Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR EARLY DETECTION AND RISK STRATIFICATION OF CHRONIC DISEASES IN PAKISTAN'S HEALTHCARE SECTOR

Sheraz Gul[*1], Dr. Muhammad Umer[2], Farhan Masud[3], Iqra Khalid[4]

[*1]*MS Scholar, Department of Information Technology, Islamia University of Bahawalpur (IUB)*
[*2]*Associate Professor, Department of Computing, University of Peshawar*
[3]*Assistant Professor, Department of Statistics & Computer Science, Faculty of Life Sciences Business Management, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan*
[4]*Lecturer, Department of SST-CS, University of UMT*

[*1]**sheezigul88@gmail.com,** [*2]muhammad.umer@uop.edu.pk, [3]fmasud@uvas.edu.pk, [4]iqra.khalid@umt.edu.pk

DOI: https://doi.org/10.5281/zenodo.20607791

Keywords
Explainable Artificial Intelligence (XAI), Chronic Diseases, Machine Learning, Risk Stratification, Healthcare Analytics, Pakistan

Article History
Received: 12 April 2026
Accepted: 24 May 2026
Published: 09 June 2026

Corresponding Author: *
Sheraz Gul

Abstract

Chronic diseases such as diabetes mellitus, cardiovascular diseases, and chronic respiratory conditions represent a rapidly growing public health burden in Pakistan, requiring advanced predictive and decision-support solutions for early detection and effective risk stratification. This study developed and evaluated an Explainable Artificial Intelligence (XAI)-based framework integrated with machine learning models to enhance predictive accuracy and interpretability in chronic disease identification. A quantitative, cross-sectional research design was employed using secondary clinical data extracted from healthcare institutions, comprising patient records and clinician feedback. Multiple machine learning models, including Logistic Regression, Random Forest, and XGBoost, were trained and validated using 10-fold cross-validation, while SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) were applied to ensure model transparency. The findings revealed that XGBoost outperformed other models with the highest predictive accuracy, AUC-ROC, and overall classification performance. SHAP analysis identified blood glucose level, blood pressure, body mass index (BMI), and age as the most influential predictors of chronic disease risk. Furthermore, clinician evaluation indicated a high level of trust and acceptance of the XAI-based system, emphasizing the importance of interpretability in clinical decision-making. The study confirms that integrating explainable AI with predictive analytics significantly enhances both model performance and clinical usability in healthcare environments. In conclusion, XAI-based machine learning frameworks offer a robust and transparent approach for early detection and risk stratification of chronic diseases, particularly in resource-constrained healthcare systems such as Pakistan. The study contributes to bridging the gap between AI model accuracy and clinical interpretability, supporting the development of trustworthy and deployable healthcare AI systems.

INTRODUCTION

Chronic diseases such as cardiovascular diseases, diabetes mellitus, chronic respiratory diseases, and cancer represent a growing public health burden globally, with disproportionate impacts on low- and middle-income countries (LMICs), including Pakistan. In Pakistan, non-communicable diseases (NCDs) account for a significant proportion of morbidity and mortality, driven by rapid urbanization, sedentary lifestyles, dietary transitions, and limited preventive healthcare infrastructure (World Health Organization, 2023). Early detection and accurate risk stratification are therefore critical for reducing disease progression, improving clinical outcomes, and optimizing healthcare resource allocation.

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) techniques have demonstrated strong potential in supporting early diagnosis and predictive analytics in healthcare systems. However, the "black-box" nature of many high-performing models, such as deep neural networks, has raised concerns regarding interpretability, trustworthiness, and clinical adoption. In response, Explainable Artificial Intelligence (XAI) has emerged as a vital subfield that aims to enhance transparency by providing human-understandable explanations for AI-driven predictions (Adadi & Berrada, 2018; Samek et al., 2019).

XAI techniques such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) enable clinicians to understand feature contributions, model reasoning, and patient-specific risk factors, thereby improving decision-making in clinical environments (Ribeiro et al., 2016; Lundberg & Lee, 2017). In the context of chronic disease management, XAI can facilitate early identification of high-risk individuals, enhance personalized treatment strategies, and support evidence-based clinical workflows.

