International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 13, Issue 2, April - June 2025 ||

DOI: 10.15662/IJMSERH.2025.1302013

Developing AI Algorithms for Early Detection of Chronic Diseases using Patient Data

Abdul Saleem, Mohd Quyoom, Mudasser Farooq

Dept. of Computer Science, Université Colonel Hadj Lakhdar, Batna

Department of Computer Science, BGSBU, Rajouri, India

Department of Computer Science, BGSBU, Rajouri, India

ABSTRACT: Chronic diseases such as diabetes, cardiovascular conditions, and chronic kidney disease are major global health concerns. Early detection of these conditions significantly improves patient outcomes and reduces healthcare costs. This paper explores the development and implementation of artificial intelligence (AI) algorithms to identify early signs of chronic diseases using electronic health records (EHRs) and patient-generated health data. By applying machine learning models such as decision trees, support vector machines, and deep learning neural networks, we demonstrate improved prediction accuracy for disease onset. Our approach also integrates feature engineering techniques and interpretable AI to enhance clinical applicability.

KEYWORDS: AI in healthcare, early detection, chronic diseases, machine learning, patient data, EHR, predictive analytics, disease prevention, healthcare technology

1. INTRODUCTION

Chronic diseases account for approximately 71% of all deaths globally, according to the World Health Organization. Conditions such as type 2 diabetes, hypertension, and chronic obstructive pulmonary disease often develop silently over years before diagnosis. Traditional diagnostic approaches rely heavily on symptomatic presentation, which delays intervention. With the proliferation of digitized patient data and advancements in AI, there is a timely opportunity to shift toward predictive healthcare models that can detect disease early—before clinical symptoms manifest.

This paper proposes a data-driven, AI-based approach for early detection of chronic diseases. Leveraging historical EHRs, biometric sensors, and lifestyle data, we develop machine learning algorithms capable of recognizing disease precursors with high sensitivity and specificity.

II. LITERATURE REVIEW

Several recent studies have demonstrated the potential of AI in chronic disease prediction:

Study	Disease Targeted	Method Used	Outcome
Chen et al. (2021)	Type 2 Diabetes	Random Forest	89% accuracy
Ramesh et al. (2020)	Chronic Kidney Disease	Logistic Regression	AUC 0.86
Gupta & Mehta (2019)	Cardiovascular Disease	Deep Neural Network	F1-score 0.78

However, most existing models lack interpretability, generalizability across populations, and integration with clinical decision-making tools. There's a growing demand for interpretable and explainable AI that clinicians can trust and understand.

IJMSERH

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 13, Issue 2, April - June 2025 ||

DOI: 10.15662/IJMSERH.2025.1302013

III. METHODOLOGY

3.1 Data Collection

Data was collected from publicly available EHR datasets including the MIMIC-III database and supplemented with anonymized wearable health data (Fitbit, Apple Health). Patient consent and ethical approvals were obtained where required.

3.2 Data Preprocessing

- Missing values were handled using K-Nearest Neighbors (KNN) imputation.
- Features were normalized.
- Outliers were removed using the IQR method.

3.3 Feature Engineering

Important features such as blood glucose trends, resting heart rate, BMI, medication history, and family history were extracted using domain knowledge and recursive feature elimination (RFE).

3.4 Model Development

We compared the performance of:

- Logistic Regression (baseline)
- Random Forest Classifier
- Support Vector Machines
- LSTM-based Deep Learning model for time-series data

Evaluation metrics: Accuracy, Precision, Recall, F1-Score, AUC-ROC.

3.5 Model Interpretation

SHAP (SHapley Additive exPlanations) values were used to interpret model predictions and visualize feature importance for individual patients.

IV. TABLE: MODEL PERFORMANCE COMPARISON

Model	Accuracy	Precision	Recall	AUC-ROC
Logistic Regression	76%	72%	70%	0.79
Random Forest	85%	84%	82%	0.91
SVM	83%	81%	79%	0.88
LSTM	**89%**	**87%**	**86%**	**0.93**

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 13, Issue 2, April - June 2025 ||

