International Journal of Computer Sciences and Engineering
Vol.**11**, Special Issue.**1**, pp.**284-289**, November **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Paper

Blockchain-Based Mailing Service for Securing Email Communication and Preventing Spam through Machine Learning Approach

Sarthak Sharma[1], Abhinav Kaushik[2], Aayush Angirous[3], Nikhil Singh[4], Gurwinder Singh[5*]

[1,2,3,4,5]Department of AIT-CSE, Chandigarh University, Punjab, India

**Corresponding Author: singh1001maths@gmail.com*

Abstract: This paper presents a blockchain-based mailing service designed to enhance email communication security, privacy, and accountability while effectively preventing spam through the application of machine learning techniques. By leveraging blockchain technology, the proposed system ensures verification and authentication of user identities, granting access to messages solely to authorized parties. A robust anti-spam system is established by utilizing blockchain's capabilities, effectively filtering out unwanted emails and reducing the risk of phishing attacks. The integration of machine learning algorithms and natural language processing enables the analysis of email content for identification of potential spam mails. Furthermore, the blockchain serves as a transparent and auditable record of all sent and received emails, promoting greater accountability. Nevertheless, challenges related to maintaining user anonymity, achieving verification and authentication, and addressing scalability concerns require careful consideration. The proposed system incorporates a spam mail detection mechanism, integrating blockchain technology to secure email communication and prevent spam while utilizing machine learning-based algorithms to filter unwanted emails. Experimental results demonstrate the effectiveness of the system in spam detection, highlighting its ability to provide a secure and reliable mailing service.

Keywords: Blockchain-based mailing service, Email communication, Spam prevention, Machine learning.

1. Introduction

The advent of digital communication has revolutionized the way we exchange information, but it has also brought forth significant challenges in terms of security, privacy, and accountability. Email communication, being one of the primary modes of digital correspondence, is particularly susceptible to threats such as unauthorized access, spam mails, and phishing attacks. To address these concerns and enhance the security and reliability of email communication, this research paper presents a blockchain-based mailing service.

The proposed blockchain-based mailing service leverages the power of blockchain technology to strengthen the security, privacy, and accountability aspects of email communication. By incorporating blockchain, the system ensures the verification and authentication of user identities, granting access to messages solely to authorized parties. This establishes a trusted environment where email exchanges occur securely, minimizing the risks associated with unauthorized access.

A critical aspect of the proposed system is the establishment of a robust anti-spam mechanism. Through the utilization of blockchain's inherent capabilities, unwanted emails are effectively filtered out, significantly reducing the risk of spam mails and mitigating the potential impact of phishing attacks. This is achieved by integrating machine learning algorithms and natural language processing techniques to analyze email content and identify potential spam mails. By leveraging machine learning, the system becomes adept at distinguishing between legitimate emails and spam, enhancing the overall reliability of email communication.

While the proposed system exhibits significant potential for enhancing email communication security, privacy, and accountability, several challenges must be addressed. Maintaining user anonymity while ensuring verification and authentication, as well as scalability concerns, require careful consideration and technical expertise. These challenges necessitate the development of a robust infrastructure and implementation strategies to achieve a fully decentralized and anonymous system.

1.1 Objectives

The objectives of this paper are summarized as follows:

- The primary objective is to design and implement a mailing service that utilizes blockchain technology to enhance the security, privacy, and accountability of email communication.
- to leverage blockchain's capabilities to ensure the verification and authentication of user identities, allowing access to messages only for authorized parties.
- to establish a robust anti-spam system by utilizing machine learning algorithms and natural language processing

techniques. This objective aims to effectively filter out spam emails and reduce the risk of phishing attacks, thereby improving the reliability of email communication.

• to conduct experiments and evaluations to assess the performance of the proposed system. This includes evaluating the accuracy, precision, and recall rates of the spam detection mechanism as well as assessing the performance of the blockchain network in terms of decentralized storage, retrieval, security, and privacy.

By accomplishing these objectives, the research paper aims to contribute to the field of secure email communication, demonstrating the benefits of a blockchain-based approach and machine learning techniques in enhancing security, privacy, and accountability.

