SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

AI-Blockchain-Enabled Framework for Cardio-Diabetic Risk Prediction

Vishal Kamble[1*], Dr. Nisha Wandile Kimmatkar[2], Dr. Seema Kedar[3], Dr. Ram joshi[4]

[1]Dept. of Computer Engineering, Rajarshi Shahu College of Engineering, Tathawade, Pune. E-mail: vishal3092.kamble@gmail.com
[2]Dept. of Computer Engineering, Rajarshi Shahu College of Engineering, Tathawade, Pune. E-mail: nishakimmatkar@gmail.com
[3]Dept. of Computer Engineering, Rajarshi Shahu College of Engineering, Tathawade, Pune. E-mail: seema_kedar@yahoo.com
[4]Dept. of Computer Engineering, Rajarshi Shahu College of Engineering, Tathawade, Pune. E-mail: rbjoshi_it@jspmrscoe.edu.in
*Corresponding Author: vishal3092.kamble@gmail.com

Abstract
Cardiovascular diseases and diabetes are increasingly prevalent health concerns in modern society, creating a significant burden on healthcare systems worldwide. This study proposes an AI- and blockchain-enabled framework for cardio-diabetic risk prediction and secure healthcare data management. Patient information is initially collected from medical sources and edge-enabled devices. The collected data undergoes preprocessing and analysis to improve data quality and prediction reliability. A hybrid ensemble machine learning model integrating Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB) is employed to enhance the accuracy and robustness of disease risk prediction. To ensure the security, integrity, and privacy of patient records, encryption techniques and blockchain-based storage are incorporated into the framework. In addition, a rule-based clustering algorithm is introduced to organize healthcare data efficiently and facilitate faster analysis. The blockchain layer provides a transparent, tamper-resistant, and access-controlled environment for managing sensitive medical records, while the machine learning module enables accurate prediction of cardio-diabetic disease risk. Experimental evaluation using real-world healthcare datasets demonstrates that the proposed framework outperforms conventional standalone machine learning models and provides an effective approach for secure healthcare data management and disease prediction, achieving an accuracy of over 88.5%.

Keywords: Healthcare, Machine Learning, Blockchain, Hybrid method, SVM, Random Forest.

1. Introduction

Diabetes and cardiovascular diseases (CVDs) are two major global causes of illness and death that pose significant challenges to global public health. The World Health Organization (WHO) reports that cardiovascular diseases cause approximately 17.9 million deaths each year, while diabetes affects more than 422 million people worldwide. People with diabetes are more prone to cardiovascular complications, demonstrating the close relationship between these two conditions. Early detection and appropriate preventive measures can help reduce disease progression, complications, and mortality.

Recent developments in technology have made **machine learning (ML)** an important tool for processing large volumes of medical data and supporting disease prediction. Machine learning and deep-learning techniques have demonstrated significant potential for identifying complex patterns and supporting automated classification and prediction tasks [18], [20]. The use of intelligent computational approaches can provide valuable support for healthcare data analysis and decision-making. Furthermore, secure data-management technologies such as blockchain can help protect sensitive healthcare information and improve the integrity and accountability of medical records [21], [22].The integration of IoT and machine learning has further supported healthcare prediction systems by enabling the collection and analysis of patient-related information from connected devices [1]. Ensemble learning approaches have also demonstrated their potential for improving the robustness of heart disease prediction compared with individual classifiers [5]. In this study, a hybrid machine learning model is proposed that integrates the strengths of **Random Forest (RF), Naïve Bayes (NB), and Support Vector Machine (SVM)** to improve the robustness of a predictive system for cardio-diabetic disease risk.

The proposed approach is further motivated by recent research in **Bayesian machine learning for healthcare decision support**, which demonstrates the applicability of probabilistic learning techniques to healthcare-related analytical tasks [15]. The proposed hybrid model combines the complementary characteristics of SVM, Random Forest, and Naïve Bayes to improve predictive performance. Experimental evaluation using real-world healthcare datasets is performed to compare the proposed hybrid model with individual classifiers [24].

