Security and Communication Networks

Security and Networking for Healthcare Information Exchange and Storage in the Big Data Ecosystem

Lead Guest Editor: Chinmay Chakraborty
Guest Editors: Hemant Ghayvat and Dr. Celestine Iwendi

Security and Networking for Healthcare Information Exchange and Storage in the Big Data Ecosystem

Security and Communication Networks

Security and Networking for Healthcare Information Exchange and Storage in the Big Data Ecosystem

Lead Guest Editor: Chinmay Chakraborty
Guest Editors: Hemant Ghayvat and Dr. Celestine Iwendi

Chief Editor

Associate Editors

Academic Editors

Contents

Retracted: Electronic Health Record Monitoring System and Data Security Using Blockchain Technology
Security and Communication Networks
Retraction (1 page), Article ID 9856473, Volume 2023 (2023)

Retracted: Machine-to-Machine Communication for Device Identification and Classification in Secure Telerobotics Surgery
Security and Communication Networks
Retraction (1 page), Article ID 9854329, Volume 2023 (2023)

Retracted: Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques
Security and Communication Networks
Retraction (1 page), Article ID 9758612, Volume 2023 (2023)

Retracted: A Chaotic-Map-Based Password-Authenticated Key Exchange Protocol for Telecare Medicine Information Systems
Security and Communication Networks
Retraction (1 page), Article ID 9839023, Volume 2023 (2023)

Retracted: Method of Cumulative Anomaly Identification for Security Database Based on Discrete Markov chain
Security and Communication Networks
Retraction (1 page), Article ID 9831958, Volume 2023 (2023)

Retracted: Prediction of IoT Traffic Using the Gated Recurrent Unit Neural Network- (GRU-NN-) Based Predictive Model
Security and Communication Networks
Retraction (1 page), Article ID 9820840, Volume 2023 (2023)

Retracted: Edge Location Method for Multidimensional Image Based on Edge Symmetry Algorithm
Security and Communication Networks
Retraction (1 page), Article ID 9809853, Volume 2023 (2023)

Retracted: Fast Extraction Algorithm for Local Edge Features of Super-Resolution Image
Security and Communication Networks
Retraction (1 page), Article ID 9783897, Volume 2023 (2023)

Retracted: A Lightweight Proxy Re-Encryption Approach with Certificate-Based and Incremental Cryptography for Fog-Enabled E-Healthcare
Security and Communication Networks
Retraction (1 page), Article ID 9780906, Volume 2023 (2023)

Retracted: Enhanced Lorenz-Chaotic Encryption Method for Partial Medical Image Encryption and Data Hiding in Big Data Healthcare
Security and Communication Networks
Retraction (1 page), Article ID 9801983, Volume 2023 (2023)

Retracted: Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm
Security and Communication Networks
Retraction (1 page), Article ID 9834267, Volume 2023 (2023)

[Retracted] Method of Cumulative Anomaly Identification for Security Database Based on Discrete Markov chain
Zhiying Xu, Ting Yang, and Moslem Lari Najafi
Research Article (10 pages), Article ID 5113725, Volume 2022 (2022)

[Retracted] Enhanced Lorenz-Chaotic Encryption Method for Partial Medical Image Encryption and Data Hiding in Big Data Healthcare
P. Rashmi, M. C. Supriya, and Qiaozhi Hua
Research Article (9 pages), Article ID 9363377, Volume 2022 (2022)

[Retracted] Electronic Health Record Monitoring System and Data Security Using Blockchain Technology
Kazi Tamzid Akhter Md Hasib, Ixion Chowdhury, Saadman Sakib, Mohammad Monirujjaman Khan, Nawal Alsufyani, Abdulmajeed Alsufyani, and Sami Bourouis
Research Article (15 pages), Article ID 2366632, Volume 2022 (2022)

[Retracted] Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm
Jiangbo Yu
Research Article (13 pages), Article ID 3272603, Volume 2022 (2022)

[Retracted] Fast Extraction Algorithm for Local Edge Features of Super-Resolution Image
Feng Chen and Botao Yang
Research Article (10 pages), Article ID 8801978, Volume 2022 (2022)

[Retracted] A Chaotic-Map-Based Password-Authenticated Key Exchange Protocol for Telecare Medicine Information Systems
Yanrong Lu and Dawei Zhao
Research Article (8 pages), Article ID 7568538, Volume 2021 (2021)

[Retracted] A Lightweight Proxy Re-Encryption Approach with Certificate-Based and Incremental Cryptography for Fog-Enabled E-Healthcare
Junaid Hassan, Danish Shehzad, Insaf Ullah, Fahad Algarni, Muhammad Umar Aftab, Muhammad Asghar Khan, and M. Irfan Uddin
Research Article (17 pages), Article ID 9363824, Volume 2021 (2021)

[Retracted] Prediction of IoT Traffic Using the Gated Recurrent Unit Neural Network- (GRU-NN-) Based Predictive Model
Sonali Appasaheb Patil, L. Arun Raj, and Bhupesh Kumar Singh
Research Article (7 pages), Article ID 1425732, Volume 2021 (2021)

Contents

[Retracted] Edge Location Method for Multidimensional Image Based on Edge Symmetry Algorithm
Chen Li
Research Article (11 pages), Article ID 1326357, Volume 2021 (2021)

[Retracted] Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques
Abolfazl Mehbodniya, Izhar Alam, Sagar Pande, Rahul Neware, Kantilal Pitambar Rane, Mohammad Shabaz, and Mangena Venu Madhavan
Research Article (8 pages), Article ID 9293877, Volume 2021 (2021)

[Retracted] Machine-to-Machine Communication for Device Identification and Classification in Secure Telerobotics Surgery
Meghana P. Lokhande, Dipti Durgesh Patil, Lalit V. Patil, and Mohammad Shabaz
Research Article (16 pages), Article ID 5287514, Volume 2021 (2021)

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9856473, 1 page
https://doi.org/10.1155/2023/9856473

Retraction

Retracted: Electronic Health Record Monitoring System and Data Security Using Blockchain Technology

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] K. T. Akhter Md Hasib, I. Chowdhury, S. Sakib et al., "Electronic Health Record Monitoring System and Data Security Using Blockchain Technology," *Security and Communication Networks*, vol. 2022, Article ID 2366632, 15 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9854329, 1 page
https://doi.org/10.1155/2023/9854329

Retraction

Retracted: Machine-to-Machine Communication for Device Identification and Classification in Secure Telerobotics Surgery

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] M. P. Lokhande, D. D. Patil, L. V. Patil, and M. Shabaz, "Machine-to-Machine Communication for Device Identification and Classification in Secure Telerobotics Surgery," *Security and Communication Networks*, vol. 2021, Article ID 5287514, 16 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9758612, 1 page
https://doi.org/10.1155/2023/9758612

Retraction

Retracted: Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] A. Mehbodniya, I. Alam, S. Pande et al., "Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques," *Security and Communication Networks*, vol. 2021, Article ID 9293877, 8 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9839023, 1 page
https://doi.org/10.1155/2023/9839023

Retraction

Retracted: A Chaotic-Map-Based Password-Authenticated Key Exchange Protocol for Telecare Medicine Information Systems

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Lu and D. Zhao, "A Chaotic-Map-Based Password-Authenticated Key Exchange Protocol for Telecare Medicine Information Systems," *Security and Communication Networks*, vol. 2021, Article ID 7568538, 8 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9831958, 1 page
https://doi.org/10.1155/2023/9831958

Retraction

Retracted: Method of Cumulative Anomaly Identification for Security Database Based on Discrete Markov chain

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Xu, T. Yang, and M. L. Najafi, "Method of Cumulative Anomaly Identification for Security Database Based on Discrete Markov chain," *Security and Communication Networks*, vol. 2022, Article ID 5113725, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9820840, 1 page
https://doi.org/10.1155/2023/9820840

Retraction

Retracted: Prediction of IoT Traffic Using the Gated Recurrent Unit Neural Network- (GRU-NN-) Based Predictive Model

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. A. Patil, L. A. Raj, and B. K. Singh, “Prediction of IoT Traffic Using the Gated Recurrent Unit Neural Network- (GRU-NN-) Based Predictive Model,” *Security and Communication Networks*, vol. 2021, Article ID 1425732, 7 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9809853, 1 page
https://doi.org/10.1155/2023/9809853

Retraction

Retracted: Edge Location Method for Multidimensional Image Based on Edge Symmetry Algorithm

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C. Li, "Edge Location Method for Multidimensional Image Based on Edge Symmetry Algorithm," *Security and Communication Networks*, vol. 2021, Article ID 1326357, 11 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9783897, 1 page
https://doi.org/10.1155/2023/9783897

Retraction

Retracted: Fast Extraction Algorithm for Local Edge Features of Super-Resolution Image

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] F. Chen and B. Yang, “Fast Extraction Algorithm for Local Edge Features of Super-Resolution Image,” *Security and Communication Networks*, vol. 2022, Article ID 8801978, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9780906, 1 page
https://doi.org/10.1155/2023/9780906

Retraction

Retracted: A Lightweight Proxy Re-Encryption Approach with Certificate-Based and Incremental Cryptography for Fog-Enabled E-Healthcare

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Hassan, D. Shehzad, I. Ullah et al., "A Lightweight Proxy Re-Encryption Approach with Certificate-Based and Incremental Cryptography for Fog-Enabled E-Healthcare," *Security and Communication Networks*, vol. 2021, Article ID 9363824, 17 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9801983, 1 page
https://doi.org/10.1155/2023/9801983

Retraction

Retracted: Enhanced Lorenz-Chaotic Encryption Method for Partial Medical Image Encryption and Data Hiding in Big Data Healthcare

Security and Communication Networks

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] P. Rashmi, M. C. Supriya, and Q. Hua, "Enhanced Lorenz-Chaotic Encryption Method for Partial Medical Image Encryption and Data Hiding in Big Data Healthcare," *Security and Communication Networks*, vol. 2022, Article ID 9363377, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9834267, 1 page
https://doi.org/10.1155/2023/9834267

Retraction

Retracted: Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm

Security and Communication Networks

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Yu, "Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 3272603, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9831958, 1 page
https://doi.org/10.1155/2023/9831958

Retraction

Retracted: Method of Cumulative Anomaly Identification for Security Database Based on Discrete Markov chain

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Xu, T. Yang, and M. L. Najafi, “Method of Cumulative Anomaly Identification for Security Database Based on Discrete Markov chain,” *Security and Communication Networks*, vol. 2022, Article ID 5113725, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 5113725, 10 pages
https://doi.org/10.1155/2022/5113725

Research Article

Method of Cumulative Anomaly Identification for Security Database Based on Discrete Markov chain

Zhiying Xu,[1] **Ting Yang**,[1] **and Moslem Lari Najafi**[2]

[1] *Shaoxing University Yuanpei College, Shaoxing 312000, China*
[2] *Pharmaceutical Science and Cosmetic Products Research Center, Kerman University of Medical Sciences, Kerman, Iran*

Correspondence should be addressed to Ting Yang; yangting663123@163.com and Moslem Lari Najafi; m.larinajafi@kmu.ac.ir

Received 7 July 2021; Revised 30 August 2021; Accepted 2 February 2022; Published 21 March 2022

Academic Editor: Chinmay Chakraborty

There exists an enormous volume of data in the database system, which is accountable for the storage of data and organization of data. The intruders can breach the security system of database and steal the important information. Therefore, it is of great significance to carry out the cumulative anomaly identification of the security database. In view of the shortcomings of traditional anomaly detection methods in detection performance and poor effect of anomaly recognition, this paper proposes a cumulative anomaly recognition method based on discrete Markov chain for security database. First, the sniffer is used to read the user access behaviour data, and then, it is processed, that is, standardized processing. Then, the segmentation method is used to extract the user behaviour features, and the normal feature data and abnormal feature data are obtained. Finally, the state sequence generated by the discrete Markov chain is used to calculate the state probability, which is used to evaluate the abnormal process behaviour. The proposed method in this paper is based on the Markov chain and can be used for better anomaly recognition. The results are obtained in terms of sensitivity score, precision score, and F1-score. The results are also compared with the results obtained by using some of the state-of-the-art traditional techniques. The comparison clearly indicated that the proposed method is more effective as compared to the tradition methods. The proposed method has the highest F1-score of 0.8586, and then the traditional methods have F1-scores of 0.7233, 0.8236, and 0.7562 for methods 1, 2, and 3, respectively.

1. Introduction

Data are becoming a powerful tool for businesses and organizations. Some of these data are worth millions of dollars, and companies take great effort to limit who has access to them, both within the company and outside the company [1]. When it comes to concerns of privacy of personal data, data security is also critical, and firms and organizations that manage such data must provide solid guarantees about the confidentiality of these data to comply with legal requirements and policies [2]. In the context of the information security system, data security plays a critical role. The availability of the data allows for an agile reaction to consumers searching for improved service that is critical for the administration of a business [3]. The proper deployment of a database in public organizations aids in the achievement of the goal, thus security measures must be implemented. Stealing of relevant information, duplication of records, denial of service, and the inability to get information on time are all issues that public entities face [4]. Cyber attackers are seeking a way in through a system breach and have a variety of tools for gaining access to an organization's systems or databases [5]. Theft of information, duplication of data, denial of service, and the difficulty to obtain information particularly heathcare data on time are all difficulties that public bodies confront [6]. Cyber intruders and hackers are looking for innovative ways to breach the security of the system and have explored several methods for breaching the security of the databases of corporate organizations [7].

The issue is that security models used in databases in public organizations are vulnerable to cyber-attacks because of flaws in their security management systems [8]. Breaches are unavoidable, threats have become more complex, and database security has become more difficult. Furthermore, many threats are undetectable by traditional policy-based or rule-based security systems [9]. Firewalls, access control levels, and rule-based management are useless in circumstances of stolen privileged user accounts or internal attacks. As a result, there is a pressing need for a new technique that can detect harmful activity beyond the capabilities of rule-based systems. Any good security solution needs an intrusion detection system (IDS) to detect anomalous access. The software monitors network data and operating system operations for malicious activity or policy violations and generates reports [10].

1.1. Background Study. Anomaly detection is a technology that generates hints of possibly incorrect data and potentially dangerous processes. In the first stage, an anomaly detector analyses a system's usual state and behaviour and generates a set of reference for healthcare data that represents its unique qualities [11]. The same computations are then performed on the operational system, and the current set is compared to the reference set. The anomaly detector indicates an anomaly, i.e., an uncommon deviation, whenever the difference exceeds a certain threshold [12]. On systems with unambiguous patterns of regularity, anomaly detection works best, i.e., creates the fewest false hints and alerts. The most challenging aspect in designing an anomaly detection system (ADS) for networks and operating systems is identifying or extracting these patterns with well-designed relational databases [13]. Many of them are available for free to identify the anomaly. Anomalies are distinct from the rest of the data in the data set by their very nature. They can be separated from other data points in multidimensional Cartesian space. Anomalies will have a greater value than typical data points if the measurement of the average distance of the nearest N neighbors is obtained [14]. This attribute is used by distance-based algorithms to find anomalies in data.

The density of a neighbourhood data point is inversely proportional to its distance from its neighbours. Anomalies are found in low-density areas, while standard data points are found in high-density areas. The reason is that the relative frequency of an external user is small compared with the regular data point's frequency [15]. Data points with a low probability of occurrence are anomalies. Consequently, it is easier to discover the anomalous data points if the sample is fitted into a statistical distribution. For modeling the data set, it can be used to calculate the mean and standard deviation of a basic normal distribution. Anomalies in a data set differ by definition from the remainder of the data [16]. They are unusual data points separated from typical data points and usually do not form a close cluster. They still have a large distance from other clusters even when they join a group. Almost all classification techniques may be utilized to discover anomalies when previously categorized data are available [17]. When using the classification model, the availability of the previously marked healthcare data is an impediment. Since outlier data are unusual, it can be hard to find the anamoly [18]. Using oversampling the outer data with the remaining data, this problem can be partially overcome by stratified samples [19].

1.2. Related Work. A lot of anomaly detection technology was concentrated in operating systems and networks. In recent years, many techniques have been established due to the importance of privacy and security of personal information in database systems.

In [5], the authors have introduced a database security anomaly detection method. This means that the user's access pattern is checked in the database log and anomalous access events are detected. They evaluated the model based on the analysis of the user's pattern, the analysis of the machine learning, and the control of the rules. Casas et al. [6] have described Big-DAMA, a big data analytics framework (BDAF) for NTMA applications. Big-DAMA is a versatile BDAF, which evaluates and saves enormous quantities, both in streaming and batch mode, of structured and unstructured heterogeneous data sources. They have used Big-DAMA to detect various forms of network assaults and anomalies, comparing numerous supervised machine learning models. The assessments are made using the WIDE backbone networks based on real network measurements, and assaults are labeled with a known MAWILab data set. The experimental analysis have been compared to a normal Apache Spark cluster, and Big-DAMA can speed up computations by a factor of ten. Michele et al. [7] have drawn attention to important emerging challenges in the computer system and network security, particularly the Internet. Li et al. [8] have proposed a kind that user security auditing solution is based on a one-class support vector machine (OCSVM). The detection rate of 3 kinds of anomalous behaviour is above 80%, which shows a higher detection accuracy, according to simulation trials.

Ranganathan et al. [9] have used the Diffie–Hellman key exchange technique and the advanced encryption standards (AES) technique to implement the concept of differential privacy, which are both quite powerful in terms of speed. The tests were conducted with Laplace and Gaussian methods, which are the techniques currently most commonly used. The methods have been examined in the context of a case in which an initial and end location had been determined, and these had been encrypted using the aforementioned techniques while maintaining anonymity. Thudumu et al. [10] have attempted to chronicle the current state of anomaly detection in high-dimensional big data by utilizing a triangle model of vertices to represent the distinct challenges: the problem (large dimensionality), techniques/algorithms (anomaly detection), and tools (big data applications/frameworks specially pertaining to healthcare data). Furthermore, the limits of old methodologies and contemporary high-dimensional data strategies are explored, as well as recent techniques and applications on big data that are necessary for anomaly detection improvement. In [12], authors have introduced an anomaly detection method based on user behaviour into the internal attack detection in the database system to address the problem of internal attack in the database system. The anomaly detection of a database system was done using the discrete-time Markov chain (DTMC). The results suggest that the proposed approach can more accurately describe user activity and detect anomalies.

Even though databases include access control methods, these alone are insufficient to ensure data security. They must be supplemented by appropriate identification measures; the deployment of such techniques is critical for preventing impersonation attacks and dangerous code placed in applications. Additionally, anomaly detection procedures may aid in the prevention of insider threats, a growing problem in today's enterprises for which few solutions have been discovered. Although developing anomaly detection systems for networks and operating systems has been a hot topic of research, there are few anomaly detection systems specially designed for databases.

1.3. Need for the Research. The purpose of the research work presented in this paper is to investigate the construction of a database anomaly identification system to meet the need of the hour. There are two basic requirements for designing and developing such identification systems. One is the database application should not act as destructive element for the network and operating system used by an organization. The second and most crucial reason behind it is that the network and operating system capabilities cannot protect databases against the threats within the organization but can protect from threats from outside world. These threats are harder to detect, and it is difficult to protect the database since these threats are raised by the system administrators or users who have direct access to information and data.

1.4. Contribution of the Research. The contribution of this work is to design a cumulative anomaly detection system using discrete Markov Chain customized mechanism for database systems:

The discrete-time Markov chain (DTMC) has been used to detect anomalies in a database system

The sniffer is used to read user access behaviour data, which is then processed in a standardized manner.

The user behaviour features are extracted using the segmentation approach, and normal and aberrant feature data are acquired.

The state probability is used to evaluate anomalous process behaviour created by the discrete Markov chain

1.5. Organization of the Paper. This research is designed as follows: background, literature, as well as the study's goal and scope are provided in Section 1. The data and its representation are then defined in the subsequent part, followed by a description of the suggested anomaly detection approach in Section 2. The experiment design is provided in Section 3, and the findings are presented in Section 4. Finally, in Section 5, the findings are examined and conclusions are drawn, as well as future research directions.

2. Basic Definitions

The main process of cumulative anomaly recognition comprises of two main processes, viz training process and detection process [19], as shown in Figure 1. It is clear from Figure 1 that the training process has four main steps in the sequence data reading, processing, sequencing, and feature extraction. The detection process mainly consists of the user data gathering, characteristic extraction, comparison of characteristics with normal, and to detect the abnormality. Both the processes are discussed in detail as follows:

(1) In the process of training, make the database system run for a period of time under normal conditions, collect data during normal operation, extract user behaviour characteristics, and establish normal user behaviour mode (the established behaviour characteristics mode should include normal system behaviours).

(2) In the detection process, make the database system run in the real environment, gathers the behaviour data of the existing user and extract the behaviour characteristics, compare the behaviour characteristics of the detected user with the normal behaviour characteristics, and judge whether there is any abnormality by comparing the deviation degree between the normal and the current behaviour characteristics.

Figure 1 shows a framework of cumulative anomaly identification method for security database based on discrete Markov chain, which is divided into four parts:

(1) Data reading part: the reading object is the behaviour data generated when the user accesses the database.

(2) Data processing part: it is to process user access behaviour data.

(3) Feature extraction part: it is used to extract the feature information from the packets with known attack types and store it in the database.

(4) Feature comparison part: it is used to compare or match the captured package information and feature information in the feature database. If the match is successful, it is an anomaly and the response module is called for processing; if the match fails, it is normal.

2.1. Reading User Access Behaviour Data. When the user accesses the database, once the access behaviour occurs, the system immediately records a record in the cache, including digital ID/user account, access time, source IP, access page, source page, dwell time, and whether to leave or not. Figure 2 displays the construction process of the access record. It is clear from Figure 2 that data from "*n*" users is collected in data base, then the cookies account is generated and account, source IP, visit time, upper page, and stay time are recorded.

The access record records all the user's behaviours, so as long as these behaviour data are read, we can analyse whether the user has abnormal behaviours according to these data. At present, the main method of reading user access behaviour data is sniffer.

Sniffer is used as a software equipment that monitors the network data and mainly focuses on the legal management

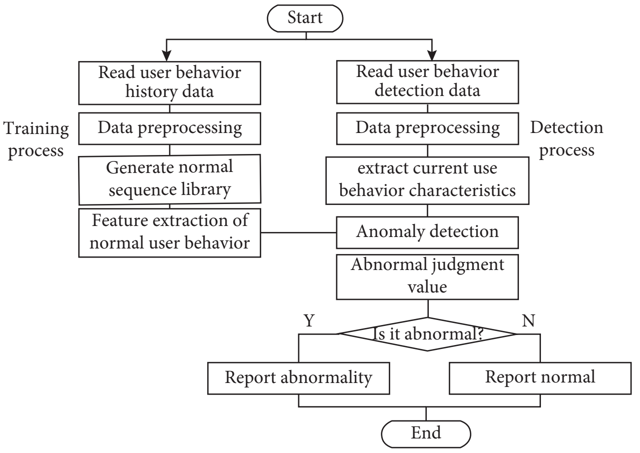

Start
Read user behavior history data
Read user behavior detection data
Training process
Data preprocessing
Data preprocessing
Detection process
Generate normal sequence library
extract current use behavior characteristics
Feature extraction of normal user behavior
Anomaly detection
Abnormal judgment value
Y
N
Is it abnormal?
Report abnormality
Report normal
End

Figure 1: Process of cumulative anomaly identification of security database.

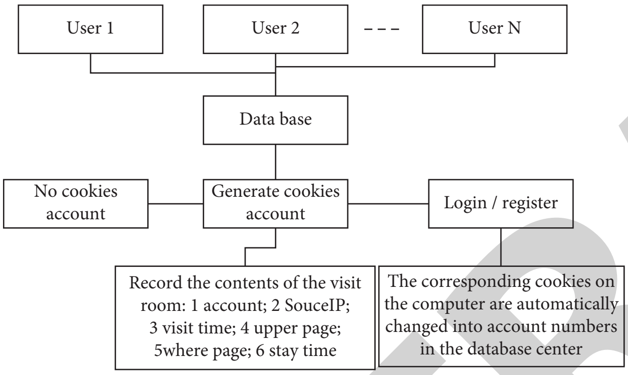

User 1
User 2
User N
Data base
No cookies account
Generate cookies account
Login / register
Record the contents of the visit room: 1 account; 2 SouceIP; 3 visit time; 4 upper page; 5where page; 6 stay time
The corresponding cookies on the computer are automatically changed into account numbers in the database center

Figure 2: Formation process of access record.

of the networks. The management of Sniffer includes monitoring of the network traffic, analysis of the data packets, monitoring of the utilization of the network resources, implementation of security rules, diagnosing of network problems, and identification and analyses of network data. Sniffer is usually composed of four parts (Figure 3): (1) network hardware equipment; (2) monitor driver: to intercept data flow, filter and store data in buffer; (3) real-time analysis program: to analyse data contained in data frame in real time to find network performance problems and faults; it is different from intrusion detection system in that it focuses on network performance and fault, rather than on discovering hacker behaviour; (4) decoding program: to decrypt the received encrypted data, construct its own encrypted data package and send it to the network.

The sniffer used in this paper is a kind of sniffer designed with Win Pcap technology. WinPcap is derived from Berkeley's group capture library. It is mainly used in 32 bit windows operation platform. WinPcap is mainly used for packet truncation and filtering the captured packets.

WinPcap technology enables the user-level data package to operate under the common windows platform. WinPcap is a kind of architecture, which uses BPF model and Libpcap function library. WinPcap mainly consists of the following parts (Figure 4):

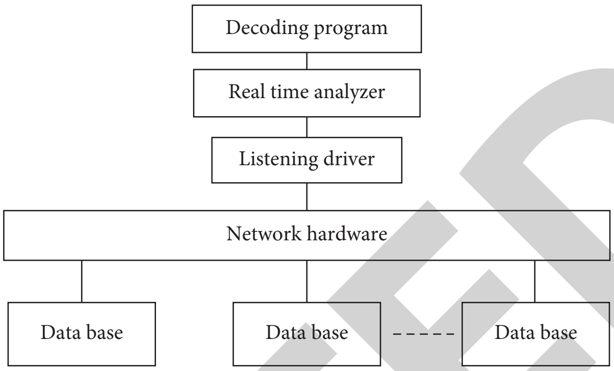


Figure 3: Structure of sniffer.

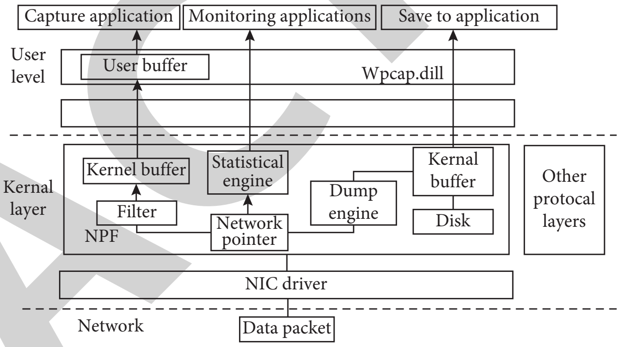


Figure 4: Structure of WinPcap.

NPF (Core Part). Net group Packet Filter, which is the network driver of the protocol, provides the function of intercepting and sending original packets for each operating system by calling NDIS. It is a virtual device driver file that filters packets and passes the original packets to the user.

Libpcap (Function Library). It is an upper level function library independent of the system and is more abstract.

Packet.dll (the Underlying Dynamic Link Section). It includes an application interface to access BPF and a function library conforming to the interface of high-level function library. Different operating systems have different kernels and user modules. This part provides a general interface for the platform in view of this phenomenon, thus saving the time of recompilation [16].

Among them, the underlying dynamic link part directly maps the kernel calls. In the dynamic link part, Wpcap.dll provides a more comprehensive and friendly function call. WinPcap's trump card lies in its standard interface for capturing packets. Moreover, WinPcap and Libpcap are compatible with each other. Therefore, for the network analysis tools supported by the original UNIX, it can be very compatible, which is very beneficial for development. At the same time, it also makes overall improvement in all aspects, making the operation more efficient. For example, it supports kernel level network packet filter and kernel state statistics mode.

WinPcap provides access to the bottom layer of the network on the application program of 32-bit operating system. It mainly includes the following aspects:

Interception function: it is used to effectively intercept the original datagram, mainly for all kinds of datagrams exchanged, sent and received by each host on the shared network

Filter function: it is used to provide user-defined rules, filter out the parts that meet the rules before sending the datagram according to the defined rules

Function of sending datagram: to support sending original datagram on shared network

Summary statistics function: in the process of active network communication, the collected information is summarized and counted

Figure 5 shows the flow chart of WinPcap sniffer reading user access behaviour data.

2.2. Data Processing. In order to make them comparable, it needs to use standardized methods to eliminate the deviation:

(1) Max-Min standardization/dispersion standardization:

Max-Min standardization, is also known as discrete standardization, is a linear transformation of the data and normalizing the values to [0,1]. The formula is shown in

$$x' = \frac{(x - \min)}{\max}, \tag{1}$$

where max represents the highest value of the sample and min represents the lowest value of the sample.

Deviation standardization keeps the relationship of the novel data and the normalized data. It is the method to eradicate the influence of dimension on the data range. The problem with this method is addition of new data that may cause changes of highest and lowest values in the sample and then the conversion function requires to be redefined [20].

(2) Z-score standardization/standard deviation standardization/zero mean standardization- Z-score is also a standard deviation standardization. The mean value is given by 0 of the processed data and the standard deviation value is 1. The formula is shown in

$$x' = \frac{(x - \mu)}{\sigma}, \tag{2}$$

where μ is the mean and σ is the standard deviation.

This method is not sensitive to outliers. It is very useful when the maximum and minimum values of the original data are unknown or the outliers control the Max-Min standardization. Z-score standardization is currently the most widely used standardization method [21].

(3) Log function conversion

By using the log function conversion, the scaling of data is also performed. The formula is shown in

$$x' = \frac{\log_{10}(x)}{\max}, \tag{3}$$

where max is the highest value of the sampling data.

2.3. Sequence Feature Extraction of User Behaviour. Feature extraction refers to the extraction of feature information from the data of known attack types and the behaviour data of current users. At present, there are multiple linear regression analysis algorithm and independent component analysis algorithm for user behaviour feature extraction. Among them, the former has a good filtering effect, but for large-scale information, the calculation process is more tedious, while the latter is within the error tolerance range, but takes a long time [22]. In view of the above situation, a user behaviour feature extraction based on time series is proposed in this section.

User access behaviour is a long series of sequential data in time sequence, so there must be some regularity, so as long as we grasp this regularity, that is to extract the sequence characteristics of user behaviour, we can achieve anomaly detection under the guidance of subsequent matching. At present, the method of feature extraction is mainly based on transform. Its principle is to transform the time series into the feature space and then use its feature mode to represent the time series. Its typical representatives are Fourier transform and discrete wavelet transform. However, this method can only be implemented on the premise of the same distribution of data groups. Once the data in the data flow are distributed differently, this method will lose its effectiveness [23]. In view of this situation, this section uses the segmentation method to extract the features of user access behaviour data. Compared with the traditional extraction method, the biggest feature of segmentation method is that it is faster and more accurate. The basic idea is the user behaviour sequence is separated into several segments and then the average value of each segment is determined. Finally, according to these average values, a vector is formed, that is, the feature representation after data dimensionality reduction, which is expressed as follows by mathematical formula:

Supposing that a time series is $G = \{g_1, g_2, \ldots, g_n\}$, where $g_{.}$ represents each data in the series and n is the number of data in the series, that is, the length of the series.

Let N represent the dimension of the feature space and $1 \leq N \leq n$, the time series with length is represented by the feature vector of N-dimension feature space as shown in

$$\overline{H} = \{h_1, h_2, \ldots, h_N\}, \tag{4}$$

where the ith element of $\overline{H}$ can be found out by

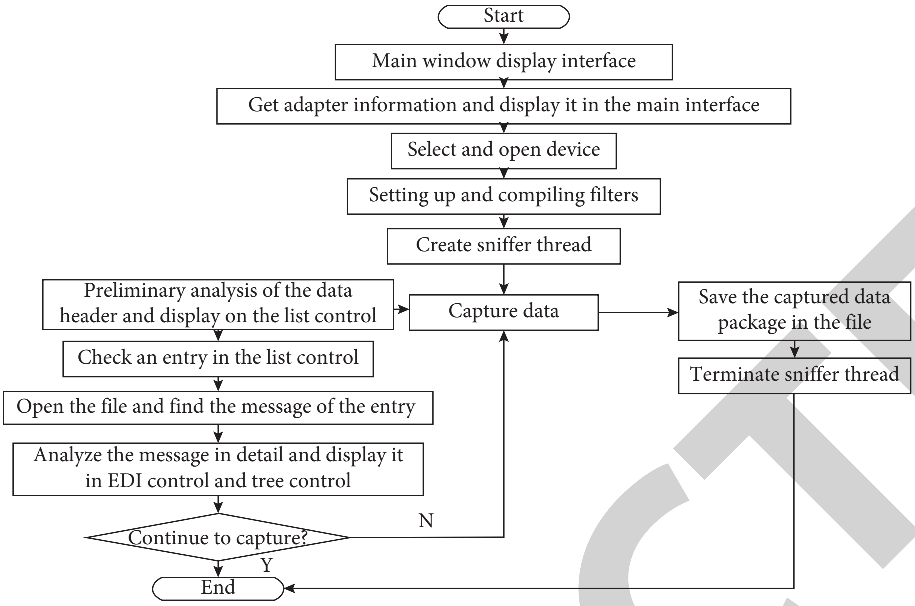

Start
Main window display interface
Get adapter information and display it in the main interface
Select and open device
Setting up and compiling filters
Create sniffer thread
Preliminary analysis of the data header and display on the list control
Capture data
Save the captured data package in the file
Check an entry in the list control
Terminate sniffer thread
Open the file and find the message of the entry
Analyze the message in detail and display it in EDI control and tree control
Continue to capture?
N
Y
End

Figure 5: Flow chart of WinPcap sniffer reading user access behaviour data.

$$h_i = \frac{N}{n} \cdot \sum_{j=(N/n)(i-1)+1}^{(N/n)\cdot i} g_i. \tag{5}$$

Here, when $n = N$, the features of time series before transformation are the same as those after transformation; when $n = 1$, the features of time series after transformation are the same as the arithmetic mean of time series before transformation.

The above is the principle basis of segmented method for feature extraction. The following describes the specific process:

Step 1: set the input parameters, that is, determine the time series set and time series, the number of sequence segments k, and define the threshold value of local change mode.

Time series set:

$$G' = [G1, G2, \ldots, G_S]. \tag{6}$$

Time series:

$$G = \{g_1, g_2, \ldots, g_n\}. \tag{7}$$

Step 2: according to the frequent sequence of big data flow, the initial characteristic matrix is constructed as follows as shown in

$$F = \begin{bmatrix} (f_{11}, Q_1), & (f_{12}, Q_2), \ldots, (f_{1n}, Q_n) \\ (f_{12}, Q_1), & (f_{22}, Q_2), \ldots, (f_{2n}, Q_n) \\ & \cdots \\ (f_{m1}, Q_1), & (f_{m2}, Q_2), \ldots, (f_{mn}, Q_n) \end{bmatrix}. \tag{8}$$

Here, $f_i\,(i = 1, 2, \ldots, n)$ is the column vector of the characteristic matrix; $Q_i\,(i = 1, 2, \ldots, n)$ is the distance. The formula represents the local features of each variable dimension in each segment of the feature sequence F of user access behaviour data.

Step 3: divide each time series in the feature series of user access behaviour data into k subseries, as shown in

$$g_i = \{z_1, z_2, \ldots, z_i, \ldots, z_k\},\ i = 1, 2, \ldots, k, \tag{9}$$

where $z_1 = \{z_{b_1}, z_{b_2}, \ldots, z_{b_i}, \ldots, z_{b_{k+1}}\}$; $z_{b_i}, i = 1, 2, \ldots, k$ is the segmentation point.

Step 4: calculate the maximum value, minimum value, slope, and slope standardization value of the kth time series in the feature series of user access behaviour data. The formula is as follows:

Maximum value:

$$\max z_i = \max\{z_{b_1}, z_{b_2}, \ldots, z_{b_{i+1}}\}. \tag{10}$$

Minimum value:

$$\min z_i = \min\{z_{b_1}, z_{b_2}, \ldots, z_{b_{i+1}}\}. \tag{11}$$

Slope p_i:

$$p_i = \frac{z_{b_{i+1}} - z_{b_i}}{b_{i+1} - b_i}, \tag{12}$$

where $1 = b_1, b_2, \ldots, b_i, \ldots, b_{k+1} = n$

Slope standardization value a:

$$a = \frac{d - \overline{d}}{v_d}. \tag{13}$$

Here, d is the sequence feature; $\overline{d}$ is the average value of the sequence feature d; and v_d is the standard deviation of the sequence feature d.

Step 5: save the results from Step 4 above to the initial matrix.

Step 6: calculate the jump value of each subsequence after the k th time series is segmented.

Step 7: judge whether the jump value u between two adjacent subsequences is greater than the threshold e. If it is greater than e, continue to the next step, otherwise terminate.

Step 8: Add the subsequence larger than d into the initial feature matrix, and stsndardize it.

Step 9: Repeat the above steps, extract the mean value, variance and slope of each time in the frequent sequence set of big data stream, and then standardize them, and list them in the feature matrix to achieva sequence feature extraction.

2.4. Cumulative Anomaly Detection Based on DTMC. Markov process is a random process with no after effect. The so-called no after effect refers to that when the state of a random process at time t_0 is known, the state of a random process at time $t\,(t > t_0)$ is only related to the state of time t_0, but not to the state of a process before time t_0 [20, 21]. Those Markov processes with discrete time and state are called Markov chains, as shown in Figure 6.

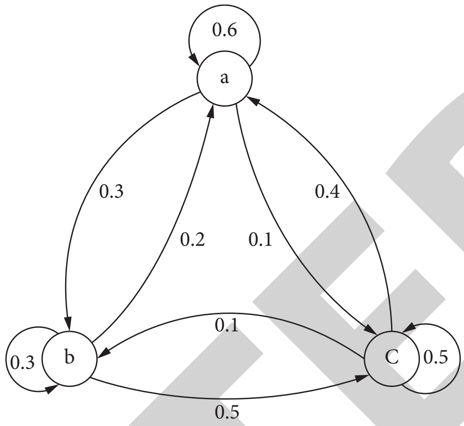


Figure 6: Markov chain.

Markov chain is a sequence of random variables with Markov property. If there is a random process $\{Y(t), t \in T\}$, the state of t at the time is Y_t, and the state of Y_{t+1} at $t+1$ is only related to the state of Y_t at t, but not to the state of $Y_{t-1}, Y_{t-2}, \ldots, Y_0$ at any time in the past, then $\{Y(t), t \in T\}$ is called Markov process. The state of Markov process is countable, as shown in

$$Z(Y_{t+1} = V_{t+1} | Y_t = V_t, Y_{t-1} = V_{t-1}, Y_1 = V_1) = Z(Y_{t+1} = V_{t+1} | Y_t = V_t), \tag{14}$$

where $V_1, V_2, \ldots, V_T \in (S_1, S_2, \ldots, S_N)$ is the value of the state and is called, as shown in

$$Y_{i,j}(t, t+1) = Y\left(Y_{t+1} = S_j | Y_t = S_i\right), 1 \le i, \quad j \le N. \tag{15}$$

$Y_{i,j}(t, t+1)$ is the probability of transition from state i to state j. i, j has N states, respectively. When $Y_{i,j}(t, t+1)$ has nothing to do with t, then Markov chain is called homogeneous Markov chain.

When the Markov chain is homogeneous and $Y_{i,j}(t, t+1)$ is recorded as b_{ij}, the state transition probability matrix is as follows, as shown in

$$B = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1N} \\ b_{21} & b_{22} & \ldots & b_{2N} \\ & \ldots & & \\ b_{N1} & b_{N2} & \ldots & b_{NN} \end{bmatrix}, \tag{16}$$

where $1 \le i, j \le N$ and $1 \le b_{ij} \le N$, $\sum_{j=1}^{N} b_{ij} = 1$ and B is called the state transition matrix.

It can be seen that matrix B represents the probability of state from t to $t+1$, but the probability of initial state distribution cannot be obtained. Therefore, in addition to matrix B, the initial probability vector $\pi = \{\pi_i\}$ must be obtained to represent the complete Markov chain process.

$$\pi_i = Y\left(V_1 = C_i\right), 1 \le i \le N, \quad 0 \le \pi_i \le 1, k4 \sum_{i=1}^{N} \pi_i = 1. \tag{17}$$

In this case, (B, π) can represent a Markov chain.

On the basis of the above Markov chain principle, the cumulative anomaly recognition of security database is carried out, and the specific process is as follows:

Step: 1: execute a system call and add it to the end of the empty queue;

Step 2: match the system call sequence in the queue with the feature pattern in the feature library. If the sequence happens to be the feature pattern, go to Step 3; if the sequence matches a feature pattern, go to Step 1; if it cannot match, go to Step 4;

Step: 3: record the corresponding status number, add the status sequence, clear the queue, and go to Step 1;

Step: 4: add the status sequence corresponding to each system call in the queue, clear the queue, and go to Step 1.

The above steps are repeated until the end of the process. The system call sequence is transformed into a state sequence, the detection is based on the probability $p\ (L)$ of L consecutive states, and the method of local frame counting is used. The frame is a window with fixed length k [24, 25]. In the detection process, the frame window will slide forward with the detection point, which is used to record the number of k state sequences with probability less than the threshold v. The number of records less than the threshold v in the frame is counted here. When the count value is greater than 2, an anomaly is considered and an alarm is given.

3. Results

In order to check the viability and the effectiveness of the proposed cumulative anomaly recognition method for security database based on discrete Markov chain, it is compared with three anomaly recognition methods in reference [3–5]. In this paper, the event log generated when the DARPA98 data set is replayed on the NT system is used as the experimental data for simulation experiment. The attack scenario of DARPA98 data set is shown in Figure 7.

The test data set of DARPA98 attack scenario comprises a series of attacks. The whole attack process is realized by DDoS attack. The invader first notices the active host through IP Sweep and then scans the port to find the host with sad-mind vulnerability. Then, the attacker attacks three hosts with this vulnerability: Pascal (172.16.112.50), Mill (172.16.112.20) and Locke (172.16.112.10) to make it a puppet machine. Then, the attacker installs the Trojan horse software to implement DDos attacks on the puppet machine and uses the controlled host to make DDos attack on the target.

3.1. Data Set. DARPA98 provided by Lincoln Laboratory of MIT is used as a data source. Because of the large amount of data, this experiment only selects part of the data for testing. In order to make the experiment comparable, five typical attacks are extracted as the experimental data of this model. Five attacks are Neptune (SYNFlooding), Satan, PortSweep, Buffe-overflow, and Guess-passwd. The attacks selected in this experiment include four categories of attacks, as shown in Table 1.

3.2. Development Environment. The development environment is java language platform (JDK1.6.2). It is an object-oriented programming language. This paper uses it as the development language mainly because it has the following characteristics:

(1) Java language is simple. Java discards redundant operations such as operator overloading, multiple-inheritance, and automatic cast. It does not make use of pointers.

(2) Java language is distributed. It supports the development of Internet applications by using network application programming.

(3) Java is portable language. In addition to it, Java strictly defines the length of each basic data type.

(4) Java language is multithreaded and provides the synchronization mechanism between multithreads (the keyword is synchronized).

3.3. Experiment Process. First, 60% of all data are used for training, including intrusion data and normal data; second, after the training, another 40% data are used to test the model; third, output results are generated.

3.4. Evaluation Index. The data in this paper can be divided into two categories after model detection, that is, positive data and negative data. Whether the payload data can be classified correctly is identified by true or false. The correct classification is true, and the error classification is false. Each model may produce four results for sample detection, which are, respectively, represented by TP, FP, TN, and FN, as shown in Table 2:

(i) TP indicates that the real category of data samples is positive, and the predicted outcome is also positive.

(ii) FP indicates that the real class of data samples is negative, but the predicted outcome is positive.

(iii) FN indicates that the real category of data samples is positive, but the final predicted outcome is negative.

(iv) TN indicates that the real category of data samples is negative, and the predicted outcome is also negative. According to the above indexes, precision and recall can be calculated, respectively.

Precision, the accuracy rate, indicates the probability of correct prediction of positive class in the prediction results and in the data samples of positive class, shown in

$$precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{18}$$

TPR, also known as recall, indicates the probability of being correctly predicted as a positive class in the positive class of the original data sample, as shown in

$$TPR = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{19}$$

In the experiment, we hope to get high precision and recall, but the precision and recall are mutually exclusive, so we need a compromise way Fl-score to express the effect of the experiment. Fl-score represents the harmonic average evaluation index of precision rate and recall rate, as shown in

$$Fl - score = \frac{2 \times \mathrm{TPR} \times \mathrm{precision}}{\mathrm{TPR} + \mathrm{precision}}. \tag{20}$$

3.5. Result Analysis. From Table 3, it can be observed that the proposed work in this paper is better than other methods in reference [3–5] in terms of cumulative anomaly recognition of security database, and the F1-score obtained is higher than the three anomaly recognition methods in other methods in reference [3–5], which shows that the recognition performance of the method in this paper is better. Table 4 shows precision achieved by all the methods, and the precision achieved by the method proposed in this article is the highest.

4. Discussion

In this paper, user behaviour anomaly recognition is establishing a normal behaviour mode of a legal user. By comparing the current behaviour and normal behaviour characteristics of the legal user, we can identify the abnormal behaviour. That is, if the present behaviour of the legal user deviates greatly from the normal behaviour characteristics in its history, it is considered that an anomaly has occurred. This anomaly may be caused by the unauthorized operation of the legal user itself, or by the illegal operation of other legal users or external intruders in the system. In the

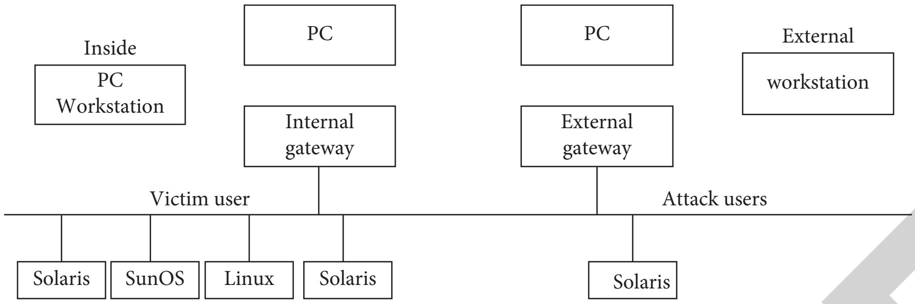


Figure 7: Attack scenario of DARPA98 data set.

Table 1: Attack data.

Attack categories	Attack selected in this experiment	Benign sample	Malicious samples
Dos	Neptune	31366	24225
Probing	PortsWeep	74463	62551
	Statan	35440	36641
U2L	Buffer-overflow	74450	82565
R2L	Guess-passwd	47123	36550

Table 2: Data classification.

Classification	Actual normal data	Actual malicious data
Forecast normal data	TP	FP
Predict malicious data	FN	TN

Table 3: F1-score.

Evaluation parameter	Article method	Reference [3] method	Reference [4] method	Reference [5] method
F1-score	0.8586	0.7233	0.8236	0.7562

Table 4: Precision obtained by various methods.

Evaluation parameter	Article method	Reference [3] method	Reference [4] method	Reference [5] method
Precision	0.92	0.75	0.85	0.78

database system, users mainly interact with the database management system through the access request to complete information query, modification, deletion, and other operations. Therefore, by analysing the execution sequence of the access request, we can more comprehensively explore the behaviour characteristics of users.

In order to improve the poor performance of traditional methods, this paper proposes a new method based on discrete Markov chain, which is proved to be more effective than traditional methods. The proposed method in this paper is based on the Markov chain and can be used for better anomaly recognition. The results are obtained in terms of sensitivity score, precision score, and F1-score. The results are also compared with the results obtained by using some of the state-of-the-art traditional techniques. The comparison clearly indicated that the proposed method is more effective as compared to the tradition methods. The proposed method has the highest F1-score of 0.8586 and then the traditional methods that have F1-score of 0.7233, 0.8236, and 0.7562 for methods 1, 2, and 3, respectively. The precision obtained by our method is 0.92, which is the highest among the comparative methods.

5. Conclusions

In this paper, a novel anamoly detection method based on discrete Markov chain is proposed to identify the cumulative anomaly in security database. This method not only considers the probability relationship between system calls but also considers the semantic relationship of system calls, that is, the short sequence of repeated system calls. After testing, the F1-score of the proposed method is higher than that of traditional methods, which proves the validity and feasibility of the method and achieves the purpose of research. This research provides a novel approach based on discrete Markov chain, which has been shown to be more successful than traditional methods in order to enhance the poor performance of traditional methods. This article's proposed method is based on the Markov chain and can be utilized to improve anomaly detection. The sensitivity score, precision

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9801983, 1 page
https://doi.org/10.1155/2023/9801983

Retraction

Retracted: Enhanced Lorenz-Chaotic Encryption Method for Partial Medical Image Encryption and Data Hiding in Big Data Healthcare

Security and Communication Networks

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] P. Rashmi, M. C. Supriya, and Q. Hua, "Enhanced Lorenz-Chaotic Encryption Method for Partial Medical Image Encryption and Data Hiding in Big Data Healthcare," *Security and Communication Networks*, vol. 2022, Article ID 9363377, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9363377, 9 pages
https://doi.org/10.1155/2022/9363377

Research Article

Enhanced Lorenz-Chaotic Encryption Method for Partial Medical Image Encryption and Data Hiding in Big Data Healthcare

P. Rashmi,[1] **M. C. Supriya**,[2] **and Qiaozhi Hua**[3]

[1]*Research Scholar Sri Siddhartha Academy of Higher Education, Tumakuru, India*
[2]*Professor Dept of ISE SSIT Tumakuru, Sri Siddhartha Academy of Higher Education, Tumakuru, India*
[3]*Computer School, Hubei University of Arts and Science, Xiangyang 441000, China*

Correspondence should be addressed to Qiaozhi Hua; 11722@hbuas.edu.cn

Received 23 August 2021; Revised 28 September 2021; Accepted 11 January 2022; Published 27 February 2022

Academic Editor: Chinmay Chakraborty

Image encryption is highly required in the big data healthcare cloud to improve the security of the medical image for remote access. Data hiding method is the process of storing the medical information of the patient in the medical image in the hidden format. Many existing data hiding methods are based on wavelet and chaotic map due to its effectiveness. Wavelet based methods have limitations of lack of phase information, poor directionality, and shift sensitivity. Chaotic map is applied to improve the security of the medical image and chaotic map has the limitation of low sensitive to control parameters and initial conditions. In this research, the Improved Chaos Encryption (ICE) is applied to improve the security based on randomness. The average energy is calculated in the images and compared with adaptive threshold to segment the Lorenz 96 model applied in the chaos encryption algorithm to improve the model security. Lorenz 96 increased the randomness of the chaos encryption method due to its high sensitivity. Medial images were used to test the performance of the ICE in the image encryption and image hiding. The proposed ICE model evaluated the quality of the recovered and decrypted image in the various embedding rate. The result shows that the proposed ICE model has the PSNR value of 104.7 dB compared to the LSB-ROI method which has 97.61 dB PSNR.

1. Introduction

In medical images stored in the cloud, the data hiding method is used to store the medical images of patient for remote access. Healthcare generates a bigger amount of data related to the patient for diagnosis purposes. The encoding method is used in the data hiding approach to store the medical record in the medical images. Decoding is performed in the client side to extract the hidden data and also find digital content that has been attacked. Two types of methods are present in data hiding, namely, steganography and watermarking [1–11]. In reversible data hiding (RDH), the original cover or the region of interest (ROI) is restored losslessly and the remaining part of the images is restored in lossy manner. The RDH plays an important role in medical image processing and in other applications such as big data healthcare, image transcoding, and multimedia archive management [12]. Image encryption method is applied to secure the image and the images can be restored based on authentication to protect the data from users or attackers. Image encryption is required in the medical images to store in cloud and to protect the privacy of the patients [13]. Electronic Patient Record (EPR) consists of patient ID's information, diagnostics reports, and vital signs. The EPR is hidden in the medical images to store in cloud for authenticated information [14]. Most of the RDH method aims to enhance the embedding rate and quality of the images for encryption.

The patient's sensitive information stored in the cloud has the chance to be leaked to hackers with malicious intent. In medical information systems, the protection of patient's privacy and medical images plays an important role and attained more attention [15–26]. Medical images transmission or storage in cloud is easily accessible by the hacker that creates privacy issue. Data hiding method prevents information from unauthorized access and protects the privacy of data. Medical images such as computed tomography (CT), ultrasound, X-ray, and magnetic resonance imaging (MRI) created by imaging devices are used as the cover image [27–32]. The chaotic method is used in the image encryption to improve the security based on the features of initial values of nonlinearity, pseudorandomness, and sensitivity. The chaotic encryption method is widely followed in the encryption process due to its features [33, 34]. The discrete chaotic maps are sensitive to control parameters and initial conditions that increase randomness being deterministic, unpredictable, and easily reproducible [35]. In this research, the ICE method is proposed to increase the privacy of the data in the cloud. Medical images were used to test the proposed model's efficiency in various embedding rate.

(1) This research aims to measure average energy in the images and, compared with adaptive threshold, to segment the image into ROI, RONI, and border area. The ROI is encrypted and decrypted in lossless manner and RONI is retrieved in lossy manner.

(2) The objective of this research is to apply Lorenz 96 model in chaos encryption method to increase the sensitivity to initial value and resilience against the various kinds of attacks. The random value is set in the chaotic sequence generator to enhance the ability in attack resistance.

(3) The PSNR value is measured for various embedding rate in the medical images to evaluate the quality of retrieved images. The proposed Lorenz-chaotic encryption method is compared with state-of-the-art method to evaluate the efficiency.

(4) The proposed ICE model has achieved the PSNR value of 104.7 dB and existing LSB-ROI method has 97.61 dB PSNR. The proposed method has achieved higher performance due to its sensitivity to initial condition.

The review of recent researches in the image encryption and data hiding methods is given in Section 2, the proposed ICE method, image embedding, and recovery were given in Section 3, results of the proposed ICE method are given in Section 4, and conclusion is given in Section 5.

2. Related Works

Medical information storing in cloud for remote access requires effective encryption method to protect the privacy of the medical data. Data hiding or embedding techniques were applied to store the sensitive information in the image. ROI-based methods show the considerable performance in encryption and preserve the quality of the image. Chaotic based encryption method increases the efficiency of the encryption.

Parah et al. [36] applied the Intermediate Significant Bit Substitution (ISBS) to embed the checksum data, watermark data, and EPR to eliminate commonly used Least Significant Bit (LSB) replacement/removal attacks. Chen and Chi [37] applied Block Truncation Coding (BTC) method for data hiding and compression method to reduce the image size and improve the security. The block classification method is applied to classify the blocks into three types such as smooth blocks, and two complex blocks. Applied separate method for three types of block for data hiding and encryption. Loan et al. [38] applied hybrid edge detection method and Pixel Repetition Method (PRM) for the data hiding in medical image. The PRM is applied to increase the small size image and hybrid edge detection method is applied to preserve the edge information. Geetha and Geetha [39] applied the Rhombus Mean Interpolation method to predict the interpolated points for data hiding method. Checksum is applied in the nonoverlapping method to embed for content authentication and tamper detection. The modified LSB based methods such as ISBS [36], BTC [37], PRM [38], and Rhombus Mean Interpolation [39] have limitations of high loss of information due to rearranging of image.

Yang et al. [40] applied adaptive threshold method to automatically segment ROI and RONI in the medical images. To enhance the ROI, the grayscale is stretched and data is embedded in the stretched histogram peak bins. Gao et al. [41] applied reversible data hiding (RDH) and contrast enhanced algorithm to improve the image quality and embedding capacity. The developed algorithm separates ROI and NROI in the medical images and stretches the ROI's gray level histogram. Balasamy and Suganyadevi [42] applied fuzzy based ROI selection method and wavelet transformation method to embed the encrypted watermark in medical images. Fuzzification method is applied to measure the critical points based on the final and central intensity with selected ROI. Wu et al. [43] applied Otsu's method for image segmentation to enhance the ROI in the medical images before contrast enhancing. The GrabCut interactive algorithm is used to accurately segment the ROI in medical images. The analysis of the proposed GrabCut method in medical images shows that the proposed method is effective in encryption and data hiding. Zhou et al. [44] proposed game theory with hidden ROI position and optimized ROI parameters for lossless medical image encryption. The Quantum Cell Neural Network (QCNN) hyperchaotic model generates random sequence to diffuse and scramble ROI. The result shows that the game theory provides optimal balance between the encryption speed and security performance. Priya and Santhi [45] proposed nonembedding image encryption to conceal the presence of the watermark in the medical image. The biometric authentication is applied to decrypt the information in the medical image. The developed method prevents the intentional and unintentional attacks. Ding et al. [46] applied Deep Learning based Encryption and Decryption Network (DLEDNet) method for the encryption and decryption in the medical images. Main learning network is applied in Cycle-GAN to transfer the medical images. The adaptive threshold methods [40, 41] have higher performance in the segmentation and has lower embedding rate in data hiding. The fuzzy [42] method is supervised method and manual features are required for segmentation in data hiding. Otsu's method [43] changes the modality of histogram due to global thresholding and degrades the images. The QCNN [44] and DLEDNet [46] model have limitation of overfitting and authentication method [45], which has lower sensitivity for data hiding.

Yin and Li [47] proposed modified quantum chaos system and Particle Swarm Optimization (PSO) with genetic simulated annealing method for the medical image encryption. The modified chaos system method is applied for key stream and genetic algorithm is applied for the selection and cross operation to process the plaintext images. Simulated annealing method is applied to scramble the image to generate the optimal sequence. The PSO method is applied to process the simulation annealing. Liu et al. [48] partition the ROI and NROI in the medical image using ROI-based reversible data hiding method. An encryption key is applied to encrypt ROI and NROI in the medical images. The LSB of the EPR and encrypted ROI is concatenate in the data hider. The LSB substitution method is applied to embed the concatenate data in the medical image. Zhang et al. [49] proposed hyperchaotic system for the encryption of medical images and embedded the patient private information in ROI based on reversible data hiding method. The developed method has low distortion and high embedding capacity in the data hiding process and also improves the security of the encryption phase. Anand and Singh [50] applied Singular Value Decomposition (SVD)-Discrete Wavelet Transform (DWT) to embed the multiwatermarks in medical images. Hamming code is applied to decrease the channel noise distortion in the text watermark. The chaotic-LZW has the higher efficiency in data hiding and security in the medical image. Kumar et al. [51] applied chaotic map on the fractional Discrete Cosine Transform (FrDCT) on the medical images. The result shows that the proposed model has the higher efficiency in improving the security in the data hiding. Ravichandran et al. [52] provided the hybrid encryption method based on chaotic map and deoxyribonucleic acid to be adaptable for selective and full medical image encryption. The result shows that the hybrid model has the higher efficiency in the security improvement. The PSO [47] method has lower convergence in the parameter settings and encryption method [49] has lower embedding performance. The SVD-DWT [50] method has limitation of lack of phase information, poor directionality, and shift sensitivity. The chaotic map [51, 52] methods have limitations of low sensitivity in initial conditions and control parameters.

3. Method

Medical images required image hiding and image encryption to increase the privacy of the patient data. In this research, the ICE method is applied to increase the security of the medical encryption. Medical images were used to test the efficiency of the proposed method in the encryption. The Lorenz 96 model is applied in the chaos encryption to increase the privacy of the image. The overview of ICE model in medical image encryption is shown in Figure 1.

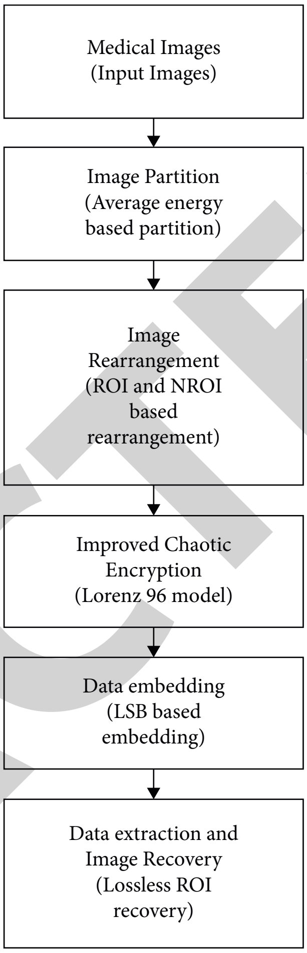

Medical Images (Input Images)
Image Partition (Average energy based partition)
Image Rearrangement (ROI and NROI based rearrangement)
Improved Chaotic Encryption (Lorenz 96 model)
Data embedding (LSB based embedding)
Data extraction and Image Recovery (Lossless ROI recovery)

FIGURE 1: The overview of the proposed ICE method.

4. Image Partition

The original input medical image I is divided into three parts such as border area, ROI, and region of noninterest (RONI). In most cases, the ROI is irregular shape in medical images. The image bottom line is the border area and ROI vertices are used to describe ROI, denoted by D_{roi} with length L_c.

The size of input medical image is $N_1 \times N_2$ and is first divided into block size of $n_1 \times n_2$. The amount of blocks is $(N_1 \times N_2)/(n_1 \times n_2)$. Every block average energy is measured using the following formula:

$$\text{average energy}\,(m,n) = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} I\,(i,j)^2}{n_1 \times n_2}, \qquad (1)$$

where Average energy (m,n) denotes the current block average energy, the image blocks position is denoted as (m,n), and the pixel value is denoted as $I\,(i,\ j)$. Every block average energy value with an adaptive threshold T determines the ROI. If average energy $(m,\ n) > T$, then it belongs to the ROI; otherwise, the blocks belong to RONI.

The ROI is carried out in lossless retrieval and RONI is carried out in lossy retrieval. The ROI is placed in front concatenated by border area and RONI. The rearrangement operation can improve the security.

4.1. Chaotic Encryption Algorithm. A chaotic system providing chaotic sequence is applied for the image encryption for medical images [53–55]. The encryption method robustness is important for the medical image encryption. Since chaotic encryption method is sensitive to initial value and resilience against various kinds of attacks, the chaotic encryption method is used for encryption.

The Chen system is applied to iterate out three chaotic sequences in the chaotic state that is applied in this encryption algorithm. The Chen system is defined in the following equation:

$$\begin{cases} x = 35(y - x), \\ y = -7x - xz + 28y, \\ z = xy - 3z. \end{cases} \tag{2}$$

The proposed encryption involves four steps; they are as follows.

Step 1: the image matrix is denoted as $P = [p]_{512\times512}$. Initial value is selected as $K' = [k_0', k_1', k_2', k_3']$, where $k_0' \leq 512 \times 512$; the scrambling algorithm iteration is set to 50; the chaotic Chen system initial value is denoted in triplet $[k_1', k_2', k_3']$, and it is randomly set to $[0.0663598, 0.45679, 0.9256]$. This random operation is applied to enhance the encryption algorithm ability to attack resistance.

Step 2: the Chen system and key group obtains the chaotic sequences of C_1, C_2, C_3 and three sequences consist of bits 1001 to $1000 + M \times N$ to chaos discard (secret image size is denoted as $M \times N$).

Step 3: the Helical scan sort matrix is denoted as A and the scrambling iteration is used to calculate A', as in the following equation:

$$A_i' \begin{cases} \text{rotation}(A, 90), & 0 < C_1(i) < 0.25, \\ \text{rotation}(A, 180), & 0.25 \leq C_1(i) < 0.5, \\ \text{rotation}(A, 270), & 0.5 \leq C_1(i) < 0.75, \\ \text{rotation}(A, 270), & 70.75 \leq C_1(i) < 1. \end{cases} \tag{3}$$

The simplest version of Lorenz 96 model based on periodic system of $K(k = 1, \ldots, K)$ is

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = -X_{k-1}\left(X_{k-2} - X_{k+1}\right) - X_k + F. \tag{4}$$

Advection term is first term on the right hand side, damping is represented as second term, and external forcing term is provided as F that is set as 10.

Two-level version of Lorenz 96 model is applied for parameter estimation and this will add another periodic variable Y. The X and Y are linked in coupling term that is last term in equation. Each X has J Y variables related with it.

X is resolved, slow variables, and Y is fast and unresolved variables. The task is to represent a parameterization on fast variables effect on X and replace last term in X equation for convenience.

Consider ignoring correlation in space and time, modelling B_k as a local function of X_k, as given in

$$-hc\overline{Y}_k \coloneqq B_k \approx (X_k). \tag{5}$$

A linear regression is applied for simplest parameterization, as given in equation

$$B_k = aX_k + b. \tag{6}$$

Advantages of Lorenz 96 model are discussed as below:

Multiscale: Lorenz 96 model is applied for parameterization in chaotic map that is divided into resolved and unresolved processes.

Encryption complexity: as Lorenz 96 model increases the randomness and sensitivity of chaotic map parameter, the complexity of the model to decrypt is high which improves the security.