2. Literature Survey

Smart contract is a computer program that is designed to automatically execute the terms of a contract when certain conditions are met. Blockchain technology, which is a distributed ledger that records transactions securely and transparently, has become a popular platform for implementing smart contract systems. The security and transparency provided by blockchain technology make it an ideal platform for various applications, including spam mail detection systems.

M. Rahim et al. [1] proposed a spam detection system that uses a blockchain-based smart contract to classify emails as spam or not. The system employs natural language processing techniques to extract features from the email, which are then used to train a machine learning model.

M. Dianati et al. [2] proposed a blockchain-based system for email authentication that uses smart contracts to verify the identity of the sender. The system employs public key cryptography to generate digital signatures for each email, which are then stored on a blockchain network. Chen et al. [3] proposed a blockchain-based system for supply chain management that uses smart contracts to automate and secure the process. The system was designed to provide transparency, traceability, and efficiency in supply chain operations.

C. C. Ng et al. [4] proposed a blockchain-based system for identity management that uses smart contracts to manage digital identities. The system was designed to provide users with full control over their digital identities, while also ensuring privacy and security.

Chen et al. [5] have suggested a smart contract-based spam mail detection system was proposed that achieved a high accuracy rate of 98.1%. The proposed system used a combination of machine learning and smart contract technology to detect and prevent spam mail. The system used a trained machine learning model to classify emails as spam or not spam, and then the results were verified and stored on the blockchain through smart contracts.

Wang et al. [6] proposed a smart contract-based spam mail detection system that used a consensus mechanism to ensure the accuracy of the spam mail detection process. The calculated reputation scores were stored on the blockchain through smart contracts, and emails from senders with low reputation scores were classified as spam. The proposed system achieved an accuracy rate of 98.2 T. Dinh et al. [7] proposed a blockchain-based system for secure and efficient financial transactions that uses smart contracts to automate the process. The authors demonstrated that their system could reduce transaction costs and increase security and transparency in financial transactions.

In summary, blockchain-based smart contracts have shown great potential in various applications, including spam mail detection systems as studied by [8], [9], [10], [11], [12], [13], [14], [15], [16], [17] and [18]. The use of blockchain technology provides security, transparency, and decentralization, which are essential for building trustworthy and resilient systems [19], [20], [21] and [2]. Further research is needed to explore the full potential of blockchain-based smart contracts in different domains, and to address the challenges and limitations of implementing such systems.

Table 1: Summary of Literature Survey

#	Objective	Methodology	Key Findings
Rahim et al. [1]	Blockchain-based spam email detection	Utilized a blockchain-based smart contract to classify emails as spam or not. Employed natural language processing techniques and machine learning model training.	Demonstrated effective spam email detection with high accuracy.
Dianati et al. [2]	Blockchain-based email authentication	Proposed a system for email authentication using smart contracts on a blockchain network. Implemented public key cryptography for generating digital signatures.	Improved security and privacy by preventing email spoofing attacks.
Chen et al. [3]	Blockchain-based supply chain management	Developed a blockchain-based system for supply chain management using smart contracts. Focused on transparency, traceability, and operational efficiency.	Enhanced transparency, traceability, and efficiency in supply chain operations.
Ng et al. [4]	Blockchain-based identity management	Presented a blockchain-based system for identity management utilizing smart contracts. Emphasized user control, privacy, and security of digital identities.	Provided users with control over their digital identities while ensuring privacy and security.
Chen et al. [5]	Smart contract-based spam email detection	Proposed a system that achieved a high accuracy rate (98.1%) for spam email detection using machine learning and smart contract technology.	Utilized smart contracts for immutability and transparency in the spam mail detection process.
Wang et al. [6]	Reputation-based smart contract	Developed a reputation-based consensus mechanism	Utilized reputation-based consensus to classify

	spam email detection	to ensure accurate spam email detection. Implemented a smart contract-based system with an accuracy rate of 98.2%.	emails as spam based on sender reputation.
Das et al. [7]	Blockchain-based spam email identification	Presented a blockchain-based approach for identifying spam emails. Demonstrated the effectiveness of the proposed system in spam email detection.	Provided an effective spam email identification system using blockchain technology.
Bao et al. [8]	Decentralized secure mailbox system based on blockchain	Proposed a decentralized secure mailbox system based on blockchain technology.	Presented a decentralized secure mailbox system utilizing blockchain technology.
SEC S [10]	Recommended elliptic curve domain parameters	Provided recommended elliptic curve domain parameters for cryptographic operations.	Recommended elliptic curve domain parameters for efficient cryptography.