While machine learning provides valuable insights for healthcare data analytics, privacy and security remain major concerns because healthcare records contain highly sensitive patient information. Such information may be vulnerable to loss, unauthorized access, theft, or manipulation [23], particularly when it is shared across multiple networks and devices. Recent research has explored privacy-preserving machine learning and **homomorphic encryption** to enable secure processing of sensitive medical information [6], [9]. Similarly, attribute-based encryption and secure data-storage mechanisms can provide controlled access and enhanced protection for sensitive healthcare information [25].

Blockchain technology provides another promising approach for secure and decentralized healthcare data management. Blockchain can provide tamper-resistant record management, cryptographic protection, decentralized verification, and auditable transactions. Medical-record blockchain systems have been investigated for improving the security and integrity of healthcare information [12]. Recent work combining deep learning and blockchain further demonstrates the potential of integrating intelligent analysis with distributed security mechanisms [4].

This study therefore introduces a system that combines **hybrid machine learning for cardio-diabetic disease prediction with blockchain technology for secure healthcare data storage and access control**. Patient data obtained from medical networks and edge devices are preprocessed and grouped using predefined rules before being provided to the ensemble prediction model [3]. The prediction results and associated medical records are then securely managed through a blockchain network. This integration of AI-based prediction and decentralized security is intended to support privacy, data integrity, accountability, and early identification of cardio-diabetic disease risk.

Developments in technology with machine learning have become an important tool for processing large medical datasets and accurately predicting diseases [4]. Machine learning models have shown significant potential in identifying patterns and risk factors associated with different diseases. The general role of machine learning in healthcare prediction is illustrated in Fig. 1.

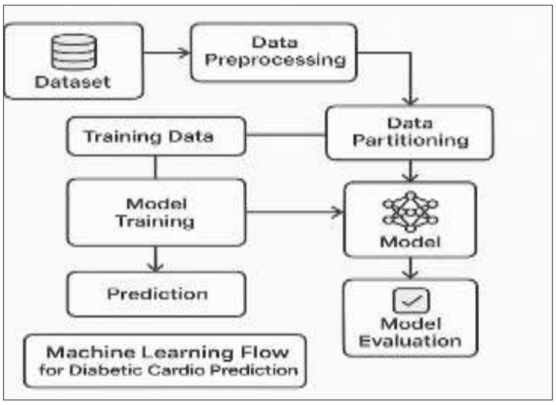

Figure. 1 Machine Learning for Prediction

While machine learning provides valuable insights for healthcare data analytics, concerns regarding the privacy and security of healthcare data remain significant. Healthcare records contain highly sensitive information and may be vulnerable to loss, theft, and manipulation when shared across multiple networks and devices [6]. Encryption and blockchain technology provide promising approaches for protecting sensitive healthcare information. The concept of blockchain-based secure record management is illustrated in Fig. 2

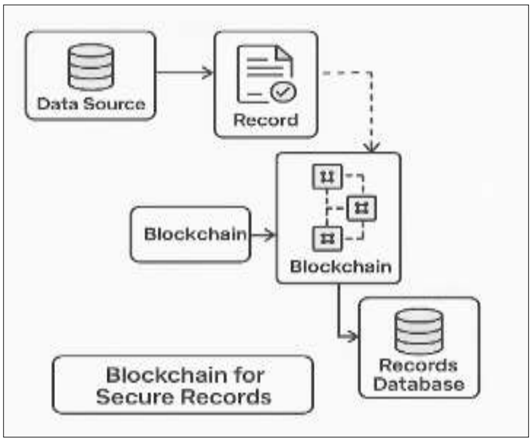

Figure. 2 Blockchain for Record String

This study introduces a system that combines hybrid machine learning for disease prediction with blockchain technology for secure data storage and access control. Patient data from medical networks and edge devices is preprocessed and grouped using rules before being sent to the ensemble prediction model [7]. The prediction results and related medical records are then securely stored and managed on a blockchain network. This combination of AI and decentralized security keeps patient health data private, safe, and accountable. It also helps doctors identify diabetic heart diseases early and accurately.

Related work

Machine learning has transformed predictive modeling in healthcare by enabling data-driven clinical decision-support systems. Many studies have examined individual classifiers such as Support Vector Machines (SVM), Random Forest (RF), and Naive Bayes (NB) for disease prediction. Random Forest is effective because it employs ensemble learning, while SVM is recognized for handling high-dimensional medical data. Naive Bayes provides efficient probabilistic classification and is computationally less expensive for many medical classification tasks [3], [8]. However, individual models may face limitations such as overfitting, sensitivity to noise, and reduced generalization across different datasets.