The Lorenz 96 model has been applied in the chaos encryption method to improve the security of the model.

$$L = (x_{m+1} - x_{m-2})x_{m-1} - x_m + F, \tag{7}$$

where $m = 1, \ldots, M$, $x = x_{M-1}$, $x_0 = x_M$, and $x_{M+1} = x_1$. This model mimics an meteorological quality of unspecified scalar time evolution, x, and latitude circle of equidistant grid M.

Let $P = [p]_{512\times512} = [p_0]_{512\times512}$, and get the chaotic sequence C_2 of index S in order. Then, $P_i' = [p_i']_{512\times512}$ is determined, as in

$$p_i' = \left(A_i'(S(j))\right) = p_{i-1}(j), \quad j = 1, 2, \ldots, 512 \times 512, \tag{8}$$

where $i = 1, 2, \ldots, k_0'$, and the scrambling process is repeated $k_0^{'}$ times. Finally, $P' = P'k_0'$.

Step 4: $P' = [p']_{512\times512}$ and C_3 using the chaotic diffusion diffuse the matrix $P'' = [p'']_{512\times512}$ as shown in the following equation:

$$p''(j) = \begin{cases} p'(j) \oplus \left(\text{mod}\left(C_3(j) \times 1000, 256\right)\right), & j = 1, \\ p'(j) \oplus p'(j-1) \oplus \left(\text{mod}\left(C_3 \times 1000, 256\right)\right), \\ j = 2, 3, \ldots, M \times N. \end{cases} \tag{9}$$

Cipher secret image P'' is obtained in four steps with a key sequence K'. Original.

Four steps help to obtain a key sequence K' and cipher secret image P''. The encryption and decryption are carried out based on key sequence and decryption is inverse of encryption.

4.2. Data Embedding. From the appointed position $L(x, y)$, D_{roi} and H are embedded into the LSBs border area and the key is used to control the position. Note that the data owner or third party cannot access the original image content without the encryption key; thus the owner content privacy is protected.

The data is embedded into the encrypted medical image.

Step 1: the LSB is read to obtain D_{roi} and embedded in the border area appointed position $L(x, y)$ and shared key is used to control the position.

Step 2: after vertex ROI information, the LSB-plane ROI is recorded by data hider, denoted as D_{lsb}.

Step 3: the EPR and D_{lsb} are concatenated to form the embedded data W, as shown in

$$W = D_{\mathrm{lsb}} + \mathrm{EPR}. \tag{10}$$

Here, the concatenation operation is indicated using "+," the LSB substitution is used by data hider to embed W, and embedding process end position is pointed out by an ending label into encrypted image except for the border area.

Embedding is performed based on single LSB plane and more information is embedded based on two or more LSB planes.

4.3. Data Extraction and Image Recovery. In the image recovery and data extraction, three cases are analyzed, namely, receiver having (i) only data-hiding key, (ii) only the encryption key, and (iii) both encryption and data-hiding key.

Case 1. The receiver with the data-hiding key can read the embedded data and the original image cannot be obtained. From the border area given position $L(x, y)$, D_{roi} can be read, and then ROI size and vertex information can be obtained. From first pixel to ending label, the LSB-plane is read and embedded W is extracted successfully. The W is separated into EPR and D_{lsb}, as given in (6).

Case 2. The receiver with encryption key roughly recovers the original image and embedded data is not obtained. From the border area of appointed position $L(x, y)$, D_{roi} is read, and then ROI size is obtained. The receiver decrypts the image with the encryption key except the LSB plane. The inverse of encryption process is decryption. The ROI is rearranged to its original position based on D_{roi}. Encrypted image most significant bits (MSB) are not altered by the data embedding operation; the decrypted MSB is same as the original MSB. The decrypted image content is similar to the original image.

Case 3. If the receiver has both encryption key and data-hiding key, both embedded data and losslessly recovered ROI.

Step 1: based on the data hiding key, D_{lsb} and EPR can be obtained.

Step 2: the encrypted ROI LSB plane is recovered with D_{lsb}.

Step 3: from the border area of appointed position $L(x, y)$, D_{roi} and H can be read.

Step 4: the embedded data is calculated in decrypted version according to the encryption key.

Step 5: ROI is losslessly recovered and returns to its original position with D_{roi}.

Step 6: integrity authentication: H' denotes recovered ROI hash message. If H is equal to H', the image is authentic; otherwise ROI has been tampered.

5. Result

Data hiding and image encryption techniques are required in the medical images in the cloud for remote access. In this research, ICE method is proposed to encrypt the images with data hiding method. The 18 medical images from online available dataset (http://imaging.cancer.gov/) were used in this method to test the efficiency of the ICE method. Input medical images in the size of 512 × 512 (16 bits). Each vertex coordinates are represented by 20 bits. The test images used to test model performance are shown in Figure 2.

The proposed ICE method is applied for the data hiding and image encryption to transfer the image in the cloud. In the client side, the decryption is performed to retrieve the original images with hidden data. The input images, rearrange image, encrypted image, encrypted and embedded image, recovered image, and directly decrypted image are shown in Figure 3. The rearrange operation increases the security of the ROI. The embedding rate of each pixel is set as 0.5 and assume that ROI size is set as 5% in the image. The PSNR value of the decrypted image is achieved as 105.12 dB.

The proposed ICE method recovered images PSNR value for various embedding rate is shown in Table 1. The three images such as Im 1, Im 5, and Im 9 were selected for the PSNR analysis in three LSB planes. The average PSNR value of 18 images is presented in the table for three LSB planes and various embedding rate. The increases in embedding rate decrease the PSNR value of the image and single LSB plane has higher quality compared to 2 or 3 planes. The single LSB plane and lower embedding rate have higher quality of image in the analysis. The lower embedding rate has less data in RONI and less distortion in extracting embedded data. The proposed ICE method of recovered images average PSNR value of 18 medical images is given in Table 1.

The increases in the embedding rate decrease the PSNR value of the recovered images. The 1 LSB plane has the higher PSNR than 2 LSB planes and 3 LSB planes. The average PSNR values for 18 images for various embedding rate and LSB plane are presented in Table 2. The less embedding rate has less distortion in the data recovery and payload is less for single LSB plane. The quality of image is high for less embedding rate and single LSB plane in data recovery. The proposed ICE method has the considerable PSNR value for the 0.5 embedding rate in the analysis. The average PSNR value of the ICE method for 0.5 embedding rate is 104.21 and average PSNR value for 0.1 embedding rate is 114.42 dB in the analysis.

The proposed ICE method of directly decrypted images for various embedding rate is shown in Table 2. The increases

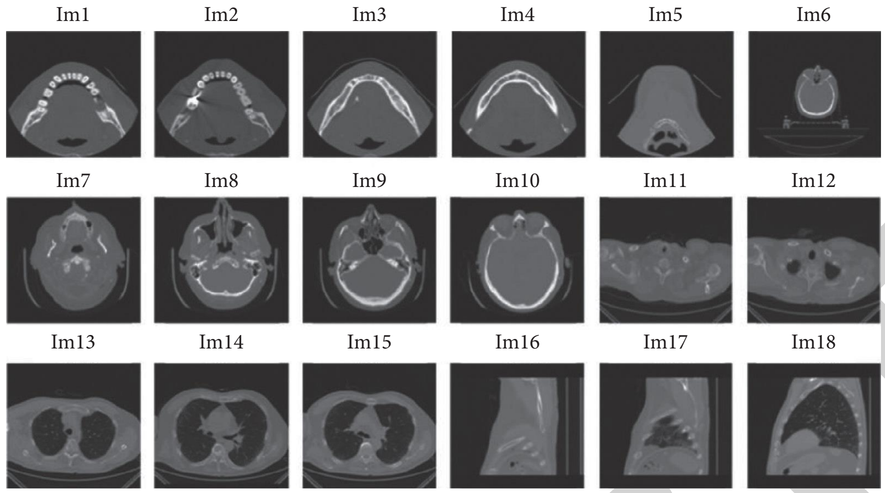


Figure 2: Test images for data hiding and image encryption.

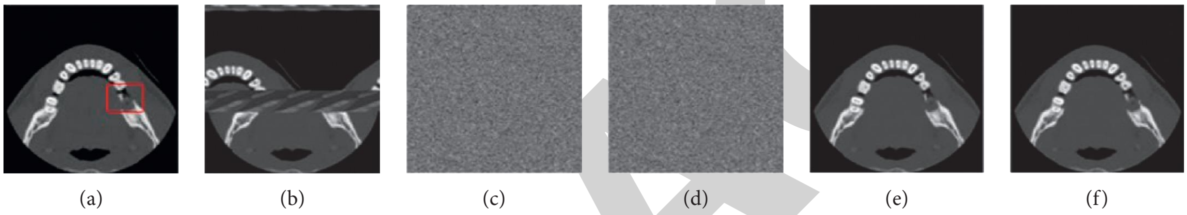


Figure 3: The proposed model: (a) input image, (b) rearranged image, (c) encrypted image, (d) encrypted and embedded image, and (e) directly decrypted image and (f) recovered image.

Table 1: Recovered image PSNR for the various embedding rate.

Embedding rate		PSNR							
		0.005	0.01	0.05	0.1	0.2	0.3	0.4	0.5
Im 1	1 LSB plane	132.41	132.14	118.1	113.06	110.06	107.08	105.6	104.07
	2 LSB planes	115.1	114.87	110.2	105.07	103.12	101.26	100.12	99.76
	3 LSB planes	106.15	105.47	101.76	99.43	97.42	95.41	94.32	93.13
Im 5	1 LSB plane	135.67	133.46	116.36	110.62	109.64	107.68	105.61	104.21
	2 LSB planes	115.61	114.28	110.21	105.61	104.62	101.32	101.16	99.16
	3 LSB planes	105.12	104.21	102.26	101.54	99.42	96.32	95.53	94.21
Im 9	1 LSB plane	134.51	133.43	116.87	113.26	110.43	107.62	106.32	106.42
	2 LSB planes	115.62	114.32	110.72	107.24	105.34	103.16	102.12	101.53
	3 LSB planes	106.21	105.54	103.31	101.41	97.62	96.37	95.21	94.17
Average value	1 LSB plane	134.52	135.47	117.61	114.42	109.21	108.43	105.12	104.21
	2 LSB planes	115.21	114.67	109.37	106.31	103.64	101.21	101.13	99.87
	3 LSB planes	106.72	105.32	102.36	99.24	97.61	95.62	95.12	94.31

in the embedding rate and LSB plane decrease the PSNR value of the directly decrypted images. The ICE model has the considerable PSNR value for the 0.5 embedding rate. The ICE model average PSNR value for 0.5 embedding rate is 105.61 dB and average PSNR of 109.61 dB for 0.1 embedding rate.

The average PSNR values for the 18 medical images of the ICE and existing methods for various embedding rate are compared in Figure 4. The result shows that proposed ICE method has the higher PSNR value compared to existing methods in the image encryption and data hiding. The existing methods [48, 49, 51, 52] have created the error value in the recovery of the images and the proposed model has lower error value compared to the existing models, which improves the image quality.

The ICE model is evaluated in the data hiding and image encryption in terms of SSIM, NPCR, UACI, and Entropy, as shown in Table 3. The proposed ICE model has higher efficiency than existing methods. The ICE method has the advantage of lossless retrieval of image in ROI region and improved the security of the model. The existing model [52] has produced the error in the retrieval process which affects the image quality. The ICE has 99.62 NPCR and existing method [52] has 99.32 NPCR in the image encryption.

Table 2: Directly decrypted images of PSNR for various embedding rate.

Embedding rate		PSNR							
		0.005	0.01	0.05	0.1	0.2	0.3	0.4	0.5
Im 1	1 LSB plane	116.12	115.21	114.16	110.26	107.32	106.41	105.34	105.12
	2 LSB planes	111.23	110.61	108.61	106.32	104.41	102.31	101.21	99.43
	3 LSB planes	105.32	105.21	103.32	101.17	99.43	96.32	95.51	94.21
Im 5	1 LSB plane	115.61	114.32	113.24	110.41	107.64	106.32	105.33	106.43
	2 LSB planes	110.62	109.14	106.62	104.61	102.44	101.47	99.71	98.12
	3 LSB planes	105.61	104.21	103.14	98.42	97.64	95.52	94.61	93.16
Im 9	1 LSB plane	115.71	116.46	114.21	109.61	107.61	105.31	105.21	103.57
	2 LSB planes	111.49	110.76	107.81	105.65	103.47	103.45	101.32	99.46
	3 LSB planes	106.21	104.24	101.26	99.24	97.45	96.78	95.51	96.12
Average value	1 LSB plane	116.45	115.52	113.24	109.61	107.61	106.61	105.52	105.61
	2 LSB planes	111.26	110.17	106.42	105.52	103.32	101.12	101.31	98.16
	3 LSB planes	105.52	103.41	103.31	101.31	98.43	95.51	94.51	94.45

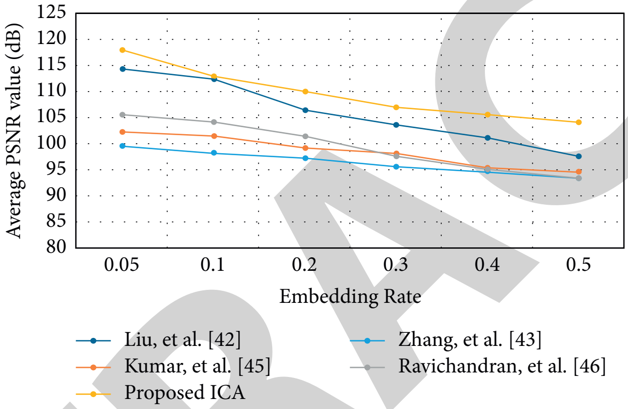


Figure 4: The average PSNR value of the proposed ICE method.

Table 3: Comparative analysis of the proposed model.

Methods	SSIM	NPCR	UACI	Entropy
Liu et al. [48]	0.9999	99.21	33.12	7.9922
Zhang et al. [49]	0.9941	99.24	33.17	7.9941
Kumar et al. [51]	0.9934	99.31	33.22	7.9953
Ravichandran et al. [52]	0.9972	99.32	33.24	7.9962
Proposed ICE	0.9999	99.62	33.41	7.9974

The proposed ICE method is compared with existing chaotic map and other encryption methods over plain-text attack, as shown in Table 4. The proposed ICE method has advantage of applying Lorenz 96 method to increase the sensitivity to initial conditions and control parameters. The existing chaotic map [41, 42] has limitation of lower sensitivity and periodic degradation on account of finite precision in the process. The wavelet transform [40] method has limitations of lack of phase information, poor directionality, and shift sensitivity. The fuzzy encryption [32] has required supervised features and PSO method [37] has lower convergence in parameter settings. The proposed ICE method has PSNR value of 134.52 and existing hybrid chaotic map has 117.3 PSNR value.

Table 4: Comparison of the proposed ICE with existing chaotic map and other encryption methods over plain-text attack.

Methods	PSNR
Fuzzy [42]	102.7
PSO [47]	106.5
SVD-DWT [50]	109.8
FrDCT [51]	115.1
Hybrid chaotic map [52]	117.3
Proposed ICE method	134.52

6. Conclusion

Medical images stored in the cloud require encryption to increase the security of the patient information. Data hiding method is applied to store the patient data in the medical image in the hidden format. In this research, the ICE method is applied to improve the security of the medical image and improves the quality of the image. The medical images were used to evaluate the performance of the proposed ICE and existing method. The proposed method is tested for quality in various embedding rate. The Lorenz 96 model is applied in the chaos method to improve the security based on increases of random value. The result shows that the proposed ICE method has 104.07 dB PSNR and the existing model has 97.61 dB PSNR value. The future work of the proposed model involves an effective data hiding technique to improve the embedding rate of the model.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Hubei Natural Science Foundation under grant 2021CFB156.

References

[1] C. Feng, B. Liu, K. Yu, S. K. Goudos, and S. Wan, “Blockchain-empowered decentralized horizontal federated learning for 5G-enabled UAVs,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3582–3592, 2021.

[2] Z. Guo, K. Yu, A. Jolfaei, F. Ding, and N. Zhang, “Fuz-spam: label smoothing-based fuzzy detection of spammers in internet of things,” *IEEE Transactions on Fuzzy Systems*, p. 1, 2021.

[3] K. Yu, L. Tan, S. Mumtaz et al., “Securing critical infrastructures: deep-learning-based threat detection in IIoT,” *IEEE Communications Magazine*, vol. 59, no. 10, pp. 76–82, 2021.

[4] D. Wang, Y. He, K. Yu, G. Srivastava, L. Nie, and R. Zhang, “Delay sensitive secure NOMA transmission for hierarchical HAP-LAP medical-care IoT networks,” *IEEE Transactions on Industrial Informatics*, p. 1, 2021.

[5] L. Tan, K. Yu, N. Shi, C. Yang, W. Wei, and H. Lu, “Towards secure and privacy-preserving data sharing for COVID-19 medical records: a blockchain-empowered approach,” *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 271–281, 2021.

[6] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, and T. Sato, “Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote E-health,” *IEEE Wireless Communications*, vol. 28, no. 3, pp. 54–61, 2021.

[7] L. Yang, K. Yu, S. X. Yang, C. Chakraborty, Y. Lu, and T. Guo, “An intelligent trust cloud management method for secure clustering in 5G enabled internet of medical things,” *IEEE Transactions on Industrial Informatics*, p. 1, 2021.

[8] L. Zhen, Y. Zhang, K. Yu, N. Kumar, A. Barnawi, and Y. Xie, “Early collision detection for massive random access in satellite-based internet of things,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 5184–5189, 2021.

[9] T. Guo, K. Yu, M. Aloqaily, and S. Wan, “Constructing a prior-dependent graph for data clustering and dimension reduction in the edge of AIoT,” *Future Generation Computer Systems*, vol. 128, pp. 381–394, 2022.

[10] L. Tan, K. Yu, F. Ming, X. Chen, and G. Srivastava, “Secure and resilient artificial intelligence of things: a HoneyNet approach for threat detection and situational awareness,” *IEEE Consumer Electronics Magazine*, 2021.

[11] Q. Zhang, K. Yu, Z. Guo et al., “Graph neural networks-driven traffic forecasting for connected internet of vehicles,” *IEEE Transactions on Network Science and Engineering*, p. 1, 2021.

[12] Z. Yin, Y. Xiang, and X. Zhang, “Reversible data hiding in encrypted images based on multi-MSB prediction and Huffman coding,” *IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 874–884, 2020.

[13] C. Qin, Z. He, X. Luo, and J. Dong, “Reversible data hiding in encrypted image with separable capability and high embedding capacity,” *Information Sciences*, vol. 465, pp. 285–304, 2018.

[14] G. Gao, X. Wan, S. Yao, and Z. Cui, C. Zhou and C. Sun, “Reversible data hiding with contrast enhancement and tamper localization for medical images,” *Information Sciences*, vol. 385-386, pp. 250–265, 2017.

[15] B. Ma, B. Li, X.-Y. Wang, C. P. Wang, J. Li, and Y.-Q. Shi, “A code division multiplexing and block classification-based real-time reversible data-hiding algorithm for medical images,” *Journal of Real-Time Image Processing*, vol. 16, no. 4, pp. 857–869, 2019.

[16] K. A. Kumari, A. Sharma, C. Chakraborty, and M. Ananyaa, “Preserving health care data security and privacy using carmichael’s theorem-based homomorphic encryption and modified enhanced homomorphic encryption schemes in edge computing systems,” *Big Data*, 2021.

[17] H. H. Attar, A. A. Solyman, A. Alrosan, C. Chakraborty, and M. R. Khosravi, “Deterministic cooperative hybrid ring-mesh network coding for big data transmission over lossy channels in 5G networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, pp. 1–18, 2021.

[18] V. Ravi, H. Narasimhan, C. Chakraborty, and T. D. Pham, “Deep learning-based meta-classifier approach for COVID-19 classification using CT scan and chest X-ray images,” *Multimedia Systems*, pp. 1–15, 2021.

[19] J. Amin, M. Sharif, N. Gul, S. Kadry, and C. Chakraborty, “Quantum machine learning architecture for COVID-19 classification based on synthetic data generation using conditional adversarial neural network,” *Cognitive Computation*, pp. 1–12, 2021.

[20] C. Feng, K. Yu, M. Aloqaily, M. Alazab, Z. Lv, and S. Mumtaz, “Attribute-based encryption with parallel outsourced decryption for edge intelligent IoV,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13784–13795, 2020.

[21] L. Tan, K. Yu, L. Lin et al., “Speech emotion recognition enhanced traffic efficiency solution for autonomous vehicles in a 5G-enabled space-air-ground integrated intelligent transportation system,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2021.

[22] Y. Sun, J. Liu, K. Yu, M. Alazab, and K. Lin, “PMRSS: privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1981–1990, 2022.

[23] Y. Gong, L. Zhang, R. Liu, K. Yu, and G. Srivastava, “Nonlinear MIMO for industrial internet of things in cyber-physical systems,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5533–5541, 2021.

[24] F. Ding, G. Zhu, M. Alazab, X. Li, and K. Yu, “Deep-learning-empowered digital forensics for edge consumer electronics in 5G HetNets,” *IEEE Consumer Electronics Magazine*, p. 1, 2020.

[25] L. Liu, J. Feng, Q. Pei et al., “Blockchain-Enabled secure data sharing scheme in mobile-edge computing: an asynchronous advantage actor-critic learning approach,” *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2342–2353, 2021.

[26] W.-L. Shang, J. Chen, H. Bi, Y. Sui, Y. Chen, and H. Yu, “Impacts of COVID-19 pandemic on user behaviors and

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9856473, 1 page
https://doi.org/10.1155/2023/9856473

Retraction

Retracted: Electronic Health Record Monitoring System and Data Security Using Blockchain Technology

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] K. T. Akhter Md Hasib, I. Chowdhury, S. Sakib et al., "Electronic Health Record Monitoring System and Data Security Using Blockchain Technology," *Security and Communication Networks*, vol. 2022, Article ID 2366632, 15 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2366632, 15 pages
https://doi.org/10.1155/2022/2366632

Research Article

Electronic Health Record Monitoring System and Data Security Using Blockchain Technology

Kazi Tamzid Akhter Md Hasib,[1] **Ixion Chowdhury**,[1] **Saadman Sakib**,[1] **Mohammad Monirujjaman Khan**,[1] **Nawal Alsufyani**,[2] **Abdulmajeed Alsufyani**,[2] **and Sami Bourouis**[3]

[1] *Department of Electrical and Computer Engineering, North South University, Bashundhara, Dhaka 1229, Bangladesh*
[2] *Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia*
[3] *Department of Information Technology, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia*

Correspondence should be addressed to Mohammad Monirujjaman Khan; monirujjaman.khan@northsouth.edu

Received 12 October 2021; Revised 24 November 2021; Accepted 13 January 2022; Published 4 February 2022

Academic Editor: Chinmay Chakraborty

Bangladesh should have owned a decentralized medical record server. We face a lot of issues, such as doctor's appointments, report organization in one spot, and report follow-ups. People now bring a large number of papers to the doctor's chamber. They carry prescriptions, reports, and X-ray files, among other things. It complicates everyone's life as a result. All of the reports must be reviewed by doctors on a regular basis. It is difficult to read old reports on a regular basis, and patients do not receive the correct medications or treatment. Doctors also find it extremely difficult to comprehend handwritten prescriptions. Data security, authenticity, time management, and other areas of data administration are dramatically improved when blockchain (smart contract) technology is linked with standard database management solutions. Blockchain is a groundbreaking, decentralized technology that protects data from unauthorized access. After smart contracts are implemented, the management will be satisfied with the patients. As a result, maintaining data privacy and accountability in the system is tough. It signifies that the information is only accessible to those who have been authenticated. This study focuses on limiting third-party engagement in medical health data and improving data security. Throughout the process, this will improve accessibility and time efficiency. People will feel safer during the payment procedure, which is the most significant benefit. A smart contract and a peer-to-peer encrypted technology were used. The hacker will not be able to gain access to this system since this document uses an immutable ledger. They will not be able to change any of the data if they gain access to the system. If the items are found to be defective, the transaction will be halted. Transaction security will be a viable option for recasting these problems using cryptographic methodologies. We developed a website where patients and doctors will both benefit because of the use of blockchain technology to ensure the security of medical data. We have different profiles for doctors and patients. In the patient profile, they can create their own account by using a unique address, name, and age. This unique address will be created from the genesis block. The unique address is completely private to the owner, who will remain fully secure in our network. After creating an account, the patient can view the doctors' list and they can upload their medical reports such as prescriptions and X-rays. All the records uploaded by the patient will be stored on our local server (Ganache). The records are stored as hashed strings of the data. Those files will also have a unique address, and it will be shown in the patient profile. After granting access, the doctors will be able to view their records in the respective doctor's profile. For accessing the options such as uploading, viewing, or editing the data, Ethereum currency (a fee) will have to be paid in order to complete the request. On the other hand, doctors can enter their profile using their name and unique address. After logging in, they can view their name, unique address, and the list of patients that have granted access to the doctor to view their files. On our website, the front end is handled by JavaScript, ReactJS, HTML, and CSS. The backend is handled by Solidity. Storage is handled by Ganache as the local host. Finally, this paper will show how to ensure that the procedure is as safe as feasible. We are also maintaining transparency and efficiency here.

1. Introduction

The electronic health record is called an EHR. As we know, we are heading towards the fourth industrial revolution. For that, we need to be ready to tackle data. In the modern world, data are an asset. Moreover, medical data is so important and it has security issues as well. Every system cannot handle this security. Critical data in today's global market has a huge impact on the worldwide economy. In the modern world, the revolution in EHRs (electronic health records) has changed medical care and management drastically. People could not access their medical data from their devices before. Because of this system, people from all over the world can access their medical data through websites. We made a website that has some revolutionary features, including doctors' and patients' different profiles. They can interact with each other using the blockchain network. A simple blockchain system can be defined by an ordered list of the nodes and links that are used, with these nodes having the ability to store information and chains connecting these links [1]. The block contains the data. It ensures the privacy of the data. Every block is integrated with each other. No one can temper the medical data easily, either. According to the blockchain security ecosystem, the confidence is based on the increased security, better transparency, and immediate traceability provided by blockchain technology. Beyond issues of trust, blockchain provides a slew of other business advantages, including cost savings as a result of improved speed, efficiency, and automation, among other things. Blockchain lowers overhead and transaction costs substantially by decreasing or eliminating the requirement for third-party or middle-man verification of transactions. Blockchain does this by drastically reducing paperwork and mistakes. The medical record systems currently used by healthcare institutions are highly susceptible to alteration, fabrication, and the risk of getting lost. Additionally, a patient with a long history of medical complications may face the hassle of having to carry a lot of reports to their doctors. Healthcare institutions may record their patients' information digitally, but it will always require a person to verify and insert it into their database. This is an unreliable way of managing such sensitive information. The whole system of record keeping requires patients to travel from one help desk to another to get the information they require. These issues are neglected and some are considered standard protocol in our healthcare industry. Moreover, there are incidents where doctors have prescribed wrong medications based on an unverified report or a patient has had to return without receiving consultation because they forgot to bring certain reports [1–4].

In [5], the authors discussed that to validate and store data, blockchain technology employs blockchain data structures. The data received from the end user are first encrypted so that it is only visible to them. It is then verified and added to blocks in a decentralized system, ensuring data security and accessibility. The data are stored in a block in key and value pairs where each of the blocks is well interconnected to each other. The system is alerted when one of the data contained in the blockchain is tampered with, and the blocks become fragmented. Moreover, all the data stored in the blockchain will be completely transparent to all entities while being encrypted. No one can view them unless the owner gives permission for it. The authors of the paper [6] emphasized that they solved the health record system because it was difficult to monitor and protect against potential threats. The writers have taken the distributed computing idea of blockchain and adapted it to a variety of applications. They highlighted a few of the potential uses for this kind of decentralization. This technology is becoming more widely used, and its primary use is data exchange, which includes access control and administration of patients' medical information. We analyzed one existing similar system and found the author [7] mentioned a cloud-based medical record system that encrypts medical data using the ABE method. These records describe the patient's condition without revealing the exact nature of the disease or the doctor's location. In reality, the article uses symmetric encryption, but one of the strategy's fundamental weaknesses is that it relies on a perfectly trustworthy global authority to handle key management, issue public system settings, and generate secret keys for the doctor. To aid the network in achieving an agreement, the method depends on a single centralized point of trust for transactions. A technique for constructing a conceptual framework for a medical health record management system is included in the study. Its main aim is to guarantee safe transactions by using blockchain and smart contracts.

In [8], the authors figured out that a blockchain is an ever-growing collection of data known as cryptographically linked and secured blocks that are cryptographically linked and protected. Cryptography is used to connect the blocks. The bulk of nodes in the blockchain network are in charge of block verification. When we go further, there are a few traits that distinguish blockchain from other technologies. Data stored on the blockchain, for example, are immutable, tamper-resistant, and based on a decentralized network, so it cannot be hacked or decoded in any way. In general, there are three sorts of blockchain: public (or unapproved), private (or allowed), and consortium (or both private and allowed) (or permitted). As a consequence of the network's geographically varied geographic area, each one has its own distinct characteristics. Smart contracts, according to [9] Nick Szabo, are "a computerized transaction protocol that complies with contract conditions." Smart contracts are code snippets written in a high-level programming language that are used to complete transactions (software or scripts). There are programming languages such as Java, C++, Python, NodeJS, Go, and Solidity. Many blockchain systems utilize a variety of high-level programming languages to create smart contracts.

In [10], the authors indicate that between 2013 and 2014, the number of office-based physicians who utilize certified electronic health record systems increased. In 2014, 74.1 percent of office-based physicians utilized a certified EHR system. According to the findings of the study, 56.8% of doctors in Alaska and 88.6% of doctors in Minnesota utilize a certified EHR system. The HITECH Act of 2009 offered monetary incentives to qualified physicians who used a certified electronic health record system, which may be one of the reasons for the continued rise in physician usage of these systems. According to the American Medical Association, physicians who utilized a certified EHR system shared patient information with clinicians outside their medical group in 2014.

The authors discussed in [11] that electronic health records (EHRs) are used for research all around the globe. The ability to analyze real-world real-patient results in near-real time, which is not feasible with other techniques, is a clear and unique benefit. Whether done prospectively or retrospectively, the cost and convenience of evaluating existing data are less expensive and more convenient than the cost and convenience of creating a human-curated dataset. The future potential for EHR in research seems to be almost unlimited, especially as the types of data collected and the ability to extract information from records continue to improve. The EHR is being gradually changed and upgraded to make research more accessible as the usage of electronic health data for research increases in popularity. The North American Association of Central Cancer Registries and the Standards for Oncology Registry are attempting to establish data format and display standards, as well as common data components, such as the National Cancer Institute. Structured data, according to E. Kim and colleagues, is more likely to contain this information. It may be found in the story (+/machine comprehensible). It is unlikely that it will show up in EHR Facts and Statistics [12].

In [13], the authors showed that the International Data Encryption Standard Algorithm (IDEA) with salt will be used to protect patient-sensitive data during transmission. The session will be protected via the usage of JSON Web Tokens, which internally use the Signature Algorithm HS256. The SHA256 algorithm with salt will be used to protect the confidentiality of passwords. Everything done so far is towards at protecting electronic medical data (EMD) against external theft. Providing integrity and privacy for sensitive data does not just relate to external invaders but also applies to data leaks that may occur inside an organization. In [14], the authors emphasize that blockchain technology is at the heart of data acquisition since it guarantees the integrity and dependability of information. Tamper-proof and open class verification features guarantee that the information on the block cannot be tampered with in any way. Even if the attacker altered part of the ledger information, the system would detect and fix the error in a short period of time. In order to create network congestion, the authors also said [15] an adversary may attempt to submit a large number of requests to endorsers at the same time. In some ways, it is comparable to denial of service (DoS) assaults. It serves no purpose since the endorser only responds to client requests by verifying that the data have been signed by the client. Even if the expense is huge, the impact is small. Additionally, nodes in the system may be attacked, crash, or even turn against one another. The consensus process and the election of endorsers may both help to maintain the stability of the system and reduce the likelihood that opponents will do significant harm. In addition to ensuring data privacy, the access control mechanism on the ledger may provide results similar to those obtained via ring signatures and zero-knowledge proof. A highly effective method of protecting the privacy of patients is through the use of encryption. In [16], the authors mentioned that our proof-of-stake method, which incorporates implicit bonding, makes it simpler to audit each entity's involvement in the system while simultaneously giving large payers (such as the Centers for Medicare and Medicaid Services) a simple option to promote participation. In reality, during the time of image collection, the institution provides resources for the study. In [17], the authors also discussed how the peer-to-peer network of nodes and associated blockchain architecture are design decisions that effectively address another interoperability criterion: the sharing of protected health information (PHI) with patient-authorized recipients over a secure, private, and tamper-resistant network infrastructure. As previously mentioned, the chaining technique produces an effectively immutable data record, meaning that any attempt to change existing blocks will be immediately visible to the observer.

In [18], the authors described how MetaMask is also used to store patient and healthcare provider keys. MetaMask is a bitcoin wallet that works with Chrome, Firefox, and Microsoft Edge. MyEtherWallet and MyCrypto are two more cryptocurrency wallets. MetaMask's goal is to make connecting to the Ethereum network as easy, reliable, and secure as possible. Furthermore, since its inception in 2016, MetaMask has not been hacked severely. MetaMask has become a useful tool in our work. In this article, we present a system that satisfies the security, privacy, interoperability, and performance requirements of an EHR system. Patient data may be shared anywhere and at any time using the recommended method as long as the patient grants permission. Only a few authors [19] have shown that the aforementioned EHR system criteria are covered by existing research. However, all of these criteria are taken into account in this project to guarantee patient data security and privacy, interoperability, and performance needs are met. To fulfill these criteria, our study employs blockchain technology, which employs a sophisticated and long-lasting cryptographic technique to allow cross-healthcare-provider data exchange while giving patients ownership over their information. The goal of the study is to create and evaluate a blockchain-based electronic health record application for security, privacy, interoperability, and performance. The authors [20] suggest a method for electronic transactions that does not depend on the confidence of the participants. As a starting point, we used the traditional structure of coins created using digital signatures, which offers tight control over ownership but is insufficient since it lacks a mechanism to prevent double-spending. A proof-of-work network, which records a public history of transactions and makes it computationally difficult for an attacker to alter the history if honest nodes hold a majority of CPU power, was suggested to address this challenge. Because of its unorganized simplicity, the network is very durable. Nodes operate in a synchronous fashion with minimal cooperation. They do not need to be identified since communications are not directed to any specific location and simply need to be delivered using the best available technology. With the help of this consensus mechanism, all necessary regulations and incentives may be implemented [21]. PHR described it has a straightforward interface that allows you to quickly go through the following options: appointments give you the freedom to schedule appointments with your physician and canceling them before the appointment if necessary. A directory of your medical history, medicines, allergies, and vaccination records is kept in your health history file. It is simple to view the records of claims that have been filed on your behalf when you use the Claims History feature [22]. Furthermore, [23] by providing you with a database containing all of the lab findings for the tests you have performed, you may save both time and effort on your part. Connecting with your provider through direct and secure messaging is made possible by secure messaging services. Demographics: it allows you to make changes to any of your personal information, such as your phone number or address.

We also notice that [24] cloud-based electronic health record platform, GenexEHR, was developed to assist clinics and hospitals in managing patient data, thus allowing informed choices. Developed with the patient in mind, this platform assists in the preservation of the patient's health information across several healthcare settings. Using its state-of-the-art healthcare software, Genex HI, hospital management and information solution, meets the software and workflow needs of hospitals, clinics, and even individual patients. GenexEHR, Inc., provides Software-as-a-Service Laboratory and Clinic Management software. It is possible that latency will be impacted if there is an excessive amount of data. Because of the low storage and computing needs, blockchain transactions will be extremely cost-effective in terms of storage and processing. Another method of increasing performance is using decentralized databases such as BigchainDB and HBasechainDB. In the case of a large-scale deployment, the system may also contain tracking devices as well as extra actors. The quantity of data that can be stored on the blockchain is growing, so we may need to use off-chain architecture to store the original data while keeping the evidence of existence on the blockchain. There is a possibility that this may become a future study subject for this inquiry [25].

In this paper, we created a website that benefits both patients and physicians since we are using blockchain technology to guarantee the confidentiality of medical data. Our website has distinct profiles for physicians and patients. They may establish their own account under the patient profile by providing a unique address, name, and age. This one-of-a-kind address will be generated from the genesis block and cannot be entered into anyone's profile. The owner's unique address is totally confidential and will stay completely safe in our network. After establishing an account, the patient may see a list of physicians and upload medical records such as prescriptions and X-rays. All of the patients' submitted records will be kept on our local server (Ganache). The records are kept in the form of hashed strings containing the data. Additionally, each file will have a unique Uniform Resource Locator (URL) that will be shown in the patient's profile. There will also be an option for patients to give physicians access to their medical records. After being granted access, physicians will be able to see their records in their profile. To get access to features such as uploading, viewing, or modifying data, Ethereum money (a charge) will need to be paid. Doctors, on the other hand, may create a profile using their name and a unique address. They may see their name, unique address, and a list of patients who have given the doctor permission to read their files after signing in. The front end of our website is powered by JavaScript, ReactJS, HTML, and CSS. Solidity manages the backend, while Ganache acts as the local host. Finally, this article will show how to preserve the method's safety while simultaneously preserving its transparency and efficiency.

Section 1 explores and briefly explains blockchain, cryptocurrency, and smart contracts. Section 2 includes a description of the technique and materials used, as well as a discussion of the issues and a diagram of the entire system and how it all works together. The findings and analysis, as well as a comparison of the EHR to current systems, are presented in Section 3. Finally, Section 4 discusses the conclusion and future prospects.

2. Method and Materials

Healthcare institutions' existing medical record systems are extremely vulnerable to modification, falsification, and the possibility of data loss. Additionally, a patient with a lengthy history of medical problems may have to deal with the inconvenience of having to bring a large number of reports to their doctor's appointments. Healthcare organizations may choose to digitize the information they collect about their patients, but a human will always be required to verify and enter the information into their database. When dealing with such sensitive material, this is an untrustworthy method of handling it. Under the current record-keeping system, patients are required to go from one help desk to another in order to get the information they need. The majority of these problems are ignored, and some of them are even considered normal practice in our healthcare sector. Furthermore, there have been instances in which physicians have given incorrect medicine based on an unconfirmed report, or in which a patient has been required to return without getting consultation because they failed to bring the necessary papers with them to the appointment. Blockchain technology makes use of blockchain data structures in order to verify and store data. The information received from the end user is first encrypted in order to ensure that it is only accessible to that individual. Data security and accessibility are ensured in a decentralized system by having it validated and added to blocks in a decentralized system. The data are kept in a block in key and value pairs, with each of the blocks being properly linked to the previous and following blocks. When one of the data blocks in the blockchain is tampered with, the system receives an alert. If the blocks become fragmented, the system is notified. Furthermore, all data saved in the blockchain will be fully visible to all entities while also being encrypted so that no one will be able to see them until the owner grants permission for them to be viewed by others. A blockchain-based system for storing and retrieving medical data will make the whole process of updating, retrieving, and displaying medical data much more frictionless. As a result, inside the blockchain architecture, the hospital's activities might be viewed as information services. In a blockchain-driven cyber environment that resembles real-world EHR operations, smart contract design may thus be viewed of as the computation of start and finish times for information services.

2.1. Outline of Full System. Figure 1 shows the EHR architecture of the proposed system in this paper. The primary users in the blockchain architecture are a doctor, a patient, and a hospital. Furthermore, in the blockchain ecosystem, everyone will have a unique power due to the system's own private keys. Doctors, patients, and hospitals may also have limited access to some functionalities. Using his or her private key, the doctor will gain access to the web app (Frontend). By using his own private key, the patient communicates with the doctor. Both parties will share medical information via a web app (Frontend). Patients must use a smart contract to transact after receiving the service. As a backend support, we are deploying the Ethereum network. Below, the roles of every character in this figure are described:

(1) Doctor: doctors can access the system by using a private key, which will be given by the patient's unique key.

(2) Patient: patients will provide private key to doctors. Patients can share personal data through our portal.

(3) Hospital: hospital admin can look into the process but not details. They can maintain the flawless communications system.

(4) Website: the website is under frontend programming. It connects with the backend through smart contracts.

(5) Smart contracts: It is the main bone of our system. Every change in a block will have a log in it.

(6) Ethereum network: for backend processing, we used the Ethereum network.

2.2. Process of System. Figure 2 shows that the patients must first visit the hospital for a doctor's consultation. Both the patient and the doctor must have a network account. A new patient must first register an account and fill out his or her profile's primary information. After filling out the form, the doctor will search the network for his or her information and consult with the patient. The hospital will update the patient's information on the blockchain network after checking with the doctor. So, all the processes are connected to the blockchain network with websites.

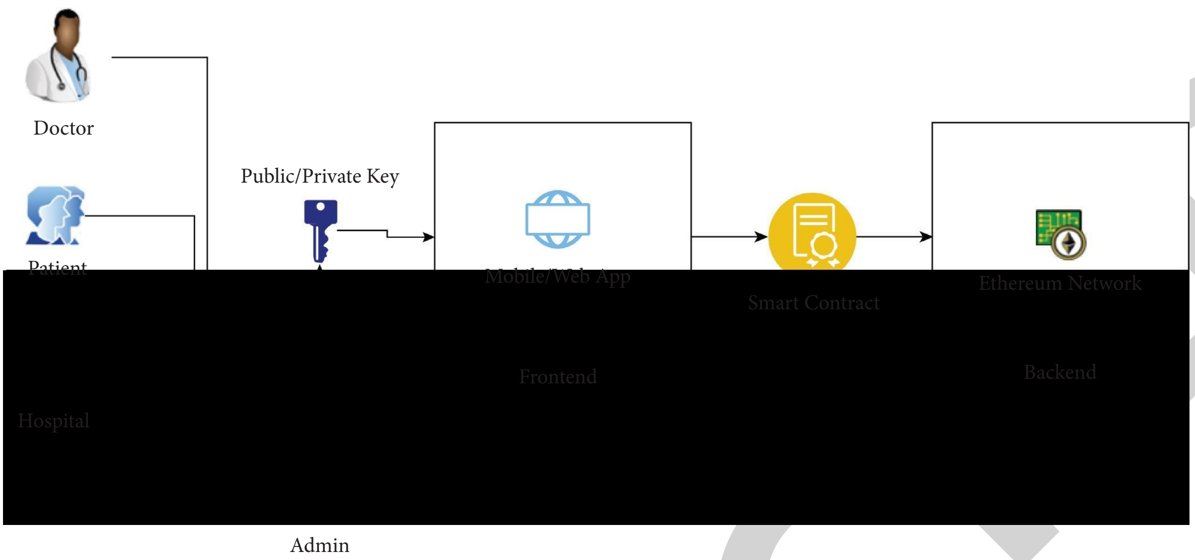
As shown in Figure 3, the hospital has access to the patient's public key, but only the patient has access to his or her own private key. If a doctor wishes to see a patient's medical records, he or she must make a request to the patient. When the patient receives the queue request in his mobile app or on his website, he or she can authorize access by inputting their private key. After the operation is completed, the blockchain network will be updated.

Figure 4 shows an example of an unauthorized user cannot access the EHR application focused on the patient since the patient's data must demonstrate its security. The SHA-3 (Keccak-256) hash, which generates a 64-character hexadecimal string, can be used to get access to the program. The patient's public key is used to encrypt each transaction and can only be decoded with the patient's private, encrypted key, thanks to cryptography. The elliptic curve digital signature algorithm protects all nodes on the Ethereum network (ECDSA).

The functions of individual entities are depicted in Figure 5. When we look into the node in our network, we observe blocks and blocks are interconnected in the blockchain network. A block, as shown in the diagram above, consists of two parts: a block header and a block body. A ledger is a collection of linked blocks. The patient's public key is contained in each block. The preceding transaction's hash, the encrypted data hash, and the service provider's digital signature must all be provided. A copy of the blockchain ledger will be sent to each node in the network. Inside every block, there is some mechanism of the blockchain ecosystem. In the block header, there is a previous block hash, signature, root hash, timestamp, and nonce. We have described it all before. Inside the block body, we have found the root hash. Every root is interconnected with every block. We see that Tx1. Tx1 means 1 transaction happened, and so, we found the hospital, doctor, diagnosis, date, and signature as it is unique.

Figure 6 shows the entire system's entity relationship (ER) diagram. We concentrated on patient, doctor, and lab report storage. First and foremost, the patient can visit the websites and create an account. Users can register and choose a hospital, but we are only proposing one. The user can view their prescription history as well as other medical information. In the future, new features will be added to our project's tasks. If we see the main features in the figures, they are the patient, doctor, and laboratory. Patients are connected with Login, Signup, Select Hospital, View Past Records, and Make an Appointment. We introduced patients to doctors by providing them with treats. Doctors are linked with the ability to view appointments, diagnose patients, make appointments with patients, and also add lab tests. Doctors are interconnected with laboratories. Because he/she can make decisions for patients, they can tell a patient if a lab test is mandatory or not. We discovered that we can view the lab test, generate a lab test report, and perform lab tests using the lab test.

Figure 7 shows the login page. It is for both the doctor and the patient. Both must fill in their names and log in by address to their respective accounts in order to log in. You can create three distinct sorts of accounts on this page. The



[redacted]

Figure 1: EHR architecture of our system.

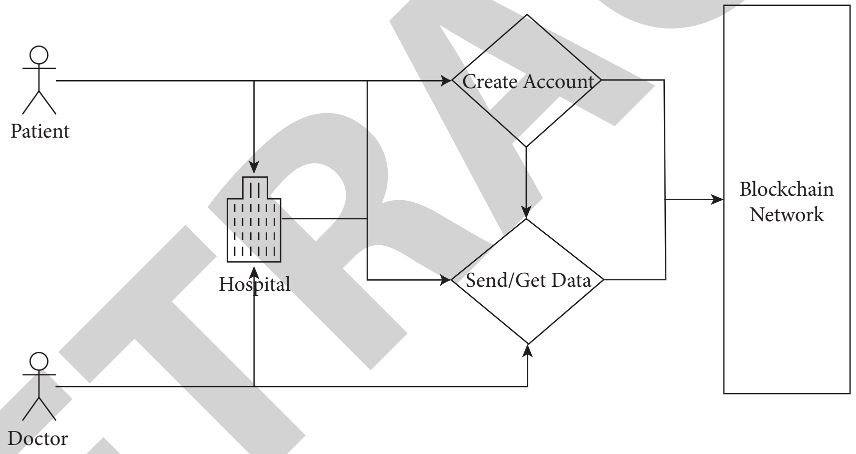


Figure 2: Process of system.

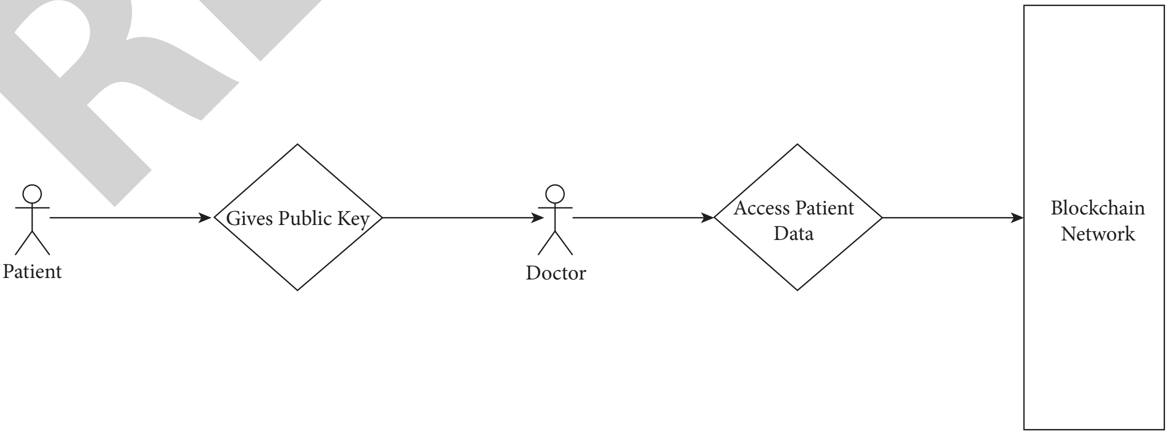
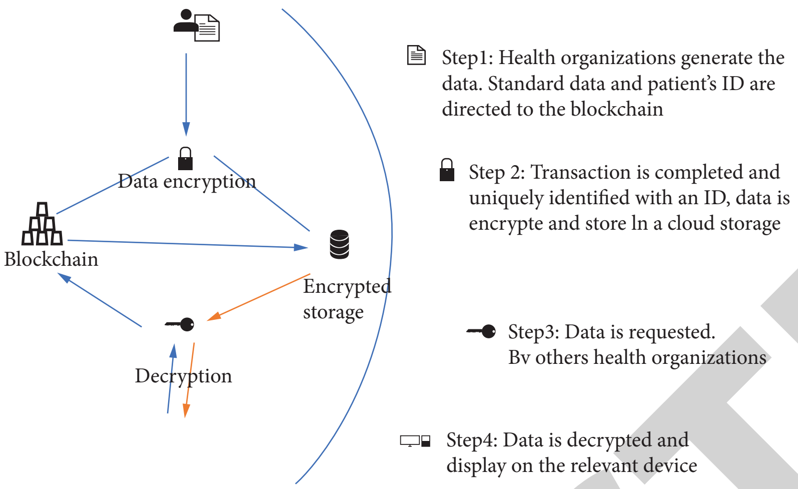


Figure 3: Authentication of public key and blockchain network.



Step1: Health organizations generate the data. Standard data and patient's ID are directed to the blockchain

Step 2: Transaction is completed and uniquely identified with an ID, data is encrypte and store ln a cloud storage

Step3: Data is requested. Bv others health organizations

Step4: Data is decrypted and display on the relevant device

Figure 4: Blockchain data storage.

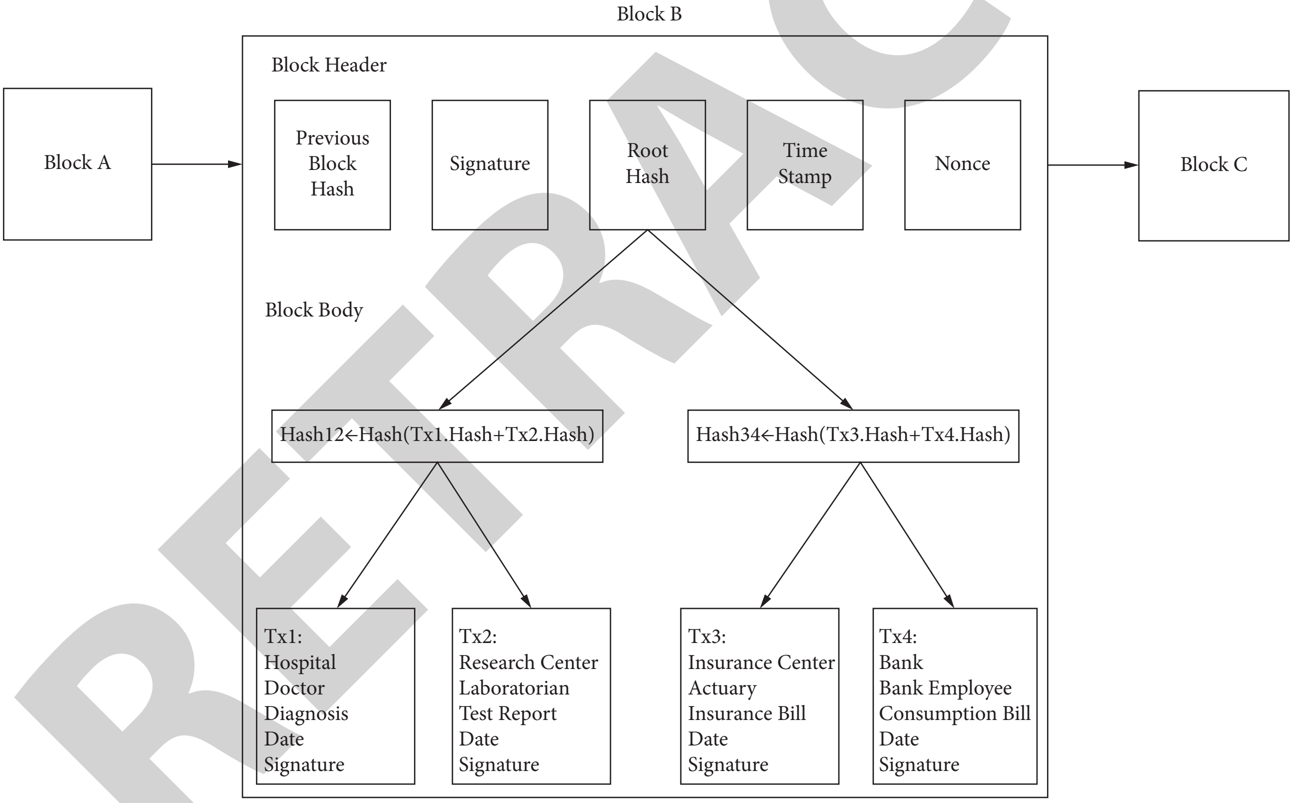

Block B
Block Header
Block A
Previous Block Hash
Signature
Root Hash
Time Stamp
Nonce
Block C
Block Body
Hash12←Hash(Tx1.Hash+Tx2.Hash)
Hash34←Hash(Tx3.Hash+Tx4.Hash)
Tx1: Hospital Doctor Diagnosis Date Signature
Tx2: Research Center Laboratorian Test Report Date Signature
Tx3: Insurance Center Actuary Insurance Bill Date Signature
Tx4: Bank Bank Employee Consumption Bill Date Signature

Figure 5: Block details in chain.

admin account, patient account, and doctor account are the three key aspects of this page. The username and private key will be required if the user already has an account. A private address is required if someone is creating one for the first time. Every action can be tracked and controlled. JavaScript, CSS, HTML, and Solidity were used to create this website system.

Figure 8 shows the signup page. Figure 8 shows how the data will be typed and stored in the local database once the user fills out the registration form with information such as their username and private address. The information may be viewed via the administrative control panel.

2.3. Data Security and Transaction in Blockchain and Smart Contract. Transactions do not appear in the block by accident. The list must be empty when one transaction occurs. Otherwise, it will not receive all of the individual data. Each block cannot contain the same data. If a majority of nodes agree on a transaction, it is included in a block. The

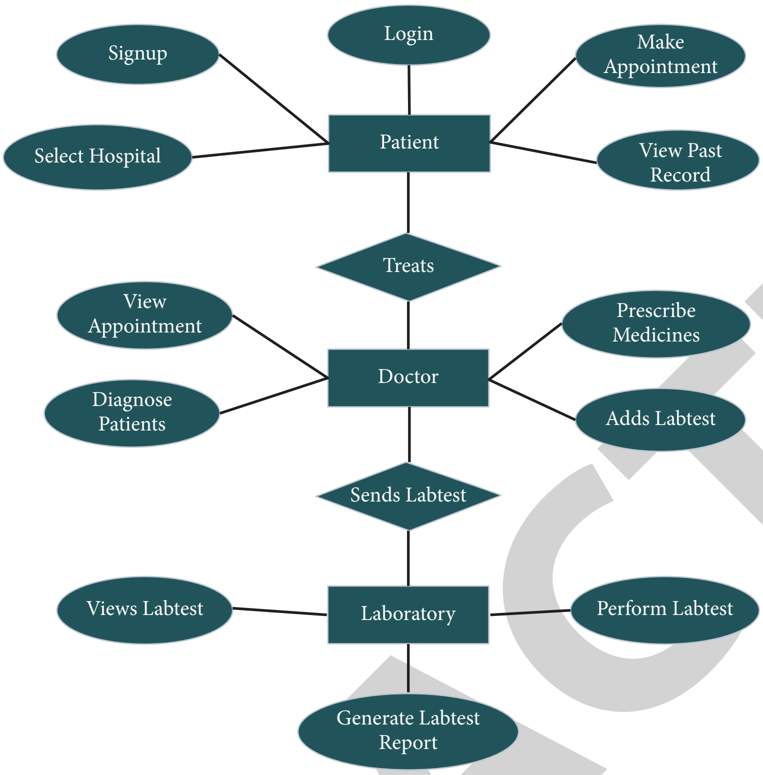


Figure 6: Entity relationship diagram of electronic health record (EHR).

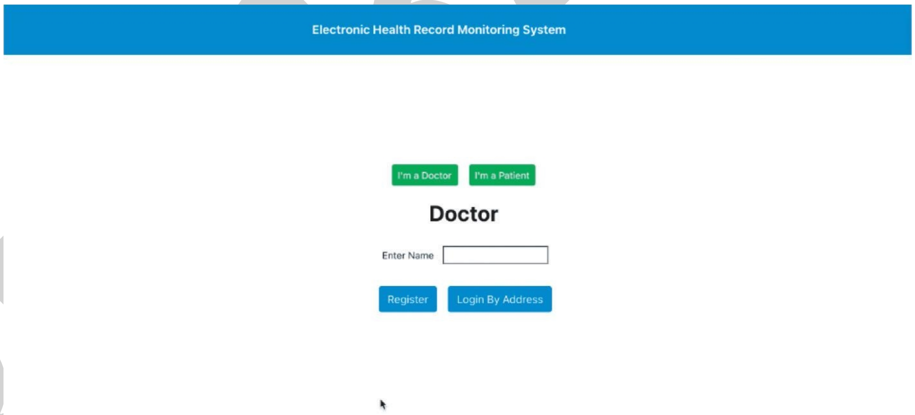


Figure 7: Login page.

blockchain is formed by each block referencing the previous block. A miner should verify whether a transaction fulfills the criteria to be processed before adding it to their block, depending on the blockchain's history. The transaction is genuine, according to blockchain history, and should be included in the block if the sender's wallet balance is sufficient. Privacy key by address is shown in Figure 9.

The data owner loses all rights to the deed if it is hacked or changed, as shown in Figure 9. It is a risky approach to possess something that may be deceptive at any moment. If only one line of data is deleted, the license for the data may be revoked. In conventional databases, client-server networks are utilized. A user (also known as a client) has the ability to make modifications to data stored on a central server. The database is still under the jurisdiction of a designated authority, which checks a client's credentials before allowing access. Because of database administration, if the authority's security is compromised, the data may be altered or even deleted. We can term blockchain technology an immutable ledger after using it here.

Figure 10 shows the immutable ledger with security. It is hard to alter the data if everyone retains all of the information in the block. Because all of the data in the block is linked to their prior hash number, if any of the data in the block is altered, the whole system is notified. It also has its own transaction mechanism that is very safe and reliable. These are

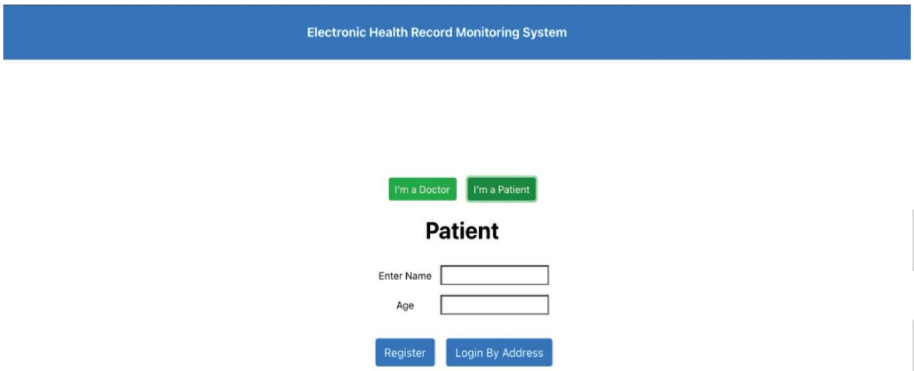
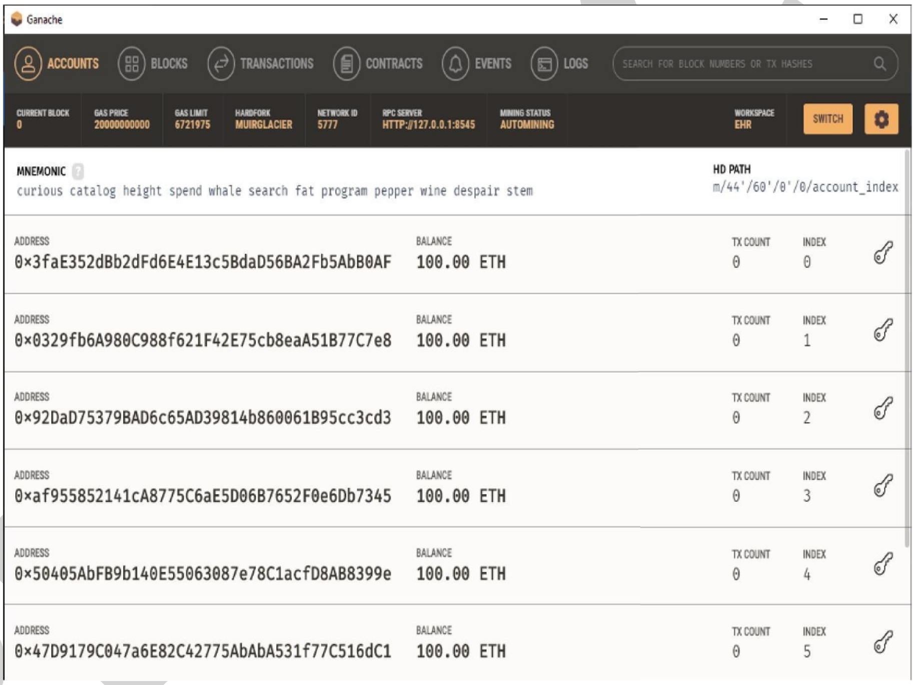


Figure 8: Sign-up page.



CURRENT BLOCK 0
GAS PRICE 20000000000
GAS LIMIT 6721975
HARDFORK MUIRGLACIER
NETWORK ID 5777
RPC SERVER HTTP://127.0.0.1:8545
MINING STATUS AUTOMINING
WORKSPACE EHR

MNEMONIC
curious catalog height spend whale search fat program pepper wine despair stem

HD PATH
m/44'/60'/0'/0/account_index

ADDRESS	BALANCE	TX COUNT	INDEX
0×3faE352dBb2dFd6E4E13c5BdaD56BA2Fb5AbB0AF	100.00 ETH	0	0
0×0329fb6A980C988f621F42E75cb8eaA51B77C7e8	100.00 ETH	0	1
0×92DaD75379BAD6c65AD39814b860061B95cc3cd3	100.00 ETH	0	2
0×af955852141cA8775C6aE5D06B7652F0e6Db7345	100.00 ETH	0	3
0×50405AbFB9b140E55063087e78C1acfD8AB8399e	100.00 ETH	0	4
0×47D9179C047a6E82C42775AbAbA531f77C516dC1	100.00 ETH	0	5

Figure 9: Private key by address.

the characteristics that distinguish any conventional ledger as immutable. A key characteristic of blockchain technology that distinguishes it from traditional database technology is public verifiability, which is achieved through integrity and transparency.

2.4. Chain and Data Component. This section explains the core notion of smart contracts by discussing the nature and types of contracts. We begin by defining the basic features of contracts and their many functions throughout the partnership's lifecycle from a legal and economic standpoint. Then, after looking at a few other definitions, we provide a general description of smart contracts. Last but not least, the importance of distributed ledger technology is discussed in the last part of this section. It displays all of the functions that were imported. When a block is produced and mined, the datetime function is used to give it its own timestamp. Because the function in disguise of a hash will be used, the function may need to hash the blocks in order for it to be effective.

Figure 11 shows us the JSON function. Before we hash the blocks, we will encode them. A class will be required. Because this will be performed through the use of a web application that will grab the message and utilize a postman to interact with the blockchain, a genesis block, a chain function, and a block function are all included in the blockchain class, and they are all responsible for adding and

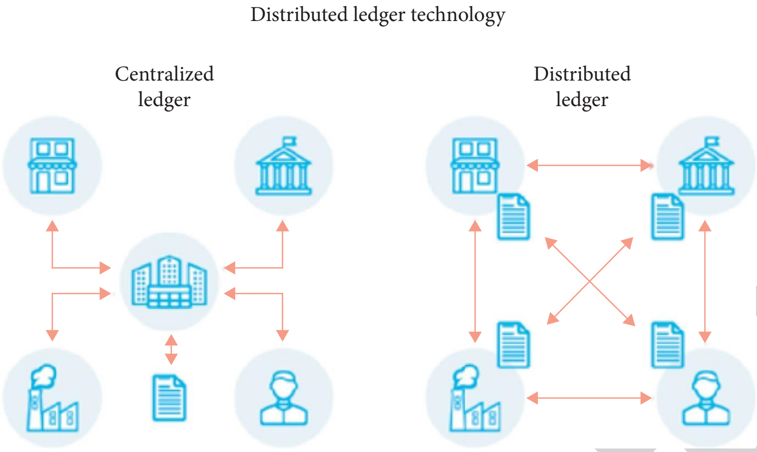
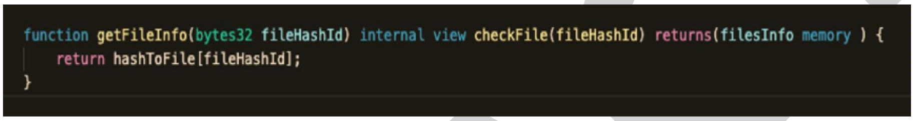


Figure 10: Immutable ledger with security.



```
function getFileInfo(bytes32 fileHashId) internal view checkFile(fileHashId) returns(filesInfo memory ) {
    return hashToFile[fileHashId];
}
```

Figure 11: Function for blockchain.

mining new blocks. The generate block function takes two arguments: a proof and the previous block's hash number. The function proof of function takes two parameters. The first is self, which is used to access a class-generated instance object. The second is an instance. In addition, there is the prior proof, which provides a path for miners to follow in order to solve the issue.

