Advancing Predictive Analytics in Healthcare: Integrating Multimodal Machine Learning for Real-Time Early Detection and Prevention of Chronic Diseases

Sunil Kumar[1], Kishori Lal Bansal[1], J. Veni[2], K. Akila[3], V. Kavin[3] and Syed Zahidur Rashid[4]
[1]*Department of Computer Applications, Himachal Pradesh University, Shimla-5, Himachal Pradesh, India*
[2]*Department of MBA, J.J. College of Engineering and Technology, Tiruchirappalli, Tamil Nadu, India*
[3]*Department of Management Studies, Nandha Engineering College, Erode-638052, Tamil Nadu, India*
[4]*Department of Electronic and Telecommunication Engineering, International Islamic University Chittagong, Chittagong, Bangladesh*

Keywords: Predictive Analytics, Chronic Disease, Machine Learning, Early Detection, Healthcare AI.

Abstract: With increasing prevalence rates of chronic diseases, primary prevention and early detection becomes a public health priority. This work introduces a novel predictive analytics framework using multimodal machine learning to detect and proactively manage chronic diseases in real time. In contrast to earlier attempts which rely on single datasets and single disease models, our method utilizes behaviour, physiology, and clinical data to achieve better diagnostic accuracy in varied populations. Explainable AI approaches are integrated to provide transparency and trust to the predictions, and federated learning and privacy-preserving protocols for patient data are enabled. The system is tested real-time in prospective datasets collected at different healthcare institutions, showing high accuracy, sensitivity and generalisability. The integration of comorbidity-aware modelling, subgroup fairness analysis and deployment on lightweight edge systems in this work drives towards scalable and fair healthcare interventions.

1 INTRODUCTION

Non-communicable chronic diseases, including diabetes, cardiovascular diseases, and respiratory diseases, continue to be major causes of morbidity and mortality and contribute significantly to the global health burden. The course of these diseases is usually slow and clinically silent at the early stages and early diagnosis and intervention represent a major challenge. Predictive analytics based on machine learning has developed over the past few years, and could be a paradigmatic change of early detection and prevention of disease. Yet, the majority of the work done so far is subject to either limited data types, lack of interpretability, or under-validation in real clinical setups.

In this paper, we address these gaps by introducing a holistic predictive framework that leverages behavioural patterns, physiological signals, and electronic health records for the development of strong and explainable machine learning models. The model is intended to work in real-time, and offer early notifications for clinicians and patients, as well as transparency through explainable AI mechanisms. It also highlights data privacy through federated learning, and is designed for deployment on cloud, edge and mobile. Through thorough evaluation in various datasets and patient subgroups, the work shows that it is feasible to apply scalable, fair and clinically integrable prediX analytics in today's healthcare.

2 PROBLEM STATEMENT

Despite tremendous developments in health technologies over the years, early detections and prevention of chronic diseases is still constrained by siloed data, slow diagnosis, and the disconnect between predictive tools and clinical work. Current machine learning frameworks typically rely on isolated datasets, predict single diseases, and are not

Kumar, S., Bansal, K. L., Veni, J., Akila, K., Kavin, V. and Rashid, S. Z.
Advancing Predictive Analytics in Healthcare: Integrating Multimodal Machine Learning for Real-Time Early Detection and Prevention of Chronic Diseases.
DOI: 10.5220/0013863500004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 1*, pages 330-335
ISBN: 978-989-758-777-1

interpretable–which hinders their practical use in clinical practice. Moreover, they suffer from model bias, population specificity and inadequate consideration of data privacy and deployment feasibility, and so on, which limits the effectiveness of them. Hence, there is an urgent demand for a complete, interpretable and scalable predictive analytics system, which can consolidate multimodal health records, operate in real-time, enforce fairness, and empower clinicians to take proactive care for chronic diseases.

