jamk

Investigation and prevention of cybercrimes using Artificial Intelligence

Godwin Stephen

Master’s thesis
May 2025
Master's Degree Programme in Information Technology, Cyber Security

jamk | Jyväskylän ammattikorkeakoulu
University of Applied Sciences

Stephen, Godwin

Investigation and prevention of cyber-crimes using Artificial Intelligence

Jyväskylä: Jamk University of Applied Sciences, May 2025, 68 Pages

Master's Degree Programme in Information Technology, Cyber Security. Master's thesis.

Permission for open access publication: Yes

Language of publication: English

Abstract

Artificial intelligence (AI) has made huge impact in cybercrime and cybersecurity over the recent years. While AI has enabled new forms of cyberattacks it has also provided newer and better performing AI powered cyber defense tools. This narrative review is aimed at reviewing the algorithms behind the prevalent AI anomaly detection tools and assessing their performance based on key metrics.

This review employed a narrative approach followed by a comparative analysis. Two databases were used to search for the relevant literature. Studies were included based on the relevance to the research questions and the time of publication. Comparison of machine learning (ML) models based on key performance metrics such as accuracy, precision, recall and F1-Score was carried out.

Fundamental concepts behind ML techniques, performance metrics, and their application within the domain of cloud security are discussed. Among selected AI-based anomaly and malware detection techniques the unsupervised learning based DBSCAN method delivered excellent performance, while deep learning methods showed significant improvement in identifying new and unknown attack patterns. Though supervised models had limitations in terms of false negative rates, they delivered better accuracy in detecting known anomaly patterns. Additionally, real-world AI driven anomaly detection tools such as the Microsoft Sentinel and the DeepLog have robust machine leaning capabilities and efficiency in combating cyberattacks. Deepfake detection tools, including Intel's FakeCatcher and DeepFake-O-Meter, also delivered excellent accuracy in identifying fake media.

ML models are well-suited for combating modern cyberattacks and integrating multiple ML models based on key performance metrics can further strengthen the cyber defense systems based on tailored needs.

Keywords/tags (subjects)

Cybercrime, Cyber defense, Artificial Intelligence, Machine Learning, Supervised Learning, Unsupervised Learning, Deep Learning

Miscellaneous (Confidential information)

Contents

Figures

Tables

1 Introduction

Competition has always been key factor for evolution. Freeze (2020) describes humans like many other species on our planet evolve in parallel, and constantly seeking competitive edge over the others. Similarly, cybercrime and cybersecurity have evolved in parallel. Cybercriminals evolve their tactics to exploit new vulnerabilities, while cybersecurity experts constantly advancing defensing mechanism against these evolving threats. Cybercriminals and attackers seek to exploit weaknesses in the system or people for personal gain and may also be state sponsored to advance state interests. Artificial Intelligence (AI) is reshaping the cybersecurity landscape in the current era, introducing new challenges and opportunities for both attackers and defenders.

AI significantly enhances the capabilities of the cybercriminals to perform the precision and large scale cyberattacks, from automating phishing schemes to deploying sophisticated malware (Sai Meghana et al., 2024). Previously, spear-phishing attacks were easier to identify due to their poor formatting. However, with advanced AI technologies like large language models (LLMs), phishing emails often appear to come from legitimate sources due to their clear formatting and effective use of language which has significantly increased the success rate of malicious phishing attacks. (Torre, 2023). Cybercriminals are increasingly using advanced generative AI technology to create realistic deepfake images and videos. These fraudulent attempts surged by up to 3000% between 2022 and 2023, with many people unable to differentiate between genuine and fake content. This opens up opportunities for cybercriminals, putting all businesses and individuals at risk of targeted attacks (McBride, 2024).

A lethal weapon employed by cybercriminals for targeting various organizations is malware. Although cyber defense systems typically identify and block such malwares, AI-powered attacks introduce self-modifying malware that rapidly evolves to evade the traditional detection systems. This advancement allows cybercriminals to bypass established security measures more effectively (Djenna et al., 2023).

In response to this, the cybersecurity industry is developing advanced systems to counteract these AI-driven cyberattacks. AI-based deep learning methods are being used to analyze and detect various malware families and enhancing detection capabilities to identify sophisticated cyber

threats (Djenna et al., 2023). Companies are now encouraged to develop their own AI-powered defense systems, including adversarial AI and anomaly detection to counter advanced cyber threats in real time (Torre, 2023).

While the field is swiftly advancing and AI powered cybersecurity tools are being introduced at a rapid pace, the algorithms behind these tools (e.g., Machine Learning (ML) models—supervised, unsupervised, and deep learning) need critical evaluation on their performance and applicability. This will help the users (individual or companies) identify the correct tools for the correct tasks.

This thesis aims to critically appraise the performance of algorithms behind the prevalent AI anomaly detection tools. This narrative review provides a comparative analysis of these algorithms based on key performance indicators based on metrics like accuracy, precision, recall, and F1-score. In addition, real-world tools such as Microsoft Sentinel and DeepLog are also reviewed.

2 Theoretical Background

AI is used by cybercriminals to carry out advanced attacks, but the same technology supports and strengthens modern cyber defense systems. A critical understanding of traditional signature-based detection method, modern ML approaches such as supervised, unsupervised, and deep learning based anomaly detection methods is needed to select the appropriate tool for specific needs. In the following sections, each of these methods are explained in terms of how they work and their contribution in classifying and identifying threats. Figure 1. systematically describes the formation of the theoretical background based on the existing literature

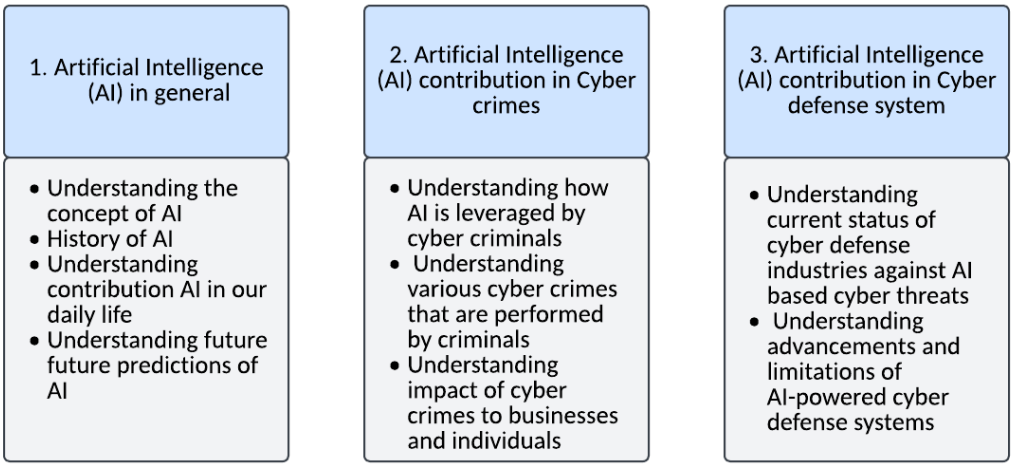

Figure 1. Literature review process for the thesis

2.1 Artificial intelligence (AI) – a brief introduction

The term "Artificial Intelligence" was first introduced by John McCarthy in 1956. Since then, numerous definitions have been proposed. For example, Gupta and Mangla (2020) describe AI as the effort to enable computers to perform tasks at which humans typically excel. AI is broadly defined as the science and engineering of creating intelligent machines, particularly intelligent computer programs. These systems are designed to mimic human intelligence by learning from experience, reasoning, solving problems, adapting through self-correction, and even demonstrating creativity (Demkovych, 2024) . AI significantly contributes to various aspects our

daily life, including healthcare, education, customer support, and cybersecurity. It offers tailored services by employing specific subsets or applications which include,

Machine Learning (ML) comprises computer algorithms developed by humans that learn and improve based on their previous experience. ML algorithms use training data to create models that can make predictions or judgments without being explicitly programmed (Hua, 2022). ML techniques play a vital role in the cybersecurity industry. They are capable of analyzing large volumes of data to identify attack patterns and types and it can predict potential threats before they occur (Roponena et al., 2021).

Artificial Neural Networks are computational algorithms, inspired by the human nervous system, and are designed to create complex data models. These algorithms communicate with each other like neurons in the human brain, working together to solve complex problems. The neural network approach enables data extraction for understanding trends and patterns in data where human analyzing capabilities may be limited (Abdolrasol et al., 2021). Neural networks also perform well to identify emerging cyber threats through intrusion detection systems (IDS). The ability of neural networks to adapt and learn newer techniques enables the detection of novel attacks where traditional cyber defense systems may face challenges (Sukhvinder Singh Dari, 2024).

Large Language models (LLMs) utilize advanced AI technology, and are capable of processing and generating text for real-time communication and performing multiple tasks simultaneously which enables efficient handling of complex interactions and processes in different applications (Naveed et al., 2024). LLMs contribute to different cybersecurity domains in fighting against threats, identifying and classifying malware families based on textural analysis, and in detecting phishing attacks by analyzing malicious content and intentions (Xu et al., 2024).

Generative Pre-Trained Transformer (GPT) models are based on neural networks. The are used to perform Natural Language Processing (NLP) tasks e.g., interpreting multiple languages, generating and analyzing text. These transformer models also have the capability to self-evaluate their outputs at various lengths, ensuring better quality and accurate results to the users (Yenduri et al., 2024).

2.2 AI in cybercrime and cyber defense

The rapid advancement in the field has increased the use of AI-powered tools by cybercriminals. In response to these sophisticated threats, the cybersecurity landscape in also evolving by leveraging the AI for defense purposes discussed in detail below.

2.2.1 The cybercrime perspective

AI is increasingly utilized by cybercriminals to conduct sophisticated attacks targeting businesses and private individuals, either for financial gain or simply to demonstrate their capabilities. With technological advancements, cyberattacks and crimes have escalated rapidly, impacting nearly every branch of science and engineering. According to a McAfee-led analysis, cybercrimes have caused significant global damage, amounting to approximately $600 billion, or around 1% of global GDP (Johns, 2022).

Current studies indicate that cybercriminals are exploiting IoT-based technologies to develop and execute malware and ransomware attacks, further augmented by AI technologies. If this trend continues, it will expand the attack surface to encompass over 2.5 million fully connected online devices, including personal, healthcare, and industrial devices (Velasco, 2022). According to Microsoft Digital Defense Report (2024) the complexity of attack tactics and techniques has doubled over the past decade, resulting in a 79% increase in attack indicators since 2020.

Misinformation is a growing trend where AI-powered bots are used to spread false information through social media, particularly targeting the younger generation who may struggle to distinguish between real and fake news. This can lead to emotional distress, crimes, and political instabilities (Velasco, 2022). While misinformation might not seem harmful on the surface, its effects can still be damaging, causing confusion and leading to serious consequences (European Commission, 2024).

Another popular cybercrime involving AI technology is the creation of deepfake content, including images, voices, and videos. The use of deepfake technology by cybercriminals has rapidly increased for malicious purposes (Velasco, 2022). The malicious use of deepfake technology poses

significant threats to society, including fraud, emotional blackmail, money extortion, and political disinformation. While the technology itself is legal, using it for harmful purposes is not (Boucher, 2021).

AI-powered cybercrimes allow criminals to perform more sophisticated attacks such as malware, spear phishing attacks, and deepfake contents that significantly impact both businesses and individuals.

Malware is malicious software designed to infect and damage information systems. Cybercriminals utilize various strategies to distribute malware across businesses for financial gain or other motives. AI-driven malware attacks employ ML techniques to train and adapt malware patterns in real time, making them increasingly difficult to detect with traditional malware detection tools (Mustafa Majid et al., 2023). This advancement allows cybercriminals to bypass the security measures more effectively (Djenna et al., 2023). Modern malware represents some of the most devastating forms of cybercrime, as it can evade detection and render the security analysis team's real-time investigations nearly impossible. The impact of these attacks can be both disastrous and unpredictable (Djenna et al., 2023).

Usage of generative AI in cyberattacks is a growing concern. Cybercriminals employ large language models (LLMs) to create malware, generate disruptive scripts, and craft convincing phishing content. The AI can be misused for such purposes even by persons with limited technical skills. An AI-generated phishing email was able to convince more than 75% of recipients to click on a harmful link demonstrating just how powerful and accessible AI-based cyberattacks have become. (Usman et al., 2024).

Spear phishing attack is an email type of cyberattack where the attacker contacts victims via email, posing as a legitimate source, to force them into actions they wouldn't normally take. These actions can lead to the disclosure of confidential information or the transfer of money to the attacker (Eze & Shamir, 2024). Earlier phishing emails were often easily identifiable due to their poor formatting, mostly these email originated from the country where English is not primary language (Torre, 2023). Traditional phishing emails can typically be identified and blocked by providing proper training to individuals and implementing URL filtering techniques (Eze & Shamir,

2024). These methods help users recognize suspicious emails and prevent access to harmful links. Phishing attacks increased by 58% in 2023, resulting in an estimated financial impact of around $3.5 billion in 2024. Between July 2023 and June 2024, approximately 775 million emails were reported to contain malware (Microsoft Digital Defense Report, 2024).

AI-driven spear phishing emails present new challenges in cybersecurity domain. These sophisticated attempts often mimic messages from legitimate sources with enhanced authenticity and improved language, making them harder to detect (Mohamed et al., 2025). Generative AI tools have simplified the process for attackers to perform large-scale spear phishing attacks by creating unique emails that are difficult to distinguish as being machine-generated or human written. The large volume of unique emails generated by AI tools makes it challenging for spam detection systems to effectively identify and block these messages (Eze & Shamir, 2024).

Advancements in language modeling have led to the development of AI systems capable of performing human-like tasks across numerous natural languages, particularly when considering their scale. Large Language Models (LLMs) have shown remarkable progress in generating personalized and convincing content, including spear phishing emails that closely mimic human communication patterns. GPT-4 and GPT-3.5 represent significant improvements in LLM technology, demonstrating enhanced capabilities in creating personalized and human-like emails (Hazell, 2023).

Deepfake content is generated using existing images or videos to produce authentic-looking fake content. The accessibility of AI-based deepfake tools has significantly lowered the technical barrier for creating convincing fake content. This ease of use enables attackers with limited technical expertise to produce deepfakes, potentially for malicious purposes (Naitali et al., 2023). AI generated and synthesized deepfake audio has become increasingly difficult to distinguish between real and fake audio (Rabhi et al., 2024) Deepfake technology employees five manipulation types to perform the tasks illustrated in the Figure 2. Identity swap or face swapping manipulation method involves replacing an existing face in a video, referred to as the source, with another person's face, known as the target. During this process, the expressions of the target person are transferred onto the source face, creating malicious video content that seamlessly integrates the target's appearance into the source video (Naitali et al., 2023).

