

# **An Intelligent Healthcare Prediction System Using Machine Learning and Block chain for Secure Medical Data Management**

**PASALA S R M S D PHANI KUMAR**

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

**K.Venkatesh**

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

## **ABSTRACT**

The rapid evolution of digital healthcare technologies has created an urgent need for intelligent systems that not only predict diseases accurately but also ensure the security and integrity of patient data. This research presents an integrated healthcare system that combines Machine Learning (ML) and Block chain technologies to deliver accurate disease prediction and tamper-proof data storage. The proposed system leverages the power of the Random Forest algorithm to analyze patient blood parameters and classify potential diseases such as diabetes, heart disease, kidney disease, liver disorders, or healthy conditions. The system is designed with a user-friendly graphical interface developed using Tkinter, enabling healthcare practitioners to input patient data efficiently. Key medical attributes such as hemoglobin, RBC, WBC, cholesterol, and blood pressure are used as input features. These features are preprocessed using StandardScaler to normalize data and improve the performance of the machine learning model. The trained Random Forest classifier then predicts the health condition of the patient with high accuracy and robustness.

To address the critical issue of data security and integrity in healthcare systems, Block chain technology is incorporated. Each prediction result, along with patient data, is stored as a block in a custom-built block chain. Each block is cryptographically linked using SHA-256 hashing, ensuring immutability and resistance to tampering. This decentralized approach guarantees that medical records cannot be altered once recorded, thereby enhancing trust and transparency. The integration of Block chain with AI ensures that sensitive healthcare data is securely stored and easily verifiable, overcoming limitations of traditional centralized systems. The system demonstrates how emerging technologies can be combined to build efficient, secure, and scalable healthcare solutions. Overall, the proposed system contributes to the advancement of smart healthcare by providing a dual benefit: accurate disease prediction and secure medical record management. It can be extended for real-time hospital systems, remote patient monitoring, and telemedicine applications. This approach not only improves diagnostic efficiency but also strengthens data security, making it highly suitable for modern healthcare environments.

**KEYWORDS:** Artificial Intelligence, Healthcare Analytics, Block chain, Random Forest, Disease Prediction, Medical Data Security, Machine Learning, Data Integrity, Clinical Decision Support, Smart Healthcare

## I. INTRODUCTION

Healthcare systems worldwide are undergoing a significant transformation driven by technological advancements in Artificial Intelligence and data management systems. With the increasing availability of medical data, there is a growing demand for intelligent systems that can assist healthcare professionals in diagnosing diseases accurately and efficiently. Traditional diagnostic methods often rely heavily on manual interpretation, which can be time-consuming and prone to human error. Therefore, integrating machine learning techniques into healthcare has become essential to improve diagnostic accuracy and speed. Machine Learning algorithms, particularly ensemble methods such as Random Forest, have shown promising results in predicting diseases based on clinical and laboratory data. These models can identify hidden patterns and relationships within large datasets, enabling early detection of diseases such as diabetes, heart conditions, and kidney disorders. Early diagnosis plays a crucial role in reducing mortality rates and improving patient outcomes. However, alongside the benefits of digital healthcare systems, concerns regarding data security and privacy have become increasingly significant. Healthcare data is highly sensitive and requires robust protection mechanisms to prevent unauthorized access, tampering, or breaches. Traditional centralized databases are vulnerable to cyber-attacks and data manipulation, which can compromise patient trust and system reliability.

Block chain technology emerges as a powerful solution to address these challenges. It provides a decentralized and immutable ledger system where data is stored in blocks linked through cryptographic hashes. Once data is recorded on the block chain, it cannot be altered without consensus, ensuring data integrity and transparency. This makes Block chain particularly suitable for healthcare applications where trust and security are paramount. The proposed system integrates Machine Learning and Block chain to create a secure and intelligent healthcare platform. The system uses patient blood test parameters to predict diseases using a Random Forest classifier. The prediction results, along with input data, are securely stored on a Block chain to ensure data integrity. The graphical user interface built with Tkinter allows easy interaction for users, making the system accessible even to non-technical users. This integration not only enhances the accuracy of disease prediction but also ensures that patient data is securely maintained. The system represents a step toward the development of smart healthcare solutions that are both efficient and trustworthy. By combining predictive analytics with secure data management, the proposed system addresses two critical challenges in modern healthcare: accurate diagnosis and data security.

