Federated Learning in Ransomware Detection: A Systematic Literature Review

Chinonso E. Ali[1] , Songfeng Lu[1,*] , Francis A. Ruambo[1] , Francelle TCHAMINI[1] , Umar Muhammad Ibrahim[1]
[1]School of Cyber Science and Engineering, Huazhong University of Science and Technology, Wuhan, 430074 China.
* Corresponding Author
Ericchinonso1@outlook.com ; Orcid: 0009-0009-8449-627X
lusongfeng@hust.edu.cn ; Orcid: 0000-0003-4489-2488
ruambof@gmail.com ; Orcid: 0000-0001-6218-8203
frantchams@gmail.com ; Orcid: 0009-0000-1281-3138)
muhammad2k16@gmail.com ; Orcid: 0009-0008-5401-565

Abstract

The exploitative and destructive challenges posed my ransomwares has continues to persist within the cyberspace industry. The increasing frequency and complexity of ransomware attacks threaten data security, resulting in substantial financial losses and operational disruptions across sectors. Recently, Federated Learning (FL) technology has been identified as a prospect for improvement in ransomware detection and mitigation. This trend is because it provides a decentralized method using machine learning (ML)/deep learning (DL) techniques to enable the collaborative training of multiple devices without providing access to their private information. This Systematic Literature Review (SLR) synthesizes the current applications of FL in ransomware detection, providing a critical evaluation of the successes and limitations of these approaches. Additionally, the review explores the evolving ransomware threat landscape and offers suggestions for future research directions to strengthen ransomware defenses. Our review began by identifying 185 relevant publications from 2019 to 2024. After thoroughly examining their abstracts, methodologies, and full texts, 53 key papers were selected for in-depth analysis. These articles were sourced from reputable databases, including Scopus, Web of Science, Springer Nature, and IEEE, among others, with the findings reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Our study addresses four critical research questions:(RQ1, RQ2, RQ3, and RQ4). Through these questions, this SLR presents a complete overview of the recent happenings in ransomware detection using FL, demonstrating valuable insights and emerging trends that can guide researchers and practitioners in crafting more effective strategies to combat ransomware attacks.

Keywords: Collaborative, Federated Learning, Machine Learning, Ransomware Attacks, Ransomware Detection, Trend Landscape

1. Introduction

We currently leave in world of destructive cyber-attack and exploitation, one of the most exploitative and dangerous malware type is called ransomware. This program poses an enormous threat to organizations of all sizes, primarily by encrypting relevant data and making it out of reach to the owners until a ransom is paid for the decryption key. Since the discovery of the first ransomware, Trojan Horse AIDS, in 1989, as noted in [1] numerous high-profile ransomware incidents have led to substantial financial extortion, amounting to billions of dollars. [2] reported that in 2021, the US-based meat production company JBS paid $11 million in ransom following an attack that disrupted its operations. Beyond the direct costs of ransom payments and operational downtime, victims of ransomware attacks may also suffer additional damages, such as data loss and reputational harm. [3] Points out that these attackers often get access to devices through phishing emails, malicious attachments, or exploiting the network and software vulnerabilities of the target platform. Phishing is a cybercrime in which criminals send spam emails with harmful links. The objective is to deceive recipients into accessing deceptive websites or downloading harmful software; initially, these messages were limited to emails; however, the recent strategy has expanded to texts, social media engagements, and phone calls.

[4] maintains that cybercrimes were not so profitable and attractive, but that has changed dramatically in recent history, primarily through the advancement of ransomware as a service (RaaS) software. Attackers no longer necessarily need technical knowledge or expertise before they carry out criminal activities since they can purchase an already developed and packaged ransomware service plan aided by an increasing community on the Darkweb. In 2021, Sophos [5]reported a significant rise of 75% in the occurrence of ransomware attacks. This increase affected 77% of organizations, a notable rise from the 44% recorded in 2020. The number of assaults decreased by 23% in 2022, suggesting that improvement in detection and mitigating technologies, heightened regulatory scrutiny, and public awareness about the modes of propagation effects have a significant impact; however, due to the constant advancement of ransomware, this didn’t persist for a long time as 2023 to early 2024 produced 4,893 victims up from 2,708 the previous year as observed in [1]. Ransomware attacks involve six stages; we have annotated this in Figure 1. At first, the attackers implore different social engineering tactics to deceive victims into installing malicious wares; once the compromised software is installed, the rest of the process will follow, including the virus searching for relevant files to encrypt, establishing a connection with the hacker's backend servers where such files will be sent before it encrypts, delete or Lock folder and then finally display a message to inform the user of their demand.