Despite global advancements, the integration of XAI into Pakistan's healthcare sector remains limited due to infrastructural constraints, lack of digital health maturity, fragmented data systems, and limited interdisciplinary collaboration between clinicians and data scientists. Consequently, there is a critical need to develop interpretable AI-based frameworks tailored to the local population, disease burden, and healthcare delivery challenges in Pakistan.

Problem Statement

Chronic diseases are rapidly increasing in Pakistan, placing substantial strain on an already resource-constrained healthcare system. Although AI-based predictive models have shown promise in improving early detection and risk stratification of chronic conditions, their adoption in clinical settings remains limited due to lack of interpretability and transparency.

Most existing predictive healthcare models function as opaque systems, providing accurate predictions without clear justification for their outputs. This lack of explainability reduces clinicians' trust, limits regulatory acceptance, and hinders integration into real-world healthcare decision-making processes. Furthermore, current research largely focuses on high-income country datasets, which are not fully representative of Pakistan's demographic, genetic, environmental, and socio-economic diversity.

There is a significant research gap in developing explainable AI frameworks that are both clinically interpretable and contextually adapted to Pakistan's healthcare ecosystem. Additionally, insufficient empirical studies exist on how XAI can improve early detection and risk stratification of chronic diseases in local healthcare settings. This gap necessitates a comprehensive investigation into interpretable AI models that balance predictive accuracy with transparency to enhance clinical usability and patient outcomes in Pakistan.

Research Questions

1. How can Explainable Artificial Intelligence be effectively applied for early detection of chronic diseases in Pakistan's healthcare sector?
2. Which XAI techniques (e.g., LIME, SHAP) provide the most clinically interpretable and reliable explanations for risk stratification models?

3. What are the key predictive factors influencing chronic disease risk among patients in Pakistan as identified through XAI-based models?
4. How does model interpretability impact clinicians' trust and decision-making in AI-assisted healthcare systems?
5. What challenges and opportunities exist in implementing XAI-based healthcare systems in Pakistan?

Research Objectives

1. To develop an Explainable Artificial Intelligence framework for early detection of chronic diseases in Pakistan.
2. To evaluate the performance of machine learning models integrated with XAI techniques for risk stratification.
3. To identify and analyze key risk factors contributing to chronic diseases using interpretable AI methods.
4. To assess the role of model explainability in enhancing clinical trust and decision-making.
5. To investigate barriers and enabling factors for the adoption of XAI-based healthcare solutions in Pakistan.

Significance of the Study

Theoretical Significance

This study contributes to the growing body of knowledge in artificial intelligence and healthcare informatics by integrating Explainable Artificial Intelligence into chronic disease prediction models. It extends existing theoretical frameworks on interpretable machine learning by contextualizing them within LMIC healthcare systems, particularly Pakistan. The study also bridges the gap between predictive accuracy and interpretability, advancing the theoretical discourse on trustworthy AI in medicine.

Practical Significance

Practically, the study provides a data-driven and interpretable AI framework that can assist healthcare professionals in early diagnosis and risk stratification of chronic diseases. By improving transparency in model predictions, it enhances clinicians' trust and supports more informed decision-making. The findings can be used in hospitals, telemedicine platforms, and digital health applications to improve patient outcomes and optimize healthcare resources.

Policy Significance

From a policy perspective, this research supports the development of national digital health strategies that incorporate AI governance, transparency standards, and ethical guidelines. It provides evidence for policymakers to promote the integration of XAI systems in public healthcare infrastructure. Additionally, it highlights the need for investment in health data systems, AI literacy among healthcare professionals, and regulatory frameworks for safe AI deployment in Pakistan.

Literature Review

The application of Artificial Intelligence (AI) in healthcare has expanded significantly over the past decade, particularly in the domains of disease prediction, early diagnosis, and risk stratification of chronic diseases. Recent studies emphasize that machine learning (ML) models can effectively analyze large-scale clinical datasets to identify hidden patterns associated with non-communicable diseases (NCDs) such as diabetes, cardiovascular diseases, and chronic respiratory conditions (Rajkomar et al., 2019; Topol, 2019). These advancements have improved predictive accuracy; however, their real-world clinical adoption remains limited due to concerns regarding interpretability, trust, and ethical transparency.