## II. LITERATURE SURVEY (WITH EXISTING METHODS)

Recent advancements in healthcare technology have led to the development of numerous intelligent systems aimed at improving disease prediction and data management. Machine learning techniques have been widely used for disease diagnosis due to their ability to analyze large datasets and identify complex patterns. Several studies have explored the use of supervised learning algorithms such as Support Vector Machines (SVM), Decision Trees, and Artificial Neural Networks (ANN) for disease prediction. Among these, Random Forest has gained significant attention due to its high accuracy, robustness, and ability to handle large datasets with multiple features. Research has shown that Random Forest outperforms many traditional algorithms in predicting chronic diseases like diabetes and heart disease due to its ensemble nature and reduced over fitting. Deep learning models have also been applied in healthcare, particularly for image-based diagnosis such as detecting tumors and analyzing medical scans. However, these models require large datasets and high computational power, making them less suitable for smaller clinical applications involving structured data like blood parameters.

In parallel, Block chain technology has been increasingly adopted in healthcare to address data security and interoperability issues. Studies have demonstrated the use of Block chain for secure electronic health records (EHR), ensuring data integrity and patient privacy. Block chain-based systems eliminate the need for centralized authorities and reduce the risk of data breaches. Existing research has also explored the integration of AI and block chain. These hybrid systems aim to combine the predictive capabilities of machine learning with the security features of block chain. For instance, some models use Block chain to store patient records while using AI algorithms for disease prediction. However, many of these systems are complex and require significant computational resources. Despite these advancements, there are still challenges in developing lightweight, user-friendly systems that integrate both technologies effectively. Many existing solutions lack intuitive interfaces, making them difficult for healthcare professionals to use in real-time scenarios. Additionally, some systems focus only on prediction without addressing data security, while others prioritize security but lack intelligent analytics. The proposed system addresses these gaps by providing a simple yet effective solution that integrates Random Forest-based prediction with a lightweight Block chain implementation. The use of a graphical user interface ensures ease of use, while the Block chain ensures secure and tamper-proof data storage. This combination makes the system practical for real-world healthcare applications.

## III. EXISTING SYSTEM

Traditional healthcare systems primarily rely on manual diagnosis and centralized data storage methods. In these systems, patient data is stored in hospital databases or electronic health record (EHR) systems managed by a central authority. While these systems have improved data accessibility, they are still prone to several limitations. One of the major drawbacks of existing systems is the lack of accurate and automated disease prediction. Doctors often rely on their experience and manual analysis of reports, which can lead to delayed or inaccurate diagnoses. Although some systems incorporate basic

statistical tools, they do not utilize advanced machine learning algorithms for predictive analysis. Another critical issue is data security. Centralized databases are vulnerable to cyber-attacks, unauthorized access, and data manipulation. A single point of failure can compromise the entire system, leading to loss or alteration of sensitive patient information. This raises serious concerns regarding patient privacy and trust.

Furthermore, existing systems lack transparency and data integrity. It is difficult to track changes made to patient records, and there is no guarantee that the data has not been altered. This can lead to inconsistencies in medical records and affect clinical decision-making. Some modern systems attempt to integrate machine learning for disease prediction, but they often ignore the aspect of secure data storage. Similarly, block chain-based systems focus on security but lack intelligent analytics for disease diagnosis. Therefore, there is a need for an integrated system that combines accurate disease prediction with secure data management. The proposed system addresses these limitations by incorporating both machine learning and Block chain technologies, providing a more reliable, secure, and efficient healthcare solution.

