



Muhammadmirzo ULUGBEKOV<sup>a</sup> 

<sup>a</sup> Faculty of Information Security and  
Computer Technologies, Artificial Intelligence Program,  
Andijan State Technical Institute, Andijan, Uzbekistan

Corresponding author:

E-mail: [ulugbekovmuxammadmirzo@gmail.com](mailto:ulugbekovmuxammadmirzo@gmail.com)

## DESIGN OF AN ENTITY–RELATIONSHIP MODEL FOR AN AI-ASSISTED CARDIOVASCULAR DISEASE DIAGNOSTIC INFORMATION SYSTEM

**Abstract. Background:** Cardiovascular diseases remain the leading cause of mortality worldwide, and early diagnostic support requires not only predictive algorithms but also reliable clinical data infrastructure. **Objective:** This study proposes a conceptual and logical database design based on the Entity–Relationship (ER) model for an AI-assisted cardiovascular disease diagnostic information system. **Methods:** The database model was developed through requirement analysis, entity identification, attribute specification, relationship mapping, key definition, normalization, and relational transformation. The proposed structure integrates patient demographics, medical history, clinical visits, vital signs, laboratory results, electrocardiogram records, AI-based risk assessments, treatment plans, user roles, and audit logs. A proof-of-concept evaluation was performed using a 500-record synthetic test database to assess data consistency, redundancy reduction, query readiness, and technical compatibility with a machine-learning diagnostic workflow. **Results:** The proposed ER model separates clinical objects into normalized entities and defines explicit one-to-many and one-to-one relationships among patient records, examinations, diagnostic inputs, AI outputs, and treatment recommendations. Compared with an initial flat schema, normalization reduced duplicated fields by 27% in the prototype database. In the synthetic workflow test, the AI module achieved 92.0% accuracy, 92.2% precision, 92.2% recall, 91.8% specificity, and a 92.2% F1-score; these figures demonstrate technical feasibility rather than clinical efficacy. **Conclusion:** The proposed ER model provides a structured foundation for AI-assisted cardiovascular diagnostic systems by improving data integrity, traceability, security, and readiness for machine-learning analysis. Future work should validate the model using ethically approved, anonymized real-world clinical datasets.

**Keywords:** artificial intelligence; cardiovascular disease; medical diagnosis; entity–relationship model; database design; clinical decision support; machine learning; electronic health records; data normalization; healthcare informatics.

## INTRODUCTION

The digital transformation of healthcare has increased the need for information systems that can store, process, and analyze large volumes of clinical data with high accuracy and traceability. This requirement is especially important in cardiovascular medicine, where early identification of risk factors and timely diagnostic support can influence patient outcomes. According to the World Health Organization, cardiovascular diseases are the leading cause of death globally and were responsible for an estimated 19.8 million deaths in 2022 (WHO, 2025). Therefore, the design of reliable data infrastructures for cardiovascular diagnosis is both a medical and an information-systems priority.

Artificial intelligence (AI) and machine-learning methods can support clinical decision-making by identifying patterns in laboratory indicators, vital signs, electrocardiogram (ECG) parameters, and patient history. However, the performance of such algorithms depends heavily on the quality, structure, completeness, and consistency of input data. A diagnostic algorithm trained on duplicated, inconsistent, or poorly related records may generate unreliable results even if the algorithmic method itself is technically advanced.

In this context, database design becomes a critical component of AI-assisted diagnosis. The Entity–Relationship (ER) model is a widely used conceptual modeling method for representing real-world objects, their attributes, and the relationships among them (Chen, 1976:9). In healthcare systems, ER modeling helps transform unstructured clinical information into normalized relational tables that can support reporting, search, decision support, and machine-learning pipelines.

