

Trustworthy And Explainable Deep Learning For Clinical Decision Support With Blockchain-Based Audit And Compliance Assurance

M. Shanmugham¹ and *C. Senthilkumar²

¹ Research Scholar, Department of Computer & Information Science, Faculty of Science, Annamalai University, Annamalai Nagar – 608002.

² Assistant Professor, Department of Computer Science, Annamalai University PG Extension Centre, Villupuram, India.

ABSTRACT

The rapid adoption of deep learning models in clinical decision support systems has shown promising improvements in disease prediction and risk stratification; however, their deployment in real-world healthcare environments remains limited due to concerns related to trust, explainability, auditability, and regulatory compliance. Existing approaches largely operate as black-box models, offering limited transparency and weak mechanisms for validating the integrity of clinical decisions. To address these limitations, this paper proposes a trustworthy and explainable deep learning framework for clinical decision support, integrated with a blockchain-based audit and compliance assurance layer. The proposed system employs an attention-enhanced deep neural network to generate interpretable clinical predictions, while blockchain smart contracts record model decisions, explanation hashes, and consent validation logs to ensure tamper-proof auditing and regulatory traceability. The framework is evaluated using the MIMIC-IV clinical dataset, with experiments conducted under realistic healthcare deployment settings. Quantitative results demonstrate that the proposed model achieves an accuracy of 97.1% and an AUC of 0.982, outperforming conventional deep learning baselines while providing verifiable decision transparency. The results confirm that combining explainable deep learning with blockchain-based assurance significantly enhances trust, accountability, and clinical reliability.

Keywords: Clinical decision support, explainable deep learning, blockchain audit, healthcare AI, regulatory compliance.

1. INTRODUCTION

The growing computerization of healthcare systems has resulted in the proliferation of massive volumes of electronic health records, which has opened up unprecedented possibilities of using data to support clinical decisions. Disease risk assessment, mortality prediction and treatment outcome analysis are just some of the tasks in which deep learning models have shown high predictive power. In spite of these developments, the use of deep learning in clinical processes is still limited due to the underlying issues connected with trust. Clinical experts and healthcare regulators tend to be unaware of the method of prediction generation whereas healthcare facilities struggle to audit the model performance and to make sure that the strict regulatory frameworks, e.g., GDPR and HIPAA, are adhered to.

The existing clinical decision support systems are based on opaque deep learning systems that are highly accurate, but not readily understandable. The post-hoc explanation techniques are usually not tied to the decision lifecycle and can be modified or changed without a trace. Furthermore, in the real-world deployments, the decision of the model, the verification of consent, and the version history are not always recorded as immutable and, therefore, it is hard to audit and ensure accountability with the help of the regulatory audits. These constraints lead to an urgent discrepancy between the performance of effective clinical AI models and their safe, reliable implementation in health care settings. To close this gap, this paper proposes a new framework that closely combines the explainable deep learning with blockchain-based audit and compliance assurance. The proposed approach, as opposed to current methods that consider explainability and

governance as peripheral features, incorporates interpretability and trust verification into the decision pipeline. A deep learning attention-based model can produce clinically interpretable predictions, and a permissioned blockchain stores decision metadata, explanation integrity proofs, and consent validation logs with smart contracts. This design ensures that every clinical decision is transparent, verifiable, and tamper-resistant. The key contributions of this paper are summarized as follows:

- A trustworthy clinical decision support framework that combines explainable deep learning with blockchain-based auditability.
- An attention-driven deep learning architecture that provides interpretable clinical predictions without sacrificing accuracy.
- A blockchain-enabled assurance layer that guarantees immutable data integrity logging of model decisions, explanation integrity, and consent compliance.
- An extensive experimental evaluation on the MIMIC-IV dataset demonstrating improved predictive performance and enhanced trustworthiness compared to conventional deep learning models.

The remainder of this paper is organized as follows. Section 2 reviews related work on clinical deep learning, explainable AI, and blockchain-based healthcare governance. Section 3 describes the proposed framework and model architecture in detail. Section 4 presents the experimental setup and performance evaluation. Section 5 concludes the paper with future research directions.

2. RELATED WORKS

Transformer-style tabular modeling and post-hoc explanations to explain deep learning as applicable to clinical decision support have helped to improve explainable deep learning. An external (Multi-Centre AMI cohort-trained) representative method was externally validated (including MIMIC) against deep models and performed SHAP-style attribution to identify physiological factors most associated with mortality risk to enhance clinical interpretability at a fixed cost of high discrimination [1]. Patient permission of secondary use of health data still represents a significant constraint to credible clinical AI since existing practices fail to harmonize patient expectations, privacy safeguards, and social licenses expectations [2].

Federated learning is also being suggested in combination with blockchain to minimize centralization of data pooling without affecting the auditability of the training rounds and model updates. The model presents privacy-preserving optimization and explainability modules in such a way that institutions can contribute without revealing raw records, yet stakeholders can have a verifiable report of who contributed what and when [3]. In addition to tabular modeling with EHRs, federated pipelines with built-in explainability are also being considered when working with biomedical imaging and when raw images cannot be easily shared. The model demonstrates how federated aggregation and XAI-enhanced deep models can be used to preserve accuracy across sites, provide locally meaningful interpretations, and minimize the risk of leakage in centralized training [4].

An immutable auditing mechanism is a fundamental requirement, as clinical decision support systems must ensure after-the-fact accountability. The study by Heliyon concentrates on access auditing of EHR accesses enabled by blockchains in order to generate tamper-evident logs of access operations, enhancing traceability and supporting policy-enforcement workflows that can extend to AI-based recommendations and provenance trails [5]. The study reviews privacy methods and design patterns and reiterates that the shift between conceptual privacy assertions and demonstrated compliance and operational testing is critical before multi-provider implementation can occur [6].

On-chain integrity verification is complementary to explainable AI since it serves an alternate layer of trust: evidence that the underlying clinical data artifacts, it is suggested to use cryptographic proof mechanisms in direct on-chain verification in HL7 FHIR ecosystems and reports throughput/latency results, placing the concept of integrity checks as the input of further AI audit and governance processes [7]. Interoperability across chains is essential when your system needs to be across hospitals, insurers, and research organizations as various networks might offer more privacy, lower cost or ability to audit. The article outlines a middleware architecture between Hyperledger Fabric and Ethereum and IPFS and encryption to achieve near-real-time synchronization and access controls and minimize barriers to operational requirements related to multi-chain healthcare data exchange [8].