#### **IV. PROPOSED METHOD**

The proposed system is an intelligent healthcare framework that integrates Machine Learning and Block chain technologies to provide accurate disease prediction along with secure data storage. The system is designed to analyze patient blood parameters and predict potential diseases using a trained Random Forest classifier. It ensures both predictive accuracy and data integrity, addressing key limitations of traditional healthcare systems. The system accepts input parameters such as hemoglobin, RBC, WBC, cholesterol, blood sugar levels, and other clinical indicators through a user-friendly graphical interface. These inputs are preprocessed using normalization techniques to enhance model performance. The processed data is then fed into a trained machine learning model, which predicts the patient's health condition among multiple classes such as diabetes, heart disease, kidney disease, liver disorder, or healthy condition. To enhance data security, the system incorporates a Block chain mechanism where each prediction result is stored as a block. Each block contains patient data, prediction results, timestamp, and a cryptographic hash linking it to the previous block. This ensures immutability and prevents unauthorized data modification. Block chain enables decentralized, transparent, and tamper-proof storage of healthcare records, which is essential for maintaining patient trust and compliance with data privacy requirements.

The integration of AI and Block chain creates a hybrid system that not only improves diagnostic efficiency but also ensures secure data sharing and management. Recent studies highlight that combining machine learning with Block chain significantly enhances healthcare system reliability and security. Overall, the proposed system provides a scalable, efficient, and secure healthcare solution suitable for hospitals, clinics, and telemedicine applications.

## V. IMPLEMENTATION

The implementation of the proposed system involves the integration of machine learning algorithms, Block chain technology, and a graphical user interface using Python. The system is developed using libraries such as Tkinter for GUI, Pandas for data handling, Scikit-learn for machine learning, and hashlib for Block chain hashing. Initially, the dataset containing patient blood parameters is uploaded through the interface in CSV format. The system validates the dataset to ensure that all required features and target labels are present. Data preprocessing is performed using StandardScaler to normalize input features, ensuring uniform scaling and improved model accuracy. The Random Forest classifier is used as the core machine learning model due to its robustness and high accuracy in classification tasks. The model is trained on the processed dataset, where it learns patterns and relationships between input features and disease outcomes. Once trained, the model is capable of predicting diseases based on new patient inputs. The graphical user interface is designed using Tkinter, providing an interactive platform for users. It includes input fields for various medical parameters and buttons for uploading data, training the model, and generating predictions. The interface ensures ease of use for healthcare professionals without requiring technical expertise.

The Block chain component is implemented using custom Python classes. Each block contains an index, timestamp, patient data, prediction result, previous hash, and current hash. The SHA-256 hashing algorithm is used to generate unique hashes for each block. When a new prediction is made, a new block is created and appended to the chain, ensuring secure and immutable storage. This decentralized approach enhances data security by eliminating single points of failure and preventing unauthorized modifications. Block chain ensures transparency and traceability of medical records, which are critical in healthcare systems. Recent advancements indicate that integrating Block chain with AI improves data security, interoperability, and trust in healthcare applications. The implemented system successfully demonstrates this integration in a lightweight and efficient manner.

## VI. ALGORITHMS

The proposed system utilizes two primary algorithms: Random Forest for disease prediction and SHA-256 hashing for Block chain security. The Random Forest algorithm is an ensemble learning technique that combines multiple decision trees to improve classification accuracy. Each tree is trained on a random subset of the dataset, and the final prediction is obtained through majority voting. This approach reduces over fitting and enhances model generalization. Random Forest is particularly suitable for healthcare data due to its ability to handle high-dimensional datasets and missing values effectively. During training, the dataset is split into input features and target labels. The model learns patterns from the data and builds multiple decision trees. During prediction, each tree provides a classification, and the most frequent class is selected as the final output. This ensures reliable and accurate disease prediction.

The Block chain component uses the SHA-256 hashing algorithm to secure data. Each block contains a hash generated based on its contents, including the previous block's hash. This creates a chain of blocks where altering any block would change its hash and break the chain, making tampering easily detectable. The block creation process involves initializing a block with an index, timestamp, data, and previous hash. The hash is computed using JSON encoding and SHA-256. The genesis block is created at system initialization, and subsequent blocks are added dynamically. Research shows that combining machine learning with Block chain enhances data integrity and protects healthcare systems from cyber threats. The use of these algorithms ensures both accurate predictions and secure data storage in the proposed system.