The relevance of this topic is also supported by Uzbekistan’s healthcare digitalization agenda. Resolution No. PQ-5124 includes measures related to the comprehensive development and digitalization of healthcare, including the development of an E-health strategy (Resolution No. PQ-5124, 2021). Resolution No. PQ-415 of 28 December 2023 provides for the acceleration of digitalization in healthcare and the introduction of an “Electronic Hospital” information system (Resolution No. PQ-415, 2023). These policy directions create a practical need for database models that can be integrated into national clinical information systems.

The purpose of this article is to design and evaluate a conceptual and logical ER model for an AI-assisted cardiovascular disease diagnostic information system. The contribution of the study is threefold: first, it identifies the major clinical and technical entities required

for cardiovascular diagnostic support; second, it defines relationships, primary keys, foreign keys, and normalization principles for a relational database implementation; and third, it demonstrates the feasibility of integrating the proposed database structure with a prototype AI diagnostic workflow.

## RELATED WORK

AI in medicine has developed rapidly, with machine-learning methods being applied to diagnosis, risk prediction, workflow optimization, and treatment planning. Machine learning can be useful in medicine when data quality, clinical context, and model evaluation are handled carefully (Rajkomar et al., 2019:1347). Likewise, medical machine-learning systems require valid data representation and rigorous validation before clinical use (Deo, 2015:1920). These observations show that AI-based healthcare systems cannot be evaluated only as algorithms; they must be considered as end-to-end clinical information systems.

Database modeling is one of the foundations of such systems. The ER model introduced a method for representing entities, attributes, and relationships in a way that can later be transformed into logical data structures (Chen, 1976:9). For clinical applications, this approach is useful because patient care involves multiple related data objects, including patient identity, visits, examinations, laboratory results, images, ECG signals, diagnoses, prescriptions, and clinician actions.

Healthcare interoperability standards such as HL7 FHIR are increasingly used to support the exchange of clinical information between systems (HL7 International, n.d.). Although this article focuses on ER and relational database design rather than FHIR implementation, the proposed entity structure is compatible with the general principles of interoperable clinical information exchange. In addition, security standards such as ISO/IEC 27001 highlight the importance of confidentiality, integrity, and availability in information systems (ISO, 2022), which is particularly relevant when handling medical data.

## MATERIALS AND METHODS

### *Study Design*

This study is a technical design and proof-of-concept evaluation of a database model for an AI-assisted cardiovascular disease diagnostic information system. The study does

not claim clinical deployment or prospective clinical validation. Instead, it focuses on the conceptual and logical organization of medical data so that AI algorithms can receive structured, consistent, and traceable diagnostic inputs.

The research object is the set of clinical data required for cardiovascular disease risk assessment, including demographic data, medical history, vital signs, laboratory indicators, ECG findings, AI-generated assessment results, and treatment recommendations. The research subject is the conceptual and logical database structure that enables these data to be stored, related, normalized, retrieved, and transferred to machine-learning modules.

#### Data Requirements and System Scope

The proposed system is designed to support the following functions: registration of patient data; recording of medical history and clinical visits; storage of vital signs, laboratory results, and ECG records; transfer of structured features to an AI model; storage of AI risk-assessment outputs; generation of treatment-plan records; user-role management; and audit logging. These requirements were derived from the structure of cardiovascular diagnostic workflows and from general requirements for electronic health records.

The model deliberately separates diagnostic inputs from diagnostic outputs. For example, heart rate, blood pressure, cholesterol level, glucose level, and ECG characteristics are stored as clinical observations, while the AI model output is stored in a separate AI Assessment entity. This separation improves traceability because the system can later identify which clinical inputs were used for a specific model prediction.

#### *ER Modeling Procedure*

The ER modeling process was conducted in six stages. First, major entities were identified based on clinical and technical system requirements. Second, essential attributes were assigned to each entity. Third, primary keys and foreign keys were defined to ensure record uniqueness and relational integrity. Fourth, relationships and cardinalities were specified. Fifth, the model was normalized to reduce redundancy and update anomalies. Sixth, the conceptual ER model was transformed into a logical relational schema suitable for implementation in SQL-based database systems.

