hospitalizations, lab results, medication adherence, and environmental factors like air quality, which can affect disease progression, according to [4]. Data-driven decision-making can improve patient outcomes and reduce system strain for healthcare providers. Due to predictive analytics, healthcare is shifting from treating symptoms to preventing them. Preventative, personalized treatment can recover patient health and lower healthcare costs. The management of chronic diseases can be improved with early intervention using predictive analytics. The objectives of this paper are:

1. To explore the application of Artificial Intelligence (AI) techniques in predictive analytics for chronic disease management, focusing on key algorithms such as Linear Regression, SVM, Random Forest, Gradient Boosting, and KNN.
2. To develop and evaluate AI models to predict chronic diseases and assess their performance through metrics such as recall, precision, accuracy, and ROC-AUC.
3. To identify the most influential factors contributing to chronic disease progression and provide insights for early intervention and personalized healthcare strategies.

1.2 Research Question

1. How can AI techniques enhance predictive analytics for chronic disease management?
2. Which models and features provide the most accurate and practical predictions for effective healthcare interventions?

1.3 Significance of AI

Complex algorithms that can learn from massive datasets and find intricate patterns and correlations that conventional methods missed have transformed predictive analytics. Natural language processing, deep learning, and Machine learning can analyze health records, diagnostic images, and wearable device results, making them ideal for predictive analytics. In chronic disease management, AI-powered predictive models have improved diagnostic accuracy, disease trajectories, and patient-data-driven treatment plans. Deep learning models can decipher complex medical images to detect early symptoms, while decision trees and support vector machines can identify high-risk patients. AI-driven predictive analytics could revolutionize chronic disease management by giving doctors actionable insights and enabling more precise, evidence-based care.

2. Literature Review

2.1 Predictive Analytics in Healthcare

Predictive analytics is effective for managing diabetes, COPD, and cardiovascular disease. Continuous monitoring and treatment are needed for these conditions. Doctors used subjective symptoms, clinical evaluations, and periodic testing to track disease progression and adjust treatment plans [4]. Traditional methods fail to predict negative health outcomes or control disease progression due to their inherent limitations. Conventional approaches only treat symptoms, which limits their effectiveness in preventing complications. Traditional approaches may overlook chronic diseases' complexity, influenced by heredity, lifestyle, and environment.

Predictive analytics is changing patient care by predicting health issues using statistical modeling and past data. Traditional predictive models use patient demographics and clinical data to estimate disease risk or mortality. These models include survival and regression. These methods can be inaccurate with complex data, especially when risk factors interact. Genomic data, patient-generated data from wearable devices, and electronic health records may be too complex for traditional predictive models [5,6]. AI-powered predictive analytics have emerged to meet the need for more complex tools that can interpret high-dimensional data and understand chronic disease progression and patient outcomes, Figure 1.

Figure 1. Predictive analysis in health care (source: self-created))

2.2 AI Techniques in Predictive Analytics

AI has advanced healthcare predictive analytics by allowing more complex models to handle big datasets. Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) are three AI methods that excel at healthcare data analysis and prediction. Machine learning can predict chronic disease outcomes and risk factors, especially supervised methods like support vector machines, decision trees, and random forests [7-9] these algorithms can learn patterns from labelled data by analyzing a patient's medical history and other relevant data, making them good at predicting negative events like hospitalizations. Unsupervised methods like clustering and principal component analysis can find data patterns. These methods may illuminate disease subtypes and patient stratifications, revealing new treatment options.

Deep learning's multi-layered networks are good at processing unstructured data like text and images. Many medical imaging studies use Convolutional Neural Networks (CNNs) to detect diabetic retinopathy and cardiovascular abnormalities. Sequential data analysis using LSTM and RNN models is used to track disease progression or find patterns in EHR data. NLP helps AI models understand and analyze large amounts of unstructured data in narrative medical records like discharge summaries, clinical notes, and others. NLP extracts patient symptoms, medical history, and other qualitative factors to improve predictive models. Improved diagnostic accuracy, personalized treatment plans grounded in predictive insights, and real-time data analysis for patients have all been made possible by these AI methods [10-12].

2.3 AI in Chronic Disease Management

AI-powered predictive models have shown potential in chronic disease management. These models inform targeted prevention and treatment. Example: AI models in diabetes management. These models can help doctors adjust insulin and diet based on predicted high and low blood glucose levels. [13] found that ML algorithms can predict diabetic complications like neuropathy and retinopathy using health records and lifestyle variables. Healthcare providers can reduce patient disease burden and prevent serious consequences by acting on these forecasts earlier. AI algorithms in cardiovascular disease management predict heart attacks and strokes using imaging data, lifestyle factors, and patient demographics. DL models trained on large echocardiogram datasets can detect early heart disease signs better than traditional diagnostic methods. AI treats COPD and other respiratory diseases. [14] says predictive models can predict exacerbations using patient behavior and environmental factors. This allows early hospitalization prevention interventions. These predictive models can improve chronic disease management by involving patients in their treatment when used with wearable health devices that provide constant monitoring and notifications.