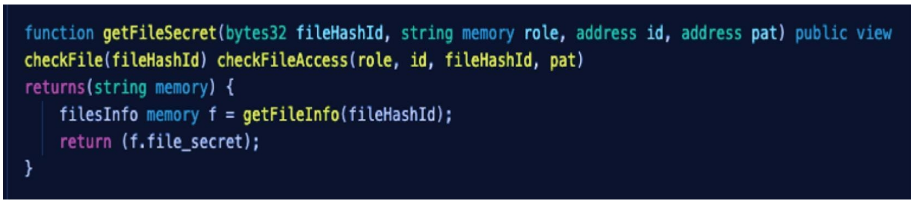
The function of file check is presented in Figure 12. Figure 12 illustrates that the new proof value is 1, indicating that after each repetition, the value of the proof must be increased by one until the correct evidence is obtained. Check proof will do the necessary checks to determine if the proof is valid or not. In addition to having four leading zeros, which makes mining for the hash operation more difficult for miners, the hash operation is made up of a string of 64 characters. The encode function will encode the string in the correct format, which is the format that the sha256 function expects to be sent. In this transaction, the transaction's function has been defined. With the help of the self, sender, and recipient keys in the argument, this add transaction method will carry out the procedure. This will complete the transaction before it is included in the block of transactions. The append method will be used to add a new transaction to the end of this collection of data. It is necessary to include the prior data for each new transaction, and this will be accomplished via the previous block function in each transaction. It will add +1 to the previous block function before returning to the previous block function. As a result, the number will automatically rise, the list will grow, and the information will be preserved.

2.5. *Transaction Component*

2.5.1. Encryption Methodology. In a blockchain, all the information provided by a doctor or patient is encrypted using cryptographic functions. In this research work, we have used the "keccak256" function built into Solidity. As shown in Figure 13, this function takes an input and converts it into a hash with a fixed length of 256, making the data provided by patients and doctors highly confidential.

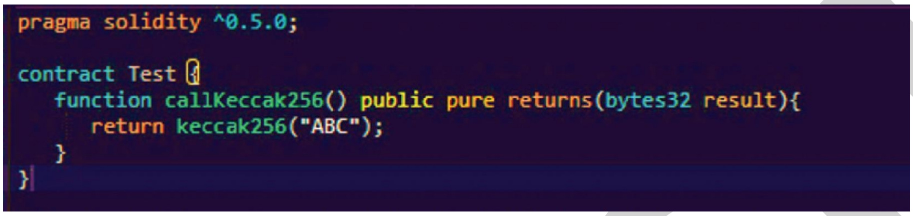
Figure 14 shows the output of the keccak256 function. What is important here is that whenever a slight change is made to the string, the hashed output changes drastically. A blockchain is formed only when each of the blocks is referenced by one another. Each block has the data and a hash pointing at the next block in the blockchain. This makes our system highly secure as even the slightest tampering with the input data will change the hashed output. Then, it will also change the hash of the previous block and the one before it, and subsequently, the whole blockchain will break. When this code is run using the keccak256 function, the string "ABC" gets converted into a hashed output.

2.5.2. Smart Contracts. Figure 15 shows the smart contract example and transactions. According to Figure 15, it is observed that smart contracts are deployed to facilitate, verify, or enforce negotiation of digital transactions. It separates the negotiating parties from any third party. Smart contracts are basically pieces of code that are programmed to verify certain conditions. If those conditions are not met,



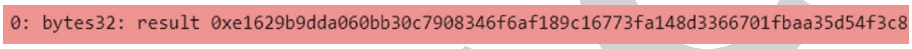
```
function getFileSecret(bytes32 fileHashId, string memory role, address id, address pat) public view
checkFile(fileHashId) checkFileAccess(role, id, fileHashId, pat)
returns(string memory) {
    filesInfo memory f = getFileInfo(fileHashId);
    return (f.file_secret);
}
```

Figure 12: Function of file check.



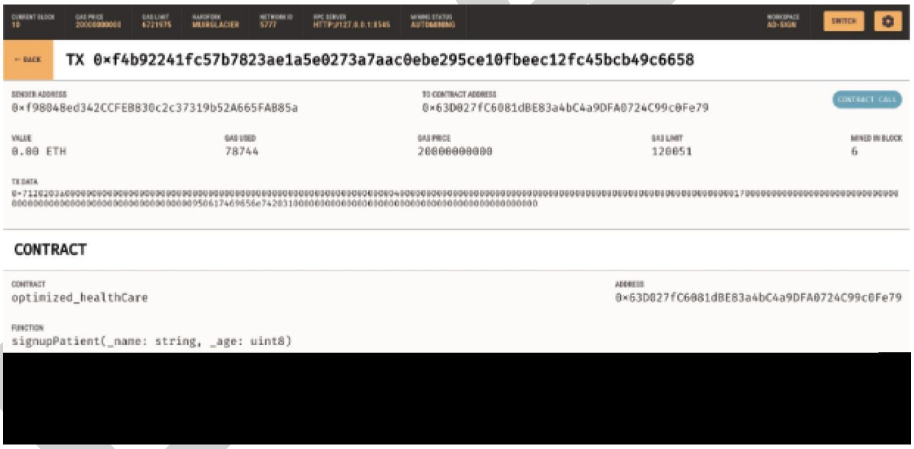
```
pragma solidity ^0.5.0;

contract Test {
    function callKeccak256() public pure returns(bytes32 result){
        return keccak256("ABC");
    }
}
```

Figure 13: The keccak256 function converting a string.



```
0: bytes32: result 0xe1629b9dda060bb30c7908346f6af189c16773fa148d3366701fbaa35d54f3c8
```

Figure 14: The output of the keccak256 function.



CURRENT BLOCK 10 | GAS PRICE 20000000000 | GAS LIMIT 6721975 | HARDFORK MUIRGLACIER | NETWORK ID 5777 | RPC SERVER HTTP://127.0.0.1:8545 | MINING STATUS AUTOMINING | WORKSPACE AD-SIGN

TX 0×f4b92241fc57b7823ae1a5e0273a7aac0ebe295ce10fbeec12fc45bcb49c6658

SENDER ADDRESS
0×f98048ed342CCFE8830c2c37319b52A665FA885a

TO CONTRACT ADDRESS
0×63D027fC6081dBE83a4bC4a9DFA0724C99c0Fe79

VALUE
0.00 ETH

GAS USED
78744

GAS PRICE
20000000000

GAS LIMIT
120051

MINED IN BLOCK
6

TX DATA
[illegible]

CONTRACT

CONTRACT
optimized_healthCare

ADDRESS
0×63D827fC6081dBE83a4bC4a9DFA0724C99c0Fe79

FUNCTION
signupPatient(_name: string, _age: uint8)

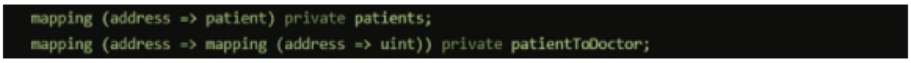
[redacted]

Figure 15: Smart contract example and transactions.

then the transaction will not happen. We have programmed many smart contracts throughout the system and those smart contracts only execute when the conditions are met.

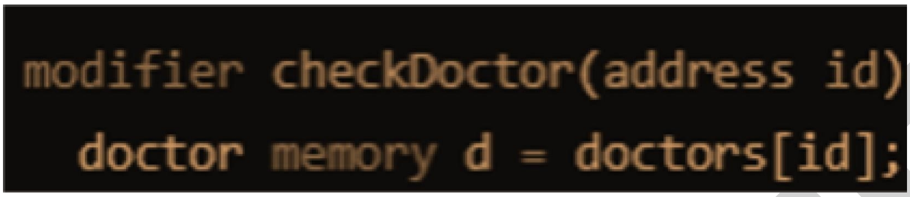
2.5.3. Mapping and Modifiers. A mapping function example is shown in Figure 16. Mapping is a function in solidity that exists in a key-variable relationship. It takes a key and returns a variable. An access modifier can be set on the returning variable. The address is the unique public key half of the Ethereum address of the interacting party. In the first mapping of Figure 16, the mapping function takes the address of the interacting party and then returns the list of doctors and list of patient profiles he can access. Since the access modifier is private, nobody outside the contract can view this information. The modifiers are used to check if that doctor exists or not, which is shown in Figure 16.

Figure 17 shows a modifier example. Figure 17 demonstrates that a large number of very strong security measures have been applied, resulting in a system that is exceptionally secure and dependable. As a consequence of their failure to address these concerns, other publishers' systems have become insecure and susceptible to hacker attacks. It satisfies the standards since it features an immutable ledger, smart contracts, blockchain-compatible transactions, and straightforward refund and return mechanisms. Almost every argument offered in earlier articles contains a fault. The inability of some of the website's administrators to correctly execute the smart contract was the primary cause of the website's collapse.



```
  mapping (address => patient) private patients;
  mapping (address => mapping (address => uint)) private patientToDoctor;
```

Figure 16: Mapping function example.



```
modifier checkDoctor(address id)
  doctor memory d = doctors[id];
```

Figure 17: Modifier example.

3. Result and Comparison Analysis

We have developed an EHR system for Bangladesh. Actually, in South Asia, EHR is much less than in North America. The main reason is the Blockchain ecosystem and structure. The study summarizes initiatives throughout Asia to deploy EHR systems for a variety of purposes. We highlight 32 pieces of research performed in 15 countries, including two that compare locations throughout Asia. This study collects data on EHR systems in a variety of countries and healthcare situations, including LMIC settings, diverse organizational structures, and various levels within health systems. It reflects a range of technical infrastructure and EHR system "maturity" levels, as well as the resulting human resource requirements.

This research analyzes the obstacles to the use of electronic health record systems in developing healthcare in subcontinental countries. Restrictions on the framework needed for EHR systems (e.g., stable electricity, wireless technology, and other mobile technologies) add another degree of complexity to system requirements and the level of EHR sophistication that may be supplied. As a result, there may be risks associated with using EHR for public health purposes in a particular context. Several of the most important hurdles are connected to organizational culture, and they underscore the crucial necessity for well-trained technical assistance in Asian hospital environments. Hospitals often discover that delays in EHR deployment arise as a result of physician and health professional nonadoption of the system. According to research conducted in Iran, organizational obstacles to EHR adoption include a lack of effective planning, a shortage of qualified personnel, and restrictions on healthcare workers' access to information technology training. In light of these concerns, potential solutions include conducting a priori assessments of organizational cultures and settings where EHR systems will be implemented in order to determine the level of technical support required; examining staff understanding, experience levels, and readiness for new software; and reviewing current data harvesting systems in order to minimize early deployment bottlenecks. In addition, before a system is implemented, it is important to address staff concerns about new information and communications technology interventions. This will help to prevent reluctance to adopt new practices and alleviate concerns about the administration of the burden associated with the new system. Implementation within a specific health system or organization may benefit from these investigations because they allow for a more customized approach to EHR interventions that are contextualized in light of unique externalities that may present obstacles but cannot be addressed at the implementation level. In addition to technological and practical challenges, research has shown that the adoption of EHR therapies is fraught with ethical concerns. As electronic health records (EHRs) become more common in low- and middle-income countries (LMICs), continuing concerns in HIC about patient confidentiality, privacy, informed consent, and data security continue to be important in resource-constrained settings. Aside from highlighting the reality that worldwide EHR systems are increasing at a rapid speed, it also highlights the problem that LMIC settings may be underprepared to deal with the difficulties connected with EHR adoption. Many smaller healthcare providers and independent hospitals are still looking for successful EHR systems, or are transitioning from fragmented applications provided by a variety of suppliers to a single, functional system. It is necessary to carefully consider patient-provider interactions in low- and middle-income countries (LMICs). These interactions must take into account cultural sensitivity, ethnic health inequalities, low levels of patient literacy, language difficulties, and the need for institutional supervision of the patient-provider connection. To ensure that efficient and flexible EHR systems are deployed to meet public health needs in the future, more systematic and comprehensive preparations should be undertaken. In the same way that technical and practical barriers must be overcome when implementing EHR interventions, ethical issues must be addressed in order for effective EHR initiatives to be implemented successfully. It is anticipated that by developing a framework for EHR implementation and providing formal instruction to healthcare professionals and support personnel on ethical issues, healthcare providers and support staff would be able to minimize patient risks. Because we did not have access to any computer calculations, we

Table 1: Comparison with other papers.

Point of this paper	Point of other papers
(1) Medical data is encrypted from end to end by the user. No third party can see the details.	(1) The data received from the end user is first encrypted so that it is only visible to them [1].
(2) We used cryptocurrency transactions. It will be revolutionary.	(2) No cryptocurrency transactions [21].
(3) We have used the "keccak256" function built into Solidity. This function takes an input and converts it into a hash with a fixed length of 256, making the data provided by patients and doctors highly confidential.	(3) Authors discussed about many variations of EHR's security [9].
(4) Patients and doctors will both benefit. Every service under our system will be faster than under the current system.	(4) Some EHR systems need more time to update the information. Sometimes, the system needs more electricity to run the server. It is expensive [20].
(5) We have used secured databases to monitor every footprint on the web. It will be stored on the server in a hash. So, no one can temper the medical data easily. It Is under smart contracts.	(5) The system is controlled by the admin. Sometimes, records cannot be undone. So, its footprint is so strong that it finds the bad side of the system [7].

Table 2: EHR lifecycle factors.

EHR lifecycle factor	Usability and safety optimization opportunities	
	For EHR developers	For healthcare providers
Safety	• Incorporate a safety practice into the organization's policies so that all team members understand its value. • Create a risk-free atmosphere in which possible risks may be reported. • To classify and act on discovered issues, use a specialized professional with expertise in patient safety and risk management. • Consider security measures in current and prospective software versions. • Allow workers and customers to report safety risks and communicate promptly about the stated danger.	• Set up a safe space where employees may come forward with concerns about potential dangers. • Prioritize safety by implementing measures to make sure that everybody on the team understands how important it is. • Authorize the use of automatic surveillance to look for potential risks in case of setup errors. • Use teams with embedded health IT knowledge, especially in big companies with internal specialists in health information technology (health IT).
Product R&D	• Develop and improve infrastructure by identifying healthcare provider requirements. • Find out what kind of training you need. • Prior to the introduction of a product, carry out testing that focuses on high-risk functionalities and involves targeted users who are representative of the target market and rigorous test scenarios.	• Provide developer employees with chances to study healthcare staff processes and technology usage. • Design and testing for usability and safety should be shaped by physicians and experts in the field.
Acquisition	• Provide healthcare providers with upfront information about the HER product's features and anticipated prices so that they may examine the product's viability, such as through comprehensive documentation.	• In order to identify prospective suppliers, evaluate clinician requirements for the EHR product as well as any budgetary restrictions.
Customization and configuration	• Create a list of known high-risk modifications that defy developer advice and explain how and why they may affect patients. • Clarify with providers the relevant definitions, resources, and responsibilities (vendor, provider, and third-party products) • Analyze the resources and expertise of the supplier to make adjustments to the product and interact with them.	• Evaluate internal capabilities and expertise to modify the product based on talks with the vendor. • Create a compelling argument for why modifications are necessary, including use cases. • Document customizations that are made and develop a risk mitigation plan.

Table 2: Continued.

EHR lifecycle factor	Usability and safety optimization opportunities	
	For EHR developers	For healthcare providers
Implementation and upgrades	• Create an implementation strategy based on your knowledge and share it with the teams in charge of implementing it at the healthcare providers.	• Implementation procedures must be supported by a sound governance framework that addresses any safety concerns that may emerge.
	• Help implementation teams at healthcare facilities, such as doctors and nurses, and understand the fundamental functions and processes.	• Set up implementation teams of experts from various fields who can come to an agreement on the functionality and process.
	• Ensure that a mechanism is in place to monitor safety by encouraging facilities to review their risk management procedures, resources, and requirements.	• Select a representative group of users to evaluate the new product.
	• A well-qualified and well-supported IT staff is essential for providing on-going assistance.	• As soon as new versions of your operating system are available, begin implementing them.
		• Identify areas that need further construction, configuration, and/or clinical risk analysis and mitigation by reviewing changes included with updates.
Training	• Incorporate training situations that are both demanding and safety-focused.	• Keep track of the training expenses and the steps you took to get there.
	• Refresher training should be provided following any significant system modifications or upgrades.	• Personalized training that meets the specific requirements of the participants is essential. Workflow simulations, refresher training, ongoing access to training materials, and the chance for improved training for individuals in need are all ways to provide this kind of training.
	• Train your employees with the help of professionals who are trained or certified in the new technology you have bought.	• Learn to take into account the finest training methods suggested by vendors.

came up with a list of advantages and disadvantages of how the system would affect the lives of healthcare and information technology professionals.

The points of this paper are compared to the points of other studies in Table 1. The facts concerning EHR are presented in this article.

Table 2 shows the analytical points that if we implement EHR, the healthcare providers will benefit and how all the structures will be organized in detail.

4. Conclusion

The purpose of this article is to improve the intelligence and security of electronic health management. This architecture is unchangeable and provides complete transaction transparency. It guards against unwanted access and data tampering on our website. Smart contracts also reduce the amount of time spent on time-consuming documentation. In general, medial data administration entails a great deal of paperwork. Smart contracts are immutable because of the blockchain, which saves the information as evidence. This paradigm has the greatest impact on transactions, immutability, and refundable procedures in chain management. An end-to-end approach was suggested in this research. The function and role of each actor have been defined. It also implies that our framework may be used in a wide range of situations. In addition, the construction of smart contracts is explored. As a consequence of the results of this article, the problems that people have with outdated processes will be permanently removed. This study creates trust to develop and reduce transaction costs, as well as improve financial inclusion by providing additional options for those who do not have easy access to financial services, among its many advantages (the most significant of which is the ability to keep data secure). This is a small-scale, one-of-a-kind piece of work. The delay may be impacted if there is a high amount of data. In terms of storage and processing expenses, blockchain transactions will be very advantageous. Another approach to increase performance is to use decentralized databases. Furthermore, tracking devices may be added to the structure in the case of a large-scale deployment. As the amount of data grows, we may employ off-chain architecture to store the original data, with the proof of existence remaining on the blockchain. This may be a possible study subject for this inquiry in the future.

Data Availability

No data were utilized to support these research findings.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

The authors are thankful for the support from Taif University Researchers Supporting Project (TURSP-2020/115), Taif University, Taif, Saudi Arabia.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9834267, 1 page
https://doi.org/10.1155/2023/9834267

Retraction

Retracted: Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm

Security and Communication Networks

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Yu, "Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 3272603, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 3272603, 13 pages
https://doi.org/10.1155/2022/3272603

Research Article

Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm

Jiangbo Yu

Business School, Luoyang Normal University, Luoyang 471934, China

Correspondence should be addressed to Jiangbo Yu; yujiangbo@lynu.edu.cn

Received 14 October 2021; Revised 4 December 2021; Accepted 9 December 2021; Published 5 January 2022

Academic Editor: Chinmay Chakraborty

A business credit risk early warning algorithm based on big data analysis and discrete selection model is presented to address the issues of poor sample fitting performance, long warning time, and low warning accuracy that plague the traditional enterprise credit risk early warning algorithm. A-share listed enterprises in China were chosen as the credit data source for screening the samples based on big data analysis. After screening, financial failure firms were coupled, and paired samples were created. The credit risk variables, which included financial and corporate governance characteristics, were chosen based on the created samples. The enterprise financial risk submodel and the nonfinancial risk submodel were built based on the enterprise credit risk variables, and the financial and nonfinancial index scores of enterprise customers were evaluated separately to develop a discrete choice model of enterprise credit risk. The algorithm's sample fitting performance was employed to achieve early warning of corporate credit risk. The algorithm based on big data analytics and discrete choice model is compared to the traditional method in order to verify its validity. The findings of the experiment reveal that the algorithm's sample fitting performance is superior to the traditional one, making it more suitable for enterprise credit risk early warning. The proposed model depicts 85% accuracy.

1. Introduction

Credit growth is to promote investment, production, and consumption. Traditional financing products such as medium- and long-term loans (mostly for infrastructure projects) and commercial-paper financing are still the major components of new loans for various enterprises at the moment. Many company construction projects have relatively easy access to a substantial number of bank credit money due to the trust in government credit. However, due to the unpredictability of many enterprise infrastructure projects' profitability or a lack of government experience, there are numerous hazards [1], including project rushes, project returns, and cost uncertainty. Because credit cards are revolving credit lines, lenders and investors have more possibilities for actively monitoring and managing them than other types of retail loans, such as mortgages. As a result, maintaining credit card portfolios might be a significant source of revenue for financial institutions. Better risk management might result in annual savings of hundreds of millions of dollars for financial organizations. For financial institutions, better risk management could lead to financial revenue of hundreds of thousands of dollars. Lenders, for example, could reduce their exposure by cutting or freezing credit lines on accounts that are likely to default. Early credit risk prediction is necessary because effective application of the above risk management measures necessitates banks' ability to identify accounts that are likely to default.

These risks could cause issues during the implementation phase, necessitating a comprehensive risk assessment and a project cost budget. At the same time, due to the nature of public interests, many projects are not heavily commercialized, putting their future earnings in jeopardy. Furthermore, large-scale projects are typically expensive, take a long time to build, and have a significant payback period, during the operation phase; it is difficult to ensure debt-paying ability. As a result, company credit risk is becoming more complex and expansive, posing a slew of new and challenging criteria for enterprise credit risk management. The risk warning system can help businesses improve their resilience, adaptability, and competitiveness, as well as preventing crises from forming or germinating in the first place.

This paper expanded the current research status and relevance of financial risk early warning, as well as the development background, current position, and future difficulties, based on reviewing and assessing prior research works. This paper proposes a novel technique that can predict the potential risk related to finances and protect the organization from potential credit risks. The detailed sections are arranged as follows: Section 2 introduces the related works; Section 3 discusses the sample screening structure and variable selection process; Section 4 discusses the experimental results of big data analytics and discrete choice model for enterprise credit risk early warning algorithm; Section 5 is the conclusion.

2. Related Works

According to relevant experts, financial intermediary services are reasonably mature, and the driving factor behind their promotion is to increase value. Bayesian model to investigate the early income mechanism in order to help organizations predict and reduce the risk of loan transactions in [2] is used. Although the system may detect early signs of business credit risk, it suffers from a lack of precision. In [3] the authors proposed an enterprise credit risk assessment method based on improved genetic algorithm, which satisfies the adaptability of improved algorithm to corporate credit risk. The algorithm can accurately forewarn the credit risk of enterprises, but it takes a long time.

Increased value can be obtained by expanding the creation of value-added services and effectively lowering various expenses, i.e., increasing income while lowering expenditure [4]. Collaboration between the supply chain and financial service providers will be varied, boost the value of both parties, and advance the financial industry to a higher level [4]. The linear regression method is used by related researchers in loan risk evaluation research. The linear regression model's underlying conceit is to create a regression equation based on individual variables in order to calculate the likelihood that consumer credit performance is "excellent" [5] but many flaws in parametric statistical methods have been identified by researchers.

Through balance sheets, credit business, and other channels, commercial banks and other financial institutions form a complicated network interaction. Risk identification, risk assessment, risk early warning, and risk treatment are the key components of financial risk early warning job, which may be further split into financial risk organization form, indicator system, and prediction method. When banking is subjected to internal and external shocks that cause debt failures, volatility risks are transferred through interbank credit channels, and risk spillovers affect the activities of other banks and financial institutions, posing systemic hazards [6].

Researchers in related fields have also studied enterprise credit, starting from the perspective of organizational security and forecasting techniques for risk management. In [4] the BP neural network model to evaluate the project risk is used. The model can effectively avoid overtraining and overfitting and has good generalization ability. Compared with the fuzzy theory, the influence of human factors is avoided. In [5] the authors discussed the comprehensive gray correlation of computational languages and their numbers and the enterprise credit as examples to analyze and compare and finally verified the practicability, rationality, and effectiveness of this method. But the above algorithm has the problem of poor sample fitting performance.

Considering the sample fitting performance of enterprise credit risk, this paper proposes an enterprise credit risk early warning algorithm based on big data analytics and discrete choice model.

3. Basic Definitions

3.1. Sample Screening Structure and Variable Selection

3.1.1. Sample Screening. Based on big data analytics, A-share listed companies in China were selected as sample sources, and nonfinancial listed companies that were specially treated (ST) for the first time for financial reasons were used as samples of financial failure companies. On this basis, small- and medium-sized listed companies and large listed companies are screened by industry [7, 8]. The industry division standards of listed companies are mainly according to the "Guidelines for the Classification of Listed Companies" issued by the Securities Futures Commission. Large, medium, and small listed enterprises are screened according to the "Interim Provisions on Standards of Small and Medium-Sized Enterprise" jointly issued by the State Economic and Trade Commission and another three ministries, and the "Supplementary Standards for the Division of Large, Medium, and Small Nonindustrial Enterprises" issued by the State-Owned Assets Supervision and Administration Commission. The "Interim Provisions on Standards of Small and Medium-Sized Enterprise" are shown in Table 1.

The "Supplementary Standards for the Division of Large, Medium, and Small Nonindustrial Enterprises" are shown in Table 2 [8].

The industry classifications in the "Guidelines for the Classification of Listed Companies" are not exactly the same as those in the "Interim Provisions on Standards of Small and Medium-Sized Enterprise" and the "Supplementary Standards for the Division of Large, Medium, and Small Nonindustrial Enterprises." Inconsistencies in the sector are quite easy to create. Some large-scale industries are more precisely matched according to their secondary industries to eliminate these industry matching errors [9, 10]. The sample was further screened as follows:

(i) Companies with significant asset reorganization during the period from (t-5) to (t-2) were excluded from the ST companies, and companies with missing data were excluded as well.

(ii) A total of 132 samples from ST companies were left, including 80 small and medium-sized listed enterprises and 52 large listed enterprises.

Table 1: Interim Provision on Standards of Small and Medium-Sized Enterprise.

Industry	Number of people engaged	Sales volume	Total assets	Explain
Industry	300–2000	30 million–3 billion	40 million–4 billion	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Construction business	600–300	30 million–3 billion	40 million–4 billion	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Wholesale business	100–200	30 million–3 billion	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Retail	100–500	10–150 million	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Communications and transportation industry	500–3000	30 million–3 billion	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Postal industry	400–1000	30 million–3 billion	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Accommodation and catering	400–800	30–150 million	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.

Table 2: Supplementary Standards for the Division of Large, Medium, and Small Nonindustrial Enterprises.

Industry	Index name	Company	Large	Medium-sized	Small-scale
Agriculture, forestry, animal husbandry, and fishery	Number of people employed sales volume	People ten thousand RMB	≥3000 ≥15000	500–3000 1000–15000	<500 <1000
Warehousing industry	Number of people employed sales volume	People ten thousand RMB	≥500 ≥15000	100–500 1000–15000	<100 <1000
Estate	Number of people employed sales volume	People ten thousand RMB	≥200 ≥15000	100–200 1000–15000	<100 <1000
Finance	Number of people employed sales volume	People ten thousand RMB	≥500 ≥50000	100–500 5000–50000	<100 <5000
Geological exploration and water conservancy environmental management	Number of people employed sales volume	People ten thousand RMB	≥2000 ≥20000	600–2000 2000–20000	<600 <2000
Sports and entertainment industry	Number of people employed sales volume	People ten thousand RMB	≥600 ≥15000	200–600 3000–15000	<200 <3000
Information transmission industry	Number of people employed sales volume	People ten thousand RMB	≥400 ≥30000	100–400 3000–30000	<100 <3000
Computer service and software industry	Number of people employed sales volume	People ten thousand RMB	≥300 ≥30000	100–300 3000–30000	<100 <3000
Leasing industry	Number of people employed sales volume	People ten thousand RMB	≥300 ≥15000	100–300 1000–15000	<100 <1000
Business and technology services	Number of people employed sales volume	People ten thousand RMB	≥400 ≥15000	100–400 1000–15000	<100 <1000
Resident service industry	Number of people employed sales volume	People ten thousand RMB	≥800 ≥15000	200–800 1000–15000	<200 <1000
Other enterprises	Number of people employed sales volume	People ten thousand RMB	≥500 ≥15000	100–500 1000–15000	<100 <1000

Table 3 gives the statistics of the ST samples.

For the purpose of determining the starting point of financial failure, the first two years of ST are used. That is, if the year of ST is *t* year, then (*t*-2) year is the starting point of financial failure [11, 12]. The warning capability of the loss in the first three years was the subject of inquiry, that is, the warning capability in the year of (*t*-5), (*t*-4), and (*t*-3) regarding whether there will be a financial failure in the (*t*-2) year, taking into consideration the timeliness of financial failure warnings. The first two years of ST are the starting point of financial failure when choosing a financial failure point.

Table 3: Statistics of ST samples.

CSRC classification	2015	2016	2017	2018	2019	Total	Proportion
Agriculture, forestry, animal husbandry, and fishery	0	1 (0/1)	1 (0/1)	2 (0/2)	0	4 (0/4)	3.03
Extractive industry	0	3 (0/3)	2 (0/2)	0	0	5 (0/5)	3.79
Manufacturing industry	11 (4/7)	20 (7/13)	25 (9/16)	13 (6/7)	10 (3/7)	79 (29/50)	59.9
Communications and transportation industry	0	0	1 (0/1)	0	0	1 (0/1)	0.76
Information technology industry	0	4 (0/4)	3 (1/2)	0	1 (1/0)	8 (2/6)	6.06
Wholesale and retail	0	3 (0/3)	2 (2/0)	1 (1/0)	0	6 (3/3)	4.55
Estate	0	11 (8/3)	3 (2/1)	2 (2/0)	3 (2/1)	19 (14/5)	14.4
Social service industry	1 (1/0)	2 (1/1)	1 (1/0)	0	0	4 (3/1)	3.03
Cultural communication industry	0	1 (0/1)	0	0	0	1 (0/1)	0.76
Production and supply of electricity, gas, and water	1 (0/1)	0	2 (1/1)	1 (0/1)	1 (0/1)	5 (1/4)	3.79
Total	13 (5/8)	45 (16/29)	40 (16/24)	19 (9/10)	15 (6/9)	132 (52/80)	100

Note. The figures in brackets are the number of St large companies and St small and medium-sized companies, respectively.

3.1.2. Sample Construction. After the sample screening, the financially failed companies were paired to construct paired samples. The non-ST samples paired with the ST sample are derived from the companies that never have an ST in the sample period and have excluded the IPO (Initial Public Offerings). Because the existing standards have a multidimensional definition of the enterprise scale, it is easy to have mismatch asset scale, which will cause the inhomogeneity in paired samples [13]. In order to reduce this mismatch, from the three-dimensional characteristics: asset size, sales revenue, and number of employees, 80 small and medium-sized enterprises and 52 large enterprises were paired in a ratio of 1 : 2 in the same year and in the same industry. The time of pairing is the first two years of the financial failing firm before ST, in the (t-2) year. After pairing, a total of 396 samples were obtained, including 240 small and medium-sized enterprises (SMEs) and 156 large enterprises (LEs) [14]. Then, 180 samples were randomly selected from the SMEs group as the estimated samples to build the model, and the remaining 60 were used as verification samples to test the warning effect of the model.

3.1.3. Variable Selection. Based on the constructed samples, corporate credit risk variables were selected, including financial variables and corporate governance variables [15, 16]. 28 financial ratios were selected from seven categories of financial indicators, which reflect the financial leverage structure, solvency, profitability, operating capacity, growth potential, investment income level, and cash flow status of the listed company, as shown in Table 4.

In order to eliminate industry differences in financial ratios, the industry median adjustments were made on an annual basis for all 28 financial ratios, as follows:

$$R_{it} = \frac{X_{it}}{X_{igt}}, \quad (1)$$

where X_{igt} represents the median of financial ratio i in the industry g in the t year, X_{it} represents the median of the industry, and R_{it} represents the annual adjustment value of financial ratio. All financial ratio data comes from the Wind database [17].

The corporate governance variables investigate the impact of corporate governance characteristics on financial failures mainly from the ownership structure, actual controller type, board structure, and executive incentives as shown in Figure 1.

Table 5 provides a full overview of the above corporate governance characteristics. The data for all corporate governance variables is derived from the "Listed Corporate Governance Structure Database" in the CCER China Economic and Financial Database [18].

3.2. Construction of Warning Model of Enterprise Credit Risk

3.2.1. Construction of Enterprise Financial Risk Submodel. The enterprise financial risk submodel was built based on the enterprise credit risk variables to provide a high-precision data source for the enterprise credit risk warning model. To eliminate the correlation between financial indicators, the 21 secondary financial indicators were first subjected to principal component analysis (PCA). Then the extracted principal components were used to construct a submodel of corporate financial risk [19].

The specific operation of principal component analysis is as follows:

(1) A data file has been created. Numerical variables X_1, X_2, X_3, . . ., X_{21} were defined. The collected 21 scalar values were then standardized.

(2) Principal component analysis was performed on the standardized data. The results are shown in Table 6. The cumulative variance contribution rate of the seven principal components has reached 84.62%, according to the cumulative variance contribution rate table and the principal component analysis matrix. From the 8th one, Eigen values are all less than 1. It was observed that 7 principal components effectively reflect 84.62% of the information in the indicator system, substantially simplifying the research process compared to 21 indicators components.

(3) The principal component score was calculated. According to the cumulative variance contribution

Table 4: Selection of financial variables.

Financial variables and symbols	Financial variables and symbols	Financial variables and symbols
Asset liability ratio (X_1)	Operating profit/total operating revenue (X_{11})	Earnings per share (X_{21})
Equity ratio (X_2)	Return on equity (X_{12})	Net assets per share (X_{22})
Current assets ratio (X_3)	Total assets ratio of retained earnings (X_{13})	Total cash to assets ratio (X_{23})
Current liability ratio (X_4)	Total assets turnover rate (X_{14})	Net cash flow growth rate (X_{24})
Working capital/total assets (X_5)	Turnover rate of accounts receivable (X_{15})	Accounts payable/total assets (X_{25})
Current ratio (X_6)	Accounts payable turnover (X_{16})	Accounts payable/current assets (X_{26})
Quick ratio (X_7)	Inventory turnover (X_{17})	Current market value debt ratio (X_{27})
Cash flow liability ratio (X_8)	Fixed assets turnover rate (X_{18})	P/E ratio (X_{28})
Interest cover (X_9)	Growth rate of total assets (X_{19})	
Return on total assets (X_{10})	Operating revenue growth rate (X_{20})	

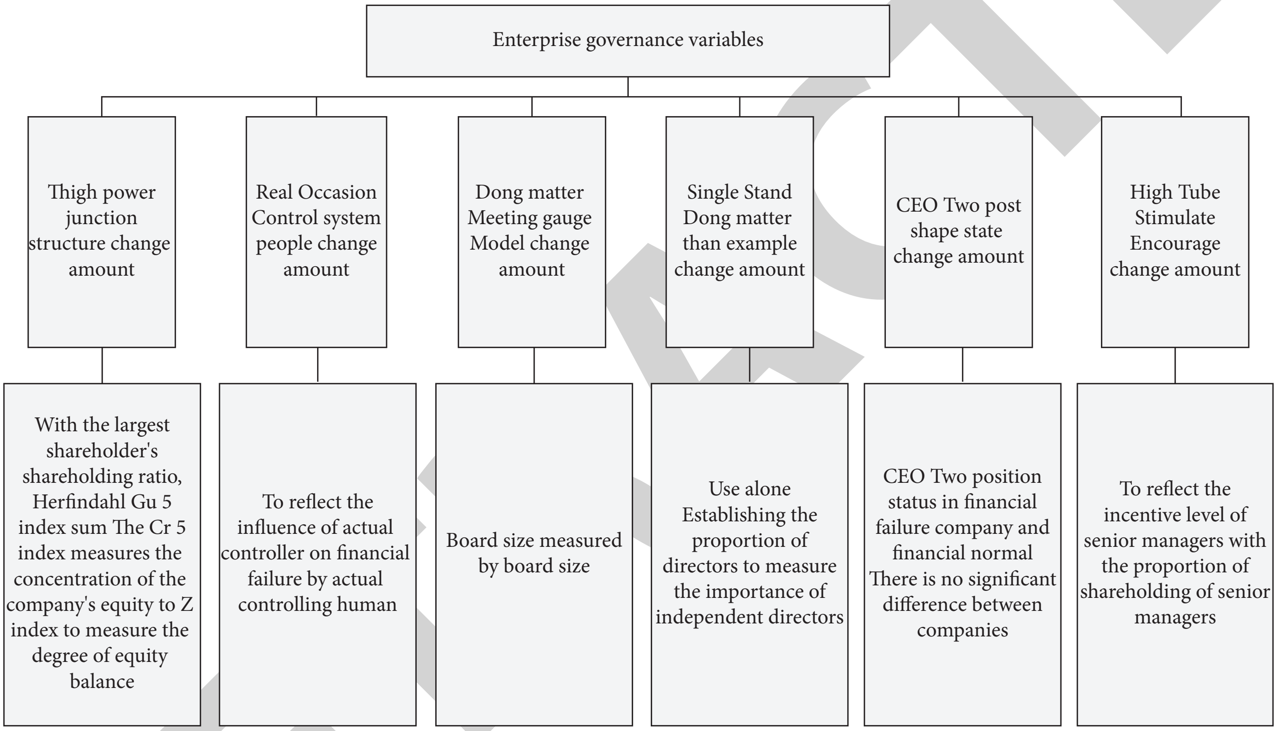

Enterprise governance variables
Thigh power junction structure change amount
Real Occasion Control system people change amount
Dong matter Meeting gauge Model change amount
Single Stand Dong matter than example change amount
CEO Two post shape state change amount
High Tube Stimulate Encourage change amount
With the largest shareholder's shareholding ratio, Herfindahl Gu 5 index sum The Cr 5 index measures the concentration of the company's equity to Z index to measure the degree of equity balance
To reflect the influence of actual controller on financial failure by actual controlling human
Board size measured by board size
Use alone Establishing the proportion of directors to measure the importance of independent directors
CEO Two position status in financial failure company and financial normal There is no significant difference between companies
To reflect the incentive level of senior managers with the proportion of shareholding of senior managers

Figure 1: Composition of enterprise governance variables.

Table 5: Enterprise governance variables and their meanings.

Variable type	Variable name and symbol	Specific meaning
Ownership structure	Shareholding ratio of the largest shareholder (G_1)	Number of shares/total share capital held by the largest shareholder
	CR_5 index (G_2)	The sum of the shareholding ratios of the top five major shareholders
	Herfindahl_5 index (G_3)	Sum of the squares of the shareholding ratio of the top five major shareholders
	Z index (G_4)	Number of shares held by the first largest shareholder/number of shares held by the second largest shareholder
Type of actual controller	Actual controller (G_5)	If the last controlling shareholder belongs to state-owned holding, take 1; otherwise take 0
Board structure	Board size (G_6)	Number of board members (natural logarithm)
	Proportion of independent directors (G_7)	Total number of independent directors/total number of directors
Executive incentive	CEO status (G_8)	The chairman and general manager shall be one person, otherwise 0
	Shareholding ratio of senior executives (G_9)	Total shares/total share capital held by senior executives at the end of the year

Table 6: Analysis results.

Component	Initial Eigen values			Extraction sums of squared loadings		
	Total	Variance contribution rate (%)	Cumulative (%)	Total	Variance contribution rate (%)	Cumulative (%)
1	5.298	26.228	26.228	5.298	25.228	25.228
2	3.599	17.136	42.364	3.599	17.136	42.364
3	2.681	12.768	55.133	2.681	12.768	55.133
4	2.048	9.752	64.885	2.048	9.752	64.885
5	1.597	7.606	72.491	1.597	7.606	72.491
6	1.501	7.147	79.638	1.501	7.147	79.638
7	1.046	4.982	84.620	1.046	4.982	84.620
8	0.670	3.193	87.812			
9	0.549	2.613	90.426			
10	0.524	2.495	92.921			
11	0.437	2.081	95.003			
12	0.390	1.855	96.858			
13	0.261	1.244	98.101			
14	0.166	0.790	98.891			
15	0.109	0.519	99.410			
16	0.063	0.302	99.712			
17	0.37	0.178	99.890			
18	0.13	0.064	99.954			
19	0.007	0.034	99.987			
20	0.003	0.013	100.000			
21	$1.29E-016$	$6.14E-016$	100.000			

rate table, the Eigen values of the coefficient matrix of the seven principal components are

$$\begin{cases} \lambda_1 = 5.298, \\ \lambda_2 = 3.599, \\ \lambda_3 = 2.681, \\ \lambda_4 = 2.048, \\ \lambda_5 = 1.597, \\ \lambda_6 = 1.501, \\ \lambda_7 = 1.046. \end{cases} \tag{2}$$

For values in each column in the principal component analysis matrix, they are divided by $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$, $\sqrt{\lambda_3}$, $\sqrt{\lambda_4}$, $\sqrt{\lambda_5}$, $\sqrt{\lambda_6}$, and the unit eigenvector corresponding to each Eigen value was obtained. Thus, seven principal components ($F_1, F_2, F_3, F_4, F_5, F_6$, and F_7) can be expressed [20–22], and the standardized variables can be used to determine each principal component's final score.

Then, using a BP neural network, a submodel of corporate financial risk was created. Figure 2 depicts the individual steps.

3.2.2. Construction of Nonfinancial Risk Submodel. A nonfinancial submodel was constructed based on corporate credit risk variables. First, the target level O: nonfinancial risk score was set. Then, the criterion level C: there are enterprise scale score, shareholding structure, and audit opinion. Last, measure level P: there are total enterprise assets, annual income, the proportion of the largest shareholder, the shareholding ratio of the two major shareholders, and the type of audit opinion. Because the target level and the measure level are linked by the relationship between the primary and secondary indicators, considering that the total assets and the annual income of the enterprise will also affect the audit opinion, the logical relationships between different levels are shown in Figure 3.

The importance level between each indication can be assessed using the hierarchical structure model of nonfinancial indicators to produce the judgment matrix of each level. The particular results are presented in Figure 4 after clicking "calculating result" in the yaahp software.

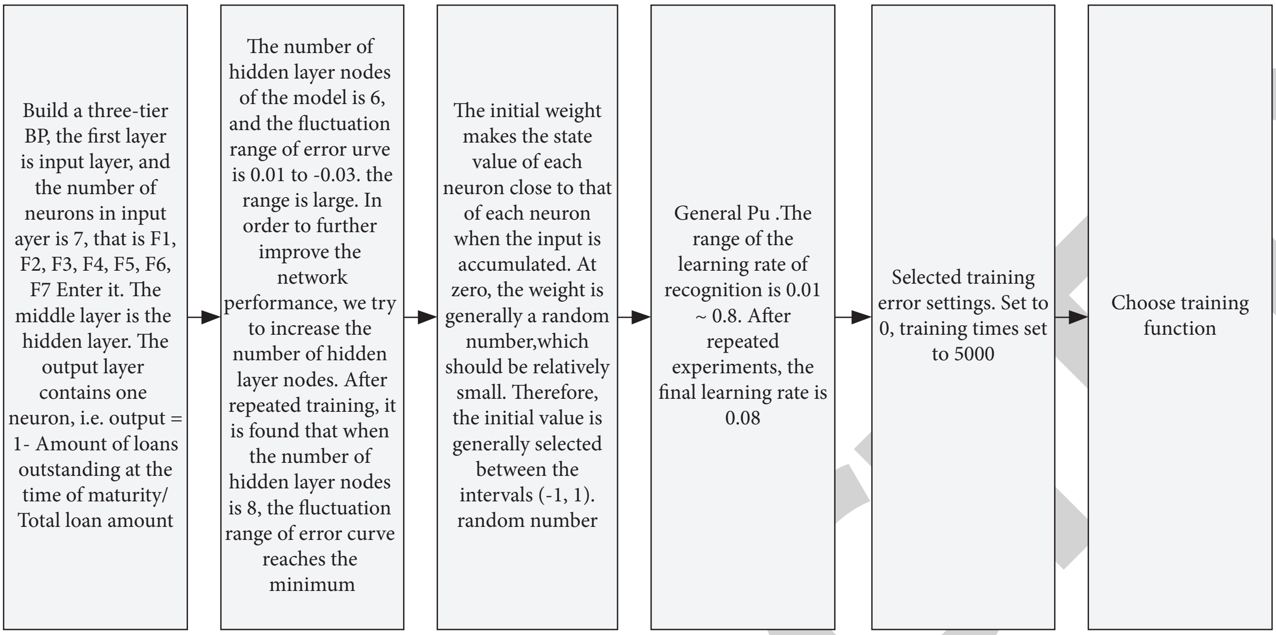
After obtaining the proportion of nonfinancial indicators, a nonfinancial submodel can be constructed as given in the following equation:

$$\begin{aligned} S = {} & \left(0.3339 \cdot Q_z\right) + \left(0.3339 \cdot Q_w\right) + \left(0.1602 \cdot V\right) \\ & + \left(0.1187 \cdot D_1\right) + \left(0.0533 \cdot D_2\right), \end{aligned} \tag{3}$$

where S represents the comprehensive score of nonfinancial indicators, Q_z represents the total assets of the enterprise, Q_w represents the annual income of the enterprise, V represents the type of audit opinion, D_1 represents the shareholding ratio of the largest shareholder, and D_2 represents the shareholding ratio of the two major shareholders.

3.2.3. Construction of Enterprise Credit Risk Early Warning Model. The input variables must be determined initially when creating a model. The output variables can be calculated using the input variables. The enterprise comprehensive credit score prediction based on the discrete choice model is shown in Figure 5.

The enterprise financial indicator submodel and the nonfinancial indicator submodel are currently being built, and the enterprise customer's financial and nonfinancial



Build a three-tier BP, the first layer is input layer, and the number of neurons in input ayer is 7, that is F1, F2, F3, F4, F5, F6, F7 Enter it. The middle layer is the hidden layer. The output layer contains one neuron, i.e. output = 1- Amount of loans outstanding at the time of maturity/ Total loan amount

The number of hidden layer nodes of the model is 6, and the fluctuation range of error urve is 0.01 to -0.03. the range is large. In order to further improve the network performance, we try to increase the number of hidden layer nodes. After repeated training, it is found that when the number of hidden layer nodes is 8, the fluctuation range of error curve reaches the minimum

The initial weight makes the state value of each neuron close to that of each neuron when the input is accumulated. At zero, the weight is generally a random number,which should be relatively small. Therefore, the initial value is generally selected between the intervals (-1, 1). random number

General Pu .The range of the learning rate of recognition is 0.01 ~ 0.8. After repeated experiments, the final learning rate is 0.08

Selected training error settings. Set to 0, training times set to 5000

Choose training function

Figure 2: Steps of model construction.

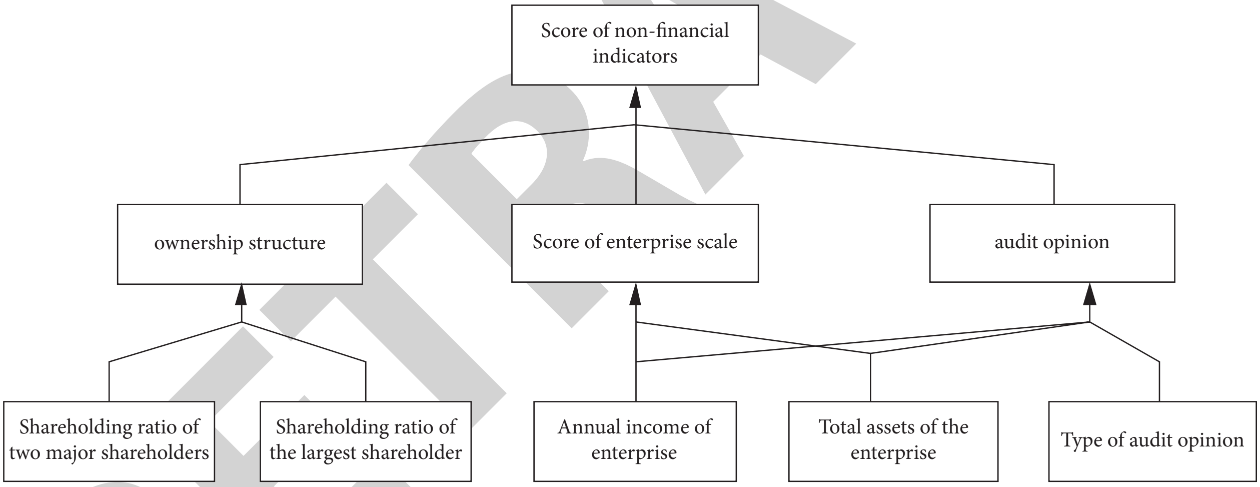

Score of non-financial indicators
ownership structure
Score of enterprise scale
audit opinion
Shareholding ratio of two major shareholders
Shareholding ratio of the largest shareholder
Annual income of enterprise
Total assets of the enterprise
Type of audit opinion

Figure 3: Hierarchy model of nonfinancial indicators.

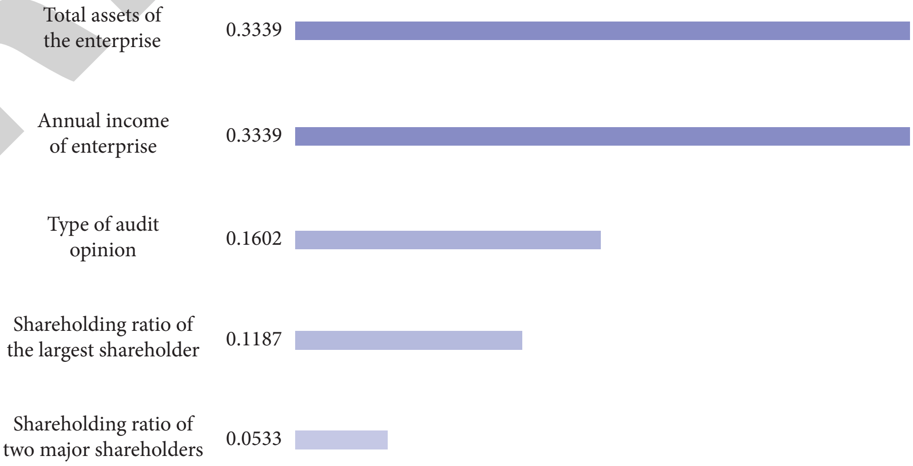


Figure 4: Proportion of nonfinancial indicators.

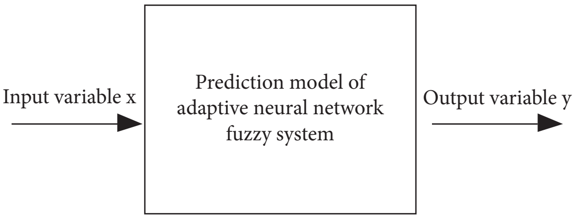


Figure 5: Prediction of enterprise comprehensive credit score.

indicator scores can be obtained using the two models. In this way, an enterprise credit risk early warning model based on discrete choice model can be constructed. The warning model has two input variables, namely, the financial indicator score and the nonfinancial indicator score of the enterprise. It has one output: "1-customer default probability." Default probability refers to the possibility that the borrower can repay the loan principal and interest or fulfill the relevant obligations according to the contract within a certain period of time in the future, and it is inversely proportional to credit risk. Usually, the probability of default refers to the one-year default probability [23].

3.3. Design of Enterprise Credit Risk Early Warning Algorithm. The sample fitting design enterprise credit risk early warning algorithm is based on the created enterprise credit risk early warning model. The specific steps of the enterprise credit risk early warning algorithm are as follows:

(1) The relative importance of each indicator was quantified: The importance of the credit risk indicator was scored. After removing the top and lowest scores, the average score of the left is the indicator's score [24, 25].

(2) A warning judgment matrix was constructed: If a_{ij} represents the relative importance scale of elements a_i and a_j to elements in the previous level, then a positive reciprocal judgment matrix can be constructed as follows:

$$A = \left(a_{ij}\right)_{n\times m} = \begin{bmatrix} \frac{a_1}{a_2} & \frac{a_1}{a_2} & \cdots & \frac{a_1}{a_n} \\ \frac{a_2}{a_1} & \frac{a_2}{a_2} & \cdots & \frac{a_2}{a_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{a_n}{a_1} & \frac{a_n}{a_2} & \cdots & \frac{a_n}{a_n} \end{bmatrix}_{n\times m}, \tag{4}$$

where A represents the positive reciprocal judgment matrix, and n and m represent constants, respectively.

Based on the positive reciprocal judgment matrix, the enterprise credit risk early warning judgment matrix was constructed, including two input early warning judgment matrices as given in matrices 5 and 6:

$$R_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad R_4 = \begin{bmatrix} 1 & \frac{1}{2} & 2 & 1 & \frac{1}{6} \\ 2 & 1 & 4 & 2 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & 1 & \frac{1}{2} & \frac{1}{8} \\ 1 & \frac{1}{2} & 2 & 1 & \frac{1}{6} \\ 6 & 4 & 8 & 6 & 1 \end{bmatrix}, \tag{5}$$

where R_3 and R_4 represent two input warning judgment matrices.

An output warning judgment matrix is as follows:

$$G = \begin{bmatrix} 1 & 1 & 2 & \frac{1}{3} & 2 \\ 1 & 1 & 2 & \frac{1}{3} & 2 \\ \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{4} & 1 \\ 3 & 3 & 4 & 1 & 4 \\ \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{4} & 1 \end{bmatrix}, \tag{6}$$

where G represents the output warning judgment matrix.

Table 7: Specific information of experimental subjects.

Features	A area	B area	C area	D area	E area	Total
Industry attribute						
Agriculture, forestry, animal husbandry, and fishery	0	1	1	4	0	6
Mining industry	0	10	3	5	3	21
Manufacturing industry	2	85	48	48	162	345
Construction business	0	0	1	3	3	7
Wholesale and retail	0	0	0	3	2	5
Leasing business service industry	0	0	0	1	0	1
Transportation, warehousing, and postal services	0	0	0	0	2	2
Production and supply of electric gas and water	0	0	0	0	2	2
Other industries	0	7	1	6	11	25
Total	2	103	54	70	185	414
Types of enterprises						
State-owned enterprise	0	0	0	0	3	3
Collective enterprise	0	4	0	3	10	17
Private enterprise	2	99	54	67	172	394
Total	2	103	54	70	185	414
Enterprise age						
Within 5 years	2	43	42	20	73	180
6–15 years	0	48	10	43	93	194
16–25 years	0	10	2	5	15	32
More than 26 years	0	2	0	2	4	8
Total	2	103	54	70	185	414
Credit rating						
Class A	0	11	9	7	29	56
Class AA	1	74	15	37	91	218
Class AAA	1	18	30	26	65	140
Total	2	103	54	70	185	414

A fuzzy layer warning judgment matrix is as follows:

$$R_5 = \begin{bmatrix} 1 & 3 & 3 & 1 & 4 \\ \frac{1}{3} & 1 & 1 & \frac{1}{3} & 2 \\ \frac{1}{3} & 1 & 1 & \frac{1}{3} & 2 \\ 1 & 3 & 3 & 1 & 4 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & 1 \end{bmatrix}, \tag{7}$$

where R_5 represents the fuzzy layer warning judgment matrix.

(3) The two input warning judgment matrices are utilized to create one output warning judgment matrix and one fuzzy layer warning judgment matrix.

The enterprise credit risk early warning was realized based on the aforementioned judgment matrix. This will assist in preventing the crisis and ensuring the company's steady development, allowing the company to grow sustainably.

4. Experimental Researches

In order to ensure the validity of the experimental results, compare the algorithm in this paper with the algorithms proposed by asset-pricing model [2] algorithm and two-stage econometric approach proposed in [3]. The sample fit, time used, and accuracy of three algorithms were compared. The sample fit is judged through the volatility of the sample fitting curve. If the volatility of the sample fitting curve is small, the sample is proved to have good fitting performance.

4.1. Experimental Process

4.1. Experimental Process. Enterprise credit risk warning algorithm based on big data analytics and discrete choice model was used to conduct experiment on enterprise credit risk early warning. The data utilized in the experiment are all from my sea data set, and the data is analyzed using the online data analysis software MOA (an experimental tool for massive online analysis). 414 enterprises from 5 regions were selected as experimental objects. The specific information is shown in Table 7.

Further, 414 experimental enterprises were subdivided and analyzed in their asset scale, number of employees, and sales revenue. The relevant statistical results are shown in Table 8.

In order to ensure the validity of the experimental results, compare the algorithm in this paper with the algorithms proposed in [2, 3]. The sample fit, time used, and accuracy of three algorithms were compared. The sample fit is judged through the volatility of the sample fitting curve. If the volatility of the sample fitting curve is small, the sample is proved to have good fitting performance.

4.1.1. Sample Fit. Figure 6 depicts the sample fit comparison findings of the algorithm in this paper with the algorithms presented by [2, 3].

Table 8: Statistical description of experimental samples.

	"A" area	"B" area	"C" area	"D" area	"E" area	Total
Asset scale (unit: 10000 yuan)						
0–200 (contain)	0	5	0	2	2	9
200–500 (contain)	0	4	0	7	11	22
500–1000 (contain)	0	25	8	12	46	91
1000–3000 (contain)	1	61	31	23	71	187
3000–5000 (contain)	0	4	7	11	22	44
5000 above	1	4	8	15	33	61
Total	2	103	54	70	185	414
Number of employees (unit: No.)						
0–50 (contain)	0	24	17	15	27	83
50–200 (contain)	2	65	33	28	98	226
200–500 (contain)	0	11	4	14	35	64
500–100 (contain)	0	3	0	10	23	36
1000 above	0	0	0	3	2	5
Total	2	103	54	70	185	414
Sales revenue (unit: 10000 yuan)						
0–200 (contain)	0	8	0	4	2	14
200–500 (contain)	0	4	1	8	7	20
500–1000 (contain)	0	18	2	5	7	32
1000–3000 (contain)	0	40	17	22	50	129
3000–5000 (contain)	1	8	15	11	33	68
5000 above	1	25	19	20	86	151
Total	2	103	54	70	185	414

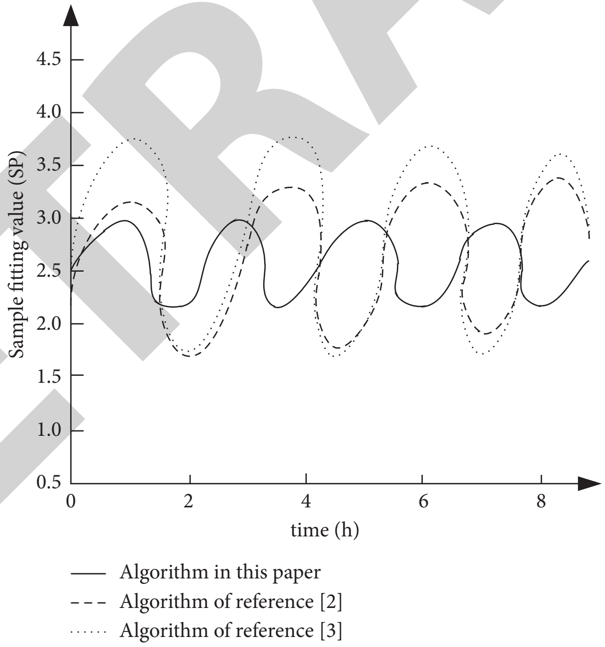


Figure 6: Experimental comparison results of sample fit.

According to Figure 6, the sample fitting curve of the algorithm in this paper has less volatility, which proves that its sample fit is better than those of the asset-pricing model [2] algorithm and two-stage econometric approach proposed in [3].

4.1.2. Time Used. A comparison was done between the time needed by the method in this paper and the time used by algorithms presented by [2, 3] based on the number of enterprises necessary to compute, and the results are displayed in Figure 7.

According to the experimental results in Figure 7, the enterprise credit risk early warning algorithm based on big data analytics and discrete choice model takes less time, and its curve gradually flattens after the number of company's reaches 1000 and does not change until the number of enterprises reaches 2000. It depicts that when the number of

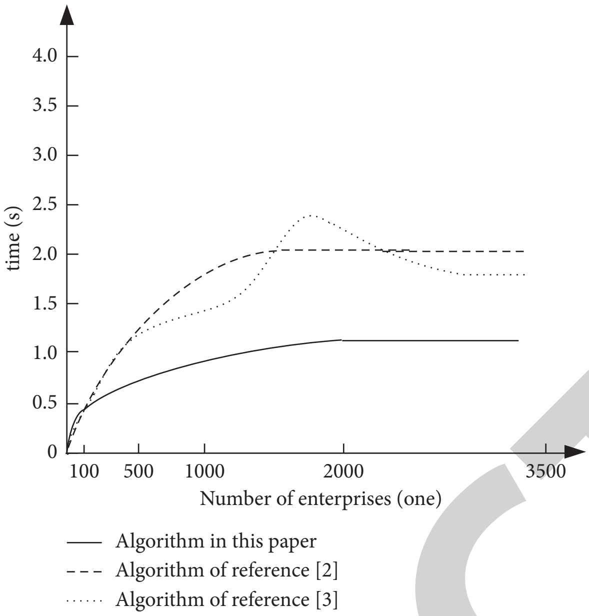


Figure 7: Experimental comparison results of time used.

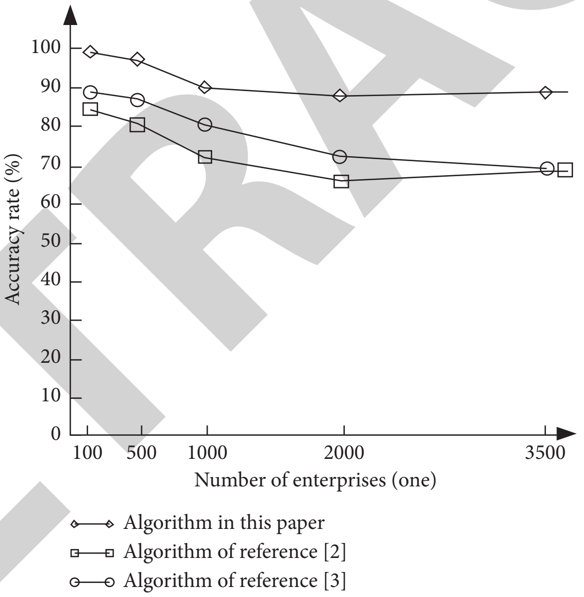


Figure 8: Experimental comparison results of accuracy.

enterprises is greater than 100, the enterprise credit risk early warning algorithm in this paper uses less time than the other two algorithms. This is because the algorithm in this paper firstly filters the enterprise credit risk before constructing the sample. When there are a significant number of enterprises that have set aside sufficient time for the prompt prevention of corporate credit risk, this step allows the algorithm to make the credit risk warning in a shorter amount of time.

4.1.3. Accuracy. The accuracy rate of three algorithms is used to assess the accuracy of their calculating results. The accuracy of the algorithm outcomes is high if the accuracy rate is high and stable. The algorithm's outputs, on the other hand, have a poor level of accuracy. Figure 8 depicts the comparing results.

Figure 8 shows that the algorithm in this paper has a higher accuracy and is more stable. Therefore, the enterprise credit risk early warning algorithm based on big data analytics and discrete choice model has higher accuracy. This is related to the sample fit experiment in 3.2.1. The higher the sample fitting performance, the higher the accuracy rates of the algorithm and the higher the accuracy of the algorithm results. Because the algorithm in this paper improves the sample fitting performance, the accuracy of the algorithm results is also improved, which makes the enterprise credit risk early warning more accurate, reduces unnecessary credit risk prevention, and avoids waste of resources.

5. Conclusions

The early warning system can protect an enterprise from the biggest loses and bankruptcy. Hence, some smart techniques need to be devised in order to warn the enterprise regarding credit risk. The early warning system for enterprise credit risk in this study is able to evaluate properly the potential credit risk of the futuristic credit related activities. The proposed mechanism is applied on the dataset and it shows good results in terms of sample fitting performance, complexity, and accuracy. The system can warn in advance regarding the potential credit risks and it has a higher accuracy as compared to the conventional techniques. It introduces a novel technique for early warning of company credit risk, as well as technical assistance to help businesses to increase their business without having fears of loses and competitiveness. With the influence of enterprise credit risk early warning on current social stability factors, the algorithm should have extensive application space. The proposed technique shows higher accuracy and takes minimal time in producing results and the data is also fitted properly. The accuracy obtained by the proposed algorithm proves the viability of the suggested method in this paper. However, the technique presented in this paper is one-sided in its application, and therefore it is not appropriate for all enterprises. As a result, the next study will concentrate on broadening the algorithm's reach.