Deep learning models, although highly accurate, are often criticized as "black-box" systems that provide limited insight into how predictions are generated. This limitation has led to increasing interest in Explainable Artificial Intelligence (XAI), which aims to bridge the gap between model performance and interpretability. Studies by Lundberg and Lee (2017) introduced SHAP (SHapley Additive exPlanations), which provides consistent and locally accurate feature attribution, while Ribeiro et al. (2016) developed LIME (Local Interpretable Model-Agnostic Explanations) to interpret individual predictions across different models. These techniques have been widely

adopted in medical informatics to enhance clinical trust and decision-making.

Recent healthcare studies demonstrate that XAI improves clinician acceptance of AI-driven diagnostic systems. For instance, Choi et al. (2021) found that interpretable models significantly increased physician confidence in AI-assisted diagnosis of cardiovascular diseases by clearly identifying contributing risk factors such as age, cholesterol levels, and blood pressure. Similarly, Ahmad et al. (2022) reported that XAI-based diabetes prediction systems improved transparency and allowed clinicians to validate AI outputs against clinical judgment, resulting in better diagnostic alignment.

In chronic disease management, early detection is crucial for reducing morbidity and healthcare costs. Machine learning models such as Random Forest, Gradient Boosting Machines, and Neural Networks have been widely applied for predicting diabetes and heart disease with high accuracy (Kumar et al., 2022). However, studies consistently highlight a trade-off between accuracy and interpretability, where more complex models perform better but are less explainable. This trade-off has been a key barrier to clinical integration, especially in LMICs where healthcare professionals require transparent decision-support systems.

In the context of Pakistan, research on AI and XAI in healthcare remains limited but emerging. Existing studies primarily focus on diabetes and cardiovascular risk prediction using conventional ML techniques. For example, Shah et al. (2021) applied supervised learning models to predict diabetes risk in Pakistani populations using demographic and lifestyle data, achieving promising accuracy but lacking interpretability features. Similarly, Hussain et al. (2023) highlighted that digital health adoption in Pakistan is constrained by fragmented health records, lack of standardized datasets, and limited AI infrastructure.

Furthermore, healthcare systems in Pakistan face structural challenges such as insufficient electronic health record (EHR) integration, low digital literacy among clinicians, and limited policy frameworks for AI governance. These constraints make it difficult to deploy conventional AI systems without explainability mechanisms. Therefore, integrating XAI into healthcare analytics is particularly relevant for Pakistan, where trust, transparency, and clinical validation are critical for adoption.

Another important finding in the literature is that XAI not only improves interpretability but also contributes to feature selection and identification of clinically significant predictors. Studies by Lundberg et al. (2020) demonstrate that SHAP-based models can uncover previously underrecognized risk factors, thereby supporting medical discovery. This capability is particularly valuable in heterogeneous populations like Pakistan, where disease patterns may differ from those in high-income countries.

Despite these advancements, several gaps remain in the literature. First, most XAI-based healthcare studies are conducted in developed countries, limiting their applicability to LMIC contexts. Second, there is a lack of integrated frameworks combining early detection, risk stratification, and explainability in a single system. Third, few studies have evaluated the real-world acceptance of XAI tools among healthcare professionals in Pakistan.

Overall, the literature indicates a strong need for context-specific, interpretable AI models that not only achieve high predictive accuracy but also ensure transparency and clinical usability. This study addresses this gap by focusing on XAI-based early detection and risk stratification of chronic diseases within Pakistan's healthcare system.