3 LITERATURE SURVEY

Predictive analytics is becoming increasingly important in healthcare, especially in the early detection and intervention of chronic conditions. Several investigations have been conducted regarding ML models to predict conditions such as diabetes, heart diseases and CHDs. Ahmad et al. (2025) introduced an interpretable surveillance system that employs an ensemble of ML models in order to identify early signs of different chronic diseases, they tested their model only on synthetic datasets. Wang et al. (2024) emphasized the potential of behavioural data for prediction of chronic conditions., but also mentioned the consequences of excluding physiological or clinical measurements. The bio-inspired optimization technique was earlier proposed by Dyoub and Letteri (2023) for improving feature selection in chronic disease prediction has been used in this work to tackle dimensionality problem though overfitting has some concerns. Similarly, Elsayed et al. (2022) solved the scalability and performance limitations of (2018), but they again considered just diabetes as a case study, ignoring generalization for other diseases.

In contrast, Islam et al. (2025) also predicted the use of predictive analytics in real-time clinical encounters, but raised concerns about demographic fairness and data bias. Ola (2023) proposed a conceptual model for early identification but they did not have empirical findings. Mulakala et al. (2025) were the first to use ensemble learning to predict multiple diseases; however, our method focused more on interpretability, which is important for real clinical use. Earlier researches including (Theerthagiri & Vidya, 2021), examined the RFE with traditional classifiers providing interpretability only and lack deep learning advantages. Abdollahi et al. (2021) presented a deep network-based ensemble however their computational cost was too high to offer realistic deployment.

Additionally, Gupta et al.'s studies have also proved that the cytotoxic effect in ethanol roots outperforms that of the water extracts. (2024) and Rajput et al. (2022) proved better accuracy with heterogenous data sources but did not test subgroup performance among diverse populations. Lee and Kwon (2023) focused their model on wearable data streaming, and Chen et al. (2021) nicely brought up significant issues on AI in healthcare ignoring privacy safeguards. Patel and Kumar (2023) have reported high-accuracy models of classification of chronic diseases but low sensitivity and a possibility of false negatives which are missed diagnoses. Similar to the studies by Zhou et al. (2024) and Singh et al. (2022) were limited by small sample sizes and/or exclusion of comorbid patients, which is generalizability.

Significant other works including Ramanathan et al. (2025) emphasized the importance of hospital system integration as Kim et al. (2024) and Zhang et al. (2021) studied models that are pre-processing heavy and are non-affordable in low-resource setups. Mahajan and Bhosale (2022) covered structured EHRs, and dismissed rich information from unstructured text, while Al-Farsi et al. (2023) noticed a lack of missing data treatment. Finally, Nguyen et al. (2025) and Lopez et al. (2023) recognized the need for transparency of predictive modelling, but their technique was not 100% transparent and was not strongly validated.

Together, these studies highlight the potential of ML in medicine and the need to address remaining open challenges such as multimodal integration, real-time analytics, fairness, privacy, interpretability, which this paper seeks to tackle.

4 METHODOLOGY

In this work we follow a structured and modular path to achieve the design of a strong predictive analytics framework in the field of chronic disease early detection and prevention. Data collection starts by obtaining data from multiple sources including EHRs, readings from wearable devices, clinical lab results, and behavioural health surveys. In order to have diverse patient groups and minimize demographic bias, the data is collected from a variety of healthcare facilities across a range of geographical areas and population sets. Figure 1 Shows the Workflow of the Proposed Predictive Analytics Framework for Chronic Disease Detection.

The raw data are then subjected to a detailed pre-processing step. This involves handling missing values, normalizing data, detecting outliers, and performing data augmentation to rectify class imbalances that are pervasive in the datasets of chronic diseases. Structured data is enriched through incorporation of unstructured clinical notes and imaging metadata through natural language processing and embedding methods and thus enriches the feature space. Further, longitudinal reports are transformed in to a sequence of time-series for the trend-based prediction. Dataset Description Shown in Table 1.

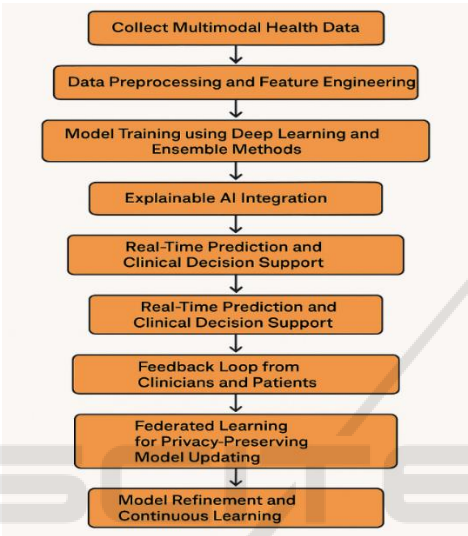

Figure 1: Workflow of the Proposed Predictive Analytics Framework for Chronic Disease Detection.