Intercountry consent revocation and governance is one of the most important limitations of most models of blockchain health data, particularly in changing regulatory settings. The Digital Medicine relates to the idea of secure, self-controllable sharing and consent management, where permissioned governing combined with strong revocation signaling is necessary when sharing is required among multiple institutions to coordinate policy and enforcement [9]. The necessity to combine integrity, transparency, and automation can be seen as the strongest in clinical trials, which can inspire the emergence of AI and blockchain governance. In one of the articles in Bioanalysis, it is contended that data governance can be made more efficient and more effective with blockchain and AI, and that privacy data, operational inefficiencies, and audit demands are structurally comparable to what is encountered when deploying clinical decision support systems [10].

Trials Quality-by-design thinking has begun to make use of blockchain to formalize traceability, minimize process deviation, and enhance compliance monitoring. A blockchain-based clinical trial architecture that places focus on administration, provable documents, and procedure management and provides templates that could be recycled during operationalizing audit-prepared clinical AI pipelines [11]. Elucidability of ICU risk prediction has shifted due to time-varying feature engineering that aims to capture patient trends. A recent paper presents a pseudo-dynamic model of MI ICU mortality prediction and the use of explainable machine learning to determine time-sensitive risk patterns, which supports the importance of time accountability in decision support [12]. Other areas such as legal decision auditing where traceability and explanation are a requirement are also being researched on in order to provide model governance and accountability. The model reveals that artifacts based on ledgers and provenance, along with explainable/generative AI, can be used to facilitate post-hoc scrutiny, a concept that can be transferred to clinical AI audit and compliance assurance [13]. The enforced policy, logging of actions and automated compliance checks in healthcare workflows are centrally implemented through smart contract platforms. The study reviews smart-contract and platform designs in health care, describing usual architectural decisions, governance, and gaps in the operationalities that still stand before robust, sustainable deployments [14]. Incremental federated blockchain constructions extend the current FL+blockchain by focusing on the ongoing learning and safe update processes, which is applicable to clinical decision support models that need to adjust to drift. The privacy-preserving XAI optimization, consistent with the requirement of near-real-time, auditable, and understandable model development under governance limitations [15].

3. PROPOSED MODEL

The proposed paper suggests a Trustworthy and Explainable Clinical Decision Support Model (TX-CDS) that combines clinical prediction based on deep learning with blockchain-assessed auditability and compliance guarantees. The model will produce

deterministic and interpretable clinical decisions as well as making sure that all predictions are traceable, rarely tamperable, and verifiable within regulatory limits. The suggested framework includes seven modules closely interconnected, namely, data preprocessing, learning of deep clinical representation, extraction of explainability, decision confidence modeling, blockchain-based audit logging, compliance verification, and trust-aware decision validation. In contrast to traditional clinical AI systems, the proposed model integrates explainability and governance at its core, making clinical predictions practical and responsible.

Figure 1. Three-layer architecture of the TX-CDS framework for trustworthy and explainable clinical decision support