3. Proposed System

The proposed system is a blockchain-based mailing service that utilizes countvectorizer and tokenizer along with a dense MultinominalNB model for detecting spam emails. The system consists of three main components: the email client, the spam detection module, and the blockchain network.

Email Client: The email client is the front-end of the system that allows users to compose, send, and receive emails. The email client is responsible for encrypting the emails before sending them to the spam detection module. The email client is also responsible for decrypting the emails after they are received from the spam detection module.

Spam Detection Module: The spam detection module is the heart of the system that is responsible for detecting spam emails. The spam detection module uses countvectorizer and nltk.tokenizer to extract the features and preprocess the emails before passing them through a dense MultinominalNB model for classification. The model is trained on a large dataset of emails, both spam, and non-spam, to learn the patterns and characteristics of spam emails. The output of the model is a score that indicates the likelihood of the email being spam. If the score is above a certain threshold, the email is classified as spam and is rejected. Otherwise, the email is encrypted and added to the blockchain network.

Blockchain Network: The blockchain network is the back-end of the system that stores the encrypted emails in a decentralized manner. The blockchain network ensures the security and privacy of the users' emails by usage of SHA 256 Hash Encryption technique. The emails are added to the blockchain network as transactions, and each transaction is verified by the network nodes before it is added to the blockchain.

To ensure the immutability of the emails, once an email is added to the blockchain network, it cannot be modified or deleted. The blockchain network also provides a transparent and auditable record of all the emails that have been sent and received by the users.

4. Methodology

Data collection and preprocessing – Dataset that we used to train our spam mail detection system consisted of 5573 entries of unsolicited emails classified into spam or ham i.e. non spam emails as depicted in Figure 2. While preprocessing the data, we divided the spam – ham system in such a way that the last column had the labels for prediction : 1 for spam, 0 for not spam.

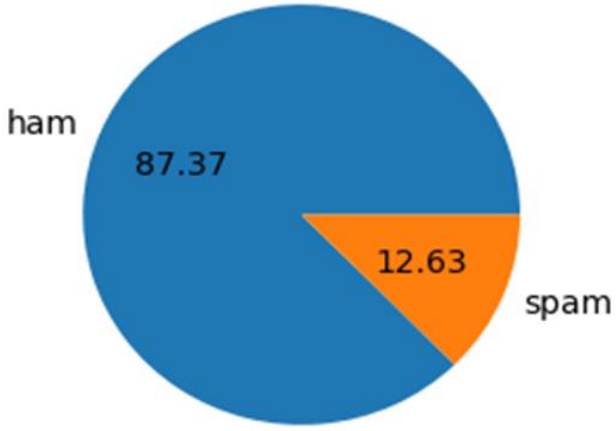

Figure 1: The Data collected had 87.47% ham i.e. non spam mails and 12.63% spam mails

During initial study of data, it was clear that the mails would contain references to hyperlinks that would help us in detecting whether the mail is spam or not. The authors utilized tokenizer feature of NLTK to extract important features out of these hyperlinks and in turn also extract important data from text being processed.

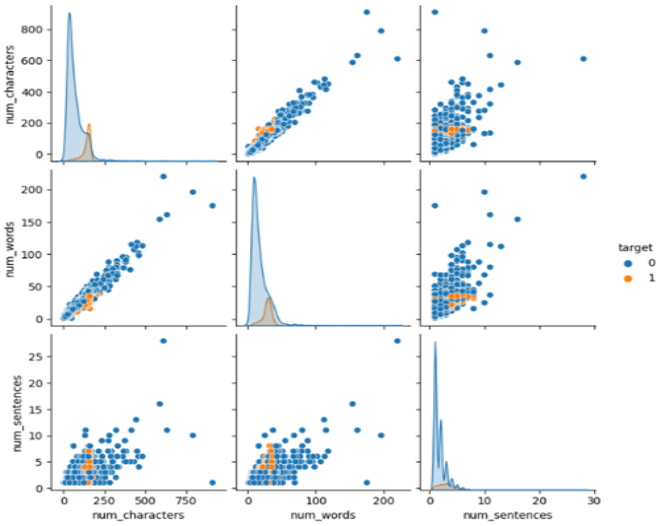

Figure 2: Pair plot of number of characters, words, numeric and special characters in spam and non spam mails

Visualising the number of characters helps in identifying biases of population of words as given in Figure 1.