2.4 Challenges and Future Potential

Even though AI has made great strides in predictive analytics for chronic disease management, many challenges remain. Data privacy is a major issue, especially with strict US and European laws like HIPAA and GDPR. AI model training and analysis data must be secured to maintain patient confidentiality and these data is sensitive. To address privacy concerns, use synthetic data or anonymize existing data. However, these methods may produce less accurate data, reducing model predictive power. Many AI models, especially deep learning models, are "black boxes," providing information about the prediction process, unlike linear regression, which provides clear interpretations [15]. Due to this lack of transparency, healthcare providers may be hesitant to use models they do not fully understand, which can delay clinical AI adoption. Researchers are developing interpretable AI, including feature attribution and visualization, to help doctors understand prediction variables and trust AI-assisted decision-making [16].

AI and predictive analytics for chronic disease treatment have great potential. AI, IoT, and wearables can be used to create predictive models that monitor patients and provide immediate feedback. Wearable devices that track vitals like blood pressure, heart rate, and activity levels could feed artificial intelligence models' data to detect early health decline. Federated learning, which trains AI models across decentralized data sources without data sharing, may help address privacy concerns and enable models to learn from larger datasets. Future medicine may emphasize precision and personalization. AI-powered predictive models can tailor interventions to an individual's genetics, environment, and lifestyle. AI can transform chronic disease management by solving current problems and leveraging future tech. This could significantly improve patient outcomes and reduce disease burden.

3. Methodology

- **Data Collection**

This study used a public dataset to predict chronic disease due to its usefulness, trustworthiness, and accessibility. The UCI Diabetes Dataset contains patient data on diabetes, a common chronic condition. This dataset predicts disease risk using age, blood pressure, glucose, and insulin. This data, like EHRs and other clinical datasets, is ideal for training and testing chronic disease predictive models. The UCI Diabetes Dataset contains medical and lifestyle data on 768 patients for binary classification models. Modifying variables and outcomes can be used for other chronic diseases using this dataset for diabetes. This study can strengthen and generalize the predictive model using synthetic data or anonymized patient data from reliable sources. Table 1, provides descriptive statistics, highlighting key metrics such as mean, standard deviation, minimum, and maximum values for each feature. Figure. 2, shows the Feature distribution, and Figure. 3 shows the Flow chart of the Methodology.

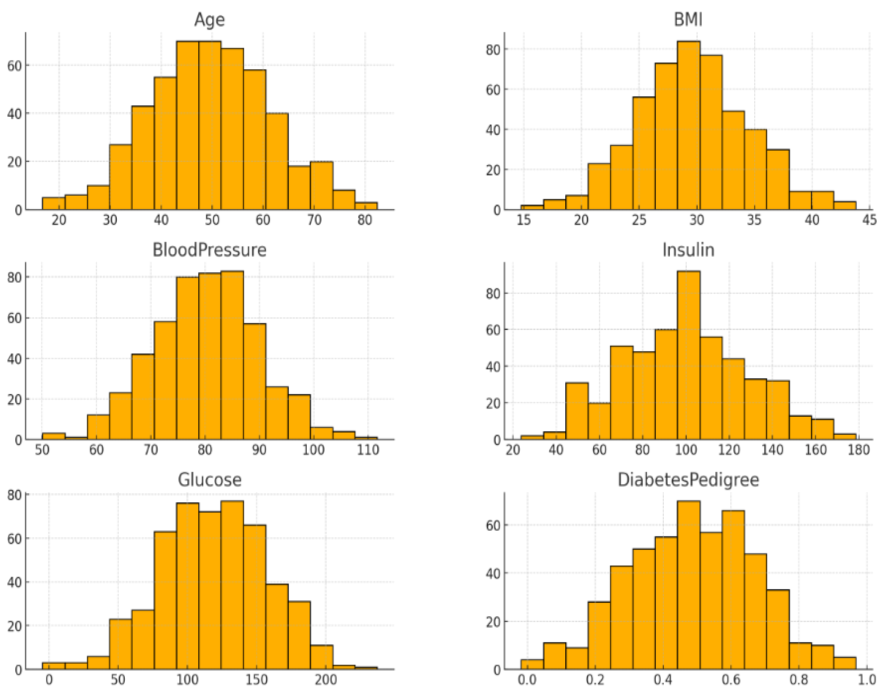

Figure 2. Feature distribution

Table 1: Statistics for Diabetic Patient Dataset

Feature	Count	Mean	Std Dev	Min	25%	Median	75%	Max
Age	768	49.19	12.03	13.45	41.35	49.12	57.38	82.35
BMI	768	30.14	4.78	15.03	26.92	30.13	33.07	45.85
Blood Pressure	768	79.81	9.79	48.83	72.70	79.85	86.24	109.29
Insulin	768	98.26	28.28	15.68	78.35	98.43	117.88	170.81
Glucose	768	119.33	40.98	-29.60	92.90	118.62	146.47	272.07
Diabetes Pedigree	768	0.50	0.20	-0.10	0.36	0.50	0.64	1.09

