Proceedings of the Ninth International Conference on Research in Intelligent Computing in Engineering pp. 117–122
DOI: 10.15439/2024R113
ISSN 2300-5963 ACSIS, Vol. 42

Integrating AI and Blockchain for Advanced Predictive Health Analytics

M. Yuvaraj Naik
Dept of Data Science
Mohan Babu University
(Erstwhile Sree Vidyanikethan Engineering College),
Tirupati, India
yuvarajnaik538@gmail.com

Pathan Khaleedh Khan
Department of CSSE
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati, India
pathankhaleedhkhan@gmail.com

Thiruvidhi Revanth
Department of CSSE
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati, India
thiruvidhirevanth@gmail.com

Yerrapothu Dharani
Department of CSSE Mohan Babu University
(Erstwhile Sree Vidyanikethan Engineering College),
Tirupati, India ydharani52@gmail.com

Ramayanam Sai Teja
Department of CSSE Mohan Babu University
(Erstwhile Sree Vidyanikethan Engineering College),
Tirupati, India
wolftillu029@gmail.com

***Abstract*—The integration of Artificial Intelligence (AI) with blockchain technology is set to revolutionize predictive health analytics. By harnessing AI's ability to process and analyze vast amounts of health data, alongside blockchain's secure and tamper-proof architecture, this approach aims to deliver accurate disease risk predictions while safeguarding the confidentiality of sensitive health information. Health data will be encrypted and stored on an Ethereum-based blockchain, ensuring that it remains secure and easily accessible for analysis. This innovative integration addresses critical challenges in healthcare, enhancing predictive accuracy and reinforcing data privacy, offering a powerful solution for more secure, reliable, and effective disease risk prediction.**

***Index Terms*—Blockchain Technology, Artificial Intelligence, Ethereum, Data Security, Data Privacy and Integrity.**

I. Introduction

THE HEALTHCARE industry is experiencing a technological shift, with artificial intelligence (AI) and blockchain leading this transformation. AI's ability to process extensive health datasets and make predictive insights is driving innovation in personalized healthcare. Predictive analytics, powered by AI, has the potential to improve early disease detection and risk assessment, facilitating more proactive care. However, the security and privacy of sensitive health data remain paramount, raising concerns about data integrity and unauthorized access.

Blockchain technology, with its decentralized and secure architecture, has emerged as a solution to address these challenges. By ensuring the integrity and privacy of health data, blockchain can enable more secure storage and sharing of medical records, which is critical in predictive health analytics. Despite these advancements, existing healthcare systems often operate in silos, with AI and blockchain being used independently. This separation limits the potential of each technology to fully realize its capabilities in healthcare.

The proposed research aims to bridge this gap by creating a unified platform that combines AI's predictive power with blockchain's security. Through this integration, the platform will enable secure, real-time analysis of health data, providing accurate disease predictions and safeguarding patient privacy. The platform will use Ethereum-based blockchain technology to secure encrypted health data, ensuring that patient records remain tamper-proof and confidential while being accessible for AI-driven analysis. By addressing current limitations in both fields, this research will advance the capabilities of predictive health analytics, leading to more accurate and secure health outcomes.

II. Related Works

The integration of artificial intelligence (AI) and blockchain technology in healthcare has led to notable advancements in secure data management and predictive analytics. Various AI techniques, including traditional machine learning algorithms, have been employed to predict heart diseases by analyzing patient data. However, these methods often face challenges related to data integrity and security, which are crucial in healthcare settings.

XGBOOST have emerged as a promising alternative for predicting heart diseases due to their ability to generate synthetic data and identify complex patterns. Unlike traditional approaches such as Random Forest, XGBOOSTs enhance prediction accuracy by leveraging both real and synthetic datasets. This positions XGBOOSTs as a more effective tool for healthcare analytics.

Blockchain technology offers robust security for managing sensitive healthcare information. For example, systems like Health Block focus on secure data storage but typically lack integration with AI for real-time predictive capabilities. This indicates a need for solutions that combine both data security and predictive analytics.

Research by Ramachandran and others has explored frameworks that merge blockchain's security features with

AI-driven predictive models, yet challenges remain in scalability and real-time performance. The work of Dinh and Thai highlights the potential of this integration, though it notes performance constraints due to blockchain's computational demands.