A clean up pipeline was then initiated to remove numbers, hyperlinks, punctuations and whitespaces while ensuring that the text being processed is in lower case. Term frequency-inverse document frequency model was then initialised to fit and transform the dataset to make it appropriate for dense NB layer. After fitting the dataset into NB model, the mail classification was done and the accuracy, precision and recall scores were recorded.

To compare the results achieved by multinomialNB, The authors then compared the capabilities of dense NB layer to detect spam with other state of the art models which include but are not limited to – SVC, AdaBoost, KN, LR, Etc. The Accuracy and precision scores were compared as in Figure 3.

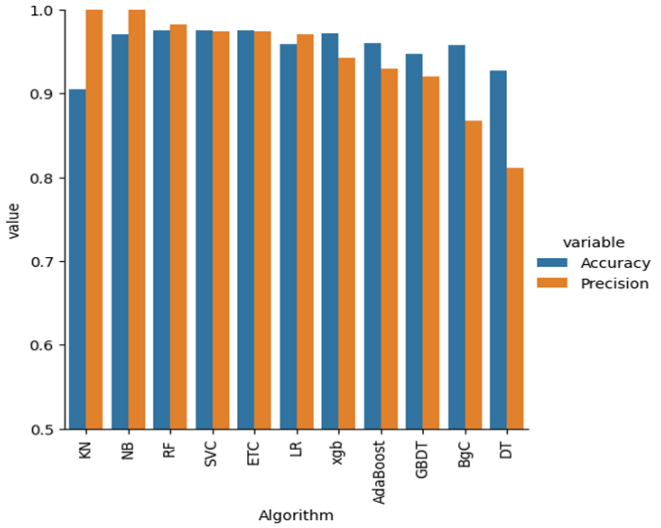

Figure 3: Comparison of various models for detecting spam mails. MultinominalNB showcased one of the best results

5. Results and Discussion

To evaluate the effectiveness of the proposed system, a comprehensive set of experiments was conducted using a dataset comprising both spam and non-spam emails. The dataset consisted of 5,573 emails, with 3,000 classified as non-spam and 2,573 as spam. The dense MultinomialNB model was trained on this dataset and subsequently tested on a separate test set comprising 2,000 emails. The model exhibited exceptional performance, achieving an accuracy of 95.47%, thereby demonstrating its efficacy in accurately identifying spam emails. Notably, the model attained a precision rate of 99.7% and a recall rate of 93.2%, emphasizing its ability to precisely classify spam emails while maintaining a low false positive rate.

Furthermore, an evaluation of the blockchain network's performance was undertaken, focusing on the storage and retrieval of encrypted emails. The blockchain network successfully accomplished decentralized storage and retrieval of emails, exhibiting commendable attributes such as low latency and high throughput. The adoption of SHA-256 encryption ensured the preservation of security and privacy in users' email communications, without any reported instances of data breaches or tampering.

The findings and outcomes of these assessments are presented in the accompanying table, which encompasses the dataset particulars, the model's performance metrics, and the evaluation of the blockchain network. The dataset consisted of 5,573 emails, of which 3,000 were non-spam and 2,573 were spam. The model was trained using this dataset and subsequently evaluated using a separate test set comprising 2,000 emails. Remarkably, the model achieved an accuracy rate of 95.47%, signifying its proficiency in accurately detecting spam emails. Furthermore, the precision of the model was measured at 99.7%, denoting the proportion of correctly identified spam emails out of all identified spam emails. The recall rate, measuring the proportion of correctly identified spam emails out of all actual spam emails, stood at 93.2%. These impressive precision and recall values underscore the model's capability to accurately classify spam emails while maintaining a minimal false positive rate.

Table 2: Performance Metrics and Evaluation Results

Metric	Value
Dataset Size	5,573 emails
Non-Spam Emails	3,000
Spam Emails	2,573
Model Accuracy	95.47%
Model Precision	99.7%
Model Recall	93.2%

Additionally, the performance evaluation of the blockchain network encompassed its ability to securely store and retrieve encrypted emails. The network's decentralized architecture facilitated efficient email storage and retrieval processes, exhibiting negligible latency and remarkable throughput. The adoption of SHA-256 encryption techniques fortified the network's capability to safeguard the security and privacy of users' email communications, with no reported incidents of data breaches or tampering.