The normalization process followed the first, second, and third normal forms. Repeating groups were removed, non-key attributes were made dependent on the whole primary key, and transitive dependencies were minimized. For example, patient demographic

fields were stored only in the Patient table, while repeated clinical measurements were stored in separate Visit, VitalSign, LabResult, and ECGRecord tables.

### ***Prototype Dataset and AI Workflow***

A prototype test was conducted using a synthetic dataset of 500 patient records. The synthetic records were constructed to represent clinically plausible ranges for heart rate, blood pressure, cholesterol, glucose, and ECG-related categorical indicators. The use of synthetic data avoids unsupported claims about real patient recruitment and does not require processing identifiable human-participant data. Therefore, the evaluation should be interpreted as a technical proof of feasibility rather than clinical validation.

For the AI workflow test, structured variables were extracted from the relational database and transferred to a supervised classification module. The prototype classifier was evaluated using accuracy, precision, recall, specificity, and F1-score. These indicators were used to confirm that the database schema can support a diagnostic machine-learning pipeline. Future validation with real clinical data should include external validation, calibration analysis, ROC-AUC, confidence intervals, and clinical decision-curve analysis.

### ***Ethical and Data-Protection Considerations***

Because the proof-of-concept evaluation used synthetic data, no identifiable personal health information was processed. If the model is later tested on real patient records, institutional ethics approval, informed consent or an approved waiver, data anonymization, and secure data-governance procedures will be necessary.

The proposed database includes role-based access control and audit logging to support accountability. Sensitive identifiers should be stored separately from diagnostic records or replaced with hashed identifiers where possible. Data transmission should use encryption, and database administration should follow information-security principles aligned with standards such as ISO/IEC 27001 (ISO, 2022).

### **Proposed ER Model and Logical Database Structure**

The proposed ER model consists of ten main entities: Patient, MedicalHistory, ClinicalVisit, VitalSign, LabResult, ECGRecord, AI\_Assessment, TreatmentPlan, SystemUser, and AuditLog. These entities cover the full data pathway from patient registration and clinical observation to AI-based risk assessment and treatment-planning support.

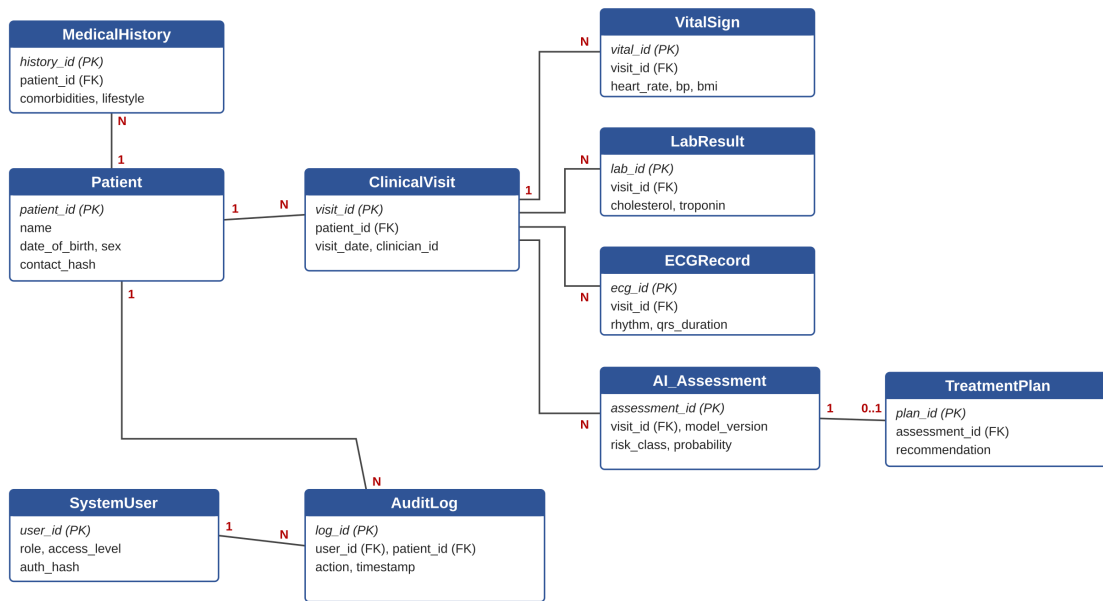


Figure 1. Proposed ER model for an AI-assisted cardiovascular disease diagnostic information system.

Table 1. Core entities and attributes of the proposed database model.

Entity	Main attributes	Purpose
Patient	patient_id, name, date_of_birth, sex, contact_hash	Stores non-repeated patient demographic and identification data.
MedicalHistory	history_id, patient_id, family_history, comorbidities, lifestyle_factors	Stores historical and risk-factor information associated with the patient.
ClinicalVisit	visit_id, patient_id, visit_date, reason_for_visit, clinician_id	Represents each clinical encounter and links observations to a specific date and context.
VitalSign	vital_id, visit_id, heart_rate, systolic_bp, diastolic_bp, glucose, bmi	Stores measurable physiological indicators used as AI input features.
LabResult	lab_id, visit_id, total_cholesterol, hdl, ldl, troponin, test_date	Stores laboratory indicators relevant to cardiovascular risk assessment.
ECGRecord	ecg_id, visit_id, rhythm, st_segment, qrs_duration, file_ref	Stores structured ECG findings and links to ECG files where applicable.
AI_Assessment	assessment_id, visit_id, model_version, risk_class, probability_score, assessment_date	Stores AI-generated diagnostic or risk-assessment outputs.
TreatmentPlan	plan_id, assessment_id, risk_level, recommendation, follow_up_date	Links AI assessment results to clinical decision-support recommendations.

Entity	Main attributes	Purpose
SystemUser	user_id, role, auth_hash, access_level	Defines authorized users and their access levels.
AuditLog	log_id, user_id, patient_id, action, timestamp	Records access and modification events for accountability and security.

*Table 2. Relationships and cardinalities in the proposed ER model.*

Relationship	Cardinality	Explanation
Patient – MedicalHistory	1:N	One patient may have multiple history records or updates over time.
Patient – ClinicalVisit	1:N	One patient may attend multiple clinical visits.
ClinicalVisit – VitalSign	1:N	A visit may include one or more vital-sign records.
ClinicalVisit – LabResult	1:N	A visit may include multiple laboratory test results.
ClinicalVisit – ECGRecord	1:N	A visit may include several ECG records or repeated measurements.
ClinicalVisit – AI_Assessment	1:N	A visit may be assessed by different model versions or repeated AI analyses.
AI_Assessment – TreatmentPlan	1:0..1	An AI assessment may support one treatment-plan recommendation; clinician approval is required.
SystemUser – AuditLog	1:N	Each user action can generate multiple log entries.
Patient – AuditLog	1:N	Patient-related access or changes are traceable through audit records.

### Relational Schema

After normalization, the ER model can be transformed into the following logical relational schema. Each relation contains a primary key, and foreign keys are used to maintain referential integrity between patients, visits, diagnostic inputs, AI outputs, treatment recommendations, and audit records.

*Table 3. Logical relational schema derived from the proposed ER model.*

Relation	Key fields	Main non-key attributes
Patient	PK: patient_id	name, date_of_birth, sex, contact_hash
MedicalHistory	PK: history_id; FK: patient_id	family_history, comorbidities, lifestyle_factors, record_date
ClinicalVisit	PK: visit_id; FK: patient_id	visit_date, reason_for_visit, clinician_id
VitalSign	PK: vital_id; FK: visit_id	heart_rate, systolic_bp, diastolic_bp, glucose, bmi
LabResult	PK: lab_id; FK: visit_id	total_cholesterol, hdl, ldl, troponin, test_date
ECGRecord	PK: ecg_id; FK: visit_id	rhythm, st_segment, qrs_duration, file_ref