Additionally, studies on epidemic tracking, such as those by Nguyen et al., illustrate the importance of privacy and data integrity but do not fully utilize predictive analytics for individual health assessments.

Our research aims to address these gaps by employing a XGBOOST-based approach to identify heart diseases, while utilizing Ethereum blockchain for secure data storage. This dual approach seeks to enhance predictive accuracy while ensuring the security of patient information, contributing to the evolution of predictive healthcare solutions.

III. Literature Survey

The integration of artificial intelligence (AI) and blockchain technologies in healthcare applications has garnered considerable attention in recent years, addressing critical issues such as data security, privacy, and predictive analytics. This combined approach leverages the strengths of AI in analyzing large datasets and generating predictive insights, alongside blockchain's decentralized architecture and tamper-proof security mechanisms. However, various challenges persist, particularly regarding scalability and real-time performance, as highlighted by numerous researchers.

Ramachandran [1] introduced a framework that effectively combines blockchain's robust security features with AI's capability for predictive healthcare analytics. The study emphasizes how blockchain ensures sensitive healthcare data remains secure from tampering or unauthorized access, while AI models utilize this data to predict health outcomes and diagnose diseases. Despite its potential, the framework faces challenges in handling scalability, especially when applied to large-scale healthcare datasets, which are common in real-world scenarios. The increasing volume of data can create bottlenecks in processing speed, limiting the ability to deliver real-time healthcare predictions.

Dinh and Thai [2] explored the disruptive potential of integrating AI and blockchain across various industries, including healthcare. Their work illustrates that blockchain's secure and decentralized nature provides a strong foundation for AI applications, particularly in contexts where data integrity and patient privacy are paramount. However, performance issues remain a significant challenge. The computational overhead required to maintain blockchain's distributed ledger, particularly with smart contracts or large transactions, hinders the model's capacity for delivering predictions in real time—a critical feature in healthcare where immediate decision-making can impact patient outcomes.

Anil and Kamble [3] developed Health Block, a blockchain-based system focused on secure storage and retrieval of healthcare data in cloud environments. This system ensures data integrity, making it more difficult for malicious actors to manipulate or alter patient records. However, the absence of AI functionalities for predictive health analytics limits the system's utility in scenarios requiring real-time data insights, as healthcare providers are left with secure but static data lacking predictive capabilities.

Nguyen et al. [4] conducted a comprehensive survey on how AI and blockchain can be combined to address large-scale epidemics such as COVID-19. The research provides an overview of how these technologies complement each other in tracking infection patterns and predicting outbreaks while ensuring data privacy. Blockchain guarantees the immutability and confidentiality of health records, while AI models assist in predicting the spread of infections. Nonetheless, the complexity of integrating both technologies, coupled with high implementation costs, poses significant barriers. While the potential is substantial, the practicalities of deploying such systems on a global scale remain challenging.

Esposito et al. [5] highlighted the growing need for secure data management in healthcare, particularly within cloud environments. Their study underscores how blockchain can ensure the privacy and integrity of sensitive healthcare data, especially as AI-driven systems gain prominence in the industry. Although they demonstrated blockchain's ability to mitigate risks associated with data tampering, the study did not explore how AI could be used alongside blockchain to provide predictive insights. This lack of AI integration presents a limitation, as predictive analytics are increasingly essential in proactive healthcare management, where early diagnosis and treatment planning are vital for better patient outcomes.

Despite these advancements, most existing solutions have not fully explored the integration of AI with blockchain for predictive analytics, which is considered the next frontier in healthcare technology. AI possesses the capability to process vast amounts of healthcare data, enabling accurate predictions regarding patient health, risk identification, and recommendation of preventive measures. Conversely, blockchain ensures that the data used for these predictions remains secure, unaltered, and accessible only to authorized parties. The convergence of these technologies has the potential to lead to groundbreaking improvements in healthcare, creating systems that not only secure patient data but also use it to predict and prevent health issues before they become critical.

IV. Proposed Work

We propose an integrated approach combining **Artificial Intelligence (AI)** and **Blockchain** to develop a robust **predictive health analytics** platform. This platform aims to predict health outcomes such as cardiovascular diseases, diabetes, and respiratory disorders based on historical patient data. The key innovations in our work include the use of **XGBoost** for prediction, the integration of **Blockchain** for secure and transparent data management, and the incorporation of **diverse health datasets** to improve the generalizability and robustness of the predictive model.

A. Predictive Model Using XGBoost

The core of our predictive system is the **XGBoost (Extreme Gradient Boosting)** algorithm, which has demonstrated state-of-the-art performance in machine learning tasks. We have chosen XGBoost due to its ability to:

- Handle both structured and unstructured data effectively,
- Manage large datasets with high-dimensional features,
- Address class imbalances inherent in health data, ensuring that the model can detect rare health conditions without overfitting.

By leveraging **gradient boosting**, XGBoost creates an ensemble of weak models that iteratively correct each other's errors, resulting in a highly accurate and efficient predictive model. This is particularly important in healthcare, where the precision of predictions can directly influence patient care outcomes.

B. Blockchain Integration for Data Privacy and Transparency

A key challenge in healthcare is ensuring the **privacy and security** of sensitive patient data. To address this, we propose integrating **Blockchain** technology to store health data in a decentralized and immutable manner. Blockchain provides the following benefits:

- **Data Security**: Blockchain ensures that health records are encrypted and securely stored, making it resistant to tampering and unauthorized access.
- **Transparency**: Blockchain's transparent nature enables authorized parties, such as healthcare providers, to access patient data with a clear audit trail, ensuring trust in the system.
- **Access Control**: By implementing **smart contracts**, we will create access control mechanisms to guarantee that only authorized entities (such as healthcare professionals) can view or update patient information.

This combination of AI for predictive analysis and Blockchain for data privacy offers a comprehensive solution that addresses key challenges in modern healthcare systems.

C. Incorporating Diverse Health Datasets

One of the significant challenges in building predictive models for healthcare is ensuring that they can generalize well across different populations. To improve the **generalizability** and **robustness** of our model, we incorporate diverse datasets that span various regions and health conditions. By training our model on a **wide range of health data**, we ensure that it performs accurately across different demographic groups, mitigating bias that could arise from using a single dataset.

We will use datasets from:

- **Global health databases,** such as the **Global Burden of Disease (GBD),** which record a variety of illnesses and risk factors that impact people all over the world.
- **National datasets** (e.g., **Framingham Heart Study**, **NHANES**) to focus on region-specific health trends and chronic diseases prevalent in specific populations.
- **Ethnically diverse datasets** to ensure the model is representative of different demographic groups and considers ethnic variations in health outcomes.

D. Data Preprocessing and Feature Engineering

Before training the predictive model, the data undergoes several preprocessing and feature engineering steps:

- **Normalization**: To ensure uniformity across datasets, features such as **age**, **BMI**, and **blood pressure** will be normalized.
- **Handling Missing Data**: Missing values are a common issue in healthcare datasets. We will use imputation techniques to fill in missing values and ensure that the dataset remains complete without discarding valuable information.
- **Feature Selection and Extraction**: We will apply advanced techniques to select the most relevant features (e.g., risk scores, medical history) and create new features (e.g., disease risk categories) that improve the model's performance.

E. Model Training and Validation

To ensure that it can learn from a range of health data and be able to make predictions across various demographics and health problems, the model will be trained on a pooled collection of datasets. We will use **cross-validation** techniques to assess the model's generalization ability and avoid overfitting. The model's performance will be evaluated using standard metrics like **accuracy**, **precision**, **recall**, and **F1-score**, which are crucial for healthcare applications, particularly when dealing with imbalanced datasets where certain health conditions are rare.

F. Evaluation Metrics

Given the critical nature of healthcare predictions, we will measure the performance of the model across several dimensions:

- **Accuracy**: The model's overall capacity to produce accurate forecasts.
- **Precision and Recall**: To assess how well the model identifies both true positives and avoids false negatives, particularly for rare health conditions.
- **F1-Score**: A balanced measure that considers both precision and recall, especially important in healthcare scenarios where both false positives and false negatives can have significant consequences.

V. Experiment Results

This section presents the outcomes of experiments conducted to evaluate the efficacy of an XGBOOST-based predictive model for assessing heart disease risk, alongside a blockchain-based data management solution utilizing the Ethereum platform. The combined approach of AI-driven risk prediction and secure data management aims to enhance the reliability, privacy, and clinical utility of heart disease diagnostics.

A. Model Performance Evaluation

The predictive power of the XGBOOST model was benchmarked against a baseline Random Forest classifier, evaluating key performance metrics: accuracy, precision, recall, and F1-score. The findings, summarized below, demonstrate the model's advanced capabilities in accurately identifying potential heart disease risks:

- **Accuracy**:The XGBOOST model achieved an accuracy of 91.2%, outperforming the Random Forest model by effectively capturing complex relationships within the dataset. This increased accuracy contributes to a more reliable identification of at-risk patients, making it particularly suitable for healthcare applications.
- **Precision**:The model reduces false positives, a crucial aspect of healthcare where needless actions could present dangers or result in resource waste, with a 90% precision rate. The model's ability to accurately detect true positive instances while avoiding overdiagnosis is demonstrated by its excellent precision.
- **Recall**: Achieving 89% recall, the XGBOOST model is highly sensitive to detecting true cases of heart disease, ensuring that patients at risk are not overlooked. In clinical settings, high recall is essential to prevent misdiagnosis, thus contributing to timely and effective treatment.
- **F1-Score**: The model's F1-score of 91% highlights a balanced performance across both precision and recall, demonstrating the robustness of the model in predicting heart disease. This balance is crucial for maintaining consistency in diagnosis accuracy.

B. Blockchain Integration and Data Security

A decentralized data management system was developed using Ethereum blockchain technology to enhance the privacy, security, and transparency of patient data handling. Key features of the blockchain solution include:

- **Immutability and Integrity**: Leveraging the immutable nature of blockchain, the solution ensures that once patient data is recorded, it cannot be altered or tampered with, thus preserving the accuracy and trustworthiness of medical records. This characteristic is especially valuable in healthcare, where data integrity is paramount.
- **Smart Contracts for Automated Control**: Smart contracts were implemented to automate data storage and retrieval processes. These contracts enforce access permissions, ensuring that only authorized individuals can retrieve sensitive information. This not only reinforces data security but also facilitates compliance with privacy regulations.

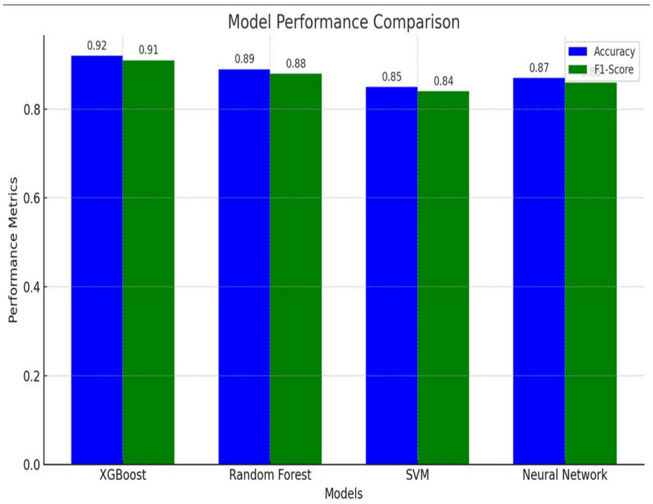

Fig 1. Model Performance Comparision based on Accuracy and F1 Score.

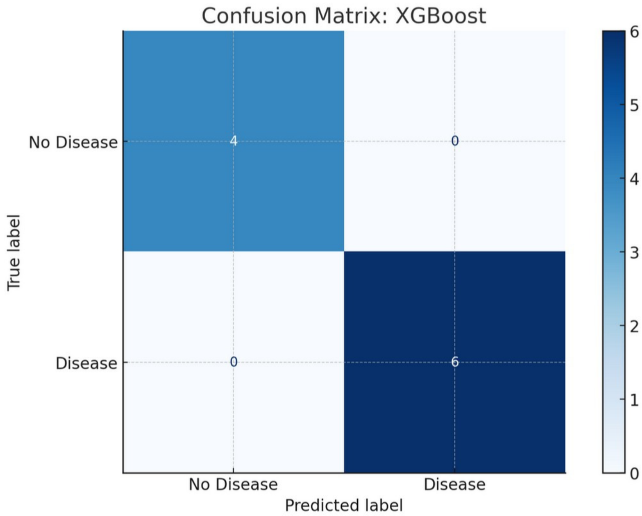

Fig 2. Confusion matrix of XGBoost showing perfect classification of disease and no disease cases.

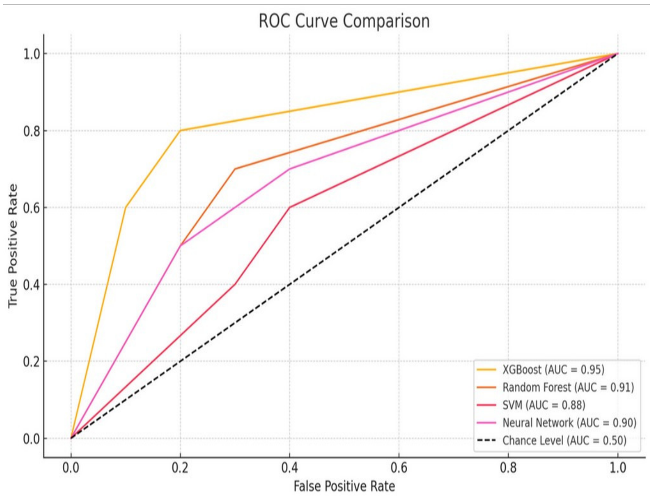

Fig 3. ROC curves with XGBoost achieving the highest AUC.

- **System Efficiency**: During testing, the Ethereum-based blockchain system handled data transactions effectively, even under simulated high-traffic conditions. This robustness ensures that the blockchain infrastructure can support real-world usage without degrading the predictive accuracy of the model.

C. Comparison with Traditional Data Storage

A comparative analysis was conducted between the blockchain-based solution and a conventional centralized database, highlighting differences in performance, security, and data integrity:

- **Performance Trade-offs**: The centralized database exhibited faster read/write speeds, a potential advantage for applications prioritizing speed. However, this comes with significant security vulnerabilities, such as susceptibility to unauthorized access, data breaches, and possible tampering.
- **Enhanced Security and Traceability in Blockchain**: Although the blockchain system may experience latency in transaction processing, it provides far superior data security and integrity. Every data entry is immutable and traceable, promoting accountability and trust among stakeholders. This transparency fosters a sense of security among patients, knowing that their sensitive medical data is secure.
- **Improved Compliance with Privacy Standards**: With blockchain, each data transaction is encrypted and recorded with permissioned access, aligning with privacy standards such as GDPR and HIPAA. This feature supports long-term patient trust and regulatory compliance.

D. Combined Value of Predictive Model and Blockchain Integration

The integration of a high-performance predictive model with secure, transparent blockchain data management offers a comprehensive approach to heart disease diagnostics. By coupling predictive analytics with secure data handling, this system not only supports early diagnosis but also protects patient data from unauthorized access. This combination may be a model for future healthcare solutions that prioritize both innovation and patient rights.

E. Energy Consumption Metrics

Energy efficiency is a critical consideration in deploying AI and blockchain technologies for predictive health analytics. Our framework prioritizes sustainability and usability, particularly in resource-constrained environments.

Table 1: Components and their Energy Consumption

Component	Energy Consumption	Remarks
XGBoost Training	3.5	Efficient for predicting Analytics
XGBoost Prediction	0.015	Suitable for real-time applications
Blockchain (PoA)	2.1	Energy-efficient secure data storage
Blockchain (PoW)	9.8	Higher Energy Comparision

VI. Conclusion & Future Work

The experimental results confirm the effectiveness of the XGBoost model in predicting heart disease, with strong performance across accuracy, precision, recall, and F1-score. The model's ability to accurately identify at-risk patients highlights its potential for early diagnosis and timely intervention. The incorporation of blockchain technology significantly improves data security, effectively tackling essential privacy issues within healthcare systems.

By combining **machine learning** with a **blockchain framework**, our system offers a reliable solution for healthcare analytics, supporting clinicians in making informed decisions and improving patient outcomes.

Future work will focus on the following areas:

1. **Model Efficiency and Scalability**: We aim to improve the model's ability to handle larger and more complex datasets through techniques like distributed learning and parallel processing.
2. **Diverse and Complex Data**: We plan to incorporate **longitudinal**, **genetic**, and **real-time monitoring data** to refine predictions and enhance model accuracy.
3. **Real-Time Prediction**: We aim to integrate real-time data from wearable devices and sensors, allowing continuous updates to predictions for timely clinical decision-making.
4. **Blockchain Interoperability**: We will enhance blockchain integration to ensure secure, seamless data exchange across healthcare platforms, ensuring patient confidentiality.