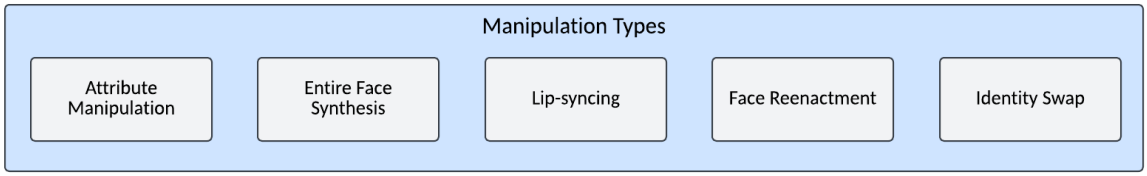

Figure 2. Deepfake manipulation types

Deepfake technology is evolving rapidly, with latest techniques like Generative Adversarial Networks (GANs) and Autoencoders transforming the criminal landscape. GANs function by having two neural networks, the generator and the discriminator to compete against each other to produce highly realistic fake images. Autoencoders, on the other hand, learn to compress and reconstruct facial features, and are used in pairs to map one person's face onto another. Alongside such technical advances, easy-to-use apps like FaceApp, Reface, DeepBrain, DeepFaceLab have made deepfake creation rather accessible to the public. As AI continues to evolve, deepfakes are becoming more realistic and rather challenging to detect, raising concerns about their potential misuse (Naitali et al., 2023). Major deepfake incidents increased by 257% from 2023 to 2024. In 2025 already, deepfake-driven fraud has resulted in $200 million in financial losses, while the use of deepfakes to impersonate public figures has led to total losses of about $350 million (securitymagazine, 2025).

2.2.2 The cyber defense perspective

As cybercrimes become more widespread and sophisticated, it is crucial that the cybersecurity field enhances its efficiency to combat these advanced threats. AI technology is a key element in addressing both current and future cybersecurity challenges, providing innovative solutions to stay ahead of cybercriminals. Due to the versatile nature of AI in counteracting cyber threats, its application varies across different sectors, each employing unique strategies tailored to their specific security needs (Khan et al., 2024). AI-powered cybersecurity tools proactively detect threats through predictive analysis, using ML to anticipate and mitigate potential cyber threats effectively (Mamidi, 2024).

AI has significantly enhanced security operations through its ability to efficiently analyze vast amounts of data and automate responses to incidents (Mamidi, 2024). In the future, AI might be integrated with blockchain, quantum computing, and other IoT systems, benefiting the cybersecurity field by enhancing protection for decentralized systems and improving data validation speeds. ML algorithms can be used to analyze patterns and deviations in network traffic, helping to prevent cybercrimes before they occur. AI-powered cyber defense tools have self-learning capabilities that enable them to detect and prevent cyber threats without human intervention (Khan et al., 2024).

AI-powered cyber defense approaches are needed to continually work towards the prevention of AI-powered attacks. One such example are the text analysis techniques explained by Eza and Shamir (2024) which can be employed to identify spear phishing attacks. Another example to guard against deepfake manipulations is the use of advanced ML models to analyze the visual inconsistencies typical of deepfakes and employing robust datasets to train these models more effectively (Naitali et al., 2023). Nonetheless, increasing public awareness through cybersecurity training can help individuals spot malicious deepfake content and such training may focus on recognizing signs such as lip desynchronization, jerky eye and body movements, and visual inconsistencies (McBride, 2024).

Limitations of AI-powered cyber defense include false positives alarms which are a significant concern impacting the efficiency of these defense tools. Additionally, the complex nature of AI models often makes them difficult to understand, posing challenges in troubleshooting and managing automated decisions effectively (Luna, 2024). The resources required to implement and maintain AI systems poses another significant barrier, particularly for smaller organizations, due to the high computational power and infrastructure needed. Integrating AI into existing cybersecurity frameworks can also be complicated and costly, requiring substantial modifications. (Luna, 2024)

2.3 Anomaly and malware detection methods

In this chapter traditional and AI/ML based ML methods and their functions are discussed, including multiple ML approaches. Traditional anomaly detection methods are not included in the performance analysis; therefore, only the signature-based malware detection method has been briefly discussed.

2.3.1 Tradtional malware detection methods

Signature-based malware detection is a traditional method that is not primarily based on AI/ML technologies. It operates by matching known malware signatures or patterns against files or network traffics to identify and analyze potential malwares. It is crucial to discuss how that traditional malware detection systems work, highlighting their strengths and limitations in the context of modern cybersecurity challenges. Signature-based detection primarily relies on predefined patterns and known threats to identify malware in systems or network traffic. This method uses a database of known malware signatures to scan and compare against files and data flows, making it effective for detecting previously identified threats (Rehman et al., 2024). Because of this, the method is capable of detecting malware with different patterns from various applications, but it requires constant updates to the database of predefined malware signatures to remain effective. Due to the rapidly evolving nature of malware families, signature-based detection can be less effective, as it struggles to identify new and unknown variants that do not match with existing signatures (Souri & Hosseini, 2018).

Many signature-based detection systems rely on binary matches, which poses a challenge in adapting to minor yet strategically significant modifications made by malware or ransomware actors. Automated updates of malware patterns have been used to address the limitations of signature-based systems, but the evolving nature of malware continues to present challenges. Additionally, more complex signatures based on hash values have improved detection efficiency with more granular detection capabilities. However, these advanced systems require more computing power and processing time, which may not be suitable for environments with limited resources (LaRocque et al., 2024). A signature in malware detection is a unique identifier that encapsulates the structure of a program, allowing each malware type to be uniquely identified.

The signature-based malware detection method is widely used for identifying common and known vulnerabilities (Aslan & Samet, 2020).

The signature-based detection method functions through four main processes, as illustrated in Figure 3. First, data from network traffic and other sources are collected using either native or third-party monitoring tools. Next, relevant features are extracted from this data, from which unique signatures are generated. These signatures are then compared against a database of known threat signatures to determine whether the traffic represents actual cyber threats or benign (Faruk et al., 2022).

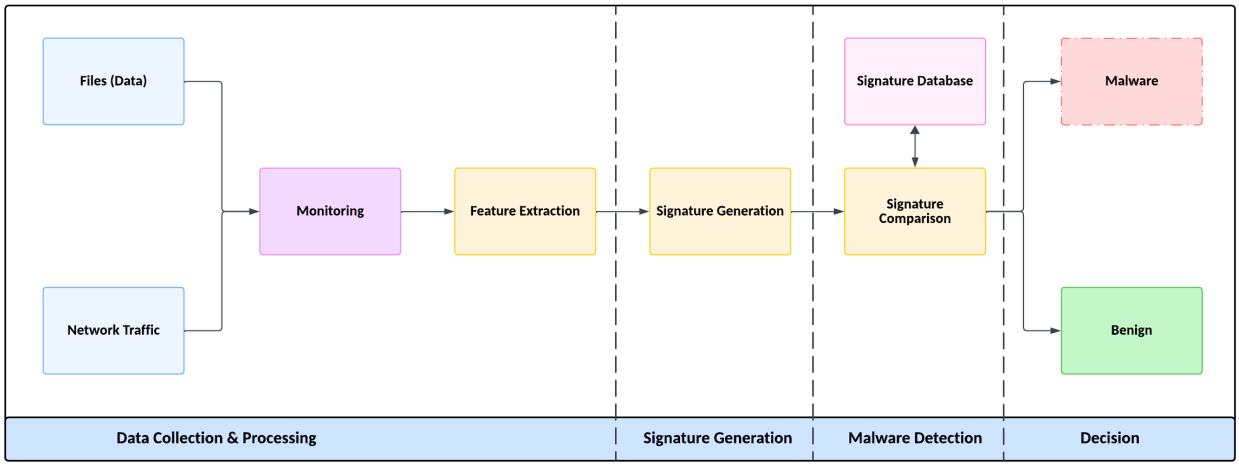

Figure 3. Signature based malware detection process

Signature generation is a core component of the signature-based malware detection method during which a generation engine extracts features from the collected data. It then creates signatures based on these features and stores them in a database. To identify malware, a sample is compared against these existing signatures. Based on this comparison, the sample is either marked as malware or benign. Several signature generation techniques are available for creating signatures, such as string scanning, top-and-tail scanning, entry point scanning, and integrity checking (Aslan & Samet, 2020).

String scanning involves comparing the byte sequences of an analyzed file with the previously uploaded files in the database. This method has been widely used for many years because it can effectively identify different signatures from the same malware family (Aslan & Samet, 2020).

Top-and-Tail scanning targets malware that attaches itself to the beginning and end of files. In this process, signatures are created from the top and tail points of the files, focusing on these areas instead of scanning the entire file (Aslan & Samet, 2020).

Entry point scanning employs a targeted approach towards malware which often alters the entry point of a program, ensuring that the malicious code executes before the actual intended code. To combat this, the entry point scanning method extracts signatures from the sequence at the program's entry point. This targeted approach allows for the detection of this specific type of malware (Aslan & Samet, 2020).

Integrity checking (Hash Signature) involves generating a cryptographic checksum such as MD5 or SHA-256, for each file within a system at regular intervals. This method is used to detect any changes to files that might be caused by malware. By comparing the current checksums to those from a previous state, any discrepancies can indicate unauthorized modifications, often signaling a malware infection (Aslan & Samet, 2020).

2.3.2 AI and ML based malware detection methods

ML is an advanced tool that can be used for anomaly detection in various environments. There are several ML models available to identify anomalies effectively. The following chapter presents three main models in ML: supervised learning, unsupervised learning, and deep learning (Wang et al., 2021).

Supervised Learning involves training models with labelled data, where each data point is classified as either anomalous or benign (Chukwuemeka Nwachukwu et al., 2024). The model learns from this past data and uses it as a reference to make accurate decisions on new, unseen inputs (Wang et al., 2021). The algorithms used in this learning method include Decision Trees (DT), Support Vector Machines (SVMs), and Neural Networks (Chukwuemeka Nwachukwu et al.,

2024). Training the model is crucial in this method, as high-quality training is essential to achieve strong performance. However, the best results depend not only on the training process but also on building a reliable and trustworthy predictor. In the Supervised Learning model, the system is first trained, and once the training is complete, it can predict outputs for further actions (Wang et al., 2021). The performance of Supervised Learning techniques relies heavily on the availability of labelled training data. In some cases, labels are generated naturally, while in others, obtaining accurate labels may require manual verification (Das et al., 2024). Models commonly used to classify data in Supervised Learning methods are, *Support Vector Classifier (SVC)* which efficiently performs non-linear classifications while requiring less computational power. *Random Forest Classifier (RFC)* is based on the concept of DTs and is considered an ensemble model because it uses a collection of classifiers. RFC selects the best parameters from DTs, where the chosen parameter is based on all available features. *k-Nearest Neighbor (KNN)* classifies data based on the distance between samples, assuming that samples with similar classifications are closer in proximity. It uses this assumption to classify new samples based on the k closest neighbors (Kimmell et al., 2021). The efficiency of supervised models depends on the size and quality of the labelled datasets—larger and higher-quality datasets lead to more accurate results. In cloud computing, these models are commonly used to identify known threats, such as unusual login attempts and abnormal network traffic. However, supervised learning models often struggle to detect new or unknown anomalies, which can be challenging in cloud environments due to their dynamic and constantly evolving nature (Chukwuemeka Nwachukwu et al., 2024).

Unsupervised Learning methods primarily work without labelled datasets. Unlike Supervised Learning, they detect anomalies by identifying inherent patterns and structures in the data, without relying on labelled samples or a prior training process (Goswami, 2024). Unsupervised Learning methods operate without training, as the model functions independently by analyzing data patterns. Since there is no labelled data for reference, the performance is difficult to evaluate. Therefore, some security experts manually verify the results against existing labelled data to assess their reliability (Wang et al., 2021). Cybersecurity systems in the modern era face an enormous number of new threats, along with a shortage of cybersecurity experts. In such cases, unsupervised ML methods are highly effective, as they require minimal supervision and do not rely on labelled datasets (Das et al., 2024). These methods can be used to recognize behavior-based anomalies and identify unknown attack patterns, which are essential to fight against Zero-Day attacks. For these purposes, the algorithms used in unsupervised methods are both as non-meta

learners and as base-level learners within meta-learning approaches (Yee Por et al., 2024). Multiple algorithms are used in unsupervised learning techniques, each serving a slightly different purpose in a cybersecurity environment. The commonly used method is k-means from the clustering algorithm family, which helps identify anomalies in network traffic and abnormal user behavior, may increase the attack surface or lead to potential security breaches. Principal Component Analysis (PCA) is another common algorithm in unsupervised learning. It reduces high dimensional data into fewer variables, making it suitable for cloud environments where large volumes of traffic are involved. After applying PCA, anomalies can be detected by observing deviations of major patterns (Chukwuemeka Nwachukwu et al., 2024).

Deep Learning is one of the most advanced and complex branches of ML. Its learning process involves multiple layers of neurons, allowing it to effectively handle and analyze vast amounts of data. Due to this capability, it can be applied efficiently across various fields (Abdallah et al., 2024). The fundamental technique behind deep learning is Artificial Neural Networks, which function similarly to synapses in the human brain. ANNs use weighted connections between neurons to process and transmit information through the network (Tayyab et al., 2022). The deep learning domain can be categorized into multiple sub-domains, as illustrated in Figure 4 (Tayyab et al., 2022). In this study, to understand the functions and performance of the deep learning methods, such as Long Short-Term Memory Networks (LSTM) and Autoencoders are selected and reviewed.

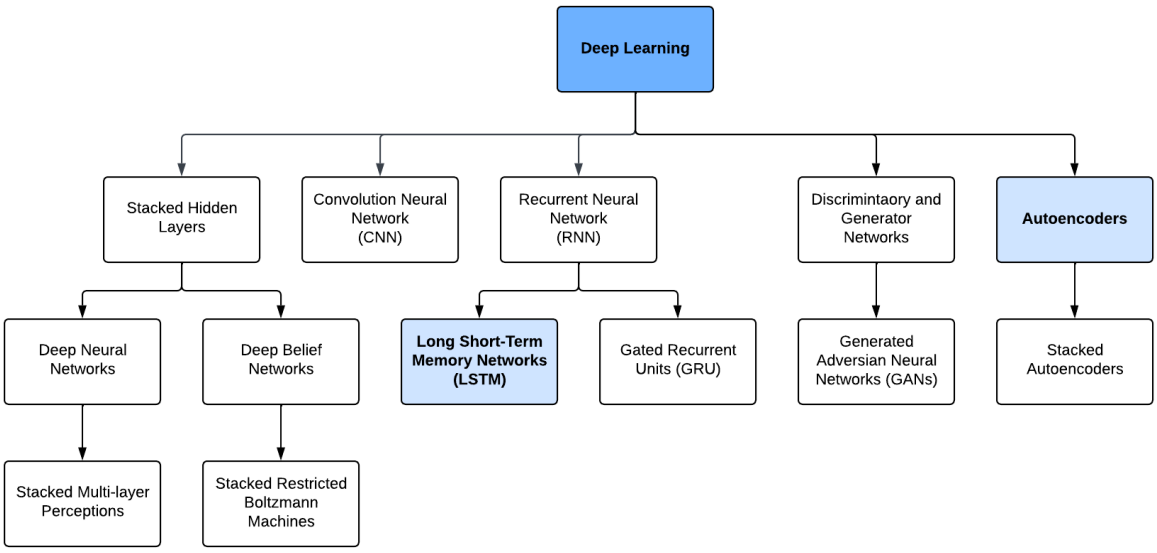

Figure 4. Deep learning sub-domains

The number of layers in neural networks is the primary factor that distinguishes different techniques within deep learning domains. These networks have the capability to automatically learn features directly from data. During feature engineering, large amounts of sample data are fed into the neural network algorithms, enabling them to identify and learn the most relevant features. Once this learning is complete, the models can classify, identify, or generate data based on the patterns they have recognized (Tayyab et al., 2022). Due to the growing dependence on cloud services, cloud security defense systems need to effectively handle modern threats. Deep learning techniques, such as CNNs, LSTMs, and Autoencoders, significantly contribute to addressing these security challenges by offering intelligent threat detection capabilities. These methods are commonly utilized to detect malware, recognize anomalies in network traffic and user behavior, and strengthen intrusion detection systems. They enable real-time monitoring and predictive analytics (Alzoubi et al., 2024).