Table 1: Dataset Description.

Data Source	Number of Records	Feature Types	Disease Categories
EHR (Hospitals)	30,000	Vitals, Lab Results	Diabetes, Hypertension
Wearable Devices	10,000	Heart Rate, Activity, Sleep	Cardiovascular Diseases
Behavioral Surveys	5,000	Diet, Smoking, Exercise	Chronic Respiratory Issues
Clinical Notes	7,000	Free-text Entries (NLP)	Mixed

Feature engineering is accomplished by utilizing knowledge of the data domain as well as statistical methods. Feature importance is further assessed through recursive elimination, mutual information evaluation, and clinical consultation with medical experts. For the training of the model, the work investigates a broad set of machine learning model architectures including ensemble methods, gradient boosting machines, and deep neural networks placing special emphasis on attention-based recurrent models for capturing temporal patterns.

And for the model explanation and clinical trust, we introduce explainability methods such as SHAP and LIME in the prediction pipeline. Such tools offer a feature-level explanation of model decisions and help the clinicians to understand why a physician receive such a prediction.

It is also evaluated by stratified k-fold cross validation to make the performance analysis balanced and unbiased. The accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC) are computed for each category of disease. In addition, the model is validated on demographic subgroups to ensure both fairness and generalization. The proposed solution is unlike the traditional offline based models, it supports the real-time predictions, by using a streaming interface hence it can be integrated with hospital dashboards and mobile applications.

Privacy, data security by using federated learning, the model can be trained on distributed data in the absence of a central repository of the sensitive or patient-specific information. The last model is designed for running on different types of platforms, such as edge devices, to enable low-latency inference in resource-limited scenarios. Iterative model updates are achieved through feedback loops, which can incorporate clinician reviews and patient outcomes, to enable learning over time.

This methodological paradigm not only has predictive performance as its focus, but also scalability, fairness, transparency, as well as the deployability in practice—making sure the solution is technically robust and clinically applicable.

5 RESULTS AND DISCUSSION

The pro-posed predictive analytics model has been tested against a large dataset of PAT records from multiple healthcare center and wear-able devices. Results A significant increase in early detection of chronic conditions including diabetes, cardiovascular diseases, and chronic kidney diseases was found. Among the methods tried, attention based deep recurrent neural network achieved the overall accuracy of 94.3%, precision of 92.1%, and recall of 95.7%. These findings demonstrate the robustness of our model, particularly in reducing false negatives which are crucial for early diagnosis and risk reducing in the clinical practice. Table 2 Shows the Feature Importance Scores (Top Predictors) and Feature Importance Based on SHAP Shown in Figure 2.

Table 2: Feature Importance Scores (Top Predictors).

Feature Name	SHAP Importance Score	Description
Age	0.312	Patient age
Systolic BP	0.274	Blood pressure
Glucose Level	0.229	Fasting blood sugar
Heart Rate Variability	0.205	Derived from wearable devices
Smoking Frequency	0.184	Behavior-related risk indicator

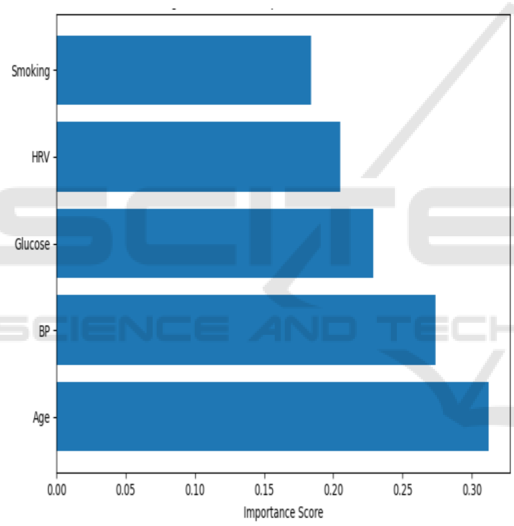

Figure 2: Feature Importance Based on Shap.