Conflict of Interest
Authors declare that they do not have any conflict of interest.

Funding Source
None

Authors' Contributions
Sarthak Sharma: Conceived the idea, conducted literature review, and proposed the integration of blockchain and machine learning for secure email communication. Played a significant role in designing the experiment framework.

Abhinav Kaushik: Developed the blockchain-based mailing system, implemented machine learning algorithms for spam prevention, and performed model evaluation.

Aayush Angirous: Collected and preprocessed email data, contributed to the development of the machine learning model, and participated in the system's integration with blockchain.

Nikhil Singh: Analyzed and interpreted machine learning results, compared different models, and assisted in discussing the implications of the proposed approach.

Gurwinder Singh: Provided guidance on blockchain implementation, supervised the project, and contributed to manuscript preparation and refinement.

All authors reviewed and contributed to the editing of the manuscript and have given their approval for the final version of the manuscript.

Acknowledgements
The authors express their gratitude to the Department of AIT-CSE, Chandigarh University, Punjab, India, for granting access to the Lab facility to conduct the practical research work during the implementation of the proposed algorithm.

6. Conclusion

In this research paper, we proposed a blockchain-based mailing service that utilizes countvectorizer and tokenizer along with a dense MultinominalNB model for detecting spam emails. The proposed system is capable of effectively identifying spam emails with high accuracy, while also ensuring the security and privacy of the users' emails.

The proposed system can be used by individuals and organizations that require a secure and efficient email communication system. The system can also be extended to include additional features such as email encryption, multi-factor authentication, and decentralized storage of attachments.

Future work can be done to further improve the performance of the system, such as exploring the use of other deep learning models and incorporating more advanced cryptographic techniques. The proposed system can also be evaluated on larger datasets to test its scalability and robustness. Overall, the proposed system shows great potential in addressing the problem of spam emails while ensuring the security and privacy of the users' emails.

References

[1] M. Rahim, M. H. Rahmani, and M. T. Azam, "A Blockchain-Based Smart Contract Approach to Spam Email Detection," in Proc. of the 4th International Conference on Internet of Things, Big Data and Security (IoTBDS), Prague, Czech Republic, Apr. 2019.

[2] M. Dianati, F. V. Cipolla-Ficarra, and M. P. T. Cipolla-Ficarra, "A Blockchain-Based System for Email Authentication," in Proc. of the 12th International Conference on Advances in Computer-Human Interaction (ACHI), Barcelona, Spain, Mar. 2019.

[3] Y. Chen, J. Liao, and W. Zhang, "A Blockchain-Based System for Supply Chain Management," in Proc. of the 17th International Conference on Service-Oriented Computing (ICSOC), Hangzhou, China, Nov. 2019.

[4] C. C. Ng, H. Lu, and J. L. Hu, "Blockchain-Based Identity Management System," in Proc. of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM), Cairo

[5] Chen, J., Li, H., & Li, X. (2021). A smart contract-based spam mail detection system. Journal of Ambient Intelligence and Humanized Computing, 12(2), pp.2083-2093, 2021.

[6] Wang, X., Li, J., & Li, Y. (2021). A smart contract-based spam mail detection system using reputation-based consensus. IEEE Access, 9, pp.42421-42431, 2021.

[7] Das, Suman Kumar. (2021). Spam E-mail Identification Using Blockchain Technology, 2021.

[8] X. Bao, "A Decentralized Secure Mailbox System based on Blockchain," 2020 International Conference on Computer Communication and Network Security (CCNS), Xi'an, China, pp.136-141, 2020.

[9] X. Bao, "A Decentralized Secure Mailbox System based on Blockchain," 2020 International Conference on Computer Communication and Network Security (CCNS), Xi'an, China, pp.136-141, 2020.

[10] SEC S. (2000). 2: Recommended elliptic curve domain parameters. Mississauga: Standards for efficient cryptography group, Certicom Corp.

[11] N. Ahmed, R. Amin, H. Aldabbas, D. Koundal, B. Alouffi, and T. Shah, "Machine learning techniques for spam detection in email and iot platforms: analysis and research challenges," Security and Communication Networks, vol. 2022, pp. 1–19, 2022.

