ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23

American Academic publishers, volume 05, issue 11,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

DEVELOPING AN AI-DRIVEN PREDICTIVE MONITORING SYSTEM FOR DIABETES AND HYPERTENSION IN UZBEKISTAN: DATA COLLECTION CHALLENGES AND PRACTICAL SOLUTIONS

Khusanov K.Kh.

Samarkand branch of Tashkent University of Information Technologies named after Muhammad al-Khwarizmi

Email: xusanovkamol01@gmail.com

Abstract: Artificial intelligence (AI) and Big Data technologies are transforming healthcare by enabling early detection and risk prediction for chronic diseases. However, their effectiveness depends on the availability of complete, high-quality, and standardized medical data. This study investigates the feasibility of establishing an AI-driven predictive monitoring system for diabetes and hypertension in Uzbekistan, focusing on barriers in clinical data collection and opportunities for improvement. Using a mixed-methods approach, electronic health records (EHRs), laboratory information systems (LIS), pharmacy data, dispensary registries, and wearable device outputs were analyzed from a pilot cohort of 112,846 patients. Results demonstrate that predictive AI models can achieve clinically meaningful accuracy, but fragmented data flows, inconsistent standardization, and incomplete laboratory records significantly limit model performance. The paper proposes a set of technical, organizational, and regulatory solutions—including federated data architecture, national interoperability standards, and strengthened data governance—to support the successful deployment of large-scale AI systems in Uzbekistan's healthcare system.

Keywords: Artificial Intelligence (AI), Big Data in Healthcare, Predictive Monitoring, Diabetes Mellitus Type 2, Arterial Hypertension, Health Data Integration, Federated Data Architecture, Electronic Health Records (EHR), Data Quality and Standardization

1. Introduction

Chronic non-communicable diseases (NCDs), particularly type 2 diabetes mellitus (T2DM) and arterial hypertension, remain leading contributors to morbidity and mortality in Uzbekistan. International epidemiological reports indicate a high prevalence of hypertension and metabolic disorders in the country, with cardiovascular mortality accounting for a substantial proportion of premature deaths [1,2]. These trends underscore the need for more effective monitoring approaches that can identify high-risk individuals, optimize resource allocation, and improve long-term clinical outcomes.

Globally, AI-powered predictive monitoring systems—exemplified by initiatives such as the UK Biobank, Kaiser Permanente's clinical analytics, and Singapore's national e-health infrastructure—have demonstrated their ability to detect disease progression, support clinical decision-making, and reduce healthcare burdens. The success of such systems relies heavily on the availability of robust, standardized, and longitudinal datasets. In Uzbekistan, large-scale digital transformation in healthcare is ongoing; however, several structural challenges impede the development of unified medical datasets [3–5].

This study examines the current state of clinical data collection in Uzbekistan and evaluates the practical steps needed to develop high-quality datasets for AI-based predictive monitoring of diabetes and hypertension. The paper focuses on identifying key problems in data completeness,

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

interoperability, standardization, and governance, while proposing feasible solutions aligned with global best practices.

2. Methods

2.1 Study Design

A mixed-methods research design was used to capture both the structural characteristics of Uzbekistan's healthcare data ecosystem and the quantitative attributes of available datasets. The study included:

- **Qualitative assessment:** evaluation of digitalization levels in healthcare facilities, analysis of EHR and LIS capabilities, and review of existing dispensary follow-up practices.
- **Quantitative analysis:** assessment of data completeness, accuracy, duplication, and coding consistency.
- **Model-driven data requirement analysis:** identification of essential data elements needed to train predictive models focused on diabetes and hypertension.

2.2 Data Sources

The pilot study involved health data from **112,846 patients** across one major city (Tashkent) and one regional area. The integrated dataset included:

- EHR records (diagnoses, vitals, encounters)
- Laboratory results (HbA1c, glucose, lipid panel, creatinine)
- Pharmacy and prescription data
- Dispensary follow-up registries
- Wearable device data (physical activity, heart rate, sleep metrics)

2.3 Data Processing

Cleaning

Data were processed to remove erroneous values, duplicate patient records, inconsistent timestamps, and physiologically impossible measurements.

Standardization

Clinical and laboratory records were mapped to:

- **ICD-10** for diagnoses
- **LOINC** for laboratory codes
- **HL7 FHIR** format for data structuring

Integration

A **federated data architecture** was developed, allowing healthcare facilities to maintain local storage while securely contributing de-identified datasets to a national data lake.

2.4 Ethical and Legal Considerations

Data governance followed internationally accepted principles of informed consent, pseudonymization, access control, and audit logging. Compliance with GDPR-like standards was maintained during all stages of data processing.

2.5 AI Models

Three predictive models were evaluated:

- **XGBoost** for forecasting diabetic complications

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

- **LSTM-based time-series model** for predicting blood pressure spikes
- **Random Forest** for estimating rehospitalization risk

Model performance was assessed using AUC, precision, recall, F1-score, and MAE metrics.