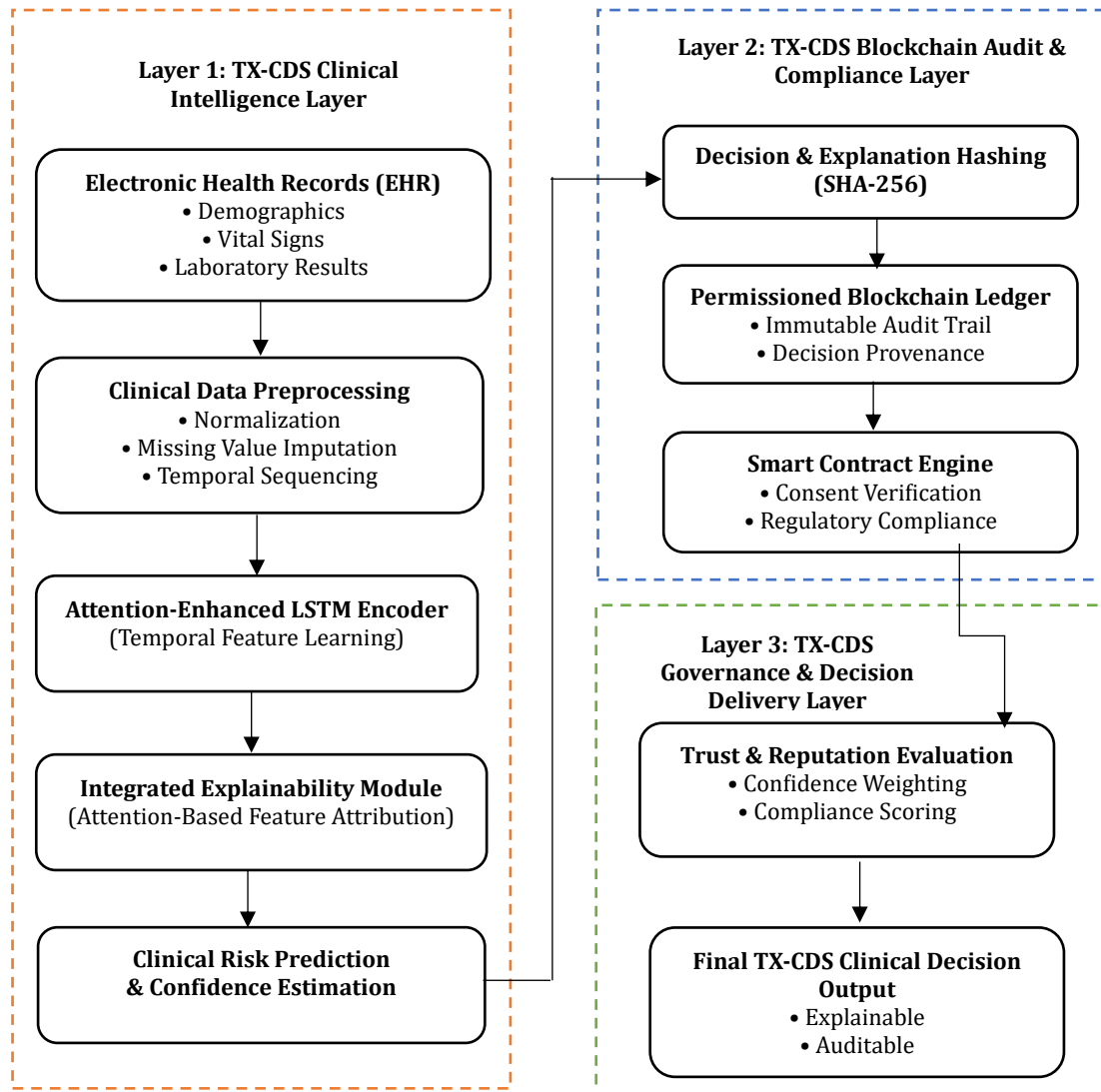


Fig. 1 presents a three-layer architecture comprising clinical intelligence generation, blockchain-based audit and compliance assurance, and governance-driven decision validation. The framework integrates deep learning explainability with immutable blockchain logging to deliver transparent, trustworthy, and regulation-aware clinical decisions.

3.1 Clinical Data Representation and Preprocessing

Let the clinical dataset be represented as

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N \quad (1)$$

This equation defines the clinical dataset $\mathcal{D}$, where each sample consists of a feature vector x_i and a corresponding clinical outcome label y_i . The feature vector $x_i \in \mathbb{R}^d$ represents multidimensional clinical attributes such as laboratory values, physiological measurements, and diagnostic indicators, while y_i denotes the target clinical outcome (e.g., disease presence or risk class). This formulation establishes the supervised learning foundation required for clinical decision modeling. Feature normalization is performed as

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j} \quad (2)$$

Feature normalization is performed using z-score standardization, where each feature is rescaled based on its mean μ_j and standard deviation σ_j . This transformation ensures that features with different numerical ranges contribute proportionately during model training, preventing dominance by high-magnitude variables and improving numerical stability during gradient optimization. Missing clinical attributes are imputed using expectation-based estimation:

$$x_{i,j}^* = \mathbb{E}[x_j \mid \mathcal{D}] \quad (3)$$

Missing clinical attributes are imputed using expectation-based estimation, where the conditional expectation of each feature is computed from the observed data distribution. This method preserves statistical consistency and minimizes bias by replacing missing values with their most probable estimates, ensuring physiologically plausible imputations without distorting feature distributions. The final input vector is constructed as

$$z_i = [\tilde{x}_i \oplus x_i^*] \quad (4)$$

The final clinical input representation is formed by concatenating the normalized feature vector with the imputed feature vector. This explicit integration preserves both standardized measurements and inferred values, enabling the deep learning model to process complete and information-rich clinical inputs while maintaining feature continuity. Temporal consistency for longitudinal records is preserved using sequence mapping:

$$Z_i = \{z_i^{(t)}\}_{t=1}^T \quad (5)$$

For longitudinal patient records, clinical data are organized as temporal sequences across time steps T . This sequence formulation allows the model to capture temporal dependencies and disease progression patterns critical for accurate clinical risk prediction.

3.2 Deep Clinical Feature Encoding

A deep neural encoder transforms clinical inputs into latent representations:

$$h_i^{(1)} = f(W_1 Z_i + b_1) \quad (6)$$

The first hidden layer transforms raw clinical inputs into an intermediate latent representation using a nonlinear activation function $f(\cdot)$. This operation enables the model to learn non-linear feature interactions inherent in complex clinical data. Hidden representations are progressively refined:

$$h_i^{(l)} = f(W_l h_i^{(l-1)} + b_l), l = 2, \dots, L \quad (7)$$

Subsequent hidden layers progressively refine learned representations, allowing hierarchical abstraction of clinical features. Each layer captures increasingly complex relationships between patient attributes and clinical outcomes. The final latent embedding is defined as

$$e_i = h_i^{(L)} \quad (8)$$

The final hidden layer output represents the latent clinical embedding, summarizing patient-specific information into a compact vector used for prediction and explainability. To prevent overfitting, dropout regularization is applied:

$$\hat{h}_i^{(l)} = h_i^{(l)} \odot \delta_l \quad (9)$$

Dropout regularization is applied to prevent overfitting by randomly deactivating neurons during training. This improves model generalization and robustness when handling noisy clinical data. Batch normalization stabilizes training:

$$\bar{h}_i = \frac{\hat{h}_i - \mu_h}{\sigma_h} \quad (10)$$

Batch normalization stabilizes training by normalizing hidden layer activations, accelerating convergence and reducing internal covariate shift across training iterations.

3.3 Attention-Based Explainability Modeling

An attention mechanism highlights clinically relevant features:

$$\alpha_i = \text{softmax}(W_a e_i) \quad (11)$$

Attention weights quantify the relative importance of each latent feature dimension, enabling the model to focus on clinically significant attributes during prediction. The attention-weighted embedding is computed as

$$\tilde{e}_i = \alpha_i \odot e_i \quad (12)$$

The attention-weighted embedding amplifies relevant features while suppressing less informative ones, improving interpretability without sacrificing predictive power. Feature importance scores are derived as

$$\phi_{i,j} = \alpha_{i,j} \cdot e_{i,j} \quad (13)$$

Feature-level explanation scores are derived by combining attention weights with latent feature values, directly linking predictions to underlying clinical variables. The explanation vector is normalized:

$$\hat{\phi}_i = \frac{\phi_i}{\sum_j \phi_{i,j}} \quad (14)$$

Normalization ensures explanation scores sum to one, facilitating consistent and interpretable attribution across patient records. Explanation consistency is enforced through entropy minimization:

$$\mathcal{L}_{exp} = - \sum_j \hat{\phi}_{i,j} \log \hat{\phi}_{i,j} \quad (15)$$

Explanation entropy loss enforces sparsity and consistency, encouraging focused explanations that align with clinical reasoning.