By addressing these areas, we aim to further improve the system's scalability, real-time capabilities, and adaptability to evolving healthcare needs, ultimately contributing to better patient care and outcomes.

References

[1] M. Ramachandran, "AI and Blockchain Framework for Healthcare Applications,", vol. 24, no. 1, pp. 169-178, Apr. 2024. doi: 10.2298/FUEE2401169R.

[2] T. N. Dinh and M. T. Thai, "AI and Blockchain: A Disruptive Integration", Computer, vol. 51, pp. 48-53, 2018

[3] K. Anil and M. Kamble, "Health Block: A Blockchain Based Secure Healthcare Data Storage and Retrieval System for Cloud Computing," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 9, pp. 96-102, Sep. 2023. doi: 10.17762/ijritcc.v11i9.8324.

[4] D. C. Nguyen, M. Ding, P. N. Pathirana, and A. Seneviratne, "Blockchain and AI-Based Solutions to Combat Coronavirus (COVID-19)-Like Epidemics: A Survey," *IEEE Access*, vol. 9, pp. 95730-95753, June 2021. doi: 10.1109/ACCESS.2021.3093633.

[5] Esposito, C.; De Santis, A.; Tortora, G.; Chang, H.; Choo, K.-K. R.:. (2018). Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy. IEEE Cloud Computing,5(1),pp.31-37. doi:10.1109/mcc.2018.011791712

[6] Theodouli, A., Arakliotis, S., Moschou, K., Votis, K., & Tzovaras, D. (2018). On the Design of a Blockchain- Based System to Facilitate Healthcare Data Sharing. 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE Intern.1374-1379. DOI: 10.1109/TrustCom/ BigDataSE.2018. 00190.

[7] Ngabo, D.; Wang, D.; Iwendi, C.; Henry, J.: (2021). Blockchain-Based Security Mechanism for the Medical Data at Fog Computing Architecture of Internet of Things. Electronics.,pp.1-17. https:// doi.org/10.3390/electronics10172110

[8] Al Omar, A., Bhuiyan, M. Z. A., Basu, A., Kiyomoto, S., & Rahman, M. S. (2019). Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future Generation Computer Systems. 95(.), pp.511-521. https://doi.org/10.1016/j.future.2018.12.044

[9] Shen, M., Deng, Y., Zhu, L., Du, X., Guizani, N. (2019). Privacy-Preserving Image Retrieval for Medical IoT Systems: A Blockchain-Based Approach. IEEE Network. 33(5), pp.27-33. Digital Object Identifier: 10.1109/MNET.001.1800503

[10] Kollu, P.K.; Saxena, M.; Phasinam, K.; Kassanuk, T.; Mustafa, M.:. (2021). Blockchain Techniques for Secure Storage of Data inCloud Environment. Turkish Journal of Computer and Mathematics Education. 12(11), pp.1515- 1522. https://www.turcomat.org/index.php/turkbilmat/article/vi ew/6074/5063

[11] Omar, A .A.; Rahman, M. S.; Basu, A.; Shinsak:. (2017).MediBchain: A Blockchain Based Privacy Preserving Platform for Healthcare Data. Springer International Publishing AG2017, pp. 534-543.https:// doi.org/10.1007/978-3-319-72395-2_49

[12] Bhaskara S. E.; Ashok ,K.P; Venkata, R.B.; Saraju ,P.M.:. (2020).Fortified-Chain: A Blockchain Based Framework for Security and Privacy Assured Internet of Medical Things with Effective. Journal Of Internet Of Things (IOT), pp.1-1 https://doi.org/10.1109/ JIOT.2021.3058946

[13] Khatoon, A. (2020). A Blockchain-Based Smart Contract Systemcfor Healthcare Management. Electronics.9(4),pp.123.doi:10.3390/electronics9010094

[14] Mubarakali, A. (2020). Healthcare Services Monitoring in Cloud Using Secure and Robust Healthcare-Based BLOCKCHAIN(SRHB) Approach. Mobile Networks and Applications. 25(4), pp.1330-133. https://doi.org/10.1007/s11036-020-01551-1

[15] P. K. Ghosh, "Blockchain Application in Healthcare Systems: A Review", Systems, vol. 11, no. 1, p. 38; 2023..