Data Availability

The data used to support the findings of this study are available on request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding

This study did not receive any funding in any form.

References

[1] D. Xiao, "International experience for SME credit risks identification: based on Monitoring-Cashflow method," in *Proceedings of the 2011 International Conference on Management and Service Science*, pp. 1–4, Bangkok, Thailand, May 2011.

[2] M. Wang, J. Yu, and Z. Ji, "Credit risk assessment of high-tech enterprises based on RSNCL-ANN ensemble model," in *Proceedings of the 2018 International Conference on Math and Art Intelligence (ICMAI)*, pp. 73–78, Chongqing, China, March 2018.

[3] J. Galindo and P. Tamayo, "Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications," *Computational Economics*, vol. 15, no. 1/2, pp. 107–143, 2000.

[4] X. Xiao-si, C. Ying, and R. Ruo-en, "Studying on forecasting the enterprise bankruptcy based on SVM," in *Proceedings of the 2006 International Conference on Management Science and Engineering*, pp. 1041–1045, Lille, France, October 2006.

[5] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable machine learning in credit risk management," *Computational Economics*, vol. 57, no. 1, pp. 203–216, 2021.

[6] R. S. Kenett and S. Salini, "Modern analysis of customer satisfaction surveys: comparison of models and integrated analysis," *Applied Stochastic Models in Business and Industry*, vol. 27, no. 5, pp. 465–475, 2011.

[7] M. Kaur and S. Kadam, "Bio-inspired workflow scheduling on HPC platforms," *Tehnički Glasnik*, vol. 15, no. 1, pp. 60–68, 2021.

[8] A. Ono, R. Hasumi, and H. Hirata, "Differentiated use of small business credit scoring by relationship lenders and transactional lenders: evidence from firm-bank matched data in Japan," *Journal of Banking and Finance*, vol. 42, pp. 371–380, 2014.

[9] L. Zhang, H. Hu, and D. Zhang, "A credit risk assessment model based on SVM for small and medium enterprises in supply chain finance," *Financial Innovation*, vol. 1, no. 14, 2015.

[10] M. Kaur, "FastPGA based scheduling of dependent tasks in grid computing to provide QoS to grid users," in *Proceedings of the 2016 International Conference on Internet of Things and Applications (IOTA)*, pp. 418–423, Pune, India, January 2016.

[11] N. Yoshino and F. Taghizadeh-Hesary, "Analysis of credit ratings for small and medium-sized enterprises: evidence from Asia," *Asian Development Review*, vol. 32, no. 2, pp. 18–37, 2015.

[12] H. S. Bhamra, C. Dorion, A. Jeanneret, and M. Weber, "Low inflation: high default risk and high equity valuations," *Social Science Electronic Publishing*, vol. 21, pp. 46–70, 2018.

[13] T. Madjia, A. Amir-Reza, D. C. Debora, and P. Maryam, "An artificial neural network and Bayesian network model for liquidity risk assessment in banking," *Neurocomputing*, vol. 275, pp. 2525–2554, 2018.

[14] K. W. Li, "Analyzing the TFP performance of Chinese industrial enterprises," *The Singapore Economic Review*, vol. 63, pp. 194–217, 2018.

[15] C. Zhao-Quan, C. Guang-Cai, X. Li-Ning, Y. Jing-Hui, and T. Xu, "Evaluating hedge fund downside risk using a multiobjective neural network," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 433–438, 2019.

[16] L. Jiang, S. R. Sakhare, and M. Kaur, "Impact of industrial 4.0 on environment along with correlation between economic growth and carbon emissions," in *International Journal of System Assurance Engineering and Management*, Springer Science and Business Media LLC, Berlin, Germany, 2021.

[17] S. Yu, G. Chi, and X. Jiang, "Credit rating system for small businesses using the K-S test to select an indicator system," *Management Decision*, vol. 57, no. 1, pp. 229–247, 2019.

[18] L. Chen, V. Jagota, and A. Kumar, "Research on optimization of scientific research performance management based on BP neural network," *International Journal of System Assurance Engineering and Management*, 2021.

[19] K. Mahajan, U. Garg, and M. Shabaz, "CPIDM: a clustering-based profound iterating deep learning model for HSI segmentation," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 7279260, 12 pages, 2021.

[20] A. Fronzetti Colladon and E. Remondi, "Using social network analysis to prevent money laundering," *Expert Systems with Applications*, vol. 67, pp. 49–58, 2017.

[21] J. Gupta, N. Wilson, A. Gregoriou, and J. Healy, "The effect of internationalisation on modelling credit risk for SMEs:

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9783897, 1 page
https://doi.org/10.1155/2023/9783897

Retraction

Retracted: Fast Extraction Algorithm for Local Edge Features of Super-Resolution Image

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] F. Chen and B. Yang, "Fast Extraction Algorithm for Local Edge Features of Super-Resolution Image," *Security and Communication Networks*, vol. 2022, Article ID 8801978, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8801978, 10 pages
https://doi.org/10.1155/2022/8801978

Research Article

Fast Extraction Algorithm for Local Edge Features of Super-Resolution Image

Feng Chen and Botao Yang

Rongcheng Campus, Harbin University of Science and Technology, Harbin 150080, China

Correspondence should be addressed to Feng Chen; chenfeng@hrbust.edu.cn

Received 21 October 2021; Revised 7 December 2021; Accepted 10 December 2021; Published 5 January 2022

Academic Editor: Chinmay Chakraborty

Image super-resolution is getting popularity these days in diverse fields, such as medical applications and industrial applications. The accuracy is imperative on image super-resolution. The traditional approaches for local edge feature point extraction algorithms are merely based on edge points for super-resolution images. The traditional algorithms are used to calculate the geometric center of gravity of the edge line when it is near, resulting in a low feature recall rate and unreliable results. In order to overcome these problems of lower accuracy in the existing system, an attempt is made in this research work to propose a new fast extraction algorithm for local edge features of super-resolution images. This paper primarily focuses on the super-resolution image reconstruction model, which is utilized to extract the super-resolution image. The edge contour of the super-resolution image feature is extracted based on the Chamfer distance function. Then, the geometric center of gravity of the closed edge line and the nonclosed edge line are calculated. The algorithm emphasizes on polarizing the edge points with the center of gravity to determine the local extreme points of the upper edge of the amplitude-diameter curve and to determine the feature points of the edges of the super-resolution image. The experimental results show that the proposed algorithm consumes 0.02 seconds to extract the local edge features of super-resolution images with an accuracy of up to 96.3%. The experimental results show that our proposed algorithm is an efficient method for the extraction of local edge features from the super-resolution images.

1. Introduction

The super-resolution technology of images is a technique for obtaining high-resolution images corresponding to scenes by using an existing method for low-resolution images without changing the image observation system [1]. Image super-resolution technology is improving day by day due to the huge demand in computer science and aligned fields [2]. It is widely used in medical imaging, video surveillance and transmission, generation of satellite remote sensing images, and HDTV [3]. In order to process images and to draw meaningful inferences, it is necessary to extract the edge's local features of super-resolution images [4]. In this paper, the features of super-resolution images are extracted fast from the point of view of local feature points. Since the points are the primitives that constitute the super-resolution image, the points constituting the image vary widely. Therefore, it needs to extract some specific features that can represent the image attributes and can assist in image feature extraction and recognition. Feature extraction is also important to identify and track the target according to different needs and to construct a three-dimensional target surface [5, 6]. In the past research endeavours, many feature point extraction algorithms are proposed and deployed, which are mainly divided into two categories. One is the curvature-based local edge feature extraction algorithm of super-resolution image, and the other is the grey gradient-based algorithm. Both algorithms have disadvantages. The first type of algorithm has a large amount of calculation, and the second type of algorithm has low accuracy. At present, there is a feature point extraction algorithm based on edge points. Compared with the performance of the first two types of algorithms, the algorithm is simple and convenient to implement, but it has the following problems: first, only the edge line is considered. It is closed and does not conform to the actual situation. Secondly, the obtained local edge features of the super-resolution image are only the convex curvature with large curvature on the edge line, and the description of the shape of the target is incomplete [7].

In [8], the authors have introduced a novel method for extracting transform characteristics from pictures or video frames. These characteristics are used to represent the local visual content of image and video frames. The projected method, such as Shot Boundary Detection (SBD), is measured using conventional methods utilizing the standard procedure. The experimental results reveal that the proposed method outperforms previous methods in terms of computational cost. In [9], the authors have investigated various picture feature extraction analysis techniques. By aggregating low-level characteristics to explore various feature data representations, this technique obtains more expressive and productive high-level information content. In [10], the authors have suggested a random deep neural network-based picture feature extraction technique. The goal of this strategy is to detect more consistent features by eliminating duplicate feature points. In [11], the authors have introduced a novel approach based on Bidimensional Empirical Mode Decomposition. This method was used to extract self-adaptive characteristics from pictures. In [12], the authors have designed a video summarization framework based on frame choice to determine only significant frames. As there are various drawbacks in the existing systems, we aim to produce an efficient algorithm to solve the above problems, which is fast and efficient than the previous approaches for local edge feature extraction of super-resolution images. The proposed algorithm will accurately reflect the shape contour of the target feature, which has important significance to extract the super-resolution image features.

The contribution of the work is as follows:

(i) In this paper, a new fast extraction algorithm for local edge features of super-resolution images is proposed

(ii) The algorithm mainly focuses on the super-resolution image reconstruction model, which is used to obtain the super-resolution image

(iii) The algorithm puts emphasis on polarizing the edge points with the center of gravity as the pole to find the local extreme points of the upper edge of the amplitude-diameter curve and determine the feature points of the edge of the super-resolution image

(iv) The algorithm is compared using a super-resolution image reconstruction model and local edge feature extraction of a super-resolution image on the basis of which we calculate the geometric center of gravity and polarization of edges

(v) So, the proposed algorithm produces efficient output than the existing traditional approaches when compared using the RPC curve and keeping *F*-measure as a calculatory parameter

(vi) The proposed algorithm gives an efficiency of around 99% that surpasses the traditional approaches with a great margin, which was about 75% and 64%, respectively

Further, the paper is divided into five sections:

(1) Section 1 gives an introduction about the existing approaches and shows the drawbacks of existing approaches

(2) Section 2 shows the various approaches to the image reconstruction model and local edge feature extraction model and the changes to be made in them to make them efficient

(3) Section 3 shows us the comparative result analysis of various algorithms with respect to the proposed system

(4) Section 4 is the discussion of results and the various approaches made to make the algorithm efficient

(5) Section 5 tells us about the obtained results and how much efficient our approach is with respect to others

2. Material and Methods

2.1. Super-Resolution Image Reconstruction Model. Degraded models for super-resolution image reconstruction can be expressed as follows:

$$y_k = DW_kF_kX + \eta_k, \tag{1}$$

where y_k is the degraded kth frame image, X is the high-resolution image, D and W_k are the downsampling matrix and motion matrix, F_k is the fuzzy matrix, and η_k is the noise.

Let the low-resolution image be Y and the corresponding high-resolution image be X. The problem that the super-resolution reconstruction needs to solve is to find the optimal approximate solution X under the condition of known Y. The common method for solving this problem is Maximum Posterior Probability (MAP) under low-resolution range image conditions. MAP estimates can be represented by

$$\hat{x} = \arg\max_{X} P\left(X|Y_1, \ldots, Y_N\right). \tag{2}$$

According to the Bayesian estimation criterion, (2) can be rewritten as follows:

$$\hat{x} = \arg\max_{X} P\left(Y_1 = y_1, \ldots, Y_N = y_N|X = x\right)P(X = x). \tag{3}$$

In (3), it is assumed that the noise is independent Gaussian white noise [10] and the variance is σ^2. Then, (4) is expressed as follows:

$$P\left(Y_1 = y_1, \ldots, Y_N = y_N | X = x\right) = \frac{1}{\left(2\pi\sigma^2\right)^{N_1N_2}} \times \ \exp\left(-\sum_{k=1}^{N}\frac{\left\|y_k - DW_kF_kX\right\|_2^2}{2\sigma^2}\right). \tag{4}$$

According to the Markov random field model, the prior probability $P(X = x)$ of the high-resolution range image is obtained, and the equivalence between MRF and Gibbs is used. The Gibbs distribution is used to explicitly describe the Markov distribution; that is, the prior probability of X can be expressed by

$$P(X = x) = \frac{1}{Z}\exp\left(-\frac{1}{T}\sum_{c\in C_X} V_c^x(x)\right). \tag{5}$$

The energy function takes the form of Li, as follows:

$$\begin{aligned}\sum_{c\in C_X} V_c^x(x) = \sum_{i=1}^{L_1}\sum_{j=1}^{L_2} 4\gamma - \\ \gamma\ \exp\left\{\frac{-[x(i,j) - x(i,j-1)]^2}{\gamma}\right\} - \\ \gamma\ \exp\left\{\frac{-[x(i,j) - x(i,j+1)]^2}{\gamma}\right\} - \\ \gamma\ \exp\left\{\frac{-[x(i,j) - x(i-1,j)]^2}{\gamma}\right\} - \\ \gamma\ \exp\left\{\frac{-[x(i,j) - x(i+1,j)]^2}{\gamma}\right\}.\end{aligned} \tag{6}$$

Here, Z is a normalized constant, T is the temperature parameter, and $V_c^x(x)$ is the potential function of the connected group; the potential function describes the interaction of a set of neighbouring pixels, and different potential functions determine the different MRF models. From (4) and (5), (3) can be rewritten as follows:

$$\widehat{x} = \arg\min_X\left(\sum_{k=1}^{N}\left\|y_k - DW_kF_kX\right\|_2^2 + \sum_{c\in C_X} V_c^x(x)\right). \tag{7}$$

Since the energy function of the prior distribution of the DAMRF model is a nonconvex function when solving the optimal solution of the objective function, it is easy to fall into the local minimization problem, and the optimal approximate solution of the reconstructed image cannot be obtained [7, 11]. Therefore, the graduated nonconvexity (GNC) optimization algorithm is used to optimize the objective function to obtain the optimal solution of the reconstruction result.

2.2. Local Edge Feature Extraction of Super-Resolution Image

2.2.1. Chamfer Matching Metrics. The Chamfer distance is used to measure the similarity of the two-edge figures. The match between the template map T and the image to be matched E is achieved by searching for their minimum Chamfer distance. The main steps are as follows:

Step 1. Calculate the Chamfer distance map of the image to be matched.

Step 2. Superimpose the template on the distance map. Calculate the Chamfer distance between the template and the image to be matched as follows:

$$D_{\mathbf{Chamfer}} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} v_i. \tag{8}$$

Here, n is the number of edges of the template and v_i is the distance value at which the template is superimposed.

Step 3. The template is translated on the distance map to obtain the Chamfer distance value distribution function $S(p)$ of the template on the image to be matched, and the position vector p of the minimum value $S(p)$ is the best matching point. In practical applications, image features are extracted by determining whether the minimum value $S(p)$ is less than a set threshold θ.

2.2.2. Local Edge Contour Feature Function Based on Class Chamfer Distance. The local edge features used in this paper are defined by a rectangular window, that is, $r = (x, y, w, h)$. Each local edge r is represented by two positional parameters (x, y) and two scaled parameters (w, h), which, respectively, represent the width and height of the rectangle. The local edge feature F_r is defined on the Chamfer distance map I of the image, concerning (8), and the eigenvalue calculation is as follows:

$$F_r(I) = \frac{1}{w\times h}\sum_{i=x}^{x+w}\sum_{j=y}^{y+h} v_{i,j}, \tag{9}$$

where $v_{i,j}$ is the value of the Chamfer distance at the corresponding point in the image.

This paper implements the fast calculation of (9) by establishing the integrated image of the Chamfer distance map. For the distance graph I, as shown in Figure 1, the integral image value at a pixel (x, y) is defined as $ii(x,y) = \sum_{x_1\le x, y_1\le y} i(x_1, y_1)$, that is, the sum of all the pixel values of the shaded portion. Once the integral image is established, the local edge feature values of any parameters can be obtained by only 4 table lookups and simple operations.

The super-resolution feature extracted by the above method is the feature edge contour, so the geometric center of gravity of the feature contour needs to be calculated to determine the feature points that meet the requirements.

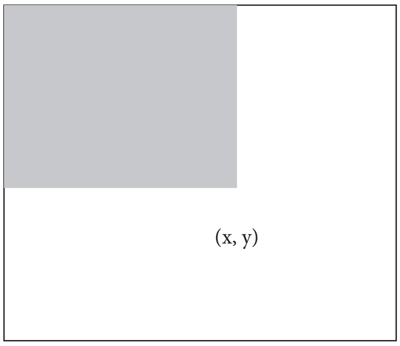


Figure 1: Integral image value at the point (x, y).

2.3. Geometric Center of Gravity Calculation. The geometric center of gravity is obtained by weighting the points in the Grassmannian space and adding them; then it is divided by the sum of ownership. There are many state-of-the-art approaches that deal with calculation of geometric center of gravity [13, 14]. The pixel points of the image have greyscale properties, but the feature points extracted in this paper are on the edge contour line, and the edge image is a binary plane image, which is independent of the grey level of the image. Therefore, the edge image can be considered as a uniform substance. So, the geometric center of gravity of the edge contour is the geometric center.

Considering whether the extracted edge lines are closed, the edge contours can be divided into two categories: closed contours and nonclosed contours, and their centres of gravity are calculated below.

(1) *Calculation for the Geometric Center of Gravity of the Closed Contour.*

For a closed irregular planar figure, let the coordinates of the edge points be (x_i, y_i) and $1 \leq i \leq n$ be the number of edge points. The geometric center coordinate (x_0, y_0) is calculated by

$$x_0 = \frac{1}{n} \sum_{i=1}^{n} x_i, y_0 = \frac{1}{n} \sum_{i=1}^{n} y_i. \tag{10}$$

The geometric center of gravity, thus, obtained is unique. The experimental results are shown in Figure 2(a). In the figure, the circle represents the center of gravity of the geometry, and the contour of the feature extraction is closed.

(2) *Calculation for the Geometric Center of Gravity of the Unclosed Contour.*

Since the three points that are not collinear on the plane are linearly independent, the coordinates for the center of gravity of the triangle are uniquely defined [15, 16]. Therefore, for a convex or concave curve, it can first form a triangle consisting of the two ends of the edge line and the middle point and then find the geometric center of gravity. If an edge line is not single convex or single concave, then it is segmented. The experimental results are shown in Figure 2(b). In the figure, the curve $\overset{\frown}{AC}$ is divided into segments from point B, and the geometric centers of gravity of the arcs $\overset{\frown}{AC}$ and $\overset{\frown}{BC}$ are, respectively, obtained. The solid line is the outline of the feature extraction, which is nonclosed, and the circle represents the center of gravity of the geometry.

2.4. Polarization of Edge Points. To conveniently calculate the length and position of each point on the edge from the geometric center of gravity, it is necessary to polarize the edge point. Taking the geometric center of gravity as the pole, as shown in Figure 3, the conversion equation is as follows:

$$\begin{cases} \rho_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \\ \theta_i = \arctan\left(\dfrac{y_i - y_0}{x_i - x_0}\right) \end{cases} \tag{11}$$

In Figure 3, the origin of the coordinates is represented by (x_0, y_0), and the points in polar coordinates are represented by (x_i, y_i).

The polarized edge points form a polar-amplitude curve, as shown in Figure 4. In Figure 4, the abscissa is the edge point angle, and the unit is the arc degree; the ordinate is the polar diameter of the edge point, and the unit is taken in micrometres (μm). The horizontal and vertical coordinates' contents in Figures 5 to 8 are the same as those in Figure 4.

2.5. Determination of the Local Feature Points of the Image Edge. Polarization simplifies the problem, making it easy to find extreme points locally and then further identifying the feature points [17, 18]. The extraction process of extreme points and feature points is described below.

2.5.1. Determination of the Extreme Point. If the maximum point in the range $[\theta_1, \theta_2)$ is represented by $P_{\max}$, then $P_{\max}$ can be described by

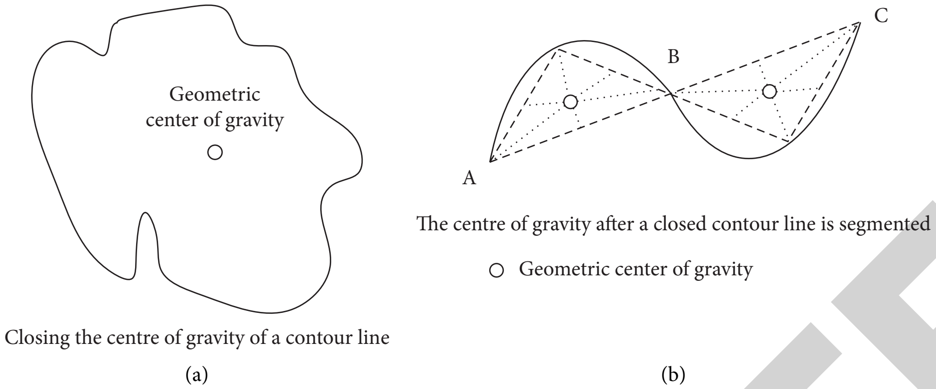
$$P_{\max} = \max_{\theta \in [\theta_1, \theta_2]} (\rho). \tag{12}$$

Similarly, if the minimum value point in the range $[\theta_1, \theta_2)$ is represented by $P_{\min}$, then $P_{\min}$ can be represented by

$$P_{\min} = \min_{\theta \in [\theta_1, \theta_2]} (\rho). \tag{13}$$

The curves of the local maximum point and the minimum point on the polar-amplitude curve are shown in Figures 5 and 6, respectively, and the interval between the two figures is 10°, that is, $1/18\,(\pi)$. In Figure 5, the maximum value of the local maximum point is about 140 μm, and the minimum value is about 20 μm. In Figure 6, the maximum value of the local maximum point is about 110 μm, and the minimum value is about 15 μm.

2.5.2. Determination of Feature Points. The final feature points are obtained based on the principle of nonmaximum (small) value suppression in the local area [19, 20].



Closing the centre of gravity of a contour line

(a)

The centre of gravity after a closed contour line is segmented

(b)

Figure 2: The geometric barycenter of the contour line of an object. (a) Closing the center of gravity of a contour line. (b) The center of gravity after a closed contour line is segmented.

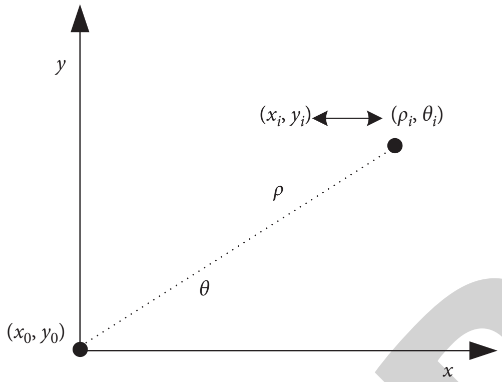


Figure 3: Polar coordinates of edge points.

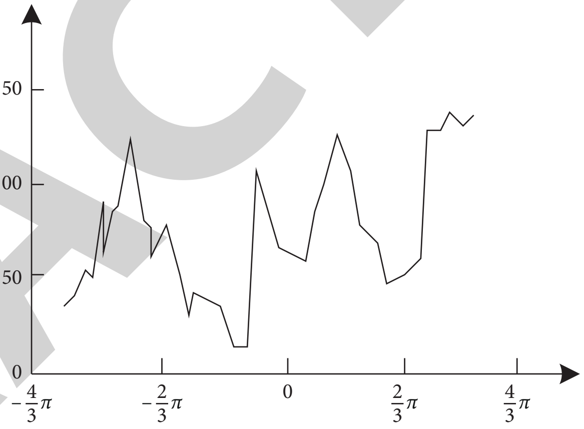


Figure 5: Polar diameter and amplitude curves of local maximum points.

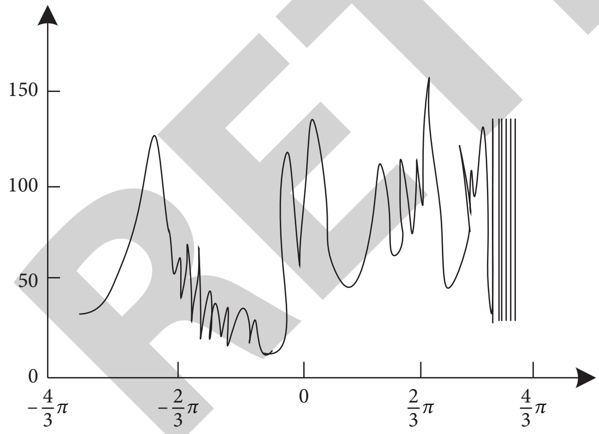


Figure 4: Polar diameter curve.

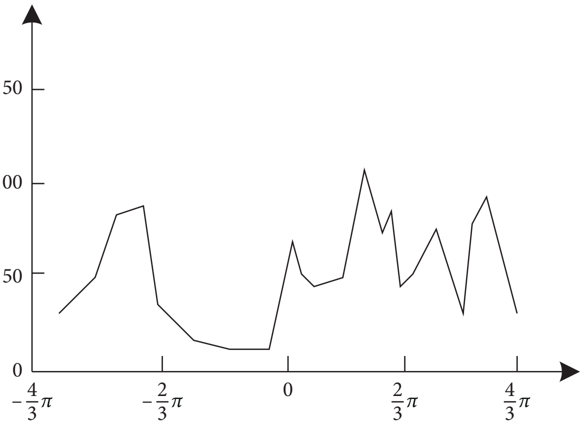


Figure 6: The polar diameter curve of the local minimum point.

Therefore, the maximum (small) value point of each interval and the maximum (small) value point in the adjacent interval are compared, and if the polar diameter of the maximum (minimum) value point is larger (smaller) than that of the maximum (minimum) value point in the two adjacent intervals, it is considered to be a feature point. The argument-polar radius curves of the two feature points further extracted from the maximum and minimum values of Figures 5 and 6 are shown in Figure 7, and the finally extracted feature points are as shown in Figure 8 in Figure 7; the argument-polar radius curve of the maximum value curve is always above the argument-polar radius curve of the minimum value curve; the distribution of the last extracted effective feature points and the argument-polar radius curve

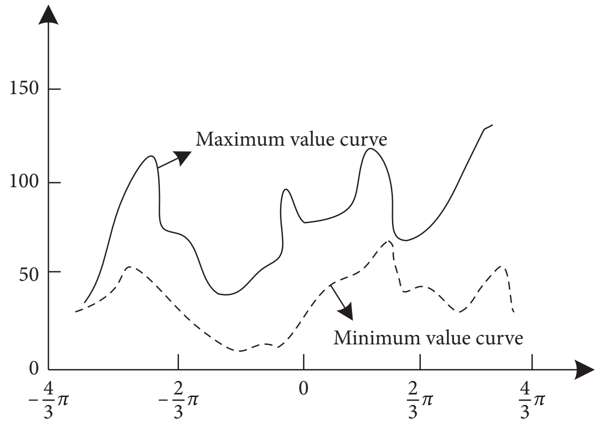


Figure 7: The maximum diameter curve of the maximum diameter (small) value.

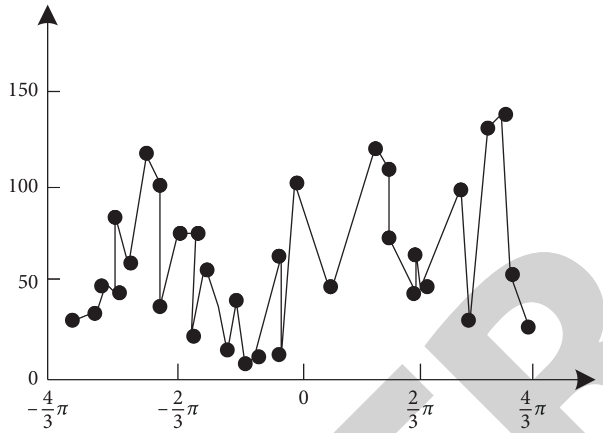


Figure 8: The polar diameter curve diagram of the edge local feature points.

of the feature points can be seen in Figure 7. In Figure 8, it can be observed obviously that the polar diameter curve diagram of the edge local feature points varies from the range of $-4/3\ \pi$to $4/3\ \pi$, representing the values of maximum value curve points and the minimum value curve points on the y-axis ranging from 0 to 150.

3. Results

3.1. Algorithm Validation. To verify the effectiveness of the proposed algorithm, the expansion feature extraction test of a single super-resolution image is selected. Figure 9 is the local edge feature result of the super-resolution image extracted by the proposed algorithm. It can be seen from the analysis of Figure 9 that the algorithm not only extracts the feature points of the super-resolution image convex, but also the feature points of the concave point can be detected. The connection of these feature points can reflect the edge contour shape of the target, which verifies the validity of the local edge features extraction by the proposed algorithm.

To verify the effectiveness of the proposed algorithm, the advantages of the proposed algorithm are highlighted. The super-resolution image of the valve pressure gauge is used as the object and 10% noise is added. The proposed algorithm, method given in paper [4], and method given in paper [5] are used to extract local edge features of the image. The original super-resolution image of the pressure meter with 10% noise is shown in Figure 10(a), and the result of feature extraction is shown in Figures 10(b)–10(d).

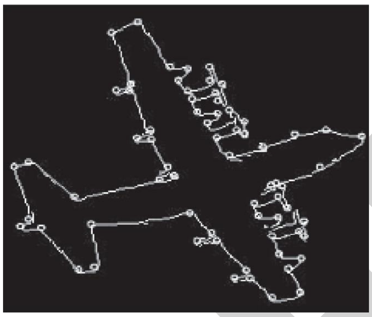


Figure 9: The feature points extracted by this algorithm.

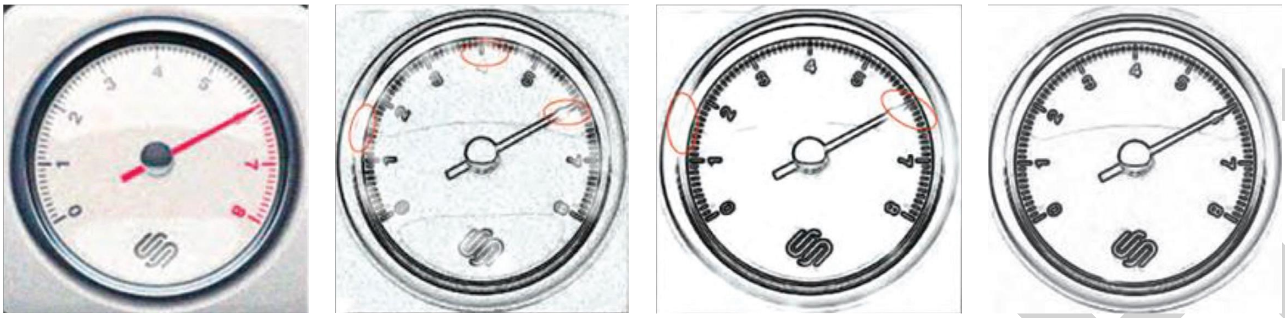
Figure 10(b) is the result of the method given in paper [4] extracting the local edge features of the super-resolution image, and Figure 10(c) is the extraction result of the method given in paper [5]. Due to the feature points extracted by the two algorithms being unclear, they are circled with red lines.

3.2. Comparison of RPC Performance and F-Measure Performance of Different Algorithms

3.2.1. Testing Set. To highlight the advantages of the proposed algorithm, it is compared with the method given in paper [4] and the method given in paper [5]. The experiment uses the super-resolution image database of vehicles in UIUC. The database consists of a training set and a testing set. The training set includes 550 positive samples with size of 100 × 40. The experiment does not increase the number of positive samples in the training set (usually, increasing the number of training samples can improve the accuracy of the classifier), and the testing set consists of two subsets, denoted by T_{I} and T_{II}, respectively. T_{I} contains 170 super-resolution images with a total of 200 vehicles. The scale of vehicle imaging is the same as the training set. T_{II} contains 108 super-resolution images with 139 vehicles. The scale of vehicle imaging is different from the T_{I} testing set. The range is between 0.8 and 2 times. Some testing images contain complex backgrounds. Some images also have partial occlusion and image blur. In general, the feature extraction of testing set T_{II} is more difficult than that of testing set T_{I}. The training set also includes 50,517 negative samples, each of which is 100 × 40 in size.

The experiment uses the recall rate, precision rate, and F-measure to evaluate the performance of the algorithm. The calculation method of the evaluation index is as follows:

(1) The recall-precision curve (RPC) is defined as follows:



Super resolution image of Pressure meter with 10% noise

(a)

Adaboost algorithm

(b)

BA algorithm

(c)

The proposed algorithm

(d)

Figure 10: Local edge feature extraction results from noisy super-resolution images.

$$\text{recall} = \frac{P_T}{P_n}, \tag{14}$$

$$\text{precision} = \frac{P_T}{P_T + P_F}. \tag{15}$$

In the equation, P_T, P_F, and P_n, respectively, indicate the number of correct extraction of features, the number of error extractions, and the total number of features.

(2) The F-measure ($F_$measure) is defined as (16), which can be considered as an equal error measure:

$$F\_\text{measure} = \frac{2\text{recall} \times \text{precision}}{\text{recall} + \text{precision}}. \tag{16}$$

3.2.2. Analysis of Experimental Results. During the experiment, some feature extraction results of the proposed algorithm are shown in Figures 11(a)–11(c).

It can be seen from Figures 11(a)–11(c) that the result of using the algorithm of this paper to extract the vehicle in the super-resolution image can identify the vehicle in the image in the case of obstacle occlusion and verify the effectiveness of the algorithm.

The feature extraction and RPC fold line comparison of the three algorithms in the testing set T_{I} are shown in Table 1 and Figure 12, respectively. The feature extraction and RPC fold line comparison of the three algorithms in the testing set T_{II} are shown in Table 2 and Figure 13, respectively.

In Table 1, the equal error measure of the proposed algorithm is 96.3%, which is 14.1% higher than that of the method given in paper [4] and 21.1% higher than that of the method given in paper [5]. The average feature extraction time of the proposed algorithm is 0.02 s, which saves 0.075 s compared with the method given in paper [4] and 0.036 s compared with the method given in paper [5].

In Figure 12, the RPC polyline arrangement of the proposed algorithm, method given in paper [4], and the method given in paper [5] can be seen. The RPC polyline of the proposed algorithm is located at the top of the line graph. The initial value of the algorithm is 60%. The rate rises linearly, the stability is about 97% in the later period, and the maximum precision is 99%; the initial value of the method given in paper [4] and the method given in paper [5] is about 8%. The maximum precision of the method given in paper [4] is about 75% and that of the method given in paper [5] is approximately 64%. From this, we can clearly state that our proposed algorithm is better in terms of precision compared to the method given in paper [4] and the method given in paper [5] as our algorithm has the maximum of 99% precision rate at its best case and the other algorithms method given in paper [4] and method given in paper [5] are having a maximum of 75% and 64% approximately as our proposed algorithm outcasts them in case of RPC curves.

It can be seen from Table 2 that the F-measure value of the proposed algorithm is higher, and the average feature extraction time is less. Compared with the case of Table 1, the proposed algorithm saves feature extraction time and has high efficiency while maintaining the highest accuracy and high precision.

In Figure 13, the RPC polyline of the proposed algorithm on the testing set T_{II} is located at the top of the image, indicating that the RPC performance of the proposed algorithm is the strongest. Like the testing set T_{I}, the initial value of the proposed algorithm is 60%, which is larger than that of the method given in paper [4] of 51% and the method given in paper [5], respectively; the highest precision of the proposed algorithm is 99%, and the highest precision of the method in paper [4] and the method in paper [5] is 82% and 68%, respectively. Comparing the three groups of data, the RPC performance of the proposed algorithm is superior to similar algorithms and has significant advantages.

4. Discussion

In Figure 10, the feature points extracted by the method given in paper [4] and the method given in paper [5] have a commonality: fuzzy, unclear, and large fragment points; it is difficult to effectively restore the local edge features of super-resolution images, affecting the image analysis effect; relatively speaking, the feature points extracted by the proposed algorithm in Figure 10(d) are clear and significant. The pointers and meter scales of the pressure gauge are clear and

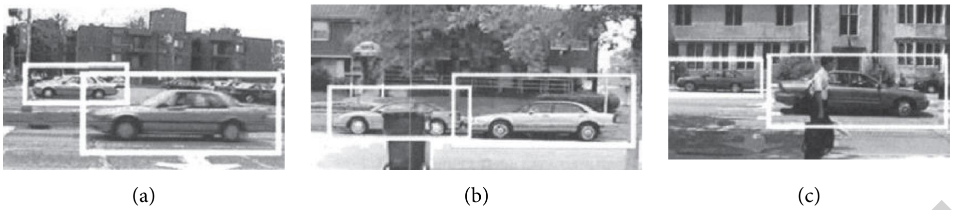


Figure 11: Part of the feature extraction results of the algorithm.

Table 1: Feature extraction of different algorithms on the test set T.

Algorithm	Number of base classifiers	F-measure (%)	Average feature extraction time (s)	Test platform
Method of paper [4]	800	82.2	0.095	1.86 GHz Core 2 Duo
Method of paper [5]	4050	75.2	0.056	1.86 GHz Core 2 Duo
Proposed method	1850	96.3	0.02	1.86 GHz Core 2 Duo

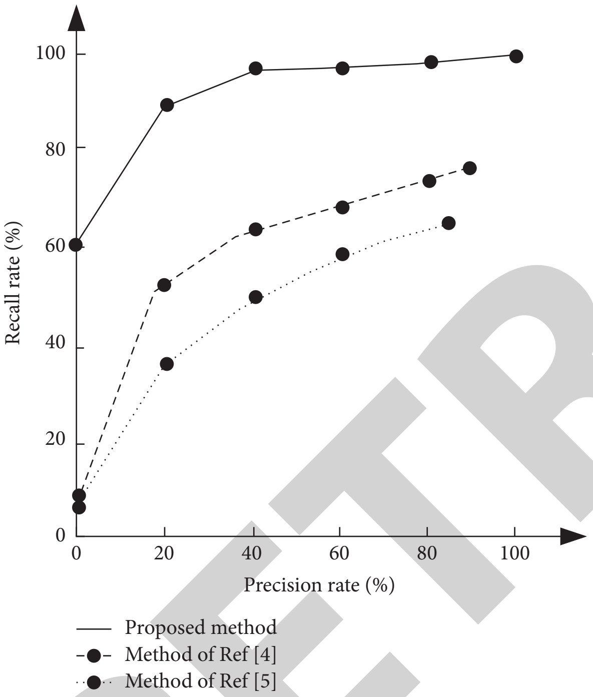


Figure 12: Comparison of RPC curves of three algorithms on $T_{I.}$

Table 2: Feature extraction performance of different algorithms on the test set T_{II}.

Algorithm	F-measure (%)	Average feature extraction time (s)
Method in paper [4]	44	0.35
Method in paper [5]	80.1	0.22
Proposed method	95.6	0.031

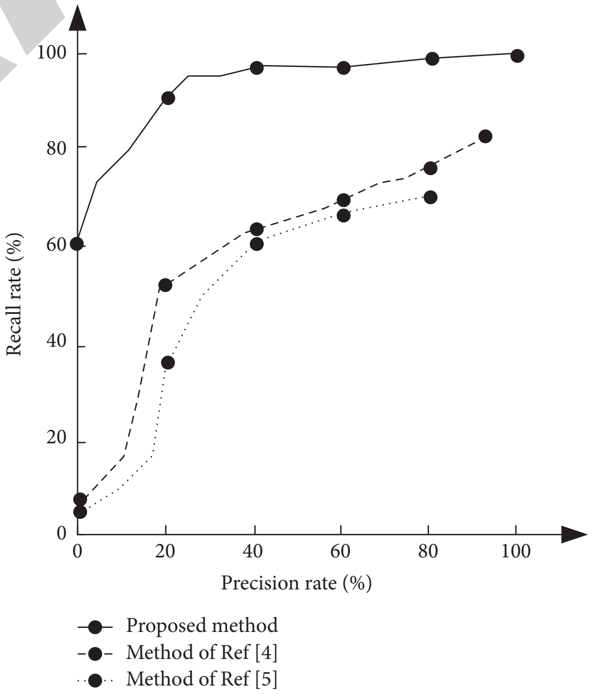


Figure 13: Comparison of RPC algorithms for three algorithms on the test set $T_{II.}$

complete, and there are only weak blur points at the edge of the instrument, which does not affect the overall image restoration effect. In summary, the proposed algorithm can effectively extract the feature points of a super-resolution image with 10% noise, and the acquired local edges features are clear, continuous, and complete.

Based on the obtained local features of the super-resolution image, the proposed algorithm calculates the geometric center of gravity of the closed contour line and the nonclosed contour line to further determine the effective feature points. Firstly, the edge points are polarized, which is convenient for calculating the length and position of each point on the edge point from the geometric center of gravity, searching for local extremum points, and further determining the feature points so that the feature points of the obtained super-resolution image are effective and reliable. The final feature points are obtained on the principle of nonmaximum (minimum) value suppression in the local area.

In Table 1, the equal error measure of the proposed algorithm is 96.3%, which is 14.1% higher than that of the method given in paper [4] and 21.1% higher than that of the method given in paper [5]. The equal error measure indicates the accuracy of feature extraction, which indicates the accuracy of the proposed algorithm to extract the local edge features of the super-resolution image is high; the average feature extraction time of the proposed algorithm is 0.02 s, which saves 0.075 s compared with the method given in paper [4] and 0.036 s compared with the method given in paper [5]. It can be seen that the average time of extracting features by using the proposed algorithm is less. In summary, compared with similar algorithms, the proposed algorithm extracts feature with the shortest time while maintaining the highest accuracy and achieves fast feature extraction with high efficiency and high precision.

In Figure 12, the RPC polyline of the proposed algorithm is located at the top of the line graph, indicating that the recall rate and precision are higher. The initial value of the proposed algorithm is 60%. The recall rate in the previous stage rises linearly and is stable in the later stage, being about 97%, and the maximum precision is 99%; the initial value of the method given in paper [4] and the method given in paper [5] is about 8%, the maximum precision of the method given in paper [4] is about 75%, and the maximum precision of the method given in paper [5] is about 64%. The recall rate and precision of the three algorithms are compared, indicating that the RPC performance of the proposed algorithm is better. In the process of local edge feature extraction of super-resolution images, the proposed algorithm has a high recall rate and high accuracy. Similar to the testing set T_{I}, the feature extraction result obtained by the algorithm on T_{II} has a high recall rate and high precision, which has advantages.

It can be seen from Tables 1 and 2 that in the process of local edge feature extraction of super-resolution images when the proposed algorithm maintains the highest accuracy, the feature extraction time is the shortest. The reason why the algorithm has good RPC performance and F-measure performance is because the algorithm uses the Chamfer matching metric to extract the local edge features of super-resolution images. Chamfer's distance is used to measure the similarity of two-edge graphics and search for similar graphics. The Chamfer distance matching is realized by searching the minimum Chamfer distance of similar graphics. Finally, the local edge features of super-resolution images are extracted based on the local edge feature function of class Chamfer distance. The obtained features are comprehensive and accurate, which provide favourable conditions for extracting the final feature points and prevent feature points from loss.

5. Conclusions

In super-resolution image processing technology, feature extraction algorithms have attracted a great attention as a potential area of interest. The feature point extraction algorithm for the local edge of super-resolution image based on edge point has been proposed. The proposed method considers the case where the edge line is closed. The obtained local edge feature points of the super-resolution image are lone convex points with large curvature on the edge line. The problem proposes a new fast extraction algorithm for extracting local edge features of super-resolution images. The experimental results show that the proposed algorithm can not only detect the points with large curvature on the edges of the image but also locate them accurately for refined extraction of the features. The equal error measure for extracting the local features of the super-resolution image is 96.3%, and the average time taken by the algorithm to produce the results is 0.02 seconds. In the final results, it shows a precision of 98% where our method outperforms the existing approaches, which shows accuracy of 75% and 64%, respectively, during comparative study. The proposed method is of great significance for the recognition of objects in the super-resolution image and for the reconstruction of the three-dimensional surfaces where accurate extraction of features plays a noteworthy role.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that no conflicts of interest exist regarding the publication of this paper.

References

[1] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems," in *Proceedings of the 2014 4th International Conference on Advanced Computing & Communication Technologies*, pp. 5–12, Rohtak, India, February 2014.

[2] J. Guo, L. Liu, W. Song, C. Du, and X. Zhao, "The study of image feature extraction and classification," in *Proceedings of the 2017 International Conference on Progress in Informatics and Computing (PIC)*, pp. 174–178, Nanjing, China, December 2017.

[3] M. C. Popescu and L. M. Sasu, "Feature extraction, feature selection and machine learning for image classification: a case study," in *Proceedings of the 2014 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*, pp. 968–973, Bran, Romania, May 2014.

[4] P. Hou, "A new feature extraction method for medical images integrity verification," in *Proceedings of the 2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, pp. 1589–1593, Chengdu, China, December 2018.

[5] S. Patil and S. R. Patil, "Enhancement of feature extraction in image quality," in *Proceedings of the 2019 3rd International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 490–495, Palladam, India, December 2019.

[6] P. Benagi, S. M. Meena, U. Kulkarni, and S. Shetty, "Feature extraction and classification of heritage image from crowd source," *(ICCTCT)*, in *Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies*, pp. 1–5, Coimbatore, India, March 2018.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9839023, 1 page
https://doi.org/10.1155/2023/9839023

Retraction

Retracted: A Chaotic-Map-Based Password-Authenticated Key Exchange Protocol for Telecare Medicine Information Systems

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Lu and D. Zhao, "A Chaotic-Map-Based Password-Authenticated Key Exchange Protocol for Telecare Medicine Information Systems," *Security and Communication Networks*, vol. 2021, Article ID 7568538, 8 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2021, Article ID 7568538, 8 pages
https://doi.org/10.1155/2021/7568538

Research Article

A Chaotic-Map-Based Password-Authenticated Key Exchange Protocol for Telecare Medicine Information Systems

Yanrong Lu[1] **and Dawei Zhao**[2]

[1] *School of Safety Science and Engineering, Civil Aviation University of China, Tianjin, China*
[2] *Shandong Provincial Key Laboratory of Computer Networks,*
Shandong Computer Science Center (National Supercomputer Center in Jinan) Qilu University of Technology (Shandong Academy of Sciences), Jinan, China

Correspondence should be addressed to Yanrong Lu; yr_lu@cauc.edu.cn and Dawei Zhao; zhaodw@sdas.org

Received 22 July 2021; Revised 23 August 2021; Accepted 30 October 2021; Published 30 November 2021

Academic Editor: Chinmay Chakraborty

Telecare medicine information systems (TMISs) provide e-health services such that patients can access medical resources conveniently and doctors can prescribe treatments rapidly. Authentication is an essential security requirement in TMISs. In particular, the growth of password-based remote patient authenticated key exchange combining extended chaotic maps has enhanced the level of secure communications for TMISs. Recently, Lee suggested an improved random-number-based password-authenticated key exchange (PAKE) using extended chaotic maps and synchronized-clock-based PAKE using extended chaotic maps on Guo and Zhang and Xiao et al.'s PAKE. Unfortunately, we found that the nonce-based scheme of Lee is insecure against known session-specific temporary information and server spoofing attacks. To cope with the aforementioned defects, this study aims to provide a new secure PAKE based on extended chaotic maps with more security functionalities for TMISs. Additionally, we show that the proposed scheme for TMISs provides high security along with low communication cost, computational cost, and a variety of security features.

1. Introduction

At present, the researches on the cloud assisted e-health are more and more in-depth. It facilitates health condition monitoring and improves efficiency for medical resources [1]. As one of the most popular applications of e-health care service, telecare medical information systems (TMISs) provide the medical or healthcare for those patients who are disabled or cannot attend hospital normally [2, 3]. With the openness of wireless environment, the security of TMISs is highlighted. How to authenticate the communication entities and thus securely transmit sensitive medical data related with patients is an urgent problem that needs to be researched and solved.

Key exchange schemes aim at establishing a shared session key between two or more communicating entities. The shared session key is used in securing subsequent communication over an insecure channel. Therefore, the key challenge of designing such a scheme is how to securely and efficiently derive a session key that is only known to the communicated entities. Hitherto, a large number of related authenticated key agreement schemes have been presented with different structures such as pure password schemes, password schemes with smart cards, dynamic schemes, and dynamic schemes with smart cards.

With the extremely studied and widely applied Chebyshev polynomials by the cryptographic research community, various password authenticated key exchange (PAKE) based on chaotic maps and related approaches have been developed recently [4–10]. Kocarev-Tasev [11] presented the first chaotic maps-based public key encryption scheme. Unfortunately, according to the periodicity of cosine function, the scheme of Kocarev-Tasev was demonstrated to be insecure by Bergamo et al. [12]. After that, Xiao et al. [13] suggested an authenticated key agreement scheme using chaotic map. Nevertheless, Alvarez [14] pointed out that the scheme of Xiao et al. could not withstand man-in-the-middle attack. Shortly after, Xiao et al. [15] introduced an enhanced PAKE to prevent the security threats. Guo-Chang [16] raised a smart card based PAKE using chaotic maps. Later, Lin [17] claimed that the Guo-Chang's scheme might easily leak the identity of communicating user by intercepting the transmitted messages. In addition, Lin [17] also pointed that the session key could be derived by an adversary during the communication in the Guo-Chang's scheme [16]. In order to negate these risks, Lin also developed an improved variant without sacrificing the efficiency.

Recently, Guo and Zhang [18] identified that the drawbacks of Xiao et al.'s scheme [13] and found that Xiao et al.'s scheme failed to satisfy the requirements with the contributory nature of key agreement. Subsequently, Guo and Zhang developed their own improved version of the remote user PAKE. Recently, Lee [19] observed that both Xiao et al. [13] and Guo and Zhangs' schemes [18] were unable to free from offline password guessing attack and achieve the session key security. As a counter measure to these sufferings, Lee developed two PAKE; that is, one uses random numbers, while the other uses timestamp. However, this study shows that Lee (the nonce based) fails to resist known session-specific temporary information and server spoofing attacks.

The merits of this paper are as follows.

(i) Our proposed scheme demonstrates that Lee's scheme has several drawbacks once the private information is leaked.

(ii) Our proposed scheme for TMISs withstands an unauthorized patient to deceive the service provided by the telecare medical server.

(iii) Our proposed scheme for TMISs satisfies high security along with a variety of attributes compared with Xiao et al. [13], Guo and Zhang [18], and Lee schemes. Extensive comparisons are conducted with related schemes to verify the performance of our schemes in terms of security and efficiency.

The remainder of this paper is organized as follows. Section 2 introduces preliminary knowledge of some Chebyshev chaotic maps that we use in our system. We describe Lee's scheme in Section 3. In Section 4, we show that Lee's scheme [19] is vulnerable to various attacks. The proposed scheme for TMISs is presented in detail in Section 5, followed by the security analysis in Section 6. In Section 7, we compare the performance and security of our scheme for TMISs with related schemes. Finally, Section 8 concludes the paper.

2. Preliminaries

This section lists the used definitions and defines Chebyshev chaotic maps and corresponding chaotic properties that are used in this paper [11, 12, 20–22].

The Chebyshev polynomial $T_s(x)$: $[-1, 1] \longrightarrow [-1, 1]$ is defined as follows:

$$T_s(x) = 2xT_{s-1}(x) - T_{s-2}(x), \qquad (1)$$

where s be an integer with $s \geq 2$, $x \in [-1, 1]$, $T_0(x) = 1, T_1(x) = x$.

The Chebyshev polynomial satisfies the semigroup property:

$$T_{st}(x) = T_s(T_t(x)) = T_t(T_s(x)). \qquad (2)$$

Chaotic property:

When $a > 1$, Chebyshev polynomial map T_s: $[-1, 1] \longrightarrow [-1, 1]$ of degree s is a chaotic map with invariant density $f^*(x) = 1/(\pi\sqrt{[2]}\,1 - x^2)$ for Lyapunov exponent $\lambda = \ln s > 0$.

Quadratic residue assumption:

If $y = x^2 \text{mod} n$ has a solution, that is, ∃ a square root for y, then y is named as a quadratic residue modulo n. It is computationally unfeasible to derive x such that $y = x^2 \text{mod} n$ under the condition of not knowing the parameters p and q because of the factoring problem $n = pq$ is NP-hard problem.

3. Review of Lee's PAKE

Lee [19] presented two authentication schemes, which are based on nonce and timestamp, respectively. Without loss of generality, we briefly review the nonce based PAKE of Lee, which includes system initialization, authentication and key agreement phases.

3.1. System Initialization.

Step 1: Server B chooses two large primes p and q as its private keys.

Step 2: B calculates $n = pq$.

Step 3: B publishes h_1: $\{0, 1\}^* \longrightarrow \{-\infty, +\infty\}$, h_2: $\{0, 1\}^* \longrightarrow \{0, 1\}^\tau$ and n, where τ is the fixed size.

3.2. Authentication and Key Agreement

Step 1: User A computes $x = h_1(y), X_1 = (y, pw)^2 \text{mod} n, TID_A = ID_A \oplus h_2\ (y, pw)$, and $X_2 = h_2\ (TID_A, (y, pw), T_r(x))$, where r and y are the random numbers. A then sends the message $M_1 = \{TID_A, X_1, X_2, T_r(x)\}$ to B.

Step 2: Once receiving M_1, B retrieves the four solutions $(y_i, pw_i)\,(0 \leq i \leq 3)$ from X_1 by using the Chinese remainder theorem (CRT) and checks whether $h_2(TID_A, (y_i, pw_i), T_r(x)) \overset{?}{=} X_2$. If successful, it means B has gotten the correct y' and pw'. After that, B checks if $pw' \overset{?}{=} pw$. If the equation is true, B derives ID_A by computing $TID_A \oplus h_2(y', pw')$ and computes $x = h_1(y'), sk = h_2(T_r(x), T_s(x), T_s(T_r(x))$ and $Auth_B = h_2(sk, pw, ID_B, ID_A)$. Next, B sends back $M_2 = \{ID_B, T_s(x), Auth_B\}$ to A.

Step 3: When receiving M_2, A computes $sk = h_2(T_r(x), T_s(x), T_r(T_s(x))$ and $Auth_B' = h_2(sk, pw, ID_B, ID_A)$. Next, A checks whether $Auth_B' \overset{?}{=} Auth_B$. If it holds, A computes $Auth_A = h_2(sk, ID_A, ID_B)$ and sends $M_3 = \{Auth_A\}$ to B.

Step 4: After receiving M_3, B verifies whether $h_2(sk, ID_A, ID_B)$ is equal to the received $Auth_A$. If it does not hold, B terminates the session; otherwise, A and B have a common session key $sk = h_2(T_r(x), T_s(x), T_{rs}(x))$.

4. Security Analysis on Lee's Scheme

Lee [19] found some severe security pitfalls in Xiao et al. [13] and Guo and Zhang's schemes [18, 21] and proposed new chaotic-based authenticated key schemes. It is claimed that their new scheme achieves many security attributes while being secure against general threats. In this part, however, we will demonstrate that Lee's nonce based scheme [19] is actually vulnerable to known session-specific temporary information attack [23], which is one of the most important security properties that most of schemes shall attain. In addition, as the result of overlooking the server is a semi-trusted party, this scheme is subject to server spoofing attack.

4.1. Known Session-Specific Temporary Information Attack. Assume the user's session random number y is corrupted by an adversary $\mathbb{U}$. The scheme will suffer the following attack.

Step 1: $\mathbb{U}$ guesses a candidate password pw^* and checks whether $h_2(TID_A, (y, pw^*), T_r(x)) \overset{?}{=} X_2$, where TID_A and $T_r(x)$ are intercepted information through the public channel by $\mathbb{U}$. If the result is true, the correct password has been gotten. Otherwise, $\mathbb{U}$ continues to execute the aforementioned procedure until he succeeds.

Step 2: Once $\mathbb{U}$ successfully owns the user A's password pw. There can be no real defense against attacks from $\mathbb{U}$. First, $\mathbb{U}$ derives ID_A by computing $TID_A \oplus (y, pw)$ and computes $X_1 = (y, pw)^2 \bmod n$. Next, $\mathbb{U}$ sends the counterfeited message $M_1 = \{TID_A, X_1, X_2, T_r^*(h_1(y))\}$ to the server B, where r^* is the random number chosen by $\mathbb{U}$.

Step 3: When the server B receives M_1, it performs the scheme without any detection since all the verification information derived from the user "A." Finally, the server B sends back the message $M_2 = \{ID_B, T_s(h_1(y)), Auth_B\}$ to $\mathbb{U}$ who masquerades as a legal user A, where $Auth_B = h_2(ID_B, ID_A, pw, sk)$ and $sk = h_2(T_r^*(h_1(y)), T_s(h_1(y)), T_s(T_r^*(h_1(y)))$.

Step 4: After receiving the message M_2, $\mathbb{U}$ computes $sk = h_2(T_r^*(h_1(y)), T_s(h_1(y)), T_{r^*}(T_s(h_1(y)))$ and verifies whether the equation $h(ID_B, ID_A, pw, sk)$ is equal to the received $Auth_B$. If the result is correct, $\mathbb{U}$ computes $Auth_A = h_2(ID_B, ID_A, sk)$ and returns $M_3 = \{Auth_A\}$ to the server B.

Step 5: Once receiving the message M_3, server B validates the correctness of the value $Auth_A$. Then, server B accepts the communication request from user "A" and agrees on the session key sk as a "confidential" session key for concealing the following messages. In this way, the subsequent communication messages seem like plain text such that $\mathbb{U}$ could do whatever he wants. This shows that, in Lee's scheme, $\mathbb{U}$ can use the unexpectedly disclosed session random number to successfully complete mutual authentication. This concludes that their scheme lacks strongly the SK-security, which is very essential in the security critical applications.

4.2. Server Spoofing Attack. In Lee's scheme, server B masters the sensitive information pw of user A, which leads to a malicious spoofing attack because the legal but malicious server B could monitor the authentication process of user A and gather information related to user A and thus become an adversary. The malicious server $\mathbb{B}$ can forge the valid request message by performing the following procedures.

Step 1: The malicious server $\mathbb{B}$ can eavesdrop the message $M_1 = \{TID_A, X_1, X_2, T_r^*(h_1(y))\}$ during authentication and key agreement phase corresponding to the legitimate user A. Then, $\mathbb{B}$ generates two random numbers r^*, y^* and calculates $x = h_1(y^*), T_r^*(x), z = (y^*, pw), X_1 = z^2 \bmod n, TID_A = ID_A^* \oplus h_2(z)$ and $X_2 = h_2(TID_A, z, T_r^*(x))$. Next, $\mathbb{B}$ sends an imitative message $M_1 = \{TID_A, X_1, X_2, T_r^*(x)\}$ to server B.

Step 2: When receiving M_1, server B derives (y', pw') from X_1 by using the Chinese remainder theorem (CRT) and examines whether $h_2(TID_A, (y', pw'), T_r^*(x))$. Because the computed result equals the received X_2, B will accept $\mathbb{B}$'s request. Next, B derives ID_A^* by computing $TID_A \oplus h_2(y', pw')$ and checks whether $pw' \overset{?}{=} pw$. If it holds, server B computes $x = h_1(y'), T_s(x), sk = h_2(T_r^*(x), T_s(x), T_s(T_r^*(x))$ and $Auth_B = h(sk, pw, ID_B, ID_A^*)$. At last, B sends the message $M_2 = \{ID_B, T_s(x), Auth_B\}$ to $\mathbb{B}$ who is masquerading as user A.

Step 3: Once receiving M_2, $\mathbb{B}$ computes $h_2(sk, pw, ID_B, ID_A^*)$ and checks it with $Auth_B$, where $sk = h_2(T_r^*(x), T_s(x), T_s(T_r^*(x))$. If they are equivalent, $\mathbb{B}$ computes $Auth_A = h_2(sk, ID_A^*, ID_B)$ and sends back $M_3 = \{Auth_A\}$ to server B.

Step 4: When receiving M_3, server B computes $h_2(sk, ID_A^*, ID_B)$ and compares it with M_3. If they are equal, B authenticates $\mathbb{B}$. In this regard, $\mathbb{B}$ and B share a common session key $sk = h_2(T_r^*(x), T_s(x), T_s(T_r^*(x))$ for securing communication. Therefore, a legal but malicious server can masquerade as a legal user to log into a remote server.

The same flaw can be applied to the timestamp based scheme of Lee. Since they work on the same principle, only nonce-based scheme is analyzed above.

5. The Proposed PAKE Scheme for TMISs

To overcome the security pitfalls found in Lee's scheme, we present efficient and secure PAKE using chaotic maps for TMISs. To achieve the patient anonymity and reduce the computation overhead at the patient's side who may take mobile device, the proposed scheme leverages the encryption function to find a trade-off between the security and the cost. The proposed scheme has the following phases: system initialization phase, patient registration phase (Algorithm 1), and authentication and key agreement phase (Algorithm 2).

5.1. System Initialization.

Step 1: The telecare medical server B chooses two large primes p and q as its private keys.

Step 2: B calculates $n = pq$.

Step 3: B publishes h: $\{0,1\}^* \longrightarrow \{0,1\}^\tau$ and n, where τ is the fixed size.

5.2. Patient Registration

Step 1: Patient A computes $h(pw, ra)$ and sends $\{ID_A, h(pw, ra)\}$ to the telecare medical server B over a private channel, where ra is a random nonce and pw is A's password.

Step 2: When B receives the message, B computes $A_{pw} = E_{K_B}[h(ID_A, h(pw, ra))]$ and stores A_{pw} in its database, where K_B is the secret key of B.

5.3. Authentication and Key Agreement

Step 1: A computes $x = h(ID_A, h(pw, ra))$, $z = (ID_A, r)^2 \bmod n$, $X_1 = h(ID_A, x, r)$ and $X_2 = E_x [T_r(x), z, X_1]$, where r is a random nonce. Next, A transmits the messages $M_1 = \{X_2\}$ to B.

Step 2: Upon receiving the message, B first derives $x = h(ID_A, h(pw, ra))$ by decrypting A_{pw} and then retrieves $[T_r(x), z, X_1]$ by decrypting X_2. Subsequently, B solves z by CRT and verifies whether $h(ID_A', r', x) \overset{?}{=} X_1$. If true, B computes $X_3 = E_{T_r(x)} [T_s(x), SID_B, h(ID_A', r')]$ and sends $M_2 = \{X_3\}$ to A.

Step 3: On receiving the message, A retrieves $[T_s(x), SID_B, h(ID_A', r')]$ by decrypting X_3, computes $h(ID_A, r)$, and checks whether $h(ID_A', r') \overset{?}{=} h(ID_A, r)$. If successful, A computes $sk = h(T_u(x), T_s(x), T_{us}(x))$, $X_4 = E_{T_{r(x)}}[T_{u(x)}]$, and $Auth_A = h(sk, T_u(x), T_s(x))$. Next, A sends back $M_3 = \{Auth_A, X_4\}$ to B.

Step 4: When receiving the message, B retrieves $T_u(x)$ by decrypting X_4 with computed $T_r(x)$. After that, B computes $sk = h(T_s(x), T_u(x), T_s(T_u(x)))$ and checks whether $h(T_s(x), T_u(x), T_s(T_u(x))) \overset{?}{=} Auth_A$. If correct, B computes $Auth_B = h(sk, T_s(x), T_u(x), 00)$ and sends $M_4 = \{Auth_B\}$ to A.

Step 5: A verifies the correctness of the value $Auth_B$. If not, A aborts the session. Otherwise, A and B share a common session key $sk = h(T_s(x), T_u(x), T_{su}(x))))$ with each other.

6. Cryptanalysis of Our Enhancement

In this section, we provide an in-depth analysis on the security features of our enhanced remote user PAKE scheme for TMISs. We will show that the proposed scheme not only provides anonymity and mutual authentication and but could also withstand the aforementioned attacks.

6.1. Full Protection for Patient's Identity. Obviously, the proposed scheme for TMISs provides patient anonymity because patient A's identity ID_A is not transmitted in plaintext via any messages traveling over insecure network. For one thing, ID_A is protected by hash function as a symmetric key only known by the patient A and the corresponding telecare medical server B. The telecare medical server could not know the real identity even if it intends to decrypt the stored value or the legal telecare medical server's private key is embezzled by an illegitimate patient or an illegitimate server to derive the hash value. The real identity ID_A is concealed by the quadratic residue assumption. As we know, the assumption is secure for chosen-plaintext attack and the identity is always a short string which could not be known by the unauthorized third-party unless it is completely learnt. Besides, the random number r is not an uncertain number which is not easily guessed. All in all, the proposed scheme for TMISs can be categorized as one preserving the patient privacy.