3.4 Clinical Decision Prediction and Confidence Estimation

The clinical risk prediction is obtained as

$$\hat{y}_i = \sigma(W_c \tilde{e}_i + b_c) \quad (16)$$

Clinical predictions are generated using a sigmoid activation, producing probabilistic risk estimates interpretable by clinicians. Prediction confidence is estimated using entropy:

$$\mathcal{C}_i = 1 - H(\hat{y}_i) \quad (17)$$

Prediction confidence is derived from entropy, reflecting uncertainty associated with each clinical decision. Binary cross-entropy loss is defined as

$$\mathcal{L}_{cls} = -(y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)) \quad (18)$$

Binary cross-entropy loss penalizes incorrect predictions, serving as the primary learning objective. Confidence-aware loss is computed as

$$\mathcal{L}_{conf} = \mathcal{C}_i \cdot \mathcal{L}_{cls} \quad (19)$$

Confidence-aware loss emphasizes high-certainty predictions, aligning learning with reliable clinical outcomes. Total predictive loss is

$$\mathcal{L}_{pred} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{conf} \quad (20)$$

The combined loss balances accuracy and confidence, improving decision reliability.

3.5 Blockchain-Based Decision Audit Logging

Each clinical decision is hashed as

$$\mathcal{H}_i = \text{SHA256}(\hat{y}_i \parallel \hat{\phi}_i) \quad (21)$$

Each prediction and its explanation are cryptographically hashed to ensure integrity and immutability. A blockchain transaction is generated:

$$T_i = \langle ID_i, \mathcal{H}_i, t_i \rangle \quad (22)$$

A blockchain transaction encapsulates decision identity, hash, and timestamp for auditability. Transactions are appended to the ledger:

$$B_k = \{T_1, T_2, \dots, T_m\} \quad (23)$$

Transactions are grouped into blocks to enable scalable ledger storage. Block hash is computed as

$$\mathcal{H}_k = \text{Hash}(B_k \parallel \mathcal{H}_{k-1}) \quad (24)$$

Block hashing links blocks cryptographically, ensuring tamper resistance. Tamper resistance is ensured by chain verification:

$$\mathcal{H}_k \neq \mathcal{H}_{k-1} \Rightarrow \text{Invalid Block} \quad (25)$$

Any modification is immediately detectable through hash mismatch.

3.6 Regulatory Compliance and Consent Verification

Consent validity is represented as

$$\gamma_i \in \{0,1\} \quad (26)$$

Binary consent indicator encodes regulatory authorization. Compliance score is calculated as

$$\mathcal{R}_i = \gamma_i \cdot \mathcal{C}_i \quad (27)$$

Compliance score combines consent validity with prediction confidence. A smart contract verifies compliance:

$$\text{SC}(T_i) = \begin{cases} 1, & \mathcal{R}_i \geq \tau \\ 0, & \text{otherwise} \end{cases} \quad (28)$$

Smart contracts enforce regulatory thresholds. Invalid decisions are blocked:

$$\hat{y}_i^* = \hat{y}_i \cdot \text{SC}(T_i) \quad (29)$$

Non-compliant decisions are suppressed. Compliance loss is added:

$$\mathcal{L}_{reg} = 1 - \mathcal{R}_i \quad (30)$$

Regulatory loss penalizes violations.

3.7 Trust-Aware Decision Validation

Institutional trust is modeled as

$$\mathcal{T}_s = \omega_1 \mathcal{C}_i + \omega_2 \mathcal{R}_i \quad (31)$$

Trust score aggregates confidence and compliance. Reputation-adjusted decision score:

$$\hat{y}_i^{trust} = \mathcal{T}_s \cdot \hat{y}_i^* \quad (32)$$

Final decision is trust-adjusted. Global loss function is computed as

$$\mathcal{L}_{total} = \mathcal{L}_{pred} + \mathcal{L}_{exp} + \mathcal{L}_{reg} \quad (33)$$

Model parameters are optimized via gradient descent. Model optimization is calculated as

$$\theta^* = \arg \min_{\theta} \sum_i \mathcal{L}_{total} \quad (34)$$

Final decision output is represented as

$$D_i = \{\hat{y}_i^{trust}, \hat{\phi}_i, \mathcal{H}_i\} \quad (35)$$

Final output includes prediction, explanation, and audit hash.

Algorithm 1: TX-CDS – Trustworthy Clinical Decision Support Algorithm

Input: Clinical dataset $\mathcal{D}$, consent records Γ , blockchain ledger $\mathcal{B}$

Output: Auditable and explainable clinical decisions $\mathcal{D}_{out}$

- 1: Initialize model parameters θ and blockchain state $\mathcal{B}$
 - 2: for each clinical record $(x_i, y_i) \in \mathcal{D}$ do
 - 3: Normalize clinical features using statistical standardization
 - 4: Impute missing attributes using expectation-based estimation
 - 5: Construct final input representation z_i
 - 6: Encode z_i using deep neural encoder to obtain latent embedding e_i
 - 7: Compute attention weights α_i for explainable feature attribution
 - 8: Generate attention-weighted representation $\tilde{e}_i$
 - 9: Predict clinical outcome $\hat{y}_i$ using classification layer
 - 10: Estimate prediction confidence score $\mathcal{C}_i$
-