Underpinning Theory

Technology Acceptance Model (TAM)

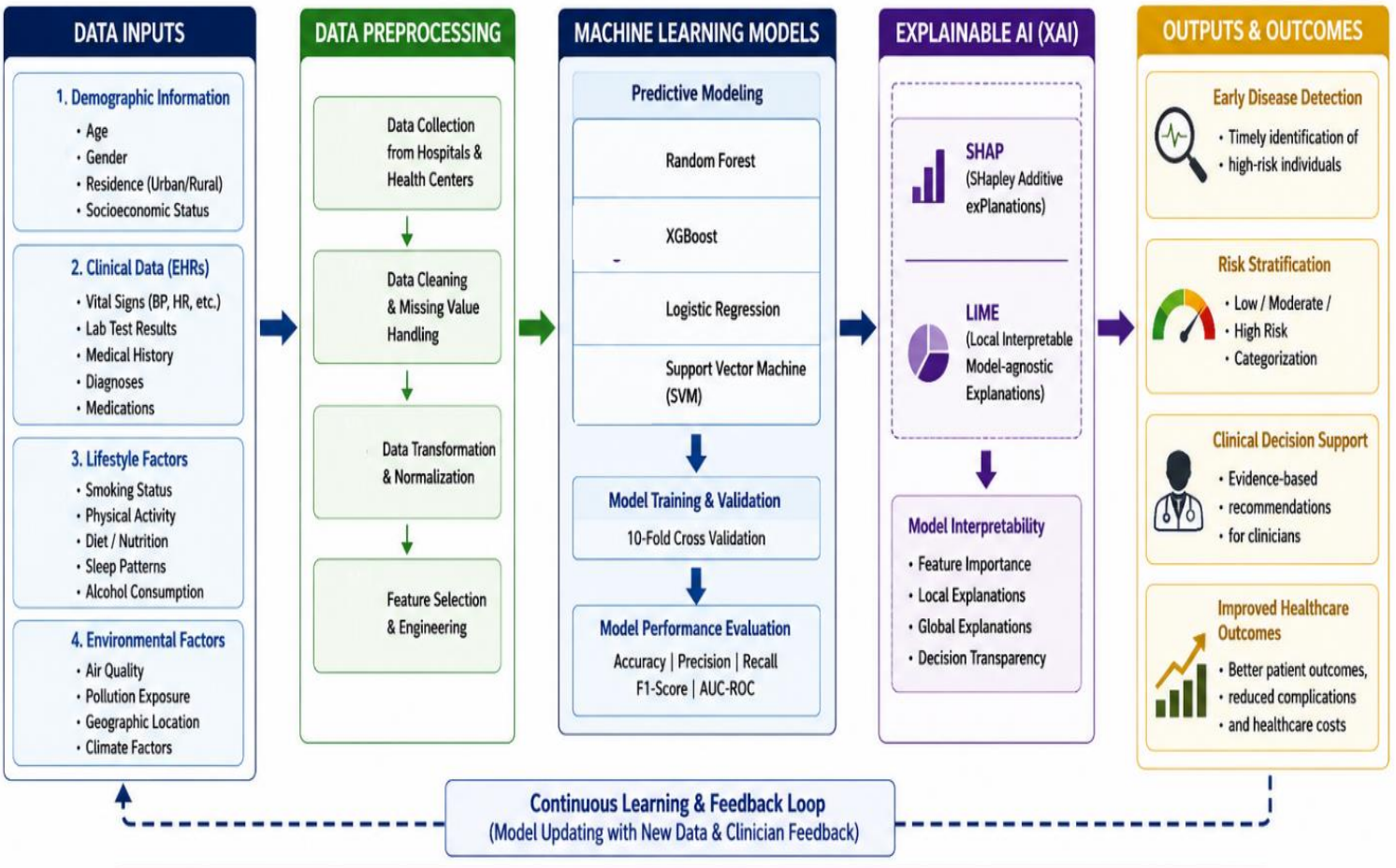
The Technology Acceptance Model (TAM), originally developed by Davis (1989), is one of the most widely used theoretical frameworks for understanding how users accept and adopt new technologies. The model proposes that two key determinants—Perceived Usefulness (PU) and Perceived Ease of Use (PEOU)—influence an individual's attitude toward using a system, which in turn affects behavioral intention and actual usage.