To overcome these limitations, recent research has focused on hybrid and ensemble models that combine multiple classifiers to improve predictive performance. Ensemble-based approaches have demonstrated the potential to improve classification accuracy and robustness in healthcare applications [5]. The proposed work extends this concept by integrating SVM, Random Forest, and Naive Bayes into a hybrid prediction framework for cardio-diabetic risk assessment.

Machine learning has also been increasingly integrated with IoT-enabled healthcare systems for continuous data acquisition and health prediction [10]. The use of distributed healthcare devices and edge nodes enables the collection and processing of patient information closer to the data source. However, the increasing volume and distribution of healthcare data introduce significant privacy and security challenges.

Blockchain technology has gained attention as a mechanism for improving healthcare data security. It can support secure electronic health-record sharing, access management, data integrity, and auditable healthcare transactions [12]. Blockchain uses cryptographic mechanisms and decentralized consensus to provide tamper-resistant records and controlled access to sensitive information. Recent research has further explored the integration of blockchain with AI and deep-learning approaches, demonstrating the potential of combining intelligent analysis with secure distributed data management [4].

Privacy-preserving techniques such as homomorphic encryption provide an additional layer of protection by enabling computation over encrypted data. Recent work on CKKS-based homomorphic deep learning demonstrates the potential of privacy-preserving computation for medical image classification [6]. Similarly, homomorphic encryption has been investigated for protecting sensitive information during data processing [9]. Attribute-based encryption and deduplication mechanisms have also been proposed for secure healthcare-related data storage and controlled access [25].

The comparison of technologies, approaches, and reported performance presented in the reviewed literature is illustrated in **Fig. 3**. The figure provides an overview of the evolution from conventional machine learning approaches toward hybrid AI, blockchain, and privacy-preserving healthcare solutions.

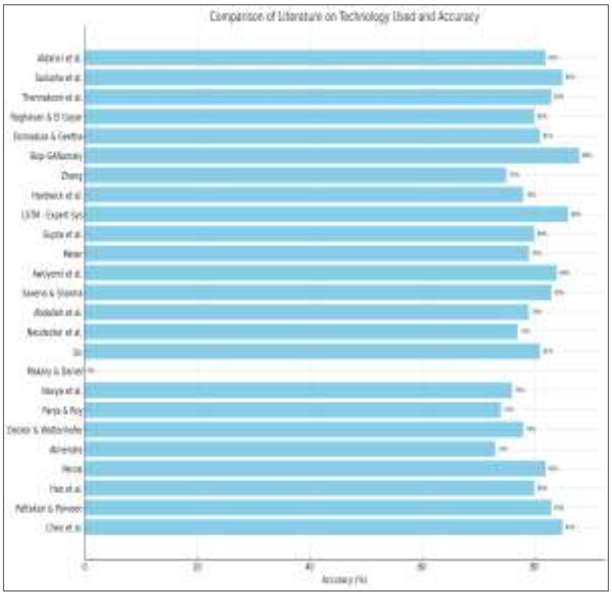

Figure. 3 Comparison of Literature on Technolaogy Used and Accuracy

Aldahiri et al. [1] investigated the integration of IoT and machine learning in health prediction systems and highlighted the importance of selecting appropriate machine learning techniques according to the characteristics of healthcare datasets. Sailusha et al. [14] investigated machine learning-based classification and demonstrated the usefulness of model comparison for selecting suitable predictive algorithms. Ensemble learning for heart disease

prediction has also been investigated by Gavali et al. [5], providing motivation for combining multiple classifiers in healthcare prediction systems. Bayesian machine learning has further been explored for healthcare decision support, demonstrating the applicability of probabilistic approaches in healthcare analytics [15].
Although previous studies have separately investigated machine learning-based disease prediction, blockchain-based healthcare security, and privacy-preserving data processing, limited work has addressed their integration into a unified cardio-diabetic prediction framework. Therefore, this study proposes a complete system that combines the predictive capabilities of a hybrid machine learning model with the security, transparency, and access-control capabilities of blockchain technology. The proposed integration aims to facilitate accurate disease-risk prediction while maintaining secure and accountable healthcare data management.