Key performance metrics of AI and ML-based malware detection methods

AI-based anomaly and malware detection methods are widely used in cloud environments. However, due to the large number of available techniques, selecting the most suitable method for specific requirements can be challenging. Key performance metrics play an important role in evaluating the effectiveness of each method and help to identify the best option for a given need.

Accuracy reflects the method's ability to correctly predict samples. It indicates the percentage of correctly classified instances, such as the proportion of labelled samples (e.g., diseased or control) that the model identifies accurately (Miller et al., 2024).

Precision evaluates the method's ability to minimize false positives, it indicates the percentage of identified anomalies behaviors that are actually anomalous (Miller et al., 2024).

Recall reflects the method's ability to correctly predicted positive samples. For example, it indicates the percentage of true anomalies behaviors (Miller et al., 2024).

F1 Score is the harmonic mean of precision and recall. For example, in a malware detection algorithm, it helps minimize both missed alerts (false negatives) and false alerts (false positives) (Miller et al., 2024).

Adjusted Rand Index (ARI) is used to measure the similarity between the predicted clusters and the ground truth (Miller et al., 2024). A score close to 0 indicates random clustering, while a score of 1 represents perfect alignment with the actual clusters (Shi et al., 2022).

Silhouette Index (SI) metric compares the similarity of data points within the same cluster to the similarity between different clusters. This index helps in identifying the best model for detecting new malware subsets (Miller et al., 2024).

Receiver Operating Characteristic (ROC) curve and *Area Under the Curve (AUC)* are important evaluation metrics. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR). The AUC value indicates how well a ML model can distinguish between different classes, such as identifying malware or its variants. A higher AUC value reflects better model performance (Ahmed et al., 2025; Miller et al., 2024).

Gini impurity is a metric primarily used in DT and random forest machine learning algorithms. It can also be used in the feature selection process to measure the chances of incorrectly classified data (Disha & Waheed, 2022).

Variance reduction method is used in DT models to minimize the variance in the leaf nodes after a split. It is commonly used in regression trees (Mienye & Jere, 2024).

Maximum tree depth is a metric in the DT algorithm. The tree continues to grow until a stopping criterion is met, it can be a predefined maximum depth or when all nodes become pure (Mienye & Jere, 2024).

Minimum sample split is a hyperparameter in DT algorithm that defines the minimum number of sample required to split the internal node (Qiang et al., 2024).

3 Research Aims and Methodolgy

3.1 Aims and objectives

The overall aim of this research is to study AI based methods in anomaly detection. A key focus is on understanding the ML algorithms commonly used in cybersecurity tools, their function, performance in anomaly detection, and utilization in cloud environment and real-world cybersecurity systems.

Specific objectives are

- What ML methods are available in the cybersecurity tools?
- How ML-based anomaly detection methods work?
- How ML-based anomaly detection methods perform?
- Understanding the application and utilization of ML-based anomaly detection methods

3.2 Research methods

A mixed research methodology, combining an extensive review of available literature and comparative analysis of various cyber defense mechanisms against selected cyberattacks has been employed in this thesis. The **narrative review methodology** systematically gathers, summarizes, and critically discusses existing research on AI-based anomaly and malware detection methods in cloud security. The focus is on exploring supervised, unsupervised, and deep learning approaches, highlighting their techniques, performance, and applicability allowing a comprehensive yet flexible review of diverse studies to identify trends, gaps, and future research opportunities. The **comparative analysis approach** is used to compare various ML based anomaly detection methods against selected cybercrimes to evaluate their effectiveness based on which recommendations are provided.

3.2.1 Data collection methods

This review is based on available literature with the data primarily sourced from academic and trade journals. Additionally, internet-based recourses have been used for supporting data. Each article was assessed based on the relevance to the research questions and the time of publication. Figure 5 illustrates the data collection methods utilized in this study.

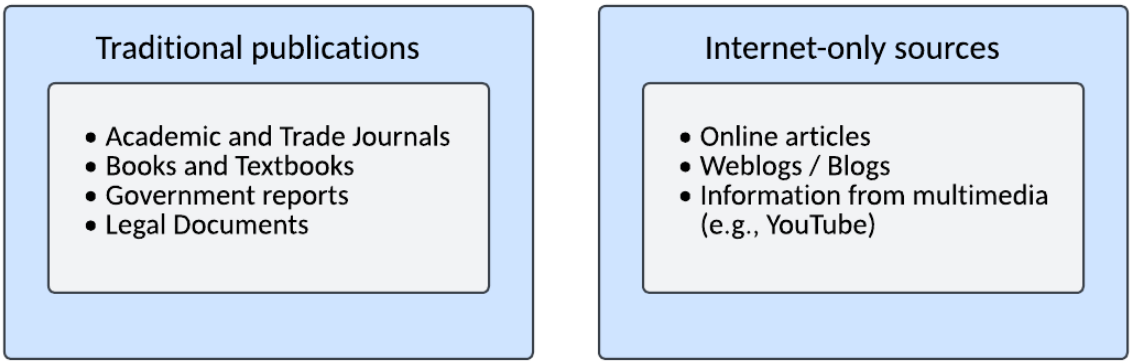

Figure 5. Data collection methods

3.2.2 Search strategy and inclusion criteria

Relevant literature including research articles, conference proceedings, e-books, reports were searched using the Google Scholar and the Janet database (JAMK's library portal which hosts e-books, e-journals and multiple databases; accessible at https://janet.finna.fi/) using keywords such as ”artificial intelligence" OR AI OR “machine learning”, supervised OR unsupervised OR "deep learning” cybersecur* OR "cyber defen*” OR detect* AND cyberattack* OR cybercrim* OR phising OR malware. The search was restricted to these two databases given the narrative approach of the review.

3.3 Research reliability and ethics

This research is based on reliable recourses from high-impact peer-reviewed scientific journals, official online cybersecurity articles and reports, opinions from cybersecurity experts and reputed cybersecurity textbooks.

Data confidentiality, consistency, and quality are ensured since this thesis only analyzes and evaluates the existing works, thereby not handling any personal data. The research data and materials is handled and stored using JAMK Office 365 apps, including MS Word, MS Excel, OneNote and OneDrive and an encrypted personal laptop. All data stored in the personal laptop will be securely destroyed after thesis is completed.

AI assitance tools such as OpenAI's ChatGPT and Microsoft Co-Pilot, were utilized during the thesis writing process primarily for language checks and proofreading. In some cases, these tools were also used to translate content from other languages, such as German and Finnish, into English. Additionally, AI tools helped with advanced searches to find specific research articles that included certain keywords. However, AI tools were not used to generate any content directly; all writing and analysis were done by the author.

Ethical considerations were strictly taken into account following the Finnish code of conduct for research integrity and procedures (TENK, 2023). Research is conducted in an ethical manner, with no intentional negative criticism on any parties and ensuring that findings and recommendations are free from personal and commercial bias. While evaluating multiple AI-based cyber defense systems, this work may highlight system limitation; however, the intention is to identify the issues and provide constructive recommendations. Any limitations discussed in this work already been acknowledged in the other research works. The findings of this review are meant to inform and enhance the cyber defense and constructive approaches and by no means are meant to benefit any ill-intended or criminal entity or activity in anyway.

4 Results

Altogether 2010 hits were retrieved, and the results are based on selected 90 research articles (based on the relevance to the research questions, the time of publication, and the narrative approach of this review) highlighting fundamental concepts behind ML techniques, performance metrics, and their application within the domain of cloud security. Selected AI-based anomaly and malware detection techniques categorized under supervised learning, unsupervised learning, and deep learning methods, are illustrated in Figure 6.

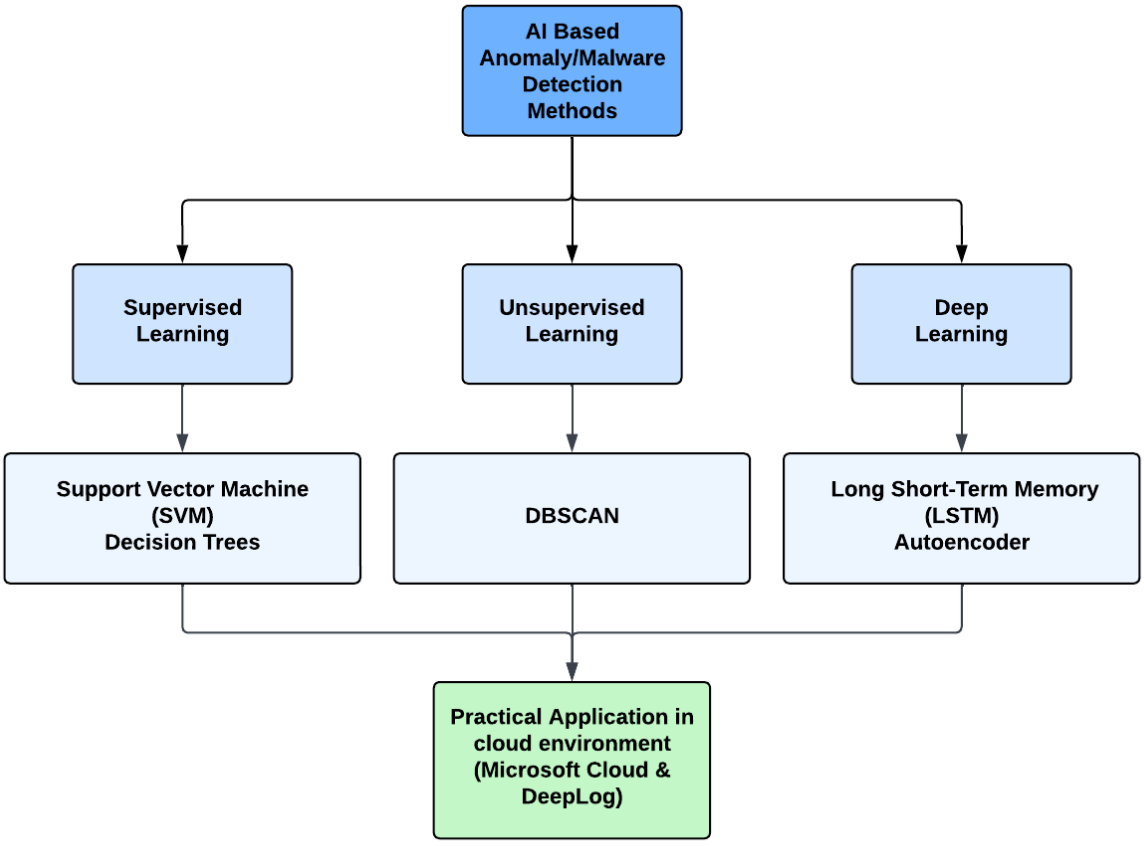

Figure 6. Selected AI techniques for review

4.1 Review of Supervised Learning Techniques

4.1.1 Support Vector Machines (SVMs)

The basic concept of SVM originates from neural network or can be viewed as a mathematical extension of neural network. SVM classifies data based on features by identifying an optimal hyperplane that maximizes the margin between two classes, such as benign and anomalous. The key decision-making data points, known as support vectors, lie closest to the decision boundary and influence its position. By increasing this margin, SVM techniques can achieve more accurate classification and reduce the probability of misclassification (Chandra & Bedi, 2021).

These classic binary classification algorithms are highly effective and robust, making them suitable for addressing real-world problems where distinguishing between normal and anomalous behavior is particularly challenging, such as in cloud environments (Wang et al., 2024). Figure 7 illustrates how the optimal hyperplane, with its maximum margin, helps to distinguish the anomalies and benign samples based on the provided features (Mustafa Majid et al., 2023).

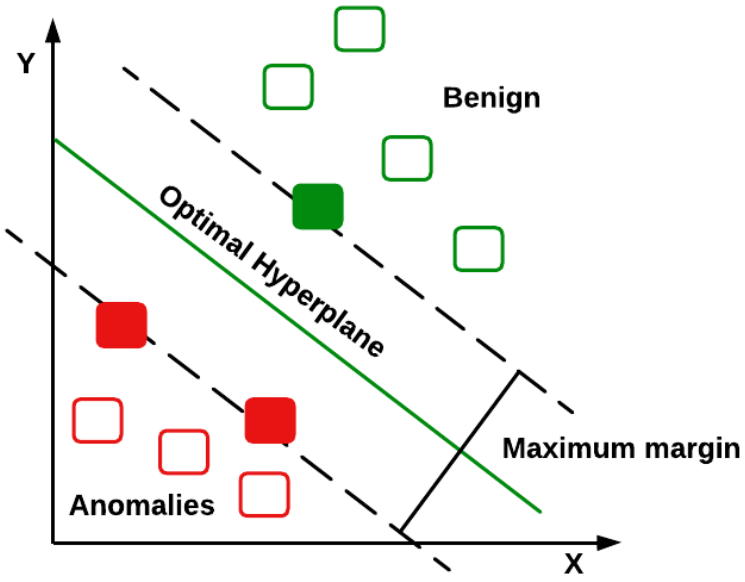

Figure 7. SVM classification concept (Mustafa Majid et al., 2023).

Figure 8 illustrates the basic workflow of the SVM anomaly detection technique. Initially, data is collected from the target environment, such as network traffic, virtual machines, or storage devices, using either native methods or third-party log collection tools. The data then goes under a dimension reduction process, where high-dimensional datasets are transformed into low dimensional datasets. It helps to reduce the volume of features, making it easier and more effective to analyze (Jia et al., 2022).

