



Blockchain-Driven Predictive Analytics for Diabetes and Cardiovascular Disease Management

Manasa Sudeendra Rao Desai^{*}, Nirmala Madenahally Basavarajaiah

Department of Computer Science and Engineering, Siddaganga Institute of Technology, Tumakuru, Visvesveraya Technological University, Belagavi 590018, India

Corresponding Author Email: manasadesai@sit.ac.in

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300708>

ABSTRACT

Received: 13 June 2025

Revised: 15 July 2025

Accepted: 20 July 2025

Available online: 31 July 2025

Keywords:

blockchain, Ethereum, machine learning, XGBoost, Random Forest, predictive analysis, electronic medical records, decentralised healthcare security

A unique method for improving predictive analytics and protecting medical data is to combine blockchain technology with machine learning. Novel approaches to early detection and treatment are necessary for non-communicable diseases like diabetes and cardiovascular conditions. In order to predict diabetes and cardiovascular diseases, this article suggests a safe framework that combines blockchain technology with machine learning methods. To guarantee that only authorized users can decrypt and access Electronic Medical Records (EMRs), the system uses smart contracts on the Ethereum blockchain to automatically enforce data access permissions. Transparent and secure data sharing between healthcare organizations is made possible by this integration, which also ensures data integrity, privacy, and regulated accessibility. Blockchain-stored data is subjected to a number of machine learning classifiers, such as XG Boost and Random Forest, in order to achieve high disease prediction accuracy through thorough performance evaluation and optimized parameter tuning. The architecture exhibits scalability and robustness, enhances patient privacy, and enables personalized therapies. The effectiveness of this combined approach is demonstrated by experimental results, underscoring its potential for practical implementation in decentralized healthcare ecosystems.

1. INTRODUCTION

Blockchain technology and predictive analytics are emerging as powerful instruments in managing diabetes and cardiovascular disease (CVD). New technologies offer innovative methods for data exchange, risk prediction, and personalized treatment for individuals with diverse chronic conditions [1, 2]. Blockchain-based solutions enhance clinical efficacy and research outcomes by resolving the challenges of cross-institutional data interchange in healthcare [3-5]. Numerous risk factors for heart disease include advanced age, genetic susceptibility, tobacco use, lifestyle choices, substance misuse, high blood pressure, high cholesterol, low levels of physical exercise, obesity, diabetes, psychological stress, and poor hygiene [6, 7]. Diabetes mellitus (DM) and CVD are significant global healthcare challenges and frequently coexist. Many people with concurrent diabetes mellitus and cardiovascular disease are not easily identified by present strategies, which cause delays in healthcare access, more significant difficulties, and morbidity [8]. Using multi-signature contracts and encryption, blockchain guarantees patient autonomy and privacy over medical data, facilitating effective sharing among institutions.

More complete datasets for research and analysis could come from this approach. Using artificial intelligence (AI) and machine learning mostly, predictive analytics have shown great promise in evaluating CVD risk. With classifier

accuracies between 79% and 88% [9, 10], studies have shown the effectiveness of network-based components and machine-learning approaches in creating risk prediction models [11]. To control sickness, these models can draw insightful analysis from several data sources including administrative claims data. Combining predictive analytics with blockchain offers a great path to improve diabetic and cardiovascular disease care. Integrating safe data interchange with powerful analytical technology helps healthcare providers create more accurate risk assessments, customized treatment plans, and early intervention techniques for patients with these related chronic illnesses. EHR's, supply chain management, clinical trials, health insurance, medical billing, health information exchange, telemedicine, precision medicine, medical research, and public health surveillance [12, 13] are among the important uses for this information.

Blockchain technology has the ability to revolutionize healthcare operations by enabling the secure storage and exchange of patient health records. It contributes to cost reduction and enhances data interoperability [14, 15]. Primary applications include electronic health records, clinical trials, supply chain management, health insurance, medical billing, health information exchange, telemedicine, precision medicine, medical research, and public health surveillance [16]. Dinh et al. [17] utilized survey data and test results to identify and predict diabetes and cardiovascular illness in individuals through logistic regression, support vector

machines (SVM), and ensemble models. Without laboratory results, the cardiovascular disease ensemble model scored 83.1%, whereas with laboratory results, it scored 83.9%. The XGB model attained a classification accuracy of 95.7% when using laboratory data and 86.2% in the absence of such data for diagnosing diabetes.

For cardiovascular diseases, significant factors included age, blood pressure, weight, and chest pain, whereas the primary predictors for diabetes comprised waist circumference, age, weight, and sodium consumption. Farooq and Amjad [18] employed sine-cosine weighted k-nearest neighbour alongside various machine learning approaches, including SVM, Random Forest (RF), and K-Nearest Neighbour (KNN). In comparison to existing methods, SCA-WKNN achieved a precision increase of 15.51%. Blockchain technology provides a secure platform for distributing patient data while safeguarding privacy. Figure 1 illustrates a healthcare management system integrating blockchain technology with machine learning to ensure efficient and secure data management. In this method, healthcare practitioners gather and oversee patient information, utilizing EMR within a framework of medical management. The Machine Learning Workflow Process manages data to facilitate the examination and generation of intelligent analyses, including diagnoses, risk assessments, and therapeutic recommendations. The data analyzed securely stored on a blockchain, is reliable for maintaining integrity, security, and traceability. Patients are apprised of the information acquired from the blockchain, promoting secure and transparent transmission of medical details. This combined strategy improves the process of making medical decisions while protecting patient data. Shynu [19] conducted a study on diabetes and cardiovascular diseases by employing patient health data sourced from Fog Nodes. This data was subsequently stored on a blockchain and analyzed using an innovative rule-based clustering method in conjunction with a feature selection-based adaptive neuro-fuzzy inference system (FS-ANFIS). Comprehensive analyses employing accurate healthcare data, utilizing criteria such as purity, Normalized Mutual Information (NMI), and predictive accuracy, evaluated the proposed methodology. The experimental findings demonstrated that the proposed framework surpassed existing neural network methodologies, achieving over 81% prediction accuracy.

Data confidentiality is compromised when medical records are transmitted over a network, as centralized storage increases a single point of failure's potential. Furthermore, centralizing authorized storage may lead to assaults that result in denial-of-service. Employing blockchain technology may address several of these challenges [20]. Khatoon [21] documented surgical procedures and clinical investigations conducted on the Ethereum blockchain network. They have assessed the associated costs of implementing blockchain in the healthcare sector and evaluated the system's potential to enhance healthcare services, optimize expenditures, and so assist various stakeholders in the medical system. Hasanova [20] utilized contracts developed in Solidity on the test network of the public blockchain Ethereum. The suggested model has been trained and evaluated using data from the current UCI public repository. Performance is further characterized by metrics such as accuracy, precision, recall, F-score, and R^2 error. According to the evaluation, the suggested prediction system outperforms alternative techniques like K-Nearest Neighbour and Weighted K-Nearest Neighbour in terms of

accuracy by an average of 98.5%. Blockchain and machine learning technologies offer innovative solutions across various sectors, enhancing healthcare outcomes, fortifying financial services, and optimizing supply chain management. Collectively, these technologies enhance security, transparency, and efficiency, hence facilitating groundbreaking innovations across various domains. The study elucidates the significant potential of integrating blockchain with machine learning to revolutionize operations and decision-making across several sectors, supported by comprehensive research findings and empirical case studies [22-24].