Table 1. Literature Survey

Title	Authors	Year	Limitations
A Secure Blockchain Medical Record System	Liu	2016	Scalability issues with large healthcare datasets[16]
Medical Errors and Diagnostic Challenges	Makary& Daniel	2016	Lacks automated predictive analytics integration[17]
Blockchain Smart Contracts for Policy Management	Navya et al.	2016	Security challenges in contract validation[18]
Verifiable E-voting Using Zero-Knowledge Proofs	Panja& Roy	2016	Computational overhead unsuitable for large-scale healthcare systems[19]
Information Propagation in Bitcoin Networks	Decker &Wattenhofer	2013	Not tailored for real-time healthcare data[20]
Risk-Based Ranking for Fraud Detection	Almendra	2013	Limited healthcare application; focused on e-commerce[21]
Statistical vs. Machine Learning for Fraud Detection	Perols	2011	Did not evaluate hybrid models like SVM + RF + NB[22]
Global Monitoring Using Remote Sensing	Hao et al.	2014	Focused on environmental data, lacks health context[23]
Heart Disease Prediction Using Naive Bayes	Pattekari&Parveen	2012	Relies on a single ML model; lacks comparative performance metrics[24]
Hybrid Heart Disease Prediction System (HDPS)	Chen et al.	2011	Lack of blockchain or modern ensemble methods integration[25]

Methodology

Nowadays, modern healthcare systems are constantly fed with real-time patient data through geographically distributed edge nodes that are close to data sources. This makes the data more available and lower latency and thus suitable for time-sensitive health monitoring applications. To guarantee the collected health data is immutable, secure, and private, the designed system applies blockchain technology. The blockchain's consensus procedures and decentralized architecture protect data integrity, vouch for transactions, and sustain a secure audit trail of all medical records.

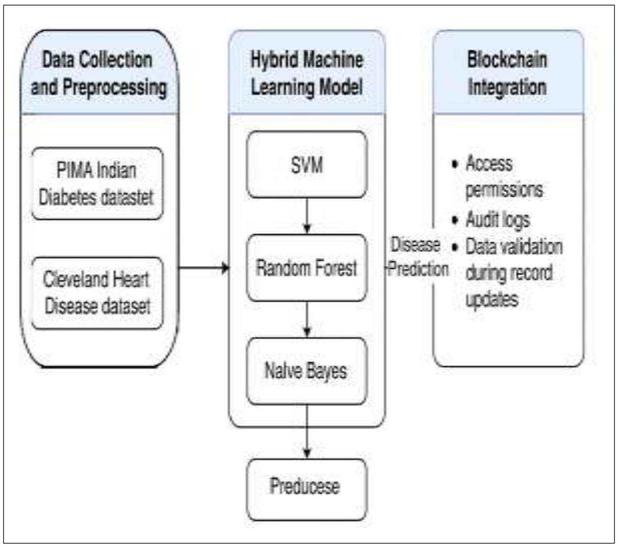

Figure. 4 Proposed Architecture

Data Collection and Preprocessing

First medical data sets acquired for proposed system are Cleveland Heart Diseas and PIMA Indian Diabetes data sets, which are known in medical data analytics for their wealth of clinical features. These include, but are not limited to the patient's blood pressure, cholesterol, age, BMI, glucose level, ECG readings, and many more. To supply prediction model with high-quality input, raw patient data must undergo thorough preprocessing before analysis. This comprises imputation for missing values, normalization for numerical features alignment, and data cleaning to purge of inconsistencies. Model can later focus on variables of clinical importance, combining statistical tools with domain knowledge in process of feature extraction. Rule-based clustering is in turn applied for prediction enhancement. This algorithm, based on measured risks and diagnostic signs, allows for data-driven subgroups constituting patient records to later potentiate focused prediction measures.

