Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 29s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

DNNLSTM Algorithms for Comparative Study on Machine Learning Approach for Malicious Study

Shabeena Nafees[1*], Anil Kumar Pandey[2], Satya Bhushan Verma[3]

[1]Department of Computer Science and Engineering, Shri ram swaroop memorial University Lucknow, India,
Email ID: qds186@gmail.com,,anipandey@gmail.com, Email ID: satyabverma1@gmail.com

Cite this paper as: Shabeena Nafees, Anil Kumar Pandey, Satya Bhushan Verma, (2025) DNNLSTM Algorithms for Comparative Study on Machine Learning Approach for Malicious Study. *Journal of Neonatal Surgery*, 14 (29s), 1012-1019.

ABSTRACT
E-commerce systems, utilizing the Internet for various functions such as banking, shopping, and sales, offer numerous advantages to customers. However, security remains a challenge, as the numerous advantages of these systems come with the potential for cyber threats. Therefore, it is crucial to carefully consider the security measures in these systems. Online transactions in the e-commerce system are increasing rapidly, posing a significant risk to consumers' personal information. To ensure the Internet safety, it is crucial to address vulnerabilities in the system's components, which can be accidental or malicious, at entry and exit points. The number of threats in the system was well managed through the usage of machine learning techniques, ensuring efficient detection and prevention. E-commerce systems require robust security measures to ensure their functionality and safety, a task that researchers working in this field are well-equipped to tackle. This paper focuses on analysing security in e-commerce systems using machine learning.

Keywords: *E-commerce systems, Security, Machine learning, Malicious Code*

1. INTRODUCTION

Technology significantly contributes to daily life by transforming traditional business methods and influencing product quality and cost. Wang et al. highlight the application of e-commerce, which involves online activities through websites, email, and other technologies . E-commerce involves electronic markets, online retailing, and online auctions, allowing customers to buy products or services remotely. However, challenges such as cyber security remain as researchers continue to enhance the process for greater profitability. Overall, technology plays a crucial role in our daily lives [1]. Internet security in e-commerce involves implementing security protocols to execute transactions safely. Without proper security, online risks and payment frauds can occur. Small e-commerce businesses face significant risks due to inadequate security measures. Despite built-in security features, many businesses close within no time due to numerous frauds and attacks. Malicious software poses a significant threat to modern technology, including viruses, Trojan horses, ransom ware, spyware, adware, rogue software, wipers, and shareware [2]. Machine learning algorithms like Naive Byes can detect harmful traffic on computer systems and improve network security [3].

2. REVIEW OF LITERATURE

The problem of URL categorization has been the subject of research in recent years, and these studies have used machine learning to learn the border between decision classes [4]. Through the use of under sampling and a reduction in the number of characteristics to nine, the work done by (M. Alazab et al., 2014) increased URL classification rates by 97% [5]. The procedure consisted of eliminating duplicates and conforming to the established security requirements, which ultimately led to a categorization process that was more precise [6]. Researchers utilizing machine learning categorize URLs as either harmful or benign. They do this by describing each category with a set of criteria and understanding the border between decision classes by using classification algorithms. The classification of online sites was accomplished by D. Xueusing URLs, which is a far quicker technique than content fetching or text parsing [7]. The URL was broken up into tokens, which made it possible to extract classification characteristics and led to an increase in the rate of accurate classification. Using the lexical and host-based characteristics of URLs, an automated approach was built to identify fraudulent websites. The method was successful in achieving a high classification rate (between 95% and 99%) and a low percentage of false positives. A real-time URL feature collection system was created in order to manage millions of URLs with ever-changing characteristics. This system was successful in reaching a classification rate of 99% on a balanced dataset. The Cost-Sensitive Online Active Learning (CSOAL) system that Zhao and Hoi developed solves class imbalance in URL detection by optimizing two cost-sensitive measures using just 0.5% of the available data [8].

