INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

# Network Security through Machine Learning-Based Anomaly Detection Systems

TIRUMURU VALLIDEVI, Research Scholar, Department of Computer Science, J.S University, Shikohabad, U.P.

Dr. VIJAYA BHASKAR K, Associate Professor, Supervisor, Department of Computer Science and Engineering, J.S University, Shikohabad, U.P.

## ABSTRACT

Finding and extracting out-of-the-ordinary elements from data has been the domain of anomaly detection for many years. In order to detect anomalies, many methods have been used. Because of its enormous relevance in this field, machine learning (ML) is a technique that is rapidly rising in popularity. The Systematic Literature Review (SLR) in this work focuses on machine learning models that may identify irregularities in their implementation. We examine the models from four perspectives in our study: the classification of anomaly detection, its applications, the process of machine learning, and the performance of machine learning models. Anomaly detection using machine learning algorithms was the subject of publications published between 2015 and 2023. Following our conclusion with the analysis of the chosen research articles, we will next list ten distinct applications of anomaly detection as discussed in those works. Additionally, 6% of all incidents use machine learning models that are employed to identify abnormalities. Lastly, we provide a large variety of datasets, including those specifically designed for anomaly detection research and a plethora of other general datasets. Unsupervised anomaly detection is also more often used by researchers than other classified approaches. Anomaly detection with ML models is a very promising area of research, and academics have already used many ML models in this area. Consequently, we provide researchers with recommendations and suggestions based on this review's findings.

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

ISSN: 2057-5688

## I. INTRODUCTION

Anomaly detection is a difficult problem, which is why researchers have been working on it for centuries. The goal of anomaly detection has inspired the development and implementation of a wide range of approaches. In this context, "anomaly detection" refers to "the problem of finding patterns in data that do not conform to expected behavior" (Naseer et al., 2018). Many individuals do this for many reasons, but one of them is to find out what's unusual. This is used, for example, to identify fraudulent activities, to approve loans, and to monitor health issues (Mulinka & Casas, 2018). Among the many possible medical uses, heart rate monitors are particularly noteworthy. Elmrabit, Zhou, Li & Zhou F. (2020) lists a plethora of other uses for anomaly detection, including cyber intrusion detection, streaming, hyperspectral imaging, and defect discovery for aviation safety research. As the associate editor overseeing the article's evaluation and final publication clearance, unpro presents a risk that calls for the detection of anomalies in a number of domains of application (Hosaain & Islam, 2023). It is possible that the discovered data contains a plethora of valuable and significant information.

For instance, according to Fourure et al. (2021), if one notices an unusual pattern in the network's traffic, it might be a sign of an attack launched by hackers. For another instance, it might be indicative of fraud to notice unusual trends in a customer's credit card spending (Pang et al., 2019). Another potential issue is that if an aircraft's sensors pick up on anything out of the norm, it might indicate a problem with one or more of its components. At its core, an anomaly is just a pattern whose behavior is out of the ordinary. Cauteruccio et al. (2021) states that anomalies may be broadly classified into three primary types. The most basic kind of anomaly is the point anomaly, however there are others. If a single data point significantly deviates from the norm, we say that data point is an outlier. If a data instance is out of the ordinary in one context but perfectly typical in another, we say that it is contextually anomalous. These kinds of differences are called contextual anomalies. A contextual anomaly will have two hallmarks. You may think of these characteristics as behavioral and environmental. The first attribute may be used to find the instance's neighborhood or context.

Information about particular

locations, such as their longitude and latitude, is a common component of geographic datasets. Within the framework of time series data, the contextual characteristic of time establishes the position of an event within the whole sequence. As a member of the behavior attribute category, the second property defines the instance's noncontextual features. For instance, the quantity of precipitation at every given location is an example of a behavioral characteristic in a global rainfall average dataset. Consideration is given to the significance of the abnormalities in the target area when deciding whether to use the contextual anomaly detection method. It is also important to think about how easy it is to get qualitative characteristics. It is prudent to implement a system to automatically determine the context when doing so is not complicated. It could be foolish, on the other hand, to imply that some approaches are hard to implement in certain situations. When many connected data instances act in an unusual way over the whole dataset, it is called a collective anomaly.

According to Hossain and Islam (2023), finding unexpected occurrences was a primary goal of statistical anomaly detection. A model is constructed using statistical approaches to depict the typical behavior of the data. The frequency of the problems should be compared to the model using a statistical reasoning test. Numerous methods exist for identifying data outliers, as stated by Naseer et al. (2018). Parametric, nonparametric, proximity-based, and semi-parametric approaches are among these methods. The usage of machine learning methods to identify anomalies is on the rise. According to Eltanbouly et al. (2020), automating the learning-from-examples process is the primary objective of machine learning. Using this strategy, we may create a model that can distinguish between groups that are considered normal and those that are considered aberrant. After reviewing the relevant anomaly detection prediction research and the benefits and drawbacks of the machine learning approach, we meticulously studied and selected the research articles.

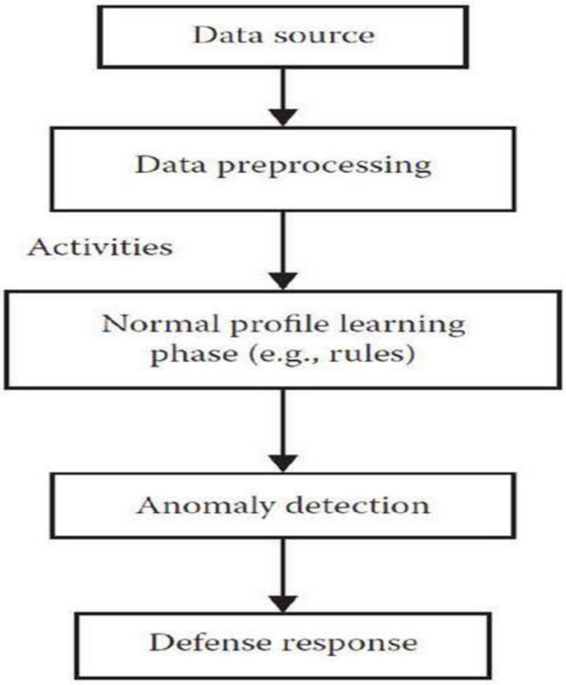

**Figure 1: Sequence of execution of modules in an anomaly detection system**

## II. RESEARCH OBJECTIVES

The primary goal of this study is to conduct a comprehensive literature evaluation of machine learning methods for anomaly identification and their practical applications. It also looks at how many research articles employ supervised anomaly detection classification and how accurate the machine learning models are Research Justification

In addition to shedding light on the most current findings in the field, we are certain that this study will provide academics a chance to get a better understanding of the many methods used for anomaly identification. The scarcity of Systematic Literature Reviews (SLR) addressing the topic of anomaly detection using machine learning techniques prompted us to conduct this research.

3. A Review of the Literature

Many areas of study and industry have looked at anomaly detection since it is such a big problem. Some anomaly detection approaches are more suited to specific use cases, while others are more general. Hosseinzadeh et al. (2021) offered a comprehensive review of anomaly detection methods and software as an example. During the board meeting, a thorough assessment of ML approaches and non-ML techniques, such as statistical and spectral detection methods, was covered in depth. In addition to that, the survey provides a plethora of other applications for identifying anomalies. Image processing, textual anomaly detection, cyber intrusion detection, medical anomaly detection, industrial damage detection, sensor networks, and fraud detection are just a few of the many applications for identification systems. A new study by Imran, Jamil, and Kim (2021) addressed the issue of finding outliers in discrete sequences. An extensive and coherent literature review on anomaly identification in symbolic or discrete sequences was presented by the

ISSN: 2057-5688

writers. Besides studies that looked at machine learning techniques and statistical anomaly detection in depth (Saba et al., 2022). In addition, the authors highlighted the benefits and drawbacks of each method by comparing and contrasting these. Regardless, it's worth studying the data mining survey proposed by Al Turaiki and El Tawijry (2021). The execution of this survey was deliberated. A small number of assessments primarily sought to identify outliers in certain domains and applications. For instance, Fourure et al. (2021) gives a thorough evaluation of popular clustering-based fraud detection techniques, contrasting and comparing them from various angles. They also supplied several models and classification techniques for the purpose of anomaly detection in automated monitoring, which is another area of interest.

The authors meticulously reviewed the research publications, taking into account the methodology, issue area, and procedure. In addition, Saba et al. (2022) summarized the three foremost methods utilized nowadays for anomaly detection in geochemical data processing. Machine learning, compositional data analysis, and fractal/multi-fractal models are a few examples of these approaches. In the meanwhile, machine learning approaches constitute the author's main area of interest. Research also suggested a summary of computer system performance, anomaly detection, and bottleneck identification, which is an added intriguing remark. After the authors identified the problem's core aspects, they categorized the numerous current remedies (Al Souci et al., 2021). Identifying unexpected invasions was the primary goal of many studies. In the meanwhile, a study examined intrusion detection approaches; however, the research mostly focused on systems that use machine learning (Ullah & Mahmood, 2021). They provided a synopsis of the machine learning methods used to address the textual issues with intrusion detection. The authors also compared related studies that used different datasets and classifier designs. Machine learning, data mining, anomaly detection, and intrusion detection strategies were all thoroughly evaluated by dini et al. (2023) for cyber intrusion detection.

Both of these experiments were conducted by researchers as part of the same overarching project. Additionally, the authors tackled the difficulties of using data mining and machine learning in the field of cyber security while

ISSN: 2057-5688

providing detailed descriptions of each strategy. In conclusion, an approach combining several machine learning algorithms with particle swarm optimization is necessary to improve the effectiveness of identifying abnormalities in network intrusion systems. A substantial amount of study has focused on discovering networks with outliers. As a result, many surveys were conducted to address the issue. As an example, consider Bahardiya's (2023) comprehensive research on network anomaly detection. They defined anomaly detection and evaluated its efficacy after they determined the typical types of threats that intrusion detection systems face. Network defenders' techniques and tools were also discussed by the authors. Also, in a comprehensive analysis, Rabel & Hussain (2023) looked at the most prevalent ways to find network abnormalities. Included in these strategies were density- and distance-based approaches, in addition to supervised and unsupervised learning techniques. Several machine learning techniques, such as deep recurrent neural networks, constrained Boltzmann machine-based deep belief networks, and others, are well suited to identifying abnormalities in networks. On top of that, the scientists provided evidence from tests that proved deep learning techniques may be useful for interpreting data from networks. In contrast to other systematic reviews, ours delves further into anomaly detection utilizing ML techniques, making it stand out. Lack of Research Additional research on the merits of supervised, semi-supervised, and unsupervised anomaly detection models, among others, would strengthen the article.

## III. RESEARCH METHODOLOGY

The primary goal of this study is to conduct a comprehensive literature evaluation of machine learning methods for anomaly identification and their practical applications. It also looks at how many research articles employ supervised anomaly detection classification and how accurate the machine learning models are Research Justification

In addition to shedding light on the most current findings in the field, we are certain that this study will provide academics a chance to get a better understanding of the many methods used for anomaly identification. The scarcity of Systematic Literature Reviews (SLR) addressing the topic of anomaly detection using machine learning techniques prompted us to conduct this research.

ISSN: 2057-5688

3. A Review of the Literature

Many areas of study and industry have looked at anomaly detection since it is such a big problem. Some anomaly detection approaches are more suited to specific use cases, while others are more general. Hosseinzadeh et al. (2021) offered a comprehensive review of anomaly detection methods and software as an example. During the board meeting, a thorough assessment of ML approaches and non-ML techniques, such as statistical and spectral detection methods, was covered in depth. In addition to that, the survey provides a plethora of other applications for identifying anomalies. Image processing, textual anomaly detection, cyber intrusion detection, medical anomaly detection, industrial damage detection, sensor networks, and fraud detection are just a few of the many applications for identification systems. A new study by Imran, Jamil, and Kim (2021) addressed the issue of finding outliers in discrete sequences. An extensive and coherent literature review on anomaly identification in symbolic or discrete sequences was presented by the writers. Besides studies that looked at machine learning techniques and statistical anomaly detection in depth (Saba et al., 2022). In addition, the authors highlighted the benefits and drawbacks of each method by comparing and contrasting these. Regardless, it's worth studying the data mining survey proposed by Al Turaiki and El Tawijry (2021). The execution of this survey was deliberated. A small number of assessments primarily sought to identify outliers in certain domains and applications. For instance, Fourure et al. (2021) gives a thorough evaluation of popular clustering-based fraud detection techniques, contrasting and comparing them from various angles. They also supplied several models and classification techniques for the purpose of anomaly detection in automated monitoring, which is another area of interest.

The authors meticulously reviewed the research publications, taking into account the methodology, issue area, and procedure. In addition, Saba et al. (2022) summarized the three foremost methods utilized nowadays for anomaly detection in geochemical data processing. Machine learning, compositional data analysis, and fractal/multi-fractal models are a few examples of these approaches. In the meanwhile, machine learning approaches constitute the author's main area of interest. Research also suggested a summary of computer system performance, anomaly detection, and bottleneck identification, which is an added intriguing remark. After the authors identified the problem's core aspects, they categorized the numerous current remedies (Al Souci et al., 2021). Identifying unexpected invasions was the primary goal of many studies. In the meanwhile, a study examined intrusion detection approaches; however, the research mostly focused on systems that use machine learning (Ullah & Mahmood, 2021). They provided a synopsis of the machine learning methods used to address the textual issues with intrusion detection. The authors also compared related studies that used different datasets and classifier designs. Machine learning, data mining, anomaly detection, and intrusion detection strategies were all thoroughly evaluated by dini et al. (2023) for cyber intrusion detection.

Both of these experiments were conducted by researchers as part of the same overarching project. Additionally, the authors tackled the difficulties of using data mining and machine learning in the field of cyber security while providing detailed descriptions of each strategy. In conclusion, an approach combining several machine learning algorithms with particle swarm optimization is necessary to improve the effectiveness of identifying abnormalities in network intrusion systems. A substantial amount of study has focused on discovering networks with outliers. As a result, many surveys were conducted to address the issue. As an example, consider Bahardiya's (2023) comprehensive research on network anomaly detection. They defined anomaly detection and evaluated its efficacy after they determined the typical types of threats that intrusion detection systems face. Network defenders' techniques and tools were also discussed by the authors. Also, in a comprehensive analysis, Rabel & Hussain (2023) looked at the most prevalent ways to find network abnormalities. Included in these strategies were density- and distance-based approaches, in addition to supervised and unsupervised learning techniques. Several machine learning techniques, such as deep recurrent neural networks, constrained Boltzmann machine-based deep belief networks, and others, are well suited to identifying abnormalities in networks. On top of that, the scientists provided evidence from tests that proved deep learning techniques may be useful for interpreting data from networks. In contrast to other systematic reviews, ours delves further into anomaly detection utilizing ML techniques, making it stand out. Lack of Research Additional research on the merits of supervised, semi-supervised, and unsupervised anomaly detection models, among others, would strengthen the article.

## IV. SELECTION CRITERIA RESULTS AND DISCUSSION

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.

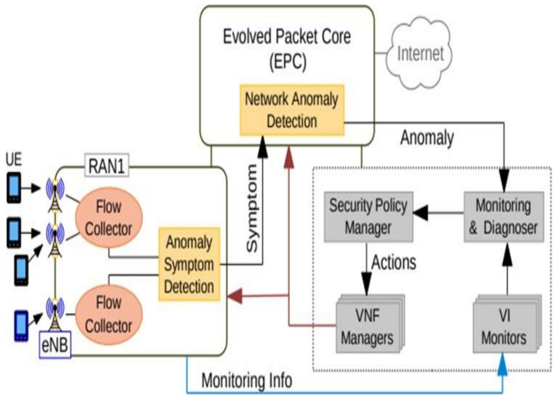

**Figure 2: Network anamoluly detection system (representing control action)**

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.

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. 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.

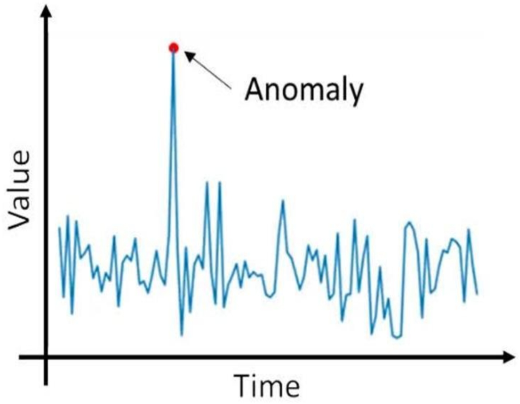

**Figure 3: Anamoly Detection via machine learning**

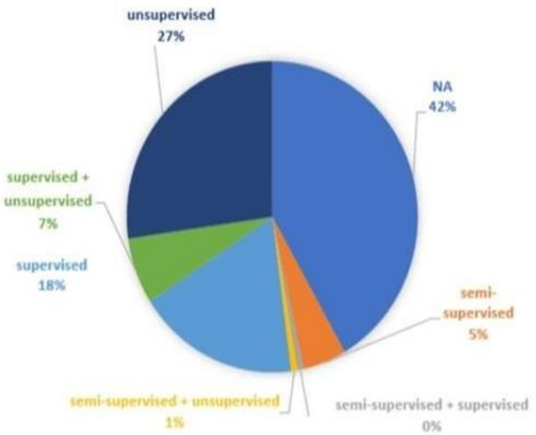

**Figure 4: Summary of articles included in the study as per their scope**

## V. CONCLUSION

We also see this as a critical area for development, given we Examining anomaly detection with ML approaches was the goal of this literature study. It assessed ML models from four perspectives: the kind of anomaly detection used, the type of ML method, the type of anomaly detection in ML theory (supervised, semi-supervised, and unsupervised), and the assessment of the accuracy of ML models. The publications published between 2015 and 2023 that were deemed relevant were evaluated by a literature review. The results show that intrusion detection, data applications, general anomaly detection, and network anomaly detection are the most popular areas of anomaly detection research. Here are some examples of anomaly detection in action from the selected publications. In addition to a large number of other generic datasets, the discovered research has made use of a diverse group of datasets that have been used in the experiments of linked papers. The bulk of studies used real-world datasets for training and testing models. Finally, a variety of research used anomaly detection methods based on categorization. Based on the research publications, we found that 6% of the selected studies utilized an unsupervised anomaly detection method, making it the most popular strategy. In second place, with 8 articles using it, was supervised anomaly detection. A combination of supervised and unsupervised anomaly detection approaches was utilized by the third percent of papers that followed.

Suggestions for enhancement We recommend more study on anomaly detection using machine learning to have a better grasp of the efficacy of these models. Researchers should also establish broad guidelines before utilizing ML models in research.

found research that did not specify which characteristics were utilized for statistical analysis. Several studies reported their results based on a single performance parameter (such accuracy), thus there has to be further research and development. It was also discovered that some other researchers were conducting their own researches using databases that were obsolete. Researchers are especially encouraged to use datasets that are more current. There are a number of restrictions on the study.

We also see this as a critical area for development, given we Examining anomaly detection with ML approaches was the goal of this literature study. It assessed ML models from four perspectives: the kind of anomaly detection used, the type of ML method, the type of anomaly detection in ML theory (supervised, semi-supervised, and unsupervised), and the assessment of the accuracy of ML models. The publications published between 2015 and 2023 that were deemed relevant were evaluated by a literature review. The results show that intrusion detection, data applications, general anomaly detection, and network anomaly detection are the most popular areas of anomaly detection research. Here are some examples of anomaly detection in action from the selected publications. In addition to a large number of other generic datasets, the discovered research has made use of a diverse group of datasets that have been used in the experiments of linked papers. The bulk of studies used real-world datasets for training and testing models. Finally, a variety of research used anomaly detection methods based on categorization. Based on the research publications, we found that 6% of the selected studies utilized an unsupervised anomaly detection method, making it the most popular strategy. In second place, with 8 articles using it, was supervised anomaly detection. A combination of supervised and unsupervised anomaly detection approaches was utilized by the third percent of papers that followed.

Suggestions for enhancement We recommend more study on anomaly detection using machine learning to have a better grasp of the efficacy of these models. Not only that, but researchers ought to establish a broad structure to begin ML model trials.

found research that did not specify which characteristics were utilized for statistical analysis. Several studies reported their results based on a single performance parameter (such accuracy), thus there has to be further research and development. It was also discovered that some other researchers were conducting their own researches using databases that were obsolete. Researchers are especially encouraged to use datasets that are more current. There are a number of restrictions on the study. Only conferences and journals devoted to anomaly detection and machine learning were consulted for this extensive literature review. At the outset of the review process, we were able to exclude a lot of research publications that were not relevant by using our search method. This ensured that the chosen research articles were appropriate for the investigation. It is possible that taking into account a broader variety of sources might have further enhanced this conclusion. Since we used a rigorous grading system for quality assurance, the same reasoning should be applied to quality assessment.

## REFERENCES:

1) Akalin, N., & Loutfi, A. (2021). Reinforcement learning approaches in social robotics. Sensors, 21(4), 1292.

2) Al-amri, R., Murugesan, R. K., Man, M., Abdulateef, A. F., Al-Sharafi, M. A., & Alkahtani, A. A. (2021). A review of machine learning and deep learning techniques for anomaly detection in IoT data. Applied Sciences, 11(12), 5320.

3) Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. Electronics, 12(6), 1333.

4) Balaji, T. K., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. Computer Science Review, 40, 100395.

5) Corallo, A., Lazoi, M., & Lezzi, M. (2020). Cybersecurity in the context of industry 4.0: A structured classification of critical assets and business impacts. Computers in industry, 114, 103165.

6) Djenna, A., Harous, S., & Saidouni, D. E. (2021). Internet of things meet internet of threats: New concern cyber security issues of critical cyber infrastructure. Applied Sciences, 11(10), 4580.

7) Dong, S. (2021). Multi class SVM algorithm with active learning for network traffic classification. Expert Systems with Applications, 176, 114885.

8) Heidari, A., & Jabraeil Jamali, M. A. (2023). Internet of Things intrusion detection systems: a comprehensive review and future directions. Cluster Computing, 26(6), 3753-3780.

9) Jacobsen, J. T. (2021). Cyber offense in NATO: challenges and opportunities. International affairs, 97(3), 703-720.

10) Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. Information Fusion, 97, 101804.

11) Klenka, M. (2021). Aviation cyber security: legal aspects of cyber threats. Journal of transportation security, 14(3), 177-195.

12) Nartin, S. E., Faturrahman, S. E., Ak, M., Deni, H. A., MM, C., Santoso, Y. H., ... & Eliyah, S. K. (2024). Metode penelitian kualitatif. Cendikia Mulia Mandiri.

13) Omolara, A. E., Alabdulatif, A., Abiodun, O. I., Alawida, M., Alabdulatif, A., & Arshad, H. (2022). The internet of things security: A survey encompassing unexplored areas and new insights. Computers & Security, 112, 102494.

14) Omuya, E. O., Okeyo, G. O., & Kimwele, M. W. (2021). Feature selection for classification using principal component analysis and information gain. Expert Systems with Applications, 174, 114765.

15) Pandey, S., Singh, R. K., Gunasekaran, A., & Kaushik, A. (2020). Cyber security risks in globalized supply chains: conceptual framework. Journal of Global Operations and Strategic Sourcing, 13(1), 103-128.

16) Quatrini, E., Costantino, F., Di Gravio, G., & Patriarca, R. (2020). Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities. Journal of Manufacturing Systems, 56, 117-132.

17) Rahmani, A. M., Yousefpoor, E., Yousefpoor, M. S., Mehmood, Z., Haider, A., Hosseinzadeh, M., & Ali Naqvi, R. (2021). Machine learning (ML) in medicine: Review, applications, and challenges. Mathematics, 9(22), 2970.

18) Saxena, N., Hayes, E., Bertino, E., Ojo, P., Choo, K. K. R., & Burnap, P. (2020). Impact and key challenges of insider threats on organizations and critical businesses. Electronics, 9(9), 1460.

19) Smith, R., Friston, K. J., & Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. Journal of mathematical psychology, 107, 102632.

20) Sturgeon, T. J. (2021). Upgrading strategies for the digital economy. Global strategy journal, 11(1), 34-57.

21) Sujith, A. V. L. N., Qureshi, N. I., Dornadula, V. H. R., Rath, A., Prakash, K. B., & Singh, S. K. (2022). A comparative analysis of business machine learning in making effective financial decisions using structural equation model (SEM). Journal of Food Quality, 2022(1), 6382839.

22) Taye, M. M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. Computers, 12(5), 91.