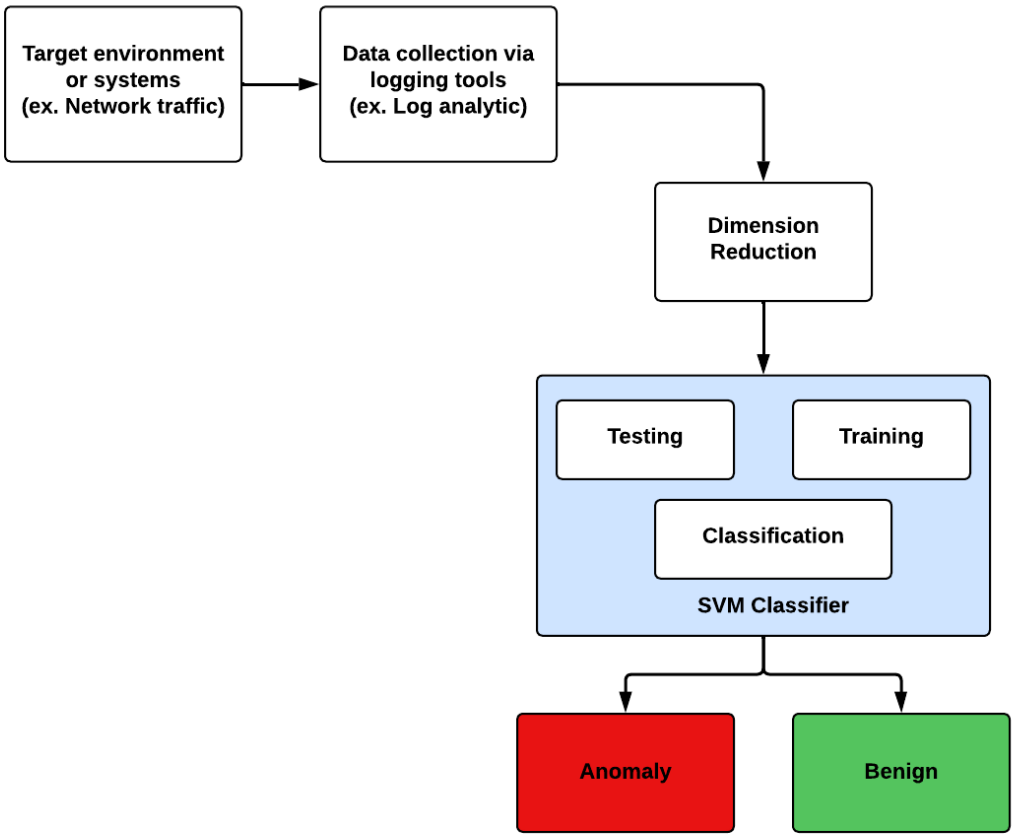

Figure 8. SVM work flow

Within the SVM classifier, the process consists of three key phases: training, testing, and classification. During the training phase, data from target sources is fed into the SVM, while in the testing phase, non-target data is introduced to refine the model (Abbas & Almhanna, 2021). Using these datasets, the SVM classifier distinguishes between anomalous and benign instances based on its learned principles.

Performance review of SVM

The common key performance metrics accuracy, precision, recall, and F1-score have been used in this study to evaluate the performance of supervised learning methods, based on practical experiments reported in various research studies. A brief overview of these metrics has been provided earlier in Chapter 2.3.2.

Baawi et al. (2025) have evaluated the SVM based malware detection technique using the Mutual Information (MI) method for feature selection to achieve the highest classification accuracy. They also tested the model with and without feature selection. For their experiment, they utilized the Meraz'18 dataset, which contains both benign and malicious samples. The SVM was trained on the preprocessed dataset following classic SVM classification principles, where a hyperplane is defined to separate benign and anomalous samples. The authors experimented with different regularization parameter values (C) to balance the trade-off between margin maximization and classification error, using C values of 0.1, 1, and 20. In this study, the highest regularization parameter (C = 20) was selected for comparison with other experiments, and the results are presented in Table 1. This experimental study shows that the SVM classifier, when trained using standard methods and proper parameter tuning, delivers robust performance with an accuracy of 95%.

Alheeti et al. (2023) have evaluated SVM algorithms for intrusion detection systems (IDS) in a cloud environment using the Cloud Intrusion Detection Dataset (CIDD), which includes both training and testing data. During the training phase, the dataset was split into 60% for training and 40% for testing. The authors assessed the model using common key performance metrics and conducted three rounds of experiments. On average, the model achieved 99% accuracy for normal behaviors and 95% for abnormal behaviors, and the results are presented in Table 1. Overall, the study concluded that an SVM-based IDS can effectively classify benign and anomalous behaviors in cloud environments.

Wang at al. (2022) have assessed SVM algorithms for anomaly detection in cloud logs alongside other ML techniques. Raw logs were collected from the cloud environment and processed using log parsing tools to separate log entries into invariant and variant components. Features were

then extracted from the parsed logs. Several traditional supervised learning models, as well as ensemble learning methods, were developed to classify anomalies and benign instances. Experiments were carried out on datasets of different sizes 2K, 100K, and 600K log records using the standard SVM anomaly detection process. The results demonstrated that smaller datasets achieved higher accuracy, while larger datasets provided better recall and F1-Score, and the results are presented in Table 1. The study concluded that although recall was lower, the higher accuracy in smaller datasets enhances the reliability of real-time anomaly detection.

Study	**Key Performance Metrics (%)**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
Baawi et al. (2025)	95	95	95	94.9
Alheeti et al. (2023)	99	96	92	94
Wang at al. (2022)	98.55	N/A	43.31	60.18

Table 1. SVM expriments results

As per the reviewed studies, SVM models demonstrate promising accuracy in cloud environments across various datasets and testing scenarios. The volume of data is a key factor influencing the performance of SVM models, as it impacts the accuracy of malware detection in cloud environments. However, a limitation of SVM models is their reliance on labelled input datasets for anomaly detection, which can pose challenges in identifying unknown or emerging malware threats.

4.1.2 Decision Tree (DT)

Decision Tree (DT) is a white-box classification model that organizes its decision-making process in a tree-like structure, where internal nodes represent test conditions based on features, and leaf nodes indicate class labels. Its core strengths include simplicity, interpretability, and a built-in feature selection mechanism, which helps during classification (Rivera-Lopez et al., 2022).

The process starts from the root node, which represents the entire dataset, and continues recursively until the model reaches a pure node or the maximum depth of the tree, so the leaf node labelled as benign or malware, its shown in Figure 9. It splits the dataset based on the best feature and threshold, using criteria such as information gain, Gini impurity, or variance reduction (Mienye & Jere, 2024). DT require no prior domain knowledge and utilize the concept of data entropy, making them both efficient and interpretable (Gorment et al., 2023).

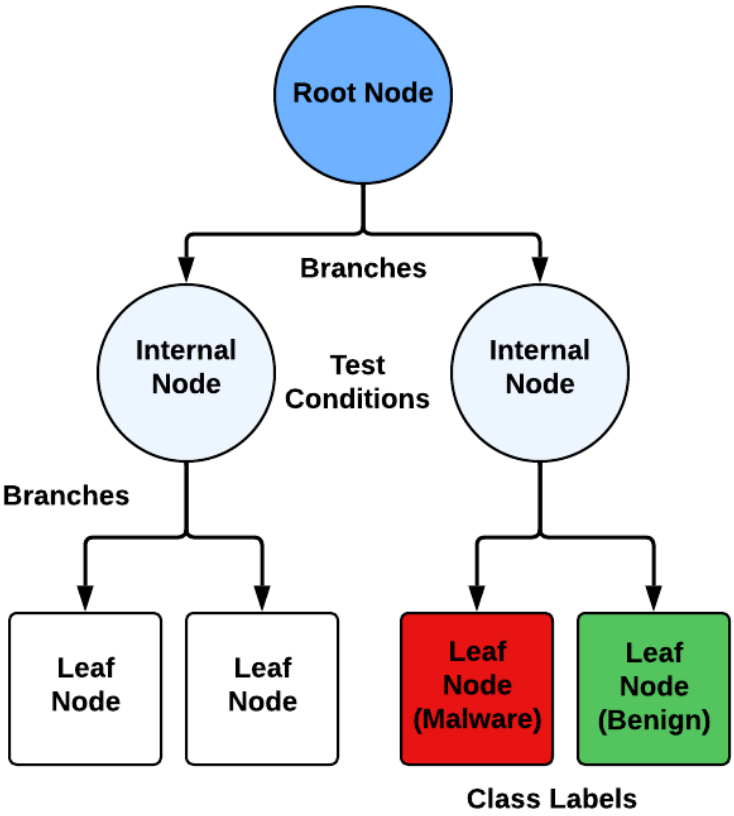

Figure 9. Basic DT stucture (Mienye & Jere, 2024)

Performance review of DT

Shahzad et al. (2022) have evaluated the DT model using a Cloud-based Anomaly Detection (CAD) approach to identify anomalies in cloud environments. The evaluation was divided into two parts: binary anomaly classification and multiclass anomaly categorization. They used the UNSW-NB15 dataset, which includes both benign and anomalous network traffic records. The dataset was split into an 80-20 ratio for training and testing purposes. In addition to the DT model, other binary malware detection models were tested for comparison. The results showed that the DT model achieved a high success rate compared to other binary models, although its accuracy was slightly lower than deep learning models, such as CNN-LSTM, as shown in Table 2. The authors concluded that cloud-based ML models, including the DT, can be effectively used to build cyber defense systems against global cybercrimes such as spam, anomalies, and network attacks.

Farzaan et al. (2024) have assessed the DT model with an AI-enabled cyber incident response system designed for cloud environments, focusing on network traffic, intrusion detection, and malware analysis. For their evaluation, they utilized benchmark datasets including UNSW-NB15, NSL-KDD, and CIC-IDS-2017, collected from cloud-based environments. Irrelevant data was removed to enhance the efficiency of the experiments. The DT classifier was implemented with standard parameters, such as Gini impurity, maximum tree depth, and a minimum sample split of 2. In the network traffic classification experiments, the DT model's accuracy varied significantly across different datasets. In the malware anomaly detection, it achieved an accuracy of around 88%, as shown in Table 2. However, the recall results revealed that the DT model had limited capability to identify most anomalous instances. Overall, the authors concluded that AI/ML models demonstrate strong capabilities in cyber threat defense systems, and deploying such systems on cloud platforms like Google Cloud and Azure Cloud could enhance scalability and versatility.

Kumar et al. (2023) have evaluated the DT model was evaluated against Zero-Day malware attacks using a dataset of Portable Executable (PE) files from the Meraz'18 database, which includes both malware and benign samples. The dataset was pre-processed to ensure an equal number of benign and malware files. Alongside the DT model, the authors analyzed other classic ML malware detection models. Similar to Shahzad et al. (2022), the dataset was split into an 80:20 ratio for training and testing, and the models were evaluated using standard key performance metrics. The

experimental results showed that the DT model achieved a high accuracy of 98.91% and a low false positive rate (FPR) of 1.34%, which is particularly important for Cloud Intrusion Detection Systems (Cloud IDS), as shown in Table 2. Although the DT model delivered strong results, the authors noted that Random Forest (RF) outperformed it overall, offering a lower FPR. They concluded that ML models are highly effective for addressing Zero-Day malware attacks.

Study	**Key Performance Metrics (%)**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
Shahzad et al. (2022)	91.86	91.65	99.78	95.54
Farzaan et al. (2024)	88.26	87.25	70.63	78.07
Kumar et al. (2023)	98.91	99	99	99

Table 2. DT expriments results

Overall, the performance of the DT model delivers promising results. However, according to Farzaan et al. (2024), experimental studies reveal that the DT model struggles to identify the majority of anomalous instances. Additionally Kumar et al. (2023) demonstrated that the Random Forest (RF) malware detection model outperforms the DT model, offering better accuracy and a lower false positive rate.

4.2 Review of Unsupervised Learning Techniques

4.2.1 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a popular clustering algorithm often used in cyber defense systems to identify anomalous behavior. It defines clusters based on areas of high data density, selecting dense regions as cluster centers. This method is particularly effective in detecting outliers or unusual patterns that may indicate potential threats (Wang et al., 2022).

The core point of a cluster is determined by a specific radius (Eps) and a minimum number of points (MinPts). The classification of core points, border points, and noise points is based on the number of points (MinPts) within the defined radius (Eps) (Singh et al., 2022) its shown in Figure 10. Core points are defined as points that have at least k number of neighboring points within a radius r. A border point is any point that has one or more core points within the radius r but does not meet the core point condition itself. Points that do not satisfy either condition are labelled as noise (Singh et al., 2022).

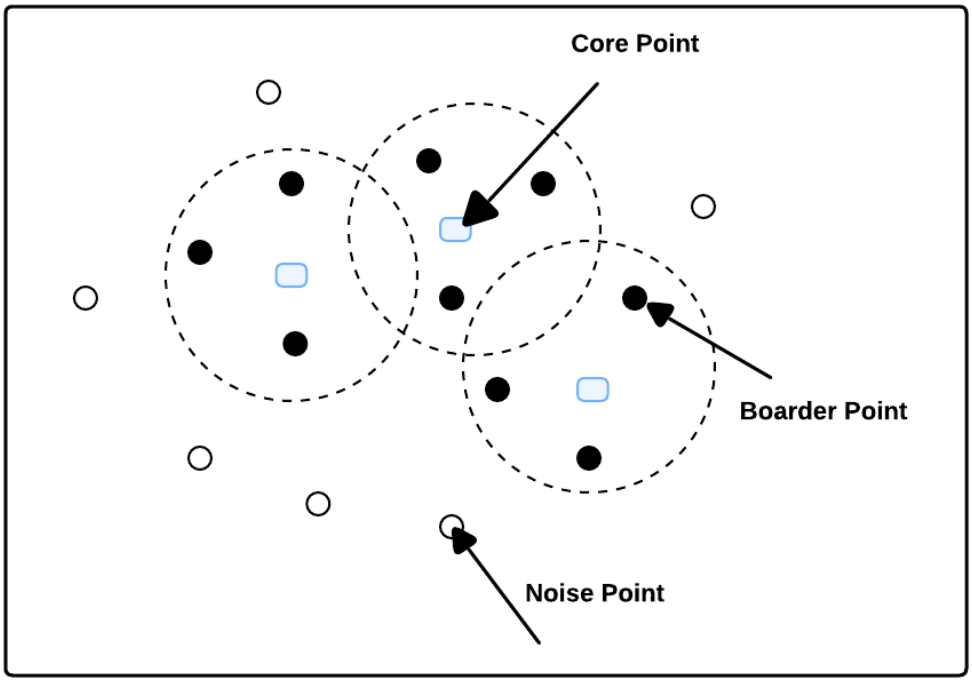

Figure 10. DBSCAN stucture (Singh et al., 2022)

Performance review of DBSCAN

Kaliyaperumal et al. (2024) evaluated the performance and importance of the DBSCAN malware detection model in their a hybrid unsupervised learning approach. In their experiment, they first used a One-Class SVM (OCSVM) supervised learning model to classify samples as either normal or anomalous. Then, a trained Deep Learning Autoencoder (dAE) model was used to identify benign traffic and reconstruct the dataset with minimal error. After completing the two-stage anomaly detection process, normal traffic was separated, and the unsupervised DBSCAN model was applied to the anomalous data to identify clustering attack patterns and similar malicious traffic. The SECIC-IDS2018 dataset was used for the experiment. The results showed that the standalone DBSCAN model achieved 97% accuracy with 79% precision, while the hybrid model achieved 99% accuracy with 99% precision, as shown in Table 3. The authors concluded that integrating DBSCAN into the hybrid model helps detect and cluster similar attack patterns more accurately.

Pitafi et al. (2022) have proposed an improved DBSCAN (I-DBSCAN) model to address common issues in Intrusion Detection Systems (IDS), such as low detection rates and high false positive rates. They also compared the performance of the improved model against the standard DBSCAN model. In their approach, the base algorithm was refined to better capture dense regions that represent authentic intrusion patterns rather than grouping noise and false intrusions. In addition to DBSCAN, they used subsequent classifiers like SVM, K-NN, and Random Forest to achieve cleaner and more accurate results. The experiments were conducted using the KDD Cup 99 and NSL-KDD Cup 99 datasets. The results showed that the I-DBSCAN model achieved a better detection rate (95.5%) compared to the standard DBSCAN model (83.3%), and the false positive rate (FPR) was also slightly improved. The combination of I-DBSCAN with the K-NN classifier provided excellent accuracy and a lower FPR. The authors concluded that integrating DBSCAN with standard classification methods can significantly enhance the performance of intrusion detection systems.