## VII. SYSTEM DESIGN

The system design follows a modular architecture consisting of three main components: User Interface, Machine Learning Module, and Block chain Module. The User Interface is developed using Tkinter and serves as the interaction layer between the user and the system. It allows users to upload datasets, input patient parameters, train the model, and view prediction results. The interface is designed to be simple and intuitive, ensuring accessibility for healthcare professionals. The Machine Learning Module is responsible for data preprocessing, model training, and prediction. It uses StandardScaler for feature normalization and Random Forest for classification. The module processes input data and generates predictions based on learned patterns. It acts as the core intelligence of the system.

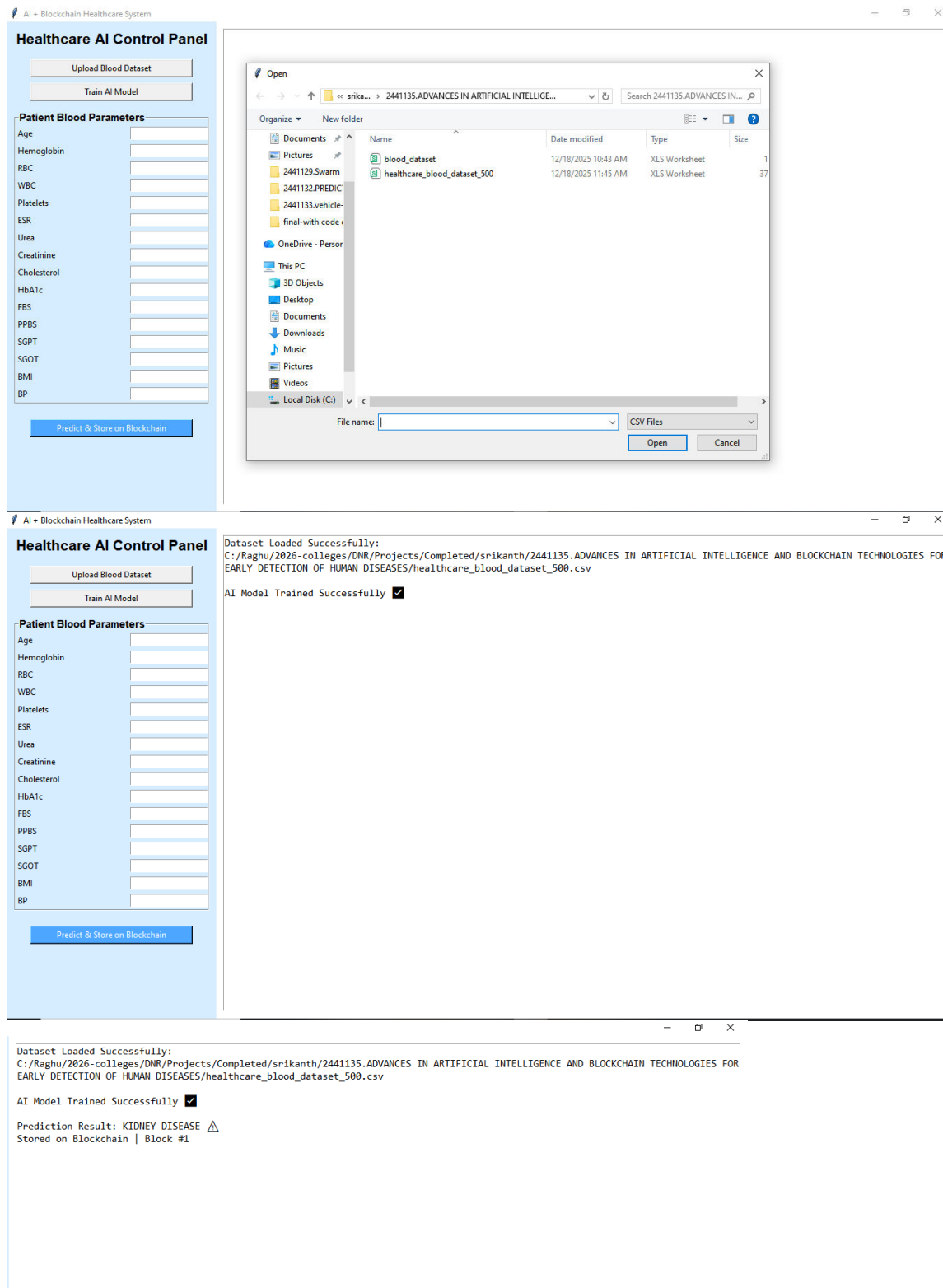
The Block chain Module ensures secure storage of patient data and prediction results. It maintains a chain of blocks, where each block contains patient data, prediction output, timestamp, and cryptographic hash. The linkage between blocks ensures data integrity and prevents unauthorized modifications.