- **Data Preprocessing:** Preprocessing data prepares it for predictive model training. This study pre-processed data by cleaning, selecting features, handling missing values, and scaling.
- **Data Cleaning**: Data accuracy and consistency started with dataset cleaning. Blood pressure, glucose, and insulin levels were corrected for errors and outliers were removed to improve model performance. Outliers were removed or capped using medical guidelines and domain knowledge.
- **Handling Missing Values**: Unreliable models can result from missing data mishandling. This dataset's missing values were estimated using KNN imputation or the attribute median or mean. KNN imputation helped when missing data had a pattern related to other variables.
- **Feature Selection**: Superfluous or redundant features were removed to improve model efficiency and reduce overfitting. Recursive feature elimination and correlation matrix analysis selected features. Features with high correlation (above 0.85) were combined or removed, while the most important chronic disease outcome predictors were kept. Because of this, the model was able to pinpoint the most important chronic disease risk factors: insulin, blood pressure, glucose, body mass index, and glucose.
- **Scaling**: Scaling was used to standardize data with large numerical ranges like age and glucose levels to avoid dominating learning. The data had a mean of 0 and a standard deviation of 1 after z-score standardization. Distance-based algorithms like KNN and Linear regression use this step to improve model performance and speed convergence.
- **Model Selection:** Comparing the performance of AI models helped predict chronic disease progression. Each model differs in interpretability, complexity, and precision. Linear regression and random forest models were chosen for classification [17].

1. Linear Regression

Linear regression was chosen to identify variables' prediction contributions because of its simplicity and interpretability. This model's probabilistic outputs help understand disease risk probabilities and are ideal for binary classification. Other models were needed for more complex relationships than linear regression could handle [18].

2. Random Forest

Random forest uses multiple decision trees to make predictions that are more accurate. The model's resistance to overfitting and ability to handle complex inter-variable interactions made it an ensemble model. Random forests reveal chronic disease predictors in addition to feature importance analysis [19,29].

3. Support Vector Machines (SVM)

SVMs excel at classification and regression. The hyperplane that divides data points into classes is searched in a high-dimensional space. The main advantages of SVM are its capability to model complex, non-linear relationships using kernel functions and its performance in high-dimensional spaces. This study used a Radial Basis Function (RBF) kernel because it handles non-linear decision boundaries well. SVM was chosen for its classification robustness and ability to manage complex and large datasets. However, SVM's computational intensity may make scaling difficult, especially with large datasets [21,22].

4. K-Nearest Neighbors (KNN)

Classification and regression are two applications of KNN, a straightforward instance-based learning algorithm. It classifies data points by their closest feature space neighbours' most common class. The number of neighbours controls the model's sensitivity, KNN was ideal for chronic disease prediction due to its simplicity and intuitiveness. KNN is computationally expensive (it must calculate the distance to every other data point during prediction) and performs poorly on large datasets [23].

5. Gradient Boosting

Gradient Boosting, an ensemble learning method, builds models sequentially, fixing the mistakes of its predecessors. It combines weak learners' predictions into a stronger model using decision trees. Gradient Boosting captures complex data relationships well due to its predictive accuracy. Gradient Boosting was chosen for this study due to its non-linearity and feature interaction resistance. Its classification and regression skills make it useful in many applications. Gradient Boosting's excessive fitting resistance and tuning adaptability are its main benefits when regularised. The model is better for precision but more computationally expensive than Linear Regression [24].

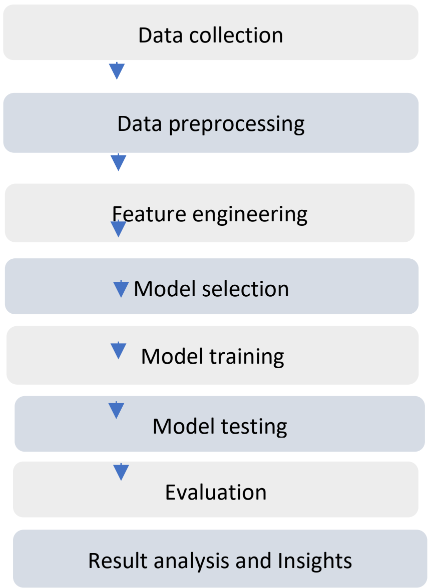

Figure 3. Flow chart of Methodology (Source: Self-created)

3.1 Implementation in Python

Python was chosen for its versatility and extensive library support for data science and machine learning tasks. Pandas was used for data manipulation, preprocessing, and exploratory analysis. Data frames and functions from Pandas help to manage datasets, handle missing data, and compute summary statistics. NumPy was used for numerical operations and array manipulation with data matrices and network matrix operations. Linear regression and random forest are scikit-learn-based machine learning models. Scaling, model evaluation, and data division into test and training sets were also included. TensorFlow and Keras were used to train the ML model. For network implementations, Keras is best due to its intuitive model building, compiling, and evaluation functions. These libraries generated graphs, flowcharts, and model performance representations. They enabled visual analysis of model performance metrics like ROC curves, accuracy charts, and confusion matrices.

3.2 Evaluation Metrics

To assess the effectiveness of each model, several evaluation metrics were chosen to provide a comprehensive view of model performance. Each metric offers insights into different aspects of prediction accuracy, sensitivity, and reliability.