Lee et al. (2024) have used the DBSCAN model on top of an LSTM model to identify malicious traffic in their experiment. DBSCAN was applied to group the trained models with similar characteristics, where clustering helped to reduce noise and prevent overfitting by identifying border points, thus narrowing the accuracy of the results. The CICDDOS2019 dataset was used for

the experiment. The evaluation was performed under both balanced and imbalanced conditions, with the conditions defined by different Dirichlet distribution (alpha) values. The results showed that the imbalanced approach achieved better accuracy (97%) compared to the balanced approach (95%). The proposed model delivered improved performance, as shown in Table 3. The authors concluded that integrating DBSCAN clustering with deep learning models like LSTM improves the IDS accuracy and helps effectively group similar attack patterns.

Study	**Key Performance Metrics (%)**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
Kaliyaperumal et al. (2024)	99.27	99.48	99.07	98.86
Pitafi et al. (2022)	99.98	-	-	-
Lee et al. (2024)	97.00	95.00	97.00	96.00

Table 3. DBSCAN experiment results

Based on the reviewed experiments from research articles, the unsupervised learning DBSCAN model performs well when combined with supervised learning or deep learning models. Very few studies have been conducted using standalone DBSCAN models in IDS to detect anomalous behaviors in network traffic or cloud environments. The reviewed results indicate that incorporating DBSCAN into ML-based IDS helps effectively group similar attack patterns and separate benign behaviors, improving the overall detection accuracy.

4.3 Review of Deep Learning Techniques

4.3.1 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) deep learning model, often considered an improved version of the RNN. It addresses key challenges faced by RNNs, such as the tendency to forget earlier information when dealing with large inputs and the problem of vanishing gradients during backpropagation (Kimmel et al., 2021).

The LSTM model operates with three main components - input, forget, and output gates, which control the flow of information through the model, as shown in Figure 11. These gates help the model retain necessary information and discard irrelevant data, allowing it to perform more effectively during the learning and detection processes (Kimmel et al., 2021). LSTM models are widely used in Intrusion Detection Systems (IDS) to combat cybercrimes at both global and local levels. Their ability to learn and remember long-term patterns makes them highly effective in detecting complex and evolving cyber threats (Yee Por et al., 2024).

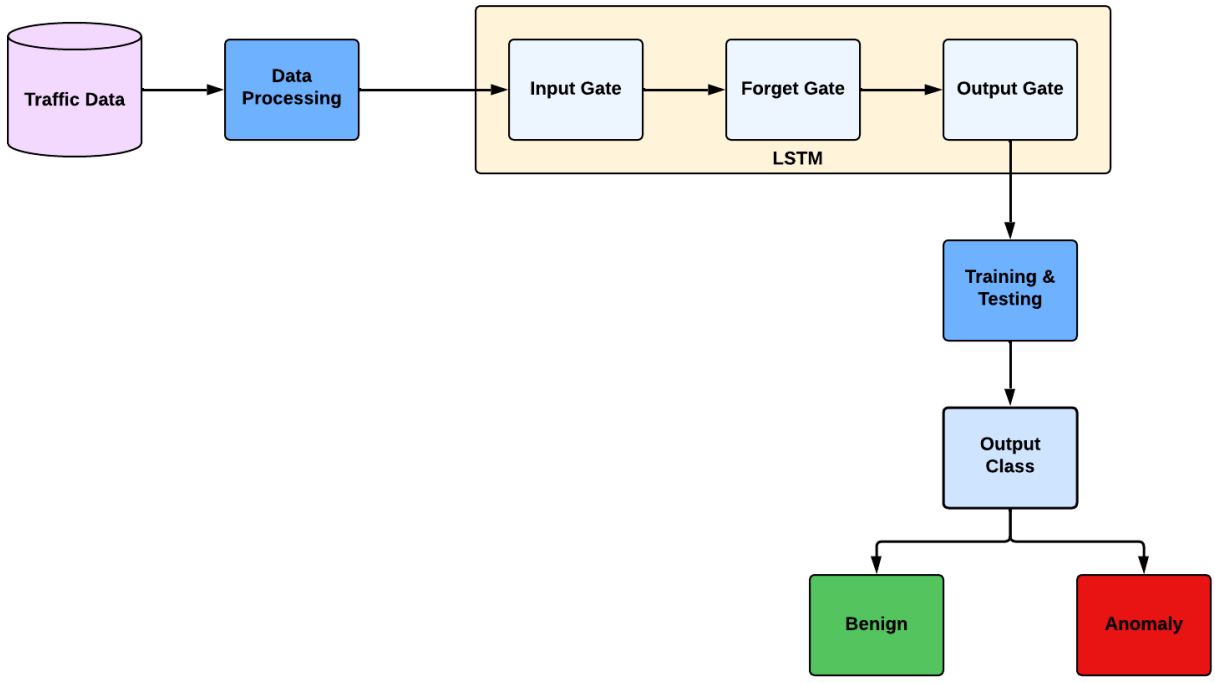

Figure 11. LSTM Process Overview (Shewale et al., 2023)

Performance review of LSTM

Kimmel et al. (2021) evaluated an LSTM-based malware detection model in an OpenStack-based cloud testbed, a popular open-source cloud platform. In their experiment, a control node and multiple compute nodes were used. A dedicated Virtual Machine (VM) injected a malware executable into one of the application servers. Data was collected continuously in two phases: the first 30 minutes represented the benign phase, followed by the malicious phase, where malware was injected at a random time between minutes 30 and 40, continuing up to minute 60. The sample malware used in this experiment was collected from 113 different Linux machines. During the malicious phase, 360 samples were collected per experiment. In the feature extraction process, non-essential features, such as ports, IP addresses, and timestamps, were excluded to maintain privacy and focus on performance. Along with the standard LSTM model, the researchers also evaluated the bidirectional LSTM model. The dataset was split into a 60:20:20 ratio for training, validation, and testing. The experimental results showed that both LSTM and bidirectional LSTM models achieved the same accuracy and precision values, both exceeding 99%. The authors concluded that LSTM models achieved these results with faster training times compared to bidirectional models, and that the input order of features did not im-pact the model's performance, though it did affect the training time.

Ahmed et al. (2025) assessed the LSTM model along with Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNNs) to detect real-time malware in cloud datacenters. The NSL-KDD and UNSW-NB15 publicly available datasets, containing benign and attack traffic in cloud datacenter environments, were used for the study. During feature selection, invalid and duplicate records were removed, and the datasets were split into an 80:20 ratio for training and testing. A comparative evaluation method was used to assess the performance of the models using standard key performance metrics, with a primary focus on accuracy and ROC-AUC to measure how well each model distinguishes between normal and anomalous traffic. The results showed that while the LSTM model achieved slightly lower accuracy than the RNN model, it achieved a higher ROC AUC score of 90% results are shown in Table 4, indicating a better ability to differentiate between normal and anomalous traffic. The authors concluded that the proactive and scalable nature of deep learning-based IDS systems can effectively detect new and known malware in real time, particularly in cloud environments like datacenters where large volumes of data are handled.

Galli et al. (2024) evaluated the performance of the LSTM model with various complex datasets. The experiment utilized three malware dataset samples: the Mal-API-2019 dataset and the Alibaba Cloud Malware dataset, which are multiclass datasets containing different malware types such as Normal, Ransomware, Miner, DDoS, Worm, Virus, Backdoor, Downloader, and Trojan, and an API Call Sequences is a structured dataset containing around 42K malware samples and 1K goodware samples. The datasets were split into a 60:20:20 ratio for training, validation, and testing. The experimental results varied across the datasets. The LSTM model achieved high accuracy (99%) with the API Call Sequences dataset, relatively lower accuracy (83%) with the Alibaba Cloud Malware dataset, and performed poorly with the Mal-API-2019 dataset, achieving only 48% accuracy. The authors concluded that the performance of the LSTM model depends heavily on the complexity of the input data. When the data is well-structured, like API call sequences, the model delivers strong performance, but with complex and noisy datasets like Mal-API-2019, the performance significantly drops. This study highlights that inputs with higher noise or unnecessary information can significantly degrade the LSTM model's performance, particularly when dealing with extremely long sequences.

Study	**Key Performance Metrics (%)**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
Kimmel et al. (2021)	99.61	99.64	99.33	99.48
Ahmed et al. (2025)	89.00	82.00	83.00	82.00
Galli et al. (2024)	99.43	95.69	91.81	93.79

Table 4. LSTM experiment results

According to the reviewed studies, LSTM-based malware detection models perform well in classifying benign and anomalous behaviors. However, the experimental results also indicate that the model's performance drops significantly when handling complex datasets, suggesting that it may be beneficial to use classifiers or feature selection techniques before feeding input to the LSTM. Additionally, the studies show that LSTM models are capable of classifying traffic in real time, making them suitable for use in cloud environments.

4.3.2 Autoencoders

Autoencoders, a subset of deep learning, utilize an unsupervised approach to identify malware by converting input sequences into encoded representations. Deep learning-based autoencoders are widely used across various applications due to their ability to efficiently learn and reconstruct data patterns (Mustafa Majid et al., 2023).

The autoencoder works in three different stages, code, encoder and decoder. The autoencoder operates in three main stages: encoder, code, and decoder. The code layer, also known as the hidden layer, is responsible for encoding the input data into a compressed representation, as shown in Figure 12. The encoder transforms the original input into this code, and the decoder then reconstructs (Faber et al., 2021), its illustrates in Figure 12.

In an ideal case, an autoencoder-based malware detection system is typically trained only with benign traffic data. During operation, incoming inputs are compared against the learned normal patterns, and any significant deviations can be easily identified as attack traffic. This approach also helps in detecting unknown attack patterns by recognizing anomalies that differ from the normal behavior (Torabi et al., 2023).

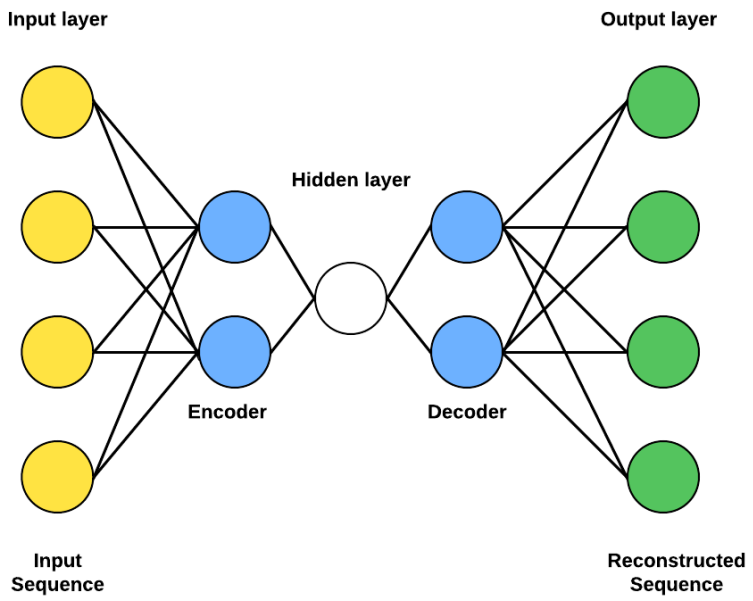

Figure 12. Autoencoder Process Overview (Faber et al., 2021; Mustafa Majid et al., 2023)

Performance review of Autoencoder

Torabi et al. (2023) evaluated the Autoencoder model using a vector reconstructive error method, where the absolute difference between each input feature and its corresponding output was recorded as an error vector. During the training phase, the model was trained only with normal traffic data, and a threshold was defined for each feature based on the maximum error observed across all normal samples. In the detection phase, if the error for any feature exceeded its threshold, the sample was flagged as anomalous. They also trained separate autoencoders for different classes, such as Normal, Attacker, Victim, Unknown, and Suspicious, to compare the performance across classes. A Multi-Class Hierarchical Classification method was used to reduce the false positive rate (FPR).

Torabi et al. (2023) used CIDDS-001 dataset for this experiment, which contains network traffic from cloud computing environment. Initial evaluations using single-class classifiers achieved excellent results for most classes, except for the Suspicious class, which had an accuracy of 69%. However, with multi-class classification, the model achieved outstanding performance across all classes, reaching 100% in all performance metrics and 0% FPR. The authors concluded that a simple autoencoder network architecture can deliver outstanding performance with cloud network traffic data, highlighting its efficiency for cloud environments. They also highlighted that selecting the right classifier is crucial, as the multi-class classifier provided excellent results across all sample types in this experiment.

Xing et al. (2022) assessed the malware detection capability of autoencoders (AE) using a grayscale image approach, where software samples, including both benign and malware files, were represented as grayscale images. Necessary information was extracted from the datasets using bytecode, and a fixed size 2D matrix was used to create a grayscale image for each software sample. Based on the network design, the authors proposed a two-stage deep learning framework consisting of AE-1 and AE-2. AE-1 was trained in an unsupervised manner to capture essential malware features, while AE-2 handled the final classification.

Xing et al. (2022) compared the performance of the AE models with other ML models, including SVM and DT, using the same dataset. The AE-2 models demonstrated excellent results, achieving

an accuracy of 96% with a false positive rate (FPR) of 3.8%. the results are shown in Table 5. The authors concluded that their proposed AE approach has low reconstruction error when comparing malware and benign samples, requires less training time, and offers quick detection, making it highly suitable for IDS applications in cloud environments.

Zhong et al. (2024) proposed the Broad Network-Based Contrastive Autoencoder (BroadCAE) approach to address the limitations of standard autoencoder models. Standard AE models operate with a fixed threshold for online detection, which restricts their adaptability to evolving cloud environments. In the BroadCAE approach, the encoder maps each input sample, whether benign or anomalous, into a latent variable, while also learning the inter-class margin between normal and anomalous samples. The authors evaluated the model's performance using different datasets and compared it with other ML models. The BroadCAE model achieved an overall accuracy of 96% with the MBD dataset, outperforming the other models tested. They concluded that this approach can significantly enhance cloud IDS systems by improving their ability to detect new and unknown malware.

Study	**Key Performance Metrics (%)**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
Torabi et al. (2023)	100	100	100	100
Xing et al. (2022)	96.22	96.14	96.20	96.17
Zhong et al. (2024)	96.11	84.44	81.36	82.87

Table 5. Autoencoder experiment results

4.4 Comparison of performance metrics of ML-based (supervised, unsupervised, and deep learning) models

Detailed evaluations of supervised learning algorithms (Support Vector Machine (SVM) and Decision Tree (DT)), unsupervised learning methods (DBSCAN), and deep learning approaches (LSTM and Autoencoders) based on insights from peer-reviewed research articles, emphasizing key performance metrics such as Accuracy, Precision, Recall, and F1-Score are presented below. The average of these performance metrics are presented below (Figures 13-16).

Figure 13 represents the mean accuracy values of the reviewed models. Overall, all ML models achieved high accuracy rates under various training conditions and datasets, indicating that ML models are good at making predictions and improving overall performance. These models effectively classify anomalous behaviors, with the unsupervised DBSCAN model demonstrating higher accuracy than other models.