Table 1: Literature Review

Ref:	*Year*	*Publisher*	*Research Contributions*	*Strengths*	*Limitations*
[18]	2022	Hindawi	Researchers have proposed a cloud-based infrastructure to provide efficient and secure transactional communication.	Efficient detection security attacks	High resource utilization of the proposed system
[19]	2022	Wiley	A threat analysis framework is designed	End-to-End secure communication	Computation overhead increases
[20]	2022	Springer	An efficient communication model is presented	Prevention against malicious entities	Communication delay increases
[21]	2022	Elsevier	An AI-based secure and efficient communication mechanism is proposed	Attack detection and Prevention	Not suitable for resource- constrained environments
[22]	2021	IEEE	AI empowered threat detection model is designed	DDoS attack detection	It demands high resource consumption
[23]	2022	IJCSIS	A reliable E-commerce management system is proposed	Botnet attack detection	Not suitable for large network
[24]	2021	Elsevier	Block chain-based security scheme is presented	Cyber-attack prevention	Computational overhead increases
[25]	2021	IEEE	A privacy-preserving mechanism is designed	Secure communication tunnel	High resource consumption
[26]	2021	IEEE	A block chain-based Secure framework is formulated	Malicious traffic analysis	Higher latencies experienced
[27]	2020	IEEE	Smart communication mechanism for E-Commerce	Malware detection	Computational overhead
[28]	2020	Elsevier	An attack detection mechanism is presented	DoD,DDoS identification	Not compatible with large networks
[29]	2021	Elsevier	An efficient attack detection model is proposed	Protection against cyber threats	High resource consumption noticed

3. SIGNIFICANCE OF STUDY

The study explores customer perceptions of security and privacy in m-commerce applications using protection motivation theory. It evaluates perceived threat severity, vulnerability, and efficacy of response. Perceived threat vulnerability affects trustworthiness, with logos, product pictures, privacy seals, and search facilities affecting trustworthiness. Malware detection can be achieved through static or dynamic analysis, with static focusing on parsing malware binaries and dynamic monitoring software [9]. Researchers should focus on finding new malware and reducing the number of characteristics needed for Detection [10].

DNNLSTM Algorithms

Val_In: Input Variable {transact, onserviceConnected and every colomn};

Target: Criteria {ServiceConnection };

Step1 : For each variable Input data,

do feature statistics

Step2 : For each instance step 2 target columns (ServiceConnection)

Step 3: Test and Score for algorithms

Find confusion metrics

Step 4: Find Rank of instances

Step 5: Predict appropriate Class using scatterplot

Class=1 Positive=Normal State

Else if (class =0)

Negative= Variable State end for

4. METHODOLOGY

Recent developments in malware detection have resulted in major improvements to the workflow, which in turn have increased both the process's efficiency and its efficacy[11]. In the research article, the workflow of standard machine learning is broken down into its individual processes and components. The challenges and limitations that such a workflow presents receive particular attention. The research also underlines the significance of deep learning approaches in this area, demonstrating the potential role that machine learning might play in the identification of malware[12]. This research article illustrates the different processes and components of a typical machine learning workflow for the detection and classification of malware[13]. This workflow is described in more detail in other research papers. It then goes on to analyse the problems and limits of such a process, review the most current advances and trends in the area, with a particular focus on deep learning approaches, and finally close with some recommendations. The research strategy that has been suggested for this research study can be viewed below in Figure 1, which can be found on the next page. It provides an example of the workflow process from beginning to finish, giving a more thorough understanding of the recommended machine learning technique for the detection of malware. This is done in order to provide more information about the process.

5. DATA DESCRIPTION

DroidFusion is an innovative classifier fusion strategy that integrates machine learning algorithms with the purpose of improving the accuracy of Android malware detection[14]. This approach is built on a multilayer architecture .The framework creates a model by training base classifiers at a lower level. To derive a final classifier, the framework then performs ranking-based algorithms on the prediction accuracies of the base classifiers at a higher level. The use of this induced multilevel DroidFusion model as an enhanced accuracy predictor for Android malware detection is possible[15] .The outcomes of experiments carried out on four different datasets demonstrate the usefulness of the suggested methodology. The DroidFusion approach may also make it possible to combine many ensemble learning methods in order to achieve higher levels of precision. The accuracy of predictions made with DroidFusion has the potential to beat those made using layered generalization, a well-known classifier fusion technique that makes use of a meta-classifier approach at its highest level. This research made use of a dataset that had feature vectors consisting of 215 different properties that were retrieved from 15,036 programs. These applications included 5,560 malicious apps derived from the Drebin project as well as 9,476 benign apps.