6.2. Mutual Authentication Thwarting Man-in-the-Middle Attack. The mutual authentication between correspondents is a basic security features for a remote PAKE. Only on the basis of the trust, two unfamiliar participates, that is, the patient and the server, are able to establish the session key for securing the following communication messages. In the proposed scheme for TMISs, patient A is authenticated by the telecare medical server by verifying the validity of X_1. This verification needs two indispensable conditions. One is the private key of the telecare medical server B to derive the hashed value including the identity ID_A, the password pw, and the random number r_a. Another are the two private values p and q, which are used to retrieve the plain-text identity ID_A and the random number r. The telecare medical server is not able to examine the received message $M_1 = \{X_2\}$ without the knowledge of the two secrets. In other words, not anyone could generate the valid message $\{X_2\}$ unless they know all the private information, such as the identity ID_A, the password pw, and even the random number r_a of the registered patient, only known by the patient itself. Additionally, the message $M_3 = \{Auth_A, X_4\}$ further consolidates the authenticity of patient A since only the real patient knows the value $T_r(x)$, which is employed to compute the authenticated messages. On the other side, following the previously mentioned discussion, only the legitimate telecare medical server retrieves the plain-text identity ID_A, the random number r, and the value $T_r(x)$,

```
Input: $pw, h, p, q, n$
Output: Store $A_{pw}$.
(1) Select $ra$,
(2) Compute $h(pw, ra)$,
(3) $A_{pw} = E_{K_B}[h(ID_A, h(pw, r_a))]$.
```

Algorithm 1: Patient registration.

which are used for checking by patient A. Similarly, message $M_4 = \{Auth_B\}$ is utilized to further confirm the legitimacy of the telecare medical server. According to the previously mentioned analysis, the man-in-the-middle attack is not launched due to lack of personal information. Any forged messages could be detected by the receivers since they have the symmetric key which is unknown by any third party. This confirms that our PAKE scheme for TMISs achieves the property of mutual authentication and thus resists man-in-the middle attack.

6.3. Resistance to Known Session-Specific Temporary Information Attack. From Algorithm 2, patient A and telecare medical server B use $T_r(x)$ to encrypt the session key $SK = h(T_u(x), T_s(x), T_{us}(x))$, where $x = h(ID_A, h(pw, r_a))$. Clearly, even if an adversary gets the temporary information r and r_a, it is incapable of computing $T_u(x)$ without having the knowledge of either ID_A or pw [19]. In this way, the proposed scheme for TMISs overcomes the drawbacks found in Lee's scheme. Moreover, without revealing the identity pw of A to B, B authenticates A through decrypting A's registered message $A_{pw} = E_{K_B}[h(ID_A, h(pw, r_a))]$. Thus, the proposed PAKE scheme for TMISs can withstand this type of attack.

6.4. Perfect Forward Session Key Secrecy. Even if the password pw of the patient A is lost, the session key is still secure since the password is not related with the computed session key. Actually, if the important values $T_u(x)$ and $T_s(x)$ are compromised, an adversary could derive the correct u and s using the approach [12]. However, the adversary has no opportunity to get the two values unless they know user A's private information, such as the identity ID_A, password pw, and the random number r_a or the private keys of the telecare medical server B, such as K_B and two large primes p, q. Unfortunately, the patient anonymity has guaranteed that it is impossible for the adversary to obtain the patient's personal information. As we know, the telecare medical server's private keys are not easily exposed. These features along with the patient anonymity confirm forward secrecy and known-key secrecy capability of our PAKE scheme.

6.5. Resistance to Patient Impersonation Attack. Evidently, the most essential goal of a secure PAKE scheme is to withstand impersonation attack, which means an interception of the transmitted messages from both sides will not lead to the serious threats on the system. In the proposed PAKE scheme for TMISs, no adversaries are able to impersonate patient A by eavesdropping the communication messages, since the secret parameters including A's identity ID_A, password pw, and random number r_a are unknown to the adversary. Additionally, it is computationally infeasible to find ID_A and r from $z = (ID_A, r)^2 \bmod n$ without the knowledge of p and q, where $n = p * q$. Therefore, the proposed PAKE scheme for TMISs provides the resilience against patient impersonation attack.

6.6. Resistance to Telecare Medical Server Spoofing Attack. Suppose an adversary plans to impersonate the telecare medical server B by eavesdropping the communication message: $M_1 = X_2 = E_x[T_r(x), z, X_1]$, where $x = h(ID_A, h(pw, r_a)), z = (r, ID_A)^2 \bmod n$ and $X_1 = h(x, ID_A, r)$. They could not pass the authentication by patient A without knowing the telecare medical server B's secret key K_B. How easy will it be to get hold of patient A's identity ID_A and r without the help of the correct value x? Hence, the proposed PAKE scheme for TMISs can withstand telecare medical server spoofing attack.

6.7. Resistance to Bergamo et al.'s Attack. The implementation of the Bergamo et al.'s attack [12] is based on the following facts: (i) Chebyshev polynomials can be alternatively defined as the cosine function, which leads to the same value due to the periodicity of the cosine function; (ii) $T_u(x), T_s(x)$, and x as the public keys are transmitted in an open channel, which can be intercepted by an adversary. However, in the proposed scheme for TMISs, patient A and the telecare medical server B transmitted the encrypted messages $Auth_A$ and $Auth_B$ over a public channel, where $Auth_A = h(sk, T_u(x), T_s(x))$, $Auth_A = h(sk, T_s(x), 00, T_u(x))$, respectively. Without knowing $T_r(x)$, no adversaries can decrypt the message X_4 and thus they cannot recover $T_u(x)$. Additionally, the value $x = h(ID_A, h(pw, r_a))$ is related with the patient A's sensitive information, and adversaries are incapable of getting such sensitive information. Therefore, the proposed PAKE scheme for TMISs is free from Bergamo et al.'s attack [12].

6.8. Resistance to Replay Attack. With the purposing of free from replay attack, we use a random number r which is only recovered by the telecare medical server S. If an adversary attempts to masquerade A by immediately replaying the previous authentication messages $M_1 = \{X_2\}$ after eavesdropping, the telecare medical server S would obviously refuse the request because the invalid random number r will be detected by checking $h(x, ID_A, r) \overset{?}{=} X_1$. Moreover, the

```
Input: $ID_A$, $pw$, $ra$
Output: true: success; false: failure
  Generate $r$,
  Compute $x = h(ID_A, h(pw.ra)), y = (ID_A, r)$
  $Z = y^2 \text{mod} n$, $X_1 = h(x, ID_A, r), X_2 = E_X[T_r(x), z, X_1]$.
  Transmit $M_1 = \{X_2\}$ to $B$.
  Decrypt $A_{pw} \longrightarrow x = h(ID_A, h(pw, ra)), X_2 \longrightarrow [T_r(x), z, X_1]$.
  Solve $z$ by CRT, determinate $(ID_A', r')$
    if $h(x, ID_A', r') \overset{?}{=} X_1$ then
  Generate $s, X_3 = E_{T_r(x)}[T_s(x), SID_B, h(ID_A, r)]$
    Transmit $X_3$ to Patient $A$.
      Decrypt $X_3 \longrightarrow [T_s(x), SID_B, h(ID_A, r)]$
        if $h(ID_A, r) \overset{?}{=} h(ID_A', r')$ then
        Generate $u$, $sk = h(T_u(x), T_s(x), T_{us}(x))$
        $X_4 = E_{T_{r(x)}}[T_u(x), ID_A]$, $Auth_A = h(sk, T_u(x), T_s(x))$.
        Transmit $Auth_A$ and $X_4$ to $B$.
        Decrypt $X_4 \longrightarrow [T_u(x), ID_A]$.
        $sk = h(T_s(x), T_u(x), T_s(T_u(x)))$.
          if $h(sk, T_s(x), T_u(x)) \overset{?}{=} Auth_A$ then
       $Auth_B = h(sk, T_s(x), T_u(x), 00)$
     Transmit $Auth_B$ to $M_4 = \{Auth_B\}$.
           if $Auth_B \overset{?}{=} h(sk, T_s(x), T_u(x), 00)$then
           return true
        else
      return false
     else
   return false
      else
   return false
     else
   return false
end if
```

Algorithm 2: Authentication and Key agreement.

patient also checks the random number which is sent from the telecare medical server to prevent the replay attack.

7. Security Attributes and Performance Comparison

In the following section, we analyze the security attributes and the computational efficiency of the proposed PAKE scheme for TMISs and compare to Xiao et al. [15], Guo and Zhang [18], and Lee [19] since they are all based on chaotic-maps PAKE schemes. Table 1 shows the security attributes comparison among our presented scheme and other schemes [15, 18, 19]. Compared with other schemes, both Guo and Zhang and Xiao et al.'s schemes cannot achieve user anonymity and perfect forward session key secrecy. Furthermore, both of their schemes cannot withstand patient impersonation attack. In addition, Lee's scheme fails to prevent known session-specific temporary information and server spoofing attacks.

Table 2 lists the computational complexity comparison of our proposed PAKE scheme with other schemes, where T_c denotes the time of executing a Chebyshev polynomial computing; T_h denotes the time of executing a hash operation; T_s denotes the time of executing a symmetric key encryption/decryption; T_{sq} denotes the time of executing a squaring; T_{sr} denotes the time of executing a squaring root solving. According to [15], the execution time for T_s is about 70 times than T_c, and T_s is almost equal to T_h in software. Therefore, our proposed PAKE scheme consumes a slightly higher computation cost than others. We think it is worth slightly sacrificing the efficiency in the hope of guaranteeing a high level security for TMISs.

Table 1: Comparison of security attributes.

	Ours	Lee [19]	Guo and Zhang [18]	Xiao et al. [15]
Provide anonymity	Yes	Yes	No	No
Provide perfect forward session key secrecy	Yes	Yes	No	No
Provide mutual authentication	Yes	Yes	Yes	Yes
Resist man-in-the-middle attack	Yes	Yes	Yes	Yes
Resist replay attack	Yes	Yes	Yes	Yes
Resist known session-specific temporary information attack	Yes	No	—	—
Resist Bergamo et al.'s attack	Yes	Yes	No	Yes
Resist patient impersonation attack	Yes	Yes	No	No
Resist server spoofing attack	Yes	No	Yes	No

Table 2: Comparison of computational cost.

	Ours	Lee [19]	Guo and Zhang [18]	Xiao et al. [15]
User	$3T_c + 7T_h + 3T_s + 1T_{sq}$	$2T_c + 6T_h + 1T_{sq}$	$2T_c + 8T_h$	$2T_c + 2T_h$
Server	$2T_c + 8T_h + 4T_s + 1T_{sr}$	$2T_c + 9T_h + 1T_{sr}$	$2T_c + 10T_h$	$2T_c + 2T_h$
No. of message communications	4	3	6	5

8. Conclusion

In this paper, we first reviewed Lee's scheme and then demonstrated that Lee's scheme is vulnerable to the known session-specific temporary information and server spoofing attacks. With the purpose of remedy of these security loopholes, we presented an improved PAKE scheme using extended chaotic maps for TMISs. We showed that our design is secure and provides more functionalities compared with the related schemes. Performance analysis showed the proposed PAKE scheme for TMISs is secure and efficient. In the future, we will further optimize the proposed scheme regarding security and performance using encryption and machine learning in order to apply to network structure to improve its availability.

Data Availability

The data are included in the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported in part by the National Natural Science Foundation of China under grant 61802276 and the Fundamental Research Funds for the Central Universities of China (no. 3122021027).

References

[1] Z. Yu, C. Y. Zhang, and J. X. Yu, "Research on mathematical model of characteristic curve of surface perception by PVDF array based on Ferguson function," *International Journal of Engineering Systems Modelling and Simulation*, vol. 12, no. 1, p. 17, 2021.

[2] X. Li, F. Wu, M. K. Khan, L. Xu, J. Shen, and M. Jo, "A secure chaotic map-based remote authentication scheme for telecare medicine information systems," *Future Generation Computer Systems*, vol. 84, pp. 149–159, 2017.

[3] A. K. Sutrala, A. K. Das, V. Odelu, M. Wazid, and S. Kumari, "Secure anonymity-preserving password-based user authentication and session key agreement scheme for telecare medicine information systems," *Computer Methods and Programs in Biomedicine*, vol. 135, pp. 167–185, 2016.

[4] S. Deng, Y. Li, and D. Xiao, "Analysis and improvement of a chaos-based hash function construction," *Communications in Nonlinear Science and Numerical Simulation*, vol. 15, no. 5, pp. 1338–1347, 2010.

[5] J.-L. Tsai and N.-W. Lo, "A chaotic map-based anonymous multi-server authenticated key agreement protocol using smart card," *International Journal of Communication Systems*, vol. 28, no. 13, pp. 1955–1963, 2015.

[6] X. Li, J. Niu, S. Kumari, M. K. Khan, J. Liao, and W. Liang, "Design and analysis of a chaotic maps-based three-party authenticated key agreement protocol," *Nonlinear Dynamics*, vol. 80, no. 3, pp. 1209–1220, 2015.

[7] X. Wang, W. Zhang, W. Guo, and J. Zhang, "Secure chaotic system with application to chaotic ciphers," *Information Sciences*, vol. 221, pp. 555–570, 2013.

[8] W.-C. Yau and R. C.-W. Phan, "Cryptanalysis of a chaotic map-based password-authenticated key agreement protocol using smart cards," *Nonlinear Dynamics*, vol. 79, no. 2, pp. 809–821, 2014.

[9] Y. Lu, L. Li, H. Peng, and Y. Yang, "Cryptanalysis and improvement of a chaotic maps-based anonymous authenticated key agreement protocol for multiserver architecture," *Security and Communication Networks*, vol. 9, no. 11, pp. 1321–1330, 2016.

[10] D. Abbasinezhad-Mood, A. Ostad-Sharif, S. M. Mazinani, and M. Nikooghadam, "Provably secure escrow-less Chebyshev chaotic map-based key agreement protocol for vehicle to grid connections with privacy protection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7287–7294, 2020.

[11] L. Kocarev and Z. Tasev, "Public key encryption based on Chebyshev maps," in *Proceedings of the IEEE Symposium on Circuits and Systems*, pp. 28–31, IEEE, Bangkok, Thailand, May 2003.

[12] P. Bergamo, P. D'Arco, A. De Santis, and L. Kocarev, "Security of public-key cryptosystems based on Chebyshev

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9780906, 1 page
https://doi.org/10.1155/2023/9780906

Retraction

Retracted: A Lightweight Proxy Re-Encryption Approach with Certificate-Based and Incremental Cryptography for Fog-Enabled E-Healthcare

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Hassan, D. Shehzad, I. Ullah et al., "A Lightweight Proxy Re-Encryption Approach with Certificate-Based and Incremental Cryptography for Fog-Enabled E-Healthcare," *Security and Communication Networks*, vol. 2021, Article ID 9363824, 17 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2021, Article ID 9363824, 17 pages
https://doi.org/10.1155/2021/9363824

WILEY | Hindawi

Research Article

A Lightweight Proxy Re-Encryption Approach with Certificate-Based and Incremental Cryptography for Fog-Enabled E-Healthcare

Junaid Hassan,[1] Danish Shehzad ,[1] Insaf Ullah ,[2] Fahad Algarni ,[3] Muhammad Umar Aftab ,[1] Muhammad Asghar Khan ,[2] and M. Irfan Uddin [4]

[1]Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad Chiniot-Faisalabad Campus, Chiniot 35400, Pakistan
[2]Hamdard Institute of Engineering & Technology, Hamdard University, Islamabad 44000, Pakistan
[3]College of Computing and Information Technology, The University of Bisha, Bisha, Saudi Arabia
[4]Institute of Computing, Kohat University of Science and Technology, Kohat 26000, Pakistan

Correspondence should be addressed to Insaf Ullah; insafktk@gmail.com and Muhammad Umar Aftab; ms.umaraftab@yahoo.com

Received 27 July 2021; Revised 1 October 2021; Accepted 9 October 2021; Published 19 November 2021

Academic Editor: Chinmay Chakraborty

Cloud computing aims to provide reliable, customized, and quality of service (QoS) guaranteed dynamic computing environments for end-users. However, there are applications such as e-health and emergency response monitoring that require quick response and low latency. Delays caused by transferring data over the cloud can seriously affect the performance and reliability of real-time applications. Before outsourcing e-health care data to the cloud, the user needs to perform encryption on these sensitive data to ensure its confidentiality. Conventionally, any modification to the user data requires encrypting the entire data and calculating the hash of the data from scratch. This data modification mechanism increases communication and computation costs over the cloud. The distributed environment of fog computing is used to overcome the limitations of cloud computing. This paper proposed a certificate-based incremental proxy re-encryption scheme (CB-PReS) for e-health data sharing in fog computing. The proposed scheme improves the file modification operations, i.e., updation, deletion, and insertion. The proposed scheme is tested on the iFogSim simulator. The iFogSim simulator facilitates the development of models for fog and IoT environments, and it also measures the impact of resource management techniques regarding network congestion and latency. Experiments depict that the proposed scheme is better than the existing schemes based on expensive bilinear pairing and elliptic curve techniques. The proposed scheme shows significant improvement in key generation and file modification time.

1. Introduction

Health monitoring has become a leading issue in modern paradigms all over the world. According to the United Nations report [1], by 2050, the world's 22% population will consist of vulnerable people who can be victims of various diseases. Nowadays, many devices are being developed to monitor vital signs such as human blood pressure, glucose level, heartbeat, and oxygen level. Readings are collected from the medical devices and then shared with the authorized health teams. These medical devices generate a large amount of data, and managing those data using traditional hardware or software has become a huge challenge [2, 3]. Since the advent of COVID-19, medical big data (MBD) has become important for the effective health monitoring systems. MBD by now has become irresistible because of its diversity and volume [4–7]. People suffer from many diseases in the world but Parkinson's disease (PD) is a serious neurodegenerative disorder. Some recent studies have suggested different techniques for diagnosing Parkinson's disease [8, 9]. These studies have improved the automated Parkinson's disease detection using neural network techniques.

Over time, in healthcare, electronic health records (EHR) have replaced the paper record system, so that data can be handled and maintained efficiently. It further improves the availability, sharing, and cost of data maintenance. A hybrid machine learning framework is proposed to predict mortality in paralytic ileus patients using (EHRs) [10]. Previously, if a person wanted to be treated by another doctor, he had to bring his report to the doctor in paper format. The EHR system has saved people from this hassle because it is abysmal to save paper reports for a long time. Now, the patient record is stored on a centralized cloud database from which all the doctors can check the patient's medical reports and history. Therefore, the patient no longer needs to bring his reports to the doctor in paper form. All the users who register to a healthcare application can share their medical reports with all the healthcare authorities, i.e., doctors, physicians, pharmacists, and laboratory technicians.

The widespread adoption of IoT devices in business and healthcare has raised serious concerns about evaluating IoT architectures regarding data communication, storage, security, and processing. In the healthcare field, a large number of IoT devices and sensors are currently being used producing a vast amount of data [11]. Therefore, we required an infrastructure that can process, store, and analyze this large amount of data.

Across the world, cloud infrastructure is used to process a large variety of data. Currently, cloud infrastructure is the only feasible option for managing and communicating between IoT devices in the healthcare field [12, 13]. All computational heavy and battery drain works are being offloaded from IoT devices to cloud infrastructure to reduce the load on IoT devices, which increases the battery life of IoT devices [14].

Healthcare IoT devices require minimum communication latency to gain expected performance. On the other hand, healthcare devices produced a massive amount of data. This large amount of data leads to high data traffic that causes network congestion and further delays. Healthcare-sensitive data becomes inadequate and meaningless for end-users due to large hop counts and large data transmission between IoT devices and cloud servers; as a result, we face two-way delays. Real-time data are required for time-sensitive healthcare applications. End users and healthcare IoT devices expect minimal latency delays, but cloud servers cannot meet these expectations. All kinds of delays, such as network delays, connection delays, and computational delays, need to be minimized during data transformation among healthcare IoT devices. The fog computing paradigm has emerged to address the challenges mentioned earlier. Fog computing has a fog node layer between IoT devices and cloud servers, as shown in Figure 1. Distributed fog devices can also contribute to cloud scalability by reducing centralized communication and processing. Compared to the cloud, fog nodes in the edge network are much closer to end devices and have lower latency [15]. The 20% percent user response time is reduced in fog computing, and 90% of data traffic between cloud and end devices is reduced. The fog computing can minimize response latency for real-time applications by up to 50% [16]. While fog computing can serve IoT devices and applications more efficiently, data protection remains a critical issue in fog computing [17].

Risk problems become even more critical as IoT devices share sensitive data because they are directly connected to the Internet [18]. Many research articles on fog computing and the Internet of things have focused on infrastructure and application development issues without providing adequate support for privacy and security protection mechanisms [19]. However, more complex and intricate security mechanisms cannot be performed on resource-constrained IoT devices, which shorten the device's battery life and lead to high computational costs and significant delays in real-time applications that require an immediate response [20]. We avoid this by offloading these complex security tasks to fog rather than resource-limited IoT devices. When the user's private data are stored on a fog node, they have no physical control over their data and face numerous privacy and security attacks. We cannot consider fog nodes as reliable storage. Fog computing can perform some security operations while transmitting and retrieving data to address IoT security and privacy concerns. We can achieve data privacy and confidentiality by encrypting our data. But encryption is a very complex and heavy function that requires more computation time and power. Therefore, we offload these cryptographic functions on fog nodes to achieve efficiency. However, fog nodes must also provide convenient, manageable, and easy communication and accurate access control.

Some incremental cryptographic schemes for the limited resource devices are proposed in [21]. Because there is a lot of literature out there that is useful for data sharing. Still, our best research shows that certificate-based incremental proxy re-encryption is the most important and secure scheme. This paper proposes a certificate-based incremental proxy re-encryption scheme (CB-PReS) for e-healthcare data sharing in fog computing. The key escrow problem is also eliminated in certificate-based proxy re-encryption. In contrast, a key escrow problem is present in the identity-based proxy re-encryption scheme [22]. The efficiency and security hardness of certificate-based proxy re-encryption and identity-based proxy re-encryption scheme (IB-PRE) is based on the standard cryptosystems such as bilinear pairing (BP), elliptic curve (EC), and Rivest, Shamir, and Adleman (RSA). The 1024-bit key is used in RSA, while 160-bit key is used in ECC. The experimental results show that the bilinear pairing is 13.65 ms worse than that of Rivest, Shamir, and Adleman (RSA) as well as 13.93 ms worse than E.C. [23], and Rivest, Shamir, and Adleman (RSA) is 14.42 ms worse than the hyperelliptic curve [24]. The proposed scheme uses a hyperelliptic curve (HEC), and the 80-bit key is used in HEC to provide the same level of security and low communication and computational cost.

1.1. Motivation.

There are some serious concerns about the security and privacy of IoT devices. Given these concerns, security measures need to be taken immediately. IoT devices are very resource-constrained devices, and the implementation of resource-intensive and complex security tasks on these devices is not feasible. This usually shortens the device's battery life and leads to higher computational costs

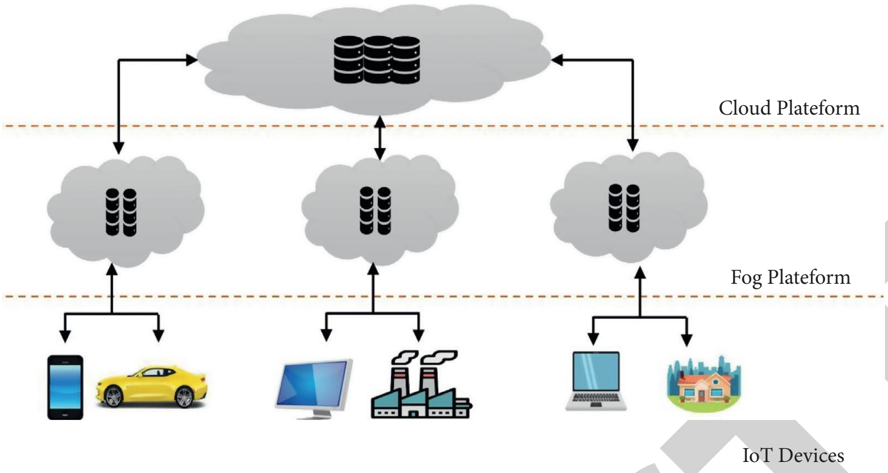


Figure 1: Cloud, fog, and end-device architecture.

and significant delays in real-time applications that require immediate response [20]. Therefore, we can avoid these problems by offloading these complex security tasks to the fog instead of processing them on resource-constrained IoT devices. The communication and computational cost of the traditional crypto system is very high because of the use of traditional cryptographic functions such as RSA that uses a 1024-bit key size. The bilinear pairing scheme is 13.65 times worse than RSA, and RSA is 14.42 times worse than the hyperelliptic curve [23]. When we need to share encrypted data with multiple participants, we can achieve significant performance by using the proxy re-encryption scheme. In some cases, when we update the EHR data, we need to calculate the hash value from scratch to reflect the updation. If we want a minor update to a large amount of EHR data, then the entire hash must be recalculated from scratch, which is not a good thing in practice. Therefore, there is a need to implement incremental cryptography that reduces the overhead of hash value recalculation from scratch.

1.2. Contribution. The contribution of our proposed scheme is listed in the following steps:

(i) We have proposed a certificate-based incremental proxy re-encryption scheme for E-healthcare data sharing on fog computing, which deals with the problem of overhead and delay of previously proposed PRE schemes

(ii) Our scheme reduces the commutation cost and communication overhead of the resource-constraint IoT devices

(iii) Our scheme is based on the hyperelliptic curve, which uses the 80-bit key instead of elliptical curves, and bilinear pairing, which uses a 160-bit key and 1024-bit key, respectively

(iv) Our scheme provides block base data modification by using the concept of incremental cryptography

(v) Our scheme also fixes the key escrow problem by using the certificate-based proxy re-encryption

(vi) We have also provided a security model and security analysis of our scheme

(vii) Our scheme provides some security services such as integrity, confidentiality, unforgeability, and anti-replay attack, respectively

2. Related Work

In this part, we will review some of the publications that are related to our proposed work. In [25], the integrity-enforcement scheme is proposed that takes advantage of trusted computing and incremental cryptographic primitive and ensures the integrity of electronic health records for mobile device users stored in the cloud computing database. However, this scheme ignores data confidentiality. The proxy re-encryption (PRE) scheme is proposed by Blaze et al. in 1998 [26], various PRE schemes have been introduced in the literature so far. Proxy re-encryption scheme may be divided into two different classes, namely, bidirectional and unidirectional. In the unidirectional PRE schemes, there are further different types of PRE schemes, i.e., (1) identity-based PRE, (2) conditional PRE, (3) attribute-based PRE, and (4) time-based PRE. In the bidirectional scheme, there are also two types of PRE schemes, i.e., (1) threshold-based PRE and (2) type-based PRE) [27]. The re-encryption scheme can provide different kinds of features, such as key optimality, collusion resistance, proxy invisibility, non-transferability, noninteractivity, and unidirectionality, and this has been debated extensively in the literature [28]. Based on a set of these features, advanced proxy re-encryption schemes with advanced features were developed. Based on the presence of these features, the proxy re-encryption schemes are evaluated.

Fine-grain access control is provided on the user's private data in the attribution-based PRE (AB-PRE) [29, 30]. In AB-PRE, user A's ciphertext is transformed under some certain set of attributes into user B's ciphertext under some other set of attributes. Conditional PRE (C-PRE) is proposed in [31]. In the C-PRE scheme, the ciphertext is only transformed into another ciphertext by the proxy when it fulfils certain conditions. In the paper [32], the author defines another type of conditional PRE in which a ciphertext is transformed into another ciphertext by a proxy that has been received from a specific sender. Identity-based encryption (IB-PRE) is defined by the author [33] in which the proxy under Alice identity transforms the ciphertext into another ciphertext under Bob's identity. Later another type of identity-based encryption is defined by the author [34] without random oracles. An extended type of IB-PRE is defined in [35], which facilitates the conditional re-encryption features. This conditional re-encryption feature protects the user's data from conditional attacks and also from identity chosen-ciphertext attacks. First time collusion-resistant unidirectional IB-PRE scheme is proposed in [36] which is secure from quantum attacks. The author in [37] proposed a type-based proxy re-encryption scheme in which ciphertexts of every delegator are associated with a specific type, and proxy transformed the ciphertext only when the public key of a delegate has the same type and through which the owner of the data gains proper delegation control. Another version of proxy re-encryption with certificate-based encryption is suggested in [38], which provides resistance to chosen-ciphertext attacks. Keyword search-based PRE scheme is proposed in [39], in which a proxy transforms the ciphertext if the ciphertext contains some keywords that match with specific information about the re-encryption key. In the paper [40], the author proposed a group-based PRE scheme in which a ciphertext is transformed by a proxy to be broadcast to a group. Therefore, all users belonging to the group can decrypt this ciphertext. Another scheme, proposed in [41], relies on conditional broadcast PRE. In this scheme, users are dynamically added to a group at runtime, and there is no need to change the public key for encryption every time.

However, many existing schemes for secure data sharing and communication in cloud computing have limitations, i.e., delays between user requests and responses from the cloud due to a drastic increase in data volume. A large number of users are involved in outsourcing and cryptographic operations [42, 43]. The purpose of fog computing is to solve cloud computing-related problems by offloading expensive computational tasks to the network's edge. Fog computing is also used to perform costly computational tasks that have reduced the computational overhead required on resource-constrained devices. However, some of the studies only focus on the privacy and security of fog computing [44]. Security issues and threats to fog computing are discussed in [16]. The PRE schemes proposed in the fog computing literature are the same ones that have been addressed in the cloud computing literature. Ciphertext-policy attribute-base-encryption (CP-ABE) is proposed in the literature [42, 45, 46] that provides secure access control to data encrypted in fog computing, which allows the data owner to define access policy to one of the features that the user needs to decrypt the ciphertext. Privacy-preserving proxy re-encryption scheme for access control is proposed by [47] which secures against chosen-plaintext attack (IND-CP-CPA). An improved version of proxy re-encryption scheme is proposed [48] in which at every hop ciphertext is transformed into a different ciphertext. We can secure the sensitive data with the usage of secure technologies mentioned in [49] such as access control, steganography, watermarking, and cryptography.

The main disadvantage of using ID-based encryption, attribute-based encryption, and CP-ABE schemes in fog computing is the expensive computational cost of the decryption, which includes the complexity of the policy as well as various pairing operations [42, 46]. The first improved scheme for Internet of Vehicles is proposed [50] in which all the vehicles communicate with each other securely. Dynamic and simultaneous data communication between IoT devices occurs in fog computing. Due to the limited resources of fog computing and IoT networks, public-key encryption and traditional key management practices to secure communications between connected devices are incompatible with fog computing. Traditional encryption mechanisms require extensive computing resources for most objects and do not meet the operational requirements of the Internet of Things in real time [43].

In [44], the authors proposed a scheme that is based on the random oracle model and uses bilinear paring cryptography. In 2013, the authors [51] proposed a random oracle-based CCA-secure proxy re-encryption scheme. In 2018, Bhatia et al. [45] proposed a certificateless proxy re-encryption scheme which is better that the scheme in [44] in computational cost because it uses elliptic curve cryptographic techniques. Another CL-PRE scheme is proposed by Xu and Chang [46] in 2012. This scheme is based on the key management and encryption-based access control for the data distributed using cloud infrastructure. Another single hope certificateless proxy re-encryption and CCA-secure unidirectional scheme is proposed in [52]. The first pairing-free certificateless proxy re-encryption scheme is proposed in 2014 by Lee and Han [53]. In 2015, another CL-PRE scheme for data distribution in cloud is proposed by Qin et al. [54]. This scheme does not provide any kind of security analysis.

Distributed secure accessibility in mobile cloud computing is proposed in [55]. Some incremental cryptographic schemes, such as sharing-based scheme (ShS), coding-based scheme (CoS), and encryption-based scheme (EnS) [56], are proposed in [21] for the limited resource devices. To ensure data security, the mobile user enters a password that is converted into a key, and then, we encrypt the data through this key. More mobile resources are used in the process of encrypting and uploading complex and heavy tasks. The incremental proxy re-encryption scheme (I-PReS) is proposed in [56]. I-PReS is a block-based data sharing scheme [57]. This I-PReS improve the file modification operations that provide data integrity and confidentiality by using advanced cryptographic functions.

The industry did not accept the incremental hash functions due to the following flaws. Firstly, a certain level of security is achieved through the use of a large number of expensive pairing operations by using known incremental hash functions [58, 59] (for example, 2^{128} or 2^{256}), which can degrade their performance with respect to hash function. Secondly, hash value size is disproportionate to the incremental hash functions security level, which is calculated in several thousand bits, unlike SHA_1 producing 160-bit message digest output [60], SHA_2 producing 512-bit message digest output [61], and SHA_3 producing 512-bit message digest output [62]. Incremental hashing is proposed in [63], which requires all hash values of intermediate nodes to be stored. Recently, the authors of [64] proposed a CL-IPRE scheme for the healthcare system based on the elliptic curve improved version of [56]. In the CL-IPRE scheme, the author compares the results of the block modification with the previous scheme and uses an elliptical curve algorithm instead of bilinear pairing for key generation.

3. Materials and Methods

3.1. Preliminaries. Hyperelliptic curve (HEC) is a family of public-key cryptosystems, and its structure is based on algebraic curves. That is why, it is also called a special class of algebraic curves and can also be seen as a generalized form of elliptic curve cryptography (ECC) [65]. HEC differs from ECC in terms of obtaining curve points from the group as shown in Figure 2. The additive abelian group is obtained from the devisor and computed by the HEC. We use a divisor class group of the curve, and there is no group law on the points of the HEC. HEC can implement all major operations of the public-key cryptosystem, such as signature, encryption, decryption, and key exchange. It is also called a successor of the RSA cryptosystem because HEC has the same security level as RSA and uses a smaller signature and key. It also provides a fast signature and fast key generation.

The hyperelliptic curve is defined over a finite field Fq (where q is a prime and $q > 3$). We can also denote the field Fq as $F2m$ (where the $q = 2m$); this means that field size is $q \times q$, and it is a square matrix, and curve points are limited to integer coordinates. When we perform any algebraic operation such as addition or multiplication, then as a result, we get another point on the curve within the field. The genus of the curve over the Fq field is denoted by "g," and the curve of genus one is called elliptic curve with the field Fq having the following values $|Fq|g.\ \log2\ q \approx 2^{160}$ and the genus two curve is called the hyperelliptic curve with the field Fq having the following values $|Fq|g.\ \log2\ q \approx 2^{80}$.

HEC is a special type of projective and nonsingular curve. Hyperelliptic curve over the field Fq with the points $(U,\ V) \in Fq$ satisfies the following equation:

$$V^2 + h(U)V = f(U), \tag{1}$$

where f and h both are polynomial in the field Fq with deg $(f) = 2g + 1$ and deg $(h) \leq g$. It also satisfies both equation (1) and partial derivative equation $h\prime(U) = 0$ and $h\prime(U)V + f'(U) = 0$.

3.2. Complexity Assumptions. In this section, we have made some assumption as follows:

(i) We can denote the field Fq as F_{2m} (where $q = 2m$); this means that field size is $q \text{ x } q$, and it is a square matrix.

(ii) D is denoted as a divisor of the curve, and it is a formal sum of the points $P \in \text{HEC}$ as shown in the following equation:

$$D = \sum p_{\epsilon\, HEC}\, n_i p, \tag{2}$$

where $n_{\text{i}} \in Fq$.

Definition 1. We have $\partial,\ \varphi\, \epsilon \{1, 2, 3, 4, \ldots.., q{-}1\}$, and $\partial,\ \varphi$ picks a random value from the given range, and then we calculate the value of N using the following equation:

$$N = \partial.\, D, \tag{3}$$

where D is a divisor from the Jacobian group.

After that, we find the value of Λ from the following equation:

$$\Lambda = \varphi.\, D, \tag{4}$$

where D is a divisor from the Jacobian group.

The probability of computing the values of ∂ and φ from equations (3) and (4) is negligible due to the Hec–Deffie–Hellman problem (HEC-DHP).

Definition 2. We have $\acute{E}\, \epsilon \{1, 2, 3, 4, \ldots.., q{-}1\}$, and $\acute{E}$ picks a random value from the given range, and then we calculate the value of N from the following equation:

$$N = \acute{E}.\, D, \tag{5}$$

where D is a divisor from the Jacobian group.

The probability of computing the values of $\acute{E}$ from equation (5) is negligible due to the Hec-discrete-logarithm problem (HEC-DLP).

3.3. Syntax of Certificate-Based Incremental Proxy Re-Encryption Scheme. Our proposed scheme is an extension of Khan et al. [56] and Bhatia et al. [64]. Our certificate-based incremental proxy re-encryption scheme consists of seven phases that are (system setup, public and private key generation, certificate generation, proxy re-encryption key generation, encryption, proxy re-encryption, and decryption). The basic symbols which are used in the construction of the proposed scheme are given in Table 1. All these phases are explained as follows.

3.3.1. Setup. In this phase, the certificate authority sets public parameters ψ, then randomly selects a master secret key δ, and computes a master public key MP_k.

3.3.2. Public and Private Key Generation. In this phase, each participant (sender, receiver, and certificate authority) first randomly selects three numbers such as α, β, and γ and calculates $X = \alpha + \beta$, $\xi = \alpha.\ (\gamma\text{-}\alpha)$ and $\eta = \gamma.\ (\beta + \gamma)$. After that,

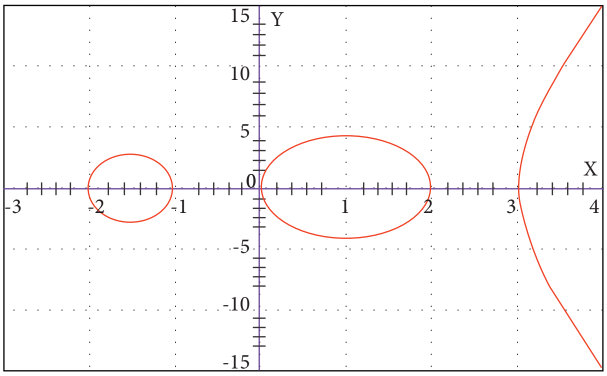


Figure 2: Hyperelliptic curve with genus = 2 [65].

each participant further computes public and private key pair (P_U, K_U).

3.3.3. Certificate Generation. In this phase, CA takes the participant identity ID_U, participant public key P_U, and the public parameters ψ as input and generates a certificate for each participant (sender and receiver) and sends to each participant.

3.3.4. Re-Encryption Key Generation. This algorithm is executed by the sender to generate re-encryption key such as $R_{A \longrightarrow B} = K^b/K_a$ and send to the resident fog proxy server without revealing any secret information about any user.

3.3.5. Encryption. In this phase, the user outsources a file F on the fog node. The user divides the file F into z blocks and encrypts each block using his public key P_U. The user chooses a random number $\chi \in \{1, 2, \ldots\ldots, q\text{-}1\}$ to encrypt the file with his public key P_U. To achieve integrity, the user applies the SHA-3 algorithm on each block of the FILE such as $\text{MAC}_k = \text{HSHA-3}\ (B_k)$, where $1 \leq k \leq z$. Then again, the hash function is applied to the concatenated hash values to get a final single hash value that verifies the integrity of the file as $\text{MAC}_{\text{final}} = \text{HSHA-3} \parallel (\text{MAC}_k)$, where $1 \leq k \leq z$.

3.3.6. Proxy Re-Encryption. In this phase, the proxy server transformed the first level ciphertext of the sender ($C_A \in C_{\text{sender}}$) to the second level ciphertext of the receiver ($C_B \in C_{\text{receiver}}$). Proxy checks the access policy control list, if user B has access rights to the uploaded file, and then the proxy server re-encrypts that file and allows user B to download it.

3.3.7. Decryption. In this phase, the receiver downloads the final hash value ($\text{MAC}_{\text{final}}$), second-level ciphertext (C, C_B), and total numbers of blocks 'z' and decrypts each block by using his private key K_B. After decrypting all the blocks, the receiver concatenates all the blocks to get the original file and checks the integrity of the FILE by calculating the hash of the FILE and matching the calculated hash with $\text{MAC}_{\text{final}}$.

4. Construction of the Proposed Algorithm

In this paper, we have proposed a certificate-based incremental proxy re-encryption data sharing scheme for fog computing. Our scheme deals with the computation and communication problem of previous PRE schemes due to the use of expensive cryptographic functions and cloud computing. In our scheme, to reduce the overhead of resource constraints IoT devices, complex and resource-intensive cryptographic functions are offloaded on fog nodes. We show that our scheme has less communication cost as compared to previous PRE schemes that are necessary for real-time devices. Therefore, each fog node is considered as a proxy node, and all the resource-intensive tasks involved in the process of re-encryption are performed on fog nodes. The structure of our proposed scheme is shown in Figure 3.

Our proposed certificate-based incremental proxy re-encryption data sharing scheme for fog computing is in [64].

4.1. System Setup. Certificate authority (CA) runs this algorithm, and it chooses a security parameter E and hyperelliptic curve (HEC) over a finite field Fq of order q. CA selects an integer $\wp$ as a divisor that is the generator point of order q on the curve. Then, CA randomly selects a master secret key $\delta \in \{1, 2, \ldots\ldots., q\text{-}1\}$ and further calculates the master public key as $\text{MP}_k = \delta.\wp$. Finally, CA announces some public parameters for encryption, decryption, and proxy re-encryption such as $\psi = \{\text{Cert}_U, Fq, \text{ID}_U, \wp, \text{MP}_k, h1, h2, h3\}$.

4.2. Public and Private Key Generation. In this phase, each participant (sender, receiver, and certificate authority) executes this algorithm, the participant first randomly selects three numbers such as $\alpha \in \{1, 2, \ldots\ldots., q-1\}$, $\beta \in \{1, 2, \ldots, q-1\}$, and $\gamma \in \{1, 2, \ldots, q-1\}$, and then calculates $X = \alpha + \beta$, $\xi = \alpha.(\gamma - \alpha)$, and $\eta = \gamma.(\beta + \gamma)$. The participant with identity IDU computes the public and private key pair (P_U, K_U). Each participant computes its private key such as $K_U = \beta.(\xi - \eta) + \text{MP}_k$, after computing private key each participant further computes its public key such as $P_U = K_U.\ \wp$. The public key of each participant is publicly announced.

4.3. Certificate Generation. In this phase, CA takes the participant identity ID_U, participant public key P_U, and the public parameters ψ as an input (ID_U, P_U, ψ) and generates a certificate (Cert_U) for each participant (sender and receiver) and sends to each participant. CA generates the certificate through the following process: first, CA takes participant public key P_U and calculates hash of the P_U such as $H_{PU} = h1\ (P_U \| \text{ID}_U) + \text{MP}_k$ after computing the hash of P_U, and CA digitally signs H_{PU} with its private key such as $\text{S} = \text{K}_{\text{CertA}}\ (H_{PU})$.

4.4. Re-Encryption Key Generation. If user A wants to share data with user B, then user A runs this algorithm to generate a re-encryption key such as $R_{A \longrightarrow B} = K^b/K_a$ using a two-party secure integer division algorithm [65] and sends it to the

Table 1: Notations of the proposed scheme.

Serial no.	Notation	Description
1	ID_U	Unique identity of user
2	E	Security parameter
3	CA	Certificate authority
4	Q	It is a large prime number
5	δ	Generator point
6	δ	Master secret key
7	MP_k	Master public key
8	Ψ	Set of public parameters
9	Cert_U	User certificate
10	α, β, γ	Random values
11	S	Digital signature
12	$H_{P\ U}$	Public key hash
13	K_U	Private key
14	P_U	Public key
15	$R_{A\longrightarrow B}$	Re-encryption key
16	HER	Electronic health record
17	Z	Total number of blocks of a file
18	B_k	k_{th} block of the FILE
19	D_j	Size of the j_{th} block
20	FS	Total size of the FILE
21	C_k	Ciphertext of k_{th} block
22	C	Ciphertext of entire file
23	C_A	A number χ PK_A
24	C_B	A number χ PK_B
25	MAC_k	MAC value of B_k block
26	$\text{MAC}_{\text{final}}$	Final mac value of all z blocks
27	B_{update}	Ciphertext of updated block
28	C_{update}	Block that user want to update
29	$h1$	Hash function
30	K_{CertA}	Certificate authority private key

resident fog proxy server without revealing any secret information about the K_a and K_b.

4.5. Encryption. In this phase, when a user wants to outsource a file F on the fog node, first, he divides the file F into z blocks, and each block has a constant size of d bits except the last block. To achieve confidentiality, the user encrypts each block using his public key $P_{U.}$ The following condition should be satisfied while dividing the file F into z blocks as follows:

$$\text{FILE} = \|_{k=1}^{z} (B_k), \tag{6}$$

$$D_j = \left[\frac{FS}{Z}\right], \tag{7}$$

where $1 \leq j \leq z–1$ and

$$D_z = FS - \left(\left[\frac{FS}{Z}\right] * z - 1\right), \tag{8}$$

where B_k represents the k th block of the file, D_j represents the size of the jth block of the file, F_S denotes the total size of the file, $[FS/Z]$ represents the mathematical floor function that removes the fraction part, and D_z denotes the size of the last block of file. Before outsourcing file on the fog node, the user encrypts all the file blocks with his public key P_U to achieve confidentiality. The user chooses a random number $\chi \in \{1, 2, \ldots, q–1\}$ to encrypt the FILE with his public key P_U. Equations (9) and (10) depict the encryption process.

$$\begin{aligned} C_A &= \chi.P_A, \\ C_K &= B_K + \chi.\wp, \end{aligned} \tag{9}$$

$$C = \|_{k=1}^{z} (C_k), \tag{10}$$

where $1 \leq k \leq z$.

To achieve integrity, the user applies the recently proposed algorithm SHA-3 on each block of the file. Afterwards, hash values of all the blocks are concatenated together. Then again, the hash function is applied to the concatenated hash values to get a final single hash value that verifies the file integrity as follows.

$$\text{MAC}_k = H_{\text{SHA-3}}(B_k), \tag{11}$$

where $1 \leq k \leq z$, and

$$\text{MAC}_{\text{final}} = H_{\text{SHA-3}}\|_{k=1}^{z} (\text{MAC}_k), \tag{12}$$

where $1 \leq k \leq z$.

The hash value of each block (MAC_k), the encrypted file (C, C_A), total number of blocks "z," and the final hash value ($\text{MAC}_{\text{final}}$) are stored on the fog node. Only the total number of blocks "z" and the file's name are saved on the local storage. The user saves the local information (z, filename) for every file.

4.6. Proxy Re-Encryption. User B requests the proxy server for the re-encryption process to get the file uploaded on the fog node by user A. In the re-encryption process, the proxy server transforms the first level ciphertext of user A ($C_A \in$ C1) to the second level ciphertext of user B ($C_B \in$ C2) and also chooses a fresh nonce. The proxy first checks the access control list if user B has access rights to the uploaded file, then the proxy server re-encrypts that file and attached a Nonce value with the file as depicted in equations (13) and (14), and then allows user B to download it.

$$\begin{aligned} C_B &= C_A R_{A\longrightarrow B} = \chi.K_a.\,\wp\left(\frac{K_b}{K_a}\right) = \chi.\,K_b.\,\wp, \\ C_k &= B_k + \chi.\wp, \end{aligned} \tag{13}$$

$$C = \|_{k=1}^{z} (Ck\|\text{Nonce}), \tag{14}$$

where $1 \leq k \leq z$.

The proxy server transfers the final hash value ($\text{MAC}_{\text{final}}$), second-level ciphertext (C, C_B), and the total numbers of blocks "z" to user B for checking the integrity of the file and decryption.

4.7. Decryption. User B downloads the final hash value ($\text{MAC}_{\text{final}}$), second-level ciphertext (C, C_B), and the total numbers of blocks "z." Then, user B decrypts the file by using C_B and his private key K_B as shown in the following equation:

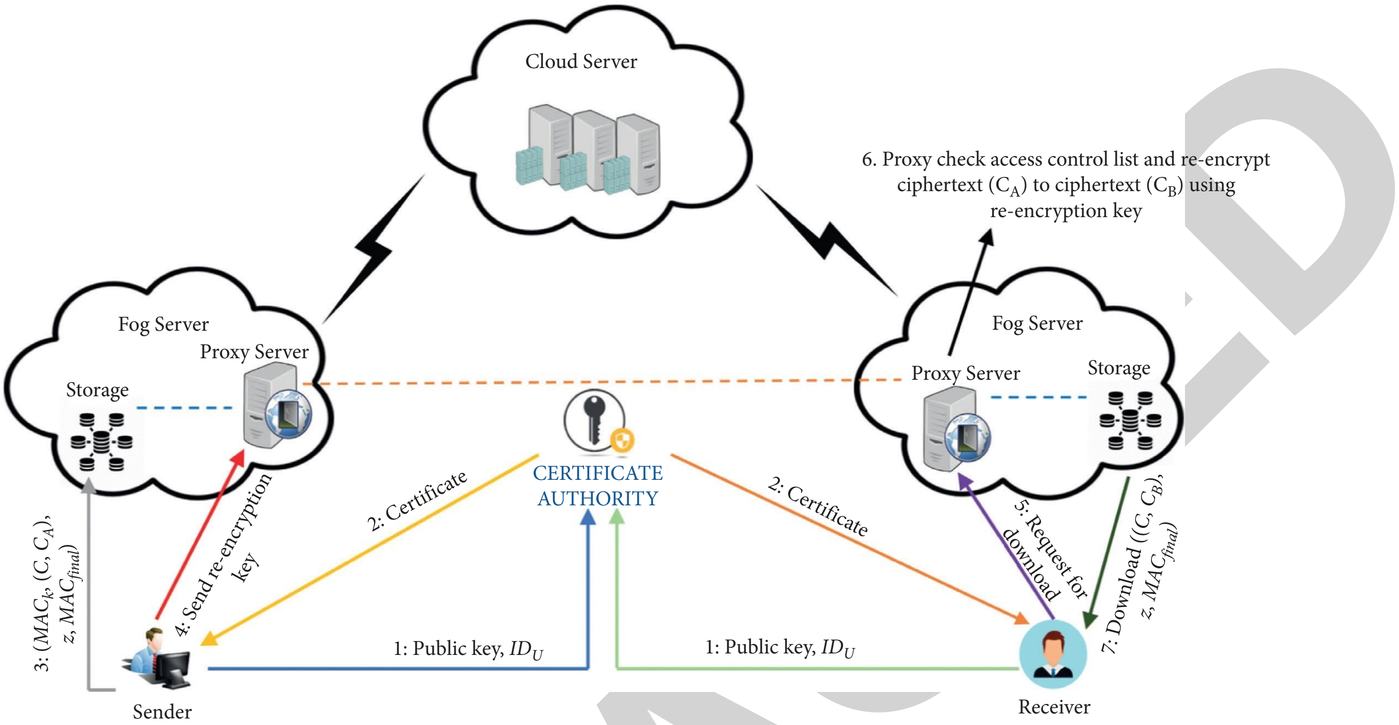

Cloud Server
6. Proxy check access control list and re-encrypt ciphertext (C_A) to ciphertext (C_B) using re-encryption key
Fog Server
Proxy Server
Storage
Fog Server
Proxy Server
Storage
CERTIFICATE AUTHORITY
2: Certificate
2: Certificate
4: Send re-encryption key
3: (MAC_k, (C, C_A), z, MAC_{final})
5: Request for download
7: Download ((C, C_B), z, MAC_{final})
1: Public key, ID_U
1: Public key, ID_U
Sender
Receiver

Figure 3: Proposed scheme architecture for secure data sharing in fog computing.

$$B_k = C_k - \left(\frac{1}{K_B}\right) C_B, \tag{15}$$

where $1 \leq k \leq z$.

User A can decrypt the file by using C_A and his private key K_A as shown in the following equation:

$$B_k = C_k - \left(\frac{1}{K_A}\right) C_A, \tag{16}$$

where $1 \leq k \leq z$.

After decrypting all the blocks, user B concatenates all the blocks to get the original file. User B checks the integrity of the file depicted as follows:

$$\text{FILE} = \Vert_{k=1}^{z} \left(B_k\right), \tag{17}$$

where $1 \leq k \leq z$,

$$\text{MAC}_k = H_{\text{SHA}-3} \left(B_k\right), \tag{18}$$

where $1 \leq k \leq z$, and

$$\text{MAC}_{\text{final}} = H_{SHA-3} \Vert_{k=1}^{z} \left(\text{MAC}_k\right), \tag{19}$$

where $1 \leq k \leq z$.

If the value of the calculated hash function is equal to the value of the final hash function ($\text{MAC}_{\text{final}}$), then the FILE integrity is confirmed; otherwise, the original file is changed by some intruder.

5. Incremental Block Modification

In this phase, we use incremental cryptographic technique shown in Figure 4. The block modification operations are performed by using the incremental cryptographic technique. The block modification consists of three phases: (1) block deletion, (2) block updation, and (3) block insertion. Dividing the FILE into blocks and calculating the hash value of each block increases the overhead, but this overhead can be overcome by using the incremental cryptographic technique during the block modification operation. There is no need to calculate the hash value from scratch while performing the block modification operations. Different phases of block modification are given in [56].

5.1. Block Deletion Operation. In this phase, the user wants to delete some blocks r of the uploaded file from different locations Idx. The user downloads hash values of all blocks (MAC_k) of the corresponding file from the fog node. The user can delete the required blocks from the file by updating the final hash value. Final hash is updated by computing the hash of the concatenated hashes of the remaining blocks as depicted in (20).

$$\text{MAC}_{\text{final}} = H_{\text{SHA}-3} \left(\Vert_{k=1}^{I\ dx-1} \text{MAC}_k \Vert_{L=I\ dx+r}^{z} \text{MAC}_L \right). \tag{20}$$

The user sends the location information (Idx) of the deleted blocks and updates the final hash value ($\text{MAC}_{\text{final}}$), total number of deleted blocks r, with delete request to the proxy server. The proxy server updates the final hash value $\mathrm{MAC}_{\mathrm{final}}$ and deletes all the requested blocks from the file given in (20) and (21).

$$\mathrm{MAC}_{\mathrm{final}} = \left(\Big\|_{k=1}^{I\ dx-1} \mathrm{MAC}_k \Big\|_{L=I\ dx+r}^{z} \mathrm{MAC}_L \right), \tag{21}$$

$$C = \left(\Big\|_{k=1}^{I\ dx-1} \mathrm{MAC}_k \Big\|_{L=I\ dx+r}^{z} \mathrm{C}_L \right). \tag{22}$$

The proxy server also updates all the values into the fog storage, and the value of the total number of blocks is also changed from z to $z = z\text{-}r$.

5.2. Block Modification Operation. In this phase, if the user wants to modify some blocks r of the corresponding file at different locations Idx, the user downloads the hash value of all the blocks (MAC_k) and the encrypted file (C, C_U) from the fog node. The user then encrypts all the blocks that need to be modified using the following process:

$$B_{\mathrm{update}_k} = C_{\mathrm{update}_k} + \left(\frac{1}{K_A} \right) C_A\,, \tag{23}$$

where $1 \le k \le z$.

The user also calculates the hash values of all the modified blocks, and afterwards, the user calculates the final hash value $\mathrm{MAC}_{\mathrm{final}}$ as follows:

$$\mathrm{MAC}_{\mathrm{update}} = H_{\mathrm{SHA-3}} \left(B_{\mathrm{update}} \right), \tag{24}$$

$$\mathrm{MAC}_{\mathrm{final}} = \left(\Big\|_{k=1}^{I\ dx-1} \mathrm{MAC}_k \Big\|_{L=I\ dx}^{I\ dx+r} \mathrm{MAC}_{\mathrm{update}_L} \Big\|_{M=I\ dx+r}^{z} \mathrm{MAC}_M \right). \tag{25}$$

The user sends the location information (Idx) of the modified blocks and the final hash value ($\mathrm{MAC}_{\mathrm{final}}$) obtained by applying hash function on the concatenated hash values of all the blocks, total number of modified blocks r, with modification request to the proxy server. The proxy server updates the final hash value $\mathrm{MAC}_{\mathrm{final}}$ and updates all the requested blocks of the FILE as follows:

$$\mathrm{MAC}_{\mathrm{final}} = \left(\Big\|_{k=1}^{I\ dx-1} \mathrm{MAC}_k \Big\|_{L=I\ dx}^{I\ dx+r} \mathrm{MAC}_{\mathrm{update}_L} \Big\|_{M=I\ dx+r}^{z} \mathrm{MAC}_M \right), \tag{26}$$

$$\mathrm{C}_{\mathrm{final}} = \left(\Big\|_{k=1}^{I\ dx-1} \mathrm{C}_k \Big\|_{L=I\ dx}^{I\ dx+r} \mathrm{C}_{\mathrm{update}_L} \Big\|_{M=I\ dx+r}^{z} \mathrm{C}_M \right). \tag{27}$$

The proxy server updates all the values into the fog storage, and the value of the total number of blocks z remains same.

5.3. Block Insertion Operation. In this phase, if the user wants to insert some new blocks r at different locations Idx of the corresponding file. Then, the user downloads the hash value of all the blocks (MAC_k) and encrypted file (C, C_U) from the fog node. The user then encrypts all the new blocks that need to be inserted in the file using the following procedure:

$$C_{\mathrm{new}} = \chi \cdot P_A C_U = B_{\mathrm{new}} \cdot \chi \cdot \wp. \tag{28}$$

The user also calculates the hash values of all the newly inserted blocks in the file, and afterwards, the user calculates the final hash value $\mathrm{MAC}_{\mathrm{final}}$ using the following equations:

$$\mathrm{MAC}_{\mathrm{insert}} = H_{\mathrm{SHA-3}} \left(B_{\mathrm{insert}} \right), \tag{29}$$

$$\mathrm{MAC}_{\mathrm{final}} = \left(\Big\|_{k=1}^{I\ dx-1} \mathrm{MAC}_k \Big\|_{L=I\ dx}^{I\ dx+r} \mathrm{MAC}_{\mathrm{update}_L} \Big\|_{M=I\ dx+r}^{z} \mathrm{MAC}_M \right). \tag{30}$$

The user sends the location information (Idx) of the newly inserted blocks in the file and final hash value ($\mathrm{MAC}_{\mathrm{final}}$) obtained by applying hash function on all the concatenated hash values of all old and new blocks, total number of newly inserted blocks r, with insertion request to the proxy server. The proxy server updates the final hash value $\mathrm{MAC}_{\mathrm{final}}$ and all the newly inserted blocks in the file using the following procedure:

$$\mathrm{MAC}_{\mathrm{final}} = \left(\Big\|_{k=1}^{I\ dx-1} \mathrm{MAC}_k \Big\|_{L=I\ dx}^{I\ dx+r} \mathrm{MAC}_{\mathrm{update}_L} \Big\|_{M=I\ dx+r}^{z} \mathrm{MAC}_M \right), \tag{31}$$

$$\mathrm{C}_{\mathrm{final}} = \left(\Big\|_{k=1}^{I\ dx-1} \mathrm{C}_k \Big\|_{L=I\ dx}^{I\ dx+r} \mathrm{C}_{\mathrm{update}_L} \Big\|_{M=I\ dx+r}^{z} \mathrm{C}_M \right). \tag{32}$$

The proxy server updates all the values into the fog storage, and the value of the total number of blocks also changes from z to $z = z + r$.

6. Security Analysis

In this phase, we present the security analysis is of our certificate-based incremental proxy re-encryption scheme. In this security analysis, we ensure various security requirements such as integrity and confidentiality.

6.1. Confidentiality. Data confidentiality means that sensitive data are protected from unauthorized users and blocks unauthorized users' access to sensitive data.

6.1.1. Level-1 Encryption. In our method, if an intruder wants to steal our sensitive data, then he must know about the level-1 encryption sender private key. An intruder can get the sender's private key by performing the following steps:

> Step 1: the intruder can get a level-1 encryption sender private key if he computes equation (33). For this, the intruder needs to find the value of β, and it is not feasible due to the hyperelliptic curve discrete logarithm problem:
>
> $$K_U = \beta \cdot (\xi - \eta). \tag{33}$$
>
> Step 2: if in any way the intruder gets the value of β, then next it wants to find the value of ξ and η from equations (34) and (35). For this, the intruder needs to find the value of α and γ, and it is also not feasible due to

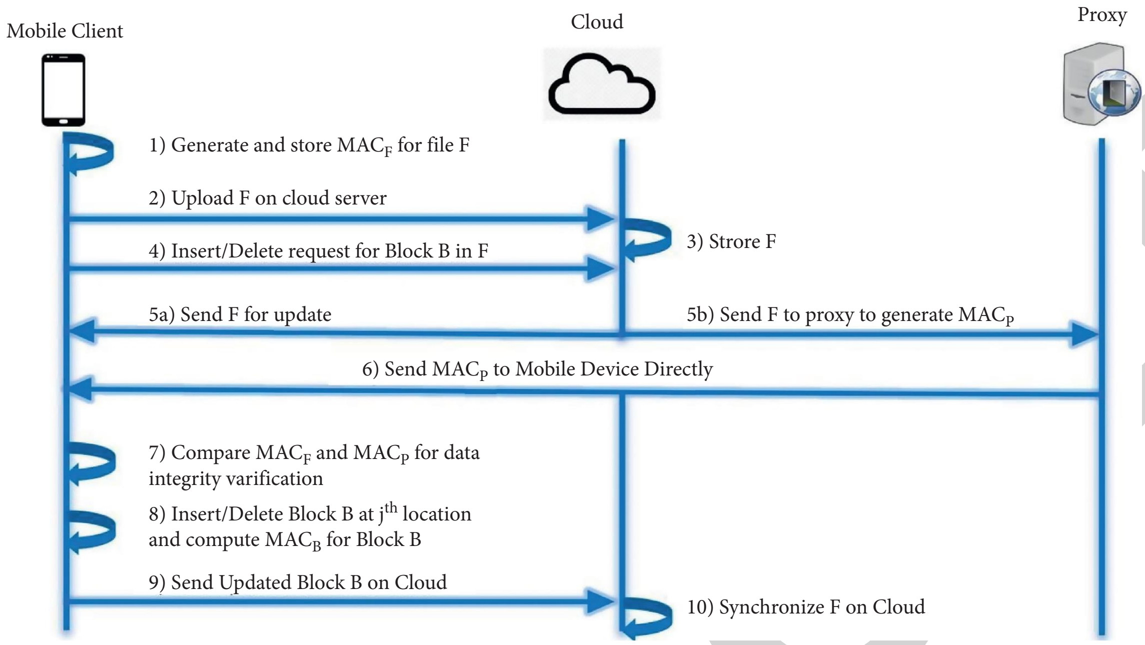

Mobile Client
Cloud
Proxy
1) Generate and store MAC_F for file F
2) Upload F on cloud server
3) Strore F
4) Insert/Delete request for Block B in F
5a) Send F for update
5b) Send F to proxy to generate MAC_P
6) Send MAC_P to Mobile Device Directly
7) Compare MAC_F and MAC_P for data integrity varification
8) Insert/Delete Block B at j^{th} location and compute MAC_B for Block B
9) Send Updated Block B on Cloud
10) Synchronize F on Cloud

Figure 4: Workflow of the incremental cryptographic scheme [66].

the hyperelliptic curve discrete logarithm problem. In fact, to get the level-1 encryption sender private key, the intruder will have to solve the hyperelliptic curve discrete logarithm problem three times, which is not feasible.

$$\xi = \alpha \cdot (\gamma - \alpha), \tag{34}$$

$$\eta = \gamma \cdot (\beta + \gamma). \tag{35}$$

6.1.2. Level-2 Encryption. In the level-2 encryption phase, we analyse the confidentiality of the proxy re-encrypted data. In this phase, there are two cases, one for the intruder.

Case 1: in the first case, the intruder wants to get the sensitive data. For this, he must know about the level-2 encryption receiver private key. An intruder can get the receiver's private key by performing the following steps:

Step 1: the intruder can get a level-2 encryption receiver private key if he computes equation (33). For this, the intruder needs to find the value of β, and it is not feasible due to the hyperelliptic curve discrete logarithm problem.

Step 2: if in any way the intruder gets the value of β, then next it wants to find the value of ξ and η from equations (34) and (35). For this, the intruder needs to find the value of α and γ, and it is also not feasible due to the hyperelliptic curve discrete logarithm problem.

6.2. Integrity. Data integrity means the overall reliability, completeness, and accuracy of the data. This means that the recipient receives the same data which is sent by the sender, and we use the hash function to ensure integrity. The sender applies the hash function on each data block and then applies final hash on the concatenated hash value shown as follows.

$$\mathrm{MAC}_k = H_{\mathrm{SHA-3}}(B_k), \tag{36}$$

where $1 \leq k \leq z$, and

$$\mathrm{MAC}_{\mathrm{final}} = H_{\mathrm{SHA-3}} \|_{k=1}^{z} (\mathrm{MAC}_k), \tag{37}$$

where $1 \leq k \leq z$.