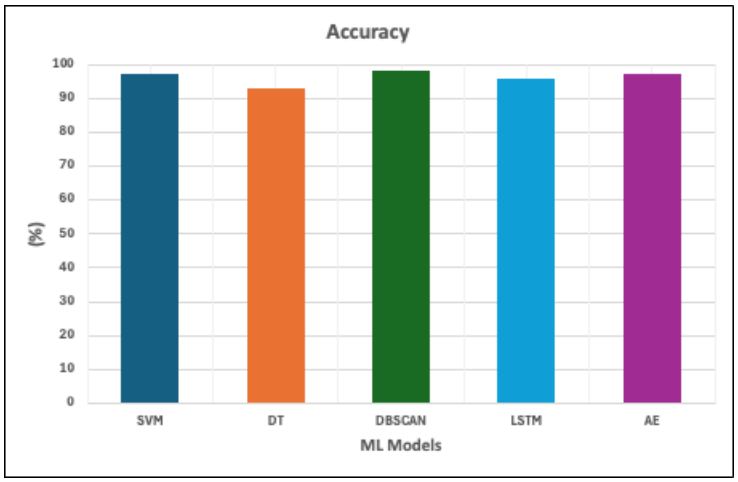

Figure 13. Mean accuracy values of the reviewed models
Each bar represents the average accuracy values of the included models from the selected research articles. SVM: Support Vector Machine; DT: Decision Tree; DBSCAN: Density-Based Spatial Clustering of Applications with Noise; LSTM: Long Short-Term Memory; AE: Autoencoder

The average precision values of the reviewed ML models are shown in Figure 14. Precision reflects the false positive rate (FPR) of the model. The reviewed models performed well in minimizing FPR, which reduces noise in intrusion detection systems. This helps security analysts focus on true positive alerts, potentially increasing the efficiency of the Security Operations Center (SOC). Among the studies, supervised and unsupervised models achieved high precision rates. Notably,

Torabi et al. (2023) reported that the Autoencoder model achieved 0% FPR with a multiclass classifier, although these results are influenced by the classifier and chosen dataset.

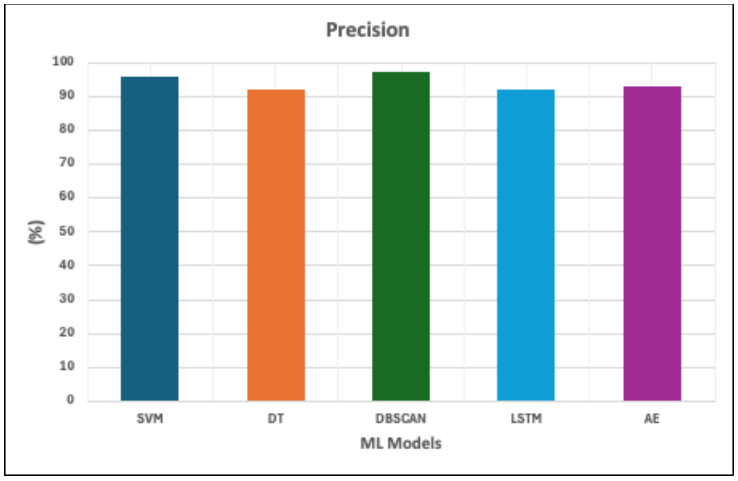

Figure 14. Mean precision values of the reviewed models
Each bar represents the average precision values of the included models from the selected research articles.
SVM: Support Vector Machine; DT: Decision Tree; DBSCAN: Density-Based Spatial Clustering of Applications with Noise; LSTM: Long Short-Term Memory; AE: Autoencoder

Mean recall values of the reviewed models are presented in Figure 15. Supervised models performed lower in recall, particularly the SVM model, which had a mean value of 77%. This results a higher false negative rate, meaning that actual anomalies might be missed. Such outcomes affect the sensitivity of IDS tools, potentially allowing real attacks to go undetected. In contrast, unsupervised and deep learning models achieved high recall rates in several experiments.

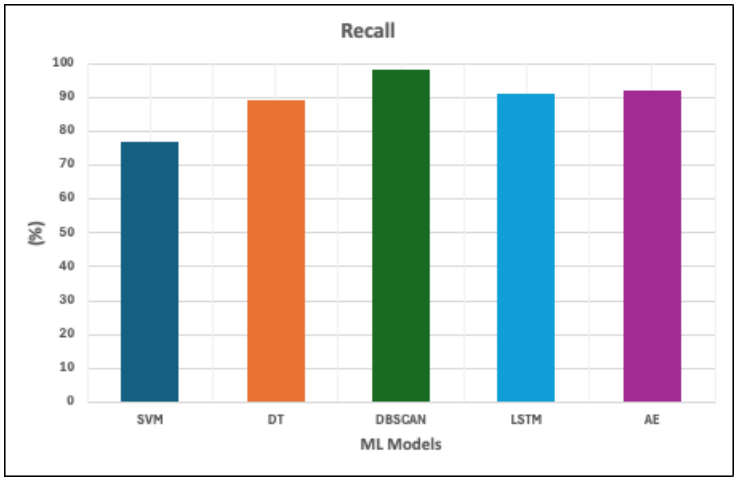

Figure 15. Mean recall values of the reviewed models

Each bar represents the average recall values of the included models from the selected research articles. SVM: Support Vector Machine; DT: Decision Tree; DBSCAN: Density-Based Spatial Clustering of Applications with Noise; LSTM: Long Short-Term Memory; AE: Autoencoder

Mean F1-Score values of the reviewed models are shown in Figure 16. As the harmonic mean of precision and recall, the F1-Score reflects the model's balance between avoiding false positives and false negatives. Lower scores indicate that the model is either missing threats or misclassifying benign activities. Supervised models showed relatively weaker performance, while unsupervised and deep learning models performed better.

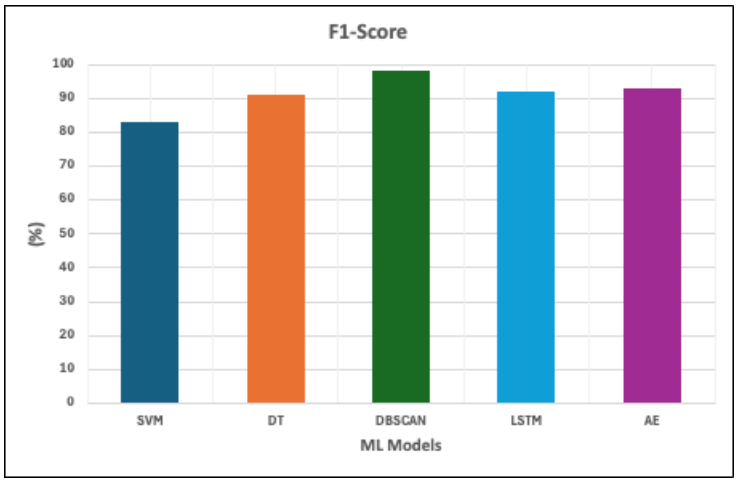

Figure 16. Mean FI-Score values of the reviewed models

Each bar represents the average F1-Score values of the included models from the selected research articles. SVM: Support Vector Machine; DT: Decision Tree; DBSCAN: Density-Based Spatial Clustering of Applications with Noise; LSTM: Long Short-Term Memory; AE: Autoencoder

Among all, the DBSCAN model from unsupervised learning achieved the best overall results across key performance metrics. Deep learning models also delivers strong results, especially when integrated with classifiers. Several studies highlighted that deep learning performance depends heavily on dataset, classifier, and model training.

4.5 Practical applications of AI/ML based cyber defense systems in cloud environment

This chapter presents the practical application of ML algorithms in cyber defense systems within cloud environments. The selected defense tools include solutions from a well-known cloud service provider, Microsoft Cloud, along with one open-source tool (DeepLog). The discussion focuses on the ML algorithms behind these tools and their real-world use cases in detecting and preventing cyber threats.

4.5.1 Microsoft Cloud Tools

Microsoft Cloud, as a public cloud provider, offers a range of native defense systems tailored to various cyber threats. Key tools include Microsoft Sentinel, Defender for Cloud, and Azure ML, which provides customizable anomaly detection solutions to enhance security monitoring and threat prevention.

Microsoft Sentinel is a cloud-native Security Information and Event Management (SIEM) and Security Orchestration Automated Response (SOAR) solution that leverages the Fusion correlation engine. It uses scalable ML algorithms to collect and correlate alerts from multiple sources, transforming them into actionable incidents. While the specific algorithm behind Microsoft Sentinel is not explicitly detailed, it operates in an unsupervised learning manner. Azure Sentinel utilizes analytical rules, such as anomaly detection rules and ML behavior rules, to identify threats across cloud environments (Microsoft, 2024c). ML- based anomaly detection rules establish baselines for normal behavior and detecting deviations from that baseline, with each rule configured using specific parameters and thresholds. Additionally, proprietary ML-based behavioral analytics detect anomalous remote connections, such as those using SSH or RDP, by analyzing factors like IP addresses or geolocations (Microsoft, 2024c).

Microsoft Sentinel also supports a Bring Your Own ML (BYO-ML) framework, allowing security analysts and researchers to integrate their preferred ML models into Sentinel to achieve the desired performance. This integration can be done using Jupyter Notebooks and Azure ML, where custom models are described. The BYO-ML framework enables users to incorporate with both supervised and deep learning models into Sentinel for advanced threat detection and response

(Microsoft, 2023). Microsoft Sentinel's ML based rules are able fight against following cybercrimes (Microsoft, 2023, 2024c)

- Credential theft and account compromise
- Lateral movement
- Insider threats through behavioral anomalies
- Malware spread
- Brute-force attacks and remote access anomalies
- Anomalous data access patterns

Microsoft Defender for Cloud is a comprehensive cloud-native application protection platform (CNAPP) that delivers unified security management and advanced threat protection across hybrid cloud environments. It uses ML algorithms to enhance anomaly detection and identify deviations from established security policies. Complex ML-based behavioral analytics are employed to detect malicious activities by analyzing logs from various cloud resources to identify compromised entities. The anomaly detection mechanism operates using deep learning techniques, and the model trained on environment-specific normal behavior, allowing the system to accurately flag deviations as potential threats (Microsoft, 2024a). Microsoft Defender for Cloud is capable to fight against following cybercrimes (Microsoft, 2024a)

- Credential Theft
- Suspicious Sign-Ins
- Malicious Activities

Azure ML offers to deploy a custom deep learning model, such as Convolutional Neural Networks (CNNs) and LSTMs, to enhance cyber defense capabilities in cloud environments (Microsoft, 2024b). Azure ML provides an end-to-end platform that supports all stages of the ML lifecycle, including data preprocessing, model training, and evaluation (Buuri et al., 2024). Its anomaly detection capabilities depend on the selected model and the quality of training, allowing for flexible integration of various supervised, unsupervised, or deep learning approaches tailored to specific security needs.

4.5.2 DeepLog

DeepLog is an open-source deep learning-based intrusion detection system that employs the LSTM algorithm for anomaly detection. The model is trained on log patterns extracted from various system log events, where each log message is represented as a log event index to learn normal sequences and identify anomalies effectively (Chen et al., 2021). Aziz and Munir (2024) evaluated the LSTM-based DeepLog intrusion detection tool for its ability to learn and detect log patterns. The model was trained on normal or expected log sequences. DeepLog uses an encoder-decoder structure that allows the model to remember past patterns and predict future log sequences. The model has been evaluated with key performance metrics where standalone DeepLog model delivered moderate results, but the hybrid model delivers outstanding performance, and the datasets also impacted the performance of the model. The authors concluded that while DeepLog is effective in detecting sequential anomaly patterns, retraining is necessary as new log patterns emerge to ensure it can detect unseen and evolving anomalies. DeepLog is capable to fight against cybercrimes, such as Insider threat, Lateral movement, Behavioral anomalies and Zero-Day attacks

4.6 AI based Deepfake image detection tools

This chapter reviews AI-based tools for detecting deepfake images and videos, focusing on the ML algorithms behind them and their performance as reported in research-based articles. The foundational concepts of ML algorithms and various detection methods

Intel's FakeCatcher, a real-time deepfake detection tool designed to identify manipulated images and videos. Sar et al. (2025) have assessed the performance of Intel's FakeCatcher. While the study does not specify the exact ML algorithm used, FakeCatcher typically incorporates deep learning techniques such as RNN, LSTM, and Autoencoders. The tool utilizes the OpenVINO framework to optimize hardware performance and OpenCV for image processing tasks, including face detection and facial landmark extraction. Unlike standard deepfake detectors that rely on learning sequential patterns from logs and frames, FakeCatcher focuses on analyzing physiological and geometric facial features using robust landmark detection. The results showed that FakeCatcher achieved an accuracy of approximately 96%. The authors concluded that while standalone FakeCatcher delivers high accuracy, integrating advanced techniques such as Eulerian Video Magnification and LSTM models for video analysis could further enhance its effectiveness in combating deepfake cybercrimes.

DeepFake-O-Meter, an open source deepfake detection tool which employs deep learning algorithms such as CNNs and Autoencoders. The tool also integrates more than 10 pre-trained deepfake detection models, including well-known algorithms like MesoNet and Xception, to analyze both deepfake images and videos. It classifies the submitted content as real or fake and produces high-quality detector output scores. Li et al. (2021) report that integrated models in DeepFake-O-Meter achieved competitive results, with accuracy rates exceeding 90% and delivers significant performance over other open source deepfake detection tools.

5 Discussion

This thesis has explored the contribution of AI in the domain of cybercrimes and cyber defense systems, with a particular focus on anomaly detection in cloud environments. A comprehensive review of supervised (SVM, DT), unsupervised (DBSCAN), and deep learning models (LSTM, Autoencoder) was carried out, analyzing their technical concepts and evaluating their performance using standard metrics such as accuracy, precision, recall, and F1-score based on selected research articles.

The findings from various experimental conditions showed that the unsupervised DBSCAN model consistently achieved high performance across all metrics. Deep learning models demonstrated strong capabilities in handling complex and large-scale data, making them effective against new and unknown threats. While supervised models showed high accuracy with well-labeled data, they faced challenges in identifying in evolving attack patterns.

This study also reviewed real-world applications of AI/ML techniques, including Microsoft cloud-native tools such as Microsoft Sentinel, Defender for Cloud, and Azure ML, as well as open-source tools like DeepLog. Microsoft's ML-based tools have shown strong capabilities in detecting unknown threats and anomalous user behaviors, leveraging advanced analytics for proactive threat identification. Similarly, open-source tools have demonstrated effectiveness in identifying new and evolving threats.

These results can help cybersecurity service providers and companies to choose appropriate ML based anomaly detection tools suited for their needs and make more informed decisions when dealing with cyber threats. By comparing different ML models, the findings offer useful insights into which methods work best for certain situations—like spotting unknown attacks, reducing false alarms, or improving overall accuracy. These results can be used to strengthen existing detection systems and support the development of smarter, AI-powered security tools, especially in cloud-based environments.

For example, Microsoft Cloud offers various ML-based anomaly detection tools that can be applied across cloud and, to some extent, on-premises environments. It also supports custom ML

integration, enabling security professionals to deploy models tailored to their specific use. Similarly, open-source IDS tools like DeepLog have demonstrated excellent results in detecting unknown anomalies. AI-powered deepfake detection tools use deep learning to identify synthetic content and mitigate cybercrimes and misinformation threats

6 Conclusion and Future Directions

The findings suggest that ML models are well-suited for combating modern cyberattacks and integrating multiple ML models based on key performance metrics can further strengthen the effectiveness of cyber defense systems through innovation of newer and better performing tools to make the AI-powered environments safe and secure for personal, corporate and societal development. As cyber threats continue to advance alongside AI technologies, future research should focus on addressing more complex threats in real-time and integrating adaptive learning models to enhance proactive threat mitigation in cloud and hybrid infrastructures.

7 References

Abbas, S. A., & Almhanna, M. S. (2021). Distributed Denial of Service Attacks Detection System by Machine Learning Based on Dimensionality Reduction. Journal of Physics: Conference Series, 1804(1), 012136. https://doi.org/10.1088/1742-6596/1804/1/012136

Abdallah, A. M., Saif Rashed Obaid Alkaabi, A., Bark Nasser Douman Alameri, G., Rafique, S. H., Musa, N. S., & Murugan, T. (2024). Cloud Network Anomaly Detection Using Machine and Deep Learning Techniques—Recent Research Advancements. IEEE Access, 12, 56749–56773. https://doi.org/10.1109/ACCESS.2024.3390844

Abdolrasol, M. G. M., Hussain, S. M. S., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., Ali, J. A., Mekhilef, S., & Milad, A. (2021). Artificial Neural Networks Based Optimization Techniques: A Review. Electronics, 10(21), 2689. https://doi.org/10.3390/electronics10212689

Ahmed, S. A., Khalifa, E. H., Nawaz, M., Abdalla, F. A., & Mahmoud, A. F. A. (2025). Enhancing Cloud Data Center Security through Deep Learning: A Comparative Analysis of RNN, CNN, and LSTM Models for Anomaly and Intrusion Detection. Engineering, Technology & Applied Science Research, 15(1), 20071–20076. https://doi.org/10.48084/etasr.9445

Alzoubi, Y. I., Mishra, A., & Topcu, A. E. (2024). Research trends in deep learning and machine learning for cloud computing security. Artificial Intelligence Review, 57(5), 132. https://doi.org/10.1007/s10462-024-10776-5

Aslan, O., & Samet, R. (2020). A Comprehensive Review on Malware Detection Approaches. IEEE Access, 8, 6249–6271. https://doi.org/10.1109/ACCESS.2019.2963724

Aziz, A., & Munir, K. (2024). Anomaly Detection in Logs Using Deep Learning. IEEE Access, 12, 176124–176135. https://doi.org/10.1109/ACCESS.2024.3506332

Baawi, S. S., Oleiwi, Z. Ch., Al-Muqarm, A. M. A., Al-Shammary, D., & Sufi, F. (2025). Efficient malware detection based on machine learning for enhanced cloud privacy protection. Evolving Systems, 16(1), 30. https://doi.org/10.1007/s12530-025-09661-5

Boucher, P. (2021, September 8). What if deepfakes made us doubt everything we see and hear? [Science and Technology podcast]. Epthinktank. https://epthinktank.eu/2021/09/08/what-if-deepfakes-made-us-doubt-everything-we-see-and-hear/

Buuri, J., Mansour, S., El-Said, M., & Wang, X. (2024). An Empirical Study Using Microsoft Azure Auto Machine Learning to Detect Zero-Day Attacks. The 25th Annual Conference on Information Technology Education, 7–11. https://doi.org/10.1145/3686852.3686860

Chandra, M. A., & Bedi, S. S. (2021). Survey on SVM and their application in image classification. International Journal of Information Technology, 13(5), 1–11. https://doi.org/10.1007/s41870-017-0080-1

Chen, Z., Liu, J., Gu, W., Su, Y., & Lyu, M. R. (2021). Experience Report: Deep Learning-based System Log Analysis for Anomaly Detection (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2107.05908

Chukwuemeka Nwachukwu, Kehinde Durodola-Tunde, & Chukwuebuka Akwiwu-Uzoma. (2024). AI-driven anomaly detection in cloud computing environments. International Journal of Science and Research Archive, 13(2), 692–710. https://doi.org/10.30574/ijsra.2024.13.2.2184

Das, B., Yadav, N., Chauhan, D., & Gupta, S. (2024). CyMac: Diving Deep into the Application of Machine Learning Algorithms in Cyber Security. International Research Journal of Innovations in Engineering and Technology, 08(01), 74–80. https://doi.org/10.47001/IRJIET/2024.801010

Demkovych, T. (2024, January 29). How AI in Cybersecurity Can Help Fight Cybercrime. Forbytes. https://forbytes.com/blog/ai-in-cybersecurity/

Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity, 5(1), 1. https://doi.org/10.1186/s42400-021-00103-8

Djenna, A., Bouridane, A., Rubab, S., & Marou, I. M. (2023). Artificial Intelligence-Based Malware Detection, Analysis, and Mitigation. Symmetry, 15(3), 677. https://doi.org/10.3390/sym15030677

European Commission. (2024, October 22). Tackling online disinformation | Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/online-disinformation

Eze, C. S., & Shamir, L. (2024). Analysis and Prevention of AI-Based Phishing Email Attacks. Electronics, 13(10), 1839. https://doi.org/10.3390/electronics13101839

Faber, K., Faber, L., & Sniezynski, B. (2021). Autoencoder-based IDS for cloud and mobile devices. 2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid), 728–736. https://doi.org/10.1109/CCGrid51090.2021.00088

Faruk, M. J. H., Shahriar, H., Valero, M., Barsha, F. L., Sobhan, S., Khan, M. A., Whitman, M., Cuzzocreak, A., Lo, D., Rahman, A., & Wu, F. (2022). Malware Detection and Prevention using Artificial Intelligence Techniques. https://doi.org/10.48550/ARXIV.2206.12770

Farzaan, M. A. M., Ghanem, M. C., El-Hajjar, A., & Ratnayake, D. N. (2024). AI-Enabled System for Efficient and Effective Cyber Incident Detection and Response in Cloud Environments (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2404.05602

Freeze, D. (2020, November 30). The History Of Cybercrime And Cybersecurity, 1940-2020. Cybercrime Magazine. https://cybersecurityventures.com/the-history-of-cybercrime-and-cybersecurity-1940-2020/

Galli, A., La Gatta, V., Moscato, V., Postiglione, M., & Sperlì, G. (2024). Explainability in AI-based behavioral malware detection systems. Computers & Security, 141, 103842. https://doi.org/10.1016/j.cose.2024.103842

Gorment, N. Z., Selamat, A., Cheng, L. K., & Krejcar, O. (2023). Machine Learning Algorithm for Malware Detection: Taxonomy, Current Challenges, and Future Directions. IEEE Access, 11, 141045–141089. https://doi.org/10.1109/ACCESS.2023.3256979

Goswami, M. J. (2024). AI-Based Anomaly Detection for Real-Time Cybersecurity. International Journal of Research and Review Techniques, 3(1), Article 1.

Gupta, N., & Mangla, R. (2020). Artificial Intelligence Basics: A Self-Teaching Introduction. Mercury Learning and Information. https://doi.org/10.1515/9781683925149

Hazell, J. (2023). Spear Phishing With Large Language Models (No. arXiv:2305.06972). arXiv. https://doi.org/10.48550/arXiv.2305.06972

Hua, T. K. (2022). A Short Review on Machine Learning. Preprints. https://doi.org/10.22541/au.166490976.66390273/v1

Jia, W., Sun, M., Lian, J., & Hou, S. (2022). Feature dimensionality reduction: A review. Complex & Intelligent Systems, 8(3), 2663–2693. https://doi.org/10.1007/s40747-021-00637-x

Johns, I. (2022). Role of AI in Tackling Cybercrime. 2(4).

Kaliyaperumal, P., Periyasamy, S., Thirumalaisamy, M., Balusamy, B., & Benedetto, F. (2024). A Novel Hybrid Unsupervised Learning Approach for Enhanced Cybersecurity in the IoT. Future Internet, 16(7), 253. https://doi.org/10.3390/fi16070253

Khan, O. U., Abdullah, S. M., Olajide, A. O., Sani, A. I., Faisal, S. M. W., Ogunola, A. A., & Lee, M. D. (2024). The Future of Cybersecurity: Leveraging Artificial Intelligence to Combat Evolving Threats and Enhance Digital Defense Strategies. Journal of Computational Analysis and Applications (JoCAAA), 33(08), Article 08.

Khattab M. Ali Alheeti, Ali Azawii Abdu Lateef, Abdulkareem Alzahrani, Azhar Imran, & Duaa Al_Dosary. (2023). Cloud Intrusion Detection System Based on SVM. International Journal of Interactive Mobile Technologies (iJIM), 17(11), 101–114. https://doi.org/10.3991/ijim.v17i11.39063

Kimmel, J. C., Mcdole, A. D., Abdelsalam, M., Gupta, M., & Sandhu, R. (2021). Recurrent Neural Networks Based Online Behavioural Malware Detection Techniques for Cloud Infrastructure. IEEE Access, 9, 68066–68080. https://doi.org/10.1109/ACCESS.2021.3077498

Kimmell, J. C., Abdelsalam, M., & Gupta, M. (2021). Analyzing Machine Learning Approaches for Online Malware Detection in Cloud. 2021 IEEE International Conference on Smart Computing (SMARTCOMP), 189–196. https://doi.org/10.1109/SMARTCOMP52413.2021.00046

Kumar, J., Rajendran, B., & Sudarsan, S. D. (2023). Zero-Day Malware Classification and Detection Using Machine Learning. SN Computer Science, 5(1), 93. https://doi.org/10.1007/s42979-023-02404-w

LaRocque, A., Gross, G., Lindholm, F., Greco, P., Dupont, B., & Kruger, J. (2024). Effective Ransomware Detection Using Autonomous Pattern-based Signature Extraction. Preprints. https://doi.org/10.22541/au.173016272.26231350/v1

Lee, Y.-C., Chien, W.-C., & Chang, Y.-C. (2024). FedDB: A Federated Learning Approach Using DBSCAN for DDoS Attack Detection. Applied Sciences, 14(22), 10236. https://doi.org/10.3390/app142210236

Li, Y., Zhang, C., Sun, P., Ke, L., Ju, Y., Qi, H., & Lyu, S. (2021). DeepFake-o-meter: An Open Platform for DeepFake Detection. 2021 IEEE Security and Privacy Workshops (SPW), 277–281. https://doi.org/10.1109/SPW53761.2021.00047

Luna, C. dela. (2024, September 16). AI and Cyber Security: Innovations and Challenges. eSecurity Planet. https://www.esecurityplanet.com/trends/ai-and-cybersecurity-innovations-and-challenges/

Mamidi, S. R. (2024). Future Trends in AI Driven Cyber Security. 8(2).

McBride, D. (2024, October 16). Deepfake Scams: Staying Safe with AI and Cybersecurity | 99Ten Business Solutions. Https://99ten.Com. https://99ten.com/articles/deepfake-scams-staying-safe-with-ai-and-cybersecurity

Microsoft. (2023, January 25). Bring your own ML into Microsoft Sentinel. https://learn.microsoft.com/en-us/azure/sentinel/bring-your-own-ml

Microsoft. (2024a, July 8). Security alerts and incidents—Microsoft Defender for Cloud. https://learn.microsoft.com/en-us/azure/defender-for-cloud/alerts-overview

Microsoft. (2024b, September 19). What is Azure Machine Learning? - Azure Machine Learning. https://learn.microsoft.com/en-us/azure/machine-learning/overview-what-is-azure-machine-learning?view=azureml-api-2

Microsoft. (2024c, November 19). Threat detection in Microsoft Sentinel. https://learn.microsoft.com/en-us/azure/sentinel/threat-detection

Microsoft Digital Defense Report. (2024). Microsoft Digital Defense Report 2024.

Mienye, I. D., & Jere, N. (2024). A Survey of Decision Trees: Concepts, Algorithms, and Applications. IEEE Access, 12, 86716–86727. https://doi.org/10.1109/ACCESS.2024.3416838

Miller, C., Portlock, T., Nyaga, D. M., & O'Sullivan, J. M. (2024). A review of model evaluation metrics for machine learning in genetics and genomics. Frontiers in Bioinformatics, 4, 1457619. https://doi.org/10.3389/fbinf.2024.1457619

Mohamed, N., Taherdoost, H., & Khashan, O. A. (2025). A Review of AI in Spear Phishing Defense: Detecting and Thwarting Advanced Email Threats. In H. Taherdoost, Y. Farhaoui, S. R. Shahamiri, T.-V. Le, M. Madanchian, & M. Prasad (Eds.), EAI 3rd International Conference on Smart Technologies and Innovation Management (pp. 177–189). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64957-8_14

Mustafa Majid, A.-A., Alshaibi, A. J., Kostyuchenko, E., & Shelupanov, A. (2023). A review of artificial intelligence based malware detection using deep learning. Materials Today: Proceedings, 80, 2678–2683. https://doi.org/10.1016/j.matpr.2021.07.012

Naitali, A., Ridouani, M., Salahdine, F., & Kaabouch, N. (2023). Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions. Computers, 12(10), 216. https://doi.org/10.3390/computers12100216

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). A Comprehensive Overview of Large Language Models (No. arXiv:2307.06435). arXiv. https://doi.org/10.48550/arXiv.2307.06435

Pitafi, S., Anwar, T., & Sharif, Z. (2022). An Improved Approach Based on Density-Based Spatial Clustering of Applications with a Noise Algorithm for Intrusion Detection. Journal of Hunan University Natural Sciences, 49(12), 67–77. https://doi.org/10.55463/issn.1674-2974.49.12.7

Qiang, X., Tang, Y., Wu, L., & Lyu, Z. (2024). Li-Ion Battery State of Health Estimation Using Hybrid Decision Tree Model Optimized by Bayesian Optimization. Energy Technology, 12(3), 2301065. https://doi.org/10.1002/ente.202301065

Rabhi, M., Bakiras, S., & Di Pietro, R. (2024). Audio-deepfake detection: Adversarial attacks and countermeasures. Expert Systems with Applications, 250, 123941. https://doi.org/10.1016/j.eswa.2024.123941

Rehman, F., Mushtaq, F., & Zaman, H. (2024). A Host-based Intrusion Detection: Using Signature-based and AI-driven Anomaly Detection for Enhanced Cybersecurity*. 2024 4th International Conference on Digital Futures and Transformative Technologies (ICoDT2), 1–7. https://doi.org/10.1109/ICoDT262145.2024.10740248

Rivera-Lopez, R., Canul-Reich, J., Mezura-Montes, E., & Cruz-Chávez, M. A. (2022). Induction of decision trees as classification models through metaheuristics. Swarm and Evolutionary Computation, 69, 101006. https://doi.org/10.1016/j.swevo.2021.101006

Roponena, E., Kampars, J., Gailitis, A., & Strods, J. (2021). A Literature Review of Machine Learning Techniques for Cybersecurity in Data Centers. 2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), 1–6. https://doi.org/10.1109/ITMS52826.2021.9615321

Sai Meghana, G. V., Saqlain Afroz, S., Gurindapalli, R., Katari, S., & Swetha, K. (2024). A Survey paper on Understanding the Rise of AI-driven Cyber Crime and Strategies for Proactive Digital Defenders. 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN), 25–30. https://doi.org/10.1109/ICPCSN62568.2024.00012

Sar, A., Sati, S., Choudhury, T., Joshi, P., Sille, R., Srihari, K., & Bansal, K. (2025). A Unified Neural Framework for Real-Time Deepfake Detection Across Multimedia Modalities to Combat Misleading Content. IEEE Access, 13, 48683–48702. https://doi.org/10.1109/ACCESS.2025.3550770

securitymagazine. (2025). Deepfake-enabled fraud caused more than $200 million in losses | Security Magazine. https://www.securitymagazine.com/articles/101559-deepfake-enabled-fraud-caused-more-than-200-million-in-losses

Shahzad, F., Mannan, A., Javed, A. R., Almadhor, A. S., Baker, T., & Al-Jumeily Obe, D. (2022). Cloud-based multiclass anomaly detection and categorization using ensemble learning. Journal of Cloud Computing, 11(1), 74. https://doi.org/10.1186/s13677-022-00329-y

Shewale, Y., Kumar, S., & Banait, S. (2023). Machine Learning Based Intrusion Detection in IoT Network Using MLP and LSTM. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), Article 7s.