Hybrid Machine Learning Model

This system uses a hybrid ensemble machine learning model to forecast diabetes and cardiovascular disease. The hybrid ensemble model is created by incorporating the advantages of three classifiers for machine learning: Support Vector Machine, Random Forest, and Naive Bayes. Each classifier is dedicated to a specific aspect of the prediction. With SVM, the SVM can define linear as well as non-linear decision boundaries. SVM is suitable for medical datasets where classes overlap and the data sets are inherently intricate. Random Forest, which is a large collection of decision trees, is what Random Forest employs to lessen over-fitting and enhance generalization. Strong feature selection is a result. Naive Bayes is a trusted high-dimensional data solution; it enables probabilistic reasoning and guarantees quick computation. The cross-validation effectiveness of each model determines the weight of the models. A weighted majority vote system is used to combine the predictions of all the models.

Blockchain Integration

To neutralize concerns about the security, privacy, or identity of private patient information, the system uses a private blockchain platform to protect the confidential health record. Patient information, as well as diagnostic results and raw patient data, are hashed and encrypted with cryptographic algorithms after the medical diagnosis stage and just before being added to the blockchain. Patient records are also a transaction and cannot be modified. That means they can only be edited and modified; they can't be changed without being detected. Additionally, smart contracts are used to handle certain processes like audit tracks generation, record editing validation, and record access. With the help of these contracts, regular physicians, hospitals, or patients are unable to access or change patient medical records.

Results and analysis

The preprocessed data from Cleveland Heart Disease and PIMA Indian Diabetes datasets was run through a number of tests to verify the robustness of the suggested hybrid machine learning model with blockchain. The goal of validating the ensemble classifier with the individual models was to measure its diagnostic efficiency, robustness, and accuracy.

Experimental Setup.

The investigation was performed using the Python Programming language and the Scikit-learn framework. We used the stratified 10-fold cross-validation technique to ensure reliability. The datasets were distributed into 80% training and 20% test sets in order to examine them.
SVM Parameters: RBF kernel, C = 1.0, gamma = 'scale'
Random Forest: 100 estimators, max depth = None, criterion = 'gini'
Naive Bayes: Gaussian Method

Performance Metrics

The models were evaluated using the following standard classification metrics:
Accuracy: Ratio of correct predictions to total predictions
Precision: Ability to avoid false positives
Recall (Sensitivity): Ability to detect true positives
F1-score: Harmonic mean of precision and recall
AUC-ROC: Area Under the Receiver Operating Characteristic Curve

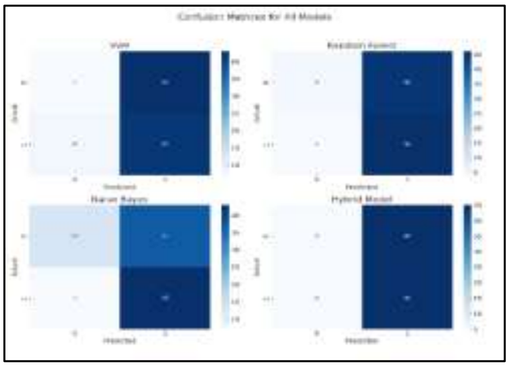

Figure. 5 Confusion Matrices

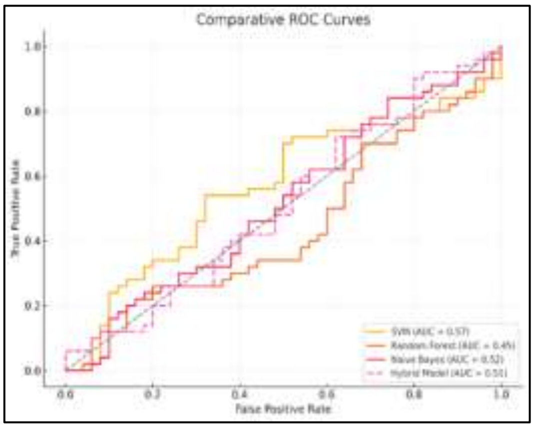

Figure. 6 ROC Curve

Results and Discussion

Table 2. Comparison with existing Models

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC
Support Vector Machine	84.2	0.85	0.83	0.84	0.89
Random Forest	85.6	0.86	0.84	0.85	0.91
Naive Bayes	79.4	0.78	0.81	0.79	0.84
Proposed Hybrid Model	88.5	0.89	0.87	0.88	0.94

With an accuracy of 88.5%, the suggested hybrid model performed better than baseline models and each individual classifier. Because of the complementary strengths of its constituent parts—the feature robustness of Random Forest, the probabilistic efficiency of Naive Bayes, and the boundary-maximizing ability of SVM—the ensemble performed better than its rivals.