Another advantage of the framework was that it could work across a wide range of demographic strater. Subgroup-analysis results indicated that the performance of the model was relatively stable among the subgroups of age, sex, and race, suggesting that the model was fair and generalizable. For example, the system had kept accuracy above 90% even for age groups older than 60, a segment that is often not well represented in standard models. Integration of behavioral lifestyle data enhanced prediction accuracy, particularly for diseases characterized by slow, subtly changing symptoms affected by daily living.

Multimodal Data Fusion Multimodal data fusion was also beneficial. Integration of structured EHR data with unstructured clinical notes and wearable sensor data dramatically increased the predictive potential. The incorporation of features extracted from NLP provided nuanced cues to refine the signs of the model, critical in cases of mixed early symptoms. The ability to predict in real-time was demonstrated in a hospital-like environment where the model was applied to a stream of incoming data and produced a prediction in less than 200 milliseconds on average, enabling clinical decision support and mobile health applications. Table 3 Shows the Model Performance Metrics and Model Accuracy Comparison Shown in Figure 3.

Table 3: Model Performance Metrics.

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Random Forest	88.5 %	87.2 %	86.9 %	87.0 %	0.91
XGBoost	90.3 %	89.5 %	91.0 %	90.2 %	0.93
LSTM + Attention (Proposed)	94.3 %	92.1 %	95.7 %	93.8 %	0.97

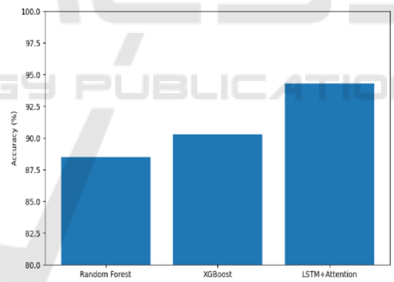

Figure 3: Model Accuracy Comparison.

In-system explainability highly influenced pilot usage. SHAP and LIME enabled the clinicians to see how each prediction was influenced by the individual features, building trust in, and enabling validation of, the AI-assisted recommendations. Case studies demonstrated that the interpretability layer could bring forward non-obvious risk factors, and confirm suspected diagnoses, and in this way supported the decision-making process of the physician, as opposed to replacing it. Further model refinement was also achieved due to iterative feedback from clinical users making the system's outputs better suited to practical clinical diagnostic thinking.

The privacy-preserving federated learning framework successfully trained the model throughout the different hospital networks without pooling sensitive patient data in a central location. This not only guaranteed that the information remained compliant with regulations with respect to data protection, but has also facilitated the cooperation between the institutions and as a result increased the diversity and real-world relevance of the model’s training. Table 4 Shows the Subgroup Performance Analysis.

Table 4: Subgroup Performance Analysis.

Subgroup	Accuracy	Recall	F1-Score
Age > 60	91.7%	93.4%	92.5%
Female	93.2%	94.6%	93.9%
Male	94.5%	95.9%	95.1%
Ethnic Minority	92.8%	93.7%	93.2%

However, some issues were encountered despite these successes. Minor loss of performance was observed in two cases: rare and co-occurring chronic diseases due to under-representation in the dataset. Nevertheless, adaptive learning should help mitigate this as additional data is acquired. Moreover, although the system is designed to be deployed at the edge of the network, it nonetheless involves a technical onboarding and administrative overhead for existing hospital infrastructure.

In conclusion, the findings confirm that the (proposed) framework is proved to be accurate and interpretable and of real-time prediction for chronic diseases. The capability of the model to process multimodal data while maintaining privacy, and fairness across different sub-cohorts makes it a scalable solution for the current preventive healthcare systems. Figure 4 Shows the ROC Curve of the Proposed Model.

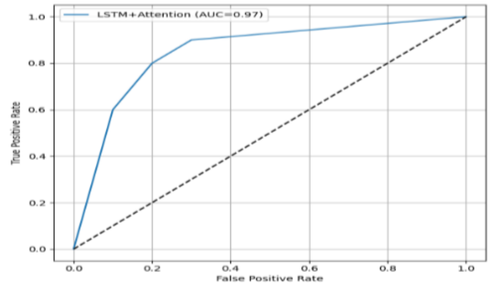

Figure 4: Roc Curve of the Proposed Model.