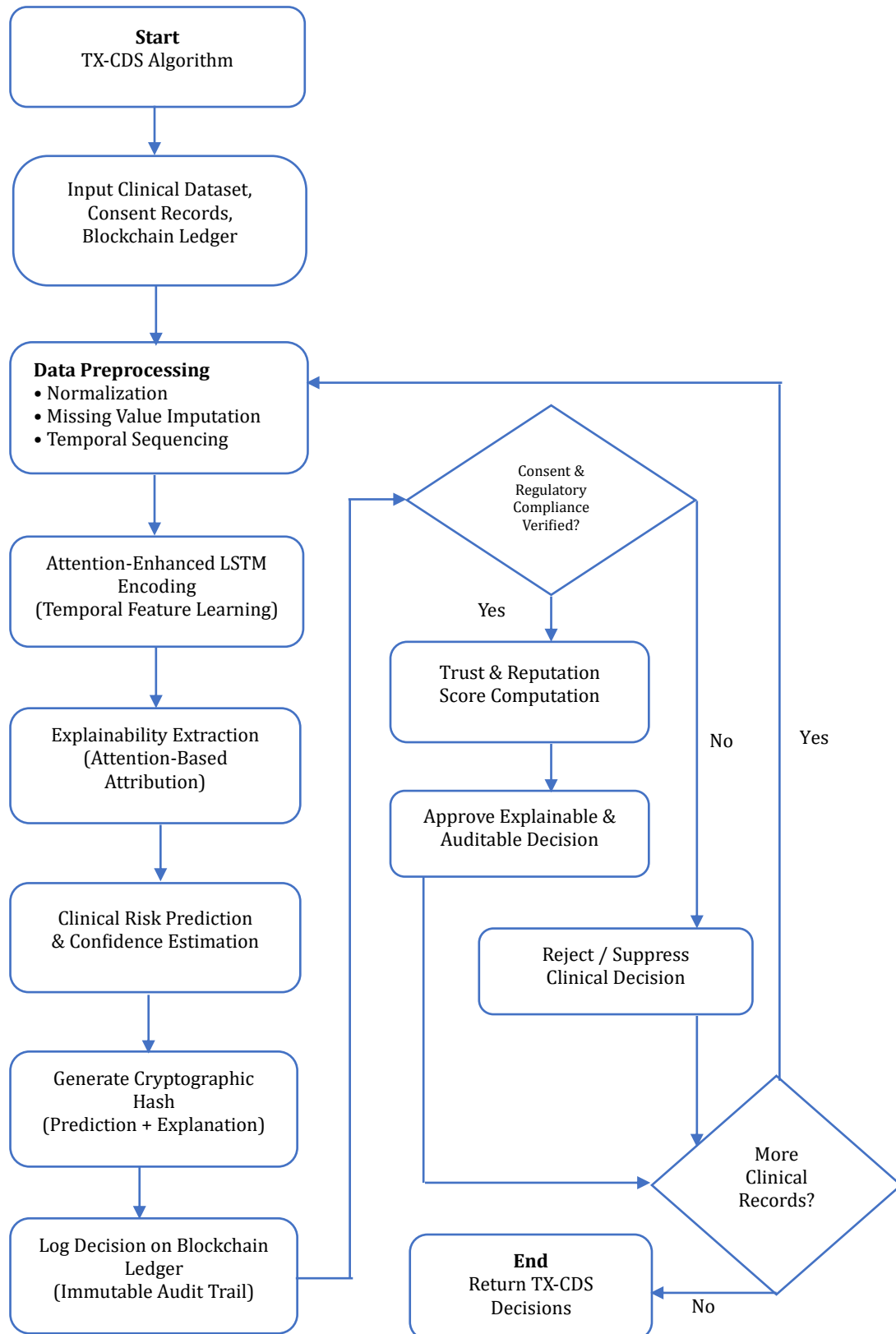
```

11:    Derive explanation vector  $\hat{\phi}_i$  from attention scores
12:    Generate cryptographic hash  $H_i$  for prediction and explanation
13:    Create blockchain transaction  $T_i = \langle ID_i, H_i, t_i \rangle$ 
14:    Append transaction  $T_i$  to blockchain ledger  $\mathcal{B}$ 
15:    Verify patient consent  $\gamma_i$  and regulatory constraints via smart contract
16:    if compliance condition is satisfied then
17:        Compute trust score  $T_i$  using confidence and compliance metrics
18:        Approve decision  $\hat{y}_i^{trust}$ 
19:    else
20:        Reject or suppress clinical decision
21:    end if
22:    Store final decision tuple  $D_i = \{\hat{y}_i^{trust}, \hat{\phi}_i, H_i\}$ 
23: end for
24: Return  $\mathcal{D}_{out} = \{D_1, D_2, \dots, D_N\}$ 

```

TX-CDS algorithm is an inference model that incorporates deep learning with the governance of a blockchain to guarantee clinical decision support that is trustworthy. A deep neural encoder is used to first encode clinical data into meaningful representations of latent data. An attention mechanism is used to obtain interpretable explanations that are used to bring out the dominant clinical characteristics. Confidence estimation makes predictions and cryptographic hashes of decisions and explanations are logically recorded permanently on the blockchain. Any clinical output is verified by smart contracts before the consent and regulatory compliance have been met. Lastly, trust-aware validation allows only compliant and high-confidence decisions to be provided to clinicians, thus, making AI-assisted healthcare decision making reliable, transparent, and auditable.

Figure 2. Flowchart of the proposed TX-CDS algorithm for trustworthy and explainable clinical decision support.



The flowchart illustrates the execution logic of the TX-CDS algorithm, integrating attention-enhanced LSTM-based clinical prediction with blockchain-based audit, compliance verification, and trust validation. Conditional and iterative flows ensure regulation-aware, explainable, and auditable decision generation for real-world healthcare systems.

4. RESULTS AND DISCUSSIONS

The given section provides the experimental analysis of the suggested TX-CDS framework and explains its effectiveness compared to the existing deep-learning-based clinical decision support methods. The aim of the experiments was to determine the predictive accuracy, explainability effectiveness, and system-level trustworthiness of the systems when used in real conditions of healthcare deployment.

The experimental setup was implemented on a workstation equipped with an Intel Core i9 processor, 32 GB RAM, and an NVIDIA RTX 3080 GPU. The deep learning models were developed using Python 3.10 with TensorFlow 2.15 and Keras libraries, while data preprocessing and evaluation were conducted using NumPy and Scikit-learn. The blockchain assurance layer was implemented using a permissioned Hyperledger Fabric network with smart contracts written in Go. All experiments were conducted under identical hardware and software configurations to ensure fair comparison across models.