If the attacker has made any changes in the ciphertext, then C is converted to C', and after decrypting C', we get the message M', and the MAC value of M' is not equal to $\mathrm{MAC}_{\mathrm{final}}$. So, we' will find out if the ciphertext has changed. Therefore, our proposed scheme verifies the integrity of the data.

6.3. Replay Attack. In our scheme, every time the proxy server generates a Nonce value Nonce when it re-encrypts the blocks and attached this Nonce value to every block such as FILE $= \| (B_k, \mathrm{Nonce})$. Nonce value Nonce is the identity of each block. The Nonce value is attached with each block to avoid the replay attack. If an intruder tries to send the previous block to the recipient, the recipient easily knows that it is the previous block, by the Nonce value, because the Nonce value is renewed in each session. In this regard, we can say that our scheme is safe from the replay attack.

7. Experimental Results

Our scheme is evaluated on the bases of turnaround time for the different file size that is given in Table 2. We have divided our file into eight blocks. Then, we performed four experiments on the given data set, and in the first experiment, we update only one block and compare the results with the

Table 2: Parameters for all scenarios.

File size in bytes	Total files
51 200	50
102,400	50
153,600	50
204,800	50
256,000	50

previously proposed schemes by Guo et al. [51], Khan et al. [21], Khan et al. [56], and Bhatia et al. [64]. Proxy re-encryption scheme proposed in [51] does not use the incremental cryptographic technique; therefore, it encrypts, decrypts, and computes the hash value for a complete file from scratch. The schemes in [21, 56] use the bilinear paring technique which uses a 256-bit key, and [64] uses the elliptic curve technique with 160-bit key. Our certificate-based incremental proxy re-encryption scheme (CB-PReS) uses the hyperelliptic curve technique with 80-bit key.

7.1. Case Scenario 1. In first case, we perform block modification operation in which we modify only one block of a file with different file sizes given in Table 2, and then we compare our results with the previously proposed schemes [21, 51, 56, 64] based on the turnaround time while block deletion, block insertion, and block updation. We can evaluate the turnaround time as follows:

$$\text{turn around time (TT)} = t_{r\ d} + t_{\text{hash}} + t_{e\ d}, \tag{38}$$

where t_{rd} represents the time that is required to read the file, t_{hash} represents the time that is required to compute the hash value, and t_{ed} represents the time that is required for encryption/decryption. Figure 5 shows the result in terms of turnaround time while modifying the block. The file size in bytes is shown along the x-axis with the total numbers of files, and the turnaround time is shown along the y-axis in milliseconds. In this experiment, every file is divided into eight blocks. The results of experiment 1 in Figure 5 clearly show that our scheme is more efficient as compared to the previously proposed schemes based on the turnaround time while block deletion, block insertion, and block updation.

7.2. Case Scenario 2. In second case, we perform block modification operation in which we modify two blocks of a file with different file sizes, and then we compare the obtained results with the previously proposed schemes based on the turnaround time while block deletion, block insertion, and block updation. We can evaluate the turnaround time as shown in equation (38).

The results of experiment 2 in Figure 6 clearly shows that our scheme is more efficient as compared to the previously proposed schemes [21, 51, 56, 64] based on the turnaround time while block deletion, block insertion, and block updation.

7.3. Case Scenario 3. In third case, three blocks are modified of a file with different file sizes, and then we compare the results with the previous proposed schemes. We can evaluate the turnaround time as shown in equation (38).

The results of experiment 3 in Figure 7 clearly shows that our scheme is more efficient as compared to the previous schemes.

7.4. Case Scenario 4. In fourth case, we update four blocks of a file with different sizes, and then we compare the results with the previously proposed schemes. We can evaluate the turnaround time as shown in equation (38).

The results of experiment 4 in Figure 8 clearly shows that our scheme is more efficient as compared to the previous proposed schemes.

8. Performance Evaluation

In this phase, we compare our proposed scheme with the previously proposed schemes by Guo et al. [51], Khan et al. [21], Khan et al. [56], and Bhatia et al. [64] and then evaluate our scheme in terms of computation overhead of each block. All the specifications about hardware and software which we use to conduct the experiments are depicted in Table 3. For the development of the fog client application, we use iFogSim simulator [67]. The major operation, i.e., hyperelliptic curve divisor multiplication (HDM), cost is computed by using a GitHub library called libg2hec [68]. This library provides a divisor group operation in the Jacobian of genus 2 (imaginary) hyperelliptic curve. We need to install V. Shoup's NTL library [69] before installing the G2HEC library. G2HEC library results have been tested with NTL 5.5 version on an x86_64 macOS big sur 11.4 operating system. The cost of the elliptic curve scalar multiplication (SM) operation is computed by using a standard cryptography library MIRACL [70] and bilinear paring modular exponential (MEXP), and bilinear paring operation (PBC) cost is computed by using the (pairing-based cryptography) library PBC [71]. PBC is built on the GMP library [72], and all the paring base mathematical operation such as bilinear paring modular exponential and pairing-based point multiplication are performed by using the GMP library. We can achieve 1024-bit RSA level security by using type A bilinear pairing over an elliptic curve with $p = 512$-bit prime and $q = 160$-bit prime number. We can also achieve the same level of security in ECC by using secp160r1 curve over a finite field Fq, and the equivalent level of security can also be achieved by using the genus two hyperelliptic curve over the finite field Fq with the following values $|Fq\ |g.\ \log 2\ q \approx 280$. To achieve the integrity of the data, we have used the SHA-3 algorithm, which is more resistant to collision and preimaging attacks than the previous SHA-2 algorithm.

8.1. Computational Overhead. In this phase, we consider some major operations such as hyperelliptic curve divisor multiplication (HDM), elliptic curve scalar multiplication (SM), bilinear paring modular exponential (MEXP), and bilinear paring operation (PBC). The cost of the major operation in millisecond is given as follows: hyperelliptic curve divisor multiplication cost is 0.36 ms, elliptic curve

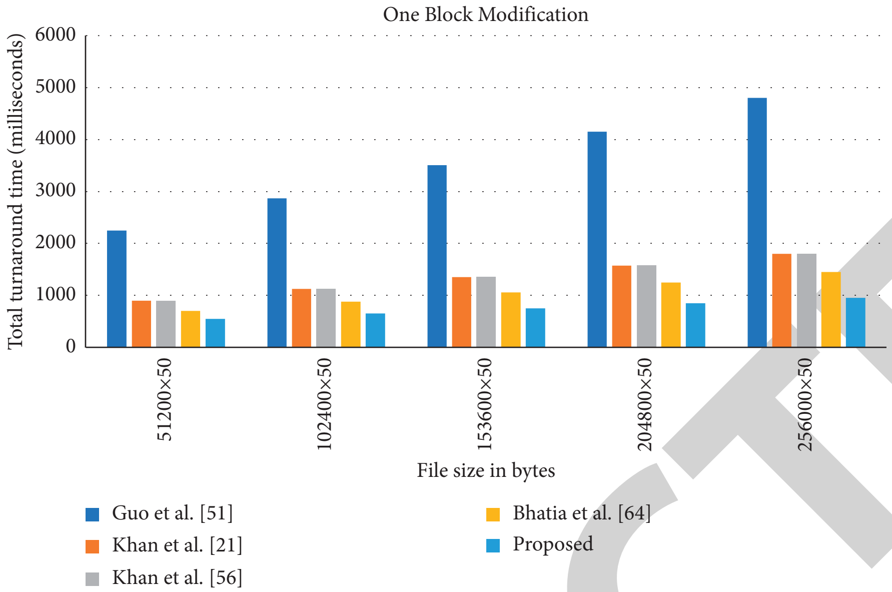


Figure 5: One block modification.

scalar multiplication cost is 0.64 ms, bilinear paring modular exponential cost is 0.84 ms, and bilinear paring operation cost is 1.02 ms. It is clearly shown from the results that the hyperelliptic curve divisor multiplication cost is much lesser as compared to the remaining cryptographic major operations. Our proposed scheme uses hyperelliptic curve divisor multiplication operations. The cost of the all-cryptographic major operations is given in Figure 9 and Table 4.

Due to the use of hyperelliptic curve divisor multiplication, our proposed (CB-PRe) scheme shows significant improvement in the block(s) modification results compared to the previously proposed [21, 51, 56, 64] schemes. The result improvement depends on the number of block(s) modification operations. The comparison of result improvements between our proposed (CB-PRe) scheme and Guo et al.'s [51] scheme is given in Table 5; it shows that the average percentage turnaround time of our proposed (CB-PRe) scheme is 78.24% better than Guo et al.'s scheme while modifying one block, 70.47% better while the modification of two blocks, 62.73% better while modifying three blocks, and 54.54% better while the modification of four blocks. The computational overhead reduction formula is given as follows:

$$\left(\frac{\text{existing scheme } cost\ -\ proposed\ scheme\ \text{cost}}{existing\ scheme\ \text{cost}}\right) * 100. \tag{39}$$

Table 6 depicts the comparison of result improvements between our (CB-PRe) scheme and Khan et al.'s [21, 56] schemes. The average percentage turnaround time of our proposed scheme is 40.66% better than Khan et al.'s [21, 56] schemes, while modifying one block, 37.51% better while the modification of two blocks, 36.62% better while modifying three blocks, and 33.48% better while the modification of four blocks.

The comparison of result improvements between our proposed scheme and Bhatia et al.'s [64] scheme is given in Table 7; it depicted that the average percentage turnaround time of our scheme is 25.67% better than Bhatia et al.'s scheme while modifying one block, 24.2% better while the modification of two blocks, 22.2% better while modifying three blocks, and 19.92% better while the modification of four blocks.

8.2. Communication Overhead. In this section, we compared the communication cost of our scheme with the existing proposed schemes, i.e., Guo et al. [51], Khan et al. [21], Khan et al. [56], and Bhatia et al. [64]. Communication costs will increase if there are some extra bits with the original message. We suppose some terms as follows:

(i) |K | represents key size of bilinear pairing is equals to 256 bits, elliptic curve is 160 bits, and hyperelliptic curve is 80 bits

(ii) |M| represents ciphertext or plaintext size and equals 100 bits

(iii) MAC represents message digest size of hash function and equals 512 bits

(iv) C_{ERT} represents user certificate size and equals 256 bits in bilinear paring, 160 bits in elliptic curve, and 80 bits in hyper elliptic curve

(v) |B| block size is equal to 13 bits

(vi) $R_{A\longrightarrow B}$ re-encryption key size of bilinear pairing equals 256 bits, elliptic curve is 160 bits, and hyperelliptic curve is 80 bits

Table 8 shows that our scheme is better than the existing related schemes, i.e., [21, 51, 56, 64] in terms of communication cost.

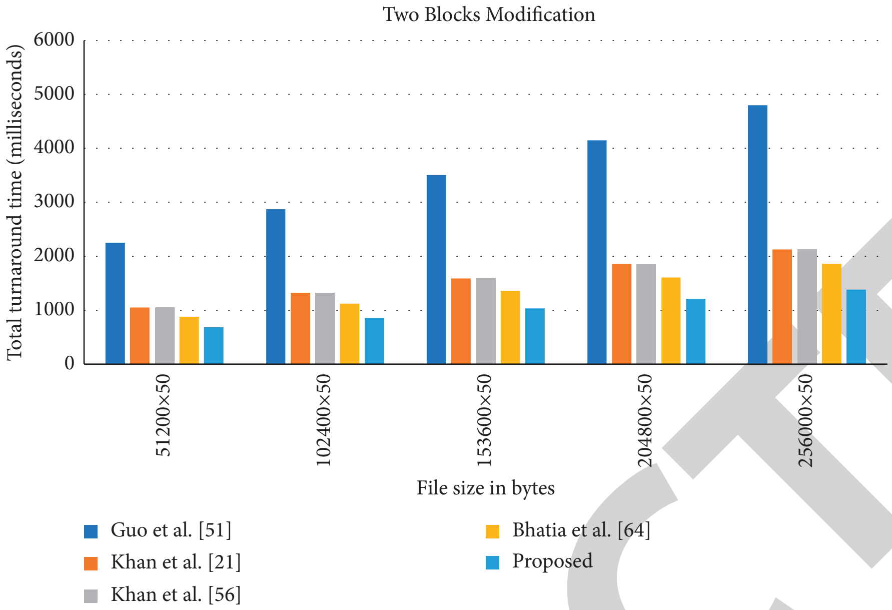


FIGURE 6: Two-block modification.

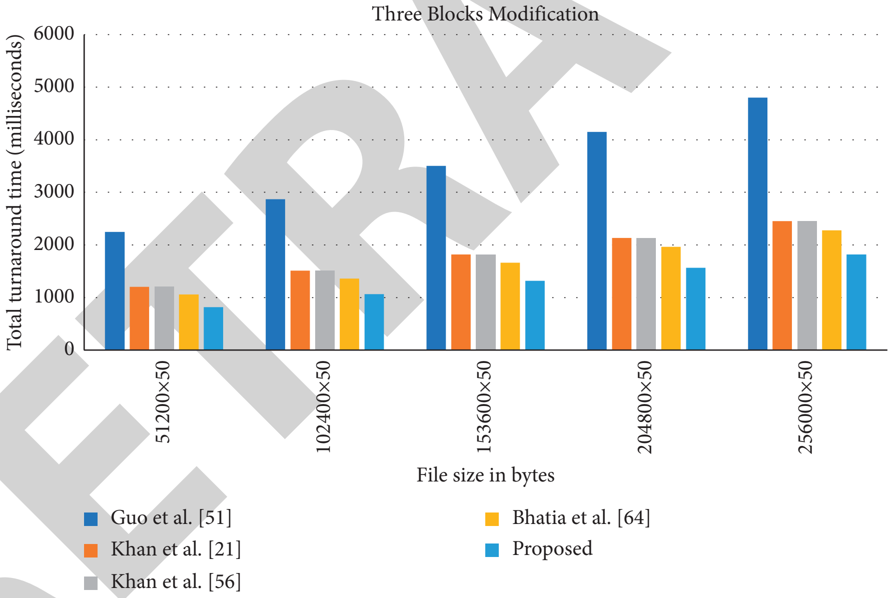


FIGURE 7: Three blocks modification.

8.3. Communication Cost Reduction. The communication cost reduction of our scheme from the existing schemes can be calculated by the following formula:

$$\left(\frac{\text{existing scheme cost} - \text{proposed scheme cost}}{\text{existing scheme cost}}\right) * 100. \tag{40}$$

Reduction from Guo et al.'s study [51] is

$$\left(\frac{3216 - 2002}{3216}\right) * 100 = 37.75\,\%. \tag{41}$$

Reduction from Khan et al.'s study [21] is

$$\left(\frac{2886 - 2002}{2886}\right) * 100 = 30.63\,\%. \tag{42}$$

Reduction from Khan et al.'s study [56] is

$$\left(\frac{2530 - 2002}{2530}\right) * 100 = 20.86\,\%. \tag{43}$$

Reduction from Bhatia et al.'s study [64] is

$$\left(\frac{2242 - 2002}{2242}\right) * 100 = 9.10\,\%. \tag{44}$$

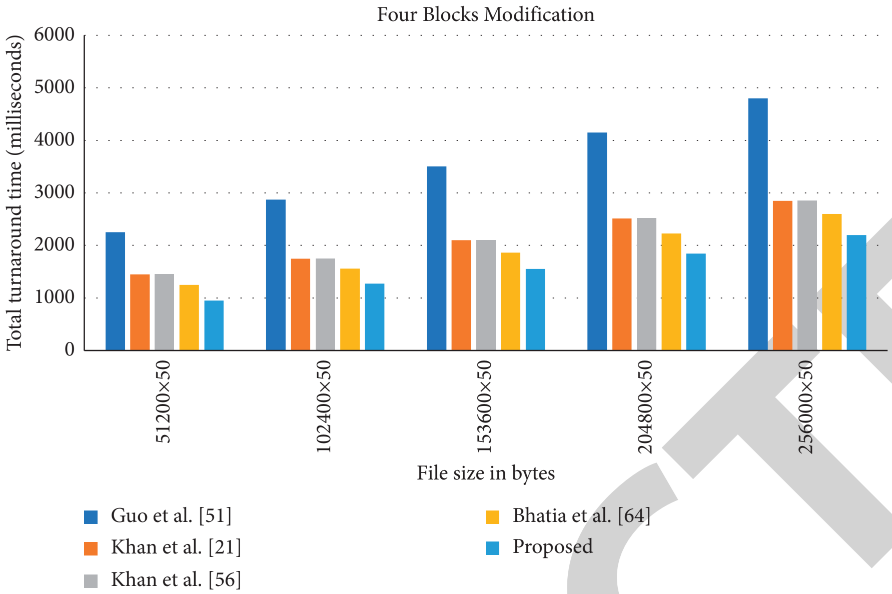


Figure 8: Four blocks modification.

Table 3: Hardware and software specification.

Hardware and software	Specification/version
System	MacBook pro-2015
RAM	16 GB 1600 MHz DDR3
Processor	2.2 GHz quad-core Intel Core i7
Storage	512 GB
OS	macOS big sur 11.4
Application	iFogSim
iFogSim	1.0
libg2hec	2.1
NTL	5.5
MIRACL	7.0.0
PBC	1.0
GMP	6.2.1

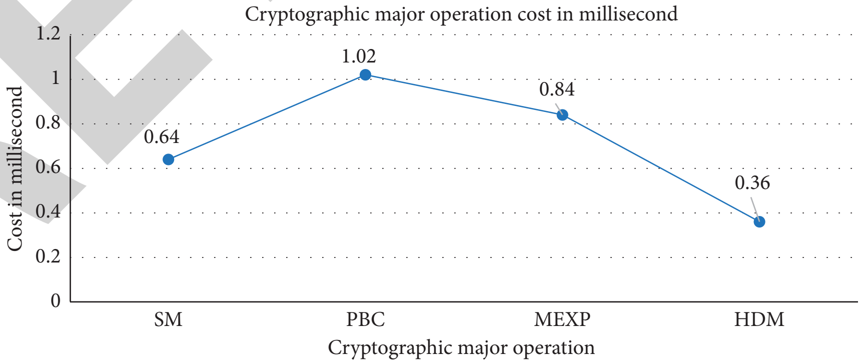


Figure 9: Cryptographic major operation cost in millisecond.

Table 4: Cryptographic major operation cost in millisecond.

Major operations	Cost in milliseconds
Bilinear paring operation	1.02
Bilinear paring modular exponential	0.84
Elliptic curve scalar multiplication	0.64
Hyperelliptic curve divisor multiplication	0.36

Table 5: Result improvements of the proposed scheme in block(s) modification operations as compared to [51].

Number of blocks	51200 × 50 (%)	102400 × 50 (%)	153600 × 50 (%)	204800 × 50 (%)	256000 × 50 (%)	Average (%)
1	75.56	77.35	78.6	79.51	80.2	78.24
2	69.64	70.1	70.52	70.89	71.18	70.47
3	61.89	62.25	62.74	63.09	63.69	62.73
4	53.12	53.61	54.35	55.47	56.16	54.54

Table 6: Result improvements of proposed scheme in block (s) modification operations as compared to [21, 56].

Number of blocks	51200 × 50 (%)	102400 × 50 (%)	153600 × 50 (%)	204800 × 50 (%)	256000 × 50 (%)	Average (%)
1	38.88	39.22	40.49	41.98	42.71	40.66
2	36.09	36.91	37.42	38.12	39.03	37.51
3	34.95	35.61	36.27	37.41	38.88	36.62
4	31.07	32.62	33.88	34.73	35.12	33.48

Table 7: Result improvements of proposed scheme in block(s) modification operations as compared to [64].

Number of blocks	51200 × 50 (%)	102400 × 50 (%)	153600 × 50 (%)	204800 × 50 (%)	256000 × 50 (%)	Average (%)
1	24.21	25.09	25.87	26.34	26.83	25.67
2	23.01	23.67	24.22	24.79	25.31	24.2
3	21.47	21.89	22.78	22.56	22.31	22.2
4	19.09	19.76	20.01	20.43	20.32	19.92

Table 8: Communication cost in bits.

Guo et al. [51]	2 (K) + 4 (M) + 4 (MAC) + 1 ($R_{A\longrightarrow B}$)	2 (256) + 4 (100) + 4 (512) + 1 (256) = 3216
Khan et al. [21]	2 (K) + 3 (M) + 4 (MAC) + 2 (B)	2 (256) + 3 (100) + 4 (512) + 2 (13) = 2886
Khan et al. [56]	2 (K) + 2 (M) + 3 (MAC) + 1 ($R_{A\longrightarrow B}$) + 2 (B)	2 (256) + 2 (100) + 3 (512) + 1 (256) + 2 (13) = 2530
Bhatia et al. [64]	2 (K) + 2 (M + 3 (MAC) + 1 ($R_{A\longrightarrow B}$) + 2 (B)	2 (160) + 2 (100) + 3 (512) + 1 (160) + 2 (13) = 2242
Proposed method	K + 1 (C_{ERT}) + 2 (M) + 3 (MAC) + 1 ($R_{A\longrightarrow B}$) + 2 (B)	1 (80) + 1 (80) + 2 (100) + 3 (512) + 1 (80) + 2 (13) = 2002

9. Conclusion

Secure EHR storage and sharing is a serious issue while using the cloud infrastructure. In certificateless cryptography, we cannot verify the public key of any user that is the major drawback of the certificateless cryptographic schemes. In this paper, we have proposed a lightweight certificate-based incremental proxy re-encryption for e-healthcare data sharing scheme in fog computing. In certificate-based schemes every user can verify the public key of other users. Recently proposed I-PRE schemes involve expensive bilinear pairing and elliptic curve operation which uses 256-bit key and 160-bit key, respectively. In this scheme, we use the hyperelliptic curve technique with 80-bit key and compare the block modification results on the basis of turnaround time with the previously proposed I-PRE schemes. Results clearly show that our scheme is more efficient as compared to the previously proposed schemes. Our scheme also provides data integrity and confidentiality and deals with the latency problem by using the fog computing paradigms. To reduce the overhead of resource constraints IoT devices, complex and resource-intensive cryptographic functions are offloaded on fog nodes. In the future, we will develop a group-based scheme that will be useful for multiple users.

Data Availability

The data used to support the findings of this study are provided in this article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] United Nations, *World Population Ageing, 2014*, Vol. 73, United Nations, Department of Economic and Social Affairs Population Division, , New York, NY, USA, 2014.

[2] M. Chen, J. Yang, Y. Hao, S. Mao, and K. Hwang, "A 5G cognitive system for healthcare," *Big Data and Cognitive Computing*, vol. 1, no. 1, p. 2, 2017.

[3] Frost & Sullivan, *Drowning In Big Data? Reducing Information Technology Complexities and Costs for Healthcare Organizations*, http://www.emc.com/collateral/analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf, Frost & Sullivan, San Antonio, TX, USA, 2012, http://www.emc.com/collateral/analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf.

[4] M. Chen, S. Mao, and Y. Liu, "Big data: a survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.

[5] M. S. Hossain and G. Muhammad, "Healthcare big data voice pathology assessment framework," *IEEE Access*, vol. 4, pp. 7806–7815, 2016.

[6] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, no. 1, pp. 8869–8879, 2017.

[7] M. Chen, P. Zhou, and G. Fortino, "Emotion communication system," *IEEE Access*, vol. 5, pp. 326–337, 2017.

[8] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019.

[9] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Systems with Applications*, vol. 137, pp. 22–28, 2019.

[10] F. S. Ahmad, L. Ali, R. U. Mustafa et al., "A hybrid machine learning framework to predict mortality in paralytic ileus patients using electronic health records (EHRs)," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3283–3293, 2021.

[11] A. A. Diro, N. Chilamkurti, and Y. Nam, "Analysis of lightweight encryption scheme for fog-to-things communication," *IEEE Access*, vol. 6, pp. 26820–26830, 2018.

[12] C. S. Nandyala and H.-K. Kim, "From cloud to fog and IoT-based real-time U-healthcare monitoring for smart homes and hospitals," *International Journal of Smart Home*, vol. 10, no. 2, pp. 187–196, 2016.

[13] M. M. Hassan, K. Lin, X. Yue, and J. Wan, "A multimedia healthcare data sharing approach through cloud-based body area network," *Future Generation Computer Systems*, vol. 66, pp. 48–58, 2017.

[14] X. Meng, W. Wang, and Z. Zhang, "Delay-constrained hybrid computation offloading with cloud and fog computing," *IEEE Access*, vol. 5, pp. 21355–21367, 2017.

[15] H. F. Atlam, R. J. Walters, and G. B. Wills, "Fog computing and the internet of things: a review," *Big Data and Cognitive Computing*, vol. 2, no. 2, pp. 1–18, 2018.

[16] J. Ni, K. Zhang, X. Lin, and X. Shen, "Securing fog computing for internet of things applications: challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 601–628, 2018.

[17] P. Zhang, J. K. Liu, F. R. Yu, M. Sookhak, M. H. Au, and X. Luo, "A survey on access control in fog computing," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 144–149, 2018.

[18] B. J. Mohd and T. Hayajneh, "Lightweight block ciphers for IoT: energy optimization and survivability techniques," *IEEE Access*, vol. 6, pp. 35966–35978, 2018.

[19] N. Farjana, S. Roy, M. J. N. Mahi, and M. Whaiduzzaman, "An identity-based encryption scheme for data security in fog computing," in *Proceedings of the International Joint Conference on Computational Intelligence,Algorithms for Intelligent Systems*, pp. 215–226, Springer, Dhaka, Bangladesh, October 2019.

[20] M. Al-Khafajiy, T. Baker, A. Waraich, D. Al-Jumeily, and A. Hussain, "IoT-fog optimal workload via fog offloading," in *Proceedings of the IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*, pp. 359–364, IEEE, Zurich, Switzerland, December 2018.

[21] A. N. Khan, M. L. M. Kiah, S. U. Khan, S. A. Madani, and A. R. Khan, "A study of incremental cryptography for security schemes in mobile cloud computing environments," in *Proceedings of the IEEE Symposium on Wireless Technology & Applications (ISWTA)*, pp. 62–67, IEEE, Kuching, Malaysia, September 2013.

[22] C. Cavanagh and U. C. Irvine, "UC irvine electronic Theses and Dissertations," vol. 228, University of California, Irvine, CA, USA, 2016, Thesis.

[23] C. Zhou, Z. Zhao, W. Zhou, and Y. Mei, "Certificateless key-insulated generalized signcryption scheme without bilinear pairings," *Security and Communication Networks*, vol. 2017, pp. 1–17, Article ID 8405879, 2017.

[24] A. Rahman, I. Ullah, M. Naeem et al., "A lightweight multimessage and multi-receiver heterogeneous hybrid signcryption scheme based on hyper elliptic curve," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, pp. 160–167, 2018.

[25] W. Itani, A. Kayssi, and A. Chehab, "Efficient healthcare integrity assurance in the cloud with incremental cryptography and trusted computing," *Cloud Technology*, pp. 845–857, 2015.

[26] M. Blaze, G. Bleumer, and M. Strauss, "Divertible protocols and atomic proxy cryptography," *Lecture Notes in Computer Science*, vol. 1403, pp. 127–144, 1998.

[27] R. Roy and P. P. Mathai, "Proxy re-encryption schemes for secure cloud data and applications: a survey," *International Journal of Computer Applications*, vol. 164, no. 5, pp. 975–8887, 2017.

[28] Z. Qin, H. Xiong, S. Wu, and J. Batamuliza, "A survey of proxy re-encryption for secure data sharing in cloud computing," *IEEE Transactions on Services Computing*, vol. 13, no. 9, p. 1, 2016.

[29] S. Kim and I. Lee, "IoT device security based on proxy re-encryption," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 4, pp. 1267–1273, 2018.

[30] M. Thangavel, P. Varalakshmi, and C. Abinaya, "A comparative study of attribute-based encryption schemes for secure cloud data outsourcing," in *Proceedings of the 2017 Ninth International Conference on Advanced Computing (ICoAC)*, pp. 261–266, IEEE, Chennai, India, December 2017.

[31] J. Weng, R. H. Deng, X. Ding, C.-K. Chu, and J. Lai, "Conditional proxy re-encryption secure against chosen-ciphertext attack," in *Proceedings of the 4th International Symposium ACM Symposium Information, Computer Communications Security ASIACCS'09*, pp. 322–332, Sydney, Australia, March 2009.

[32] P. Zeng and K.-K. R. Choo, "A new kind of conditional proxy Re-encryption for secure cloud storage," *IEEE Access*, vol. 6, pp. 70017–70024, 2018.

[33] M. Green and G. Ateniese, "Identity-based proxy re-encryption," in *Proceedings of the International Conference on Applied Cryptography and Network Security*, Springer, Zhuhai, China, June 2007.

[34] C. K. Chu and W. G. Tzeng, "Identity-based proxy re-encryption without random oracles," in *Proceedings of the International Conference on Information Security*, Valparaíso, Chile, October 2007.

[35] K. Liang, Z. Liu, X. Tan, D. S. Wong, and C. Tang, "A CCA-secure identity-based conditional proxy re-encryption without random oracles," *Lecture Notes in Computer Science*, vol. 7839, pp. 231–246, 2013.

[36] P. Dutta, W. Susilo, D. H. Duong, and P. S. Roy, "Collusion-resistant identity-based proxy re-encryption: lattice-based constructions in standard model," *Theoretical Computer Science*, vol. 871, pp. 16–29, 2021.

[37] Q. Tang, "Type-based proxy re-encryption and its construction," *Progress in Cryptology-Indocrypt*, vol. 5365, pp. 130–144, 2008.

[38] C. Sur, Y. Park, S. U. Shin, K. H. Rhee, and C. Seo, "Certificate-based proxy re-encryption for public cloud storage," in *Proceedings of the Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, pp. 159–166, IEEE, Taichung, Taiwan, July 2013.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9820840, 1 page
https://doi.org/10.1155/2023/9820840

Retraction

Retracted: Prediction of IoT Traffic Using the Gated Recurrent Unit Neural Network- (GRU-NN-) Based Predictive Model

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. A. Patil, L. A. Raj, and B. K. Singh, "Prediction of IoT Traffic Using the Gated Recurrent Unit Neural Network- (GRU-NN-) Based Predictive Model," *Security and Communication Networks*, vol. 2021, Article ID 1425732, 7 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2021, Article ID 1425732, 7 pages
https://doi.org/10.1155/2021/1425732

Research Article

Prediction of IoT Traffic Using the Gated Recurrent Unit Neural Network- (GRU-NN-) Based Predictive Model

Sonali Appasaheb Patil,[1] **L. Arun Raj**,[1] **and Bhupesh Kumar Singh**[2]

[1] *Department of CSE, BSAR Crescent Institute of Science and Technology, Chennai, India*
[2] *Arba Minch Institute of Technology, Arba Minch University, Ethiopia*

Correspondence should be addressed to Bhupesh Kumar Singh; dr.bhupeshkumarsingh@amu.edu.et

Received 26 June 2021; Revised 20 July 2021; Accepted 15 September 2021; Published 5 October 2021

Academic Editor: Chinmay Chakraborty

Prediction of IoT traffic in the current era has attracted noteworthy attention to utilize the bandwidth and channel capacity optimally. In this paper, the problem of IoT traffic prediction has been studied, and solutions have been proposed by using machine learning method ARIMA and learning time series algorithms such as LSTM and gated recurrent unit (GRU-NN) based on neural networks. The proposed GRU-NN predicts the traffic on the basis of transfer learning. The advantage of the GRU-NN over LSTM is also highlighted by solving the problem of gradient disappearance. The proposed GRU-NN memorizes the traffic characteristics of the IoT environment for a long time which eventually helps the system to forecast the upcoming traffic from the existing traces of the traffic. The proposed GRU-NN makes use of the transfer learning technique to handle the problem of insufficient IoT traffic data along with the gradient boosting training method for achieving better accuracy in predicting the network traffic in the IoT environment. The results reveal that the proposed GRU-NN model outperforms the other traffic predictors in terms of statistical performance evaluation parameters such as MAE, RMSE, MRE, and MSE. The results show that the GRU-NN provides the most accurate predictions followed by the LSTM predictor and then ARIMA and other approaches taken up for the comparative study.

1. Introduction

The Internet-of-Things (IoT) traffic is impacted by the regular changes in the IoT devices, their topology, the switching of the channel links during transmission, and the dynamic change of the connectivity of the devices to the internet. A major driver for the success of IoT networks is the prediction of upcoming traffic to handle the channel utilization and resource utilization optimally [1]. The factors that determine the health of the IoT network are cost and energy, and the accurate prediction of traffic can effectively save cost and energy. To prevent congestion on IoT channels and to improve the consumption of IoT resources, IoT traffic prediction and controlling are very important aspects for addressing. The prediction of IoT traffic can forecast the changing characteristics and tendencies of IoT traffic in advance [2]. The accurate forecasting of traffic can certainly allow the IoT users to avail the uninterrupted services [2]. Hence, it is mandatory to suggest a model that can forecast the IoT traffic in advance in an accurate manner. The Internet of Things (IoT) offers a plethora of interesting applications. One of the potential uses is in the sphere of healthcare, where data linked to healthcare are monitored by smart devices and transmitted to medical specialists. Wireless sensor network made up of smart healthcare monitoring devices records certain health indicators and sends them to a local personal digital assistant (PDA), which might be a smartphone or a bespoke device. Eventually, data are forwarded to a backend server through the internet. Remote monitoring in the healthcare industry is now feasible, thanks to Internet-of-Things- (IoT-) enabled gadgets, which have the ability to keep patients safe and healthy while also allowing clinicians to provide superior treatment. As contacts with doctors have grown easier and more efficient, it has also improved patient involvement and satisfaction. Furthermore, remote monitoring of a patient's health helps to shorten hospital stays and avoid readmissions. IoT has a huge influence on lowering healthcare expenses and increasing treatment results.

In the near future, traffic generated by IoT devices will increase tremendously, and the requirement of resources will also increase eventually. IoT channels have to deal with the futuristic traffic demands and must provide QoS to the users [3]. Therefore, effective utilization of network bandwidth, cloud resources, and available channels is crucial for the IoT environment. This can be achieved by forecasting the upcoming traffic on the channels well in advance accurately. In the IoT environment, there are important issues that require advance prediction of incoming traffic: (i) demand for high transmission rate, (ii) minimum energy usage by saving battery life and by using channels optimally [4], (iii) minimal latency during wired or wireless communications [2], and (iv) optimal utilization of data-intensive and computational resources. The solution to all these issues is efficient prediction of traffic which can certainly address the aforementioned issues.

This paper explores different kinds of predictors and proposes the GRU-NN-based predictor which gives better accuracy in predicting real-time traffic in the IoT environment. The predictors considered for the study are compared with the proposed GRU-NN for the prediction of IoT traffic. The traffic predictors in IoT are widely used to save power by integrating traffic predictors in switches and handoff mechanisms. With the increase in IoT traffic, the demand for the computational and data resources is increasing. The cost of using these resources also increases exponentially. The prediction of traffic can save energy and cost optimally if the predictor produces the results accurately. If the IoT traffic can be projected precisely, the bandwidth can be used optimally, the need for processors in core switches can also be reduced, and the channels can be allocated effectively for forwarding the IoT traffic to intended nodes.

The accurate predictions can also assist in congestion control over the channels, allocation of bandwidth, utilization of resources, detection of anomaly, and reduction in network latency. With the emergence of internet applications such as YouTube, Google Meet, Microsoft Teams, Zoom, and Netflix, video traffic has also been increasing tremendously. There are many state-of-the-art techniques that can predict the traffic, but still, there is a scope for improvement in the existing techniques and a room for exploring new techniques which can handle dynamic video traffic on channels and satisfy the needs of the IoT users. Furthermore, it is challenging to detect and prevent network abuse with the growing IoT traffic at rapid rate and due to diversity in networks.

The objective of the research study is to explore the traffic predictors suitable for IoT applications. The characteristics of the predictor should include accurate prediction of unseen traffic, lower computational complexity, lower space complexity, and lower consumption of power.

The following are the primary contributions of our study in this paper:

(i) The existing techniques are explored and implemented in order to make a comparison of the existing methods with the proposed GRU-NN method. A detailed comparison with respect to accuracy of results is presented in the paper.

(ii) To use the computing resources optimally and to diminish the computational complexity, a GRU-NN is proposed which is more efficient than LSTM; it is able to store long-term state with lesser complexity. GRU-NN can significantly enhance the training efficiency by remembering the states.

(iii) For the problem of insufficient data of IoT devices and unavailability of historical data, a transfer learning model has also been introduced in this paper. By learning from the accumulated data from IoT channels, the transfer learning (TL) method allows the predictor to get trained from accumulated data and the application of the TL-based model in the local domain.

(iv) Statistical performance evaluation metrics are used to verify the accuracy of the proposed GRU-NN with other conventional predictors.

This paper is organized into five sections. The paper begins with the introductory material where background details are highlighted, need of this research study is mentioned, and contributions and objectives are clearly stated. The next section provides details on the existing work similar to our research study and also highlights the research gaps. The third part elaborates the proposed GRU-NN-based predictive model. The next section provides details on the obtained results and comparative study. The last part of the paper concludes the work and result outcomes of the GRU-NN-based predictive model.

2. Related Work

The quick development in IoT traffic promotes huge research advancements in the field of the prediction of network traffic. IoT requires improved and proficient ways to deal with the immense and dynamic traffic [2]. The traffic prediction literature is surveyed as demonstrated in this section.

Abdellah et al. [2] proposed an ANN to predict IoT traffic and to improve the accuracy of IoT traffic projection. The predictive accuracy of the model has been calculated with statistical techniques. Compared to the other predictors, the MSE performance function has demonstrated the best prediction accuracy of the proposed method, and according to comprehensive study and simulation, MAPE has demonstrated the best prediction accuracy for the packet identifiers. Lopez-Martin et al. [3] introduced a new deep learning architecture that can be used to solve the supervised regression issue. It is based on an additive network architecture comprising learning blocks which works with the gradient boosting technique. The results of the proposed new model have been compared with a variety of emerging methods and significantly enhance prediction performance metrics as well as training/prediction processing times. In terms of machine learning methodologies, Mozo et al. [4] used the CNN to anticipate short-term changes in the quantity of traffic that goes through a data center network. The experimental results have shown that CNN's nonlinearity outperforms ARIMA's results.

In [2], authors presented a model that targets base stations using spatiotemporal information from nearby cellular stations. These features are used to predict traffic over time using deep learning techniques such as 3D convolutional networks. The technology employed yielded promising findings that outperformed those of existing traffic forecasting systems. Artificial intelligence (AI) techniques are required for a successful 5G network. The application of machine learning (ML) to traffic forecasting has been successful. In [5], a deep learning-based technique is explored to implement the traffic predictor using time series. The predictive accuracy is measured with the RMSE score and the mean absolute percentage of error (MAPE) as the MAE score. In [6], NARX time-series recurrent neural networks have been utilized to anticipate IoT communication. Three neural network training techniques have been used to test the predictability, trainlm, traincgf, and trainrp, with MSE, RMSE, and MAPE performance evaluators. As compared to others, the model forecast with the trainlm training module shows the best accuracy, while the model projected with the trainrp training module has the least predictive accuracy than other models, depending on the outcomes of the simulation.

In [7], the research is revolved around the deep neural framework for single-step time series forecasting that combines wavelet transformations (WTs), 2D CNNs, and LSTM- stacked autoencoders (SAEs). According to the findings, the suggested model has surpassed the other models in terms of prediction accuracy. In [8], the authors provided a brief overview on the contextual and existing literature on deep learning methods with possible features. The study has also provided an overview of a few strategies and technologies that assist in deploying deep learning techniques on mobile devices for the prediction of traffic in wireless networking. In [9], authors projected a feature selection strategy based on random forests to tackle the tough challenge of acquiring spatial data. The Gini score is used to indicate the spatial relationship between intersections in a data-driven network graph. The experimental results have suggested that using random forest feature selection and the RCF model, traffic forecast accuracy can reach 90%.

In [10], authors have demonstrated three deep learning models for forecasting the network traffic. CNN and RNN models with raw traffic data are proficient for accomplishing accurate outcomes when compared to two other baseline solutions. To forecast forthcoming transfer learning (TL) and congestion in the network, Tang et al. [11] introduced a unique deep learning architecture including a TL prediction method. To create a unique intelligent channel assignment technique, a deep neural-based predictive model with partially overlapping channel allocations is investigated. Finally, the paper suggested a unique intelligent channel assignment technique that smartly prevents future blocking of channels with huge traffic and quickly provides relevant channels to SDN IoT. The results of the simulation reveal that the approach is far superior to traditional algorithms for the channel assignment. The study in [12] uses machine learning approaches to accurately identify the IoT network. They have deployed the multivariate classifier for segregating IoT and non-IoT traffic. Every IoT device is assigned to a certain IoT device class in the second step. The model's overall IoT categorization accuracy has been analyzed as 99.281%.

In [13], authors described a machine learning method by examining streams of packets delivered and received for distinguishing the kind of IoT devices. They created a model to represent IoT device network activities based on the collected data. The network traffic created can be distinguished by various IoT devices by using the t-SNE approach to represent the data. The compliance data of the network will then be utilized to train distinct ML classifications to envisage which IoT device is responsible for the network traffic. The experiments have shown promising results with an overall accuracy of 99.9% on the test dataset. In [14], authors introduced the system identifier (SysID), a system for automatically classifying device features based on network data. They employed GA to identify key features in various protocol headers and then used ML classifiers for device identification with over 95% accuracy. In [15], authors presented a framework that extracts network flow characteristics to identify the type of traffic. The experimental analysis has shown a device-type recognition accuracy of 94.5%, traffic-type classification accuracy of up to 93.5%, and abnormal traffic detection accuracy of up to 97%. In [16], the authors proposed a spatiotemporal sensing approach with deep neural networks as a network traffic prediction method. This is crucial for traffic forecasting to include shorter-range dependent modelling. The proposed prediction approach has outperformed three current methods in simulation.

Alqudah and Yaseen [17] discussed different ML approaches for traffic projections. Rapid IoT traffic and AI development necessitate new methods for detecting intrusions, analysing virus activity, and categorizing IoT traffic. The proposed methods are able to achieve the predictions of dynamic traffic. In [18], the authors made use of reinforcement learning for the network traffic predictions. Markov decision process and Monte Carlo methods are used to predict network traffic. They evaluated the effectiveness of their mechanism using real network traffic. In [19], authors addressed wireless network traffic's spatiotemporal properties and constructed a recurrent neural network to predict network traffic. The experts not only have considered the long-term dependence on traffic flows but also the short term. In [20], the authors used optical data center networks and LSTM technique for traffic flow predictions. The proposed method has outperformed the existing traditional algorithms according to experimental results. In recent years, artificial neural network (ANN) and machine learning along with statistical analysis have been used in various areas of research such as medicine, engineering, mathematics, meteorology, neurology, and economics [21, 22].

With the great success of deep learning, researchers are exploiting the usage of deep NN algorithms for traffic prediction. However, the question about which type of deep neural network is best for traffic flow prediction remains unanswered. To overcome the gradient vanishing problem, certain RNN structures, such as LSTM and GRU, have been designed to overcome the problems of RNN-based models. Hence, we are also making use of the GRU-NN method in our proposed work to predict the traffic more accurately by overcoming the drawbacks of conventional deep learning models.

3. Prediction Framework for IoT Traffic

To predict the IoT-based traffic, this manuscript proposes a GRU-NN predictor based on collaborative transfer learning. The GRU-NN predictor works in three stages: data processing, training of the model, and transfer phase. The data processing stage assists in preprocessing of the data, and it converts the continuous data into discrete records to suffice the input needs of the GRU-NN model. The training phase is the most important phase of the GRU-NN predictor. In this manuscript, a GRU-NN model is proposed to train the model. The transfer phase is also a vital phase which transfers a huge amount of offline data to the training module to handle the issue of insufficient data of IoT traffic in an online mode. Finally, the GRU-NN traffic predictor is framed.

3.1. Data Preprocessing Phase. The data processing begins with the collection of the continuous data at regular time interval t and then converts the continuous stream of data into discrete chunks. The discrete data are distributed into a fixed time window of m size, and then the traffic data are acquired as $A = [a1, a2, a3, a4, \ldots\ldots, a-1, an]$. The data *'an'* of the last n time is the output B of the predictor. Next, the sequence of the data is distributed into the training test set, and then the dataset for the predictor training test is attained.

3.2. Model Building Module. In the traffic predictor, the second phase is of prime importance which builds the model. It begins with the designing of a single-layered GRU structure. The single-layered structure helps in reducing the time complexity as it takes lesser time in the optimization of the factors. The overall GRU-NN is a three-layered architecture, the very first layer represents the input layer, and the number of neurons in the input layer is equal to the dimension of the input IoT traffic. The 2nd layer is the hidden layer, and the number of neurons in this layer is decided by the empirical study of the obtained output. The last layer represents the output, and this layer finally predicts the output or predicts the traffic.

3.2.1. Training Module. The training of the model refers to the square loss function optimization. This training module minimizes the loss function value with the adjustment of the weight matrix. Generally, the gradient method is used for the optimization of the weight matrix, but this may result in a local optimal solution. We have used the ion filter method to escape from this problem of the local optimal solution.

3.2.2. Fine-Tuning Process of the Model. The fine-tuning of the network structure enhances the generalization ability of the model, resolves the problem of overfitting, and helps to minimize the model training time. In order to handle the issue of inconsistent and imbalanced data, normalization procedure takes place before the execution of the activation function. Although the imbalanced data distribution has been resolved in [12], these techniques are not capable to save the loss of data. Two learning parameters are introduced to handle this problem which are termed as β and γ.

The training of the GRU-NN method can be described as the optimization of GRU-NN parameter θ so that the variance between the actual and the predicted value of the method can be reduced as far as possible as shown in equations.

$$\theta = \operatorname{argmin}_{\theta}\frac{1}{N}\sum_{i=1}^{N}\operatorname{loss}\left(A_i, B_{i,}, \theta\right), \tag{1}$$

$$\operatorname{loss} = \frac{1}{N}\sum_{i=1}^{N}\left(A_i - \widehat{A}_i\right)^2, \tag{2}$$

where $A_1, B_1, A_2, B_2, \ldots, A_N, B_N$ are the training datasets and θ is GRU-NN's weighting parameter. The loss function of the GRU-NN is the mean square error, and $\widehat{A}_i$ is the forecasted value.

A dropout layer has been added before the hidden layer.

$$p_j^l \sim \text{Bernoulli } (p), \tag{3}$$

$$\widetilde{a}^l = p_j^l * a^l, \tag{4}$$

where p_j^l is the probability of Bernoulli as shown in equation (3) and is specifically designed for IoT traffic data, $\widetilde{a}^l$ is the randomly discarded value on the basis of input a^l with probability p_j^l as shown in equation (4), and the output is zero of discarded neurons. The normalization mean for a sample is represented as shown in the following equation:

$$\widehat{a}_i = \frac{a_i - \mu}{\sigma}, \tag{5}$$

where $a = \{a_1, a_2, \ldots, a_d\}$ is the IoT data, μ is the expected input of IoT data a, and σ is the standard deviation of the input data a. This procedure can resolve the issue of discrepancies in data, but the direct input can be represented in the following equation:

$$B_i = \gamma_i \widehat{a}_i + \beta_i. \tag{6}$$

The model training involves the optimization of the square loss function. This module optimizes the loss function value using the weight matrix. In our problem statement, the gradient technique is used to avoid the problem of local optima.

4. Results and Discussion

4.1. Experimental Results. This section provides insights into the evaluation metrics considered for the research study. In order to evaluate the prediction capability of the proposed GRU-NN method, the statistical error analysis techniques are utilized.

(1) Mean absolute error (MAE) is the average of all absolute errors, and the formula is as follows:

$$\text{MAE} = \frac{1}{n}\sum_{a=1}^{n}\left|z_a - z\right|, \tag{7}$$

where n depicts the number of errors, Σ depicts the summation of all values, and $|z_a - z|$ is the absolute errors.

The MAE score is presented in Figure 1 obtained by the proposed GRU-NN and other conventional techniques such as ARIMA (autoregressive integrated moving average), LSTM (long short-term memory), and VAR (vector autoregression). The results show that the proposed GRU-NN forecasts the IoT traffic with least error and highest accuracy, and it outperforms the traditional techniques such as ARIMA, VAR, and LSTM.

(2) Root mean square error (RMSE) shows the standard deviation of the forecasted errors, and the formula is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{a=1}^{n}\left(z_a - \hat{z}_a\right)^2}{N}}. \tag{8}$$

(3) N is the number of nonmissing data points, a represents the variable, z_a are actual observations of time series, and $\hat{z}_a$ represents predicted time series.

The RMSE scores are presented in Figure 2, and it can be observed that the proposed GRU-NN shows best results with respect to the RMSE score and outperforms the other three techniques taken up for the research study.

(4) Mean squared error (MSE) summarizes the prediction ability and forecast accuracy of the proposed GRU-NN model. It is calculated using equation (9) and is shown in Figure 3.

$$MSE = \frac{1}{n} * \sum(\text{actual} - \text{predicted})^2. \tag{9}$$

In order to exhibit the viability of the proposed GRU-NN method-based predictor in this paper, three comparative techniques are taken up which are benchmarked methods for forecasting the traffic. Compared to the VAR-based traffic prediction algorithm, GRU-NN performs very well as it has the ability to forecast by retaining relevant information in its layers. ARIMA is capable to process short-term time series, whereas the GRU-NN can consider long-term series also. Statistical continuity also shows that LSTM does not support nonlinear fitting capability. LSTM performs well in forecasting traffic, but sometimes, the relevant information is lost in the hidden layers. The proposed GRU-NN is well suited on time-series data and is able to control data of IoT-based network traffic very well. Both methods LSTM and GRU-NN perform well for forecasting the IoT traffic, but the GRU-NN gives more accurate results as compared to LSTM as shown in the results.

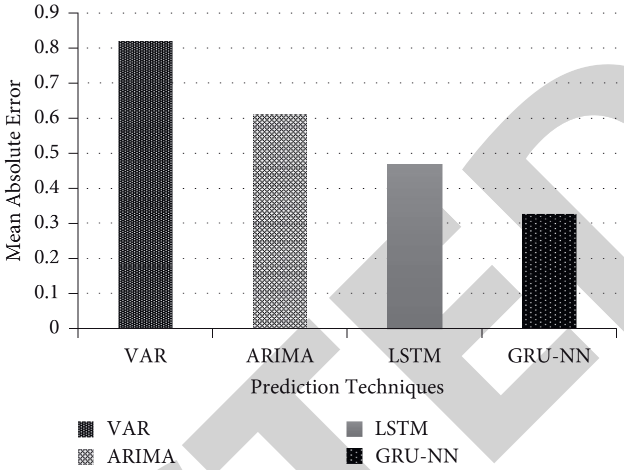


Figure 1: MAE score obtained by the proposed and other models.

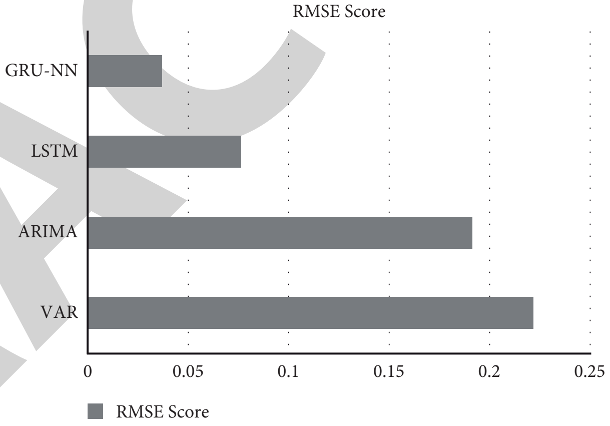


Figure 2: RMSE score obtained by the applied algorithms.

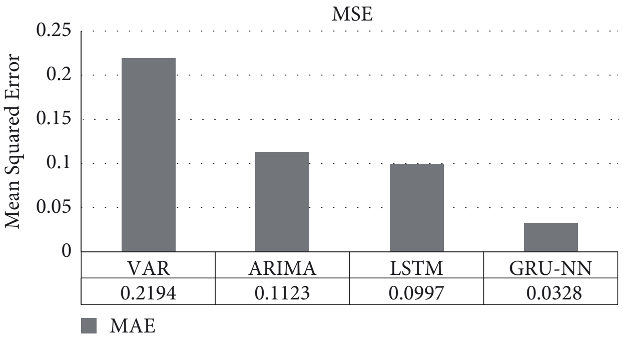


Figure 3: MSE scores obtained by the applied algorithms.

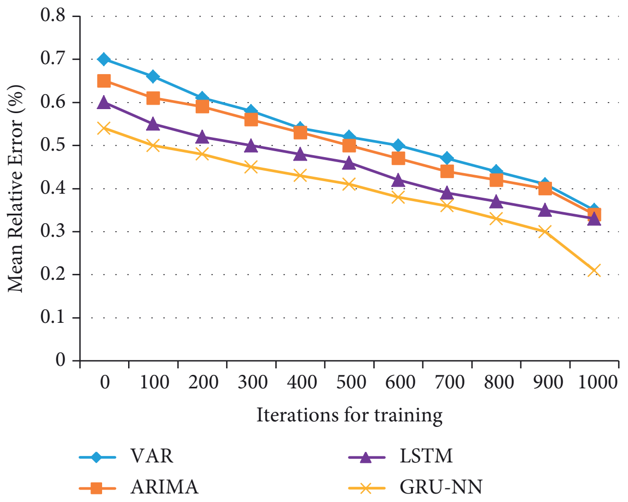


Figure 4: MRE values obtained by diverse techniques in defined iterations.

For verifying the space complexity and efficiency of convergence of the algorithms, the iterations for the training set are defined similarly for the algorithms considered for the research study, and it is observed from Figure 4 that the GRU-NN outperforms other techniques with respect to MRE scores. MRE is the ratio of the absolute error of a reading to the measurement being taken and is expressed in % as it has no units. However, in the beginning, the relative error of the proposed GRU-NN is higher, and it gradually decreases. The other techniques also behave in the same manner, but the GRU-NN shows the best performance with respect to the MRE scores.

5. Conclusions

In order to handle IoT traffic, it is inevitable to predict the traffic in advance for better utilization of resources and bandwidth. The IoT devices have attracted great attention in this decade, and traffic prediction is mandatory to enhance the channel capacity and to reduce the network latency. In this paper, the problem of IoT traffic prediction has been examined, and the GRU neural network-based solution has been proposed. Three well-established traffic predictor techniques have also been studied and considered for the comparative study. The proposed GRU-NN predicts the traffic accurately as depicted in results based on statistical performance evaluation metrics such as RMSE, MAE, and MSE. The advantage of the GRU-NN over LSTM is that it is capable of solving the problem of gradient disappearance and loss of information in hidden layers. The proposed GRU-NN memorizes the long correlation and other traffic characteristics of the IoT environment. The proposed GRU-NN outperforms ARIMA, VAR, and LSTM for predicting the IoT dynamic traffic. In a future study, a hybrid method for traffic forecasting will be researched upon, which may improve the performance efficiency of existing predictors by combining the characteristics of different methodologies.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] J. Mejia, A. Ochoa-Zezzati, and O. Cruz-Mejía, "Traffic forecasting on mobile networks using 3D convolutional layers," *Mobile Networks and Applications*, vol. 25, no. 6, pp. 2134–2140, 2020.

[2] A. R. Abdellah, O. A. K. Mahmood, A. Koucheryavy, and K. Andrey, "IoT traffic prediction using multi-step ahead prediction with neural network," in *Proceedings of the 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 1–4, Dublin, Ireland, October 2019.

[3] M. Lopez-Martin, B. Carro, and A. Sanchez-Esguevillas, "Neural network architecture based on gradient boosting for IoT traffic prediction," *Future Generation Computer Systems*, vol. 100, pp. 656–673, 2019.

[4] A. Mozo, B. Ordozgoiti, and S. Gómez-Canaval, "Forecasting short-term data center network traffic load with convolutional neural networks," *PloS One*, vol. 13, no. 2, Article ID e0191939, 2018.

[5] A. R. Abdellah and A. Koucheryavy, "Deep learning with long short-term memory for IoT traffic prediction," *Lecture Notes in Computer Science*, vol. 12525, pp. 267–280, 2020.

[6] A. R. Abdellah, O. Abdulkareem Mahmood, and A. Koucheryavy, "Delay prediction in IoT using machine learning approach," in *Proceedings of the 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 275–279, Brno, Czech Republic, October 2020.

[7] A. Essien and C. Giannetti, "A deep learning framework for univariate time series prediction using convolutional LSTM stacked autoencoders," in *Proceedings of the 2019 IEEE International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, pp. 1–6, Sofia, Bulgaria, July 2019.

[8] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: a survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[9] Y. Zhou and Z. Zhang, "Prediction of traffic flow based on deep learning," *International Journal of Advanced Computer Technology*, vol. 9, no. 2, pp. 05–11, 2020, Retrieved from https://ijact.org/index.php/ijact/article/view/42.

[10] X. Wang, S. Chen, J. Su, and M. Picone, "Real network traffic collection and deep learning for mobile app identification," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 4707909, 14 pages, 2020.

[11] F. Tang, Z. M. Fadlullah, B. Mao, and N. Kato, "An intelligent traffic load prediction-based adaptive channel assignment algorithm in SDN-IoT: a deep learning approach," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 5141–5154, 2018.

[12] M. Kaur and S. Kadam, "Bio-inspired workflow scheduling on HPC platforms," *Tehnički glasnik*, vol. 15, no. 1, pp. 60–68, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9809853, 1 page
https://doi.org/10.1155/2023/9809853

Retraction

Retracted: Edge Location Method for Multidimensional Image Based on Edge Symmetry Algorithm

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C. Li, “Edge Location Method for Multidimensional Image Based on Edge Symmetry Algorithm,” *Security and Communication Networks*, vol. 2021, Article ID 1326357, 11 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2021, Article ID 1326357, 11 pages
https://doi.org/10.1155/2021/1326357

Research Article

Edge Location Method for Multidimensional Image Based on Edge Symmetry Algorithm

Chen Li

College of Geophysics and Petroleum Resource, Yangtze University, Wuhan 430223, China

Correspondence should be addressed to Chen Li; lichen_me@163.com

Received 26 June 2021; Revised 20 August 2021; Accepted 31 August 2021; Published 15 September 2021

Academic Editor: Chinmay Chakraborty

The most basic feature of an image is edge, which is the junction of one attribute area and another attribute area in the image. It is the most uncertain place in the image and the place where the image information is most concentrated. The edge of an image contains rich information. So, the edge location plays an important role in image processing, and its positioning method directly affects the image effect. In order to further improve the accuracy of edge location for multidimensional image, an edge location method for multidimensional image based on edge symmetry is proposed. The method first detects and counts the edges of multidimensional image, sets the region of interest, preprocesses the image with the Gauss filter, detects the vertical edges of the filtered image, and superposes the vertical gradient values of each pixel in the vertical direction to obtain candidate image regions. The symmetry axis position of the candidate image region is analyzed, and its symmetry intensity is measured. Then, the symmetry of vertical gradient projection in the candidate image region is analyzed to verify whether the candidate region is a real edge region. The multidimensional pulse coupled neural network (PCNN) model is used to synthesize the real edge region after edge symmetry processing, and the result of edge location of the multidimensional image is obtained. The results show that the method has strong antinoise ability, clear edge contour, and precise location.

1. Introduction

The most basic feature of an image is edge, which is the junction of one attribute area and another attribute area in the image. It is the place where the regional attributes mutate. It is the most uncertain place in the image and the place where the image information is most concentrated. The edge of an image contains rich information. Edge widely exists between objects and background, objects and objects, and primitives and primitives, so it is an important feature of image segmentation [1]. There are three kinds of common image edges. The first one is step edge, which is from one gray level to another gray level much higher than it. The second one is roof edge, which gradually increases to a certain degree and then decreases. Another is line edge, whose gray level changes from one level to another and then returns. Edge location is one of the most basic contents in image processing and recognition. An image is an information system, and a lot of information is provided by its contour edge. Therefore, edge location plays an important role in image processing, and its positioning method directly affects the image effect [2]. Multidimensional images are not susceptible to external environmental impact and have more feature information. The recognition accuracy is not affected by single feature change, and it has natural symmetry [3]. Symmetry is a common feature of many natural objects and artificial phenomena. It is widely used in image processing to describe the shape features of objects. Symmetry detection plays an important role in object recognition and location, multidimensional object reconstruction, and other fields [4].

Image edge location methods mainly include the spatial moment location method and the subpixel location method, each of which has its own advantages and disadvantages. Spatial moment location is a common method in a fixed background. It can generally provide complete edge data, but its location effect depends on the merits of background model updating algorithm. It is particularly sensitive to scene changes, such as illumination and interference from external unrelated events [5]. The subpixel location method has strong adaptability to scene changes, but, generally, it cannot locate all the relevant edge pixels completely. It produces holes in the image entity, and it is easy to miss the location of some image edges [6]. In view of the shortcomings of the above two methods, this paper proposes an edge location method for the multidimensional image based on edge symmetry algorithm to improve the accuracy of edge location for the multidimensional image. This model was defined and tested for pictures in order to assess the performance of the suggested model. One of the test parameters was the number of iterations. The result is evaluated using the Pratt quality factor value. The suggested model's output is compared to the performance of the two approaches described above.

2. Edge Symmetry Algorithm for Locating Multidimensional Image's Edges

2.1. Edge Detection and Statistics. Firstly, the image is preprocessed to improve the image quality; secondly, the edge gradient in the edge region of the multidimensional image is calculated by using the vertical direction of the Sobel operator [7]; then, the candidate region is determined according to the characteristics of large jump and the large number of edge changes between the edge and the background.

2.1.1. Image Preprocessing

(1) Region of interest: in order to simultaneously detect a large range of image regions and reduce the computational complexity, the algorithm sets multiple regions of interest according to the number of pixels occupied by the image. The resolution of the image is 720 × 288, and a region of interest with a size of 200 × 50 is set in the three regions at the bottom of the image.

(2) Image preprocessing: in order to improve the accuracy of location, the image in the region of interest is preprocessed [8]. The color image is transformed into the gray image, and the noise is filtered by Gauss smoothing. The Gauss filter is a kind of linear smoothing filter which chooses weights according to the shape of Gauss function (i.e., normal distribution function). Gauss smoothing filter is very effective for removing noise which obeys normal distribution.

One-dimensional zero-mean Gauss function is

$$g_{(x)} = e^{-\left(x^2/2\sigma^2\right)}, \tag{1}$$

where the Gaussian distribution parameter σ determines the width of the Gaussian filter. For image edge processing, two-dimensional zero-mean discrete Gauss function is often used as the smoothing filter:

$$G(x, y) = Ae^{-\left(x^2+y^2/2\sigma^2\right)} = Ae^{-\left(r^2/2\sigma^2\right)}. \tag{2}$$

Sampling and quantifying the continuous Gauss distribution above and normalizing the template, the discrete template is obtained:

$$G^3 = \frac{1}{16}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}. \tag{3}$$

The Gauss filter can solve the problem of spatial distance weighting and pixel gradient. The pixel gradient reflects the image edge characteristics, which is very helpful for edge location of the multidimensional image.

2.1.2. Vertical Edge Detection. Vertical edge detection is applied to the preprocessed image. The essence of edge detection is to find the fast-changing region of brightness in the image, that is, the region whose first derivative of brightness is larger in magnitude than the specified range value. Because the vertical edge of the image has the strongest symmetry, in order to eliminate interference and reduce the amount of calculation, the proposed algorithm only processes the vertical component of the edge [9]. The vertical mask of the Sobel operator is used to digitally approximate the first derivative; that is, the vertical gradient of the image is calculated by formula (1) and presented as

$$\begin{aligned} G(x, y) = |&(f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1)) \\ &- (f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1))|, \end{aligned} \tag{4}$$

where $f(x, y)$ is the gray value of the pixel (x, y) and $G(x, y)$ is the approximate first derivative of the point in the vertical direction, i.e., the vertical gradient value.

2.1.3. Gradient Statistics. The vertical gradient values of each pixel are projected in the vertical direction:

$$B(i) = \sum_{j=0}^{H-1} G(i, j), \tag{5}$$

where H is the height of the detection block, B is the vertical superposition projection of the detection block, and i is the column position of the vertical superposition projection.

The average value of vertical gradient projection is

$$\overline{B} = \sum_{i=0}^{W-1} \frac{B(i)}{W}, \tag{6}$$

where W is the width of the region of interest and B is the mean value.

The mean value of vertical gradient projection is greater due to the large leap between edge and background, the significant number of edge modifications, and the quantity of vertical edges. In addition, the edge of the image is not susceptible to external environmental impact. Shadows, lights, and other interference factors cannot form a strong vertical gradient value, which will not have a great impact on the accuracy of location [10]. Therefore, when the mean value of vertical gradient projection exceeds a certain threshold, it can be determined as a candidate image region.