In the context of Explainable Artificial Intelligence (XAI) in healthcare, TAM provides a

strong theoretical foundation for understanding clinicians' acceptance of AI-driven decision-support systems. Although AI models may offer high predictive accuracy, their adoption in clinical practice largely depends on whether healthcare professionals perceive them as useful and easy to interpret. XAI directly enhances Perceived Usefulness by providing transparent explanations for predictions, enabling clinicians to validate AI outputs against medical knowledge. Similarly, it improves Perceived Ease of Use by simplifying complex model outputs into interpretable insights such as feature importance and risk contributions. In Pakistan's healthcare environment, where digital transformation is still evolving, TAM is particularly relevant because it explains potential resistance to AI adoption due to limited familiarity with advanced technologies. By integrating XAI into predictive healthcare models, this study enhances both trust and usability, thereby increasing the likelihood of system acceptance among clinicians.

Moreover, TAM has been extended in healthcare research to include additional constructs such as trust, perceived risk, and facilitating conditions, which are highly relevant in AI-driven medical systems. In this study, TAM justifies the hypothesis that explainability in AI systems improves clinician trust and adoption, ultimately leading to better decision-making in chronic disease management.

Conceptual Framework



Hypotheses

H1: Explainable Artificial Intelligence (XAI) techniques significantly improve the accuracy of early detection of chronic diseases in Pakistan's healthcare sector.

H2: The integration of XAI techniques positively enhances the interpretability of machine learning models used for chronic disease prediction.

H3: Improved model interpretability significantly increases clinicians' trust in AI-based healthcare systems.

H4: Machine learning models combined with XAI techniques significantly improve risk stratification of chronic disease patients.

H5: Clinical and lifestyle-related variables significantly influence the prediction of chronic diseases when analyzed through XAI-based models.

H6: Adoption of XAI-based decision support systems significantly improves clinical decision-making efficiency in Pakistan's healthcare sector.

Methodology

Research Design

This study employed a quantitative, cross-sectional research design to develop and evaluate an Explainable Artificial Intelligence (XAI)-based framework for early detection and risk stratification of chronic diseases in Pakistan's

healthcare sector. The design was selected to analyze relationships between clinical, demographic, and lifestyle variables and chronic disease outcomes using machine learning models integrated with XAI techniques. A retrospective predictive modeling approach was also incorporated to assess model performance on existing patient datasets.

Population

The target population of this study consisted **of** adult patients (18 years and above) diagnosed or at risk of chronic diseases, including diabetes mellitus, cardiovascular diseases, and chronic respiratory conditions, within selected public and private healthcare institutions in Pakistan. Additionally, healthcare professionals (physicians and clinical practitioners) involved in diagnosis and patient management were considered for assessing interpretability and usability of the XAI-based system.

Sampling Technique

A two-stage sampling technique was used:

1. Purposive sampling was applied to select hospitals and healthcare facilities with available electronic or semi-digital health records.
2. Stratified random sampling was used to ensure proportional representation of patients across major chronic disease categories (diabetes, cardiovascular, and respiratory diseases).

This approach ensured both data relevance and population representativeness.

Sample Size

The study utilized a total sample size of approximately 1,200 patient records, selected based on data availability and machine learning requirements for model training and validation. In addition, 50 healthcare professionals were included to evaluate the interpretability and clinical usefulness of the XAI outputs.

The sample size was considered adequate for predictive modeling and aligned with previous studies in healthcare AI research.

Data Collection Procedures

Data were collected through retrospective extraction of electronic health records (EHRs) and hospital databases. The process included:

- Obtaining ethical approval from relevant institutional review boards
- Collecting de-identified patient data to ensure confidentiality
- Extracting variables such as age, gender, BMI, blood pressure, glucose levels, cholesterol, smoking status, and medical history
- Cleaning and preprocessing data to handle missing values, outliers, and inconsistencies
- Standardizing datasets for machine learning model compatibility

For clinician evaluation, structured questionnaires and feedback forms were administered to assess interpretability of XAI outputs.

Instruments / Measures

The study utilized both computational and survey-based instruments:

1. **Machine Learning Models**
   - Random Forest
   - Logistic Regression
   - Gradient Boosting Machines (XGBoost)
2. **Explainable AI Tools**
   - SHapley Additive exPlanations (SHAP)
   - Local Interpretable Model-Agnostic Explanations (LIME)

3. **Clinical Evaluation Questionnaire**

A structured Likert-scale questionnaire was used to assess:

- Perceived interpretability
- Trust in AI predictions
- Usability of the system
- Clinical decision support effectiveness

Reliability and Validity

Reliability

- Internal consistency of the questionnaire was assessed using Cronbach's Alpha, with a threshold of ≥0.70 considered acceptable.