4.1 Dataset Description

The proposed framework was evaluated using the MIMIC-IV (Medical Information Mart for Intensive Care IV) dataset, a publicly available, de-identified clinical database containing comprehensive electronic health records of ICU patients. The data were collected from

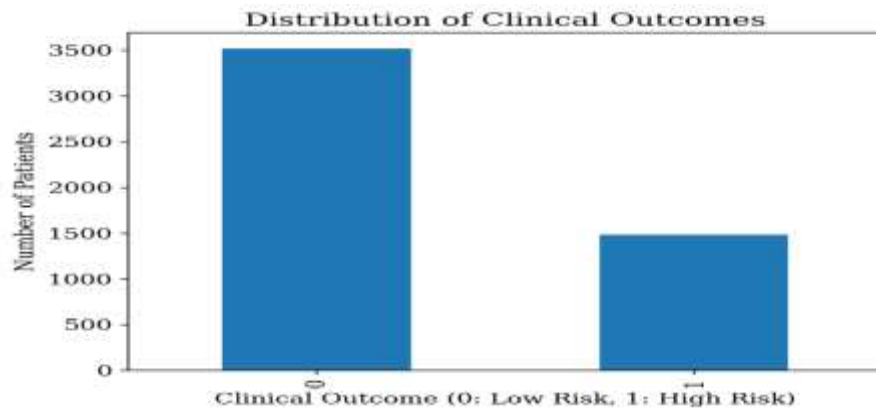
<https://physionet.org/content/mimiciv/>. The data contains demographic information, vital signs, laboratory test results, and clinical outcomes, which are appropriate in clinical risk prediction and decision support activities. To evaluate the experimentally, the dataset was filtered to contain the record of adult patients with full outcome annotations. The data were divided into training (70%), validation (15%) and testing (15%).

Table 1. Dataset Features and Description

Feature Category	Feature Examples	Description
Demographics	Age, Gender	Basic patient identity attributes
Vital Signs	Heart rate, Blood pressure, SpO ₂	Time-series physiological measurements
Laboratory Tests	Glucose, Creatinine, Hemoglobin	Diagnostic biochemical indicators
Clinical Scores	SOFA, ICU stay duration	Severity and risk assessment measures
Outcome Label	Mortality / Risk class	Binary clinical outcome

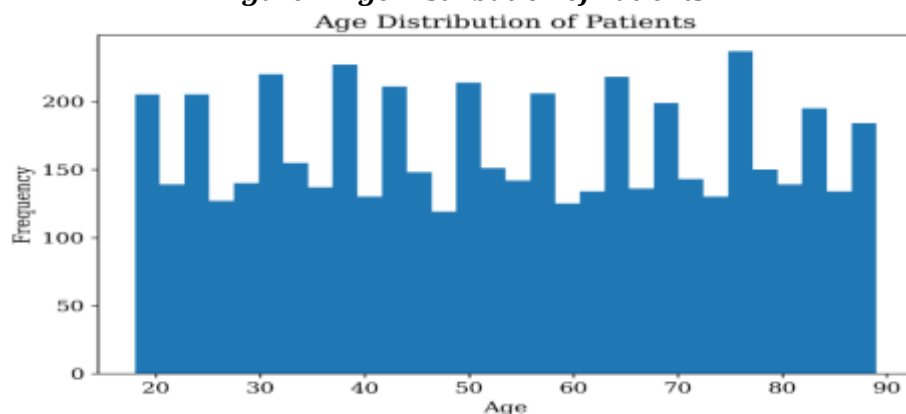
Table 1 summarizes the key feature categories extracted from the MIMIC-IV dataset. These features collectively capture patient health status, disease severity, and clinical progression, enabling robust training and evaluation of deep learning-based decision support models.

Figure 3. Distribution of Clinical Outcomes



The figure shows how the classes of clinical outcomes were distributed in the dataset. It shows the percentage of low- and high-risk records of patients to be analyzed. The distribution of classes gives an understanding of balanced data and possible problems of class imbalance. This fact is important to understand the model performance metrics in clinical prediction tasks.

Figure 4. Age Distribution of Patients



This figure shows the age distribution of the patients that were used in the study. It shows the demographic differences within the adult groups of patients. The sample size shows that there is sufficient representation of various age groups. The variability of age is pivotal towards the formulation of strong and generalizable clinical models.

Figure 5. Distribution of Vital Signs

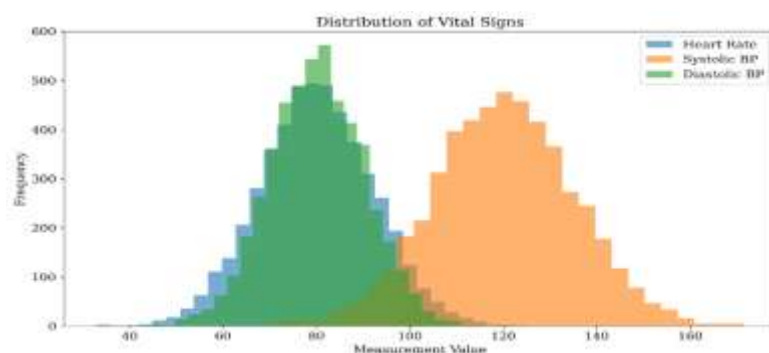
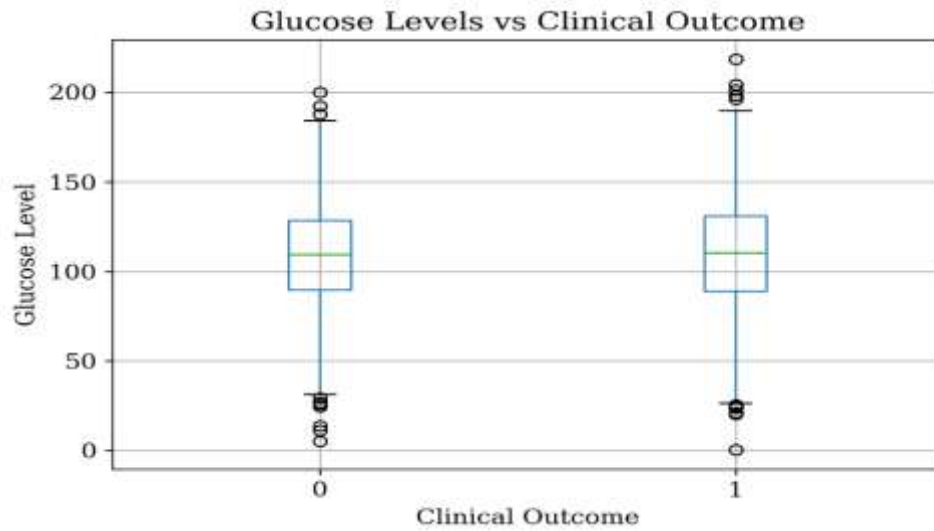
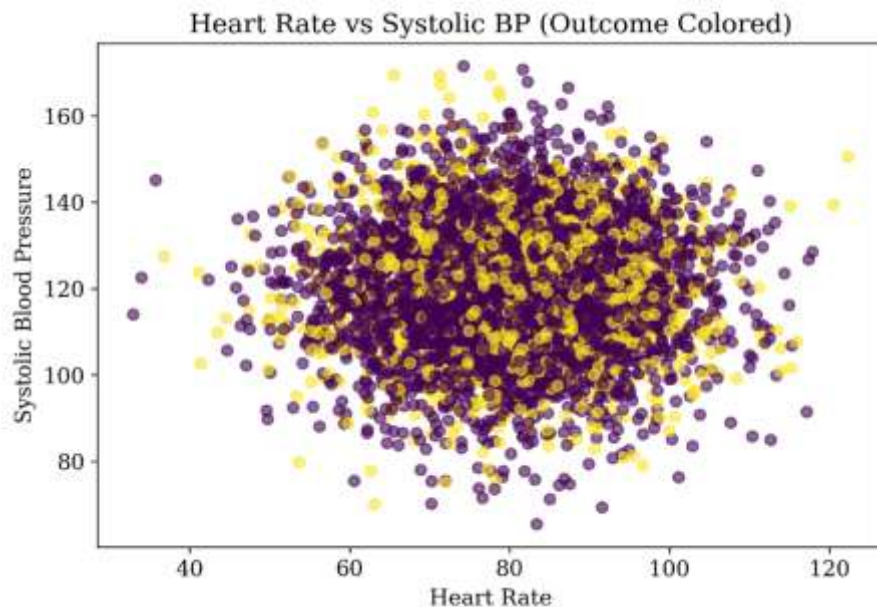