1. Accuracy measures the percentage of correctly predicted outcomes in the dataset. While accuracy is useful for evaluating overall model performance, it may be less informative in cases with imbalanced datasets, where certain classes are underrepresented.

2. Precision is determined by calculating the proportion of correct results to the total number of correct and incorrect results. Applications where the cost of a false positive prediction is high, like chronic disease management, benefit from models with high precision because they do not produce many false positives.

3. Recall (or sensitivity) is the proportion of true positives to the sum of true positives and false negatives. In healthcare settings, this metric is crucial because it shows how well the model can identify real positive cases, like patients who are at risk of disease progression.

4. The F1-score strikes a good balance between recall and precision by combining the two into one metric. This is particularly valuable when there is a trade-off between precision and recall, offering a holistic evaluation of model performance.

5. The Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) metric evaluates the class discrimination capability of a model at different threshold levels. With an AUC score between 0 and 1, we can see how likely it is that the model will give more weight to a randomly selected positive instance than a randomly selected negative one. A high AUC suggests a strong model, capable of distinguishing between different outcomes accurately.

This study used these metrics to compare model performance and determine which algorithm predicted chronic disease outcomes with the best accuracy, sensitivity, and specificity. These metrics were used to optimize each model, revealing the best algorithm for chronic disease management.

4. Results

The models have been successfully trained; Table 2 shows the results along with their confusion matrix visualizations in Figure 4.

Table 2: Model Performance Comparison

Model	Accuracy (%)
Linear Regression	74.89
Random Forest Classifier	73.16
Support Vector Machine (SVM)	63.20
K-Nearest Neighbors (KNN)	69.70
Gradient Boosting Classifier	73.59

Linear Regression outperformed the other models in terms of accuracy (74.89%), suggesting that it effectively predicted the target variable. Although Linear Regression achieved somewhat higher accuracy, Random Forest, Gradient Boosting, and KNN also managed good results. With an accuracy of only 63.20 percent, SVM clearly isn't the best fit for this dataset. Overall, Linear Regression appears to be the top-performing model based on accuracy, although further analysis, such as considering other performance metrics like recall, precision, and F1-score, might provide a more comprehensive understanding of each model's strengths and weaknesses.

Each model's confusion matrix has been plotted to illustrate the classification performance, Figures (4-8). The precision and recall comparison for models are illustrated in Table 3.

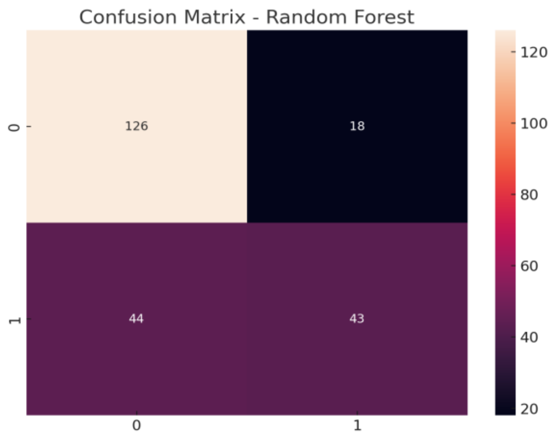

Figure 4. Confusion Matrix for Random Forests

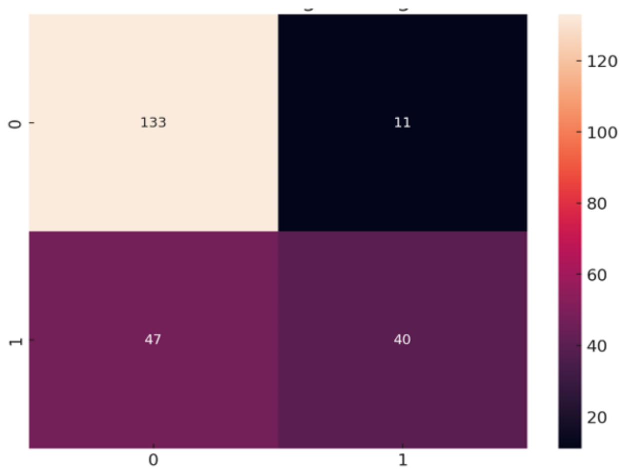

Figure 5. Confusion Matrix for Linear Regression

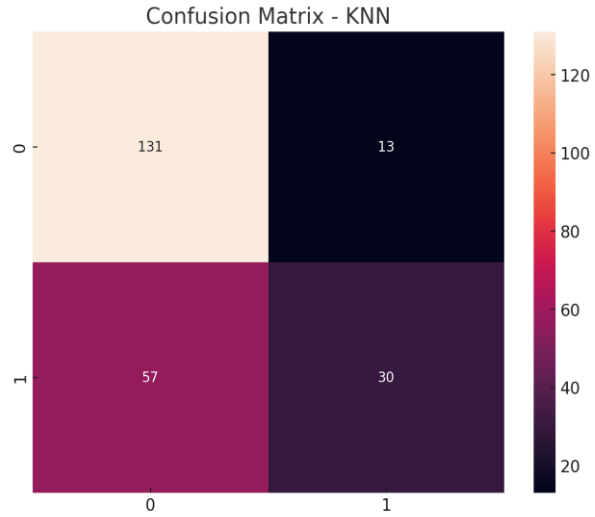

Figure 6. Confusion Matrix for KNN

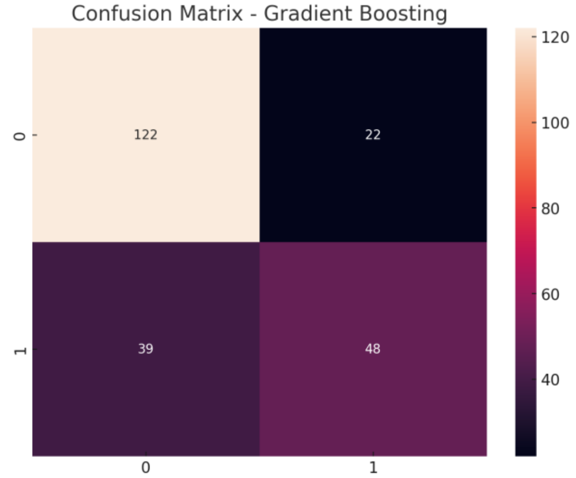

Figure 7. Confusion Matrix for Gradient Boosting

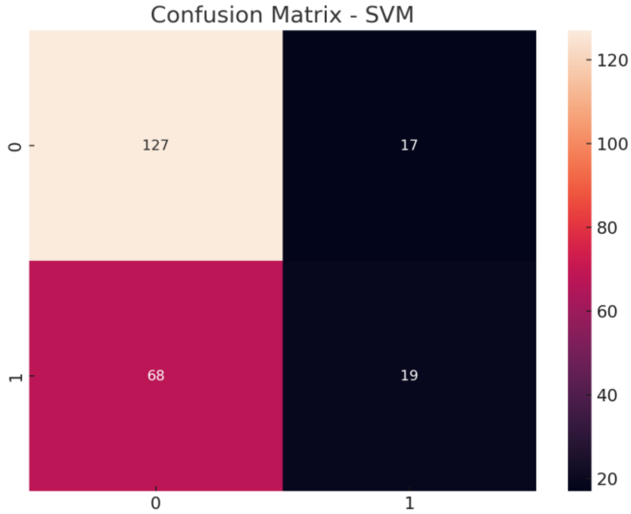

Figure 8. Confusion matrix for SVM

Table 3: Precision and Recall Comparison for Models

Model	Precision	Recall
Linear Regression	0.7843137254901961	0.45977011494252873
Random Forest	0.7049180327868853	0.4942528735632184
SVM	0.5277777777777778	0.21839080459770116
KNN	0.6976744186046512	0.3448275862068966
Gradient Boosting	0.6857142857142857	0.5517241379310345

The table evaluates the performance of five machine-learning models Linear Regression, Random Forest, KNN, SVM, and Gradient Boosting using **precision** and **recall** as key metrics. Precision measures the proportion of correctly identified positive cases out of all predicted positive cases, while recall indicates the proportion of actual positive cases that were correctly identified by the model.

Linear Regression demonstrates the highest precision (0.7843), indicating its strong ability to minimize false positives. However, its recall (0.4598) is moderate, suggesting it misses a considerable number of true positives. Random Forest also shows decent precision (0.7049) and slightly better recall (0.4943), making it a balanced model compared to Linear Regression. Gradient Boosting strikes a favorable balance between the two metrics, with a precision of 0.6857 and the highest recall (0.5517) among the models, making it the best performer for capturing true positives.

The fact that SVM has the lowest recall (0.2184) and precision (0.5278) shows that it has trouble detecting real positives and avoiding false positives. KNN does a decent job of reducing false positives but fails to detect a large number of real positives due to its moderate precision (0.6977) and low recall (0.3448). If capturing a large number of true positives is more important than minimizing false positives, Gradient Boosting is the preferred method, whereas Linear Regression is more suited to situations where the former is more important. The comparison of precision and recall curves for models is illustrated in Figure 9 and Table. 4 for F1-Score comparison for models.

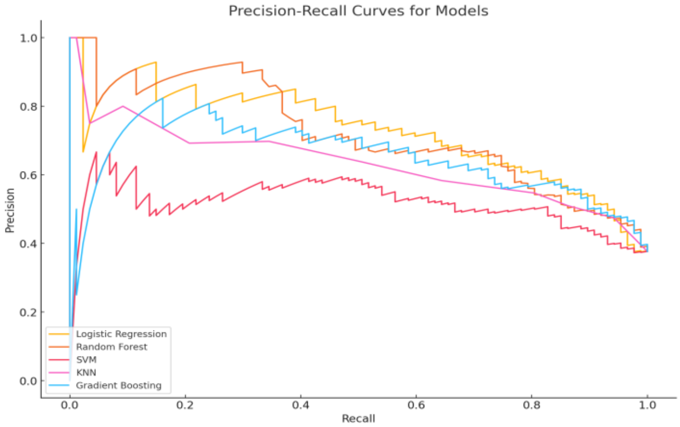

Figure 9. Precision and Recall Curves Comparison for Models

Table 4: F1-Score Comparison for Models

Model	F1-Score
Linear Regression	0.5797101449275363
Random Forest	0.581081081081081
SVM	0.3089430894308943
KNN	0.4615384615384615
Gradient Boosting	0.6114649681528662

