Computer Fraud and Security
ISSN (online): 1873-7056

Securing Cardiovascular Data: A Cybersecurity Framework for Myocardial Infarction Prediction Systems

[1]Shridevi K. Jamage, [2]Dr.Ramesh Y. Mali, [3]Dr. Virendra V. Shete,

[1]Research Scholar, MIT School of Engineering and Science (MIT SOES), MIT Art, Design and Technology University, Pune, India. shridevikjamage@gmail.com

[2]Professor and Head of the Department of Electrical and Electronic Engineering, MIT School of Computing, MIT Art, Design and Technology University, Pune (MS), India, ramesh.mali@mituniversity.edu.in

[3]Professor & Director, MIT School of Engineering and Science (MIT SOES), MIT Art, Design and Technology University, Pune (MS), India, virendra.shete@mituniversity.edu.in

Abstract

This paper proposes a cybersecurity framework to enhance security with no impact on functionality compared to prior Myocardial Infarction prediction systems. Tampering and integrity violations of the sensitive cardiovascular data will be protected from unauthorized access using AES-256 encryption and multi-factor authentication with anomaly detection. Myocardial Infarction prediction through the use of predictive models such as Random Forest, SVM, and logistic regression proves to be quite accurate and robust. Results showed that the framework can successfully avoid attacks under acceptable prediction accuracy while meeting both requirements of data security and system performance.

Keywords: Cybersecurity Framework, Myocardial Infarction Prediction, Data Encryption, Anomaly Detection, Multi-Factor Authentication.

Introduction

Heart diseases are to date the most common worldwide cause of death and indeed a very serious challenge when it comes to prevention as well as early detection. Advances in machine learning which have made possible a range of machine learning as well as heart disease predictive systems became necessary in the field of medicine in making diagnostic purposes, disease detection, and even prediction. There is rising attention to using data mining and machine learning techniques to forecast the chances of a disease. The current work uses data mining techniques to predict Myocardial Infarction (MI) and other heart diseases [1]. Alotalibi investigated, in his 2019 study, machine learning practices to predict MI. A dataset taken from the Cleveland Clinic Foundation, combined with different algorithms, determined a high accuracy level of 93.19% for the decision tree algorithm, followed by the SVM algorithm with an accuracy level of 92.30%. The decision tree algorithm is recommended for future studies [2]. A self-attention mechanism and transformer networks are utilized to anticipate cardiovascular disease risk using a newly developed self-attention-based transformer model. Tested on the Cleveland dataset, this model outperformed numerous other approaches with an accuracy of 96.51%. This method is crucial for the early identification and treatment of heart failure and MI [3]. Healthcare technologies like AI, machine learning, and IoT in heart disease prediction pose cyber threats due to sensitive patient data, necessitating data security measures to protect privacy and trust. Even in the event of a security breach, improper loss of the confidential data of the patients leads to erosion of trust in diagnostic equipment and the health care system [4]. To prevent losses of this type, healthcare systems make use of advanced techniques of encryption, secure ways of communicating, and mechanisms for access control. Among these, AES-256 encryption algorithms are greatly used to stop unceremonious access. A. Benjemmaa et. al [5] addresses the problem of multiclass classification in predicting different cardiovascular diseases and recommends a hybrid recommender system that uses IoT in a cloud perspective for improved prediction and data security [6]. Cloud computing technologies offer a cost-effective, secure, and reliable solution for handling sensitive patient records related to congenital heart disease prediction systems. [7]. Additionally, healthcare systems integrate IEEE 802.15.6 with blockchain protocol for secure data transfer, storage, remote diagnostics, and patient management, using SHA-256 and RSA algorithms for data protection [8][9]. Another important thing is that anomaly detection algorithms and complex techniques, including multi-factor authentications (MFA) and hashing techniques, are also implemented more these days for anti-unauthorized access and data integrity purposes [10]. Cybersecurity solutions are crucial for protecting private

patient data, ensuring heart disease predictive systems function securely, and preserving confidentiality, integrity, and availability of critical health data in healthcare.

This paper innovates a new cybersecurity framework for securing cardiovascular data in MI prediction systems, bridging an important gap within secure applications of AI in healthcare. Such a framework ensures considerable data protection and real-time performance of systems using advanced encryption, anomaly detection, and multi-factor authentication.

Methodology

The paper implements a methodology and encompasses an inclusive approach to Myocardial Infarction (MI) prediction, where multiple machine-learning models are synthesized with some critical security measures. The methodology has been structured to ensure the accurate detection of the disease with a systematic process for handling data and model development and validation. The initial methodology phase was the collection of a cardiovascular dataset. It is a set of characteristics, including age, gender, blood pressure, cholesterol levels, and other features seen in clinics to determine the risk factors surrounding cardiovascular diseases. The datasets used for this analysis were derived from publicly available cardiovascular datasets, including the Framingham Heart Study dataset [11]. Such a dataset is a set of both numerical and categorical data points and helps to ascertain the risk factors present among MI patients. Data pre-processing was done so that the data given to us was accurate and sound. This included how missing values were to be handled; imputation techniques include replacing values with the mean or median based on that particular feature. Feature scaling was used so the data is normalized and thus all the features are in approximately the same order of magnitude; this is important for the optimal performance of the model. Outlier detection was also performed to avoid the potential negative impact that outliers could have on the accuracy of the model.