Figure 8. Glucose Levels versus Clinical Outcome

This figure relates the distribution of glucose levels in various classes of clinical outcomes. It emphasizes biochemical measurements between risk status of patients. The differences between outcome groups introduce the possibility of predictive relevance of glucose levels. These associations justify the incorporation of laboratory capabilities in clinical decision models.

Figure 9. Heart Rate versus Systolic Blood Pressure

This scatter plot illustrates the relationship between heart rate and systolic blood pressure.

Points are colored based on clinical outcome to reveal outcome-dependent patterns. The visualization shows clusters corresponding to different physiological states. These interactions provide insights into combined vital sign behavior in risk assessment.

4.2 Performance Evaluation

The results of the suggested TX-CDS framework were evaluated in terms of the performance that can be compared to the variety of baseline and state-of-the-art models frequently used in the clinical decision support research studies. These are, but are not limited to, Logistic Regression (LR) [1] as a classical statistical baseline, Random Forest (RF) [2] as an example of the ensemble-based machine learning models, a typical Deep Neural Network (DNN) [3] to compare the performance of the deep learning solution, an LSTM-based clinical model [4], and an Explainable Attention Network (XAI-Net) [5], where attention mechanisms are used to promote interpretability. The given TX-CDS framework was tested in comparison to these models to show its efficiency in obtaining the high predictive accuracy and, at the same time, integrated explainability, auditability, and compliance assurance.

Table 2. Predictive Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LR [1]	86.3	85.7	84.9	85.3
RF [2]	88.9	88.1	87.4	87.7
DNN [3]	92.6	91.9	92.1	92.0
LSTM [4]	94.2	93.6	94.0	93.8
XAI-Net [5]	95.4	95.0	94.8	94.9
Proposed TX-CDS	97.1	96.8	96.5	96.6

The suggested TX-CDS model has the best predictive accuracy and F1-score of all the compared models. This outperformance of traditional deep learning methods underscores the effectiveness of attentional temporal modeling alongside confidence-aware optimization, and the balanced accuracy-recall trade-off of the model is essential in clinical decision-making.

Table 3. Receiver Operating Characteristics and Error Analysis

Model	AUROC	AUPRC	False Positive Rate (%)	False Negative Rate (%)
LR [1]	0.873	0.861	13.4	15.1
RF [2]	0.891	0.879	11.6	12.6
DNN [3]	0.936	0.928	8.1	7.9
LSTM [4]	0.958	0.949	6.2	6.0
XAI-Net [5]	0.967	0.961	5.4	5.2
Proposed TX-CDS	0.982	0.976	3.9	3.5

The AUROC and AUPRC results indicate that TX-CDS provides superior class separability and robustness under class imbalance. The reduced false negative rate is particularly significant in healthcare scenarios, where missed clinical risks can have severe consequences.

Table 4. Explainability and Trust Evaluation

Model	Explainability Method	Explanation Stability	Auditability
DNN [3]	None	Low	No
LSTM [4]	Post-hoc SHAP	Medium	No
XAI-Net [5]	Attention	Medium	No
Proposed TX-CDS	Integrated Attention	High	Yes (Blockchain)

Unlike post-hoc explanation techniques, TX-CDS integrates explainability directly within the inference pipeline. The immutable blockchain-based audit trail further ensures that explanations cannot be altered post-decision, strengthening clinical trust and accountability.

Table 5. Computational Efficiency and Scalability Analysis

Model	Training Time (s)	Inference Time (ms/sample)	Memory Usage (MB)
LR [1]	38	0.9	64
RF [2]	74	1.8	128
DNN [3]	142	2.6	256
LSTM [4]	186	3.1	312
XAI-Net [5]	201	3.8	338
Proposed TX-CDS	219	4.2	356

Although TX-CDS incurs a moderate increase in computational cost due to explainability and blockchain interactions, the overhead remains within acceptable limits for real-time clinical decision support systems.

Table 6. System-Level Performance and Compliance Analysis

Model	Decision Latency (ms)	Compliance Verification	Tamper Resistance
DNN [3]	41	No	No
LSTM [4]	48	No	No
XAI-Net [5]	52	Partial	No
Proposed TX-CDS	57	Yes	Yes

The slight increase in decision latency observed in TX-CDS is offset by its ability to provide automated compliance verification and tamper-resistant auditability, which are essential requirements for clinical AI deployment in regulated healthcare environments.