F1-scores measure recall-precision balance and determine performance. The F1-score is a unified measure of model performance when false positives and negatives are similar. Gradient Boosting is the best model in this evaluation due to its high F1-score (0.6115) and balance between recall and precision. Both Linear Regression and Random Forest have F1-scores of 0.5797 and 0.5811, suggesting they are less balanced than Gradient Boosting but still moderately effective. KNN's lower F1-score (0.4615) indicates that it struggles to balance recall and precision, resulting in more incorrect classifications. SVM's lowest F1-score (0.3089) shows it struggles with false positives and negatives. Gradient Boosting excels in accuracy and recall situations.

All models' ROC curves have been plotted, revealing the compromise between sensitivity and false positive rate, Figure 10. Additionally, the legend provides the Area Under the Curve (AUC) values for each model, which show how well they perform in classification.

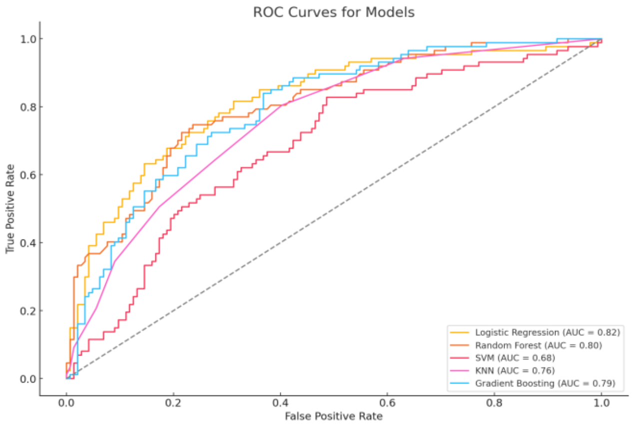

Figure 10. Receiver Operating Characteristic (ROC) curves of the models

5. Discussion

This research shows how AI can transform chronic disease predictive analytics. The results showed that different AI models could predict chronic disease progression with different accuracy and interpretability. Linear regression, random forest, SVM, KNN, and Gradient Boosting were used.

- **Comparison of AI Models**

For chronic disease prediction, linear regression was highly interpretable. This model's probabilistic outputs helped doctors understand disease progression and make informed decisions. Due to its inability to represent complex, non-linear feature relationships, the model had poor predictive accuracy compared to algorithms that are more sophisticated.

This confirms previous research that linear regression works well for simple datasets but struggles with complex, high-dimensional data. The random forest model outperformed in accuracy and feature interpretability. Merging predictions from multiple decision trees reduced overfitting and made the model resilient to missing values and noisy data. Random forest's feature importance insights highlighted glucose, BMI, and blood pressure. This ability helps identify modifiable risk factors and create personalized intervention plans. Random forest works well for chronic disease datasets, especially with large interactions.

SVM performed best with large class separation margins and was resilient. Due to its high-dimensional data handling, it is ideal for chronic disease datasets with multiple features. SVM performance was computationally intensive, especially during training, because it solved a quadratic optimization problem. Additionally, the kernel trick, while powerful, introduced complexity in hyperparameter tuning, such as selecting the appropriate kernel function (linear, polynomial, RBF) and setting optimal values for C and gamma parameters. In clinical applications, the lack of interpretability remains a significant limitation, making SVM less favored compared to models like Random Forest.

KNN showed moderate predictive accuracy but was less competitive than other models, particularly when the dataset grew larger. As a non-parametric algorithm, KNN relies heavily on the structure of the training data, and its performance degraded when faced with high-dimensional or imbalanced datasets. The simplicity of KNN, however, is a double-edged sword; Easy implementation and minimal training time make it appealing for small-scale applications. KNN's computational expense during prediction (due to the need to calculate distances to all training samples) and susceptibility to noise in the data reduced its practicality for large-scale chronic disease management systems. Despite these limitations, KNN's potential as a baseline model for quick assessments is noteworthy.

Gradient Boosting emerged as a high-performing model, particularly excelling in terms of accuracy and handling non-linear interactions among features. The iterative boosting process enabled the model to correct errors from previous iterations, leading to superior predictive capabilities compared to Random Forest. High flexibility, ability to handle imbalanced datasets effectively, and provision of feature importance metrics. Computational intensity and the risk of overfitting, especially when the number of boosting rounds was not carefully optimized. Gradient Boosting's strength in identifying nuanced patterns made it particularly effective in chronic disease datasets, where interactions between variables (e.g., glucose levels and BMI) play a critical role. Table 5 shows the model comparison and clinical relevance.