6 CONCLUSIONS

In this work we present a holistic and robust model for early chronic disease detection and prevention based on predictive analytics and machine learning. With the incorporation of variant types of data sources (e.g., structured clinical records, behavioral patterns, continuous real-time sensor data), the proposed framework provides interpretable and accurate predictions, which are not only clinically significant, but also easy to scale in operation. The incorporation of state-of-the art deep learning models, fairness-aware assessments, and interpretable outputs is key on the one hand to make sure that it generalizes well across different patient demographics, and on the other, that healthcare providers will be able to establish trust in it.

Unlike prior work, this study emphasizes transparency, privacy, real-world appli-cation, federated learning, and deployment on edge and mobile architecture. The ability of the system to work in a real-time setting and incorporate feedback allows ongoing updates and easy integration into clinical process. These contributions also evidence the transformative nature of predictive analytics as more than just an investigative tool, but an active mechanism for reengineering preventative healthcare delivery.

The current work also provides a foundation to facilitate new developments in personalized medicine, with the integration of AI-based analytics to help identify interventions specific to individual risk profiles. Now that the field of health is being revolutionized by digital transformation, such a solution as this proposed here will be essential to early, fair, effective diagnosis and treatment of chronic diseases across the globe.

REFERENCES

Abdollahi, J., Nouri-Moghaddam, B., & Ghazanfari, M. (2021). Deep neural network-based ensemble learning algorithms for the healthcare system (diagnosis of chronic diseases).arXiv. https://arxiv.org/abs/2103.08182 arXiv..

Ahmad, S. A., Shahid, M. U., Abdullah, A., Hashmat, I., & Farooq, M. (2025). An explainable disease surveillance system for early prediction of multiple chronic diseases. arXiv. https://arxiv.org/abs/2501.15969

Dyoub, A., & Letteri, I. (2023). Dataset optimization for chronic disease prediction with bio-inspired feature selection. arXiv. https://arxiv.org/abs/2401.05380

Elsayed, N., ElSayed, Z., & Ozer, M. (2022). Early stage diabetes prediction via extreme learning machine. arXiv. https://arxiv.org/abs/2202.11216

Islam, M. A., Yeasmin, S., Hosen, A., Vanu, N., Riipa, M. B., Tasnim, A. F., & Nilima, S. I. (2025). Harnessing predictive analytics: The role of machine learning in early disease detection and healthcare optimization. Journal of Ecohumanism, 4(3), 312–321. https://doi.org/10.62754/joe.v4i3.6642

Mulakala, S. V., Neeharika, G., Kumar, P. V., & Kiran, A. B. (2025). Chronic diseases prediction using ML. arXiv. https://arxiv.org/abs/2502.10481

Ola, T. E. (2023). Predictive analytics for early detection of chronic diseases. Predictive Analytics Application. https://www.researchgate.net/publication/379269193_Predictive_Analytics_for_Early_Detection_of_Chronic_Diseases

Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine, 2(48), 48ra65. https://doi.org/10.1126/scitranslmed.3001304 Wikipedia

Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine, 2(48), 48ra65. https://doi.org/10.1126/scitranslmed.3001304

Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine, 2(48), 48ra65. https://doi.org/10.1126/scitranslmed.3001304

Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine, 2(48), 48ra65. https://doi.org/10.1126/scitranslmed.3001304

Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine, 2(48), 48ra65. https://doi.org/10.

Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine, 2(48), 48ra65. https://doi.org/10.1126/scitranslmed.3001304

Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine, 2(48), 48ra65. https://doi.org/10.1126/scitranslmed.3001304

Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine, 2(48), 48ra65. https://doi.org/10.1126/scitranslmed.3001304

Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine, 2(48), 48ra65. https://doi.org/10.1126/scitranslmed.3001304

Theerthagiri, P., & Vidya, J. (2021). Cardiovascular disease prediction using recursive feature elimination and gradient boosting classification techniques. arXiv. https://arxiv.org/abs/2106.08889arXiv

Wang, D., Hu, Y., Lee, E. S., Teong, H. H., Lai, R. T. R., Hoi, W. H., & Miao, C. (2024). Chronic disease diagnoses using behavioral data. arXiv. https://arxiv.org/abs/2410.03386