3. Results

3.1 Data Completeness

The pilot dataset exhibited varying degrees of completeness:

- Blood pressure measurements — **86%**
- HbA1c — **61%**
- Glucose — **74%**
- Lipid profile — **55%**
- Prescription data — **91%**
- Dispensary visit history — **83%**

Laboratory data were the most incomplete, representing a critical challenge for predictive model training.

3.2 Data Quality Issues

Cleaning procedures identified:

- Duplicate patient records — **6.1%**
- Inconsistent or incorrect ICD-10 coding — **4.3%**
- Physiologically impossible vitals — **1.8%**
- Timestamp inconsistencies — **1.2%**

3.3 Interoperability Outcomes

FHIR-based structuring yielded:

- Patient — **100%**
- Observation — **84%**
- Encounter — **79%**
- MedicationRequest — **91%**

Integration rates were:

- EHR ↔ LIS — **52%**
- EHR ↔ Pharmacy — **83%**
- EHR ↔ Dispensary registry — **71%**

Laboratory systems represented the weakest integration point.

3.4 AI Model Performance

- **XGBoost** for diabetic complications:
- AUC = **0.82**
- Precision = **0.76**, Recall = **0.71**
- **LSTM** for hypertension spikes:
- AUC = **0.87**
- MAE = **7.3 mmHg**
- **Random Forest** for rehospitalization:
- AUC = **0.78**, F1 = **0.72**

Models demonstrated strong potential, though performance was limited by missing laboratory values and regional dataset imbalances.

4. Discussion

The findings indicate that Uzbekistan possesses significant potential to develop AI-enabled monitoring systems for chronic diseases; however, several structural constraints must be

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

addressed. High completeness in prescription data and encounter history shows progress in digitalization, yet laboratory data fragmentation hinders longitudinal analysis and risk prediction. The limited integration between EHRs and LIS is consistent with common challenges in lower- and middle-income countries, where heterogeneous vendors and non-standardized systems impede seamless data exchange [4,5]. Implementing national interoperability standards (HL7 FHIR, LOINC, ICD-11) is essential to ensure data consistency.

The federated data architecture tested in the pilot proved feasible and aligns with global approaches such as the EU Health Data Space. This model is particularly suitable for Uzbekistan's geographically diverse healthcare system, offering improved privacy protection and reduced infrastructure barriers.

AI models yielded clinically valuable results, but their performance varied across regions. Urban datasets produced higher accuracy, while rural data gaps led to reduced model reliability. This highlights the need for equitable data collection strategies to avoid algorithmic bias.

Regulatory and ethical considerations remain underdeveloped. Comprehensive data protection laws, AI-specific health regulations, and professional training in data ethics are necessary prerequisites for nationwide implementation.

5. Conclusion

This study demonstrates that Uzbekistan has the infrastructural and demographic capacity to develop AI-based predictive monitoring systems for diabetes and hypertension. Nevertheless, the successful deployment of such systems depends heavily on addressing systemic challenges in data quality, completeness, and interoperability. Key recommendations include:

- Establishing national standards for data collection (FHIR, LOINC, ICD-11)
- Strengthening LIS–EHR integration
- Adopting federated data architectures for secure scalability
- Enhancing data governance and privacy regulations
- Training healthcare professionals in digital and AI competencies

If implemented, these measures will significantly enhance early detection, reduce complications, and improve the overall effectiveness of chronic disease management in Uzbekistan.

References:

1. World Heart Federation. Uzbekistan – Cardiovascular Disease Country Profile, 2024.
2. World Health Organization. Hypertension: Uzbekistan Country Profile, 2023.
3. Yuz.uz. Unified Electronic Health Card Introduced in Uzbekistan, 2024.
4. Fraser H. et al. Electronic Health Records in LMIC: Data Quality Determinants, JMIR Public Health, 2024.
5. Abdul-Rahman T. et al. Challenges of Medical Data Archiving in LMICs, ScienceDirect, 2023.
6. Lam B. Wearable Trackers for Predicting Type 2 Diabetes, JMIR Diabetes, 2021.
7. Scientific Reports. AI Models for Diabetes Risk Prediction, Nature Publishing Group, 2024–2025.
8. Hoshimova Nilufar, Husanov Kamoliddin "SUN'IY INTELLEKTNING TIBBIYOTDA QO 'LLANILISHI VA AFZALLIKLARI" International conference on multidisciplinary science vol. 1, issue 5, pp 78-81, 2023/11/22
9. Hoshimova Nilufar, Husanov Kamoliddin "Sun'iy intellektning kasalxona boshqaruvida qollanilishi" TECHNICAL SCIENCE RESEARCH IN UZBEKISTAN vol. 1, issue 4, pp 268-272, 2023/11/22

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

10. Khusanov Kamoliddin, Akhrorov Muxammad, Hoshimova Nilufar " RAQAMLI TEXNOLOGIYALARNI INSON HAYOTIDAGI O'RNI" Research Focus vol.2, issue 1, pp592-600, 2023
11. KX Xusanov, NF Nematov "ADVANCED AI-DRIVEN FRAMEWORK FOR LUNG CANCER DETECTION USING CT SCAN IMAGES AND MOBILE APPLICATION INTEGRATION" Medicine, pedagogy and technology: theory and practice vol. 2, issue 11, pp 397-406, 2024/11/30