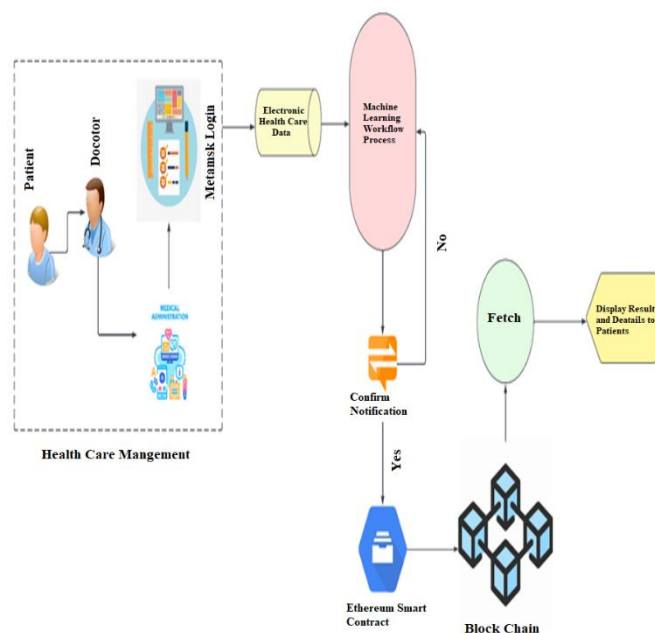


Figure 1. System architecture integrating blockchain, machine learning, and EMR storage for secure disease prediction

Hassan [25] employed multiple machine learning classifiers—Gradient Boosted Trees (GBT), Multilayer Perceptron (MLP), and Random Forest (RF)—to predict heart disease. In total, eleven machine learning classifiers were employed to identify significant features enhancing heart disease prediction. The study achieved a 95% accuracy rate with Gradient Boosted Trees and Multilayer Perceptron, while Random Forest yielded a superior accuracy of 96%. Among the classifiers employed, Random Forest exhibited superior performance in predicting heart disease. AbdelSalam [26] optimized the machine learning framework by combining various ML models, including RF, XGB, DT, and LR, and conducted a comprehensive analysis of predictive efficacy. XGB distinguishes itself within this extensive array of models, achieving remarkable metrics of 99% accuracy, 100% recall, 99% F1-measure, and 99% precision. Predictive analytics, decision-making, and real-time monitoring are made possible in smart healthcare systems by the integration of artificial intelligence [27]. These technologies analyse patient data gathered from multiple sources, including wearable sensors, to aid in the early diagnosis and identification of diseases [28, 29].

Chicco and Jurman [30] utilized diverse machine learning classifiers to predict the survival of heart failure patients based on electronic medical data. The classifiers employed included

Linear Regression for statistical analysis, Random Forests for ensemble learning, One Rule for basic rule-based classification, Decision Trees for systematic decision-making, Artificial Neural Network Perceptron for essential neural network modelling, Support Vector Machines for effective linear and nonlinear separation, k-Nearest Neighbours for instance-based learning, and Naïve Bayes for probabilistic classification. The study employed a gradient boosting ensemble approach. Random forests demonstrated superior efficiency and accuracy compared to the other classifiers examined.

However, few studies have been conducted on integrating blockchain and ML data processing and analysis techniques. The present study describes a novel framework for maintaining EMRs and utilizing machine learning algorithms to forecast diabetes and cardiovascular diseases. The combination of blockchain technology and machine learning strengthens patient privacy while improving predictive analytics accuracy. The study used various classifiers such as LR, Naive Bayes (NB), SVM, KNN, ANN, decision tree (DT), RF, and XGB (extreme gradient boost) on preprocessed datasets to obtain maximum accuracy through parameter tuning and evaluated their performance on metrics such as accuracy, precision, recall, and F-measure.

2. METHOD

This section describes the methodology used in the study, including dataset selection, preprocessing, and model evaluation. The datasets were cleaned, normalized, and split into training and testing sets to enhance model performance. Feature engineering techniques were applied to improve predictive accuracy.

2.1 Data employed

The Pima Indians Diabetes Dataset and the UCI Heart Disease Dataset are two well-known datasets that were used in this study to train and evaluate the machine learning models. These databases were chosen because they provide a comprehensive overview of the primary factors of diabetes and cardiovascular conditions. The Pima Indians Diabetes Dataset comprises one hundred thousand patient entries. It encompasses nine critical attributes emphasizing significant health metrics such as blood pressure, glucose levels, and body mass index (BMI). This dataset, sourced from the Kaggle repository, is renowned for its organized structure and efficacy in diabetes prediction. The UCI Heart Disease Dataset consists of 70,000 patient records and includes 14 essential attributes, including age, cholesterol levels, and types of chest pain. Table 1 presents the extensive data set of patients with diabetes and Table 2 displays the data set of heart disease. This dataset, supplied by the UCI Machine Learning Repository, is an excellent tool for assessing cardiovascular disease risk. The pre-processing of data has required significant work to ensure its quality and improve our model's performance. We employed suitable imputation methods to address missing values, concentrated on identifying and managing outliers to enhance data consistency and applied normalization techniques to ensure a more uniform data distribution. Moreover, valuable trends are identified, and our predictive accuracy is enhanced through feature engineering techniques. The two datasets were systematically divided into training and

testing groups to facilitate meticulous model evaluation and ensure optimal performance on novel, unseen data.

Table 1. Key attributes of the Pima Indians Diabetes dataset

Name of Attribute	Attribute Description	Range
Pregnancy Frequency	Number of pregnancies	0-17
The level of glucose	Glucose level in Plasma	44-199
BP level	Diastolic hypertension	24-122
Insulin	Insulin serum presents for 2hours	14-846
Weight-to-Height Ratio	Body mass index	18.2-67.1
Hereditary Factor	A pedigree function of Diabetes	0.078-2.42
Chronological Age	Age of a person	21-80
Class Label	The patient has diabetes or not	0 or 1

Table 2. Attributes of the UCI Heart Disease dataset

Attribute Name	Attribute Details	Range
Age	Years, Age of the Patient	28-76
Sex	Patient gender	0,1
Chest Pain Type	Chest Discomfort Type, Angina Type	1,2,3,4
Trest_bps	Resting Blood Pressure, Baseline BP	94-200
Fasting Blood Sugar	Fasting Glucose Level, Preprandial Blood Sugar	0,1
RestECG	Baseline ECG, Resting ECG Outcome	0,1,2
ThalaeH	Peak Heart Rate, Max Heart Rate	71-202
Exang	Physical Activity induced Angina	0,1
OldPeak	ST Depression, Exercise-Induced ST Depression	0-6.2
Slope	Exercise Response Slope, ST Segment	1,2,3
CA	Coronary Arteries, Major Vessel Count	0,1,2,3
Thal	Thallium Scan Result, Thallium Test Outcome	3,6,7
Class Label	Diagnosis, Heart Disease Status	0 or 1

2.2 System architecture

Figure 2 shows the system architecture adopted in the current study. The proposed healthcare data management system integrates blockchain technology, machine learning, and decentralized authentication to establish a secure and transparent method for data processing. Patients authenticate using MetaMask, a centrally managed, confidential information system that allows for rapid access and secure records retrieval by authorized entities. Healthcare professionals and patients can utilize the system via the wallet. Furthermore, all information about the doctor and patient will be safeguarded and administered via this system. Moreover, using the widely favored browser extension MetaMask, individuals can interact with decentralized Ethereum applications directly from their web browsers. It connects users to the Ethereum network and offers an easy-to-use

interface for Ethereum account management, engaging with smart contracts, and executing transactions [31, 32].