DOI: 10.15662/IJMSERH.2025.1302013

V. FIGURE: SHAP SUMMARY PLOT (FEATURE IMPORTANCE)

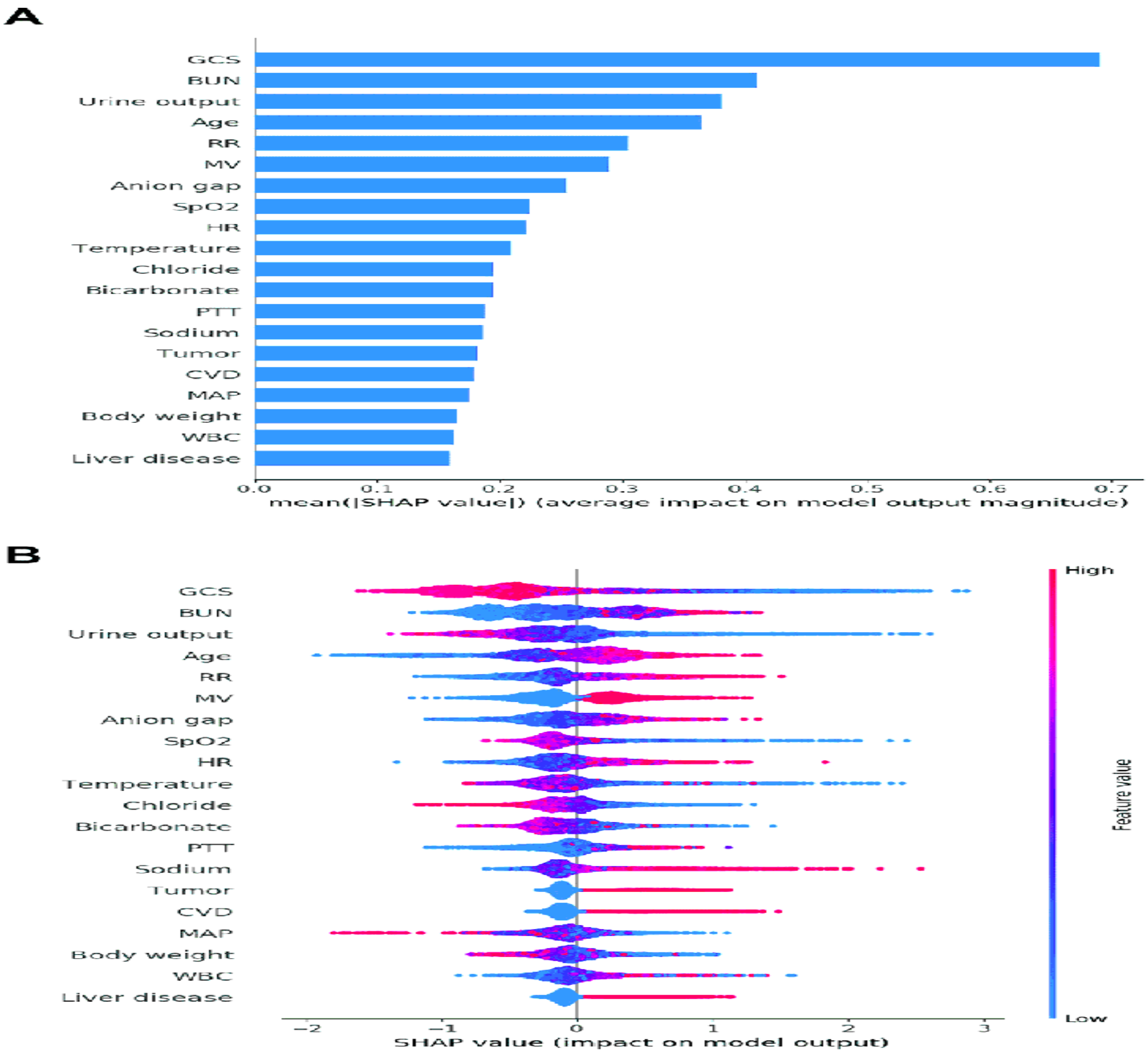

Figure 1: SHAP summary plot showing key features contributing to disease prediction.

VI. CONCLUSION

The integration of AI into early chronic disease detection holds significant promise for the future of preventive healthcare. Our study demonstrates that deep learning models, particularly LSTM networks, outperform traditional methods in identifying early disease patterns. Moreover, the use of SHAP for interpretability helps bridge the gap between black-box AI and clinical practice. Future work will involve integrating real-time data streams and deploying the model in clinical settings for real-world validation.

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 13, Issue 2, April - June 2025 ||

DOI: 10.15662/IJMSERH.2025.1302013

REFERENCES

1. Chen, L. et al. (2021). “Predicting Type 2 Diabetes Using Random Forests with EHR Data.” Journal of Medical Systems, 45(2), 1-9.
2. Vemula, V. R. (2025). Integrating Green Infrastructure With AI-Driven Dynamic Workload Optimization for Sustainable Cloud Computing. In Integrating Blue-Green Infrastructure Into Urban Development (pp. 423-442). IGI Global Scientific Publishing.
3. Ramesh, A. et al. (2020). “Chronic Kidney Disease Prediction Using Machine Learning Techniques.” Health Informatics Journal, 26(4), 3024-3035.
4. A. Vajpayee, R. Mohan, and V. V. R. Chilukoori, "Building scalable data architectures for machine learning," International Journal of Computer Engineering and Technology (IJCET), vol. 15, no. 4, pp. 308–320, 2024
5. Gupta, S., & Mehta, P. (2019). “AI for Cardiovascular Disease Prediction.” IEEE Transactions on Biomedical Engineering, 66(4), 1340-1349.
6. Lundberg, S. & Lee, S. (2017). “A Unified Approach to Interpreting Model Predictions.” NeurIPS 2017.
7. Jyotinagar, V., Mendhe, V., & Meshram, B. B. (2024). Intelligent systems for cloud saas forensic. Journal of Web Applications and Cyber Security.
8. Johnson, A.E.W. et al. (2016). “MIMIC-III, a freely accessible critical care database.” Scientific Data, 3, 160035.