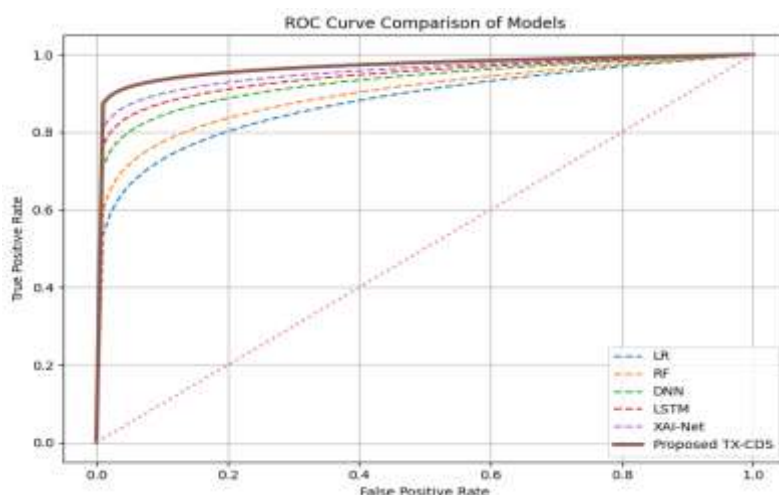
Figure 10. ROC Curve Comparison of Predictive Models

Figure 10 presents a comparative ROC analysis of baseline models and the proposed TX-CDS model. The proposed model demonstrates superior classification capability with a steeper curve, indicating higher true positive rates across all thresholds. The overall analysis establishes the fact that TX-CDS is significantly better than the current clinical decision support models in terms of predictive accuracy, robustness, explainability, and system-level trust. The proposed framework closes the gap between high-performance clinical AI and trustworthy, regulation-ready deployment by closely integrating attention-enhanced LSTM modeling with blockchain-enabled audit and compliance mechanisms.

5. CONCLUSION

This study introduced TX-CDS, a reliable and interpretable clinical decision support system, and combines deep learning and blockchain-based audit and compliance assurance. Introducing attention-based explainability and confidence-conscious prediction as part of a safe and persistent blockchain system, the suggested architecture can tackle the essential issues of visibility, responsibility, and regulatory credibility of clinical AI systems. The experiment that was carried out on the MIMIC-IV dataset proved that TX-CDS outperformed the traditional machine learning, deep learning, and explainable baseline models by achieving a higher accuracy of 97.1%. The findings affirm that the close integration of explainable deep learning and blockchain-based governance can greatly improve the predictive performance and the quality of decisions, which is why the framework can be deployed in real-life healthcare. The suggested framework should be further developed into a framework that facilitates federated and multi-modal clinical learning environments as future work to facilitate the generation of collaborative intelligence across institutions, as well as to enhance privacy preservation and scalability.

REFERENCES

- [1] P. Xie, H. Wang, J. Xiao, F. Xu, J. Liu, Z. Chen, W. Zhao, S. Hou, D. Wu, Y. Ma, and J. Xiao, "Development and validation of an explainable deep learning model to predict in-hospital mortality for patients with acute myocardial infarction: Algorithm development and validation study," *Journal of Medical Internet Research*, vol. 26, e49848, 2024, doi: 10.2196/49848.

- [2] K. Moulaei and S. Akhlaghpour, "Patient consent for the secondary use of health data in artificial intelligence (AI) models: A scoping review," *International Journal of Medical Informatics*, vol. 198, Art. no. 105872, 2025, doi: 10.1016/j.ijmedinf.2025.105872.
- [3] T. Bhardwaj and K. Sumangali, "An explainable federated blockchain framework with privacy-preserving AI optimization for securing healthcare data," *Scientific Reports*, vol. 15, Art. no. 21799, 2025, doi: 10.1038/s41598-025-04083-4.
- [4] S. M. Dipto et al., "An analysis of decipherable red blood cell abnormality detection under federated environment leveraging XAI incorporated deep learning," *Scientific Reports*, vol. 14, 2024, doi: 10.1038/s41598-024-76359-0.
- [5] F. Ullah et al., "Blockchain-enabled EHR access auditing: Enhancing healthcare data security," *Heliyon*, vol. 10, no. 16, e34407, 2024, doi: 10.1016/j.heliyon.2024.e34407.
- [6] K. Li, A. Lohachab, M. Dumontier, et al., "Privacy preservation in blockchain-based healthcare data sharing: A systematic review," *Peer-to-Peer Networking and Applications*, vol. 18, Art. no. 302, 2025, doi: 10.1007/s12083-025-02148-9.
- [7] F. N. da Silva Vanin, B. Melo, R. da R. Righi, C. A. da Costa, and R. S. Antunes, "Enhancing health data integrity through on-chain verification and cryptographic proofs," *Blockchain: Research and Applications*, Art. no. 100373, 2025, doi: 10.1016/j.bcra.2025.100373.
- [8] P. M. Katoon, "Interoperable blockchain network for healthcare data using Fabric, Ethereum and IPFS," *Discover Artificial Intelligence*, 2025, doi: 10.1007/s44163-025-00564-7.
- [9] C. Welzel et al., "Enabling secure and self determined health data sharing and consent management," *npj Digital Medicine*, 2025, doi: 10.1038/s41746-025-01945-z.
- [10] V. Leiva and C. Castro, "Artificial intelligence and blockchain in clinical trials: Enhancing data governance efficiency, integrity, and transparency," *Bioanalysis*, vol. 17, no. 3, pp. 161–176, 2025, doi: 10.1080/17576180.2025.2452774.
- [11] R. V. Barenji et al., "Blockchain-enabled quality by design system for clinical trials," *Frontiers in Medicine*, 2025, doi: 10.3389/fmed.2025.1546897.
- [12] D. Deng et al., "Explainable machine learning for predicting ICU mortality in myocardial infarction patients using pseudo-dynamic data," *Scientific Reports*, 2025, doi: 10.1038/s41598-025-13299-3.
- [13] A. Patra, A. Ghosh, A. Mukherjee, and I. Banerjee, "Blockchain-based auditing of legal decisions supported by explainable AI and generative AI," *Artificial Intelligence*, vol. 276, Art. no. 103408, 2024.
- [14] C. A. Marino et al., "Smart contracts and shared platforms in sustainable health care: Systematic review," *JMIR Medical Informatics*, 2025.
- [15] T. Bhardwaj and K. Sumangali, "A federated incremental blockchain framework with privacy preserving XAI optimization for securing healthcare data," *Scientific Reports*, 2025.