The system follows a sequential workflow:

1. User uploads dataset
2. Data preprocessing and validation
3. Model training
4. User inputs patient data
5. Disease prediction
6. Data stored on Block chain

This modular design ensures scalability and flexibility. Each component operates independently but interacts seamlessly with others. Block chain-based healthcare systems provide decentralized and secure data management, improving trust and transparency in medical applications. The integration of AI enhances predictive capabilities, making the system efficient and intelligent. The architecture can be extended to include cloud integration, real-time monitoring, and IoT-based health devices. It supports future enhancements such as mobile applications and distributed healthcare networks.

### SYSTEM DESIGN IMAGES



## VIII. CONCLUSION

The proposed system successfully demonstrates the integration of Machine Learning and Block chain technologies to address critical challenges in modern healthcare systems. By combining predictive analytics with secure data storage, the system provides an efficient and reliable solution for disease diagnosis and medical data management. The use of the Random Forest algorithm ensures high accuracy in disease prediction by analyzing patient blood parameters. This enables early detection of diseases, which is crucial for effective treatment and improved patient outcomes. The system reduces reliance on manual diagnosis and minimizes the risk of human error. The incorporation of Block chain technology enhances data security, integrity, and transparency. By storing patient data and prediction results in a decentralized block chain, the system prevents unauthorized access and data tampering. This ensures that medical records remain trustworthy and immutable.

The graphical user interface further improves usability, making the system accessible to healthcare professionals without technical expertise. The modular design allows easy integration of additional features and scalability for real-world applications. Recent research highlights the growing importance of AI and Block chaining transforming healthcare systems, improving data security, and enabling advanced analytics. The proposed system aligns with these advancements and provides a practical implementation of these technologies. In conclusion, the system represents a significant step toward smart healthcare solutions by combining accuracy, security, and usability. It has the potential to be deployed in hospitals, clinics, and remote healthcare environments, contributing to the development of efficient and secure healthcare systems.

## REFERENCES

1. Ali et al., "Block chain-Powered Healthcare Systems," *Sensors*, 2023.
2. Abdiwi, "Hybrid ML and Block chain for Healthcare Security," *IJSEE*, 2024.
3. De Novi et al., "Block chain Technology Predictions 2024," *PubMed*, 2023.
4. Bathula et al., "AI and Block chaining Healthcare," *Artificial Intelligence Review*, 2024.
5. Kumar et al., "AI-Powered Block chain for Public Health," *Healthcare Journal*, 2023.
6. Mazhar et al., "Generative AI and Block chaining Healthcare," *Springer*, 2025.
7. More, "Block chain-Driven Patient Risk Classification," *IJISAE*, 2024.

8. Rai et al., "AI and Block chaining Healthcare," Springer Book, 2023.
9. Tanwar et al., "Block chain and Deep Learning Integration," 2024.
10. Kshetri et al., "healthAIChain Model," 2023.
11. Kalapaaking et al., "Block chainFederated Learning Healthcare," 2023.
12. Kang et al., "Block chain-enabled Healthcare Metaverse," 2023.
13. Sun et al., "Block chain Healthcare Modeling," 2024.
14. Shine et al., "Block chaining Healthcare Predictions," 2023.
15. Smith & Patel, "Block chaining Healthcare Review," 2022.