2.2. Symmetry Analysis. After edge identification and statistics, the symmetry of potential picture areas is examined. The image's symmetry axis direction has been established as the vertical direction. It is required to establish the symmetrical axis's position and quantify its symmetrical intensity [11]. The symmetry of the candidate image region's vertical gradient projection is examined to determine whether the candidate region is a true edge region.

2.2.1. Symmetry Analysis. The gradient vertical projection value obtained in the previous section is regarded as one-dimensional function $g(x)$, assuming that its symmetric axis coordinate is $X_S\,(W/2 \leq X_S \leq W - W/2)$ and W is the width of the symmetric region, which can be selected according to the width of the image [8]. Let $u = X - X_S$; $g(u)$ is a function with X_S as the origin of the coordinate. Since an arbitrary function $f(x)$ can be expressed as the sum of an odd function and an even function, as mentioned in formula (7), the odd function component and even function component of the definable function $g(x) = g(X_s + u)$ are, respectively,

$$\begin{cases} O(u, X_s) = \dfrac{g(X_s + u) - g(X_s - u)}{2}, \\[2ex] E(u, X_s) = \dfrac{g(X_s + u) - g(X_s - u)}{2}, \\[2ex] -W/2 \leq u \leq W/2. \end{cases} \tag{7}$$

The even function components are normalized to make the mean value as 0 by using

$$E_n(u, X_s) = E(u, X_s) - \frac{1}{W}\int_{-w/2}^{+w/2} E(u, X_s)\mathrm{d}u. \tag{8}$$

Then, the symmetry is measured by comparing the energy of odd component function with that of an even component function:

$$S(X_S) = \frac{\int_{-w/2}^{+w/2} E_n(u, X_s)^2 du - \int_{-w/2}^{+w/2} O(u, X_s)^2 \mathrm{d}u}{\int_{-w/2}^{+w/2} E_n(u, X_s)^2 du + \int_{-w/2}^{+w/2} O(u, X_s)^2 \mathrm{d}u}, \quad -1 \leq S(X_S) \leq 1, \tag{9}$$

where $S = 1$, and it is completely symmetric, and when $S = -1$, it is completely asymmetric.

2.2.2. Verification of Candidate Region. According to the size of the candidate region, the symmetric axis is searched by changing the X_s value of the symmetric coordinate axis in a certain range. In the search interval, X_s is taken as the coordinate of the symmetrical axis of the candidate region when the maximum value of the symmetrical evaluation function $S(X_S)$ appears. When $S(X_S)\rangle 0.5$ is located on the symmetrical axis, it is considered to be in line with the candidate region.

In addition, the vertical projection value of the edge should be the largest on both sides of the image edge, so locating the maximum value on both sides of the symmetric axis of the even function component $E(u, X_s)$ can determine the approximate left and right boundaries of the image edge. Since the edge width of image satisfies certain constraints [12–15], it can be further verified whether it is a real image edge.

2.3. PCNN Edge Location of Multidimensional Image. The real edge region processed by edge symmetry is synthesized by PCNN for edge location. The PCNN model plays an important role in image denoising, smoothing, segmentation, edge extraction, and feature extraction. As multidimensional images provide richer target information than two-dimensional images, the edge location of the multidimensional image has attracted more and more attention [16–19]. At present, PCNN can only deal with two-dimensional images directly, which have great limitations. This paper introduces vector matrix and multidimensional convolution (⊗) to extend PCNN, which is called the multidimensional PCNN model for short, and the application of PCNN is extended so that PCNN can directly realize multidimensional image's edge location [20–23].

Figure 1(a) represents the process of two-dimensional PCNN, and Figure 1(b) represents the multidimensional PCNN. It can be seen from Figure 1 that a multidimensional PCNN is equivalent to the edge location of PCNN image with three real edge regions. The multidimensional PCNN model satisfies the following formulas:

$$F_{ij}[n] = s_{ij}, \tag{10}$$

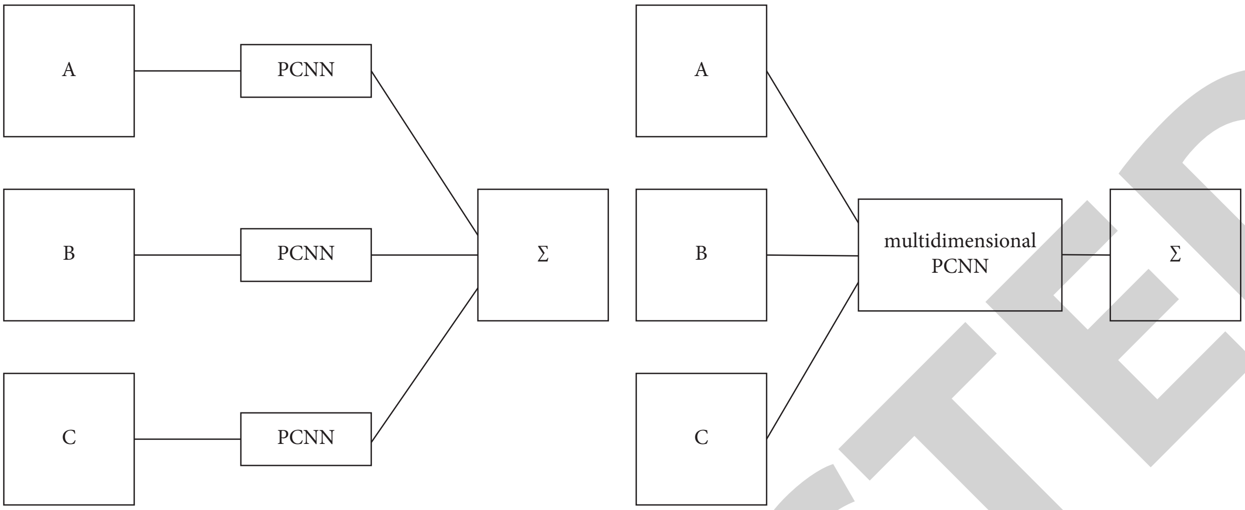
$$L_{ij}[n] = Y_{ij}[n-1] \otimes W, \tag{11}$$

$$U_{ij}[n] = F_{ij}[n-1]\left(1 + \beta L_{ij}[n]\right), \tag{12}$$

$$Y_{ij}[n] = \begin{cases} 1, & U_{ij}[n]\rangle E_{ij}[n], \\ 0, & \text{other}, \end{cases} \tag{13}$$

$$E_{ij}[n] = \exp(-\alpha_E)E_{ij}[n-1] + \nu_E Y_{ij}[n-1], \tag{14}$$

where *S, F, L, U, Y*, and *W* are vector matrices, s_{ij} is the external input stimulus signal (here is the gradient vector $(G_{ij}^a, G_{ij}^b, G_{ij}^c)$ of the input image at (i, j), $F_{ij}[n]$ is the nth feedback input of the (i, j)th image, $L_{ij}[n]$ is the nth connection input of the(i, j)th image, $U_{ij}[n]$ is the internal active item of the image, E_{ij} is the dynamic threshold, Y_{ij} is the output sequence of PCNN, W is the intensity constant of adjacent connected images in the connection domain, ν_E is the intrinsic potential in $E_{ij}[n]$, α_E is the attenuation time constant of $E_{ij}[n]$, n is the iteration number, which can be selected according to actual needs, and ⊗ is the multidimensional convolution; if convolution is the multidimensional convolution and conv2 is the two-dimensional



(a)

(b)

Figure 1: (a)Two-dimensional and (b) multidimensional PCNN processing process.

convolution, then $Y\,\mathrm{con}\,vnW = (Y_A\mathrm{conv}2W_A, Y_B\mathrm{conv}2\,W_B, Y_C\mathrm{conv}2W_C)$.

According to the principle of vector gradient and the definition of multidimensional PCNN, we can design an edge location algorithm for the multidimensional image based on the real-edge region obtained from edge symmetry. The flowchart of the algorithm is shown in Figure 2.

(1) Vector gradient operator DV is used to compute the gradient of the real-edge region processed by the symmetry of the input edge, which reflects the change of color information in the neighborhood of the pixel and provides the basis for PCNN edge location [24, 25].

(2) Initialization of PNCC: E_{ij} is the maximum value in gradient graph, which can suppress the ignition of small gradient pixels, and only large gradient can ignite. Y_{ij} is all zero.

(3) The vector gradient obtained in step 1 is input into the multidimensional PNCC model, and the threshold E_{ij} is attenuated to determine whether the value of the internal activity item is greater than the threshold value. If the value is greater than the threshold value, Y is set to 1, otherwise 0 is set.

(4) If the number of iterations is less than the prescribed maximum number of iterations, all the pixels will return to step 3 to continue iteration after judging, otherwise the iteration will end [26–28].

(5) Finally, the best processing effect is selected from the real edge regions and obtained when the number of iterations is small, and the final multidimensional image's edges are obtained by adding the three channel results of the multidimensional PCNN model. Since the threshold is attenuated from the maximum gradient value, when the iteration times are small, the small gradient is suppressed [29]. Only those large gradient values will ignite first, the sharp change of pixel points, namely, edge points, can be detected.

Because the convolution in formula (11) is multidimensional convolution and PCNN is vector operation, there is no need for three channels to run separately. Parallel operation reduces the operation time. Finally, the edge subgraphs of three channels are added together to obtain the final edge location results. The location results are determined by the output of all channels, which can locate the edges of multidimensional images more accurately.

3. Results

In order to verify the accuracy of edge location for multidimensional images, the proposed method is compared with the subpixel method and the spatial distance method.

The Pratt quality factor is a representative objective evaluation index of edge detection results and a more comprehensive evaluation parameter. Therefore, the Pratt quality factor is selected to objectively evaluate the performance of image's edge location of different edge location methods. At the same time, the vertical standard map is used to test. Noise and signal-to-noise ratio are added to analyze different methods under different iterations. The result of edge location of the multidimensional image is shown in Figures 3 and 4.

Figure 3 represents the variation of quality factor with respect to the signal-to-noise ratio (SNR) for single iteration, i.e., $N = 1$. The results show that the Pratt quality factor value of the proposed method is higher than that of the other two methods. It is clear from Figure 3 that the edge location effect of the proposed method is better than that of the other two methods, but the effect is not obvious.

Figure 4 represents the Pratt quality factor vs. signal-to-noise ratio (SNR) for three number of iterations, i.e., $N = 3$. The Pratt quality factor for the proposed method is superior to the other two methods. The infrared image House and

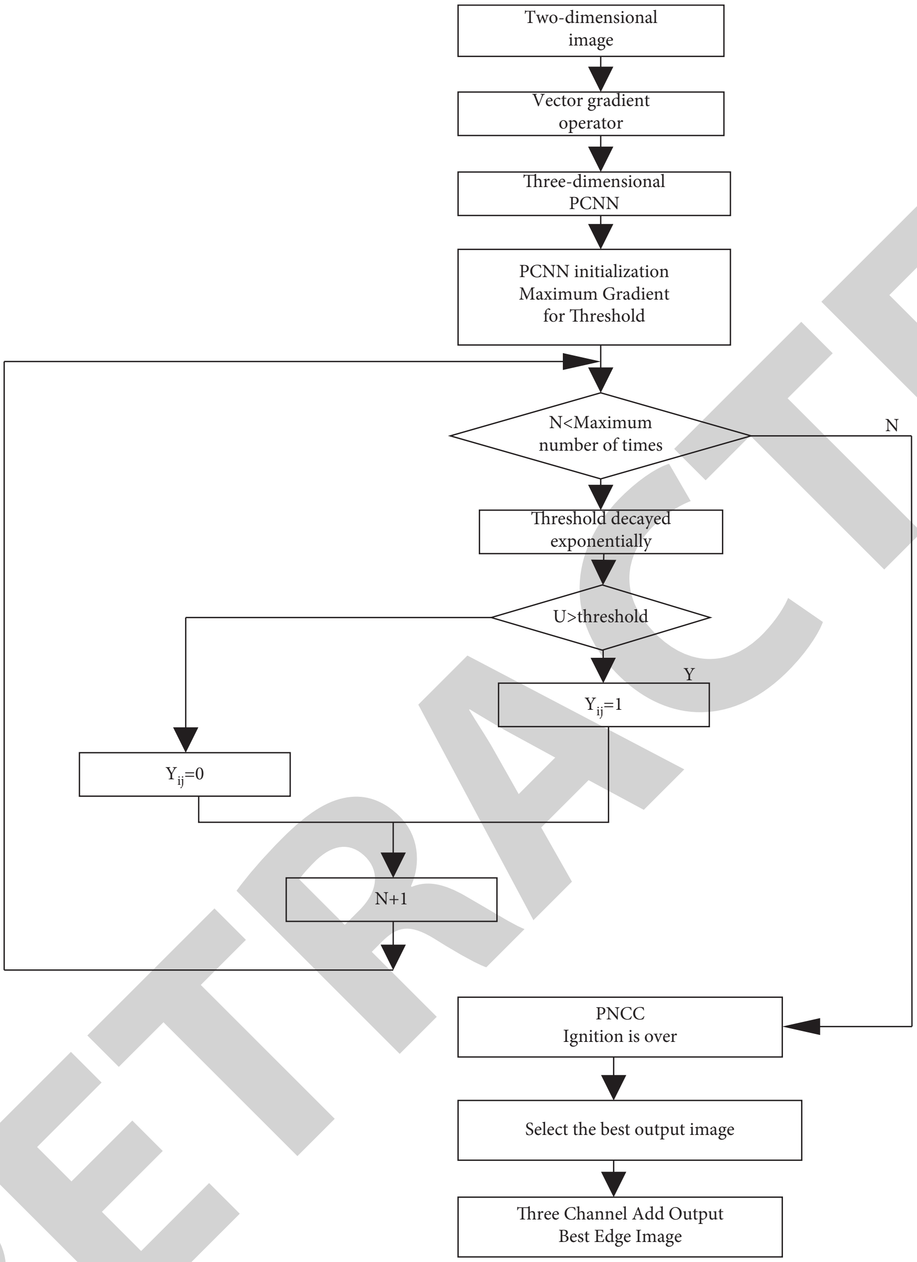

Two-dimensional image
Vector gradient operator
Three-dimensional PCNN
PCNN initialization Maximum Gradient for Threshold
N<Maximum number of times
N
Threshold decayed exponentially
U>threshold
Y
Y_{ij}=1
Y_{ij}=0
N+1
PNCC Ignition is over
Select the best output image
Three Channel Add Output Best Edge Image

Figure 2: Algorithm flowchart.

Lena are selected to simulate and compare the proposed method with the subpixel method and the space distance method. The simulation results are shown as follows.

As can be seen in Figures 5–14, the image edge image localized by the proposed method is better than the image edge localized by the nonenhanced edge image, the subpixel method, and the spatial moment method. The image processed by the proposed method not only has clearer layers between regions but also enhances the gray contrast on both sides of the edge. The simulation results also show that the edge location of this method is more precise.

In order to verify the edge location effect of this method in multidimensional images, two multiband remote-sensing images are selected for edge location, and the location results are shown in the figures. Figure 15 is the IKONOS satellite image, with the size of 544 × 342, and membership function parameters of $a = 23$, $b = 159$, and $T = 0.85$. Figure 16 is the IKONOS pseudocolor remote-sensing image, with the size of 600 × 450, and membership function parameters of $a = 11$, $b = 173$, and $T = 0.79$. The subpixel method, the spatial moment method, and the proposed method describe the results of edge location in Figure 15, by using Figures 17–19. The subpixel method, the spatial moment method, and the proposed method describe the results of edge location in Figure 16, by using Figures 20–22.

According to the above figures, we can see that the multidimensional image's edges located by the subpixel method and the spatial moment method contain many false

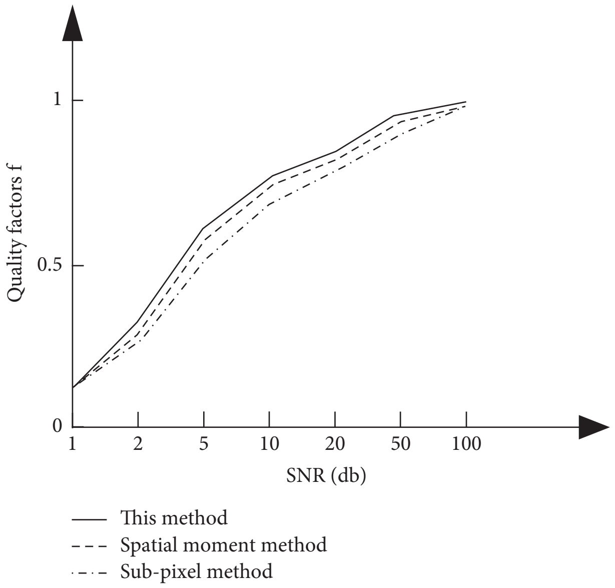


Figure 3: Evaluation results, $N = 1$.

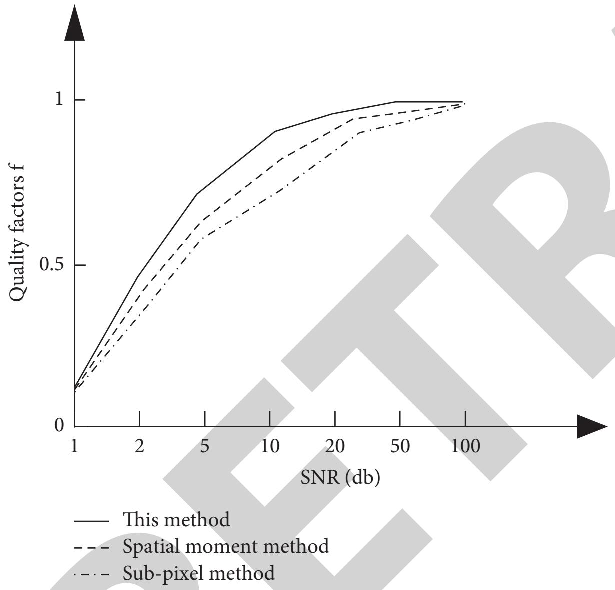


Figure 4: Evaluation results, $N = 3$.

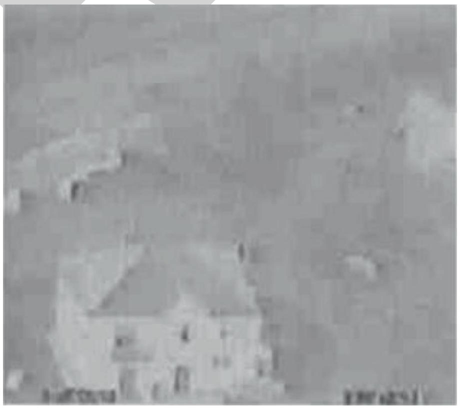


Figure 5: House original.

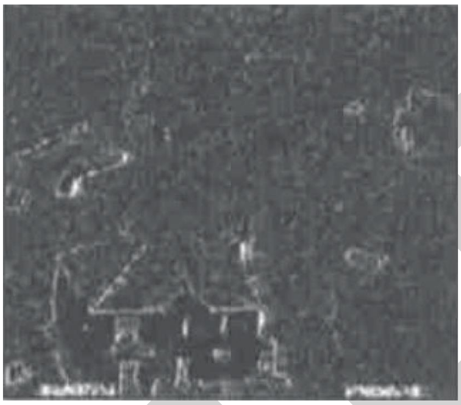


Figure 6: Edge diagram without enhancement ($N = 0$).

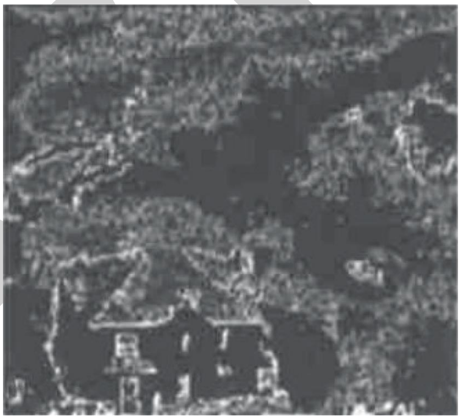


Figure 7: The method edge location graph in this paper.

Figure 8: Subpixel method edge map.

edges and secondary edges, and the effect is not ideal. However, the method in this paper is accurate and effective in multidimensional image's edge location.

In order to verify the location accuracy of the proposed method, the test image is selected as the multidimensional image of the microgear scanned by the multidimensional profiler. Considering the smoothing effect of the actual optical system and sensors on the image, the multidimensional image of the microgear is smoothed by 3×3 pixel

Figure 9: Spatial moment method edge location map.

Figure 10: Original map of Lena.

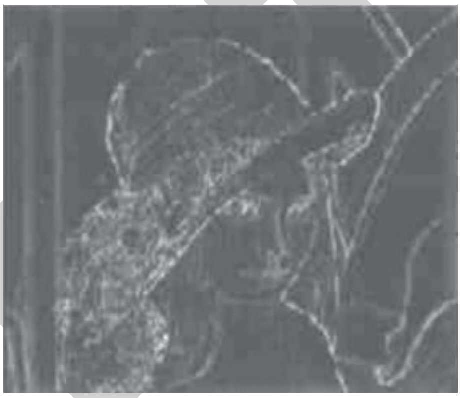


Figure 11: Edge diagram without enhancement ($N = 0$).

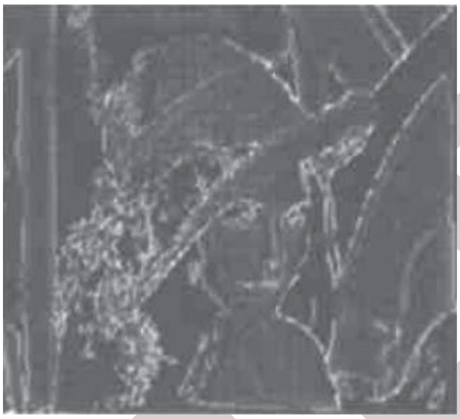


Figure 12: The method edge location diagram in this paper.

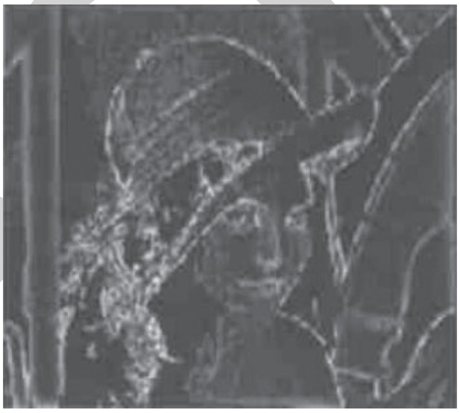


Figure 13: Subpixel method edge map.

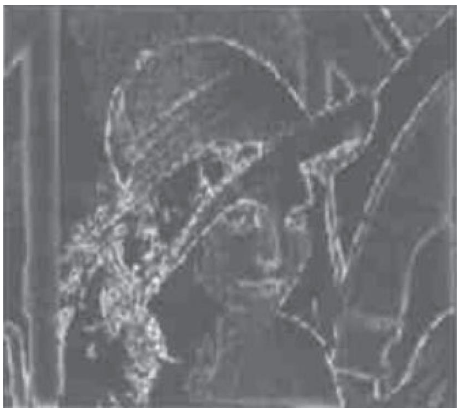


Figure 14: Spatial moment method edge mapping.

neighborhood filter, and then, the standard multidimensional test image for experiment is generated. Multidimensional profilometer is a precision-measuring instrument. Its dimension measuring accuracy can reach the level of nm. In this paper, the measured data are taken as actual data. Tables 1–3 are the results of edge location of multidimensional image of microgears by three methods in simulation experiments.

By comparing and analyzing Tables 1–3, the average error of the proposed method is 2.4 μm when locating the multidimensional image edge of the microgear, while the average error of the subpixel method and the spatial moment method is 18.2 μm and 13.4 μm when locating the multidimensional image edge. Compared with the subpixel method and the spatial moment method, the proposed method is more accurate than the subpixel method and the

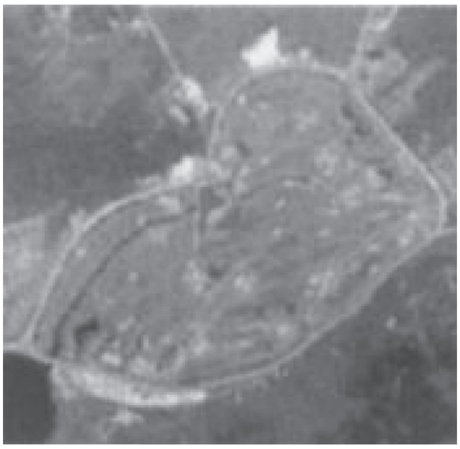


Figure 15: Original image of IKONOS satellite image.

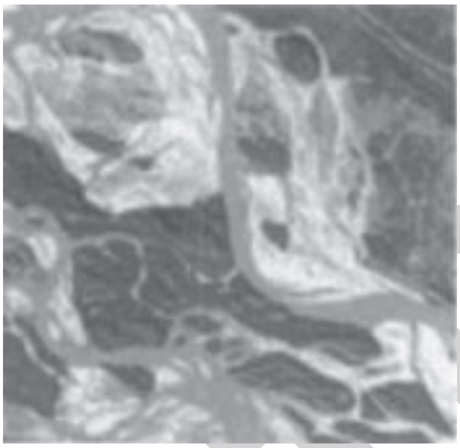


Figure 16: Original image of IKONOS false color remote-sensing images.

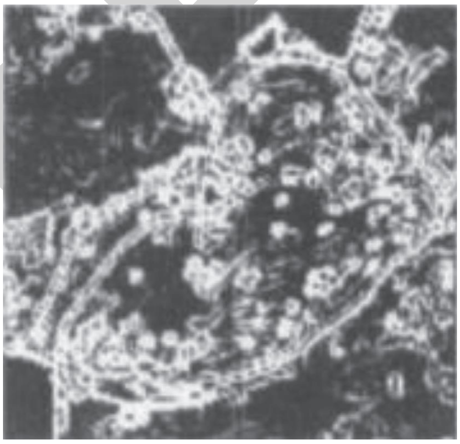


Figure 17: Subpixel method edge positioning.

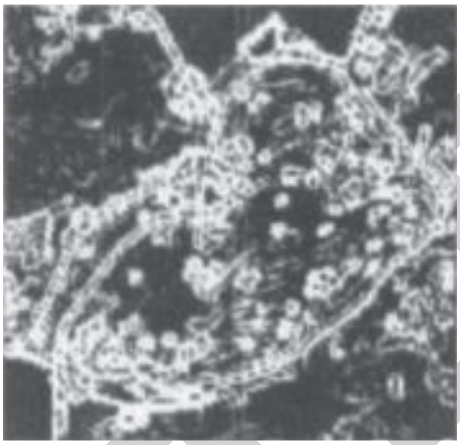


Figure 18: Spatial moment method edge positioning.

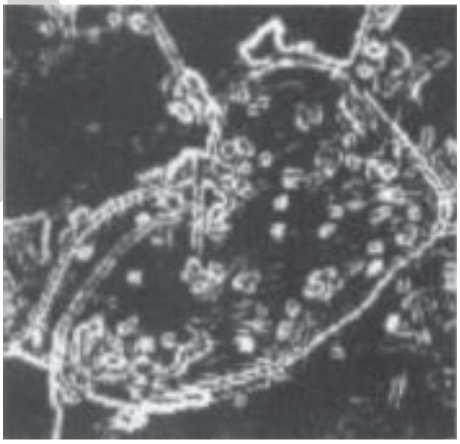


Figure 19: Edge positioning of this method.

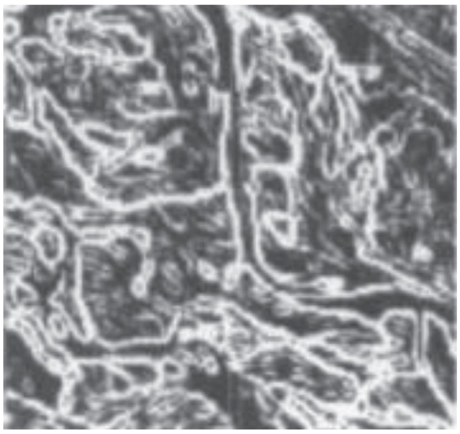


Figure 20: Subpixel method edge positioning.

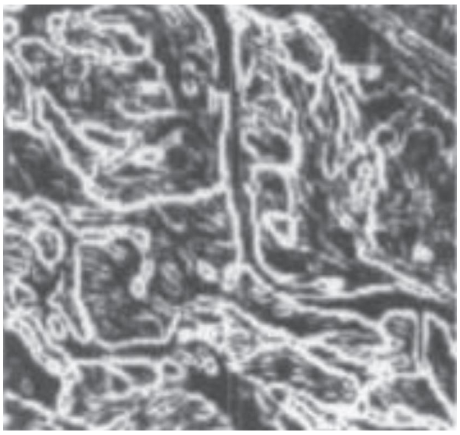


Figure 21: Spatial moment method edge positioning.

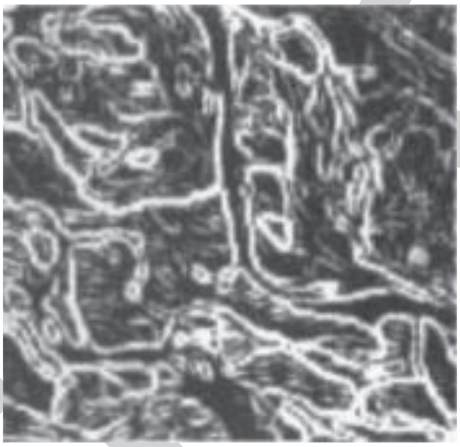


Figure 22: Edge positioning of this method.

Table 1: The method of edge positioning of the multidimensional image of the microgear.

Label point	Actual distance (μm)	This method distance (μm)	Error against actual value (μm)
D1	749	750	1
D2	752	750	2
D3	741	745	4
D4	771	773	2
D5	761	765	4
D6	1954	1956	2
D7	1952	1949	3
D8	1987	1986	1
D9	1961	1964	3
D10	1974	1972	2

spatial moment method by 15.8 μm and 11 μm. It can be seen that the proposed method can locate the multidimensional image edge accurately.

4. Discussion

In view of the above analysis and discussion, this paper proposes an edge location method for the multidimensional image based on edge symmetry, which can process the multidimensional image with strong antinoise and can locate the multidimensional image's edge more precisely and clearly, and the effect is obvious. The analysis is from the following three aspects:

(1) In this paper, a Gauss filter is introduced to preprocess multidimensional images. The problem of spatial distance weighting and pixel gradient is well solved by the Gauss filter. The pixel gradient reflects the image edge features, which is very helpful for multidimensional image's edge location.

Table 2: Subpixel method for edge positioning of microgear multidimensional image.

Label point	Actual distance (μm)	This method distance (μm)	Error against actual value (μm)
D1	749	733	16
D2	752	739	13
D3	741	727	14
D4	771	759	12
D5	761	746	15
D6	1954	1941	13
D7	1952	1940	12
D8	1987	1975	12
D9	1961	1946	15
D10	1974	1962	12

Table 3: Spatial moment method for edge positioning of microgear multidimensional image.

Label point	Actual distance (μm)	This method distance (μm)	Error against actual value (μm)
D1	749	767	18
D2	752	771	19
D3	741	762	21
D4	771	787	16
D5	761	779	18
D6	1954	1970	16
D7	1952	1969	17
D8	1987	2005	18
D9	1961	1980	19
D10	1974	1994	20

(2) In this paper, we analyze the symmetric axis position of candidate image region and measure its symmetry intensity. Then, the vertical gradient projection symmetry of the candidate image region is analyzed to verify whether the candidate region is a real edge region. The symmetry analysis of the multidimensional image is carried out. The accuracy of image's edge location and robustness of the algorithm are effectively improved by means of the verification of vertical gradient projection's mean value, symmetry detection, and width constraint.

(3) After introducing vector matrix and multidimensional convolution into PCNN, this paper generalizes PCNN, which is called the multidimensional PCNN model for short, and expands the application of PCNN so that PCNN can directly realize multidimensional image's edge location. The multidimensional PCNN model is used to synthesize the real-edge region after edge symmetry processing, and the result of edge location of the multidimensional image is obtained.

5. Conclusions

The proposed method uses the Gauss filter to preprocess the multidimensional image, which can not only suppress noise but also better preserve image edge and high-frequency detail information so that the edge image outline is clear, the edge position is precise, and the error is small, which is of great benefit to image location research. Moreover, the real edge region after edge symmetry processing is synthesized by using the multidimensional PCNN model to obtain accurate edge location results of the multidimensional image. The experimental results show that the proposed method is superior to the subpixel method and the spatial distance method in terms of noise immunity and multidimensional image's location accuracy and clarity. The location accuracy of the multidimensional image of microgears by using the proposed method is 2.4 μm on an average, which is more accurate than the subpixel method and the spatial moment method with 15.8 μm and 11 μm. The results for the proposed method are superior to the conventional methods; also, as the number of iterations increases, the quality of results also increases. It can be seen that the proposed method has clear edge location and high accuracy in the multidimensional image.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] Z. Wang, Z. Tang, and X. Zhang, "Reflection symmetry detection using locally affine invariant edge correspondence," *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1297–1301, 2015.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9758612, 1 page
https://doi.org/10.1155/2023/9758612

Retraction

Retracted: Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] A. Mehbodniya, I. Alam, S. Pande et al., “Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques,” *Security and Communication Networks*, vol. 2021, Article ID 9293877, 8 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2021, Article ID 9293877, 8 pages
https://doi.org/10.1155/2021/9293877

Research Article

Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques

Abolfazl Mehbodniya,[1] **Izhar Alam**,[2] **Sagar Pande**,[2] **Rahul Neware**,[3] **Kantilal Pitambar Rane**,[4] **Mohammad Shabaz**,[5,6] **and Mangena Venu Madhavan**[2]

[1]*Kuwait College of Science and Technology (KCST), Doha, Area, 7th Ring Road, Kuwait*
[2]*School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India*
[3]*Department of Computing, Mathematics and Physics, Høgskulen på Vestlandet, Bergen, Norway*
[4]*KCEs COEM Jalgaon, Maharashtra, India*
[5]*Arba Minch University, Arba Minch, Ethiopia*
[6]*Department of Computer Science and Engineering, Chandigarh University, Ajitgarh, India*

Correspondence should be addressed to Mohammad Shabaz; mohammad.shabaz@amu.edu.et

Received 28 July 2021; Revised 21 August 2021; Accepted 23 August 2021; Published 11 September 2021

Academic Editor: Chinmay Chakraborty

Healthcare sector is one of the prominent sectors in which a lot of data can be collected not only in terms of health but also in terms of finances. Major frauds happen in the healthcare sector due to the utilization of credit cards as the continuous enhancement of electronic payments, and credit card fraud monitoring has been a challenge in terms of financial condition to the different service providers. Hence, continuous enhancement is necessary for the system for detecting frauds. Various fraud scenarios happen continuously, which has a massive impact on financial losses. Many technologies such as phishing or virus-like Trojans are mostly used to collect sensitive information about credit cards and their owner details. Therefore, efficient technology should be there for identifying the different types of fraudulent conduct in credit cards. In this paper, various machine learning and deep learning approaches are used for detecting frauds in credit cards and different algorithms such as Naive Bayes, Logistic Regression, K-Nearest Neighbor (KNN), Random Forest, and the Sequential Convolutional Neural Network are skewed for training the other standard and abnormal features of transactions for detecting the frauds in credit cards. For evaluating the accuracy of the model, publicly available data are used. The different algorithm results visualized the accuracy as 96.1%, 94.8%, 95.89%, 97.58%, and 92.3%, corresponding to various methodologies such as Naive Bayes, Logistic Regression, K-Nearest Neighbor (KNN), Random Forest, and the Sequential Convolutional Neural Network, respectively. The comparative analysis visualized that the KNN algorithm generates better results than other approaches.

1. Introduction

The popularization of credit cards is across many fields, and healthcare is one among them. Because of credit cards, the online transaction has become more convenient and more accessible. However, fraud transaction impacts the massive loss of capital every year which might increase in the coming year. The system for detecting the fraud might be composed of a manual process and the expertise algorithm for detecting the fraud automatically. The automatic operation can be based upon all previous ways of fraud transactions happened. The manual method is estimated by different fraud investigators who check the separate transaction and generate binary feedback on every transaction. Fraud cases in the transaction are the primary barrier while enhancing the e-commerce and also cause a massive loss in the economy. Hence, detection of fraud is essential while doing transactions in an online environment.

Detection of fraud is the process of analyzing the behavior of card holders' transactions to know whether the conducted transaction is genuine. Frauds in credit cards signify the illegal use of information in credit cards and completing a transaction. While transacting physically, the involvement of credit card is there while the digital transaction is conducted utilizing the Internet or a telephone as information such as card number, its verification number, and its expiry date is collected through different means. Commonly, two different methodologies are conducted for anomaly detection in a transaction that has been conducted digitally. First, classification is used for determining whether the conducted transaction is genuine or fraudulent. Such approaches help identify the aforementioned types of conducted fraud, which helps construct the different models based upon all earlier patterns of fraud. Detection of the anomaly was conducted by the comparative analysis of data based upon the historical data of the transaction and the newly conducted transaction. It helps to identify all the possible potential of fraud transaction as fraud transaction shows deviation in its behavior from the average transaction. However, detection of fraud through anomaly requires a massive amount of successive data of different behaviors of average transaction of that cardholder. Different frauds in a credit card can be categorized as fraud in external card or inner card. Fraud in inner card happens due to commitment of false identity between the bank and the cardholders, and fraud in external card includes the usage of a stolen credit card to withdraw the cash by dubious means. However, different expertises use different computational methodologies for detecting the frauds in credit cards. Credit card fraud detection is associated with many challenges, such as dynamic or the fraudulent behavior of credit cardholders. Such kinds of activities can be identified using the popular technology called artificial intelligence through machine learning and deep learning algorithms. In particular, in this scenario, one needs to identify whether the cardholder is genuine or fraudulent, i.e., classifying the cardholders. Classification of related applications can be made through some of the ML (machine learning) algorithms such as KNN (K-Nearest Neighbor), Random Forest, Decision Tree, Logistic Regression, Naïve Bayes, and Neural Networks.

This paper consists of comparative analysis conducted between sequential convolutional neural networks and the many machine learning algorithms such as KNN (K-Nearest Neighbor), Random Forest, Decision Tree, Logistic Regression, and Naive Bayes. This paper enhanced the handling of the massive amount of imbalanced data collected from different frauds happened in credit cards, and the dataset is publicly available. In this paper, the main contribution can be summarized as dealing with the different problems related to fraud detection with the help of different machine learning approaches, and at last, from the obtained result, some suggestions and the future work related to detecting the fraud in credit cards are presented.

2. Related Work

This section reviews different fraud detection technologies with a sequential model and the different machine learning approaches. Many credit card financial applications with their transaction history are reviewed. Classification of different transactions related to a credit card mainly falls under the problem of binary classification as it will be a legitimate transaction (false class) or a genuine transaction (true class).

Awoyemi et al. in 2017 [1] investigated severely distorted credit card fraudulent information; this research analyzes the efficiencies of various methodologies such as Naive Bayes, KNN, and Logistic Regression. Credit card transaction information-based data including 284,807 transactions were gathered from European customers. On the distorted information, a combination strategy of undersampling and oversampling is used. The original and preprocessed data are subjected to three procedures. Python is used to carry out the task. The findings reveal that Naive Bayes, K-Nearest Neighbor, and Logistic Regression classifiers have an optimum accuracy of 97.92%, 97.69%, and 54.86%, respectively. KNN outperforms Naive Bayes and Logistic Regression methods, according to the comparison findings. Dal Pozzolo et al. in 2017 [2] proposed three key contributions. First, with the aid of their research assistance, the authors offer a formalization of the fraud-identification issue that accurately reflects the working circumstances of FDSs that monitor enormous flows of credit card transactions daily. The authors also showed how to utilize the most relevant evaluation metrics for fraud identification. Second, the authors devised and tested a unique learning approach for dealing with class imbalance, idea drift, and verification delay. Third, the authors illustrated the influence of class imbalance and idea drift in a real-world information stream with more than 75 million transactions permitted over three years in their studies. To train the behavior characteristics of regular and anomalous transactions, two types of random forests are employed. The framework proposed by Xuan et al. in 2018 [3] compared and analyzed the effectiveness of various random forests with various classification models in terms of credit fraud identification. The data for these tests came from a Chinese e-commerce firm. To include transactional sequences, Jurgovsky et al. in 2018 [4] framed the fraud identification issue as a sequence classification job in their article and used long short-term memory networks. In addition, the system incorporates cutting-edge attribute aggregation techniques and reports the framework findings using standard retrieval measures. The LSTM increases identification accuracy on offline transactions where the cardholder is present physically at merchants when compared to a benchmark Random Forest classifier. Manual attribute aggregation techniques are beneficial to both sequential and nonsequential learning systems. Following a review of true positives, it was discovered that both techniques tend to detect distinct types of frauds, indicating that the two should be used together.

The study by Varmedja et al. in 2019 [5] demonstrated several methods for identifying transactions as fraudulent or legitimate. The dataset utilized in the study was the credit card fraud identification dataset. The SMOTE method was employed to oversample the dataset since it was highly unbalanced. In addition, attributes were chosen and the dataset was divided into two fragments: training data and test data. Logistic Regression, Random Forest, Naive Bayes, and Multilayer Perceptron were the technologies utilized in the research. The study demonstrates that each technology is capable of detecting credit card fraud with high accuracy. The developed framework may be used to identify additional anomalies. Credit card fraud identification systems that utilize supervised learning methodologies rely on the idea that fraudulent tendencies may be learned from an examination of prior transactions.

Nevertheless, the process gets complicated when it must account for modifications in customer behavior and criminals' capacity to develop new fraud patterns. Unsupervised learning methodologies can aid fraud identification models in detecting abnormalities in this situation. Carcillo et al. in 2019 [6] offered a hybrid approach for improving fraud identification accuracy by combining supervised and unsupervised methodologies. On a real, labeled credit card fraud identification dataset, unsupervised anomaly ratings generated at various degrees of granularity are analyzed and evaluated. The combination is effective, as evidenced by experimental findings, and improves identification accuracy. Machine learning techniques are employed to identify credit card fraud in the research proposed by Randhawa et al. in 2018 [7]. First, conventional methodologies are utilized. After that, hybrid approaches based on AdaBoost and popular voting are utilized. A publicly accessible credit card dataset is utilized to test the framework's effectiveness. The data are then evaluated using a real-time credit card dataset from a financial organization. In addition, distortion is introduced into the data samples to test the techniques' resilience. The experimental findings show that the popular voting approach detects credit card fraud instances with a high degree of accuracy.

De Sá et al. in 2018 [8] proposed the Fraud-BNC methodology to identify credit card fraud issues. The proposed methodology is based on the Bayesian network classification model. Fraud-BNC was created naturally using a dataset from PagSeguro, Brazil's most prominent online payment platform, and evaluated alongside two cost-sensitive categorization methods. The acquired findings were compared to seven other methodologies and assessed for the data classification issue and the methodology's financial efficiency. Fraud-BNC emerged as the most robust methodology for achieving a good balance between the two points of view, increasing the existing organization's financial efficiency by up to 72.64%. A credit card fraud identification model was created by Sailusha et al. in 2020 [9] to identify fraudulent actions. The goal of this research is to concentrate on machine learning methodologies. The Random Forest methodology and the AdaBoost methodology were utilized. The two methodologies' accuracy, precision, recall, and $F1$-score are utilized to compare their outcomes. The confusion matrix is utilized to generate the ROC curve. The performance metrics such as accuracy, precision, recall, and $F1$-score of these two methodologies were compared. The methodology with the best performance metrics is deemed the best methodology for identifying fraud.

Economic fraud has proven to be a threat and has had a significant influence on the financial system. Data mining is one of the approaches that has proven effective in identifying credit card fraudulence in online transactions. Credit card fraudulent identification has proven difficult due to two issues: the characteristics of fraudulent and regular behavior vary over time and the datasets utilized are highly biased. The framework proposed by Bagga et al. in 2020 [10] intended to compare the efficiency of various methodologies such as Logistic Regression, Naive Bayes, Random Forest, KNN, AdaBoost, Multilayer Perceptron, Pipelining, and Ensemble Learning on the information of credit card fraudulence. The variables utilized and the approach employed to identify fraud impact the effectiveness of fraud identification.

2.1. Credit Card Fraud. The comprehensive analysis of different technologies related to fraud detection is essential while solving the different problems related to detecting fraud in credit cards. The most popular algorithm for detecting frauds in credit cards is inspired by nature. Application fraud relates to the criminal who owns a credit card from different issuing companies by spreading false data related to the cardholder [11]. In behavior frauds, the criminal thieves the detail related to the account and the password related to that account and uses that for withdrawing the money. Credit card fraud is more accessible as more money can be earned with less amount of risk in less duration of time.

Recently, many commercial banks adopted the method of fraud detection based upon the behavior related to the cardholder. Mostly, the fraud detection works upon the cardholder behavior pattern of using the card and relates all the transactions based upon the pattern for detecting the unknown transaction [12]. The sequence pattern of credit card transactions mainly relates to the Hidden Markov Model (HMM), which identifies the effectiveness based on credit card fraud [13]. Initially, the HMM is trained with a typical pattern of transaction of the cardholder. Then, when a so ever transaction happens, the new transaction is compared with the pattern of the trained model and if it is denied by the HMM [14], then it signifies that there is a fraud transaction. There is also a two-level sequence alignment technique where both anomaly detection and misuse sequence detection are combined. Here, in these models, profile analyzers were implemented for analyzing and determining the typical pattern between all the transaction sequences related to the cardholder with the past sequence of transactions. Then, the profile analyzer detects the unusual transaction that happened based upon the past and possible transactions and finally states whether the happened transaction is genuine or fraud. Most of the applications related to e-commerce use the signature-based technique for making the deviation related to user's behavior and consequently generalized all the potential behaviors of the fraud [13]. However, mostly they rely upon the clickstream of the signature which used multiple features of the transaction as it generates better results than a single transaction feature. The aggregating profile method exploits the pattern inherent concerning the transaction in time series which detects the fraud of all the transactions online at the end of a particular duration [15]. Here, they evaluate the data with various techniques such as Random Forest, Logistic Regression, and Support Vector Machine to predict the different frauds related to the credit card with the aggregation technique. However, this aggregation method fails to detect the real-time fraud that happens in the transaction with the credit card.

2.2. Feature Selection. The fraud detection system basis relies upon the behavior analysis of the cardholder. The profile of total expenditure is analyzed with help of optimal variable selection which focuses on the unique behavior of the transaction done through credit cards. The variable compares the current transaction with all the past transactions through which it has been trained. It falls under the following five different variable types: statistics of all transactions, merchant statistics, regional statistics, number of transactions, and the statistics related to the amount of transaction. Thus, through optimal variable selection, both the legitimate and the fraudulent profile can be separated easily, which helps distinguish between the transactions and enhances the system for detecting the frauds related to a credit card [15].

Currently, payment through both online and offline modes has become more common using the credit card. Hence, the rate of fraud accelerates, which brings a massive loss for financial and e-commerce companies. Fraud detection through the traditional method consumes a lot of time; thus, it needed some artificial intelligence models for detecting and tracking out the fraud in credit cards [16]. These techniques of intelligence hold many techniques based upon computational intelligence. The fraud detection system is based upon the supervised and the unsupervised learning method. The fraud detection through the supervised technique depends on the transaction based on fraudulent and legitimate and then newly occurred transaction classified based upon the learned model, while in an unsupervised model of fraud detection, the transactions that lie in outliers are the mainly considered transactions related to the fraud. The algorithms such as backpropagation of error signal with its forward pass and backward pass are implemented for fraud detection [17].

2.3. Comparative Studies. It includes the study of different related issues associated while detecting the frauds in credit cards. Comparative studies help in investigating different credit card-related fraud and nonfraud-related transactions, leading to better accuracy with great learning of an algorithm. The result visualized from the original dataset which leads towards training data is balanced with the help of a metalearning classifier which enhanced the performance of the model. Naive Bayes and Logistic Regression are compared in [18].

In minimum cases, it was observed that the performance of Logistic Regression is less than that of Naive Bayes. However, such a scenario is observed in data with fewer attributes and a small dataset [19]. Three classification methods (Logistic Regression, Decision Tree, and Neural Networks) are compared and tested for fraud detection applicability. The result visualized that the Neural Network classification technique generates better results compared to two other algorithms [20]. The Bayesian learning and the theory of Dempster–Shafer are fused for investigating frauds in credit cards, and it is observed that it has nearly 5% of false-positive rate [21]. Support Vector Machines with Decision Tree are investigated for detecting the fraud, and the result visualized that the classifier of Decision Tree outperforms SVM approaches [22]. The performance of Logistic Regression is evaluated with different approaches of data mining such as Random Forest and Support Vector Machine, and the result visualized that the performance of Logistic Regression is in undersampling level while the performance of the SVM trends to enhance in training data with the lower proportion. In paper [23], there is a comparison between the different classification models such as Logistic Regression and different artificial neural networks are developed to train and test on a dataset of fraud detection with highly skewed data [22]. Its results visualized that the artificial neural network performs better than the Logistic Regression for investigating the fraud related to the credit card. Classifier with Logistic Regression overfits the data while training due to insufficient data, which is a significant issue that causes a fall in its accuracy [24].

The classifier techniques such as Naive Bayes, Neural Network, and Decision Tree are trained and tested. It was observed that a huge database classifier such as Neural Network generates better results than another algorithm [25]. However, usually training and testing the huge dataset with a Neural Network consume a lot of time. Classifiers such as Bayesian take minimum time for training, but it is suitable for the lower or average data size [26]. The problem related to both classification and regression can be solved using a Support Vector Machine by arranging the sample to the category or many classifiers of binary-linear that consist of the nonprobabilistic sample [27]. In the probabilistic sample, the HMM is mainly used for representing different models of classification and regression. In sequential data, the HMM is used for learning succession patterns in abnormal and standard data. The likelihood transaction is used to generate a score for detecting the anomaly [28]. The recurrent neural network (RNN) lies in a nonprobabilistic model. Discriminatively, the RNNs are trained to predict the label of transactions and later generate the sequence of transactions for detecting fraud in credit cards [29]. Scalar variables are linked with Linear Regression by locating the observed data in the linear equation modeled by the function of linear predictor and unknown parameters calculated from the fraud detection data [11,30]. The summary of different machine learning techniques and their limitations is given in Table 1.

3. Experimental Setup and Methods

This section explains using a dataset in the experiment and different deep learning and machine learning classifiers such as Logistic Regression, Naive Bayes, Decision Tree, KNN, and the sequential model. All these algorithms perform different stages before generating the classifier such as data

Table1: Limitations of machine learning techniques.

Model	Strength	Limitations
Bayesian	Provide better results in problems of binary classification and suitable for analyzing the real-time data	Required better detection related to the abnormal and expected behavior of fraud cases
Neural Network	Suitable for problems related to binary classification, mostly used for detecting the fraud	Required huge computation, can be denied for real-time operation, and retraining is essential in terms of newly arrived fraud cases
Decision Tree	Implementation is more straightforward with low power of computation and suitable for analyzing the real-time data	Overfitting may rise if the information of the underlying domain does not set in training data
Logistic Regression	Implementation is easy and fraud detection is based on historical data	Performance of classification is lacking when compared with methods of data mining
Linear Regression	When dependent and independent variables have an almost linear relationship, it generates an optimal result	Sensitive for the outliers and numeric value limitation
Support Vector Machine	The nonlinear problem of classification is solved with low power of computation and suitable for analyzing real-time data	Input data transformation results in difficulties while processing the data

collection, data preprocessing, analyzing of data, data training with different classifiers, respectively, and later testing the data. During the stage of preprocessing, the entire data are transformed into a useable format. The hybrid undersampling (negative class) and oversampling (positive class) techniques were performed using two different data distribution sets. In the stage of training, the classifier algorithm is fed with preprocessed data. Later, the testing data are evaluated to find the accuracy for detecting out the fraud related to a credit card. Finally, all the different models are evaluated based upon accuracy and their best performance. The legal ratio with a total number of fraud transactions is a subset and used to conclude which model performs better when tested in the real-time scenario.

3.1. Dataset. The source of the dataset is the UCI Machine Learning Repository. The dataset holds the information-related transaction conducted through credit cards as a default payment gateway of the different customers in Taiwan. The accuracy is probably compared to six different data mining techniques. The dataset has the detail of transaction which has occurred in the year 2015 and consists of 30000 different customer data and nearly 3 lakhs of transaction data. The characteristic of the dataset is multivariate, and its entire attributes are accurate and integer. The dataset seems to be highly unbalanced and more biased about positive class. It contains the continuous variable (numerical) as input variable was Principal Component Analysis. Altogether 30 different input features are used for training and testing the model. The detailed information related to the transaction's background and its features is not provided due to the issue of confidentiality. The preprocessing of the dataset is carried out using hybrid oversampling and undersampling techniques to achieve the two different sets of distribution in an unbalanced dataset.

The experimental setup used for performing fraud detection in credit cards is Python v3 language setup with i5 8th Gen Processor and 240 GB of SSD with 8 GB of DDR4 RAM with the processor variant of 1050 H which has the clock speed of 2.6 GHz–5.0 GHz with the turbo boost, and the frequency of RAM is 2565 MHz for training and testing the model in minimum duration of time.

3.2. Sequential Model. The sequential model generates its sequential value by estimating the input values for the series which can be time-series data. A 2D convolutional neural network is applied for passing 2D signals using more cost, time, and resources for gaining the state-of-art level of performance. It is easier to train the dataset through a sequential model as it requires minimum computation complexity and generates a better result.

3.3. Naive Bayes Classifier. Naive Bayes is the statistical method that relies on Bayesian theory, where the result is obtained based on the highest probability. It estimates the probability of the unknown value based upon the known value. The logic and prior knowledge can be applied to predict unknown probability. Naive Bayes mainly depends on binary classes and conditional probabilities.

$$\text{prob}\left(\text{class}_j|\text{feature}_k\right) = \frac{\text{prob}\left(\text{feature}_k|\text{class}_j\right) * \text{prob}\left(\text{class}_j\right)}{\text{prob}\left(\text{feature}_k\right)}, \tag{1}$$

$$\text{prob}\left(\text{feature}_k|\text{class}_j\right) = \prod_{j=1}^{m} \text{prob}\left(\text{feature}_k|\text{class}_j\right). \tag{2}$$

In equations (1) and (2), n indicates the maximum amount of features, $\text{prob}\left(\text{feature}_k|\text{class}_j\right)$ indicates the probability of generating feature value feature_k provided in class_j, and $\text{prob}\left(\text{feature}_k\right)$ and $\text{prob}\left(\text{class}_j\right)$ indicate the probability of occurrence of feature value feature_k and the occurrence of class class_j, respectively. This classifier was utilized for binary classification with the aid of the Bayesian principle.

3.4. K-Nearest Neighbor (KNN). The KNN classifier is an instance approach of learning where classification is conducted based on the measure of similarity calculated by

Table 2: Comparison of performance metrics among the utilized models.

Model	Accuracy (%)	Precision (%)	Recall (%)	$F1$-score (%)
Naive Bayes	96.1	92.4	91.86	92.13
Logistic Regression	94.8	93.16	93.07	93.11
K-Nearest Neighbor	95.89	93.78	91.42	92.58
Random Forest	97.58	96.5	96.7	96.60
Sequential CNN	92.3	90.3	90.43	90.36

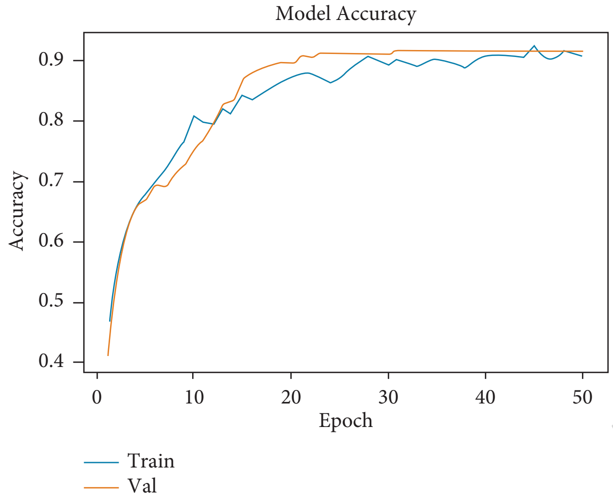


Figure 1: Representation of training accuracy vs. validation accuracy.

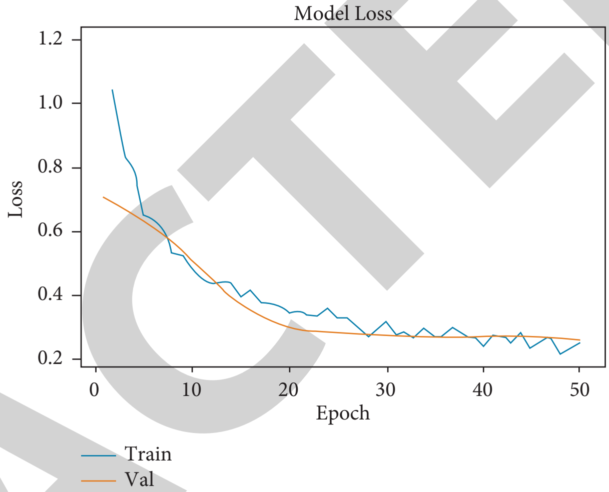


Figure 2: Representation of training loss vs. validation loss.

Manhattan or Euclidean and the Minkowski distance function. Manhattan or Euclidean function mainly deals with continuous variable, while the Minkowski deals with categorical data. The Euclidean function is used for measuring the distance in the KNN classifier. The Euclidean function (D_{ij}) between two vectors (X_i and X_j) is calculated by

$$\mathrm{Dist}_{ij} = \sqrt{\sum_{l=1}^{m} \left(X_{il} - X_{jl}\right)^2}. \tag{3}$$

3.5. Logistic Regression. Logistic Regression is a functional approach for measuring the probability for binary classes based on particular or more features. It generates the best parameter for the sigmoid nonlinear function. The input vector (x) and the sigmoid function () are shown below.

The input data are a vector (z), and w is the best coefficient, when multiplied together and summarized to generate the targeted class classification classifier. If its value crosses 0.5, then it is known as 1, otherwise it is considered as 0. Then, the gradient ascent optimizer is applied in training for knowing the best performance of the classifier.

$$\begin{gathered}\mathrm{Sig}_f(x) = \frac{1}{\left(1 + e^{-x}\right)}, \\ x = w_0 z_0 + w_1 z_1 + w_2 z_2 + \cdots + w_n z_n.\end{gathered} \tag{4}$$

3.6. Result. In this research, the sequential model and the other four models of a classifier based on KNN, Naive Bayes, Logistic Regression, and Support Vector Machine are developed. For evaluating all these classifier models, training is conducted using 70% of the entire dataset, while for testing and validating, 30% of the dataset is used. Accuracy, specificity, sensitivity, precision, and the Matthews correlation coefficient (MCC) with the rate of balance classification are applied for measuring the performance of all these classifier models. The performance of all these classifier models is evaluated. The sequential model visualizes the better performance. The technique of the sequential model generates superior performance for the evaluation metrics applied. It generates the highest value for precision and specificity. The obtained performance metrics are presented in Table 2.

The obtained results were plotted to visualize the comparison in terms of performance metrics. First, training accuracy vs. validation accuracy is represented in Figure 1. Second, training loss vs. validation loss is represented in Figure 2. Lastly, all performance metrics' comparison graph is represented in Figure 3.

4. Future Scope and Conclusion

The proposed methodology provides the information that Random Forest performs better than Sequential CNN. The drawback of this methodology is that anyone would expect Sequential CNN can outperform any of the conventional ML methodologies, but it is not happening here. It may happen because the dataset is not enough to train and identify the

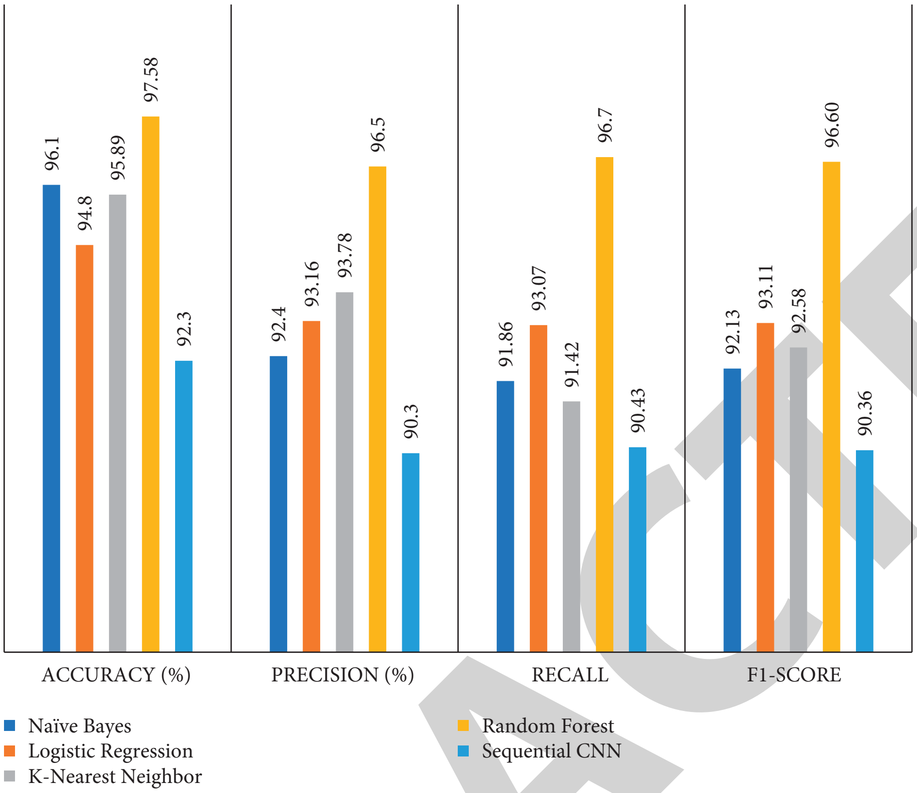


Figure 3: Comparison of performance metrics among the utilized models.

hidden patterns to predict the future or upcoming data and the initialization of weights was very random that might affect the training process. It can be further improved in two ways. The first way is to tune the hyperparameters through optimization, and the second method is to apply the transfer learning methodology so that the performance of the proposed methodology is improved to detect the fraud transaction through credit cards in the healthcare sector.

The study on fraud detection related to a credit card using deep learning and the machine learning techniques has been introduced in this paper. The different standard models such as Sequential Model, Decision Tree, Random Forest, and Naive Bayes are introduced and cast for empirical evaluation. The dataset related to a credit card is available publicly. Different standard models are trained and tested to generate the accuracy, and the model which performs better with stored and real-time data is identified. Sequential model and machine learning classifiers are trained and tested on the dataset, and their performance is evaluated with many relevant metrics for detecting fraud in credit cards. Our study indicates that online and offline transactions have different qualities when compared with the sequential pattern of earlier predicted fraud detection data.

The different algorithms presented in this paper can be extended towards the online learning approach of machine learning in the future. They can be investigated in both offline (collected data) and real-time scenario for obtaining better results with reasonable accuracy. The model of online learning will detect fraud cases in real time with the minimum time required for processing. This helps to predict the fraudulent transaction in an earlier stage (before conducted), which has positive impacts towards reducing the number of loss cases in the financial sector.

Data Availability

The data can be made available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully thank I-Cheng Yeh from the Department of Information Management of Chung Hua University, Taiwan, and the Department of Civil Engineering, Tamkang University, Taiwan, for the credit card fraud dataset and description related to it in the public domain.

References

[1] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, “Credit card fraud detection using machine learning techniques: a comparative analysis,” in *Proceedings of the 2017 International Conference on Computing Networking and Informatics (ICCNI)*, pp. 1–9, IEEE, Lagos, Nigeria, Oct. 2017.