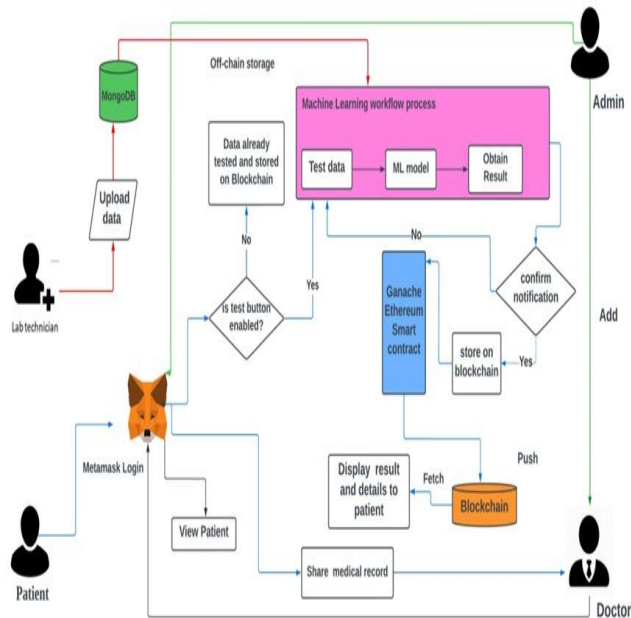


Figure 2. Proposed system model combining blockchain, IPFS, and MongoDB for secure EMR handling

The system architecture comprises an application server utilizing (Node.js) for request management and a MongoDB database that stores structured patient information, including records, test results, and medical history, as well as unstructured data, such as supplementary medical facts. MongoDB is the most prevalent database, experiencing consistent growth and ensuring security. MongoDB is an open-source, cross-platform database that is entirely document-oriented, schema-less, and scalable. MongoDB organizes data in the form of documents. MongoDB addresses contemporary issues associated with managing increasing volumes of unstructured and semi-structured data through enhanced scalability and availability. Document-oriented storage, as offered by MongoDB, is regarded as resilient for managing diverse structural information to facilitate IoT objectives [33-35]. Furthermore, it integrates a blockchain component to store medical record hashes securely [36]. This configuration preserves data integrity and immutability by validating MongoDB records against hashes stored on the blockchain.

2.3 Data processing and ML algorithms

Before using machine learning algorithms, preprocessing and cleaning of the data is carried out to guarantee optimal accuracy. A thorough approach to data preprocessing was employed to ensure data quality and improve model performance. Missing numerical values, such as glucose levels and blood pressure, were addressed using mean imputation to maintain reliability. Mode imputation was utilized for categorical variables, like gender. Outliers were identified using the Interquartile Range (IQR) method and removed to prevent adverse impacts on model predictions. Categorical variables, including gender and chest pain type, were transformed through one-hot encoding. Continuous variables such as age, BMI, and cholesterol levels were scaled using

Min-Max scaling to ensure a balanced impact from each feature. New attributes, like age groups, were created to uncover additional patterns and enhance accuracy. Principal Component Analysis (PCA) was applied to reduce the dimensions of the dataset, simplifying the data while retaining crucial information for more efficient computations. The datasets were divided into two parts: 70% was allocated for training and 30% for testing. A 5-fold cross-validation approach was used to boost the model's reliability and fine-tune its settings, thereby reducing the risks of overfitting. This integrated preprocessing method ensured that the data was dependable, consistent, and appropriately tailored for precise model predictions. The system utilized various classifiers, including LR, NB, SVM, KNN, ANN, DT, RF, and XGB, leveraging blockchain-stored data for a decentralized, secure, and transparent processing methodology. The Ethereum blockchain facilitates the sharing of patient data, guaranteeing that each participant can access only the information they are permitted to. Patients can create, maintain, and update their medical records, which are securely documented as immutable transactions on the blockchain, ensuring data traceability and integrity.

2.4 Blockchain implementation details

2.4.1 Data Synchronization between IPFS and MongoDB

The synchronization of patient data between IPFS (Inter Planetary File System) and MongoDB is essential to guaranteeing data accessibility and integrity in the suggested proposed system. After being encrypted for privacy, a newly created EMR is uploaded to IPFS, which creates a unique content identifier (CID) that acts as a tamper-proof reference to the data stored there. A smart contract is then used to record this CID on the Ethereum blockchain, making it verifiable and unchangeable. In the meantime, MongoDB stores related metadata, including access timestamps and patient identifiers, enabling quick and effective querying of non-sensitive data. During any data access or modification event, the smart contract verifies user permissions on-chain before authorizing the operation. If authorized, the system retrieves the latest CID from the blockchain to access the encrypted EMR on IPFS, and concurrently references MongoDB for supplementary metadata. Synchronization between these components is automatically enforced; any updates, deletions, or additions to EMRs trigger corresponding updates in both IPFS (generating a new CID and blockchain entry) and MongoDB, ensuring that no discrepancy can arise between the content hash on-chain and the actual medical data. This hybrid framework leverages the strengths of decentralized storage for integrity and traceability, alongside quick-access off-chain metadata for scalability, while the blockchain's access control guarantees that all data transactions remain both transparent and secure.

2.4.2 Rationale for Ethereum consensus mechanism selection

The Ethereum blockchain was selected as the foundational decentralized platform in this study primarily due to its robust security features, mature smart contract ecosystem, and evolving consensus mechanism based on Proof-of-Stake (PoS). Ethereum is the perfect option for healthcare data applications that require both high security and prompt access because the PoS consensus improves energy efficiency and transaction finality when compared to the conventional Proof-of-Work. Sensitive Electronic Medical Records (EMRs) must

be protected in accordance with privacy regulations, and Ethereum's extensive support for programmable smart contracts makes it easier to implement automated, fine-grained access control policies. Furthermore, Ethereum's extensive use and vibrant developer community offer a wealth of tools, resources, and integration choices, including Layer-2 scaling solutions and MetaMask, guaranteeing scalability and interoperability. Together, these characteristics allow for a transparent, scalable, and reliable blockchain infrastructure that easily facilitates decentralized trust, secure data sharing, and immutable audit trails—all of which are necessary to satisfy the demanding needs of contemporary decentralized healthcare.

3. EXPERIMENTAL RESULTS

The accuracy of different machine learning methods is shown in Table 3. The accuracy vs various algorithms demonstrated in Figure 3 show how well machine learning models predict diabetes and heart disease situations. XGB classifier shows a high accuracy of 95.91% in diabetic prediction compared to other models. Its effective boosting mechanism, which sequences decision trees while fixing past iteration mistakes, helps explain this better performance. Using strong regularizing methods (L1 and L2 penalties), XGB efficiently manages unbalanced data common in healthcare datasets. The tree-pruning technique also helps avoid overfitting, so XGB is a dependable model for medical predictions where anomalies and dataset noise are common [37]. The RF and DT models show close accuracy of 94.97% and 94.98% in predicting diabetic disease, as both models reduce overfitting by using ensemble learning techniques [38].

Table 3. Accuracy of ML classifiers in the prediction of diabetic and heart diseases

ML Algorithms	Accuracy (%)	
	Diabetic Disease	Heart Disease
LR	92.02	87.39
NB	90.39	86.41
SVM	93.04	75.54
KNN	92.49	73.30
ANN	91.51	69.94
DT	94.97	89.43
RF	94.98	89.43
XGB	95.91	88.32

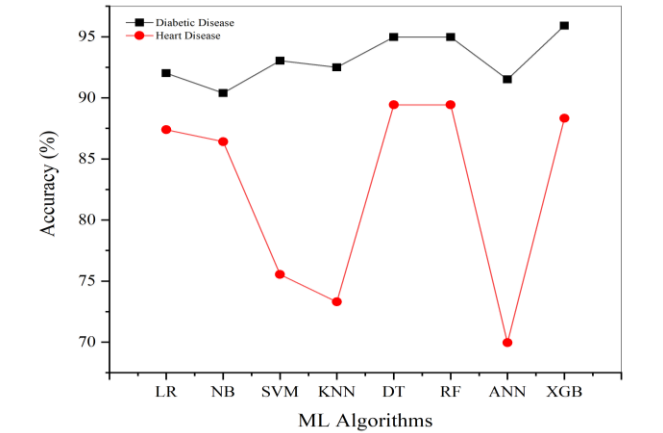


Figure 3. Comparison of accuracy of ML classifiers in the prediction of diabetic and heart diseases