Shi, Y., Zhang, L., Peterson, C. B., Do, K.-A., & Jenq, R. R. (2022). Performance determinants of unsupervised clustering methods for microbiome data. Microbiome, 10(1), 25. https://doi.org/10.1186/s40168-021-01199-3

Singh, H. V., Girdhar, A., & Dahiya, S. (2022). A Literature survey based on DBSCAN algorithms. 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 751–758. https://doi.org/10.1109/ICICCS53718.2022.9788440

Souri, A., & Hosseini, R. (2018). A state-of-the-art survey of malware detection approaches using data mining techniques. Human-Centric Computing and Information Sciences, 8(1), 3. https://doi.org/10.1186/s13673-018-0125-x

Sukhvinder Singh Dari, E. Al. (2024). Neural Networks and Cyber Resilience: Deep Insights into AI Architectures for Robust Security Framework. Journal of Electrical Systems, 19(3), 78–95. https://doi.org/10.52783/jes.653

Tayyab, U.-H., Khan, F. B., Durad, M. H., Khan, A., & Lee, Y. S. (2022). A Survey of the Recent Trends in Deep Learning Based Malware Detection. Journal of Cybersecurity and Privacy, 2(4), 800–829. https://doi.org/10.3390/jcp2040041

TENK, T. (2023). The Finnish code of conduct for research integrity and procedures for handling alleged violations of research integrity in Finland 2023.

Torabi, H., Mirtaheri, S. L., & Greco, S. (2023). Practical autoencoder based anomaly detection by using vector reconstruction error. Cybersecurity, 6(1), 1. https://doi.org/10.1186/s42400-022-00134-9

Torre, R. D. L. (2023). How AI Is Shaping the Future of Cybercrime. https://www.darkreading.com/vulnerabilities-threats/how-ai-shaping-future-cybercrime

Usman, Y., Upadhyay, A., Gyawali, P., & Chataut, R. (2024). Is Generative AI the Next Tactical Cyber Weapon For Threat Actors? Unforeseen Implications of AI Generated Cyber Attacks (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2408.12806

Velasco, C. (2022). Cybercrime and Artificial Intelligence. An overview of the work of international organizations on criminal justice and the international applicable instruments. ERA Forum, 23(1), 109–126. https://doi.org/10.1007/s12027-022-00702-z

Wang, B., Hua, Q., Zhang, H., Tan, X., Nan, Y., Chen, R., & Shu, X. (2022). Research on anomaly detection and real-time reliability evaluation with the log of cloud platform. Alexandria Engineering Journal, 61(9), 7183–7193. https://doi.org/10.1016/j.aej.2021.12.061

Wang, S., Balarezo, J. F., Kandeepan, S., Al-Hourani, A., Chavez, K. G., & Rubinstein, B. (2021). Machine Learning in Network Anomaly Detection: A Survey. IEEE Access, 9, 152379–152396. https://doi.org/10.1109/ACCESS.2021.3126834

Wang, Y., Liao, W., Shen, H., Jiang, Z., & Zhou, J. (2024). Some notes on the basic concepts of support vector machines. Journal of Computational Science, 82, 102390. https://doi.org/10.1016/j.jocs.2024.102390

Xing, X., Jin, X., Elahi, H., Jiang, H., & Wang, G. (2022). A Malware Detection Approach Using Autoencoder in Deep Learning. IEEE Access, 10, 25696–25706. https://doi.org/10.1109/ACCESS.2022.3155695

Xu, H., Wang, S., Li, N., Wang, K., Zhao, Y., Chen, K., Yu, T., Liu, Y., & Wang, H. (2024). Large Language Models for Cyber Security: A Systematic Literature Review (No. arXiv:2405.04760). arXiv. https://doi.org/10.48550/arXiv.2405.04760

Yee Por, L., Dai, Z., Juan Leem, S., Chen, Y., Yang, J., Binbeshr, F., Yuen Phan, K., & Soon Ku, C. (2024). A Systematic Literature Review on AI-Based Methods and Challenges in Detecting Zero-Day Attacks. IEEE Access, 12, 144150–144163. https://doi.org/10.1109/ACCESS.2024.3455410

Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., Raj, G. D., Jhaveri, R. H., Prabadevi, B., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2024). GPT (Generative Pre-Trained Transformer)—A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. IEEE Access, 12, 54608–54649. https://doi.org/10.1109/ACCESS.2024.3389497

Zhong, G., Liu, F., Jiang, J., Wang, B., Yao, X., & Chen, C. L. P. (2024). Detecting Cloud Anomaly via Broad Network-Based Contrastive Autoencoder. IEEE Transactions on Network and Service Management, 21(3), 3249–3263. https://doi.org/10.1109/TNSM.2024.3353772

8 Appendix

Research articles were used this study.

Publication Year	Title	Publication Title
2020	The History Of Cybercrime And Cybersecurity, 1940-2020	Cybercrime Magazine
2024	Artificial Intelligence Toolkit	
2023	The Future of Cybercrime: AI and Emerging Technologies Are Creating a Cybercrime Tsunami	SSRN Electronic Journal
2024	A Survey paper on Understanding the Rise of AI-driven Cyber Crime and Strategies for Proactive Digital Defenders	2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN)
2022	The Emerging Threat of Ai-driven Cyber Attacks: A Review	Applied Artificial Intelligence
2023	How AI Is Shaping the Future of Cybercrime	
2023	Artificial Intelligence-Based Malware Detection, Analysis, and Mitigation	Symmetry
2023	The Finnish code of conduct for research integrity and procedures for handling alleged violations of research integrity in Finland 2023	
2022	Malware Detection and Prevention using Artificial Intelligence Techniques	
2023	Unmasking Cybercrime with Artificial-Intelligence-Driven Cybersecurity Analytics	Sensors
2023	Breaking Alert Fatigue: AI-Assisted SIEM Framework for Effective Incident Response	Applied Sciences
2019	Literature review as a research methodology: An overview and guidelines	Journal of Business Research
2020	Artificial Intelligence Basics: A Self-Teaching Introduction	
2024	How AI in Cybersecurity Can Help Fight Cybercrime	Forbytes
2020	AI Watch, historical evolution of artificial intelligence: analysis of the three main paradigm shifts in AI.	
1975	Delineation of the intimate details of the backbone conformation of pyridine nucleotide coenzymes in aqueous solution	Biochemical and Biophysical Research Communications
1975	Kinetic and spectral studies of type I and type II compounds with rat hepatic microsomes in the presence of the major metabolite of diphenylhydantoin	Drug Metabolism and Disposition: The Biological Fate of Chemicals
2022	Role of AI in Tackling Cybercrime	

2022	Cybercrime and Artificial Intelligence. An overview of the work of international organizations on criminal justice and the international applicable instruments	ERA Forum
2024	Tackling online disinformation | Shaping Europe's digital future	
2021	What if deepfakes made us doubt everything we see and hear? [Science and Technology podcast]	Epthinktank
2024	Analysis and Prevention of AI-Based Phishing Email Attacks	Electronics
2025	A Review of AI in Spear Phishing Defense: Detecting and Thwarting Advanced Email Threats	EAI 3rd International Conference on Smart Technologies and Innovation Management
2023	Spear Phishing With Large Language Models	
2023	Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions	Computers
2024	Audio-deepfake detection: Adversarial attacks and countermeasures	Expert Systems with Applications
2023	A review of artificial intelligence based malware detection using deep learning	Materials Today: Proceedings
2024	The Future of Cybersecurity: Leveraging Artificial Intelligence to Combat Evolving Threats and Enhance Digital Defense Strategies	Journal of Computational Analysis and Applications (JoCAAA)
2024	Future Trends in AI Driven Cyber Security	
2024	AI and Cyber Security: Innovations and Challenges	eSecurity Planet
2022	A Short Review on Machine Learning	
2021	A Literature Review of Machine Learning Techniques for Cybersecurity in Data Centers	2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)
2021	Artificial Neural Networks Based Optimization Techniques: A Review	Electronics
2024	Neural Networks and Cyber Resilience: Deep Insights into AI Architectures for Robust Security Framework	Journal of Electrical Systems
2024	A Comprehensive Overview of Large Language Models	
2024	Large Language Models for Cyber Security: A Systematic Literature Review	
2024	GPT (Generative Pre-Trained Transformer)—A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions	IEEE Access

2024	A Host-based Intrusion Detection: Using Signature-based and AI-driven Anomaly Detection for Enhanced Cybersecurity*	2024 4th International Conference on Digital Futures and Transformative Technologies (ICoDT2)
2018	A state-of-the-art survey of malware detection approaches using data mining techniques	Human-centric Computing and Information Sciences
2024	Effective Ransomware Detection Using Autonomous Pattern-based Signature Extraction	
2020	A Comprehensive Review on Malware Detection Approaches	IEEE Access
2017	Intrusion Detection in Contemporary Environments	Computer and Information Security Handbook
2021	Machine Learning in Network Anomaly Detection: A Survey	IEEE Access
2024	AI-driven anomaly detection in cloud computing environments	International Journal of Science and Research Archive
2024	CyMac: Diving Deep into the Application of Machine Learning Algorithms in Cyber Security	International Research Journal of Innovations in Engineering and Technology
2021	Analyzing Machine Learning Approaches for Online Malware Detection in Cloud	2021 IEEE International Conference on Smart Computing (SMARTCOMP)
2024	AI-Based Anomaly Detection for Real-Time Cybersecurity	International Journal of Research and Review Techniques
2024	A Systematic Literature Review on AI-Based Methods and Challenges in Detecting Zero-Day Attacks	IEEE Access
2024	Cloud Network Anomaly Detection Using Machine and Deep Learning Techniques—Recent Research Advancements	IEEE Access
2022	A Survey of the Recent Trends in Deep Learning Based Malware Detection	Journal of Cybersecurity and Privacy
2024	Research trends in deep learning and machine learning for cloud computing security	Artificial Intelligence Review
2021	Survey on SVM and their application in image classification	International Journal of Information Technology
2024	Some notes on the basic concepts of support vector machines	Journal of Computational Science
2022	Feature dimensionality reduction: a review	Complex & Intelligent Systems
2021	Distributed Denial of Service Attacks Detection System by Machine Learning Based on Dimensionality Reduction	Journal of Physics: Conference Series
2024	A review of model evaluation metrics for machine learning in genetics and genomics	Frontiers in Bioinformatics
2022	Performance determinants of unsupervised clustering methods for microbiome data	Microbiome

2025	Efficient malware detection based on machine learning for enhanced cloud privacy protection	Evolving Systems
2023	Cloud Intrusion Detection System Based on SVM	International Journal of Interactive Mobile Technologies (iJIM)
2022	Research on anomaly detection and real-time reliability evaluation with the log of cloud platform	Alexandria Engineering Journal
2022	Induction of decision trees as classification models through metaheuristics	Swarm and Evolutionary Computation
2024	A Survey of Decision Trees: Concepts, Algorithms, and Applications	IEEE Access
2023	Machine Learning Algorithm for Malware Detection: Taxonomy, Current Challenges, and Future Directions	IEEE Access
2022	Cloud-based multiclass anomaly detection and categorization using ensemble learning	Journal of Cloud Computing
2024	AI-Enabled System for Efficient and Effective Cyber Incident Detection and Response in Cloud Environments	
2023	Zero-Day Malware Classification and Detection Using Machine Learning	SN Computer Science
2020	Research on Anomaly Detection Method Based on DBSCAN Clustering Algorithm	2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT)
2024	International Journal of Advanced Research in Computer and Communication Engineering	SSRN Electronic Journal
2022	A Literature survey based on DBSCAN algorithms	2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)
2024	A Novel Hybrid Unsupervised Learning Approach for Enhanced Cybersecurity in the IoT	Future Internet
2022	An Improved Approach Based on Density-Based Spatial Clustering of Applications with a Noise Algorithm for Intrusion Detection	Journal of Hunan University Natural Sciences
2024	Harnessing DBSCAN and auto-encoder for hyper intrusion detection in cloud computing	Bulletin of Electrical Engineering and Informatics
2024	FedDB: A Federated Learning Approach Using DBSCAN for DDoS Attack Detection	Applied Sciences
2021	Recurrent Neural Networks Based Online Behavioural Malware Detection Techniques for Cloud Infrastructure	IEEE Access
2023	Machine Learning Based Intrusion Detection in IoT Network Using MLP and LSTM	International Journal of Intelligent Systems and Applications in Engineering

2025	Enhancing Cloud Data Center Security through Deep Learning: A Comparative Analysis of RNN, CNN, and LSTM Models for Anomaly and Intrusion Detection	Engineering, Technology & Applied Science Research
2024	Explainability in AI-based behavioral malware detection systems	Computers & Security
2021	Autoencoder-based IDS for cloud and mobile devices	2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)
2023	Practical autoencoder based anomaly detection by using vector reconstruction error	Cybersecurity
2022	A Malware Detection Approach Using Autoencoder in Deep Learning	IEEE Access
2024	Detecting Cloud Anomaly via Broad Network-Based Contrastive Autoencoder	IEEE Transactions on Network and Service Management
2022	Approaches to Qualitative Comparative Analysis and good practices: A systematic review	Swiss Political Science Review
2007	The Research Imagination: An Introduction to Qualitative and Quantitative Methods	
2024	Threat detection in Microsoft Sentinel	
2023	Bring your own ML into Microsoft Sentinel	
2024	Security alerts and incidents - Microsoft Defender for Cloud	
2024	What is Azure Machine Learning? - Azure Machine Learning	
2024	An Empirical Study Using Microsoft Azure Auto Machine Learning to Detect Zero-Day Attacks	The 25th Annual Conference on Information Technology Education
2021	Experience Report: Deep Learning-based System Log Analysis for Anomaly Detection	
2024	Anomaly Detection in Logs Using Deep Learning	IEEE Access
2025	A Unified Neural Framework for Real-Time Deepfake Detection Across Multimedia Modalities to Combat Misleading Content	IEEE Access
2021	DeepFake-o-meter: An Open Platform for DeepFake Detection	2021 IEEE Security and Privacy Workshops (SPW)
2024	Microsoft Digital Defense Report 2024	
2025	Deepfake statistics 2025: how frequently are celebrities targeted?	Surfshark
2025	Deepfake-enabled fraud caused more than $200 million in losses | Security Magazine	
2024	Is Generative AI the Next Tactical Cyber Weapon For Threat Actors? Unforeseen Implications of AI Generated Cyber Attacks	