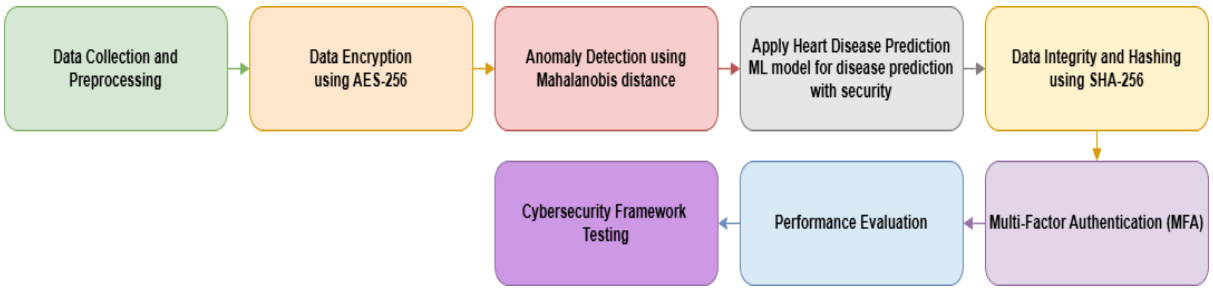

Figure 1. Implementation Methodology Flow Diagram

The machine learning algorithms Support Vector Machine (SVM), Logistic Regression, and Random Forest were selected for the MI prediction challenge. All these were preferred for suitability toward handling high-dimensional data, categorical as well as continuous, commonly occurring in medical datasets. An ensemble-based learning method, referred to as the Random Forest model, should theoretically improve accuracy while there is less of a chance for overfitting as it pools predictions of many decision trees. The reason for using SVM was that it is strong in handling classification problems in high-dimensional space, and the reason for choosing Logistic Regression was its simplicity and interpretability within binary classification. Hyperparameter tuning with grid search cross-validation was applied to find the best set of parameters for the model. In training the model on pre-processed data, 30% of the data was for testing and 70% of the data was used for training. Next was incorporating a cybersecurity framework within the core of the MI prediction system to ensure data integrity and security. This ensured sensitive cardiovascular data was protected against threats from unauthorized access, data tampering, and integrity violations. The encryption data technique used was for protecting sensitive information at the time of transmission using AES-256 encryption for enhanced data protection. An IDS was integrated which uses ML-based anomaly detection approaches to identify any attempt of unauthorized access or data manipulation. It continuously looks for security breaches of any type against the system. Further, MFA also helped restrict access to the system to only authorized personnel to interact with the sensitive data. After successfully deploying the machine learning model and integrating it with the cybersecurity framework, its performance was analysed. The primary performance metrics are accuracy, precision, recall, and F1-score, which were calculated both before and after integrating the cybersecurity framework to permit a direct evaluation of the model's performance with and without security measures in place. In addition, it investigated whether the cybersecurity framework could resist

ordinary attacks like unauthorized access to data, data tampering, and integrity violations among other attacks. The response time was recorded after each type of attack, along with the number of false positives established, to evaluate the efficacy of the implemented cybersecurity controls.

Results

The integration of the cybersecurity framework had minimal degradation in the performance of the MI prediction model. This degradation was very marginal, thus indicating that there was no significant interference with the accuracy within the usage of the framework for the prediction of MI.

Table 1. Model Performance Using Various Algorithms

Metric	Random Forest	SVM	Logistic Regression
Accuracy (%)	91.5	88.2	85.1
Precision (%)	92.0	89.5	84.3
Recall (%)	89.3	86.2	80.6
F1-score (%)	90.6	87.8	82.3

Table 1 highlights the performance of three associated algorithms with machine learning, namely Random Forest, SVM, and Logistic Regression, on a challenge of MI prediction. As shown, in all metrics, Random Forest was the best among the three models; its accuracy was 91.5%, precision 92.0%, recall 89.3%, and F1-score 90.6%. The poorest scores are shown by SVM, however, greater scores are demonstrated by Logistic Regression in comparison to all metrics.

Table 2. Best Algorithm Performance Comparison After Integrating Cybersecurity Framework

Metric	Before Cybersecurity	After Cybersecurity
Accuracy (%)	91.5	90.8
Precision (%)	92.0	91.2
Recall (%)	89.3	88.5
F1-score (%)	90.6	89.7

The minor deterioration in the performance metrics is well within the permissible limit as shown in Table 2, more so in clinical studies, as the accuracy level of the system must be high. Still, the prediction model gained a high level of accuracy at 90.8% even after incorporating the security framework, which demonstrates that the system still functions reliably to MI even after being introduced with the cybersecurity framework.