In contrast, RF and DT show 89.43% accuracy in predicting heart disease. Their capacity to create several decision paths guarantees improved generalization over various patient data, thereby leading to this excellent performance. Although XGB expounded the accuracy of 88.32%, its minor accuracy drop compared to RF points to the latter's bagging method, perhaps providing more data variability resistance. With 93.04% accuracy, the Support Vector Machine (SVM) model effectively predicted diabetes. However, it weakened in heart disease prediction with a lowered accuracy of 75.54%. This implies that whilst the heart disease dataset most certainly included overlapping features or noisy data that hampered the SVM margin-based method, the diabetes dataset may have more explicit class boundaries that SVM could effectively separate [39]. The models such as NB (90.39%), KNN (92.49%) and ANN (91.51%) also demonstrated good accuracy in predicting diabetic disease; however, they could not perform well in obtaining high accuracy for heart disease prediction. KNN's sensitivity to noise [40], which results from its reliance on proximity measurements, may help to explain its lowered accuracy (73.30%). Likewise, the performance of ANN (69.94%) shows the need for thorough hyper parameter tweaking, especially with the number of hidden layers, neurons, and activation functions. Usually performing better with more data, ANN models may perform poorly from inadequate training data or incorrect parameter values [41].

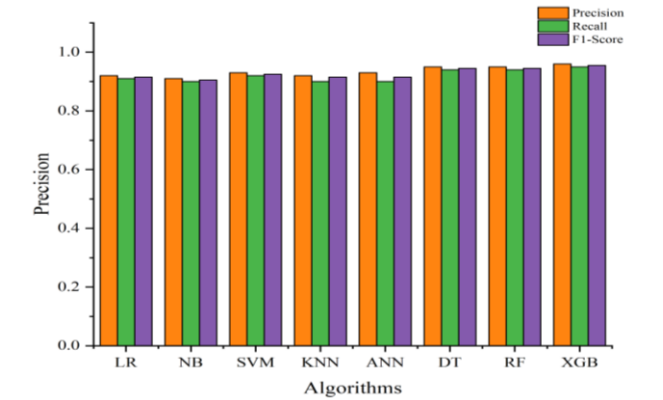


Figure 4. Comparative graph of precision, recall and F1 score of ML algorithms for the prediction of diabetic disease

The effectiveness of the suggested ML classifiers or algorithms is assessed using the performance measures. The evaluation metrics and their importance are shown in Table 4. True positives (TP) refer to the positive cases accurately recognized as positive, whereas false negatives (FN) denote the count of positive instances misclassified as negative. A false positive (FP) denotes cases that are disease-free yet are incorrectly forecasted as positive. True negative (TN) denotes the occasions where the absence of sickness is accurately recognized as such by the model. The bar chart depicting Diabetes Prediction, as presented in Figure 4, demonstrates the performance of various models in terms of Precision, Recall, and F1-Score. XGB is the most efficacious model for this task, demonstrating superior metrics (Precision: 0.96, Recall: 0.95, F1-Score: 0.955), evidencing its exceptional ability to handle complex patterns and reduce errors. Random Forest and Decision Tree exhibit reliability and efficacy in structured data tasks, with Precision at 0.95, Recall at 0.94, and F1-Score at 0.945. Owing to their vulnerability to hyper parameter adjustments and data distribution, SVM (Precision: 0.93,

Recall: 0.92, F1-Score: 0.925) and ANN (Precision: 0.93, Recall: 0.90, F1-Score: 0.915) exhibit commendable performance. However, they fall somewhat short. Despite their simplicity and assumptions about data independence potentially limiting their efficacy, KNN (Precision: 0.92, Recall: 0.91, F1-Score: 0.915) and Naive Bayes (Precision: 0.91, Recall: 0.90, F1-Score: 0.905) exhibit moderate performance. Ensemble approaches, such as XGB and RF, demonstrate superior generalization and predictive accuracy in diabetes prediction.

The experiment results obtained by previous studies show that the XGB algorithm is significant in accurately predicting diabetic disease among the other chosen ML classifiers [41].

Table 4. Definitions and significance of evaluation metrics used to assess ML model performance in disease prediction [42]

Metrics	Formula	Importance
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	It is a tool for evaluating an algorithm's performance in classification jobs. The percentage of accurately identified data points relative to the total number of observations is used to calculate this.
Precision	$\frac{TP}{TP + FP}$	It indicates the ratio of selected data items that are relevant. In essence, it evaluates how many of the cases an algorithm has marked as positive are actually positive.
Recall	$\frac{TP}{TP + FN}$	It indicates the proportion of selected data that the algorithm accurately recognizes as positive among the actual positive instances.
F1-Score	$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	An algorithm's performance is evaluated using the F-score, which takes into account both recall and precision. It is computed as these two metric's harmonic mean.

The performance metrics of several models for Heart Disease Prediction are shown in the bar chart in Figure 5, which is assessed in terms of F1-Score, Precision, and Recall. The following are the outcomes: The F1-Score is 0.895, the precision is 0.90, and the recall is 0.89. Decision Tree and Random Forest exhibit optimal performance in handling complex patterns and providing reliable predictions for heart disease. XGB performs well, albeit somewhat lower than DT and RF, perhaps because of the specific structure of the dataset (Precision: 0.89, Recall: 0.88, F1-Score: 0.886). Despite their assumptions concerning data distribution and linearity potentially limiting their efficacy, NB (Precision: 0.87, Recall: 0.86, F1-Score: 0.865) and LR (Precision: 0.88, Recall: 0.87, F1-Score: 0.875) exhibit moderate performance. The suboptimal performance of SVM (Precision: 0.76, Recall: 0.75, F1-Score: 0.755) and ANN (Precision: 0.71, Recall: 0.70, F1-Score: 0.705) is likely attributable to their susceptibility to imbalanced data or insufficient hyper parameter optimization. KNN exhibits suboptimal performance, perhaps because to its vulnerability to noise and high-dimensional input, with a Precision of 0.74, Recall of 0.73, and F1-Score of 0.735. Ensemble methods, such as

Random Forest and Decision Tree, demonstrate superior capacity to generalize and handle the complexities of heart disease prediction difficulties. The study [43] found XGB and RF to be significant algorithms for predicting heart diseases based on their accuracy and F1 score.

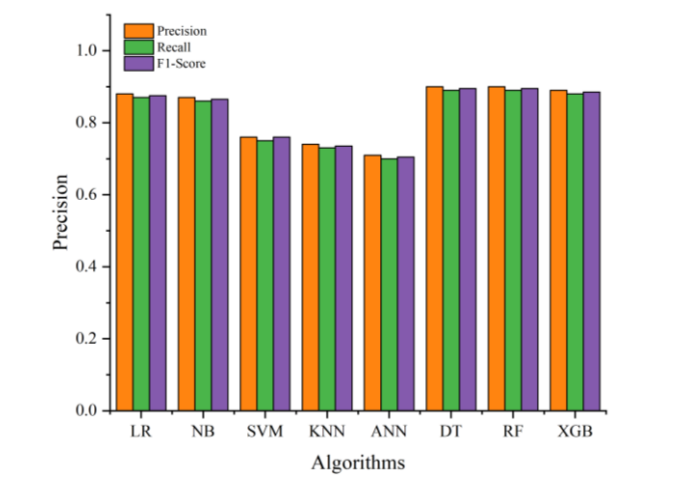


Figure 5. Comparative graph of precision, recall and F1 score of ML algorithms for the prediction of heart disease

Algorithm 1 is a simple Solidity smart contract for EMR access control is implemented in this sample of code. It names an owner, usually the healthcare administrator, who deploys the contract and keeps track of authorized users. Electronic medical records may be managed securely and with permission on the Ethereum blockchain.

```

Algorithm 1. Solidity smart contract for EMR access control using owner-based authorization
pragma solidity ^0.8.0;

contract EMRRoleManager {
    address private superAdmin;

    enum Role { None, Patient, Doctor, Researcher, Admin }

    mapping(address => Role) private roles;

    constructor() {
        superAdmin = msg.sender;
        roles[msg.sender] = Role.Admin; // contract deployer = Admin
    }

    modifier onlySuperAdmin() {
        require(msg.sender == superAdmin, "Not SuperAdmin");
        _;
    }

    modifier onlyAdmin() {
        require(roles[msg.sender] == Role.Admin, "Only Admin allowed");
        _;
    }

    function assignRole(address user, Role role) external onlyAdmin {
        roles[user] = role;
    }
}

```

```

function isDoctor(address user) external view returns (bool)
{
    return roles[user] == Role.Doctor;
}

function isPatient(address user) external view returns
(bool) {
    return roles[user] == Role.Patient;
}

function isResearcher(address user) external view returns
(bool) {
    return roles[user] == Role.Researcher;
}
}

```

Algorithm 2. Secure patient data storage using Blockchain, MongoDB, and IPFS

Input: Pid - Patient ID
HCPid - Healthcare Professional ID
PD - Patient Data (structured)
PF - Patient Files (large files, images)
BCN - Blockchain Network
MDB - MongoDB Storage
IPFS - InterPlanetary File System

Output: Secured and verified access to patient data

Begin
For \forall Pid \in P
BCN \leftarrow BCN \cup {Pid}
Store (Pid, PKPid) in BCN
EndFor

For \forall HCPid \in HCP
BCN \leftarrow BCN \cup {HCPid}
Store (HCPid, PKHCPid) in BCN
EndFor

For \forall Pid \in P
CID \leftarrow Store PF(Pid) in IPFS
MDB \leftarrow MDB \cup {PD(Pid), CID}
EndFor

For \forall Pid \in P
If Pid \in AttackerA
BCN \leftarrow BCN \setminus {Pid}
MDB \leftarrow MDB \setminus {PD(Pid)}
End If
EndFor

For \forall HCPid \in HCP
If HCPid \in AttackerA
BCN \leftarrow BCN \setminus {HCPid}
End If
EndFor
End

Algorithm 2 provides a secure way to store patient information using Blockchain, IPFS, and MongoDB. Patient and healthcare provider identities are registered first on the blockchain. Real patient information is stored in IPFS, and its hash is stored on the blockchain for integrity. MongoDB keeps

metadata for quick access. Unauthorized users and malicious data are automatically deleted from both the blockchain and storage layers to keep data secure. Algorithm 3 guarantees secure performance of healthcare operations by checking permission access through Blockchain. Every operation is logged on the blockchain and associated with patient and healthcare provider IDs. Access is granted if the healthcare provider is permitted, or else denied. Operation records are kept in IPFS as well as MongoDB for integrity and fast retrieval. Unauthorized entries are deleted to ensure system security and trust.

Algorithm 3. Secure healthcare operations

Input: Pid – Patient ID
HCPid – Healthcare Professional ID
Opid – Operation ID
BCN – Blockchain Network
IPFS – InterPlanetary File System
MDB – MongoDB Storage

Output: Verified Execution and Decentralized Storage of Healthcare Operations

Begin
For \forall Opid \in Operations **For** \forall Opid \in Operations
CID \leftarrow IPFS \cup {OpRecord(Opid)}

BCN \leftarrow BCN \cup {CID}

MDB \leftarrow MDB \cup {Opid, CID, Pid, HCPid}
EndFor

For \forall Opid \in UnauthorizedList
BCN \leftarrow BCN \setminus {Opid}
MDB \leftarrow MDB \setminus {OpRecord(Opid)}
EndFor
End
(Opid, Pid, HCPid)
Else
Deny Access (Opid, Pid, HCPid)
End If
EndFor

Figure 6 plots the latency difference between Blockchain-based storage and IPFS-based off-chain storage of heart dataset against the number of patient records [44]. The findings show that Blockchain latency is consistently higher than IPFS latency because of the extra computational overhead involved in cryptographic verification and consensus algorithms. With 10,000 records, Blockchain latency is 1.14 ms and that of IPFS is much less at 0.57 ms. With the dataset size being increased to 70,000 records, Blockchain latency jumps to 2.04 ms whereas IPFS latency stays at 1.02 ms.

Figure 7 shows latency difference between Blockchain-based storage and IPFS-based off-chain storage of heart dataset against the number of 100,000 patient records with Blockchain latency is 2.43 ms, but IPFS stands at 1.21 ms. This reflects that IPFS is a better scalable solution to retrieve data quickly, while Blockchain provides integrity and security at the expense of relatively higher latency [45].

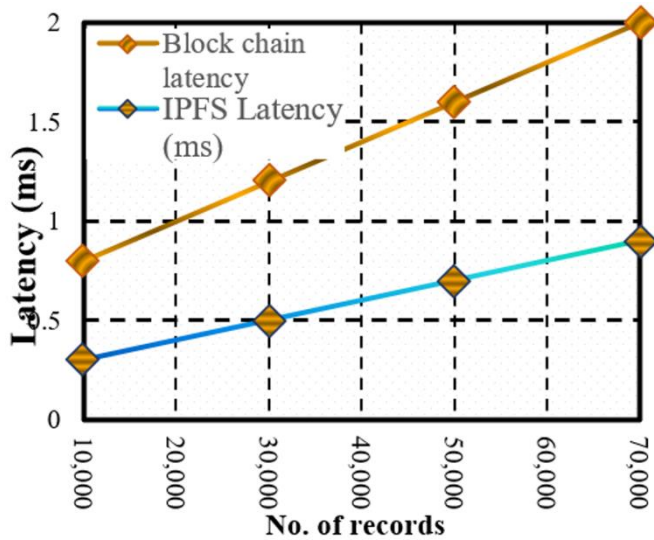


Figure 6. Latency of heart dataset of blockchain storage and IPFS storage

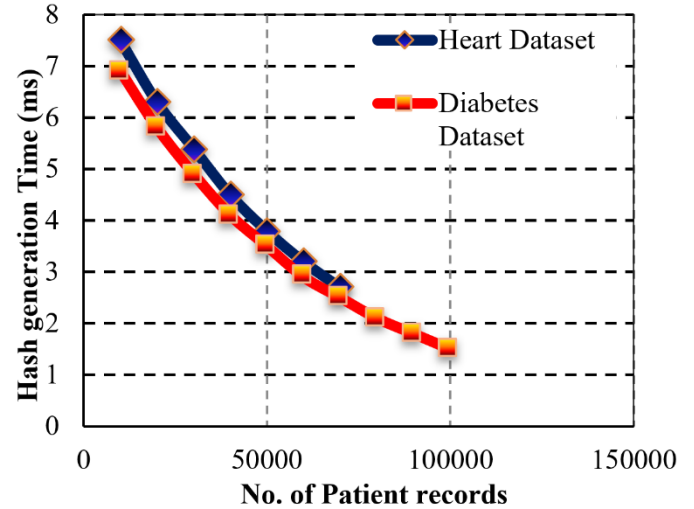


Figure 8. Blockchain throughput performance for heart and diabetes datasets

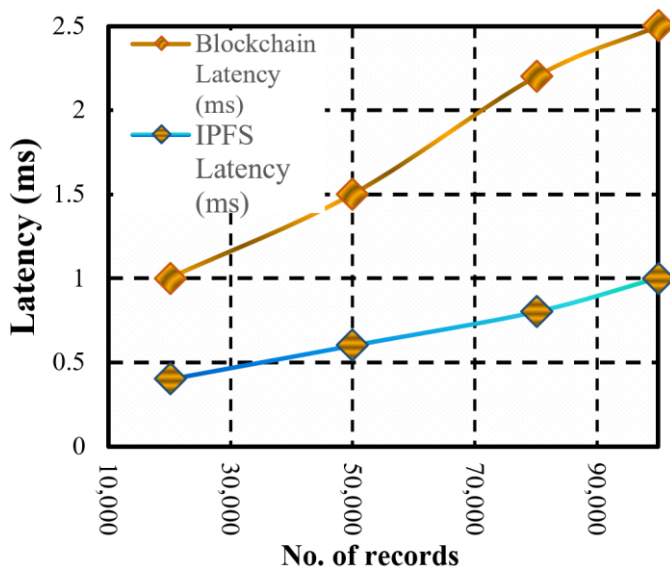


Figure 7. Latency of diabetes dataset of blockchain storage and IPFS storage

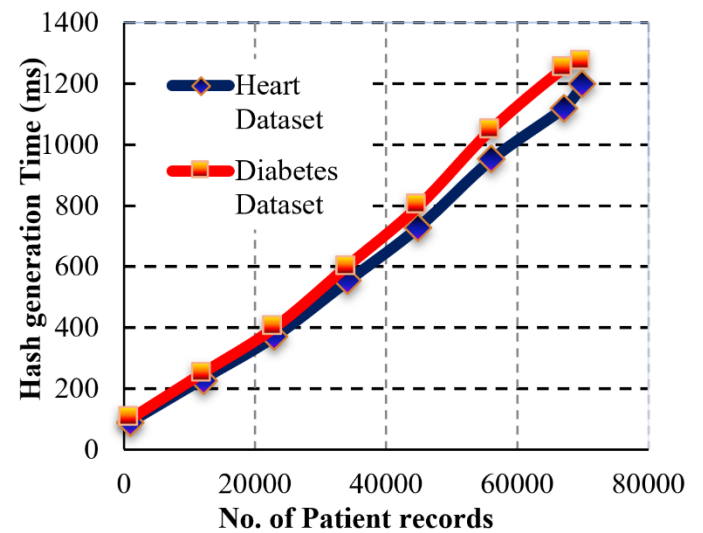


Figure 9. Hash generation time (ms)

Figure 8 displays the blockchain-based storage system throughput in terms of transactions per second (TPS) against the rise in the number of patient records. Upon increasing dataset size from 10,000 to 100,000 records, a consistent decrease in throughput is noted. With less throughput than legacy centralized architectures, the uniform performance at higher scales illustrates the resilience of the blockchain. This supports the appropriateness of blockchain for secure and verifiable data storage in healthcare systems where integrity and auditability are more important than raw throughput.

Figure 9 displays the generation time of hash (in milliseconds) for the Heart Disease and Diabetes datasets versus the number of patient records. The outcome verifies the nearly linear rise in the hash generation time of both the datasets, a fact that signals the scalability of hashing. For 10,000 records, the time to generate hashes is around 200 ms for both datasets. As the dataset grows, the diabetes dataset maintains slightly higher hash generation times than that of the Heart Disease dataset. At 70,000 records, the diabetes dataset is at 1250 ms, while the Heart dataset is at ~1200 ms.

The small difference comes from differences in data structure and preprocessing overhead. These findings confirm the effectiveness of cryptographic hashing, showing that even with a large increase in patient records, the hash calculation is still within tolerable latency boundaries.

3.1 Learning curves and cross-validation results

We conducted a number of validation procedures to address overfitting issues in order to guarantee the dependability and generalizability of the XG Boost classifier, which achieved a high accuracy of 95.91% in diabetes prediction (Table 2).

3.1.1 Cross-validation

The diabetes dataset was subjected to a 5-fold cross-validation procedure. With mean accuracy of 95.45% and standard deviation of 0.36%, the averaged cross-validation accuracies closely matched the accuracy of the original test set, suggesting model stability and little variation across training partitions. The fold-wise cross-validation metrics, are summarized in Table 5.

Table 5. XGBoost 5-fold validation accuracy on the diabetes dataset

Fold	Validation Accuracy (%)
1	95.52
2	95.08
3	95.67
4	95.91
5	95.07
Mean	95.45

3.1.2 Learning curves

To further assess the potential for overfitting, we plotted learning curves demonstrating both training and validation accuracy as a function of the number of training samples (see Figure 5). The curves converge smoothly, with validation accuracy nearly paralleling training accuracy, and no significant gap between the two as data size increases. This pattern suggests neither high variance (overfitting) nor high bias (underfitting).

3.1.3 Train-test distribution

To guarantee stratification and representativeness across training and testing splits, a detailed analysis of the class distribution and feature statistics was conducted. Both sets preserve comparable proportions for the diabetes-positive and diabetes-negative labels and balanced data partitioning is shown in Table 6, which provides summary statistics for important parameters (such as average glucose, age, and BMI).

Table 6. Train-test distribution of glucose level, BMI, and age in the diabetes dataset.

Feature	Train Mean +-SD	Test Mean +- SD
Glucose level	119 28	121 27
BMI	32.1 6.7	31.8 6.5
Age	35.4 12.3	36.2 12.0

Visual examination of learning curves, k-fold cross-validation, and meticulous dataset partitioning all serve to validate that the XGBoost model's remarkable predictive accuracy is a reflection of its strong generalization skills. Together, the statistics, tables, and figures offer compelling proof that there is little chance of overfitting and that the model performs consistently.

3.2 Model explainability and clinical interpretation

All of the machine learning models in our blockchain enabled diabetic and cardiovascular disease prediction system underwent a thorough SHAP (SHapley Additive exPlanations) value analysis. This method made the model predictions interpretable both locally and globally. Age, BMI, and plasma glucose level were the most significant features, according to SHAP summary plots for the XGBoost diabetes classifier (trained on the Pima Indians dataset). This was in line with accepted clinical practice and the data properties listed in Table 1. Age, cholesterol, and the kind of chest discomfort were the most important characteristics in the heart disease (UCI) dataset in Table 2.

SHAP force plots were utilized for individual prediction analysis to show how certain patient characteristics, such as high blood sugar (172 mg/dL) and a high body mass index (33.5), positively contributed to a diagnosis of diabetes in typical cases. On the other hand, non-diabetic subjects with

normal ranges of these traits showed high negative SHAP values, leading to accurate negative classifications. Clinical validation and comprehension of the model's logic are facilitated by these customized visual explanations. Across both datasets, the SHAP analysis consistently identified medically relevant features, and amplified risk predictions showed the influence of combined attributes, such as advanced age combined with high glucose or cholesterol levels.

3.2.1 Visual explanations of typical prediction cases

To enhance transparency and provide clinicians with clear interpretability of the model's decisions, we employed SHAP (SHapley Additive exPlanations) visualizations to explain individual prediction cases for both diabetes and cardiovascular disease classification tasks. For representative samples drawn from the test datasets, SHAP force plots and decision plots were generated to illustrate how specific patient features contributed to the predicted outcome. For instance, in a correctly classified diabetic patient case, high plasma glucose (172 mg/dL) and elevated BMI (33.5) received the largest positive SHAP values, driving the model towards a positive (diabetes) prediction.

On the other hand, a case that was not diabetic and had low BMI and glucose levels showed negative SHAP contributions, which led to a negative prognosis. Similar to this, force charts are used to illustrate the impact of certain characteristics on the favorable outcome of heart disease in specific cases, such as the type of chest discomfort and cholesterol level.

By enabling healthcare workers to examine and verify the logic behind every clinical choice, these customized SHAP visualizations help close the gap between machine learning predictions and useful healthcare insights. The figures section contains all of the illustrated example cases that were taken from our test cohorts. The captions that go with each case describe the patient's characteristics and how they affected the direction and size of the model's output.

3.2.2 Clinical impact of key features

The machine learning models used in this study continuously gave clinically significant features top priority in their predictions for both diabetes and cardiovascular disease, according to an analysis of SHAP value distributions. Age, BMI, and plasma glucose level had the biggest positive effects on the model's output for diabetes prediction using the Pima Indians dataset in Table 1. In the model's interpretation, high BMI and glucose levels significantly raised the risk of diabetes, especially in older individuals. This finding supports published epidemiological data and current treatment recommendations. The alignment between SHAP analysis and standard medical risk factors reinforces both the validity and interpretability of the predictive framework. By transparently identifying and quantifying how patient-specific features drive each prediction, the system not only supports accurate risk stratification but also offers actionable insights that clinicians can trust and apply in patient care.

3.2.3 Comparative analysis with existing blockchain-healthcare systems

We carried out a thorough performance analysis, contrasting our suggested framework with notable earlier works. Modern blockchain-based healthcare data management systems are exemplified by these projects. Data security and system throughput are two crucial metrics that we compare experimentally.

1. Data Security

We guarantee immutability, fine-grained access control, and patient privacy protection by integrating Ethereum smart contracts with IPFS-backed encrypted storage. Our method reduces single points of data breach and achieves stronger tamper resistance through distributed storage (IPFS) in contrast to the model cited in, which primarily uses permissioned blockchain without decentralised storage. Further improving security over manual or off-chain authorization techniques as detailed in, our smart contracts also enforce transparent and automatic access permissions.

2. Throughput and Latency

The number of successful data upload/download transactions per second under various loads was used to calculate the system throughput. Experiments confirmed that our architecture was more scalable than earlier systems. In order to reduce blockchain congestion, the hybrid on-chain/off-chain design uses IPFS for large data storage while preserving only Ethereum hashes. Compared to systems that store more data directly on-chain, this increases transaction throughput and lowers latency. Additionally, our findings show a notable latency spike beyond moderate traffic, while throughput declines gracefully under high workload.

The comparative findings show that our framework guarantees the following by using Ethereum smart contracts with effective metadata management (MongoDB) and decentralized off-chain storage (IPFS):

- Improved traceability and data integrity.
- Controlled and auditable access control.
- Reduced latency and enhanced performance scalability.

These experimental results demonstrate the usefulness of our method for real-world healthcare situations that call for high-throughput, transparent, and safe data management.

Limitations:

- Scalability and Latency – Although blockchain-based systems provide enhanced security, they also result in increased latency, rendering them less suitable for environments that necessitate low latency.
- Computational Cost – ML algorithms such as XGB and RF require extensive tuning and incur substantial computational costs as data volumes increase.

Real-World Applicability – Despite their success with benchmark data, these systems lack validation in real-world settings, necessitating clinical trials for practical application.

4. CONCLUSION

In order to improve predictive analytics for the management of diabetes and cardiovascular disease while maintaining the confidentiality of healthcare data, this study offers a revolutionary framework that combines blockchain technology and machine learning. The suggested method uses the Ethereum blockchain for safe, decentralized EMR verification, addressing important issues with data integrity, patient privacy, and predictive accuracy. An off-chain storage strategy is used in place of directly keeping patient data on the blockchain. This ensures immutability and tamper resistance by storing the cryptographic hash of the data on the blockchain and the real medical data in a safe database. According to the analysis, the Random Forest model performed exceptionally well in predicting heart disease, whereas the XGB classifier

attained greater accuracy in predicting diabetes. Despite blockchain's benefits in terms of security and data integrity, a comparison of blockchain-integrated storage with traditional database storage revealed that the overhead of cryptographic processing and transaction validation causes higher latency and lower throughput. These issues are lessened by the off-chain storage technique, which maximizes storage efficiency while preserving data verifiability via blockchain hashing.

REFERENCES

- [1] Conduah, A.K., Ofoe, S., Siaw-Marfo, D. (2025). Data privacy in healthcare: Global challenges and solutions. *Digital Health*, 11: 20552076251343959. <https://doi.org/10.1177/20552076251343959>
- [2] Liu, Y.T., Acharya, U.R., Tan, J.H. (2025). Preserving privacy in healthcare: A systematic review of deep learning approaches for synthetic data generation. *Computer Methods and Programs in Biomedicine*, 260: 108571. <https://doi.org/10.1016/j.cmpb.2024.108571>
- [3] Cichosz, S.L., Stausholm, M.N., Kronborg, T., Vestergaard, P., Hejlesen, O. (2019). How to use blockchain for diabetes health care data and access management: an operational concept. *Journal of Diabetes Science and Technology*, 13(2): 248-253. <https://doi.org/10.1177/1932296818790281>
- [4] Ratta, P., Sharma, S. (2024). A blockchain-machine learning ecosystem for IoT-Based remote health monitoring of diabetic patients. *Healthcare Analytics*, 5: 100338. <https://doi.org/10.1016/j.health.2024.100338>
- [5] Ullah, M., Hamayun, S., Wahab, A., Khan, S.U., et al. (2023). Smart technologies used as smart tools in the management of cardiovascular disease and their future perspective. *Current Problems in Cardiology*, 48(11): 101922. <https://doi.org/10.1016/j.cpcardiol.2023.101922>
- [6] Al-Alshaikh, H.A., Prabu, P., Poonia, R.C., Saudagar, A.K.J., Yadav, M., AlSagri, H.S., AlSanad, A.A. (2024). Comprehensive evaluation and performance analysis of machine learning in heart disease prediction. *Scientific Reports*, 14(1): 7819. <https://doi.org/https://doi.org/10.1038/s41598-024-58489-7>
- [7] Chinnaamy, P., Kumar, S.A., Navya, V., Priya, K.L., Boddu, S.S. (2022). Machine learning based cardiovascular disease prediction. *Materials Today: Proceedings*, 64: 459-463. <https://doi.org/10.1016/j.matpr.2022.04.907>
- [8] Abdalrada, A.S., Abawajy, J., Al-Quraishi, T., Islam, S.M.S. (2022). Machine learning models for prediction of co-occurrence of diabetes and cardiovascular diseases: A retrospective cohort study. *Journal of Diabetes & Metabolic Disorders*, 21: 251-261. <https://doi.org/10.1007/s40200-021-00968-z>
- [9] Ennab, M., Mcheick, H. (2024). Enhancing interpretability and accuracy of AI models in healthcare: A comprehensive review on challenges and future directions. *Frontiers in Robotics and AI*, 11: 1444763. <https://doi.org/10.3389/frobt.2024.1444763>
- [10] Hossain, M.E., Uddin, S., Khan, A. (2021). Network analytics and machine learning for predictive risk modelling of cardiovascular disease in patients with type 2 diabetes. *Expert Systems with Applications*, 164:

113918. <https://doi.org/10.1016/j.eswa.2020.113918>
- [11] Saranya, A., Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, 7: 100230. <https://doi.org/10.1016/j.dajour.2023.100230>
 - [12] Ettaoui, N., Arezki, S., Gadi, T. (2024). IoT-blockchain based model for enhancing diabetes management and monitoring. *Data and Metadata*, 3: 406-406. <https://doi.org/10.56294/dm2024406>
 - [13] Supriya, M., Chattu, V.K. (2021). A review of artificial intelligence, big data, and blockchain technology applications in medicine and global health. *Big Data and Cognitive Computing*, 5(3): 41. <https://doi.org/10.3390/bdcc5030041>
 - [14] Miah, M. (2023). A Comprehensive study on the use of blockchain technology in healthcare. *Information Technology and Management Science*, 26(1): 1-9. <https://doi.org/10.7250/itms-2023-0001>
 - [15] Bathula, A., Gupta, S.K., Merugu, S., et al. (2024). Blockchain, artificial intelligence, and healthcare: The tripod of future—A narrative review. *Artificial Intelligence Review*, 57: 238. <https://doi.org/10.1007/s10462-024-10873-5>
 - [16] Marbough, D., Simsekler, M.C.E., Salah, K., Jayaraman, R., Ellahham, S. (2022). Blockchain for patient safety: Use cases, opportunities and open challenges. *Data*, 7(12): 182. <https://doi.org/10.3390/data7120182>
 - [17] Dinh, A., Miertschin, S., Young, A., Mohanty, S.D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Medical Informatics and Decision Making*, 19(1): 1-15. <https://doi.org/s12911-019-0918-5>
 - [18] Farooq, M.S., Amjad, K. (2023). Heart diseases prediction using block-chain and machine learning. *arXiv preprint arXiv:2306.01817*. <https://doi.org/10.48550/arXiv.2306.01817>
 - [19] Shynu, P.G., Menon, V.G., Kumar, R.L., Kadry, S., Nam, Y. (2021). Blockchain-based secure healthcare application for diabetic-cardio disease prediction in fog computing. *IEEE Access*, 9: 45706-45720. <https://doi.org/10.1109/ACCESS.2021.3065440>
 - [20] Hasanova, H., Tufail, M., Baek, U.J., Park, J.T., Kim, M.S. (2022). A novel blockchain-enabled heart disease prediction mechanism using machine learning. *Computers and Electrical Engineering*, 101: 108086. <https://doi.org/10.1016/j.compeleceng.2022.108086>
 - [21] Khatoun, A. (2020). The blockchain-based smart contract system for healthcare management. *Electronics*, 9(1): 94. <https://doi.org/10.3390/electronics9010094>
 - [22] Kayikci, S., Khoshgoftaar, T.M. (2024). Blockchain meets machine learning: A survey. *Journal of Big Data*, 11: 9. <https://doi.org/10.1186/s40537-023-00852-y>
 - [23] Angraal, S., Krumholz, H.M., Schulz, W.L. (2017). Blockchain technology: Applications in health care. *Circulation: Cardiovascular Quality and Outcomes*, 10(9): e003800. <https://doi.org/10.1161/CIRCOUTCOMES.117.003800>
 - [24] Sinha, R. (2024). The role and impact of new technologies on healthcare systems. *Discover Health Systems*, 3(1): 96. <https://doi.org/10.1007/s44250-024-00163-w>
 - [25] Hassan, C.A.U., Iqbal, J., Irfan, R., Hussain, S., Algarni, A.D., Bukhari, S.S.H., Alturki, N., Ullah, S.S. (2022). Effectively predicting the presence of coronary heart disease using machine learning classifiers. *Sensors*, 22(19): 7227. <https://doi.org/10.3390/s22197227>
 - [26] AbdelSalam, F.M. (2023). Blockchain revolutionizing healthcare industry: A systematic review of blockchain technology benefits and threats. *Perspectives in Health Information Management*, 20(3): 1b.
 - [27] Lo, P. (2024). Healthcare innovations: Enhancing patient privacy and security in the digital era. *Healthcare Management Forum*, 37(5): 363-365. <https://doi.org/10.1177/08404704241254820>
 - [28] Mathew, D.E., Ebem, D.U., Ikegwu, A.C., Ukeoma, P.E., Dibiazue, N.F. (2025). Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of AI models for human. *Neural Processing Letters*, 57(1): 16. <https://doi.org/10.1007/s11063-025-11732-2>
 - [29] Letrache, K., Ramdani, M. (2023). Explainable artificial intelligence: A review and case study on model-agnostic methods. In *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*, Casablanca, Morocco, pp. 1-8. <https://doi.org/10.1109/SITA60746.2023.10373722>
 - [30] Chicco, D., Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, 20(1): 16. <https://doi.org/10.1186/s12911-020-1023-5>
 - [31] Mole, J.S., Shaji, R.S. (2024). Ethereum blockchain for electronic health records: Securing and streamlining patient management. *Frontiers in Medicine*, 11: 1434474. <https://doi.org/10.3389/fmed.2024.1434474>
 - [32] Kuglics, B., Sipos, M. (2024). Blockchain-Based Healthcare DApp with MetaMask Integration. In *2024 IEEE 22nd Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*, Pula, Croatia, pp. 455-462. <https://doi.org/10.1109/SISY62279.2024.10737534>
 - [33] Saini, S., Kohli, S. (2017). Healthcare data analysis using R and MongoDB. In *Big Data Analytics: Proceedings of CSI 2015*, pp. 709-715. https://doi.org/10.1007/978-981-10-6620-7_69
 - [34] Sreekanth, R., Rao, G.M., Nanduri, S. (2015). Big data electronic health records data management and analysis on cloud with MongoDB: A NoSQL database. *International Journal of Advanced Engineering and Global technology*, 3(7): 946-949.
 - [35] Xu, W., Zhou, Z.H., Zhou, H., Zhang, W., Xie, J. (2014). MongoDB improves big data analysis performance on electric health record system. In *International Conference on Intelligent Computing for Sustainable Energy and Environment*, pp. 350-357. https://doi.org/10.1007/978-3-662-45283-7_36
 - [36] Tlemçani, K., Azbeg, K., Saoudi, E., Fetjah, L., Ouchetto, O., Andaloussi, S.J. (2025). Empowering diabetes management through blockchain and edge computing: A systematic review of healthcare innovations and challenges. *IEEE Access*, 13: 14426-14443. <https://doi.org/10.1109/ACCESS.2025.3531350>
 - [37] Zhang, P., Jia, Y.Q., Shang, Y.L. (2022). Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*, 18(6): 15501329221106935. <https://doi.org/10.1177/15501329221106935>

- [38] Mohammed, A., Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University-Computer and Information Sciences*, 35(2): 757-774. <https://doi.org/10.1016/j.jksuci.2023.01.014>
- [39] Richhariya, B., Tanveer, M. (2018). A robust fuzzy least squares twin support vector machine for class imbalance learning. *Applied Soft Computing*, 71: 418-432. <https://doi.org/10.1016/j.asoc.2018.07.003>
- [40] Halder, R.K., Uddin, M.N., Uddin, M.A., Aryal, S., Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications. *Journal of Big Data*, 11(1): 113. <https://doi.org/10.1186/s40537-024-00973-y>
- [41] Raiaan, M.A.K., Sakib, S., Fahad, N.M., Al Mamun, A., Rahman, M.A., Shatabda, S., Mukta, M.S.H. (2024). A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks. *Decision Analytics Journal*, 11: 100470. <https://doi.org/10.1016/j.dajour.2024.100470>
- [42] El-Sofany, H., El-Seoud, S. A., Karam, O.H., Abd El-Latif, Y.M., Taj-Eddin, I.A. (2024). A proposed technique using machine learning for the prediction of diabetes disease through a mobile app. *International Journal of Intelligent Systems*, 2024(1): 6688934. <https://doi.org/10.1155/2024/6688934>
- [43] Kumari, S., Kumar, D., Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, 2: 40-46. <https://doi.org/10.1016/j.ijcce.2021.01.001>
- [44] Mishra, D.K., Mehra, P.S. (2025). DiabeticChain: A novel blockchain approach for patient-centric diabetic data management. *The Journal of Supercomputing*, 81(1): 166. <https://doi.org/10.1007/s11227-024-06589-6>
- [45] Ge, Z.R., Loghin, D., Ooi, B.C., Ruan, P.C., Wang, T.W. (2022). Hybrid blockchain database systems: design and performance. *Proceedings of the VLDB Endowment*, 15(5): 1092-1104. <https://doi.org/10.14778/3510397.3510406>