[2] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, “Credit card fraud detection: realistic modeling and a novel learning strategy,” *IEEE transactions on neural*

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9854329, 1 page
https://doi.org/10.1155/2023/9854329

Retraction

Retracted: Machine-to-Machine Communication for Device Identification and Classification in Secure Telerobotics Surgery

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] M. P. Lokhande, D. D. Patil, L. V. Patil, and M. Shabaz, “Machine-to-Machine Communication for Device Identification and Classification in Secure Telerobotics Surgery,” *Security and Communication Networks*, vol. 2021, Article ID 5287514, 16 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2021, Article ID 5287514, 16 pages
https://doi.org/10.1155/2021/5287514

Research Article

Machine-to-Machine Communication for Device Identification and Classification in Secure Telerobotics Surgery

Meghana P. Lokhande,[1,2] **Dipti Durgesh Patil**,[3] **Lalit V. Patil**,[4] **and Mohammad Shabaz**[5,6]

[1] *Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India*
[2] *Research Scholar, Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, India*
[3] *Department of Information Technology, MKSSS's Cummins College of Engineering for Women, Pune, India*
[4] *Department of Information Technology, Smt. Kashibai Navale College of Engineering, Pune, India*
[5] *Arba Minch University, Arban Minch, Ethiopia*
[6] *Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India*

Correspondence should be addressed to Mohammad Shabaz; mohammad.shabaz@amu.edu.et

Received 21 July 2021; Revised 19 August 2021; Accepted 20 August 2021; Published 28 August 2021

Academic Editor: Chinmay Chakraborty

The capacity of machine objects to communicate autonomously is seen as the future of the Internet of Things (IoT), but machine-to-machine communication (M2M) is also gaining traction. In everyday life, security, transportation, industry, and healthcare all employ this paradigm. Smart devices have the ability to detect, handle, store, and analyze data, resulting in major network issues such as security and reliability. There are numerous vulnerabilities linked with IoT devices, according to security experts. Prior to performing any activities, it is necessary to identify and classify the device. Device identification and classification in M2M for secure telerobotic surgery are presented in this study. Telerobotics is an important aspect of the telemedicine industry. The major purpose is to provide remote medical care, which eliminates the requirement for both doctors and patients to be in the same location. This paper aims to propose a security and energy-efficient protocol for telerobotic surgeries, which is the primary concern at present. For secure telerobotic surgery, the author presents an Efficient Device type Detection and Classification (EDDC) protocol for device identification and classification in M2M communication. The periodic trust score is calculated using three factors from each sensor node. It demonstrates that the EDDC protocol is more effective and secure in detecting and categorizing rogue devices.

1. Introduction

Currently, wireless and wired systems interacting with other devices having similar functionality have become one of the fastest-growing areas of research. Machine-to-machine communication (M2M) is a new technology that allows machines to communicate without human intervention [1]. Intelligent software applications are the process that collects data and provides the end user with a set of intelligent services and practical interfaces [2]. The idea of implementing telematics and telemetry is known, but in connection with the proliferation of the Internet and the ubiquitous trend to connect, especially through a wireless communication system, the M2M system has attracted the attention of both academia and industry [3].

In the medical device industry, M2M communication is one of the fastest-growing sectors. According to Global Info Research, the connected medical device market is projected to expand from $939 million in 2018 to $2.7 billion by 2023, with the largest growth forecast for the United States [4]. M2M communication faces various security challenges. Most of the vulnerability issues arise from a lack of a central authority and a wireless medium of transmission. Route creation and data transmission are two important functions of the routing algorithm. These two stages must be protected from attackers. The routing protocol must be strong enough to withstand various attacks. Thus, reliable communication needs a secure routing algorithm. Proper identification and classification of devices before any operation is required.

Machine-to-machine (M2M) communication system is an emerging technology for next-generation communication networks. M2M system facilitates ubiquitous communication among smart devices with minimum human intervention. The key characteristic of M2M is decreasing the cost of human resources and providing great research in the medical and industrial fields. The telerobotic system operates at different levels of topologies. It starts from direct control to command tracking telerobots. The major challenges are communication delay, access control, and stability [5]. M2M communication works for long-distance communication and can be easily incorporated for telerobotic surgeries. The existing telerobotic challenges are solved with M2M security measures. In M2M, malicious nodes are identified and provide a reliable route for data transmission in robotic surgery. Thus, it reduces communication delay and synchronization problems.

Remote healthcare is an evolving area of research as the world moves from remote surveillance to real-time and rapid detection of diseases. Robotic surgery gives birth to telerobotic surgery [6]. In robotics, surgeons perform operations while sitting near a patient's console. Instead, telerobotic surgery allows surgeons to remotely control patients using surgical robots and a communication network between them. In telerobotic systems, the surgeons control the slave which is at a different location. Telerobotic surgery offers several products and benefits, including high-quality assistance for people in developing countries and providing the surgical needs for soldiers [7]. Similarly, it can overcome the limits and inconsistency of the systems of public health in developing countries and developed countries and regions. These types of procedures are a major barrier to patient safety, data security, and privacy concerns. The main challenge is reliability in performing telerobotic surgery due to the unavailability of a secure mechanism for device identification and classification in M2M communication.

Wireless communication in M2M makes the attacker easily monitor the network traffic and discovers vulnerable machine-type devices even though network is secured through encryption [8]. Instead of passively monitoring the traffic and identifying vulnerable devices, the malicious devices can be identified based on their network behaviour. In this context, the medical devices perform unexpected activities by monitoring surgical environment and create denial of service attack. To improve security, it is important to know the type of devices connected to the network. Medical tools or equipment used during surgeries need to be identified. It helps the system to specify filtering rules or block the access for particular devices from which unexpected or irrelevant data is transmitted. The devices that make the network vulnerable need to be identified and classified to make secured surgical environment. As network traffic from medical field will have high priority than other industry or enterprise data, for medical device configuration, administrator has to configure different filtering and access rules depending on type of device. This manual configuration is time-consuming for unscalable IoT network. To ensure security and quality of services, the access rules are defined based on device type and its priorities. However, device classification is nontrivial task as IoT consists of heterogeneous devices with dynamic nature of network traffic.

In M2M communication, devices interact with each other and exchange information autonomously to perform the necessary tasks. During surgery, it is important to protect the network against various types of attacks. A small communication delay can create a threat to patient life. It may also slow down the entire system in the operating room. To make telerobotic surgery efficient and reliable, the proposed system is designed. The main objective is to design the system to incorporate machine-to-machine communication in a telerobotic surgical environment to meet security requirements. To make the network secure and reduce energy consumption, nonmalicious devices are identified and classified.

The proposed secured telerobotic surgery is based on robust and beneficiary control techniques for providing a more efficient, precise, and cheaper alternative for medical surgeries as compared to existing technologies. Secured surgical environments are tested through performance parameters which create technical limitations and challenges in surgeries. Standard energy-efficient protocol LEACH is analyzed for network parameters in the medical sensor node network. Fuzzy inference system-based energy-efficient protocol is adapted to measures performance parameters with and without attack. Efficient Device type Detection and Classification (EDDC) protocol is proposed and designed for device identification and classification in M2M communication for secure telerobotic surgery. This protocol identifies the node as legitimate or attacker. Finally, the trust score computation technique using three parameters of each sensor node to compute the periodic trust score of each node n is designed. The parameters are Successful Packet Delivery (SPD), Energy Level (EL), and Node Degree (ND) selected to correctly estimate the malicious behaviours of attackers in the network.

One of the motivations in healthcare is the need for long-distance medical surgeries. Under resourced locations such as semiurban and areas near the border, there is often a lack of medical equipment and expert surgeons. It provides a technological and clinical solution in the robot-assisted surgical techniques to improve the quality and results of surgical intervention. Providing this technology to surgeons has led to the development of new surgical techniques that would otherwise be impossible.

The traditional surgical techniques are being enriched by robot-assisted surgery especially in long-distance surgeries. Unsecured network affects the functionality of telesurgery. The proposed research work shows novel research for making the telerobotic system robust, reliable, and attack-resistant.

(a) The system identifies the legitimate device and attacking device based on trust parameters by using the device identification algorithm.

(b) The available energy of each legitimate device is checked periodically and based on energy available, the devices will be provided with access to the available resources in the operating room. Here, end-to-end security of devices is ensured.

(c) The legitimate and attacker nodes are identified at each interval and accordingly it will be involved in the routing mechanism.

(d) Thus, only the legitimate and authorized device can only access the resources. This makes the system robust for attack and improves the quality of service parameters in terms of packet loss and delay as well.

The main objective of the work is to propose an energy-efficient protocol to meet the security requirement in medical practice. Though several researchers worked with available energy-efficient protocols, this idea is being introduced in the telerobotic surgery with the objective of providing a secured and reliable environment during surgeries. A robust and reliable network is possible by minimizing the energy utilization of medical nodes and involving the communication of trusted nodes in data delivery.

The reason behind choosing the medical application is twofold. First of all, the authors would like to provide an excellent medical facility to underresource locations. Secondly, the proposed model makes the existing healthcare system robust and attack-resistant. It also improves quality services to medical staff and patients. However, it can be extended for military application and industrial automation.

This paper addresses the concern of ensuring secure network communication with low energy consumption. Data and device protection is the primary function of this concept, which has not been fully explored by researchers. The simulation study shows the optimum use of EDDC protocol in achieving secured and reliable communication which allows the surgeon to perform the surgery without any threat to patient life.

The rest of the paper is organized as follows. Section 2 reviews related work and Section 3 covers the proposed M2M protocol. In Section 4, experimental results are discussed and presented. In Section 5, the performance metric of the proposed protocol is discussed. Finally, Section 6 summarizes points in conclusion.

2. Related Work

This section summarizes the existing work on M2M communication and telerobotic surgery and communication networks, the device identification and classification, and their secure communication. Regardless of the clear growth rate, researchers need to work to develop an innovative solution and come out with different device characteristics and requirements. The second problem arises that the technical solutions for the entire functioning of M2M system are quite diverse [9]. Perhaps the use of M2M or a wired application may be related to research on equipment technology for wireless connectivity, or short distances, communication, or special standards or specialized communication technologies [10].

Smart healthcare M2M is a new emerging paradigm that offers promising solutions [11]. Various types of smart applications use the M2M data communication approach [12]. An M2M-based healthcare system is proposed in [13]. It illustrates the monitoring of patient health but has less focus on security requirements. The main problem and aspect is security, where M2M communications must be standardized and widely considered before they can be fully involved in practical life [14]. M2M communication model without centralized management is shown in Figure 1. This model facilitates discussion of specific wireless communications and security approaches for M2M applications [15]. Its design specifically covers two separate communication areas that support M2M wireless communications for the Internet. Its integration with remote sensing devices and the other connection supports the rest of Internet communications. Gateway technology is a way to connect end devices to back-end platforms. Communication between the two domains can be mediated by a security gateway that implements a filtering policy for communication according to the requirements of each application. This device can also have other control and safety functions. It serves as an internal system or control unit for M2M remote sensing applications.

Wireless communication between devices in the M2M domain can take place in an unsupervised manner, raising important security concerns such as authentication and trust between devices without knowing each other beforehand. Many applications also require communication with a back-end or gateway device. The gateway unit communication model can actually support the role of personal or industrial control devices or electrical devices, according to the requirements of a specific M2M wireless remote sensing application; it also supports communication between M2M remote sensing devices and the Internet.

Since applications for M2M include devices with low size, independent power supply, and energy restriction, the security solution must take into account the size of the key, the complexity of encryption algorithms, and the key algorithms used for authentication [15]. Also, M2M applications should deal with security threats [16] and other attacks that can negatively affect functionality.

(a) Incorrect network attack: when the M2M device is disabled, an attacker can pass the identification (impersonation) of the M2M device to other network components and obtain confidential information.

(b) False network response: since some M2M devices (such as the Mote sensor) operate with a low-power battery and turn off the radio to save power, an attacker can continuously put the device to sleep by sending a fake network trigger to waste power.

(c) Tamper attack: the triggered indicator may contain the IP address of the application server to which the

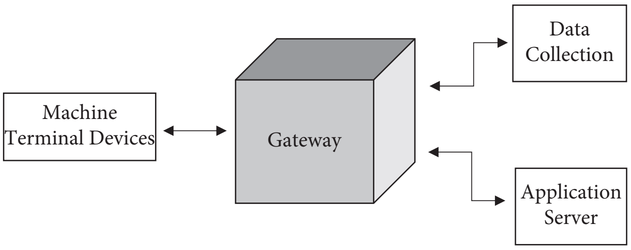


Figure 1: M2M communication without a centralized system.

M2M device should connect. If the IP address is amplified by an attacker, the M2M device may establish a connection to the wrong server, thereby failing to communicate with the correct server and wasting energy and losing the data.

M2M applications for connecting mobile users and home networks are designed in [17]. The authentication and key-based algorithm are used to ensure secure communication between network nodes. However, the proposed security scheme is not suitable for more complex scenarios and has dynamic connections between users and M2M devices. Privacy and information functions [18] are used to preserve hash integrity.

IoT device classification [19] for network security makes the device turned on a white list of permitted devices while connecting to the network. The device classification method is developed to identify and examine their network traffic using IoT devices [20]. The classification of graph-based method proposes [21] correlation graphs of objects by a random walk. It proposes a probabilistic multifunctional model for classifying heterogeneous objects according to a classification scheme with multiple labels. The researchers also worked on detecting device types for security in the IoT [22]. The new devices are designed to find things that have the appearance of the Internet when they are added to the network, so they can easily customize the way the device is detected. The neural network is introduced in [23] to recognize and control the user's intentions of the robot. In both cases, devices are classified under deep neural networks, especially repetitive neural networks.

Despite advances in robotic surgery, safety for telerobotic surgery is challenging. Complete security system, particularly the design and development of the telerobotic surgery security aspects is yet to be designed. It is important to find out not only legal and technical requirements for telerobotic surgery but also security [24]. A security requirement for telerobotic surgery is not yet proposed [25]. Secure ITP uses open source software and Federal Information Processing Standards (FIPS) to develop prototypes that meet strict security requirements in telerobotic surgery. While securing ITP is of reasonable construction, it cannot solve such critical problems as steal personal information from a patient or other administrative and legal issues [26].

Safety issues of telerobotic surgery are identified and divided into telerobotic surgical procedures [27]. Researchers [28] point to the security; availability, price, and legal liability of surgeons are major obstacles to the success of remote robotic surgery. Mechanisms were demonstrated in [29], when the remotely handled software proved the possibility of surgical telerobotic. The author proposed a new way of integrating light privacy and reliability with uniform protocol and compliance performance of cryptographic AES systems [30]. Security attacks with advanced remote-controlled robotic surgical systems are analyzed and investigated [31]. Reference [32] presented a two-way generalized predictive controller associated with QoS-friendly IP security protocol for telerobotic systems. An approach for using incorrect commands in surgery using an ML algorithm is proposed [33].

In addition, it is proposed to ensure the public safety of several patients in a network with data from remote health monitoring systems. The author proposed a multimedical framework [34]. Data is analyzed using remote sensing sensors during patient anonymity and monitoring. The framework does not ensure the security policy. Data security algorithms for the sensors in the body, which are necessary for intelligent systems in healthcare, are developed [35]. The system does not guarantee security policy and is not error-tolerant. Reference [36] proposed an easy authentication method for IoT-enabled medical environments. The system ignores anonymity and does not apply a security policy.

The current standard [37] allows handling or ensuring a reliable connection to a healthcare environment. These standardized protocols and security mechanisms may be used for scenarios of M2M communication but may be modified according to application requirements. Therefore, a new mechanism is required to ensure the security and confidential information via M2M communication. On the contrary, many common security protocols and access control approaches can be adapted to meet the requirements of security [38]. The author proposes real-time data mining for body area network in the wireless network [39, 40]. Several researchers offer device classification and security protocols for efficient route discovery. The authors exclusively focused on resource allocation for a limited number of devices. In most of the research, there are lack of centralized administrative control, low mobility of nodes, and failure to integrate and manage information in the IoT environment. The authors provided security solutions but there in need to provide efficient solutions that will help to enhance security in medical field.

The various researchers propose techniques for telerobotic surgery using mobile edge computing and machine-to-machine network. So, the author proposes the best security protocol called Efficient Device type Detection and Classification (EDDC) protocol and the practical design principle provides a complete safety guide. Protocol-based authentication ensures confidentiality, integrity, anonymity, and responsibilities. The 5 G network for robotic surgery is discussed in [41]. The M2M network performance in presence of malicious nodes is presented for the telerobotic surgery [42].

2.1. Attacks in IoT and Telerobotic Surgery. Telerobotic surgery is subject to various types of active and passive attacks [43]. These attacks and investigation scenarios are shown in Table 1.

Table 1: Attacks in IoT and telerobotic surgery.

Type of attacks	Investigation scenarios
Replay attack	The enemy will play a legitimate message/command illegally later
Eavesdropping	Attackers passively enter conversations and disclose personal information about patients without permission
Masquerading attack	Authorized authentication mechanisms and methods will try to break by bypassing users
Session hijacking	An attacker could gain control of consoles or settings for a session
Brute force attack	To recover the private key with all possible keys in the keyspace
Data removal	Enemies remove the data
Forgery attack	The enemy will play the command illegally and try to play it illegally
Viruses/worms	Virus, worm, crash the operating system
Data theft	Static and dynamic data

2.2. Challenges in Telerobotic Surgery. Several IoT devices are deployed in an open field where security and privacy are of most important concerns [44]. Security leaks in remote control systems pose an existential threat to the area of surgical robots in general, as mounted attacks can break the robot or damage other nearby devices in the operating room. Even if the attack is minor, the damage caused by the surgical robots can weaken the public's trust.

It has attacks as follows:

(a) Node hardening: an attacker replaces nodes on the entire device and either connects it directly or changes access to confidential information, and so on [45].

(b) Fake node: attackers adjust the fake node and access the information [46].

(c) Physical damage: enemies can physically damage your device with Internet of Things that cannot be serviced. The device's Internet stuff can be deployed in both open and closed places, which makes them vulnerable to physical harm from attackers.

(d) Malicious code injection: an attacker physically invades a node by inserting malicious code into the node, which gives unauthorized access to the system.

(e) Sensor data protection: data privacy requirements for sensors are low because an adversary can place a sensor near an IoT system sensor and feel the same value.

(f) Authentication of node: many medical sensors in IoT systems face authentication problems [44]. Thus, a huge amount of network communication affects performance.

(g) Congestion in node: big sensor data communicating with many device authentications can cause network congestion.

(h) RFID interference: the RF signal used by RFID is distorted by the noise signal, which causes a denial of service.

3. Proposed Protocol in M2M Communication for Secure Telerobotic Surgery

3.1. System Architecture. As represented in Figure 2, clustering is carried out autonomously for identifying the node as cluster head (CH) and cluster member (CM). Further to improve the clustering process, nodes are identified as legitimate or attacking devices based on trust score parameters. The weighting parameters are used to get the trust score value between 0 and 1.

Based on the trust-score value, nodes are identified as legitimate and attacking devices. The cluster head list is updated with legitimate devices and the cluster member list with attacking devices. At the periodic interval, the trust-score value of each node in the cluster head list is calculated to decide the cluster head for that interval. The selected cluster head advertises and announces the time slot for data transmission. At each interval, the remaining energy for the cluster head is checked, and accordingly, the cluster cycle is updated.

3.2. Proposed Protocol. M2M is a network where a large number of intelligent devices create, share, and collaborate information without humans. Having a variety of applications and many benefits, the design of the M2M network faces several technical problems. One of the key issues underlying economic growth is its security. However, it is very important to provide correct and secure information to the end user. So, the author proposes an Efficient Device type Detection and Classification (EDDC) protocol for device identification and classification in M2M communication for secure telerobotic surgery. N number of medical devices are introduced in the network their size $X * Y$. Considerations for developing the proposed protocol EDDC are given as follows:

(a) A medical device capable of sensing medical data and transmitting them to a single medical station is called a base station (BS)

(b) Medical devices are randomly deployed in medical services in different locations

(c) Every medical device is static and works evenly in the network

(d) The position of each device is calculated using the received signal strength indicator (RSSI)

(e) All nodes are grouped into clusters using conventional K-means

(f) Data transfer from the CMs to their assigned CHs is carried out in a multipass method

(g) BS nodes that are outside the network area are not restricted

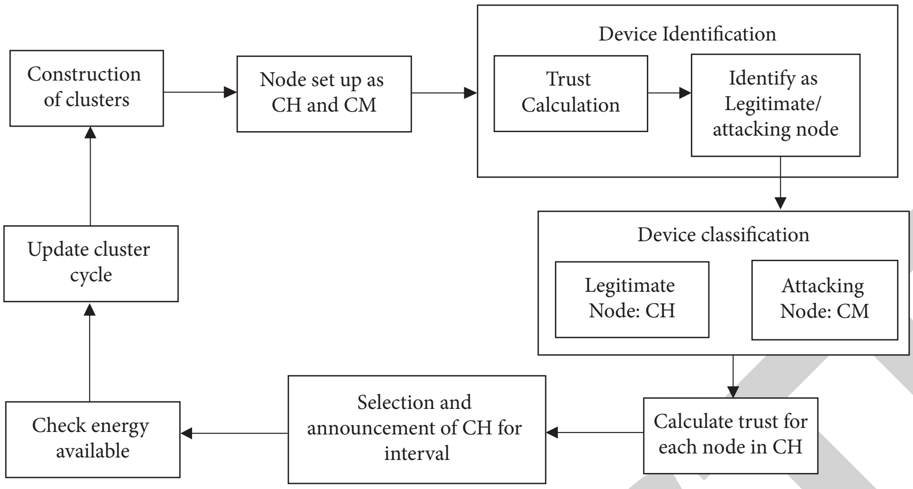


Figure 2: System architecture for Efficient Device type Detection and Classification (EDDC).

(h) All medical sensor nodes are limited by limited computing power and battery

(i) In M2M, it is assumed that nodes can perform data receiving, data forwarding, and data forwarding

3.3. Device Type Identification. This stage identifies that the device type is a legitimate sensor device or an attacker sensor device, which designs the trust-score computation technique to compute the periodic trust score of each node *n*. First, calculate the trust score (TS); if TS value of a node is larger than the global trust threshold value, then the node is identified as a legitimate medical device; otherwise, the node is identified as an attacked medical device.

To identify nodes *s* legitimate or attacking, the author has developed a method for calculating the confidence score using three parameters for each sensor node to calculate the periodic trust score for each node. The parameters like Successful Packet Delivery (SPD), Energy Level (EL), and Node Degree (ND) were selected to correctly estimate the malicious behaviours of attackers in the network. In Algorithm 1, the TS^n = get Trust Score (n) computes the trust scores using these three parameters.

3.3.1. SPD Trust. Since the attacking node performs a malicious task on the network and the SPD does not work well on the network, first exchange HELLO packets with neighbouring nodes and calculate the SPD trust score for each device. The SPD of node *n* on the current time interval $t-1$ to t is calculated as follows:

$$SPD^n = \frac{n^{rcv(t-1,t)}}{n^{ge(t-1,t)}}, \tag{1}$$

where $n^{rcv(t-1,t)}$ and $n^{ge(t-1,t)}$ are the total number of packets received and generated in time interval $t-1$ to t. Nodes with higher SPD values are more likely to be identified as legitimate medical devices.

3.3.2. EL Trust. Nodes that drain this energy faster and still advertise themselves as candidates for data transfer or CH selection processes are also considered malicious nodes. Therefore, it is very important to calculate the trust according to the current level of remaining energy of each device:

$$EL = \frac{E_{\text{rem}}\left(n^t\right)}{E_i(n)}, \tag{2}$$

where $E_{\text{rem}}(n^t)$ indicates the remaining energy at time t and $E_i(n)$ indicates the initial level of energy. Node with high EL values is more likely to be identified as legitimate medical devices.

3.3.3. ND Trust. ND is the third reliability parameter calculated in this study to determine the reliability of medical devices in the clustering and data transfer phase. An attacker, such as DDoS or eavesdropping, can attract a source of information with false claims enough for neighbours to send information over a small geographical distance to their intended destination. The number of neighbours *NC* of node *n* at time *t* is calculated as using RSSI:

$$NC = \text{count}\left[\frac{n}{\text{distance}\,(n, pi)} < rssi\right], \tag{3}$$

where $n \neq pi$ and distance(n, pi) calculates location distance of n and $pi^{th} \in N$ using RSSI of n. The NC is used to calculate the trust score value as

$$ND^n = 1 - \left(\frac{1}{NC}\right). \tag{4}$$

In the above equation, the number of nearest or one-hop nodes of the currently investigated node *n* is computed. This is done by checking the RSSI limits (i.e., the distance between current node *n* to its nearer node *pi*) should be below the RSSI value of the investigated node.

```
    $GT$: global trust threshold value
$TS^n$: trust-score value of $n^t$ sensor device
$TS^n$: trust-score value of $n^t$ sensor device
Inputs
    $N$: number of medical devices
$TS^n = 0.35$: threshold value of trust score
    $S$: simulation time
Output
    $LS\ D$: node identified as legitimate
    $AS\ D$: node identified as legitimate
While ($k$)
At each periodic interval $pi \in S$ Deploy
$N$ number of medical devices
For each node $n \in N$
$TS^n$ = get Trust Score ($n$)
    If ($TS^n > GT$)
"node identified as legitimate medical device"
$LS\ D \leftarrow n$
    Else
"node identified as attacked medical device"
)$AS\ D \leftarrow n$
    End If
End For
$pi++$
End While
Stop
```

Algorithm 1: Device type identification.

The count parameter represents the total number of nodes NC that satisfy the criteria of RSSI counted as the nearest nodes of the currently investigated node n.

The final trust score of node n is calculated using a weighted approach:

$$TS = \left(w^1 \times SPD^n\right) + \left(w^2 \times EL^n\right) + \left(w^3 \times ND^n\right). \tag{5}$$

Here, the values for w^1, w^2, and w^3 are chosen as 0.4, 0.3, and 0.3 so that the sum is 1. The trust score TS for node n now ranges from 0 to 1. The node with a high trust score value is considered as a legitimate medical device.

3.4. Device Classification. In this stage, the output of the device identification stage is the input of the device in the classification stage; we perform the next steps in Algorithm 2, using a periodically calculated trust score.

Each sensor node is identified and labelled under legitimate nodes and attacking nodes. Initially, the network is classified into clusters consisting of cluster head and cluster members. For each node in the network, the trust-score value is calculated based on the packet delivery ratio, energy available, and number of neighbours. The final trust-score value decides device type as legitimate or attacking nodes. The legitimate nodes considered for further cluster head selection and attacking nodes act as member in the cluster. For each node in the cluster head list, the trust score is calculated, and based on the trust-score value, the cluster head for the current interval is selected and announced in the network. All the cluster members transmit data to the announced cluster head in their time slots. A periodical trust score is calculated at each interval to find the behaviour of nodes in the network. Identified legitimate nodes are classified for cluster head and candidate for cluster head. As the cluster head needs to be active all the time during data transmission, it consumes maximum energy. Periodically, trust scores are calculated and cluster heads are chosen. Due to some network circumstances, the energy of the cluster head starts draining, and the candidate for the cluster head list is considered to further decide the cluster head for that period. The new cluster head advertises and the cluster member node joins the cluster. Thus, the nodes in the candidate cluster head list further take the responsibilities if cluster head energy is minimum and not able to aggregate data.

3.5. Data Transmission. The proposed model is applied to the CMS system, where sensory medical data are periodically classified based on their similarity. The samples described will be trained and tested using the artificial neural network (ANN) machine learning classifier to build a machine learning model. As a result, the classification model can recognize new samples and classify them accordingly. CH, where all the heterogeneous data readings are transmitted, consumes the minimum energy for each set of redundant data to be discarded. Thus, the sensed readings redirected to the CH from the CMs are restricted thereby saving energy of individual sensor nodes.

```
CC: current cluster
∈∈NCCM: cluster member of CC
CH: cluster head of CC
Inputs
   LS D: node identified as legitimate
   AS D: node identified as the attacker
   S: simulation time
   NC: number of clusters
Output
   CH: list of selected CH nodes at intervalS
While (S)
At each periodic interval $pi \in S$
For each cluster $CC \in NC$
   For each node $l \in CC$
   If ($l \neq AS\ D$)
      'classify into CCH'
$CCH \leftarrow l$
   Else
'classify into CM'
$CM \leftarrow l$
   End If
   For each node $i \in CCH$
$R^i$ = Fetchscore ($i$)
val ($i$)$\leftarrow R^i$
   End For
index = max (val)
$CH \leftarrow$index
   Announce CH Selection
End For
$pi++$
End While
Stop
```

Algorithm 2: Device classification.

4. Result and Discussion

This section defines the results obtained after running the performance metrics. The system is build using the NS-2 parameter. A 1000 × 1000 m square area is created with 100 nodes. Nodes are considered as medical devices used in the medical operating room. The study involves the simulation of nodes such as cameras, health checkup machines, and other medical sensors. The authors use sensory channels like haptic input, visual information, and auditory information. Compare the proposed EDDC protocol, LEACH, and ECFU protocol under various attacks like DDoS attack, replay attack, and eavesdrop attack using performance parameters such as throughput, packet delivery speed (PDR), delay, overhead, and energy consumption. Experiments have made it possible to increase the service life of the network by about 20–25% using the proposed method. High-density sensor network consists of a huge number of small and power constrained nodes. The proposed model can be enhanced for large number of power constrained nodes. The model uses IEEE 802.11 medium access control protocol for communication. The model works on cluster based approach. It coordinates communication among nodes and determines cluster head and members based on available energy. Even though model is introduced in high-density network, it handles communication among heterogeneous nodes and classifies the nodes into different clusters. The cluster based approach simplifies network monitoring and determines residual energy of nodes. The initial experimentation is done for 100 to 600 medical sensor nodes. The quality of service parameters is measured in presence of various attacks. In traditional protocols, CH is created on the probability function, and CHs are selected in each round. The continuous simulation process provided that all nodes are consumed with energy values. Simulation is performed from 100 to 600 nodes (Table 2).

Telerobotic surgeries in a machine-to-machine (M2M) communication network carry data over long distances. The signal needs to travel for longer distances with sensor and control nodes that are distinctly located. To accommodate M2M aspect in telerobotic surgery, the author proposed using IEEE 802.11ah standard. It is a wireless standard designed to be utilized in the Internet of things (IoT) network.

LEACH is a routing protocol that organizes a cluster in such a way that energy is evenly distributed across all sensor nodes in the network. The LEACH protocol creates multiple clusters of sensor nodes and one node at the head of a particular cluster and acts as a routing node for all other nodes in the cluster. Before the communication has been started, select the cluster head (CH). The CH is responsible

Table 2: Simulation factors.

Parameter	Values
Number of IoT/machines/sensor nodes	100, 200, 300, 400, 500, 600.
Number of attackers	10% of nodes
Traffic pattern	CBR
Number of connections	10
Sink	Base station (BS)
Area	1000 × 1000 (long-distance communications)
MAC	802.11
Topology	Random deployment
Routing protocol	EDDC, ECFU, LEACH
Initial energy	0.5 J
Transmitter energy consumption	16.7 nJ
Receiver energy consumption	36.1 nJ
Simulation time	200 seconds

for routing the entire cluster. The LEACH protocol for randomization and cluster heads is chosen from a group of nodes, but this selection of CH from multiple nodes is more likely to result in longer latency. CH is responsible for collecting the data and transmitting them to the BS. In process of data transmission, more energy is consumed [46]. If the threshold value is less, the node fits as CH. The threshold *T (n)* is [44]

$$T(n) = \begin{cases} \dfrac{1}{1 - p(r\bmod (1/p))}, & \forall n \in G, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where p is the percentage of sensor nodes, r shows a circle, and G is a group of nodes in $1/p$ round. ECFU clustering technology, together with machine learning technology, made it possible to achieve efficient network operation using fuzzy updating. The author used performance parameters such as throughput, packet delivery speed (PDR), delay, communication overhead, and energy consumption.

The authors considered 200 seconds to represent simulation. However, we verified the quality of service parameters for different time frames. The authors incorporated various attacks and measured network performance during these 200 seconds. The literature indicates that the performance of the system can be measured at a smaller time scale. It is observed that analysis of the network with respect to packet loss can be possible in the mentioned time frame.

5. Performance Metrics

The wireless sensor network (WSN) is becoming increasingly used in the medical field. As WSN continues to expand, it provides many opportunities but also creates security challenges. The sensor network is vulnerable to various attacks. One of the most common types of attacks carried out recently is distributed denial of service (DoS) attack, replay attack, and eavesdropping.

In Figure 3(a), the number of packets with different packet sizes is sent to the network and their response time is recorded. Delay is one of the parameters responsible for reducing network performance. The response time in the presence of a DoS attack in the same network is measured and shown in Figure 3(b) which shows delay due to misbehaved nodes present in the network.

5.1. Influence of Attack on Network Lifetime.

5.1. Influence of Attack on Network Lifetime. Security in sensor networks is an issue that has been raised over a period of time. As WSN continues to grow, it opens the door to vulnerability. The most common and threatening type of attack carried out is distributed denial of service (DDoS) and replay attack. The attacks occur from multiple end of the sensor network and compromise legitimate nodes. These attacks make network resources unavailable to legitimate devices. These attacks affect the network performance and eventually lead to complete network failure.

An energy-efficient protocol is required in a medical sensor network that can increase network lifetime. The proposed EDDC protocol improves energy consumption and provides nodes for secured data transmission. The protocol extends network lifetime using a cluster based approach.

The network performance of the proposed algorithm is measured for 100 to 600 nodes. The malicious nodes are introduced and service parameters are measured. The author considered legitimate and attacking nodes in the medical sensor network. The effect of malicious nodes with respect to packet size and response time is measured and shown in Figure 3. The algorithmic performance is verified with 10% malicious nodes. The performance is measured based on node malicious nodes behaviour. The effect of malicious nodes on network performance for DDoS, replay, and eavesdrop attack is shown in Figures 4–6. The key service parameters such as delay, energy consumed, communication overhead, and packet delivery ratio are measured which are challenging to achieve in a telerobotic environment.

Experimental results are enhanced by introducing the impact of attacks on packet response time. For variable packet size, the response time of packets with and without attack is shown in Figure 3. Energy consumption is the key parameter that decides the network lifetime.

The performance metrics such as throughput, packet delivery ratio, delay, communication overhead, and energy consumption are considered for analysis and the efficiency

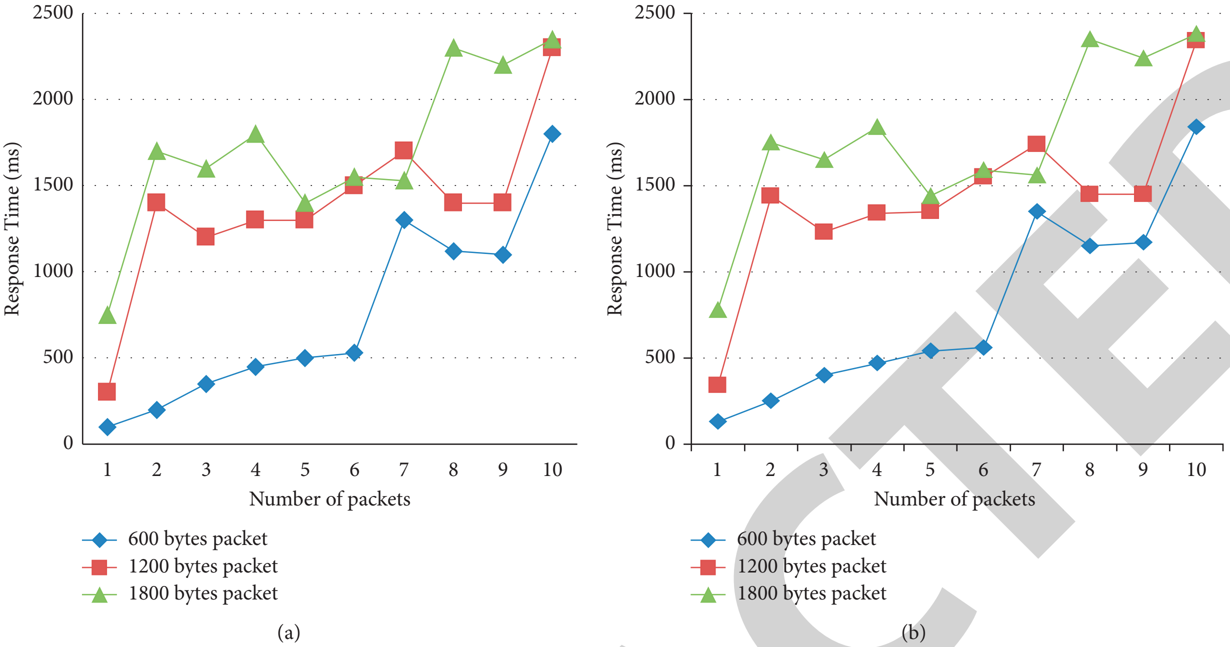


Figure 3: Node response time with respect to packet size. (a) Without attack. (b) With attack.

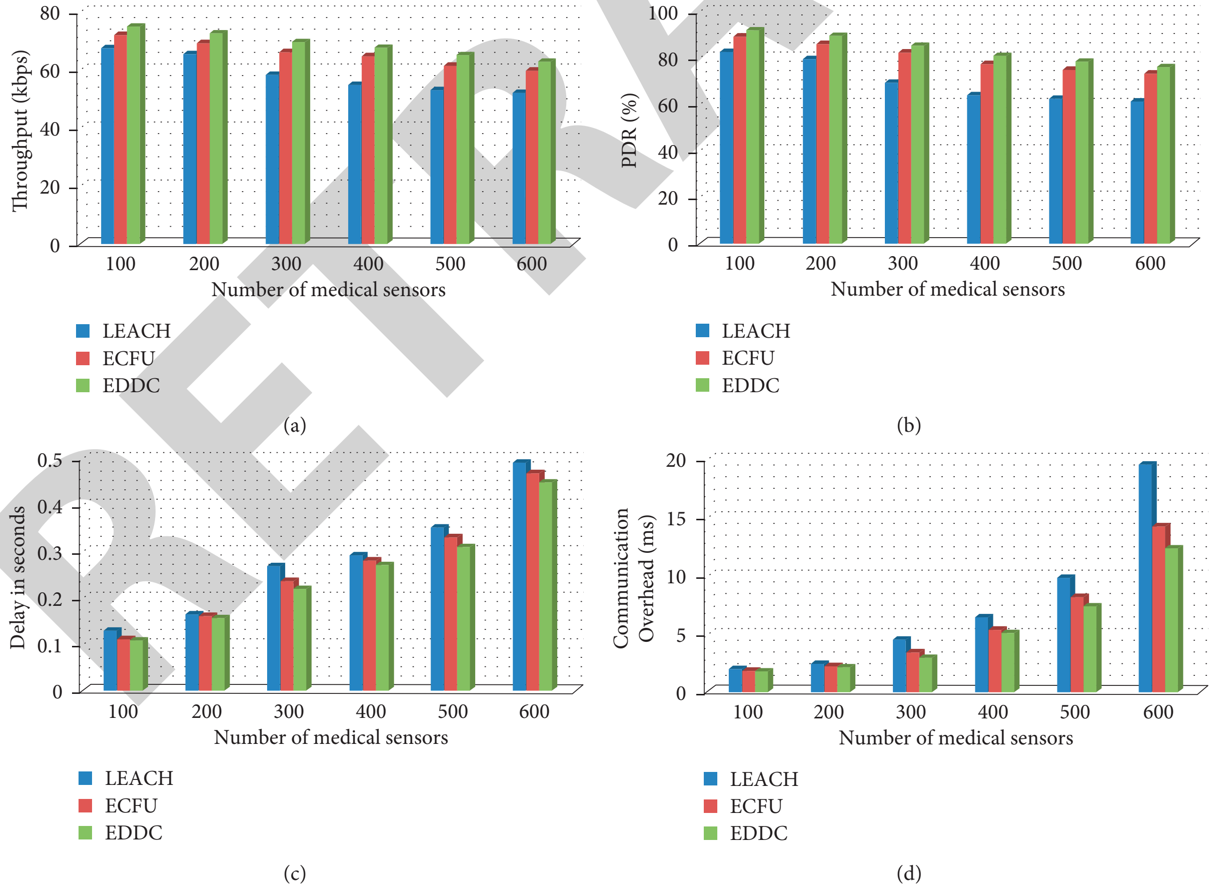
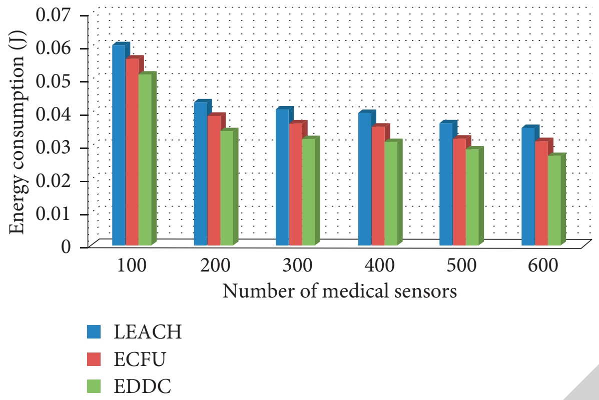


Figure 4: Continued.



(e)

Figure 4: Performance measurement under DDoS attack. (a) Throughput. (b) PDR. (c) Delay. (d) Communication overhead. (e) Energy consumption.

of the proposed EDDC system was compared with existing energy-efficient systems. The graphs are plotted for LEACH, ECFU, and EDDC for 100 to 600 medical sensor nodes network.

Throughput is the difference between send and receive data at the base station. The higher the ratio, the better the performance. The throughput of the system under DDoS attack is shown in Figure 4(a), the throughput of the system under replay attack is shown in Figure 5(a), and the throughput of the system under eavesdrop attack is shown in Figure 6(a). The number of medical sensors is defined on the x-axis representing the medical sensor nodes such as cameras, health checkups machines, computers, and printers, and throughput values are defined on the y-axis measured by EDDC protocol, LEACH, and ECFU protocol. Figures 4–6 show that the throughput value of the LEACH is less than the ECFU protocol and the throughput value of the proposed EDDC protocol is higher than LEACH and ECFU protocol. It is cleared from the graph that the proposed EDDC protocol outperforms better in increased sensor nodes and under DDoS, replay, and eavesdrop attacks. Throughput (T) is calculated by [47]

$$T = \frac{\sum_{I=1}^{\text{node}} \left(P_{\text{Suceessful delivered}}\right) \times \left(P_{\text{Average Size}}\right)}{P_{\text{sent time}}}, \qquad (7)$$

where $(P_{\text{Suceessful delivered}})$ is a packet sent successfully, $(P_{\text{Average Size}})$ = average packet size, and $P_{\text{sent time}}$ is the total time to the sent packets.

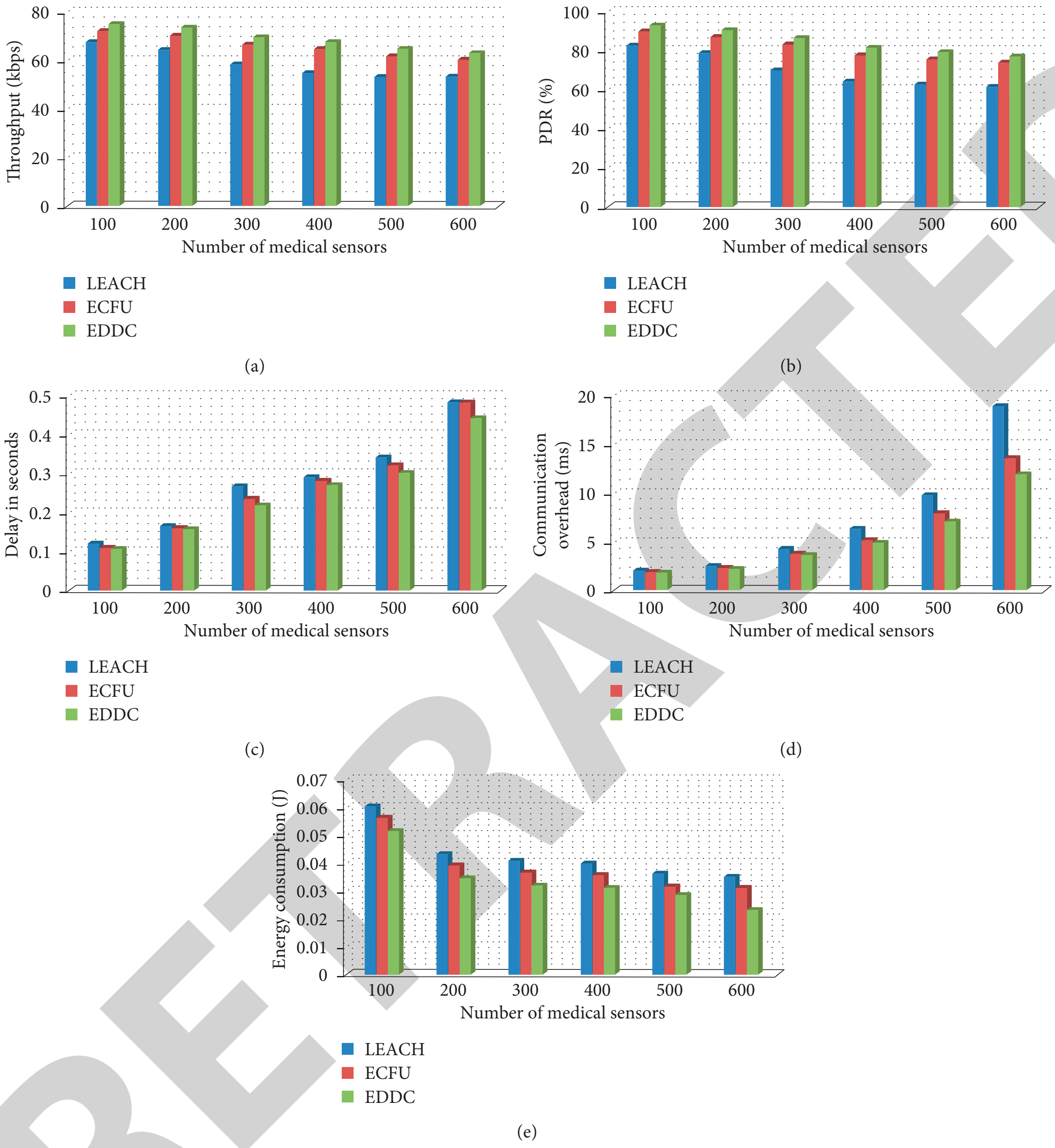
PDR calculation is the percentage of the receiving to sent packets. The PDR of the system under DDoS attack is shown in Figure 4(b), PDR of the system under replay attack is shown in Figure 5(b), and the PDR of the system under eavesdrop attack is shown in Figure 6(b). The number of medical sensors is defined on the x-axis; it represents the medical sensor nodes such as cameras, health checkups machines, computers, printers; and PDR values are defined on the y-axis measured by EDDC protocol, LEACH, and ECFU protocol. The figures show that the PDR value of the LEACH is less than the ECFU protocol and the PDR value of the proposed EDDC protocol is higher than LEACH and ECFU protocol. It is clear from the graph that the proposed EDDC protocol delivers more packets under DDoS, replay, and eavesdrop attacks. The PDR is calculated by [47]

$$P\ DR = \frac{\sum_{I=1}^{\text{node}} \left(P_{\text{delivered}}\right)}{P_{\text{sent}}}. \qquad (8)$$

The delay value suggests that the simulator has all kinds of time spent sending packets from the data source of the node to the destination node of the cluster. The system delay under DDoS attack is shown in Figure 4(c), delay of the system under replay attack is shown in Figure 5(c), and the delay of the system under eavesdrop attack is shown in Figure 6(c). The number of medical sensors is defined on the x-axis; it represents the medical sensor nodes such as cameras, health checkups machines, computers, printers; and delay values are defined on the y-axis measured by EDDC protocol, LEACH, and ECFU protocol. The figures show that the delay value of the LEACH is more than the ECFU protocol and the delay value of the proposed EDDC protocol is less than LEACH and ECFU protocol under DDoS, replay, and eavesdrop attacks. So, the delivery of packets is fast by EDDC protocol. The delay (D) is calculated by [47]

$$D = \sum_{I=0}^{\text{node}} T_{\text{time}} + R_{\text{time}} + W_{\text{time}}, \qquad (9)$$

where T_{time} = transmission time of packets, R_{time} = receiving time of packets, and W_{time} = waiting time of packets. Communication overhead is the ratio of actual communication time and computed communication time for real communication. The communication overhead of the system under DDoS attack is shown in Figure 4(d), the communication overhead of the system under replay attack is shown in Figure 5(d), and the communication overhead of the system under eavesdrop attack is shown in Figure 6(d). The number of medical sensors is defined on the x-axis; it represents the medical sensor nodes such as cameras, health



(a)

(b)

(c)

(d)

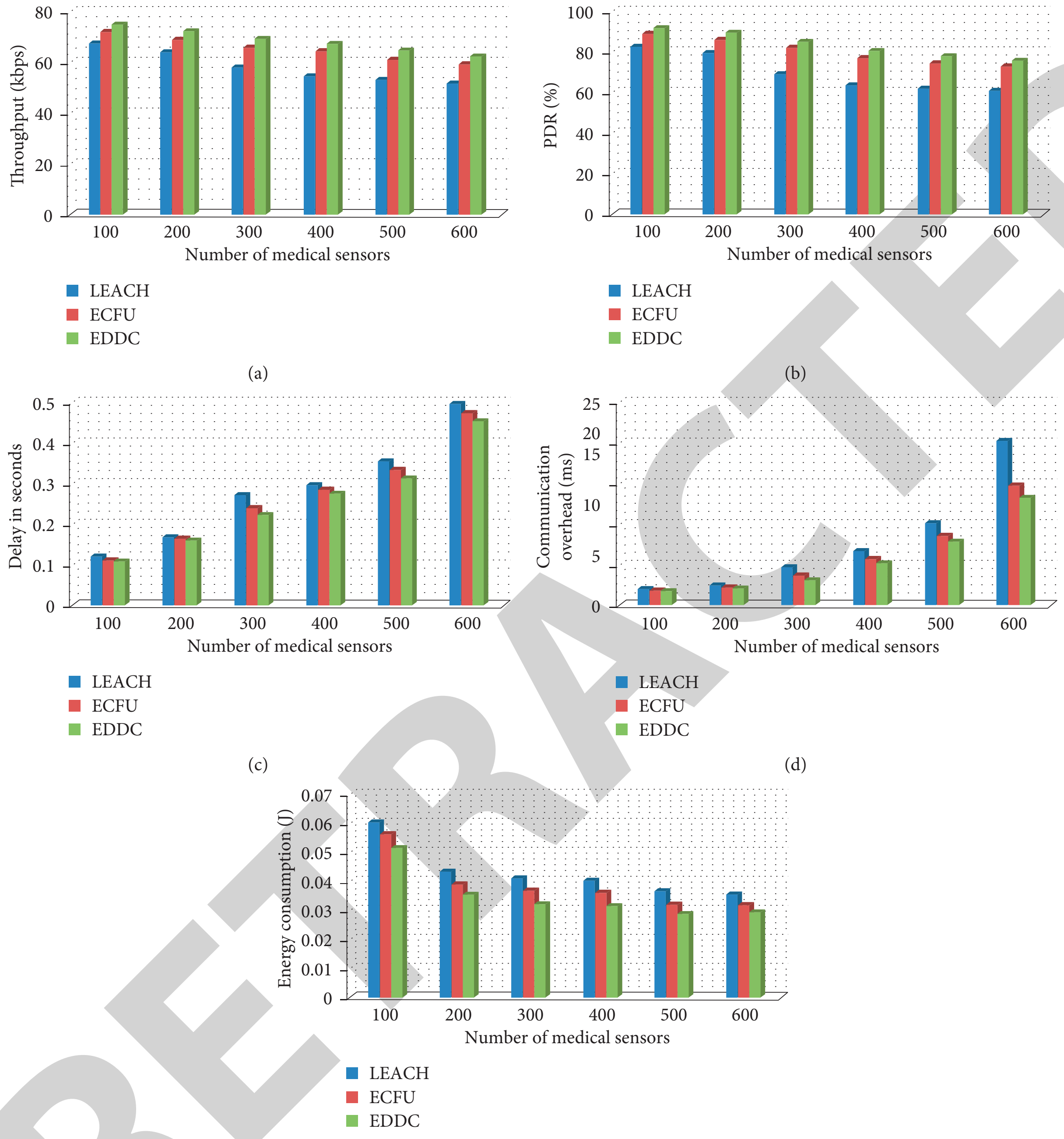
(e)

Figure 5: Performance measurement under replay attack. (a) Throughput. (b) PDR. (c) Delay. (d) Communication overhead. (e) Energy consumption.

checkups machines, computers, printers; and communication overhead values are defined on the *y*-axis measured by EDDC protocol, LEACH, and ECFU protocol. The figures show that the communication overhead time of the LEACH is more than the ECFU protocol and the communication overhead time of the proposed EDDC protocol is less than LEACH and ECFU protocol under DDoS, replay, and eavesdrop attacks. The communication overhead (CO) is calculated by [47]

$$\mathrm{CO} = \frac{\left(\text{Communication time}_A - \text{Communication time}_b\right)}{\text{Communication time}_A}. \tag{10}$$

The energy consumption of the system under DDoS attack is shown in Figure 4(e), the energy consumption of the system under replay attack is shown in Figure 5(e), and the energy consumption of the system under eavesdrop attack is shown in Figure 6(e). The number of medical



(a)

(b)

(c)

(d)

(e)

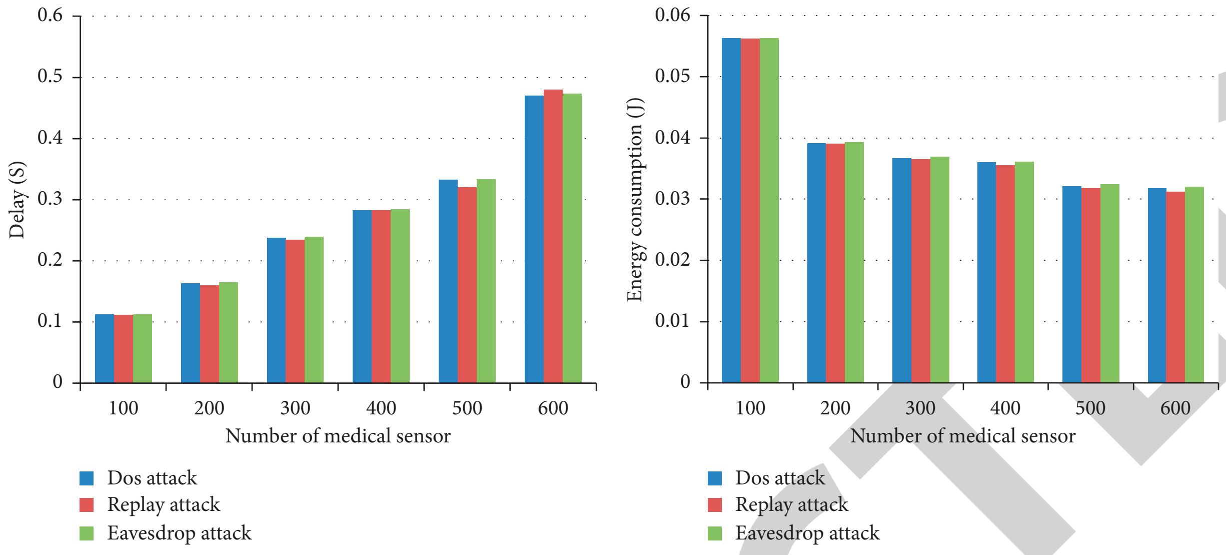
FIGURE 6: Performance measurement under eavesdrop attack. (a) Throughput. (b) PDR. (c) Delay. (d) Communication overhead. (e) Energy consumption.

sensors is defined on the x-axis; it represents the medical sensor nodes such as cameras, health checkups machines, computers, printers; and energy consumption values are defined on the y-axis measured by EDDC protocol, LEACH, and ECFU protocol. The figures show that the energy consumption of the LEACH is more than the ECFU protocol and the energy consumption of the proposed EDDC protocol is less than LEACH and ECFU protocol under DDoS, replay, and eavesdrop attacks. The energy consumption is calculated by [47]

$$P_{\text{con}} = \sum_{i=1}^{\text{node}} T_e + R_e + W, \tag{11}$$

where T_e = energy consumption while packet is sent, R_e = energy consumption while receiving packets, and W_e = energy consumption while waiting for packets.

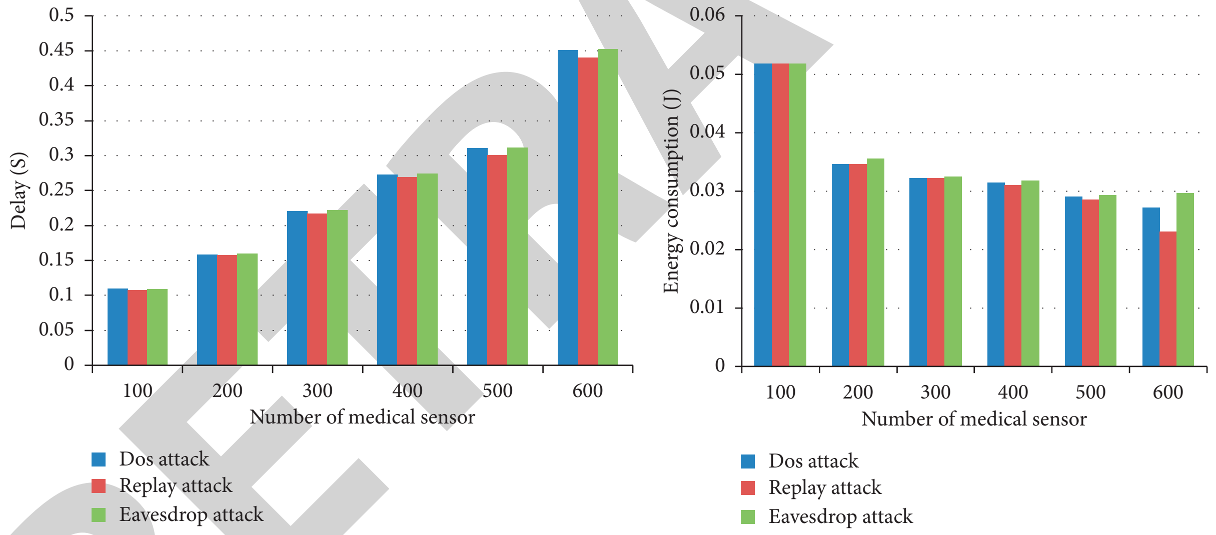
The performance measurements for ECFU and EDDC for DDoS, replay, and eavesdrop attacks are analyzed separately and shown in Figures 7 and 8. To increase network



(a)

(b)

Figure 7: Performance measurement under DDoS, replay, and eavesdrop attack using ECFU algorithm. (a) Delay. (b) Energy consumption.



(a)

(b)

Figure 8: Performance measurement under DDoS, replay, and eavesdrop attack using EDDC algorithm. (a) Delay. (b) Energy consumption.

lifetime and reduce transmission delay, the performance of the network for delay and energy consumption in presence of attacking nodes are discussed and analyzed. The comparative results indicate proposed system shows improved performance metrics in presence of attacks.

6. Conclusion

The proposed Efficient Device type Detection and Classification (EDDC) protocol is designed for device identification and classification in M2M communication for secure telerobotic surgery. The designed protocol meets the strict security requirement of the telerobotic surgical system. The contribution of this article is multifaceted. This study shows accurately detection of malicious devices in M2M communication. To identify the node as legitimate or attacker, design the trust-score computation technique to compute the periodic trust score of each node. Then, perform classification on identified devices at periodic intervals based on their similarities. The machine learning classifier artificial neural network (ANN) is built for efficient data transmission. The article presents the effectiveness of this method with operational parameters. It shows the proposed EDDC protocol performs better than LEACH and ECFU protocol under DDoS, replay, and eavesdrop attacks. The proposed algorithm for device identification and classification is energy-efficient and scalable. We conducted a simulation as a proof of the concept where we showed the quality of service parameters is influenced by changing the number of nodes, packet size, and simulation time. Malicious and nonmalicious nodes classification for secure telerobotic surgery further contributes to efficient and secure communication. Future plan is to calculate the trust score of each device for designing access control to telerobotic devices for high-density networks.

Data Availability

Data are available on request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Z. Meng, Z. Wu, C. Muvianto, and J. Gray, “A data-oriented M2M messaging mechanism for industrial IoT applications,” *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 236–246, 2017.

[2] M. Kaur and S. Kadam, “Discovery of resources over Cloud using MADM approaches,” *International Journal for Engineering Modelling*, vol. 32, pp. 83–92, 2019.

[3] K. Jairath, N. Singh, V. Jagota, and M. Shabaz, “Compact ultrawide band metamaterial-inspired split ring resonator structure loaded band notched antenna,” *Mathematical Problems in Engineering*, vol. 2021, Article ID 5174455, 12 pages, 2021.

[4] A. Kishor, C. Chakraborty, and W. Jeberson, “Reinforcement learning for medical information processing over heterogeneous networks,” *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 23983–24004, 2021.

[5] M. Kaur and S. Kadam, “A novel multi-objective bacteria foraging optimization algorithm (MOBFOA) for multi-objective scheduling,” *Applied Soft Computing*, vol. 66, pp. 183–195, 2018.

[6] M. Islam, D. A. Atputharuban, R. Ramesh, and H. Ren, “Real-time instrument segmentation in robotic surgery using auxiliary supervised deep adversarial learning,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 2188–2195, 2019.

[7] A. A. Shvets, A. Rakhlin, A. A. Kalinin, and V. I. Iglovikov, “Automatic instrument segmentation in robot-assisted surgery using deep learning,” in *Proceedings of IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 624–628, IEEE, Orlando, FL, USA, December 2018.

[8] M. Kaur, “Elitist multi-objective bacterial foraging evolutionary algorithm for multi-criteria based grid scheduling problem,” in *Prpoceedings of the 2016 International Conference on Internet of Things and Applications (IOTA)*, pp. 431–436, Pune, India, January 2016.

[9] OECD Report, “machine-to-machine communications: connecting billions of devices,” *OECD Digital Economy Papers*, vol. 192, 2012.

[10] T. Kaur and D. Kumar, “Computational intelligence-based energy efficient routing protocols with QoS assurance for wireless sensor networks: a survey,” *International Journal of Wireless and Mobile Computing*, vol. 16, no. 2, pp. 172–193, 2019.

[11] M. Chen, J. Wan, and F. Li, “Machine-to-Machine communications: architectures, standards and applications,” *KSII Transactions on Internet & Information Systems*, vol. 6, pp. 480–497, 2012.

[12] Z. Yan Zhang, Y. Rong Yu, X. Shengli Xie, Y. Wenqing Yao, X. Yang Xiao, and M. Guizani, “Home M2M networks: architectures, standards, and QoS improvement,” *IEEE Communications Magazine*, vol. 49, no. 4, pp. 44–52, 2011.

[13] S.-J. Jung, R. Myllyla, and W.-Y. Chung, “Wireless machine-to-machine healthcare solution using android mobile devices in global networks,” *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1419–1424, 2013.

[14] J. Granjal, E. Monteiro, and J. S. Silva, “Security issues and approaches on wireless M2M systems,” in *Wireless Networks and Security*, S. Khan and A.-S. Khan Pathan, Eds., Springer, Berlin Germany, pp. 133–164, 2013.

[15] C. Lai, L. Hui, Z. Yueyu, and C. Jin, “Security issues on machine to machine communications,” *KSII Transactions on Internet & Information Systems*, vol. 6, pp. 498–514, 2012.

[16] M. P. Lokhande and D. D. Patil, “Security threats in M2M framework of IoT,” *International Journal of Advanced Science and Technology*, vol. 29, no. 8, pp. 1809–1823, 2020.

[17] X. Sun, S. Men, C. Zhao, and Z. Zhou, “A Security authentication scheme in machine-to-machine home network service,” *Security and Communication Networks*, vol. 8, pp. 2678–2686, 2012.

[18] W. Ren, L. Yu, L. Ma, and Y. Ren, “RISE: a RelIable and SEcure scheme for wireless Machine to Machine communications,” *Tsinghua Science and Technology*, vol. 18, no. 1, pp. 100–117, 2013.

[19] M. Yair, M. Bohadana, A. Shabtai, M. Ochoa et al., “Detection of unauthorized IoT devices using machine learning techniques,” 2017, https://arxiv.org/abs/1709.04647.

[20] A. Sivanathan, D. Sherratt, H. Hassan Gharakheili, R. Adam, C. Wijenayake, and A. Vishwa, “Characterizing and classifying IoT traffic in smart cities and campuses,” in *Proceedings f the IEEE INFOCOM Workshop Smart Cities Urban Computing*, pp. 1–6, Atlanta, GA, USA, May 2017.

[21] V. Bhatia, S. Kaur, K. Sharma, P. Rattan, V. Jagota, and M. A. Kemal, “Design and simulation of capacitive MEMS switch for ka band Application,” *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 2021513, 8 pages, 2021.

[22] M. Miettinen, S. Marchal, I. Hafeez, A.-R. Sadeghi et al., “Iot SENTINEL: automated device-type identification for security enforcement in IoT,” in *Proceedings of the IEEE 37th International Conference on Distributed Computing Systems*, pp. 2177–2184, Atlanta, GA, USA, June 2017.

[23] X. Zhang, L. Yao, C. Huang, Q. Z. Sheng, and X. Wang, “Intent recognition in smart living through deep recurrent neural networks,” in *Proceedings of the International Conference on Neural Information Processing (ICONIP)*, pp. 748–758, Guangzhou, China, November 2017.

[24] B. M. Dickens and R. J. Cook, “Legal and ethical issues in telemedicine and robotics,” *International Journal of Gynecology & Obstetrics*, vol. 94, no. 1, pp. 73–78, 2006.

[25] H. H. King, K. Tadano, R. Donlin et al., “Preliminary protocol for interoperable telesurgery,” in *Proceedings of the international conference on advanced robotics*, pp. 1–6, Munich, Germany, June 2009.