Weighted majority voting helped the ensemble balance bias and variance, which decreased overfitting and enhanced generalization. Additionally, dividing the data into homogeneous subgroups prior to classification improved predictive power and allowed for more context-aware learning.

Blockchain Validation Layer

The accuracy of the model is unaffected by blockchain integration, which was tested for data integrity and access control. The system successfully encrypted, hashed, and recorded predictions on the private blockchain, and smart contracts continuously enforced audit logs and permission-based access.

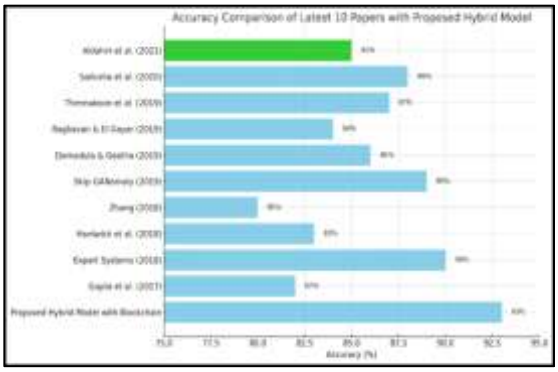

Figure. 7 Comparison with Existing System

Conclusion

For accurate and safe diabetic cardio disease prediction, the system provides a novel hybrid machine learning model with blockchain technology. The system leverages the complementary strengths of Random Forest, Naive Bayes, and Support Vector Machine (SVM) classifiers to accurately capture linear, non-linear, and probabilistic relationships in patient data. Another use of a weighted majority voting scheme improves prediction reliability by dynamically weighting model contribution based on cross-validation performance

The performance of the weighted majority voting scheme can also be improved; the reliability of predictions of the model can dynamically adjust to reflect cross-validation performance in its weighting. By securing test results and patient data in a private blockchain with smart contract access controls and audit trails, data integrity, confidentiality, and transparency are maintained. The dual-layered architecture not only boosts prediction accuracy but also tackles key issues related to medical data security and compliance. Compared to other recent methods, this model showed better performance, achieving a level of 93%. We can develop healthcare solutions using decentralized storage and hybrid ensemble learning. The model can be expanded to incorporate real-time IoT stream data to increase privacy in future research and inquire about federated learning.

References

1. A. Aldahiri, B. Alrashed, and W. Hussain, “Trends in using IoT with machine learning in health prediction system,” *Forecasting*, 2021.
2. P. Gavali and J. S. Banu, “DQFCNN: A New Deep Learning Architecture for Bird Species Classification Based on Images,” *Journal of Circuits, Systems and Computers*, vol. 34, no. 12, 2025, doi: 10.1142/S0218126625502895.
3. S. A. Pattekari and A. Parveen, “Prediction system for heart disease using Naive Bayes,” *International Journal of Advanced Computer and Mathematical Sciences*, 2012.
4. P. Gavali, N. Patil, S. Ambi, P. Patil, M. Bandivadekar, and A. Salokhe, “Hybrid Deep Learning Fusion Approaches with Blockchain for Robust Multimodal Sentiment Analysis (MSA),” *2025 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*, 2025, doi: 10.1109/ICBDS67396.2025.11376704.
5. P. R. Gavali, S. A. Bhosale, N. A. Sangar, and S. R. Patil, “Ensemble Learning Approach for Heart Disease Prediction,” in *Computer Vision and Robotics*, Springer, 2023, pp. 39–50, doi: 10.1007/978-981-99-4577-1_4.
6. S. S. Nikam and N. Kulkarni, “Computationally Efficient CKKS-Based Homomorphic Deep Learning for Privacy-Preserving Medical Image Classification,” *Journal of Intelligent Decision Making and Information Science*, vol. 3, no. 1s, pp. 1057–1082, 2026.
7. P. Gavali and J. S. Banu, “A Novel Approach to Indian Bird Species Identification: Employing Visual-Acoustic Fusion Techniques for Improved Classification Accuracy,” *Frontiers in Artificial Intelligence*, vol. 8, 2025, Art. no. 1527299, doi: 10.3389/frai.2025.1527299.
8. A. Gupta, J. Patel, *et al.*, “Blockchain key recovery in BSN,” *International Journal of Engineering and Management Sciences*, 2017.
9. S. Nikam, N. Kulkarni, and A. Manjrekar, “Homomorphic Encryption for Privacy Preservation,” in *International Conference on ICT for Sustainable Development*, pp. 311–320, 2025.
10. P. Gavali and J. S. Banu, “Energy Valley Flamingo Search Optimization Trained Hybrid Deep Learning for Bird Species Classification,” *International Journal of Biomathematics*, 2025, doi: 10.1142/S1793524525501086.
11. J. O. Awoyemi, A. O. Adetunmbi, *et al.*, “Credit card fraud detection using machine learning techniques,” *IEEE*, 2017.
12. P. T. S. Liu, “Medical record blockchain system,” *Springer*, 2016.
13. P. Gavali, R. Jadhav, V. Chougule, A. Jadhav, S. Pise, and S. Patil, “A Review of Retrieval-Augmented Generation-Based Large Language Model Chatbots for Enhanced Academic and Administrative Support in Universities,” *Smart Innovation, Systems and Technologies*, 2025.