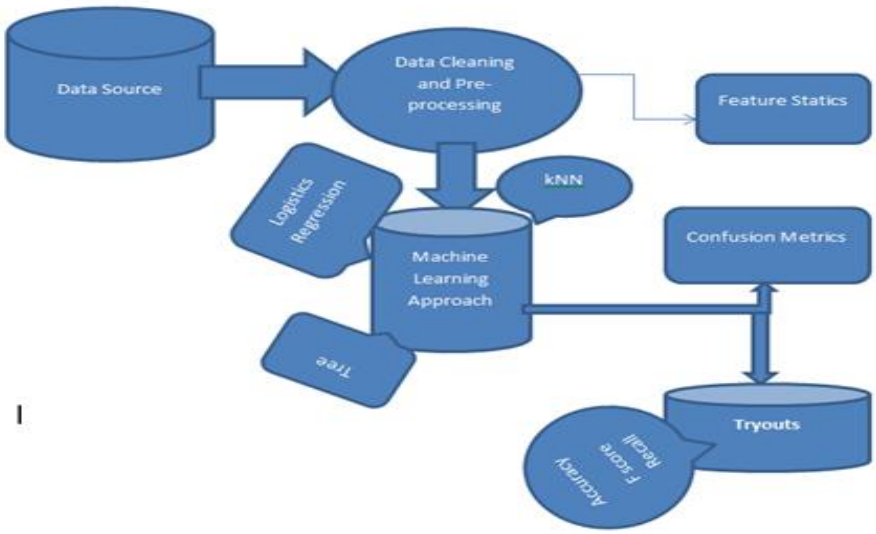

Figure 1: Prediction Flow Chart

6. DATA PRE-PROCESSING

Our study objective was accomplished via the use of data pre-processing, which included filling in missing variables in the diabetes dataset. For instance, the use of the features' nominal values is not acceptable for either machine learning or deep learning algorithms when it comes to diabetes prediction. We change the nominal attribute values into their corresponding numeric values, such as 1 for male and 0 for female in the sex group; 1 for yes and 0 for no in the other attribute group; and 1 for positive value in the class group and a negative value in the remaining attribute group, where 1 indicates a positive value and 0 indicates a negative value. In the field of machine learning, a method known as "data pre-processing" refers to the process of transforming raw data into a format that is either logical or understandable. The data that comes from the actual world is almost always missing values, inconsistent, untrustworthy, and redundant, amongst other issues. The elimination of such issues, which are often referred to as noise, is a common goal of the process known as data preparation. The term "pre-processing" refers to a series of operations that include cleaning the data, integrating the data, transforming the data, reducing the data, discretizing the data, and cleaning the data. In this step, the dataset is inspected to look for issues such as type mismatches, missing entries, and duplicate values. During the stage known as the data pre-processing step, each and every one of these inconsistencies is removed from this dataset. It is essential to clean the dataset before training it on a classifier in order to improve the classifier's ability to learn the hidden patterns included within the dataset. The relevant feature vectors that are input into the classifier enable it to acquire knowledge more precisely and in a short amount of time.

Feature Statistics

One of the eight characteristics of the data set that refers to age is used as part of the process of categorizing the personal information, and this is done as one of the steps. The following seven traits are considered critical due to the fact that they include important pieces of information: It is necessary to have all harmful data at one's disposal while attempting to diagnose malicious code and determine the level of the problem. The process of picking features includes both the steps of understanding the datasets and selecting the characteristics that would provide the essential data required to infer the information that has been sought. Feature selection is another term for this component of data analysis (Figure 2), and it refers to the process of finding a subset of data from a massive dimension of data. This process is known as feature selection.

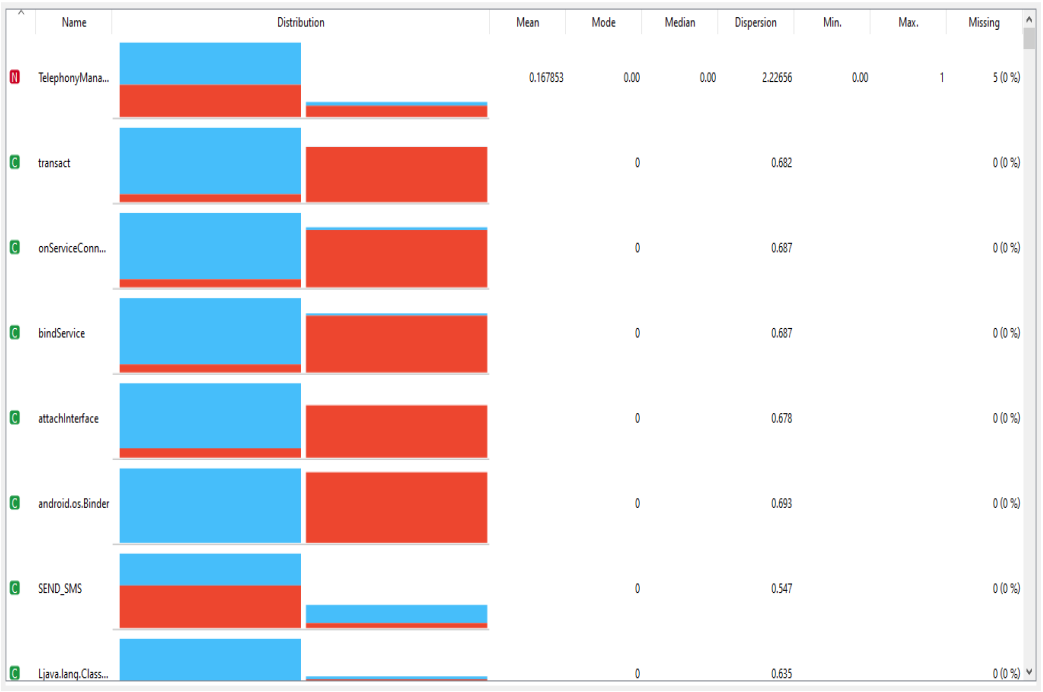

Figure 2: Feature Statistics

Figure 3: Target Column

Select Columns allows users to create data domains, distinguishing between ordinary attributes, optional class attributes, and meta-attributes. Models consist of a set of attributes and a discrete class attribute, while Meta attributes serve as instance labels.

Comparative Evaluations

The confusion matrix for the model shows that the vast majority of incorrect classifications originated from the class that was predicted, with accuracy values of 0.999. The predicted class had the lowest MSE and RMSE, which was due to the fact that its R2 value was the lowest. It was difficult to identify the anticipated class since there was overlap between the normal class and the diabetic class. The model had 215 instances in which it correctly predicted both classes, however it had the largest number of false positives from the class that was predicted. The accuracy of the classifier was negatively impacted as a result of this substantial class overlap.

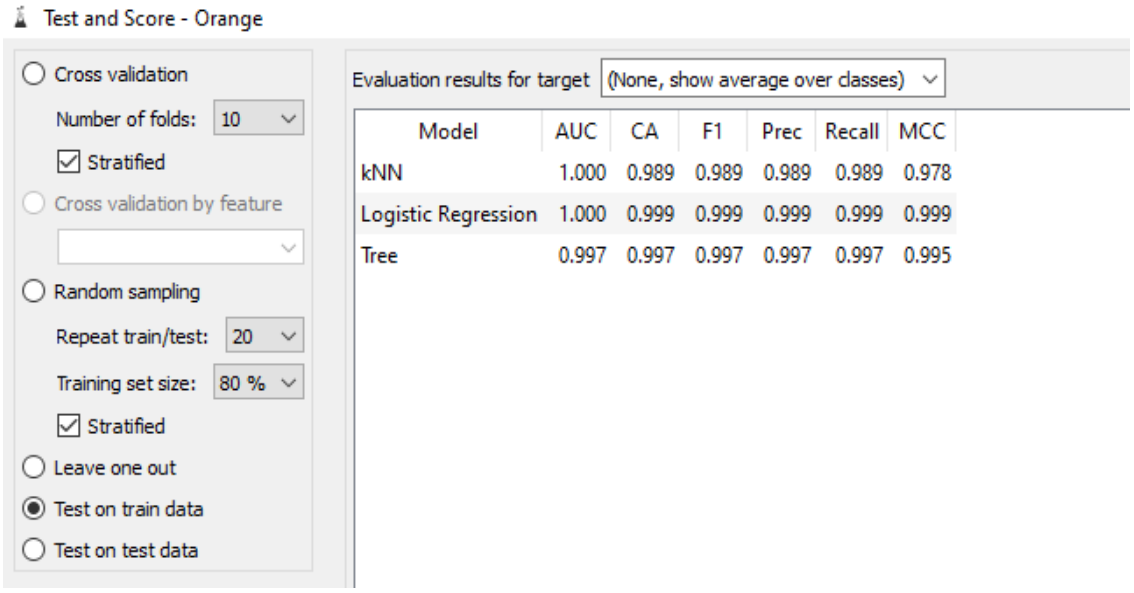

Fig 3: Evaluation Result on Train data

The models' cross-validation performance (Figure 3 and 4) was verified through 10 experiments, using the average and standard deviation of their accuracy, precision, recall, and F1-score values as evaluation metrics.

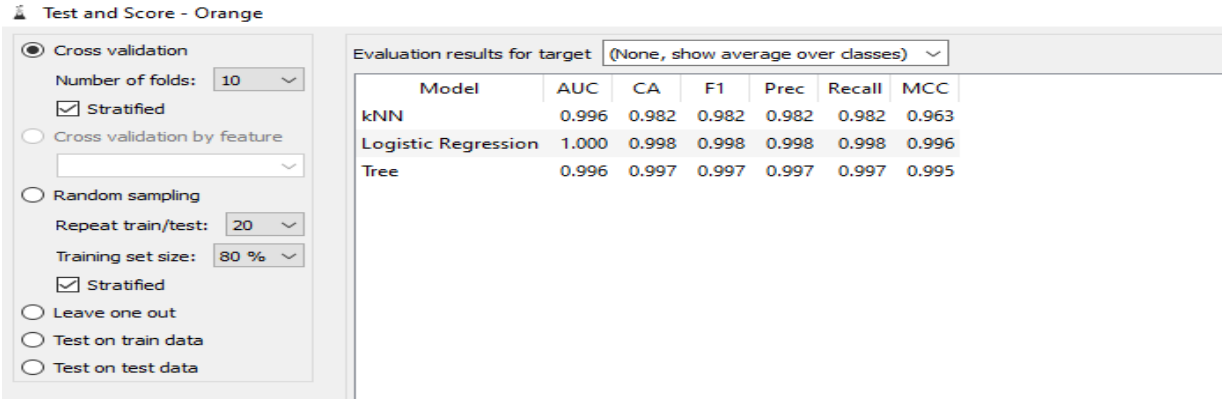

Fig 4: Evaluation Result on Cross fold

7. RESULT AND DISCUSSION

The training and testing stages of the categorization procedure were the two most important stages. In order to train a system, it was given files that were both safe and dangerous. The use of a learning algorithm allowed for the training of automated classifiers. Every classifier (KNN, LR, and DT) improved its accuracy with every new piece of data that it annotated and represented in the sequence shown below. During the phase of testing, a classifier was given a variety of new files, some of which were malicious and some of which were not. The classifier was tasked with determining whether or not the files included malicious code.

Table 2: Result of classification

Model Used	Accuracy	Precision	Recall	F1-Score
DECISION TREE	90	95	86	90
LOGISTIC REGRESSION	94	94	96	95
KNN	99	99	98	99
SVM	99	99	99	99

Fig 5: classification result

The malware and clean ware that we acquired served as the subjects of an experiment in which our DNNLSTM approach for the classification and detection of malware was examined[16]. When investigating malware and attempting to characterize it, we made use of supervised machine learning methods or classifiers (kNN, LR, and DT). The findings of the classifiers' accuracy (KNN = 99%, LR = Near 94 %, and DT = 90 %) led us to the conclusion that the LR model was the

most effective one for the malware detection method[4]. This was determined by doing statistical analysis on the data shown in Table 2. It is evident that when identifying malware using the three most ideal algorithms (KNN = 99%, LR = Near 94 %, and DT = 90 %) which had a much higher F1 (99.8%) rate and accuracy, LR accuracy is the greatest and LR is a superior option for malware detection.

8. CONCLUSION

The purpose of this study is to examine the use of machine learning (ML) methods to the identification of malware[17]. After analyzing the performance of three machine learning algorithms, researchers found that LR achieved 99% accuracy. Static analysis was used in this research project to evaluate the sensitivity of a machine learning classifier's detection capabilities. The solution based on machine learning achieved the highest level of accuracy (99%). In trial findings, static analysis, which is based on facts and data that have been carefully picked, showed promise, offering the highest detection accuracy and correctly classifying malware. Using the newly collected dataset, the three machine learning models (kNN, LR, and DT) were trained and then compared against one another.

REFERENCES

[1] Akhtar, M.S.; Feng, T. Detection of sleep paralysis by using IoT based device and its relationship between sleep paralysis and sleep quality. EAI Endorsed Trans. Internet Things 2022, 8, e4.

[2] Wang, J., Deng, P., Fan, Y., Jaw, L., Liu, Y.: Virus detection using data mining techniques. In: Proceedings of IEEE International Conference on Data Mining (2003).

[3] Chen, X., Andersen, J., Mao, Z., Bailey, M., Nazario, J.: Towards an understanding of anti-virtualization and anti-debugging behavior in modern malware. In: DSN (2008).

[4] P. Laskov, C. Gehl, S. Krüger et al., "Incremental support vector learning: analysis, implementation and applications," Journal of Machine Learning Research, vol. 7, no. 3, p. 2006, 2006.

[5] M. Alazab, "Profiling and classifying the behavior of malicious codes," Journal of Systems & Software, vol. 100, pp. 91–102, 2014.

[6] G. Liang, J. Pang, Z. Shan, R. Yang, and Y. Chen, "Automatic benchmark generation framework for malware detection," Security and Communication Networks, vol. 2018, Article ID 4947695, 8 pages, 2018.

[7] D. Xue, J. Li, T. Lv, W. Wu, and J. Wang, "Malware classification using probability scoring and machine learning," IEEE Access, vol. 7, pp. 91641–91656, 2019.

[8] Zhao, K.; Zhang, D.; Su, X.; Li, W. Fest: A feature extraction and selection tool for android malware detection. In Proceedings of the 2015 IEEE Symposium on Computers and Communication (ISCC), Larnaca, Cyprus, 6–9 July 2015; pp. 714–720.

[9] D. Xue, J. Li, W. Wu et al., "Homology analysis of malware based on ensemble learning and multifeatures," PLoS One, vol. 14, no. 8, Article ID e0211373, 2019.

[10] M. Ristin, M. Guillaumin, J. Gall, and L. Van Gool, "Incremental learning of random forests for large-scale image classification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 3, pp. 490–503, 2016.

[11] Nikam, U.V.; Deshmuh, V.M. Performance evaluation of machine learning classifiers in malware detection. In Proceedings of the 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 23–24 April 2022; pp. 1–5.

[12] Sharma, S.; Krishna, C.R.; Sahay, S.K. Detection of advanced by machine learning techniques. In Proceedings of the SoCTA 2017, Jhansi, India, 22–24 December 2017.

[13] Chandrakala, D.; Sait, A.; Kiruthika, J.; Nivetha, R. Detection and classification of malware. In Proceedings of the 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 8–9 October 2021; pp. 1–3.

[14] Sanz, B., Santos, I., Laorden, C., Ugarte-Pedrero, X., Bringas, P.G., Alvarez, G.: PUMA: permission usage to detect Malware in Android. In: Advances in Intelligent Systems and Computing (AISC) (2012)

[15] Akhtar, M.S.; Feng, T. IOTA based anomaly detection machine learning in mobile sensing. EAI Endorsed Trans. Create. Tech. 2022, 9, 172814.

[16] Sethi, K.; Kumar, R.; Sethi, L.; Bera, P.; Patra, P.K. A novel machine learning based malware detection and classification framework. In Proceedings of the 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Oxford, UK, 3–4 June 2019; pp. 1–13.

[17] Sethi, K.; Kumar, R.; Sethi, L.; Bera, P.; Patra, P.K. A novel machine learning based malware detection and

classification framework. In Proceedings of the 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Oxford, UK, 3–4 June 2019; pp. 1–13.

[18] F. Behgounia, B. J. I. J. o. C. S. Zohuri, and I. Security, "Machine Learning Driven An E-Commerce," *International Journal of Computer Science and Information Security (IJCSIS),* vol. 18, no. 10, 2020.

[19] M. Zhang, L. Lin, and Z. J. C. C. Chen, "Lightweight security scheme for data management in E-commerce platform using dynamic data management using blockchain model," *Lightweight security scheme for data management in E-commerce platform using dynamic data management using blockchain model." Cluster Computing (2021): 1-15.*, pp. 1-15, 2021.

[20] N. L. Bhatia, V. K. Shukla, R. Punhani *et al.*, "Growing Aspects of Cyber Security in E-Commerce." pp. 1-6.

[21] M. Li, L. Zhu, Z. Zhang *et al.*, "Anonymous and verifiable reputation system for E-commerce platforms based on blockchain," *IEEE Transactions on Network and Service Management,* vol. 18, no. 4, pp. 4434-4449, 2021.

[22] M. J. Girsang, R. Hendayani, and Y. Ganesan, "Can Information Security, Privacy and Satisfaction Influence The E-Commerce Consumer Trust?." pp. 1-7.

[23] Z. Zhu, Y. Bai, W. Dai *et al.*, "Quality of e-commerce agricultural products and the safety of the ecological environment of the origin based on 5G Internet of Things technology," *Environmental Technology & Innovation, 22, 101462.*, vol. 22, pp. 101462, 2021.

[24] L. T. T. J. J. o. R. Tran, and C. Services, "Managing the effectiveness of e-commerce platforms in a pandemic," *Journal of Retailing and Consumer Services 58 (2021): 102287,* vol. 58, pp. 102287, 2021.

[25] A. R. Khan, M. Kashif, R. H. Jhaveri *et al.*, "Deep Learning for Intrusion Detection and Security of Internet of Things (IoT): Current Analysis, Challenges, and Possible Solutions," vol. 2022, 2022.

[26] Y. Otoum, D. Liu, and A. J. T. o. E. T. T. Nayak, "DL-IDS: a deep learning–based intrusion detection framework for securing IoT," vol. 33, no. 3, pp. e3803, 2022.

[27] L. Sama, H. Wang, and P. Watters, "Enhancing System Security by Intrusion Detection Using Deep Learning." pp. 169-176.

[28] M. Imran, N. Haider, M. Shoaib *et al.*, "An intelligent and efficient network intrusion detection system using deep learning," vol. 99, pp. 107764, 2022.

[29] I. Ullah, and Q. H. J. I. A. Mahmoud, "Design and development of a deep learning-based model for anomaly detection in IoT networks," vol. 9, pp. 103906-103926, 2021.