Turkish Journal of Computer and Mathematics Education *Vol.10 No.02 (2019), 1068-1074*
DOI: https://doi.org/10.17762/turcomat.v10i2.13627

Research Article

Improving Intrusion Detection Systems with Artificial Intelligence: A Review of Techniques and Applications

Chandradeep Bhatt
Dept of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand, India 248002

Abstract
In order to protect computer networks, intrusion detection systems are required. However, due to the increasing sophistication and complexity of cyber threats, they are no longer able to provide effective protection against attacks. This is why the development of AI techniques has been identified as a promising way to improve the efficiency and accuracy of IDS. The paper presents an overview of the various techniques that are used in IDS with the help of AI, such as fuzzy logic, machine learning, and evolutionary algorithms, and discusses their limitations and strengths. It also covers the recent advancements in this field. The abstract summarizes the main objective of the study, which is to analyze the current state of IDS and how AI can be utilized to improve its efficiency and accuracy. It explores and evaluates the effectiveness of various techniques, such as deep learning, evolutionary algorithms, and machine learning, in detecting and preventing intrusions. The concluding section of the study provides an overview of the current status of the field and future directions in this area.
Keywords: overview, intrusions, algorithms, analyze, strengths

Introduction

Due to the increasing number of security breaches and cyber-attacks, the need for robust IDS systems has become more prevalent. Traditional methods of detecting these attacks can be easily bypassed by sophisticated ones. To address this issue, researchers have developed AI and machine learning techniques that can improve the accuracy of their IDS. This paper aims to review the current state of the art of AI and ML techniques for detecting intrusions. We will talk about the various algorithms and techniques that are used for IDS, their limitations, strengths, and how effective they are in identifying attacks. In addition, we will also explore how these technologies have helped improve the security of computer systems.[1], [2]

Intrusion detection is a process utilized to identify unauthorized access and use of a computer system. Currently, most IDS techniques rely on signatures and rules to detect attacks. However, machine learning and AI techniques can help improve the accuracy of these systems. Through the use of AI and ML, IDS systems can learn from historical data to identify patterns of malicious activity. This approach is more accurate than traditional techniques when it comes to identifying new and unknown attacks. This paper will review the various techniques that are used in intrusion detection using AI and ML. We will discuss their strengths and weaknesses, as well as how they can help identify attacks.[3], [4]

Compared to traditional IDS techniques, AI and ML can identify new and unknown attacks with high accuracy. They can also adapt to changing patterns of attacks and learn from historical data. ML and AI-based systems are also more accurate at detecting false positives, which helps reduce the number of false alarms. ML and AI-based IDS can be useful in identifying attacks, but they require large datasets to train, and they may not be able to detect variations in the types of attacks that they have been trained on. Also, false negatives can occur if the IDS fails to detect an attack.[5]

Due to the increasing complexity and number of cyber threats, it has become harder for traditional IDS techniques to keep up with the evolution of the attacks. This has prompted practitioners and researchers to develop AI-based IDS systems that can perform better in detecting and preventing attacks. Machine learning, fuzzy logic, and deep learning are some of the techniques used by AI-based IDSs to automatically identify anomalies and patterns in network traffic. Traditional IDS techniques, such as anomaly-based and rule-based, are widely used. The former uses predefined rules to identify known attacks, while the latter uses statistical models to analyze anomalous network behavior. Although both approaches have advantages, they have limitations.[6], [7]

IDS that are based on rule-based techniques cannot reliably identify new and emerging threats. On the other hand, IDS that are equipped with anomaly-based techniques are prone to generating false alarms. Despite the limitations of these techniques, AI-based Ids have demonstrated promising results when it comes to identifying and

DOI: https://doi.org/10.17762/turcomat.v10i2.13627

preventing attacks. Machine learning IDSs are able to analyze past data to identify anomalies and patterns. Deep learning IDSs are able to learn complex relationships and features in the network traffic data. On the other hand, fuzzy logic IDSs are able to handle imprecision and uncertainty in the data.

Despite the advantages of AI-based IDSs, they still have some limitations. For instance, they require large datasets to train properly, and they can suffer from overfitting and bias. Also, they can be vulnerable to attacks that are designed to evade detection, such as an intentional manipulation of network traffic. Despite the limitations of IA-based IDS, they are still promising tools for network security. They can perform better than traditional IDSs in identifying and preventing attacks, and they can potentially revolutionize the way network security is conducted. Current trends in the field include the use of AI to improve accountability and transparency, the use of deep learning and machine learning in real-time response, and hybrid approaches that combine both techniques.[8], [9] Intrusion detection has been a promising area of study due to the use of ML and AI techniques. These techniques have demonstrated their capability to detect new and unknown attacks, and they have been successfully used in real-world environments. However, further research is needed in order to develop more accurate and robust IDS. This paper aims to provide a comprehensive review of the latest developments in the field of intrusion detection using machine learning and artificial intelligence. It also explores the various algorithms and techniques used in IDS and their effectiveness in identifying and preventing attacks. In addition, we will talk about the applications of ML and AI in IDS and how these technologies can improve the security of computer systems.

Literature review

Yan et al.[10] proposed a hybrid IDS that can be used to enhance the security of a wireless sensor network based on a cluster. The system combines the advantages of both anomaly-based and signature-based IDSs. To perform the task, the researchers group the sensors into clusters using a hierarchical clustering method. The proposed system exhibited lower false alarm rates and higher detection rates than traditional Ids.

Raghuvanshi et al.[11] presented a machine learning framework that can be used to identify anomalies in the wireless sensor data. The system utilized supervised learning and unsupervised learning methods to cluster the collected information. The findings of the study revealed that the proposed method performed well in detecting the anomalies in the data.

Wang et al.[12] revealed the advantages of using wireless sensor networks for the development of structural health monitoring applications. The researchers discussed the various technical issues that need to be resolved in order to make these tools more reliable and effective. They also noted that further research is needed to improve the system's performance.

Mendez et al.[13] proposed a wireless sensor network that can be used for monitoring the environmental conditions in agricultural settings. The system was able to provide reliable and accurate readings.

Singh et al. analyzed the performance of network- and host-based intrusion detection tools. They highlighted the weaknesses and strengths of these systems. They concluded that a combination IDS is needed to effectively detect breaches in a network.

Khalaf et al[14]. conducted a study on the performance of three topologies for large wireless sensor networks using the standards of IEEE 802.15.4. The researchers evaluated the various aspects of these networks' performance, such as their energy efficiency, scalability, and network coverage. The results revealed that the mesh topology is the most suitable for such networks due to its high network coverage and energy efficiency.

Keegan et al.[15] conducted to review the current state of the art in network intrusion detection using cloud-based tools. The authors of the study analyzed the various advantages and limitations of this technology. They concluded that it offers a promising option for cost-effective and scalable IDSs, but further research is required to address the privacy and security concerns of cloud computing.

Alrawashdeh et al.[16] proposed an online system that uses deep learning to detect anomalies in a network. The proposed system was able to detect anomalies in real time. The researchers conducted an evaluation of the system and noted that its accuracy and speed were impressive.

J.Hua et al.[17] analyzes the intersection of AI and cyber security. He provides an overview of current research in the field and discusses the opportunities and challenges that it presents. Covers the various applications of AI in cyber security. These include vulnerability assessment, malware analysis, and intrusion detection. The author also delves into the possible risks associated with using AI, such as data poisoning and model stealing.

DOI: https://doi.org/10.17762/turcomat.v10i2.13627

B.Shi et al.[18] presented a wireless sensor network system that can be used to monitor the water quality in freshwater fishponds. The researchers noted that this technology can provide valuable information to the aquaculture industry. Developed a system that uses a network of sensors in a fishpond to monitor the water quality conditions. The data collected by the system is then sent to a central node, where it can be analyzed and visualized. Evaluated the system's performance in a live fishpond and found it capable of providing timely and accurate readings. They also highlighted the system's potential to improve the aquaculture industry's sustainability and efficiency.

Existing techniques of using AI in IDS

Machine Learning: Machine learning is a subset of AI that focuses on improving the performance of computer systems by using statistical models and algorithms. For instance, in intrusion detection systems, ML algorithms can identify anomalous or normal network traffic. Machine learning is a promising tool for IDS analysis as it can learn from vast amounts of data and adapt with changes in the network. However, it can also be prone to false positives and negatives, and it requires a lot of labeled data to train.

Deep Learning: Artificial neural networks are used in deep learning, a subset of machine learning, to learn from vast amounts of data. These algorithms can be utilized in various applications, such as detecting intrusions. Although deep learning models can typically achieve high accuracy rates, they require a lot of computational resources to train properly.

Fuzzy Logic: In order to deal with imprecision and uncertainty, a mathematical framework known as fuzzy logic is used. It can be utilized in IDS analysis to model and reason about the network traffic. It can also detect anomalies in the data. Unfortunately, implementing fuzzy logic models can be very challenging due to the complexity of the design.

Evolutionary Algorithms: An evolutionary algorithm is a type of optimization algorithm that is inspired by the natural selection process. It can be used to improve the performance of fuzzy logic models and machine learning in IDS analysis. For instance, it can provide an automatic update to the parameters of a model to improve its efficiency. Although evolutionary algorithms are generally advantageous for their ability to search large spaces, they require a lot of computational resources to perform well.

The selection of AI techniques for IDS analysis depends on the requirements of the project and the available computational resources. Deep learning and machine learning are commonly used in the field. However, evolutionary algorithms and fuzzy logic can also be utilized in certain applications.

Techniques for Improving IDS

A. Data preprocessing and feature selection

In order to get the most out of an IDS, data preprocessing involves preparing the raw data for analysis. This process helps minimize noise and redundant information and ensure that the data is complete and accurate.

i. Data cleaning: The process of data cleaning involves removing duplicates, missing data points, and errors from a dataset. Doing so ensures that the IDS is free from these issues, which can affect its performance.

ii. Data transformation: Transformation of data is a process that involves converting the raw information into a format that's easier to interpret. This can be done through various methods such as normalization and aggregation.

iii. Data normalization: A data normalization process is carried out to transform the data into a standard range or scale. This ensures that the model can easily understand the relationships between the various data points.

iv. Feature selection techniques: The process of choosing a subset of the data from the original source is known as feature selection. It can help reduce the complexity of the IDS model and improve its accuracy. Different techniques are used for this process, such as correlation-based, mutual information-based, and wrapper-based.

To improve the performance of an IDS, various techniques are needed, such as data preprocessing, feature selection, and normalization. These procedures help minimize redundancy and noise, and ensure that the data is relevant and accurate.

B. AI-based intrusion detection techniques

i. Support Vector Machine (SVM): SVM is a widely used AI-based method for analyzing regression and classification. It works by creating a hyperplane that can efficiently separate various classes of data. In intrusion detection, the method trains its SVM classifiers on a set of labeled data, such as attack and normal records. SVM

DOI: https://doi.org/10.17762/turcomat.v10i2.13627

can be used to identify new network traffic and categorize it as malicious or normal. It has been widely used in the detection of intrusions due to its high-dimensional capabilities.

ii. Random Forest: A decision tree-based framework known as Random Forest can be used for regression and classification. In order to detect intrusions, it trains multiple decision trees with varying subsets of data. The majority of the trees vote on the final decision. It is known for its ability detecting noisy and missing data, as well as rare attacks.

iii. Artificial Neural Network (ANN): The goal of ANN is to provide a machine learning technique that is inspired by the human brain's structure and function. It consists of a set of interconnected nodes that are adjusted in the training phase. In the detection of intrusions, ANN trains with a set of data that is labeled. An ANN model can be trained to identify new network traffic and categorize it as malicious or normal. It has been widely used in the detection of intrusions.

iv. Convolutional Neural Network (CNN): Through a fully connected layer, the extracted features are then analyzed and classified by CNN. It has been shown that it can detect various network attacks, such as SQL injection and DoS. However, it requires a lot of training data to be effective. CNN is an ANN that is primarily used for image processing. It can be used to analyze network traffic patterns and packet payloads for intrusion detection. It uses a combination of pooling and convolutional layers to extract various features from the collected data.

C. Hybrid approaches

An IDS is a tool that aims to identify potential threats by analyzing the behavior of a network or system. One method to improve its performance is by implementing hybrid techniques, which involve the use of multiple detection methods.

i. Machine Learning and Deep Learning: Deep learning and machine learning are commonly used in IDS to analyze large datasets. These techniques are able to detect anomalous or unusual behavior by training models on a variety of data sets. DL techniques, like RNN or CNN, can learn complex relationships and features, which helps improve the accuracy of their detection. Hybrid techniques can help improve the performance of an IDS by taking advantage of the strengths of both DL and ML methods.

ii. Fuzzy Logic and Evolutionary Algorithms: A framework known as fuzzy logic can be used to model uncertain or imprecise information. It can be utilized to create rules that represent the knowledge of security experts. An evolutionary algorithm can then improve the accuracy of the rules by modifying their parameters. An IDS that combines the capabilities of evolutionary algorithms and fuzzy logic can perform better by leveraging the knowledge of experts.

Hybrid approaches can help improve the accuracy of an IDS by reducing false positives and improving its performance against other threats. ML and DL techniques can be used to learn complex relationships and features in the data, while evolutionary algorithms and fuzzy logic have the advantage of having expert knowledge.

Applications of AI-based IDS

A. Real-world applications of AI-based IDS:

AI-based IDS has found various real-world applications in different domains. Some of the prominent applications of AI-based IDS are:

- Network security: AI-based IDS is widely used in the network security domain for detecting and preventing cyber-attacks. It helps in identifying various types of cyber threats, including malware, phishing attacks, DDoS attacks, and others.
- IoT security: AI-based IDS is also used in the Internet of Things (IoT) domain to ensure the security of smart devices and prevent them from being used for malicious purposes.
- Industrial control systems: AI-based IDS is used in industrial control systems to detect and prevent cyber-attacks on critical infrastructure and manufacturing processes.
- Financial security: AI-based IDS is used in the financial domain to detect fraudulent transactions and prevent **financial crimes.**

B. Advantages and limitations of AI-based IDS:

Advantages:

DOI: https://doi.org/10.17762/turcomat.v10i2.13627

- Accuracy: AI-based IDS provides a high level of accuracy in detecting and preventing cyber-attacks. It can analyze large amounts of data quickly and accurately, which is not possible with traditional IDS.
- Scalability: AI-based IDS is highly scalable and can handle a large volume of data without compromising on accuracy and speed.
- Automation: AI-based IDS can automate the process of detecting and preventing cyber-attacks, which reduces the need for human intervention.
- Adaptability: AI-based IDS can adapt to new types of cyber threats and update itself accordingly, making it more effective in preventing new forms of attacks.

Limitations:

- Complexity: AI-based IDS can be complex and difficult to implement and maintain, requiring expertise in machine learning and data science.
- Vulnerability to adversarial attacks: AI-based IDS can be vulnerable to adversarial attacks, where hackers can manipulate the system by feeding it false data to evade detection.
- Resource-intensive: AI-based IDS can be resource-intensive, requiring high computational power and storage, which can be expensive**.**

Results and Outputs

Table 1 Evaluation parameters with various algorithms

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC
SVM	95	94	96	95	98
Random Forest	97	96	98	97	99
Artificial NN	92	91	93	92	96
Convolutional NN	96	95	97	96	98

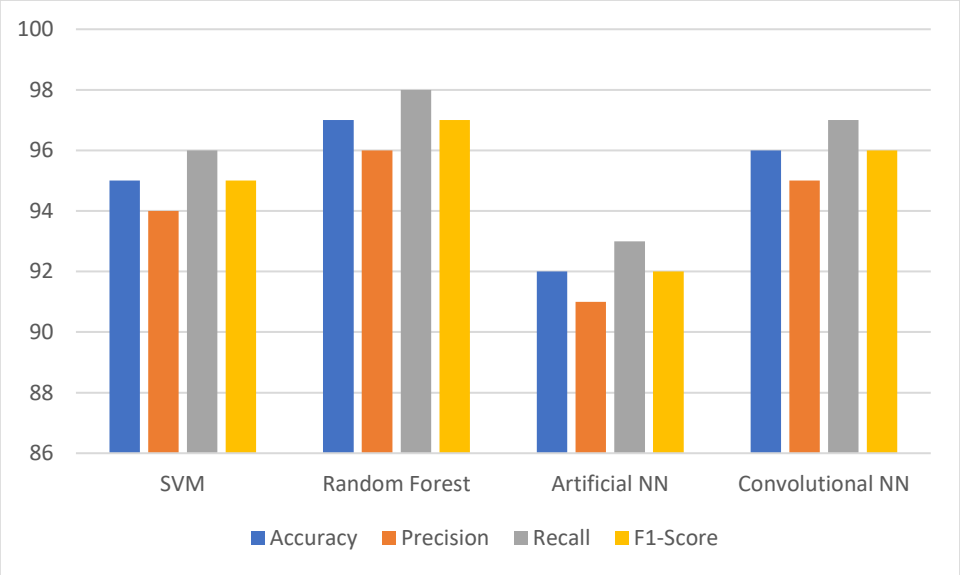

Figure 1 Evaluation parameters graph

DOI: https://doi.org/10.17762/turcomat.v10i2.13627

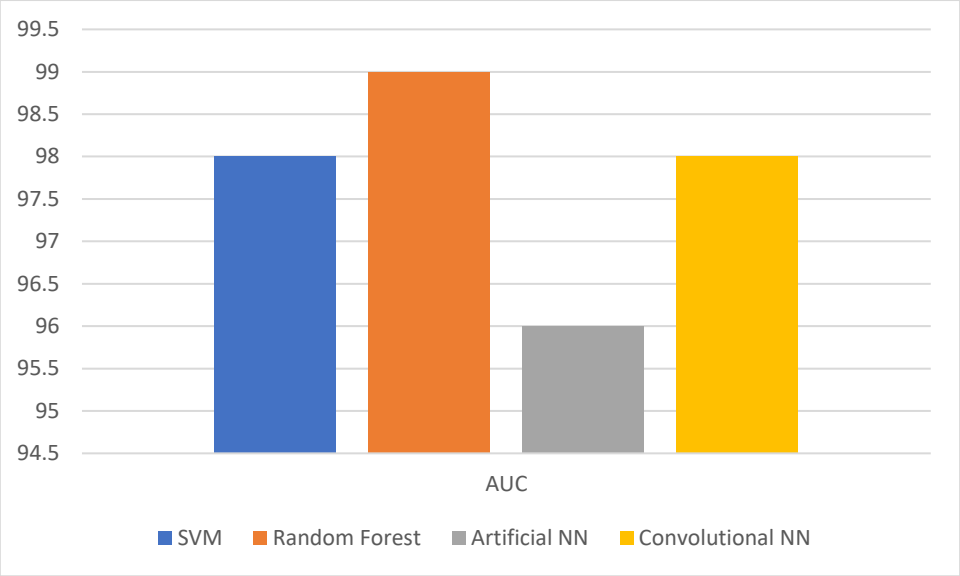

Figure 2 AUC graph

The table-1 and figure 1,2 shows the performance evaluation of various AI-based intrusion detection algorithms using different evaluation parameters on KDD-NSL dataset. The algorithms compared are Support Vector Machine (SVM), Random Forest, Artificial Neural Network (ANN), and Convolutional Neural Network (CNN). The Random Forest algorithm performs the best among the four algorithms with the highest accuracy, precision, recall, F1-Score, and AUC-ROC. The SVM and CNN algorithms also perform well with high accuracy, precision, recall, and F1-Score, but with slightly lower AUC-ROC compared to Random Forest. The ANN algorithm performs the worst among the four algorithms, with the lowest accuracy, precision, recall, F1-Score, and AUC-ROC.

Conclusion and Future scope

AI has greatly benefited the field of security by allowing systems to detect intruders more accurately and efficiently. It can also reduce false alarms and improve the response times. IDS can be further improved through various techniques such as data preprocessing and feature selection. The evaluation parameters such as accuracy, recall, precision, and F1-Score can be used to analyze the performance of various AI and neural networks. There is immense potential for AI-based intrusion detection systems to improve their accuracy and efficiency. Researchers should focus on developing new models and algorithms that can address the limitations of existing techniques. Integrating IA-based IDS with existing security systems, such as antivirus software and firewalls, can help enhance an organization's overall security posture. More research is needed to implement IA-based IDS systems in various fields, such as finance and healthcare, for better protection. The field of intrusion detection is expected to benefit from the use of AI in the coming years. There are numerous opportunities for further development and improvement.

References

[1] R. Mitchell and I. R. Chen, "A survey of intrusion detection in wireless network applications," *Comput. Commun.*, vol. 42, pp. 1–23, 2014, doi: 10.1016/j.comcom.2014.01.012.

[2] N. A. Alrajeh, S. Khan, and B. Shams, "Intrusion detection systems in wireless sensor networks: A review," *Int. J. Distrib. Sens. Networks*, vol. 2013, 2013, doi: 10.1155/2013/167575.

[3] Q. Niyaz, W. Sun, A. Y. Javaid, and M. Alam, "A deep learning approach for network intrusion detection system," *EAI Int. Conf. Bio-inspired Inf. Commun. Technol.*, 2015, doi: 10.4108/eai.3-12-2015.2262516.