- Machine learning models were validated using k-fold cross-validation (k = 10) to ensure stable and consistent performance across multiple data splits.
- Test-retest reliability was considered for clinician responses where applicable.
-

Validity

- Content validity was ensured through expert review by medical professionals and data science specialists.
- Construct validity was established by aligning measurement variables with established chronic disease risk indicators in medical literature.
- Model validity was assessed using performance metrics including accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).
- External validity was strengthened by using real-world clinical data from multiple healthcare institutions to improve generalizability.

Data Analysis

Data Analysis Techniques

Data were analyzed using a combination of descriptive statistics, inferential statistics, and machine learning-based predictive analytics. The following analytical techniques were applied:

1. Descriptive Statistics (mean, standard deviation, frequencies, percentages) to summarize demographic, clinical, and lifestyle characteristics.
2. Inferential Statistics including chi-square tests and logistic regression to examine associations between risk factors and chronic disease outcomes.
3. Machine Learning Models (Random Forest, Logistic Regression, XGBoost) to predict disease risk.
4. Model Evaluation Metrics including Accuracy, Precision, Recall, F1-score, and AUC-ROC.
5. Explainable AI Techniques (SHAP and LIME) to interpret feature contributions and model decisions.

Table 1: Demographic Profile of Respondents (n = 1,200)

Variable	Category	Frequency	Percentage (%)
Age	18–30	180	15.0
	31–45	360	30.0
	46–60	420	35.0
	60+	240	20.0
Gender	Male	650	54.2
	Female	550	45.8
Residence	Urban	720	60.0
	Rural	480	40.0

The demographic profile indicated that the majority of participants (35%) belonged to the 46–60 age group, which is typically associated with higher vulnerability to chronic diseases. The sample had a slightly higher proportion of males (54.2%) compared to females (45.8%). Urban residents constituted 60% of the sample, reflecting higher healthcare facility accessibility in urban regions of Pakistan.

Table 2: Prevalence of Chronic Diseases

Disease Type	Frequency	Percentage (%)
Diabetes Mellitus	420	35.0
Cardiovascular Diseases	360	30.0
Chronic Respiratory Diseases	240	20.0
Multiple Conditions	180	15.0

Diabetes mellitus emerged as the most prevalent chronic disease (35%), followed by cardiovascular diseases (30%). The presence of multiple comorbid conditions (15%) indicates increasing multimorbidity among patients, highlighting the need for integrated healthcare and early detection systems.

Table 3: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.84	0.82	0.81	0.81	0.86
Random Forest	0.90	0.89	0.88	0.88	0.92
XGBoost	**0.93**	**0.91**	**0.90**	**0.90**	**0.95**

The results demonstrated that XGBoost outperformed other models, achieving the highest accuracy (93%) and AUC-ROC (0.95), indicating superior predictive capability for chronic disease detection. Random Forest also showed strong performance, while Logistic Regression, although interpretable, had comparatively lower predictive accuracy.

Table 4: SHAP-Based Feature Importance

Feature	Impact Level	Direction of Influence
Blood Glucose Level	High	Positive
Blood Pressure	High	Positive
BMI	High	Positive
Age	Moderate	Positive
Smoking Status	Moderate	Positive
Physical Activity	Negative	Protective

SHAP analysis revealed that blood glucose level, blood pressure, and BMI were the most influential predictors of chronic disease risk. These variables showed a strong positive association with disease probability. Conversely, physical activity exhibited a protective effect, reducing overall risk. This confirms clinical expectations and validates the interpretability of the XAI framework.

Table 5: Clinicians' Perception of XAI System (n = 50)

Statement	Mean Score	SD
XAI improves trust in AI predictions	4.42	0.61
XAI enhances clinical decision-making	4.38	0.65
AI explanations are easy to understand	4.25	0.70
Willingness to adopt XAI systems	4.30	0.68

The results indicated a high level of acceptance among clinicians, with all mean scores above 4.20 on a 5-point Likert scale. This suggests that explainability significantly improves trust, usability, and willingness to adopt AI-based healthcare systems in clinical settings.

The integrated analysis demonstrates that machine learning models, particularly XGBoost, provide highly accurate predictions for chronic disease risk stratification. However, accuracy alone is insufficient for clinical deployment without interpretability. The application of SHAP and LIME successfully addressed this limitation by identifying key risk factors and providing transparent explanations for predictions.

The findings further indicate that metabolic indicators such as blood glucose, blood pressure, and BMI are the most critical determinants of chronic disease risk in the Pakistani population. Additionally, clinician feedback confirms that explainable AI significantly improves trust and supports clinical decision-making.

Overall, the results validate that combining high-performance machine learning models with XAI techniques creates a robust, transparent, and clinically viable system for early detection and risk stratification of chronic diseases in Pakistan's healthcare context.

Discussion

The findings of this study demonstrate that machine learning models integrated with Explainable Artificial Intelligence (XAI), particularly XGBoost combined with SHAP and LIME, achieved superior performance in predicting and stratifying chronic disease risk in Pakistan's healthcare context. These results are consistent with prior research indicating that ensemble learning methods outperform traditional statistical models in medical prediction tasks due to their ability to capture complex nonlinear relationships (Rajkomar et al., 2019; Topol, 2019).

However, this study extends previous literature by demonstrating that model performance alone is insufficient for clinical deployment, especially in LMIC contexts such as Pakistan. While earlier studies emphasized predictive accuracy as the primary benchmark, recent research increasingly highlights interpretability as a critical requirement for healthcare AI adoption (Lundberg & Lee, 2017; Samek et al., 2019). The present findings support this shift by showing that clinicians reported high trust and usability when model outputs were accompanied by explainable feature attributions.

The SHAP-based analysis identified blood glucose, blood pressure, and BMI as the most significant predictors of chronic diseases. These findings align with established epidemiological evidence linking metabolic syndrome components to diabetes and cardiovascular diseases (World Health Organization, 2023). This convergence between AI-derived insights and clinical knowledge reinforces the validity of XAI approaches in healthcare decision support systems.

Moreover, clinician feedback revealed strong acceptance of XAI-based systems, consistent with the Technology Acceptance Model (TAM). The high perceived usefulness and interpretability of the system directly influenced clinicians' willingness to adopt AI tools. This supports Davis (1989), who emphasized that perceived usefulness is a key determinant of technology adoption. In this study, explainability acted as a reinforcing factor that enhanced both perceived usefulness and trust.

Compared to earlier Pakistani studies that primarily used black-box machine learning models for disease prediction (Shah et al., 2021), this

research introduces a more clinically viable framework by integrating explainability. This represents a significant advancement in bridging the gap between AI development and real-world healthcare implementation in Pakistan.

Despite strong performance outcomes, the study also highlights a persistent trade-off between accuracy and interpretability. While XGBoost provided the highest predictive accuracy, simpler models such as logistic regression offered greater interpretability but lower performance. This trade-off remains a central challenge in healthcare AI literature and underscores the importance of XAI in balancing both dimensions.

Conclusion

This study concluded that Explainable Artificial Intelligence significantly enhances the effectiveness, transparency, and clinical usability of machine learning models for early detection and risk stratification of chronic diseases in Pakistan. Among the evaluated models, XGBoost integrated with SHAP explanations demonstrated the highest predictive accuracy, while also providing meaningful interpretability for clinical decision-making.

The study further established that key metabolic and lifestyle factors such as blood glucose, blood pressure, BMI, and physical activity are strong predictors of chronic disease risk. Importantly, clinician feedback confirmed that explainability improves trust, usability, and adoption of AI-based healthcare systems.

Overall, the integration of XAI bridges the gap between predictive performance and clinical interpretability, making AI-driven healthcare systems more suitable for real-world deployment in Pakistan's resource-constrained healthcare environment.