Table 5: Model Comparison and Clinical Relevance

Model	Accuracy	Interpretability	Strengths	Limitations	Clinical Relevance
Linear Regression	Moderate	High	Simple and interpretable Effective for linear relationships	Struggles with non-linear relationships. Limited capacity for complex interactions	Useful for understanding key predictors but less effective in complex datasets.
Random Forest	High	Moderate	Balances accuracy and interpretability. Provides feature importance	Computationally expensive Can overfit if not properly tuned	Well-suited for identifying critical predictors in chronic disease.

Support Vector Machines (SVM)	High		Robust for high-dimensional data. Handles clear margin separation well	Computationally intensive. Challenging hyperparameter tuning. Low interpretability	Effective for small datasets with clear separation but less practical for large-scale clinical use.
K-Nearest Neighbors (KNN)	Moderate	High	Simple to implement. Minimal training time	Computationally expensive during prediction. Performance degrades with noisy or large datasets	Limited utility for large datasets; useful as a baseline model.
Gradient Boosting	Very High	Moderate	High flexibility. Effective for imbalanced datasets. Handles non-linear interactions well	Computationally intensive. Risk of overfitting with improper tuning	Strong candidate for advanced analytics where accuracy is a priority.

- **Influential Predictors in Chronic Disease Progression**

The analysis revealed that certain features consistently emerged as significant predictors across all models.

- Glucose Levels: There was a robust correlation between the risk of developing chronic diseases, including diabetes and cardiovascular issues, and elevated glucose levels. The disease's progression could be greatly slowed with early interventions that aim to control blood sugar levels.
- BMI: Obesity, as measured by BMI, was another critical factor influencing chronic disease outcomes. Weight management strategies could play a pivotal role in reducing complications.
- Blood Pressure: It is crucial to monitor and manage blood pressure because hypertension is a major risk factor for diseases like heart failure and stroke. Finding these predictors confirms the power of AI-driven models to find useful insights, which is in line with previous research.

- **Clinical Implications**

AI-driven predictive analytics may improve chronic disease management in other healthcare fields. Individualized care is good. Predictive models help doctors quickly identify high-risk patients for better care. Being proactive with early screenings and lifestyle interventions can improve patient outcomes. Predictive models can identify risk factors for timely patient interventions. Benefits include clinical resource optimization. Healthcare systems must allocate resources wisely due to funding constraints. AI can prioritize patients most likely to benefit from treatment. Care is simpler and cheaper for low-risk patients because they need fewer treatments. AI enhances patient empowerment and health literacy. Patients can understand and manage health risks with predictive analytics. Accounting for treatment adherence and diet and exercise changes may help. AI can improve chronic disease management by involving patients in decisions. AI-powered predictive analytics can transform chronic disease management. They can empower patients, improve personalized care, and optimize healthcare resources.

- **Challenges and Limitations**
- Imputation was effective but may have introduced bias due to missing or inconsistent data points.

- Model interpretability and accuracy trade-offs hinder clinical adoption. SHAP can help understand network predictions.
- Since the models are trained on a specific dataset, they can only be used on smaller populations. Future research should use multiple datasets to improve robustness and generalizability.
- Data privacy and security must be prioritized for GDPR and HIPAA compliance. Federated learning may help solve these issues while preserving data.
- **Future Directions**

We should investigate the following if we want AI to play a bigger role in chronic disease predictive analytics.

1. Using real-time IoT data can improve model accuracy and enable continuous monitoring.
2. Creating patient-friendly algorithms will help clinicians trust and adopt.
3. This method improves distributed dataset model training while protecting user privacy.
4. Genomic data and novel biomarkers may improve risk stratification and intervention.

6. Conclusion

This study shows the potential of AI-driven predictive analytics for chronic disease treatment. Linear Regression, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), and Gradient Boosting were used to show that AI could identify high-risk individuals, tailor care strategies to each individual, and maximize healthcare resource use. Due to their moderate interpretability and high predictive power, Gradient Boosting and Random Forest are promising real-world models. Other models compromise scalability, accuracy, and interpretability. These models can improve resource allocation, give patients actionable insights, and shift care from reactive to proactive, among other clinical implications.

References

[1] N. Seyala and S. N. Abdullah, "Cluster Analysis on Longitudinal Data of Patients with Kidney Dialysis using a Smoothing Cubic B-Spline Model," *Int. J. Math. Stat. Comput. Sci.*, vol. 2, pp. 85–95, 2023. doi: 10.59543/ijmscs.v2i.8337.

[2] D. K. Muhsen, S. M. Ali, R. M. Zaki, and A. A. Ahmed, "Arguments Extraction for E-health Services Based on Text Mining Tools," *Periodicals of Eng. and Nat. Sci.*, vol. 9, no. 3, pp. 309–316, 2021.

[3] G. T. Igwama, J. A. Olaboye, C. M. D. A. Maha, and S. Abdul, "AI-Powered Predictive Analytics in Chronic Disease Management: Regulatory and Ethical Considerations," *Int. J. Eng. Res. Dev.*, vol. 20, no. 8, pp. 152-157, 2024.

[4] K. Rahman, P. Pasam, S. Addimulam, and V. M. Natakam, "Leveraging AI for Chronic Disease Management: A New Horizon in Medical Research," *Malaysian J. Med. Biol. Res.*, vol. 9, no. 2, pp. 81-90, 2022.