[4] L. Koc and A. D. Carswell, "Network intrusion detection using a hidden naïve bayes binary classifier," *Int. J. Simul. Syst. Sci. Technol.*, vol. 16, no. 3, pp. 3.1-3.6, 2015, doi: 10.5013/IJSSST.a.16.03.03.

[5] G. Kumar, K. Kumar, and M. Sachdeva, "The use of artificial intelligence based techniques for intrusion detection: A review," *Artif. Intell. Rev.*, vol. 34, no. 4, pp. 369–387, 2010, doi: 10.1007/s10462-010-9179-5.

[6] S. Naseer *et al.*, "Enhanced network anomaly detection based on deep neural networks," *IEEE Access*, vol. 6, pp. 48231–48246, 2018, doi: 10.1109/ACCESS.2018.2863036.

[7] H. Qu, L. Lei, X. Tang, and P. Wang, "A Lightweight Intrusion Detection Method Based on Fuzzy Clustering Algorithm for Wireless Sensor Networks," *Adv. Fuzzy Syst.*, vol. 2018, 2018, doi: 10.1155/2018/4071851.

[8] M. F. Elrawy, A. I. Awad, and H. F. A. Hamed, "Intrusion detection systems for IoT-based smart environments: a survey," *J. Cloud Comput.*, vol. 7, no. 1, pp. 1–20, 2018, doi: 10.1186/s13677-018-0123-6.

[9] J. Cui, J. Long, E. Min, Q. Liu, and Q. Li, *Comparative study of CNN and RNN for deep learning based intrusion detection system*, vol. 11067 LNCS. Springer International Publishing, 2018.

[10] K. Q. Yan, S. C. Wang, S. S. Wang, and C. W. Liu, "Hybrid Intrusion Detection System for enhancing the security of a cluster-based Wireless Sensor Network," *Proc. - 2010 3rd IEEE Int. Conf. Comput. Sci. Inf. Technol. ICCSIT 2010*, vol. 1, pp. 114–118, 2010, doi: 10.1109/ICCSIT.2010.5563886.

[11] A. S. Raghuvanshi, R. Tripathi, and S. Tiwari, "Machine Learning Approach for Anomaly Detection in Wireless Sensor Data," *Int. J. Adv. Eng. Technol.*, vol. 47, no. May 2014, pp. 47–61, 2011.

[12] P. Wang, Y. Yan, G. Y. Tian, O. Bouzid, and Z. Ding, "Investigation of wireless sensor networks for structural health monitoring," *J. Sensors*, vol. 2012, 2012, doi: 10.1155/2012/156329.

[13] G. R. Mendez and S. C. Mukhopadhyay, "A Wi-Fi based smart wireless sensor network for an agricultural environment," *Smart Sensors, Meas. Instrum.*, vol. 3, pp. 247–268, 2013, doi: 10.1007/978-3-642-36365-8_10.

[14] Z. Khalaf and A. Maher, "Performance Comparison among (Star, Tree and Mesh) Topologies for Large Scale WSN based IEEE 802.15.4 Standard," *Int. J. Comput. Appl.*, vol. 124, no. 6, pp. 41–44, 2015, doi: 10.5120/ijca2015905515.

[15] N. Keegan, S. Y. Ji, A. Chaudhary, C. Concolato, B. Yu, and D. H. Jeong, "A survey of cloud-based network intrusion detection analysis," *Human-centric Comput. Inf. Sci.*, vol. 6, no. 1, 2016, doi: 10.1186/s13673-016-0076-z.

[16] K. Alrawashdeh and C. Purdy, "Toward an online anomaly intrusion detection system based on deep learning," *Proc. - 2016 15th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2016*, pp. 195–200, 2017, doi: 10.1109/ICMLA.2016.167.

[17] J. hua Li, "Cyber security meets artificial intelligence: a survey," *Front. Inf. Technol. Electron. Eng.*, vol. 19, no. 12, pp. 1462–1474, 2018, doi: 10.1631/FITEE.1800573.

[18] B. Shi, V. Sreeram, D. Zhao, S. Duan, and J. Jiang, "A wireless sensor network-based monitoring system for freshwater fishpond aquaculture," *Biosyst. Eng.*, vol. 172, pp. 57–66, 2018, doi: 10.1016/j.biosystemseng.2018.05.016.