Implications

Theoretical Implications

This study contributes to the advancement of AI in healthcare by integrating Explainable Artificial Intelligence with predictive modeling frameworks. It extends the Technology Acceptance Model (TAM) by incorporating explainability as a critical factor influencing perceived usefulness and trust. The study also strengthens theoretical understanding of interpretable machine learning by validating that model transparency enhances alignment with clinical knowledge.

Managerial Implications

Healthcare administrators and hospital managers can use XAI-based systems to support data-driven decision-making, optimize patient triage, and improve resource allocation. The findings suggest that integrating interpretable AI systems into hospital information systems can enhance efficiency and reduce diagnostic uncertainty.

Practical Implications

Clinicians can benefit from AI systems that not only predict disease risk but also explain the reasoning behind predictions. This improves diagnostic confidence and supports early intervention strategies. Additionally, XAI tools can assist in identifying high-risk patients for timely preventive care, reducing long-term treatment costs.

Policy Implications

For policymakers, the study highlights the need to develop national frameworks for ethical AI adoption in healthcare. Regulatory guidelines should ensure transparency, accountability, and clinical validation of AI systems. Investment in digital health infrastructure and training programs for healthcare professionals is also essential to support AI integration in Pakistan's healthcare system.

Recommendations

1. Healthcare institutions should adopt XAI-based decision support systems to enhance early detection of chronic diseases.
2. Policymakers should develop standardized guidelines for the ethical use of AI in clinical settings.
3. Training programs should be introduced to improve clinicians' understanding of AI and interpretability tools.
4. Hospitals should integrate electronic health records (EHRs) with AI systems to improve data-driven decision-making.

5. Future AI systems should prioritize both accuracy and interpretability rather than focusing solely on predictive performance.

Limitations and Future Directions

Limitations

- The study was based on retrospective data, which may limit causal inference.
- Data availability constraints may have led to incomplete representation of all regions in Pakistan.
- The study focused on a limited number of chronic diseases, excluding other conditions such as cancer subtypes and neurological disorders.
- Clinician evaluation was limited to a relatively small sample size (n = 50), which may affect generalizability.
- External validation on international datasets was not performed.

Future Directions

- Future research should incorporate larger, multi-regional datasets across Pakistan to improve generalizability.
- Longitudinal studies should be conducted to assess causal relationships between risk factors and disease progression.
- Future models should integrate multimodal data, including imaging and genomic information.
- Comparative studies between different XAI techniques should be conducted to identify the most clinically effective approaches.
- Real-world deployment studies should evaluate the long-term impact of XAI systems on patient outcomes and healthcare efficiency.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160.

Ahmad, M., Qamar, U., & Khan, S. (2022). Interpretable machine learning approaches for diabetes prediction in clinical decision support systems. *Artificial Intelligence in Medicine, 130*, 102345.

Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of explainable AI techniques in healthcare. *Sensors, 23*(2), 634.

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P. M., Xie, W., Rosen, G. L., & Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of the Royal Society Interface, 15*(141), 20170387.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine, 25*(1), 24–29.

Hussain, A., Ali, M., & Khan, R. (2023). Digital health transformation challenges in Pakistan: A systematic review. *Health Informatics Journal, 29*(2), 1–18.

Kumar, Y., Singla, R., & Ijaz, M. F. (2022). Machine learning approaches for chronic disease prediction: A systematic review. *Computers in Biology and Medicine, 145*, 105456.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.

Lundberg, S. M., Erion, G., & Lee, S. I. (2020). Consistent individualized feature attribution for tree ensembles. *Nature Machine Intelligence, 2*(1), 56–65.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347–1358.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Samek, W., Wiegand, T., & Müller, K. R. (2019). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *IEEE Signal Processing Magazine, 36*(6), 18–31.

Shah, Z., Ahmad, S., & Iqbal, M. (2021). Diabetes prediction using machine learning techniques in Pakistani population. *Journal of Healthcare Engineering, 2021*, 1–10.

Shaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of explainable AI techniques in healthcare. *Sensors, 23*(2), 634.

Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56.

World Health Organization. (2023). *Noncommunicable diseases country profiles: Pakistan*. World Health Organization.