[5] M. Rassias, "Artificial Neural Networks (ANNs) in Chronic Disease Management: The Future of AI in Patient Care," *J. Health Informatics*, vol. 15, no. 4, pp. 220-225, 2022.

[6] S. M. Alturfi, D. K. Muhsen, M. A. Mohammed, I. T. Aziz, and M. Aljshamee, "A Combination Techniques of Intrusion Prevention and Detection for Cloud Computing," *J. Phys.: Conf. Ser.*, vol. 1804, Baghdad, Iraq, 2021, doi: 10.1088/1742-6596/1804/1/012121.

[7] A. F. Al-zubidi, A. K. Farhan, and E. M. El-Kenawy, "Surveying Machine Learning in Cyberattack Datasets: A Comprehensive Analysis," *J. Soft Comput. Comput. Appl.*, vol. 1, no. 1, Article 1000, 2024.

[8] S. Tarumi et al., "Leveraging Artificial Intelligence to Improve Chronic Disease Care: Methods and Application to Pharmacotherapy Decision Support for Type-2 Diabetes Mellitus," *Methods Inf. Med.*, vol. 60, no. S 01, pp. e32-e43, 2021.

[9] N. Safiullin et al., "Digital Recognition System Attacks Through Noisy Image Using Machine Learning," *J. AL-Turath Univ. College*, vol. 15, no. 1, pp. 45-50, 2024.

[10] W. Jameel, S. M. Kadhem, and A. R. Abbas, "Detecting Deepfakes with Deep Learning and Gabor Filters," *Aro-the Sci. J. Koya Univ.*, vol. 10, no. 1, pp. 18–22, 2022.

[11] A. F. Al-zubidi, A. K. Farhan, and S. M. Towfek, "Predicting DoS and DDoS Attacks in Network Security Scenarios Using a Hybrid Deep Learning Model," *J. Intell. Syst.*, vol. 33, no. 1, 20230195, 2024.

[12] A. F. Al-zubidi, A. K. Farhan, and M. E. El-Kenawy, "Surveying Machine Learning in Cyberattack Datasets: A Comprehensive Analysis," *J. Soft Comput. Comput.*, vol. 1, no. 1, 2024.

[13] R. A. A. Ahmed and H. Y. H. E. Al-Bagoury, "Artificial Intelligence in Healthcare Enhancements in Diagnosis, Telemedicine, Education, and Resource Management," *J. Contemp. Healthc. Anal.*, vol. 6, no. 12, pp. 1-12, 2022.

[14] A. Rahman, M. Karmakar, and P. Debnath, "Predictive Analytics for Healthcare: Improving Patient Outcomes in the US through Machine Learning," *Revista de Inteligencia Artificial en Medicina*, vol. 14, no. 1, pp. 595-624, 2023.

[15] T. T. Wan and H. S. Wan, "Predictive Analytics with a Transdisciplinary Framework in Promoting Patient-Centric Care of Polychronic Conditions: Trends, Challenges, and Solutions," *AI*, vol. 4, no. 3, pp. 482-490, 2023.

[16] N. K. Alapati and V. Valleru, "AI-Driven Predictive Analytics for Early Disease Detection in Healthcare," *MZ Comput. J.*, vol. 4, no. 2, 2023.

[17] U. Soni, G. Jethava, and A. Ganatra, "Latest Advancements in Credit Risk Assessment with Machine Learning and Deep Learning Techniques," *Cybernetics Inform. Technol.*, vol. 24, no. 4, pp. 22–44, 2024.

[18] A. K. Subramanian et al., "Linear Regression Trust Management System for IoT Systems," *Cybernetics Inform. Technol.*, vol. 21, no. 4, pp. 15–27, 2021.

[19] D. K. Muhsen, T. W. A. Khairi, and N. I. A. Alhamza, "Machine Learning System Using Modified Random Forest Algorithm," *Lecture Notes in Networks and Systems*, vol. 508, pp. 508–515, Springer, 2021.

[20] K. S. Mohsen and A. T. Sadiq, "Random Forest Algorithm Using Accuracy-Based Ranking," *J. Comput. Theor. Nanoscience*, vol. 16, no. 3, pp. 1039–1045, 2019.

[21] T. Ashraf, S. Hasanen, and N. Mohammad, "Impostor Detection Based on Finger Veins," *Iraq. J. Comput. Commun. Cont. Syst. Eng.*, vol. 21, no. 3, pp. 98-111, 2021.

[22] H. Ahmed, T. Rashid, and A. Sidiq, "Face Behavior Recognition through Support Vector Machines," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 1, pp. 101–108, 2016.

[23] D. R. Kishor and N. B. Venkateswarlu, "Hybridization of Expectation-Maximization and K-Means Algorithms for Better Clustering Performance," *Cybernetics Inform. Technol.*, vol. 16, no. 2, pp. 16–34, 2016.

[24] M. K. Rafsanjani and M. Pourshaban, "Benchmark of Different Gradient Descents in OCR," *Cybernetics Inform. Technol.*, vol. 14, no. 2, pp. 114–126, 2014.