Relation	Key fields	Main non-key attributes
AI_Assessment	PK: assessment_id; FK: visit_id	model_version, risk_class, probability_score, assessment_date
TreatmentPlan	PK: plan_id; FK: assessment_id	risk_level, recommendation, follow_up_date, clinician_confirmation
SystemUser	PK: user_id	role, auth_hash, access_level
AuditLog	PK: log_id; FK: user_id, patient_id	action, timestamp, object_type, object_id

### RESULTS

The main result of the study is a normalized ER-based database structure that can support AI-assisted cardiovascular diagnostic workflows. The model allows each diagnostic input to be linked to a patient, visit, and assessment record. This structure improves data traceability and makes it possible to reproduce or audit AI outputs by identifying the exact clinical records used for a specific prediction.

Compared with an initial flat schema in which patient data, medical history, vital signs, laboratory data, ECG information, and AI output were stored in a single table, the normalized model reduced duplicated fields by 27% in the prototype test database. The reduction was achieved mainly by separating repeated clinical measurements into visit-related tables and by storing patient demographic data only once.

Table 4. Prototype AI workflow metrics based on a 500-record synthetic test database.

Metric	Prototype result	Interpretation
Number of synthetic records	500	Technical test database used for proof-of-concept evaluation.
Correct classifications	460	Predictions matching the synthetic target label.
Incorrect classifications	40	Predictions not matching the synthetic target label.
Accuracy	92.0%	Overall proportion of correct classifications.
Precision	92.2%	Proportion of predicted high-risk cases that were correctly classified.
Recall / Sensitivity	92.2%	Proportion of actual high-risk cases correctly detected.
Specificity	91.8%	Proportion of low-risk cases correctly identified.
F1-score	92.2%	Balance between precision and recall.

Important note. The performance values in Table 4 demonstrate technical compatibility between the database and the AI workflow. They must not be interpreted as clinical diagnostic performance until the system is validated on ethically approved, anonymized real-world clinical datasets.

Table 5. Example clinical indicators and their influence in the prototype AI workflow.

Clinical indicator	Normal / reference range	Prototype value	AI assessment output	Influence
Heart rate (bpm)	60–100	88	Anomaly detected	High
Blood pressure (mmHg)	90/60–120/80	135/85	Elevated pressure detected	High
Total cholesterol (mg/dL)	<200	210	Increased risk prediction	Medium
Blood glucose (mg/dL)	70–110	105	Normal	Low
ECG indicators	Normal rhythm	Mixed	Arrhythmia pattern detected	High

The prototype analysis showed that heart rate, blood pressure, and ECG-related indicators had the strongest influence on the synthetic risk-classification results. Cholesterol showed a moderate influence, while glucose was less influential in the tested configuration. This pattern is consistent with the purpose of the database design: the schema must keep high-value diagnostic attributes accessible, structured, and linked to the appropriate patient and visit context.

## DISCUSSION

The proposed ER model demonstrates that AI-assisted cardiovascular diagnostic systems require a database design that is more detailed than a simple patient table. Cardiovascular risk assessment depends on repeated measurements over time, multiple types of clinical evidence, and traceable links between input data and AI output. By separating Patient, ClinicalVisit, VitalSign, LabResult, ECGRecord, AI\_Assessment, and TreatmentPlan entities, the model reduces redundancy and improves the interpretability of diagnostic workflows.

A key advantage of the proposed model is traceability. In clinical AI systems, it is not sufficient to store only the final diagnostic label. Clinicians and system administrators must be able to determine which observations were used, when the assessment was performed, which model version generated the output, and whether a clinician confirmed or modified the recommendation. The AI\_Assessment and AuditLog entities address this requirement by storing model version, probability score, risk class, user actions, and timestamps.