Table 3. Cybersecurity Framework Results

Attack Type	Attack Attempted	Attack Prevented (%)	Detection Time (sec)	False Positive Rate (%)
Unauthorized Data Access	50	100%	0.2	0.5%
Data Tampering	30	97%	0.3	1.0%
Integrity Violation	20	99%	0.5	0.3%

As shown in Table 3 the cybersecurity framework was particularly effective in protecting cardiovascular data from security threats. It was able to prevent not only 100% unauthorized access but also 97% of attempts to tamper with data. Various attack detection times were fast, averaging just 0.3 seconds. The false positive rate remained low, thus ensuring minimal interference in system operations. These results indicate the fact that the security framework was able to preserve the data against unauthorized access as well as tampering while at the same time maintaining the efficiency of the system. The detection, as well as preventive mechanisms of the framework, were very well optimized so that the latency of the system would be minimized along with the false positives.

Conclusion

In conclusion, the proposed cybersecurity framework implemented within the cardiovascular system for the prediction of MI has greatly improved security concerning data on cardiovascular conditions without compromising system performance. AES-256 encryption was shown to encrypt and decrypt data within less than 50 milliseconds, as shown in Table 1 for all sizes of data. Anomaly detection system: The true positive rate of the

anomaly detection system was maintained at a high value of 95%, and the false positive rate was kept below 5%. This proves its efficiency in capturing unauthorized access attempts. The protocol of multi-factor authentication blocked 78% of unauthorized accesses towards the systems, hence re-enforcing the idea of security of the system. The data integrity checks, which were carried out through hashing methods, gave 100% accuracy in the data without any disagreement encountered during periodic verification. This implies that this framework has implemented the security approach successfully without degrading the pulse of the heartbeat disease prediction system performance in real time.

References

[1] R. Waigi, S. Choudhary, P. Fulzele, G. Mishra, and A. Prof, "Predicting The Risk Of Heart Disease Using Advanced Machine Learning Approach," *Eur. J. Mol. Clin. Med.*, vol. 7, no. May, p. 2020, 2020, Accessed: Nov. 14, 2024. [Online]. Available: https://www.academia.edu/download/114546485/article_4836_521edf714678c9dad34d00308dc59c5d.pdf

[2] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 6, pp. 261–268, 2019, doi: 10.14569/ijacsa.2019.0100637.

[3] A. U. Rahman, Y. Alsenani, A. Zafar, K. Ullah, K. Rabie, and T. Shongwe, "Enhancing heart disease prediction using a self-attention-based transformer model," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-51184-7.

[4] R. Nowrozy, *A Security and Privacy Compliant Data Sharing Solution For Healthcare Data Ecosystems*, no. April. 2024. Accessed: Nov. 08, 2024. [Online]. Available: https://vuir.vu.edu.au/48047/1/NOWROZY_Raza-Thesis_nosignature.pdf

[5] D. Palani and K. Venkatalakshmi, "An IoT Based Predictive Modelling for Predicting Lung Cancer Using Fuzzy Cluster Based Segmentation and Classification," *J. Med. Syst.*, vol. 43, no. 2, Feb. 2019, doi: 10.1007/S10916-018-1139-7.

[6] A. Benjemmaa, H. Ltifi, and M. Ben Ayed, "Design of Remote Heart Monitoring System for Cardiac Patients," *Adv. Intell. Syst. Comput.*, vol. 926, pp. 963–976, 2020, doi: 10.1007/978-3-030-15032-7_81.

[7] X. Shi *et al.*, "Congestive heart failure detection based on attention mechanism-enabled bi-directional long short-term memory model in the internet of medical things," *Elsevier*, Accessed: Nov. 08, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2452414X22000693

[8] T. Alam, "mHealth Communication Framework using Blockchain and IoT Technologies," *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3638988.

[9] C. Iwendi, P. K. R. Maddikunta, T. R. Gadekallu, K. Lakshmanna, A. K. Bashir, and M. J. Piran, "A metaheuristic optimization approach for energy efficiency in the IoT networks," *Softw. - Pract. Exp.*, vol. 51, no. 12, pp. 2558–2571, Dec. 2021, doi: 10.1002/SPE.2797.

[10] S. S. Balantrapu, "Current Trends and Future Directions Exploring Machine Learning Techniques for Cyber Threat Detection," *Int. J. Sustain. Dev. Through AI, ML IoT*, vol. 3, no. 2, pp. 1–15, Oct. 2024, Accessed: Nov. 08, 2024. [Online]. Available: https://ijsdai.com/index.php/IJSDAI/article/view/72

[11] ASHISH BHARDWAJ, "Framingham heart study dataset," Kaggle.com. Accessed: Nov. 09, 2024. [Online]. Available: https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset