International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Network Security through Machine Learning-Based Anomaly Detection Systems

Salam Allawi Hussein[1] Sándor R. Répás[2]

Submitted: 03/02/2024 **Revised**: 11/03/2024 **Accepted**: 17/03/2024

Abstract: For decades, anomaly detection has been used to discover and extract aberrant components from data. Several techniques have been employed to spot irregularities. Machine learning (ML) is a method that is gaining importance due to its significant significance in this area. Machine learning models that detect anomalies in their application are the focus of this study's Systematic Literature Review (SLR). In our investigation, we look at the models from four angles: how anomaly detection is classified, what it's used for, how machine learning is done, and how well machine learning models perform. In this study, we looked for papers published in 2015–2023, which deal with the topic of anomaly detection using machine learning techniques. After we've finished analyzing the selected research papers, we'll go on to outline 10 different uses of anomaly detection that were found in those publications. The number of machine learning models used to detect anomalies is also identified, accounting for 6% of all instances. Finally, we offer available a wide range of datasets used in anomaly detection studies as well as many other generic datasets. Furthermore, compared to other categorized anomaly detection methods, researchers are more likely to employ unsupervised anomaly detection. The application of machine learning models for anomaly detection is one of the most promising fields of study, and researchers have utilized several ML models in this regard. Therefore, based on the results of this review, we advise and suggest things to researchers.

Keywords: *Machine Learning, Anomaly Detection, Network Security, Data privacy and protection.*

Introduction

Research into the identification of anomalies has persisted for millennia since it is a substantial challenge. A large variety of methods have been developed and put into use for the objective of anomaly detection. What is meant by "anomaly detection" is "the problem of finding patterns in data that do not conform to expected behavior" (Naseer et al., 2018). Finding out what's out of the ordinary is something that many people do and has several uses. The detection of fraudulent actions, the approval of loans, and the tracking of health problems are all instances of this (Mulinka & Casas, 2018). Among the many potential medical applications, heart rate monitors stand out. Cyber intrusion detection, streaming, hyperspectral imaging, and defect identification for aviation safety studies are just a few of the many more applications that utilize anomaly detection (Elmrabit, Zhou, Li & Zhou F, 2020). Recognized by Hosaain & Islam (2023) as the associate editor responsible for managing the assessment of this article and granting approval for publication, unpro poses a risk that necessitates the identification of abnormalities in various application areas. The data that has been found may include a wealth of important and useful information. As an example, discovering a suspicious trend in network traffic can indicate a hacker-initiated assault (Fourure et al., 2021). Another example may be seeing suspicious patterns in a credit card's transaction history that could point to fraud (Pang et al., 2019). A further risk is that if an airplane's sensors detect something out of the ordinary, it might mean that one or more of the plane's parts are flawed. An anomaly, at its most basic, is any pattern whose behavior deviates from the generally accepted norm. According to Cauteruccio et al. (2021), there are mainly three ways to categorize anomalies. An anomaly can take several forms, but the simplest of them is the point anomaly. An outlier is considered a point anomaly if it stands out from the rest of the data in some way. A data instance is considered contextually anomalous if it is unusual in one setting but normal in another. A contextual anomaly describes this kind of discrepancy. Contextual anomalies are characterized by two things. These traits are environmental and behavioral. An instance's context (or neighborhood) may be determined by using the first characteristic.

Geographical datasets often include contextual details like the longitude and latitude of specific places. Time is a contextual attribute that determines the placement of an occurrence throughout the complete sequence in the context of time series data. The instance's noncontextual traits are defined by the second property, which is part of the behavior attribute category. In a geographical dataset that characterizes the average rainfall across the world, the

[1,2] *Department of Telecommunications Széchenyi István University Györ, 9026, Hungary*
[1,2]*E-mail: salam.allawi@sze.hu repas.sandor@sze.hu*
[1]*Department of Computer Science, College of Computer Science & Information Technology, University of Al-Qadisiyah Al Diwaniyah, 58001, Iraq*
[1]*E-mail: salam.allawi@qu.edu.iq*

amount of rainfall that occurs at any particular site is an example of a behavioral attribute. To determine whether to use the contextual anomaly detection approach, the relevance of the anomalies in the target region is taken into account. Another crucial consideration is the accessibility of qualitative attributes. In cases when identifying the context is straightforward, it makes sense to put a system in place to do just that. Conversely, there are cases where it may be unwise to suggest that specific methods are difficult to put into practice. A collective anomaly happens when a cluster of related data instances collectively display an out-of-the-ordinary behavior throughout the whole dataset.

Hossain and Islam (2023) say that statistical anomaly detection was one of the first ways to find events that were not predicted. Statistical methods are used to create a model which describes how the data usually behaves. A statistical reasoning test needs to be done to see if the problem frequency fits with the model. According to Naseer et al. (2018), there are a lot of different ways to look for data outliers. Some of these methods are proximity-based strategies, parametric, nonparametric, and also semi-parametric techniques. Machine learning techniques are being used more and more to find outliers. Eltanbouly et al. (2020) say that the main goal of machine learning is to automate the process of learning from examples. We can make a model that can tell the difference between normal and abnormal groups by using this method. We carefully read through and chose the research papers after carefully looking at the important prediction research in anomaly detection as well as the pros and cons of the machine learning method. The article is broken up into the following parts: The second part will tell you more about the task that goes with it. In Section 3 we look at how the study was completed. The findings and commentary are laid up in Section 4. The limitations of this evaluation are addressed in Section 5. Finally, the study's conclusion should be stated.

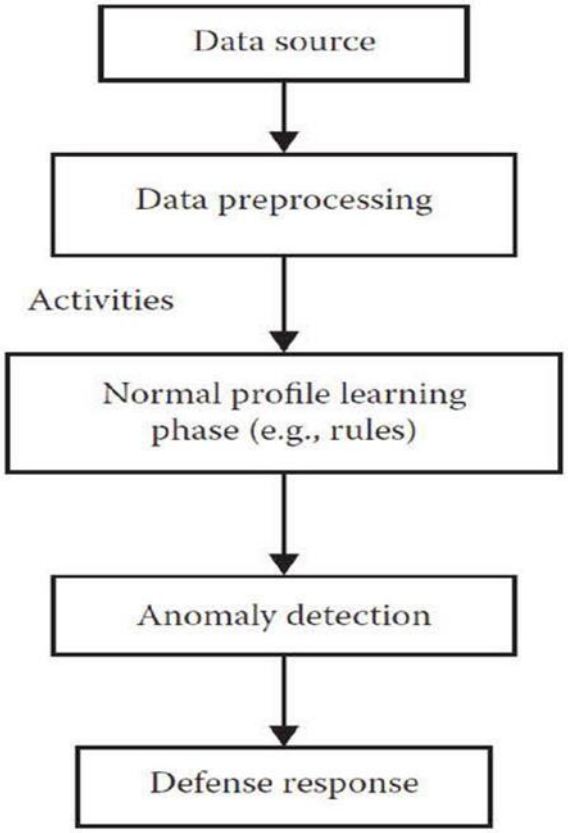

Figure 1: Sequence of execution of modules in an anomaly detection system

Sorce: Megantara, A.A., & Ahmad, T. (2021). A hybrid machine learning method for increasing the performance of network intrusion detection systems. Journal of Big Data, 8.

Research objectives

Performing a systematic review that is representative of an all-encompassing investigation of machine learning approaches for anomaly detection and the applications of these techniques is the major purpose of this work. In addition, this evaluation investigates the degree of accuracy of the machine learning models as well as the proportion of research publications that make use of supervised anomaly detection categorization.

Study Rationale

We are certain that this study will provide researchers with the opportunity to get a deeper comprehension of the various approaches to anomaly detection and will also serve as a guide for them to examine the most recent research that has been conducted on this topic. This study was inspired by the fact that, to the best of our knowledge, there are relatively few Systematic Literature Reviews (SLR) on the subject of identifying anomalies through the use of machine learning algorithms.

Literature Review

Anomaly detection is a major topic that has been investigated in many fields of research and practical applications. Several anomaly detection methods exist, some with narrower applications in mind and others with broader scopes. An example of this is the all-encompassing evaluation of anomaly detection technologies and applications that Hosseinzadeh et al., (2021) provided. A comprehensive evaluation of non-ML techniques, including statistical and spectral detection methods, and ML approaches was discussed in detail during the board meeting. Beyond that, the survey offers a multitude of other uses for anomaly identification. You may find identification systems for a variety of uses, including cyber intrusion detection, fraud detection, medical anomaly detection, industrial damage detection, image processing, textual anomaly detection, and sensor networks. Imran, Jamil, and Kim (2021) offered a different research that dealt with the detection of outliers in discrete sequences. The authors provided a comprehensive and well-organized review of previous research on the topic of anomaly detection in discrete or symbolic sequences. Besides research that thoroughly examined statistical anomaly detection and machine learning methods (Saba et al., 2022). Furthermore, the writers compared and contrasted each technique, highlighting its advantages and disadvantages. The data mining survey that Al Turaiki and El Tawijry (2021) suggested is worth considering, nevertheless. It was planned to carry out this survey. Finding outliers in certain fields and uses was the major focus of a few evaluations. An article that exemplifies this is Fourure et al. (2021), which provides a comprehensive analysis of widespread clustering-based fraud detection methods and compares and contrasts them from several perspectives. Anomaly detection in automated surveillance is another area of interest; they provided several models and classification methods for this purpose.

The writers carefully examined the research articles, considering the technique, methodology, and topic domain. Furthermore, Saba et al., (2022) provided a synopsis of the three most popular anomaly identification algorithms used nowadays in geochemical data processing. Some examples of these methods are ML, compositional data analysis, and fractal/multi-fractal models. The author's primary focus, meanwhile, is on methods related to machine learning. An extra interesting note is that research proposed a synopsis of computer system performance, anomaly detection, and bottleneck identification. The writers classified the several existing solutions after identifying the fundamental features of the problem (Al Souci et al., 2021). Several research mainly aimed at identifying unusual invasions. Meanwhile, a study reviewed intrusion detection methods; however, the focus was on machine learning techniques (Ullah & Mahmood, 2021). To overcome the textual challenges of intrusion detection, they summarised the machine learning techniques that were created to do so. The authors also looked at similar research that differed in terms of datasets, classifier architectures, and other features. dini et al. (2023) conducted a thorough evaluation of intrusion detection and anomaly detection methodologies, as well as machine learning and data mining techniques for cyber intrusion detection.

In the course of the same overall investigation, researchers performed both of these investigations. In addition to outlining each method in great depth, the writers addressed the challenges of applying data mining and machine learning to the domain of cyber security. To sum up, increasing the efficacy of detecting anomalies in network intrusion systems requires a strategy that integrates many machine learning methods with particle swarm optimization. Finding outliers in networks has been the subject of a great deal of research. This led to several surveys focusing on the matter. One example is the extensive study of anomaly detection in networks carried out by Bahardiya (2023). After they figured out what kinds of attacks intrusion detection systems often encounter, they defined anomaly detection and tested its effectiveness. The writers also covered the methods and resources used by network defenders. Similarly, Rabel & Hussain (2023) examined popular methods for detecting anomalies in networks in a thorough study. Both supervised and unsupervised learning methods, as well as density- and distance-based approaches, were part of these methods. Among these methods were machine learning approaches well-suited to detecting anomalies in networks, as well as deep recurrent neural networks, restricted Boltzmann machine-based deep belief networks and others. Beyond this, the authors presented experiments that demonstrated the practicality of deep learning methods for analyzing network data. Our systematic review stands out from the rest since it presents a thorough investigation of anomaly detection using machine learning methods. Research Deficit The paper may benefit from a more thorough examination of the pros and cons of various machine learning approaches, such as supervised, semi-supervised, and unsupervised anomaly detection models.

Research Methodology

Qualitative research methodology is being incorporated in the study along with the systematic search of the literature on enhancing network security through machine learning-based anomaly detection systems was conducted by reviewing relevant literature from authentic journals. The search procedure for this systematic review adhered to meta-analysis. The topic phrases "network security," "machine learning," "anomaly detection" and "network privacy" were all combined in the electronic search method for each database. For instance, the search approach started with the important topic phrases "enhancing network security" and "machine learning." Then, Boolean keyword combinations were searched (without any restrictions) for the terms "network security*," "machine learning*," and "anomaly detection system*. After merging the search results from many databases, duplicates were removed and the remaining entries were examined.

Selection criteria

The initial stage of the exclusion procedure was applying the selection criteria to the abstracts. Papers were read in their entirety if abstracts were missing or provided insufficient information. After this first screening of abstracts, the remaining articles were carefully scrutinized using the subsequent selection criteria to determine the final sample of records that met all the necessary criteria:

- Works published between the years 2015 and 2023. A preliminary database search that produced a larger amount of publications on the topic "network security through machine learning-2015" led to the selection of this time frame.
- The papers have to be translated from their native language or published in English.
- Research that detailed the use of network security in a methodical manner within the framework of a machine learning-based anomaly detection system was mandated.

Results and Discussion

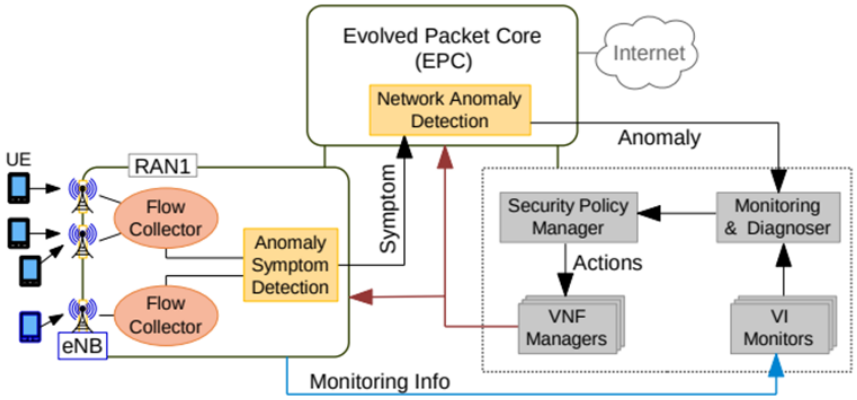

Figure 2: Network anamoluly detection system (representing control action)

Source: said et al, 2020

In this part, we'll talk about the outcomes of this review. A summary of the papers that were selected for analysis in this review may be found in this paragraph. A thorough explanation of the study's conclusions is given. Relevant studies or articles that used machine learning were chosen to facilitate anomaly detection. These academic articles were released in tandem between 2015 and 2023. A list of these published works may be found in Table 1. As previously mentioned, the process of prioritizing the articles based on their validity and dependability involves using a quality evaluation criterion.

Table 1: Related studies summary

Year	Summary	Work difference
2015	In this survey, the authors addressed the causes and aspects of network anomalies. They add performance metrics and intrusion detection systems evaluation and provide a list of tools and research issues.	Our work differs in that it is more general, and includes an estimation of the accuracy of each ML model as well as the type of anomaly detection used.

2016	In this survey, the authors present machine learning methods in network intrusion detection systems with particle swarm optimization for anomaly. It covers network anomaly detection in particular.	The researcher has covered machine learning and particle swarm detection. The paper provides optimization techniques for intrusion detection system types and presents each technique's advantages and disadvantages.
2018	In this survey, the authors provide a comprehensive analysis of performance anomaly detection and identification of bottlenecks.	In contrast, the particular study has provided machine learning systems, identifying various types of common anomalies and the techniques and strategies for detecting them.
2020	In this survey, the authors review various clustering-based anomaly detection techniques and they provide a comparison between the techniques.	Our work has presented the same but has used a machine learning model that helps sort and detect the over all learning.
2021	Various studies have covered machine learning scope and trends.	Our work articulates recommendations and guidelines using relevant yet recent database techniques.

The research was conducted in the area of prediction studies related to anomaly detection. The methods for identifying anomalies may be broadly divided into two groups: machine learning-based methods and non-machine learning-based methods. The methods that aren't machine learning-based may be split into two groups: knowledge-based and statistical. In particular, articles that investigate the use of machine learning algorithms for abnormality identification are included in this review. On the other hand, about 10 publications focus on approaches independent of machine learning technology.

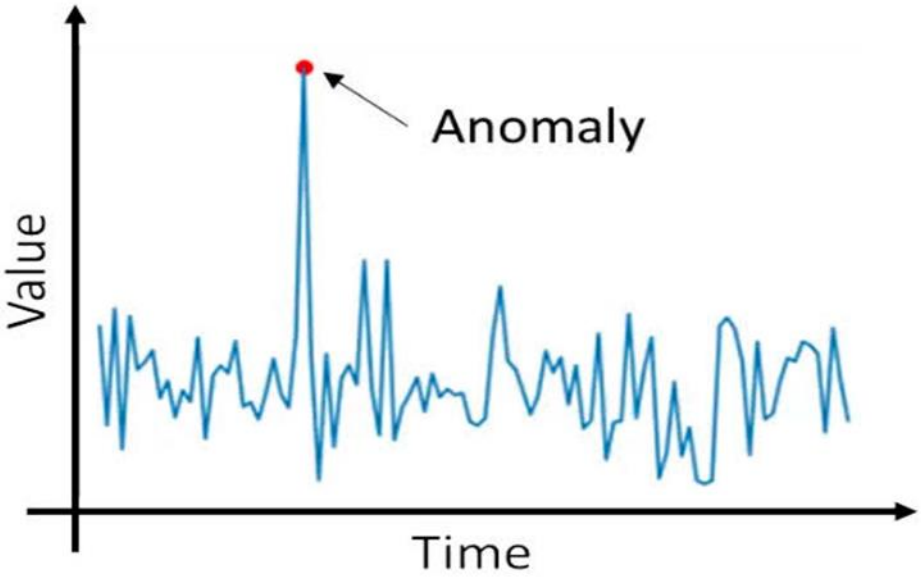

Figure 3: Anamoly Detection via machine learning

Source: Haji and Aameen, 2021

Applications for anomaly detection may be found in a variety of settings. In the course of our investigation, we found 10 different applications in the selected articles. A list of the publications that were used in the studies appears to be included in Figure 1. Furthermore, the study offers a wealth of information on the frequency at which the selected articles employ the anomaly detection algorithm. Furthermore, the assessment shows that, from 2015 to 2020, researchers began to use anomaly detection in a greater percentage of these applications.

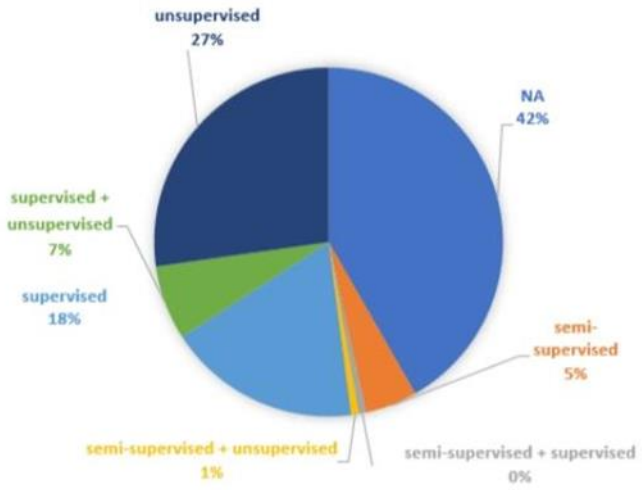

Figure 4: Summary of articles included in the study as per their scope

Conclusion

This systematic literature study set out to explore the use of machine learning (ML) techniques for anomaly identification. The application of anomaly detection type, ML method type, ML model accuracy assessment, and ML type of anomaly detection (supervised, semi-supervised, and unsupervised) were the four perspectives from which it analyzed machine learning models. To assess the relevant publications published between 2015 and 2023, a review was carried out. The results show that intrusion detection, network anomaly detection, general anomaly detection, and data applications are the studies that are used in the anomaly detection area the most frequently. These are the anomaly detection uses that were discovered in the selected articles. In addition to a large number of additional generic datasets, the discovered research has made use of a variety of datasets that have been used in the experiments of connected articles. Real-world datasets were used in most studies as training or testing datasets for their models. The last characteristic is the classification kind of anomaly detection that was used in a few research papers. About the research publications, we found that the form of unsupervised anomaly detection that was used was employed in 6% of the selected papers, making it the most often used method. With 8% of the articles using it, supervised anomaly detection was the second most popular technique. Three percent of the papers used both supervised and unsupervised anomaly detection classification after this.

Recommendations

We recommend that more research be done on machine learning studies of anomaly detection based on the findings of this study to obtain more information about the effectiveness and performance of machine learning models. Furthermore, it is highly recommended that researchers create a generic framework to start doing experiments using machine learning models. Furthermore, this area is crucial for development since we found research articles that did not specify the type of feature extraction or selection. Furthermore, several research studies only used one performance indicator—accuracy, for example—when publishing their conclusions, which calls for further research and improvement. We also learned that other researchers were still working with old datasets in their study. Using more latest datasets is highly suggested for researchers.

Limitations of the study

The only publications considered in this systematic literature review are those from journals and conferences that deal with anomaly detection and machine learning. Several research papers that were not relevant to the study were eliminated from consideration by using our search approach early in the review process. This ensured that the chosen research articles matched the demands of the investigation. However, we believe that by drawing from a larger variety of possible sources, this judgment may have been enhanced much further. Since we employed a strict quality assurance rating, the same principle holds for quality assessment.

References

[1] Alsoufi, M. A., Razak, S., Siraj, M. M., Nafea, I., Ghaleb, F. A., Saeed, F., & Nasser, M. (2021). Anomaly-based intrusion detection systems in IoT using deep learning: A systematic literature review. *Applied sciences*, *11*(18), 8383.

[2] Al-Turaiki, I., & Altwaijry, N. (2021). A convolutional neural network for improved anomaly-based network intrusion detection. *Big Data*, *9*(3), 233-252.

[3] Bharadiya, J. (2023). Machine learning in cybersecurity: Techniques and challenges. *European Journal of Technology*, *7*(2), 1-14.

[4] Elmrabit, N., Zhou, F., Li, F., & Zhou, H. (2020, June). Evaluation of machine learning algorithms for anomaly detection. In *2020 international conference on cyber security and protection of digital services (cyber security)* (pp. 1-8). IEEE.

[5] Eltanbouly, S., Bashendy, M., AlNaimi, N., Chkirbene, Z., & Erbad, A. (2020, February). Machine learning techniques for network anomaly detection: A survey. In *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)* (pp. 156-162). IEEE.

[6] Fourure, D., Javaid, M. U., Posocco, N., & Tihon, S. (2021, September). Anomaly detection: how to artificially increase your f1-score with a biased evaluation protocol. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 3-18). Cham: Springer International Publishing.

[7] Haji, S. H., & Ameen, S. Y. (2021). Attack and anomaly detection in IoT networks using machine learning techniques: A review. Asian J. Res. Comput. Sci, 9(2), 30-46.

[8] Hossain, M. A., & Islam, M. S. (2023). Ensuring network security with a robust intrusion detection system using ensemble-based machine learning. *Array*, *19*, 100306.

[9] Hosseinzadeh, M., Rahmani, A. M., Vo, B., Bidaki, M., Masdari, M., & Zangakani, M. (2021). Improving security using SVM-based anomaly detection: issues and challenges. *Soft Computing*, *25*(4), 3195-3223.

[10] Imran, Jamil, F., & Kim, D. (2021). An ensemble of prediction and learning mechanisms for improving the accuracy of anomaly detection in network intrusion environments. *Sustainability*, *13*(18), 10057.

[11] Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z., … & Akoglu, L. (2021). A comprehensive survey on graph anomaly detection with deep learning. IEEE Transactions on Knowledge and Data Engineering, 35(12), 12012-12038.

[12] Mulinka, P., & Casas, P. (2018, August). Stream-based machine learning for network security and anomaly detection. In *Proceedings of the 2018 workshop on big data analytics and machine learning for data communication networks* (pp. 1-7).

[13] Naseer, S., Saleem, Y., Khalid, S., Bashir, M. K., Han, J., Iqbal, M. M., & Han, K. (2018). Enhanced network anomaly detection based on deep neural networks. *IEEE Access*, *6*, 48231-48246.

[14] Pang, G., Shen, C., & Van Den Hengel, A. (2019, July). Deep anomaly detection with deviation networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 353-362).

[15] Peterson, K. T., Sagan, V., & Sloan, J. J. (2020). Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. GIScience & Remote Sensing, 57(4), 510-525.

[16] Poornima, I. G. A., & Paramasivan, B. (2020). Anomaly detection in wireless sensor network using a machine learning algorithm. Computer communications, 151, 331-337.

[17] Rafi, H., Rafiq, H., & Farhan, M. (2021). Inhibition of NMDA receptors by agmatine is followed by GABA/glutamate balance in benzodiazepine withdrawal syndrome. Beni-Suef University Journal of Basic and Applied Sciences, 10(1), 1-13.

[18] Rafi, H., Ahmad, F., Anis, J., Khan, R., Rafiq, H., & Farhan, M. (2020). Comparative effectiveness of agmatine and choline treatment in rats with cognitive impairment induced by AlCl3 and forced swim stress. Current Clinical Pharmacology, 15(3), 251-264.

[19] Rebel, J., & Hussain, S. Machine Learning Approaches for Anomaly Detection in Network Security.

[20] Saba, T., Rehman, A., Sadad, T., Kolivand, H., & Bahaj, S. A. (2022). Anomaly-based intrusion detection system for IoT networks through deep learning model. *Computers and Electrical Engineering*, *99*, 107810.

[21] Said Elsayed, M., Le-Khac, N. A., Dev, S., & Jurcut, A. D. (2020, November). Network anomaly detection using LSTM-based autoencoder. In Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks (pp. 37-45).

[22] Ullah, I., & Mahmoud, Q. H. (2021). Design and development of a deep learning-based model for anomaly detection in IoT networks. *IEEE Access*, *9*, 103906-103926.