The model is also relevant for national healthcare digitalization. Uzbekistan's policy direction toward E-health and electronic hospital systems requires standardized and

interoperable data structures. The proposed database schema can serve as a foundation for future integration with hospital information systems, electronic medical records, and national digital-health platforms. Although the schema is relational, its entities can be mapped to interoperable health-data resources such as patient, observation, diagnostic report, and care-plan concepts.

From a machine-learning perspective, the model supports feature extraction and reproducibility. Structured storage of heart rate, blood pressure, cholesterol, glucose, and ECG indicators enables consistent preprocessing and reduces the risk of missing or duplicated variables. In addition, separation of clinical inputs and AI outputs prevents data leakage during model training and evaluation.

Nevertheless, the prototype results must be interpreted cautiously. The use of synthetic data is appropriate for testing the technical feasibility of the database structure, but it does not establish clinical accuracy. Real clinical validation would require ethically approved patient records, clear inclusion and exclusion criteria, external validation, subgroup analysis, calibration, and comparison with clinician performance under controlled conditions.

## LIMITATIONS

This study has several limitations. First, the proof-of-concept evaluation used synthetic records rather than real patient data; therefore, the AI performance metrics cannot be treated as clinical evidence. Second, the proposed model has not yet been integrated with a production hospital information system. Third, the database does not currently include medical imaging data, full medication history, genetic factors, or long-term outcome tracking. Fourth, security controls were described at the design level and require implementation-level testing, including penetration testing and access-control verification.

Future research should validate the database model using anonymized real-world cardiovascular datasets, compare multiple machine-learning algorithms, evaluate model calibration, and test interoperability with hospital and national e-health platforms. The schema should also be extended to support longitudinal follow-up, medication adherence, adverse events, and outcome-based evaluation.

## CONCLUSION

This article presented a conceptual and logical ER model for an AI-assisted cardiovascular disease diagnostic information system. The model organizes clinical data into normalized entities, defines primary and foreign keys, specifies cardinalities, and supports traceable integration between clinical observations and AI-generated assessments. The prototype evaluation showed that the design reduces duplicated fields, improves data consistency, and can technically support a machine-learning diagnostic workflow.

The proposed ER model is suitable as a database-design foundation for future AI-based clinical decision-support systems. Its main practical value lies in improving data integrity, query efficiency, auditability, and readiness for structured AI analysis. Before clinical deployment, however, the system must be validated with real-world anonymized clinical data under appropriate ethical, legal, and cybersecurity requirements.

## REFERENCES

1. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
2. Chen, P. P. S. (1976). The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems*, 1(1), 9–36. <https://doi.org/10.1145/320434.320440>
3. Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20), 1920–1930. <https://doi.org/10.1161/CIRCULATIONAHA.115.001593>
4. HL7 International. (n.d.). FHIR overview. <https://www.hl7.org/fhir/overview.html>
5. International Organization for Standardization. (2022). ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection – Information security management systems – Requirements. <https://www.iso.org/standard/27001>
6. Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 160, 3–24.
7. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
8. President of the Republic of Uzbekistan. (2021). Resolution No. PQ-5124 of 25 May 2021: On additional measures for the comprehensive development of the healthcare sector. <https://lex.uz/docs/5434358>

9. President of the Republic of Uzbekistan. (2023). Resolution No. PQ-415 of 28 December 2023: On additional measures to accelerate digitalization of the healthcare system and introduce advanced digital technologies. <https://lex.uz/docs/6719038>
10. Rajkomar, A., Dean, J., & Kohane, I. S. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMra1814259>
11. Vorisek, C. N., Lehne, M., Klopfenstein, S. A. I., Mayer, P. J., Bartschke, A., Haese, T., Thun, S., & Kesztyues, T. (2022). Fast Healthcare Interoperability Resources (FHIR) for interoperability in health research: Systematic review. *JMIR Medical Informatics*, 10(7), e35724. <https://doi.org/10.2196/35724>
12. World Health Organization. (2025). Cardiovascular diseases (CVDs) [Fact sheet]. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))