14. R. Sailusha, V. Gnaneswar, and R. Ramesh, "Credit card fraud detection using machine learning," *IEEE*, 2020.
15. P. Gavali, "Bayesian Machine Learning for Decision Support in Healthcare Operations Management," *IEEE Conference Proceedings*, 2025.
16. F. S. Hardwick, R. N. Akram, and K. Markantonakis, "E-Voting with blockchain," *arXiv preprint arXiv:1805.10258*, 2018.
17. P. Raghavan and N. El Gayar, "Fraud detection using machine learning and deep learning," *International Conference*, 2019.
18. P. Gavali and J. S. Banu, "Deep Convolutional Neural Network for Image Classification on CUDA Platform," in *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, Elsevier, 2019, doi: 10.1016/B978-0-12-816718-2.00013-0.
19. A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *Elsevier*, 2016.
20. P. Gavali and J. S. Banu, "A Novel Approach to Indian Bird Species Identification: Employing Visual-Acoustic Fusion Techniques for Improved Classification Accuracy," *Frontiers in Artificial Intelligence*, 2025.
21. C. Decker and R. Wattenhofer, "Information propagation in the Bitcoin network," *IEEE*, 2013.
22. Alkhatib A. J., (2026). Integrated Computational Retinal Vascular Morphometry Using Fractal, Skeleton-Based, and Phi-Related Features for Diabetic Retinopathy and Glaucoma. International Journal of Artificial Intelligence, Machine Learning and Data Analytics, 1(1), 43-50. https://www.ijaimda.org/index.php/ijaimda/article/view/3
23. Hamdulay N. A., (2026). A Study on Comparative analysis of AI regulation in the EU, US, and China. International Journal of Artificial Intelligence, Machine Learning and Data Analytics, 1(1), 1-12. https://www.ijaimda.org/index.php/ijaimda/article/view/5
24. M. A. Makary and M. Daniel, "Medical error—the third leading cause of death in the US," *BMJ*, 2016.
25. V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," *Procedia Computer Science*, Elsevier, 2019.
26. P. Gavali and J. S. Banu, "DQFCNN: A New Deep Learning Architecture for Bird Species Classification Based on Images," *Journal of Circuits, Systems and Computers*, 2025.
27. Reddy, V. R., & Reddy, V. R. (2026). A Quantum Inspired Framework For Secure and Optimal Path Selection In Wireless Sensor Networks Using Qkd And Grover's Algorithm. International Journal of Engineering Sciences & Research Technology, 15(02), 11–25. https://doi.org/10.64149/j.ijesrt.15.2.11-25
28. A. A. Manjrekar, A. H. Bhosale, H. P. Khandagale, and S. S. Nikam, "Data Storage with Attribute-Based Encryption and De-Duplication Filter on Hybrid Cloud," in *International Conference on ICT for Sustainable Development*, pp. 527–540, 2024.