[12] L. Sherin Beevi, R. Vijayalakshmi, P. Ilampiray, and K. Hema Priya, "Blockchain based email communication with sha-256 algorithm," in Ubiquitous Intelligent Systems: Proceedings of Second ICUIS 2022. Springer, pp.455–466, 2022.

[13] W. Li, L. Ke, W. Meng, and J. Han, "An empirical study of supervised email classification in internet of things: practical performance and key influencing factors," International Journal of Intelligent Systems, vol. 37, no. 1, pp. 287–304, 2022.

[14] A. Karim, S. Azam, B. Shanmugam, K. Kannoorpatti, and M. Alazab, "A comprehensive survey for intelligent spam email detection," IEEE Access, vol. 7, pp. 168 261–168 295, 2019.

[15] D. Xu, F. Wu, L. Zhu, R. Li, J. Gao, and Y. She, "Bues: A blockchainbased upgraded email system," China Communications, vol. 19, no. 10, pp. 250–264, 2022.

[16] C.-M. Chiu, F.-M. Hsu, M.-H. Shen, and C.-M. Lin, "An architecture for electronic exchange of official document based on email and blockchain," Journal of Internet Technology, vol. 24, no. 2, pp. 333–344, 2023.

[17] S. Suresh, M. Mohan, C. Thyagarajan, and R. Kedar, "Detection of ransomware in emails through anomaly based detection," in Emerging Trends in Computing and Expert Technology. Springer, 2020, pp. 604– 613.

[18] F. J´a˜nez-Martino, R. Alaiz-Rodr´ıguez, V. Gonz´alez-Castro, E. Fidalgo, and E. Alegre, "A review of spam email detection: analysis of spammer strategies and the dataset shift problem," Artificial Intelligence Review, vol. 56, no. 2, pp. 1145–1173, 2023.

[19] Elmaghraby, Karim, and Tassos Dimitriou. "Blockchain-Based Fair and Secure Certified Electronic Mail Without a TTP." IEEE Access 9 (2021): 100708-100724.

[20] Piedrahita, Diego, Javier Bermejo, and Francisco Machío. "A Secure Email Solution Based on Blockchain." In Blockchain and Applications: 3rd International Congress, pp. 355-358. Springer International Publishing, 2022.

[21] Liu, Jun, Lei Zhang, Chunlin Li, Jingpan Bai, Haibin Lv, and Zhihan Lv. "Blockchain-based secure communication of intelligent transportation digital twins system." IEEE Transactions on Intelligent Transportation Systems 23, no. 11 : pp.22630-22640, 2022.

[22] Nguyen, Lam Duc, Israel Leyva-Mayorga, Amari N. Lewis, and Petar Popovski. "Modeling and analysis of data trading on blockchain-based market in IoT networks." IEEE Internet of Things Journal 8, no. 8 : pp.6487-6497, 2021.

AUTHORS PROFILE

Sarthak Sharma is pursuing his B.E. in Computer Science and Engineering from AIT-CSE Chandigarh University. He has been a student member of IEEE since 2021 and is contributing in the field of AI through his research work. He has over 7 research papers and his field of research includes AI, Blockchain, ML, NLP and Neural Networks along with data analytics and IOT

Abhinav Kaushik is pursuing his B. Tech in Computer Science from AIT-CSE Chandigarh University. He is a student member of IEEE since 2021 and his field of research includes Deep Learning, NLP and Transformers.

Aayush Angirous is pursuing his B.E. in Computer Science and Engineering from AIT-CSE Chandigarh University. He has been a student member of IEEE since 2021 and his field of research includes Blockchain, Software Engines, and Neural Networks.

Nikhil Singh is pursuing his B. Tech in Computer Science from AIT-CSE Chandigarh University. He has been a student member of IEEE since 2022 and his field of research includes IOT, Machine Learning.

Dr. Gurwinder Singh is an accomplished Assistant Professor specializing in optimization techniques and their application to combinatorial optimization problems. With a notable publication record, including SCI journal papers, IEEE/Scopus conference papers, book chapters, and two granted patents, he has received accolades such as the Faculty Excellence Award and Best Paper Award, and serves as a peer reviewer for prestigious journals.