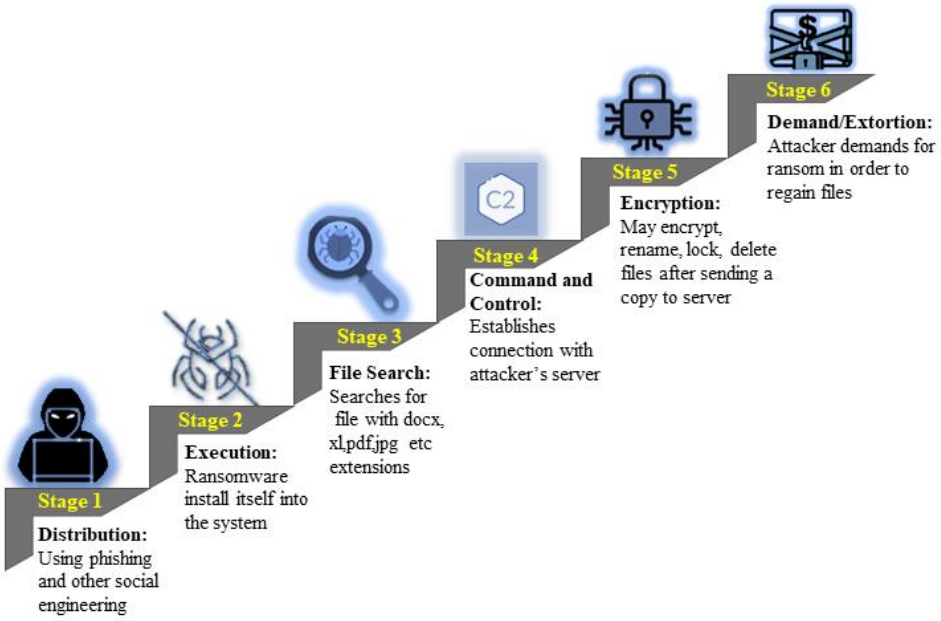

Fig 1 Stages of a ransomware attack showing the step-by-step procedures that a typical cyber actor can implore to propagate ransomware before demanding payment

Meanwhile, the observed trend in Figure 2, as analyzed by [5], shows the percentage of ransomware attacks in recent years (2020 - 2024), indicating continuous growth of these cyber-attacks yearly. It also shows the need to continue developing more aggressive anti-ransomware systems, primarily through a precise and collaborative methodology that will reduce threats and exploitation by hackers.

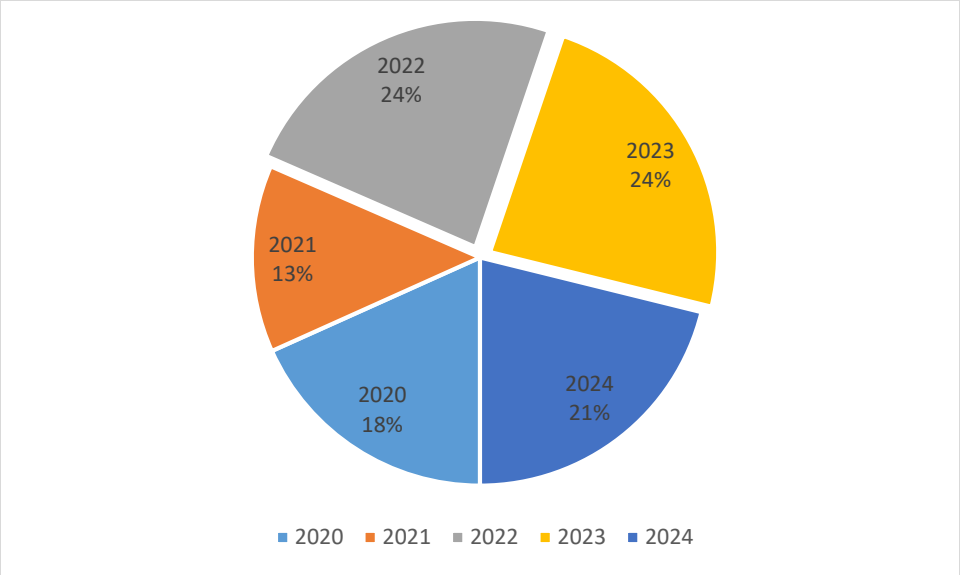

Fig 2 Percentage of ransomware attacks [5] from the year 2020 to 2024, with the attack remaining at the same level throughout 2022 to 2023

Existing literature reviews [6][7][8] offer valuable contributions through analysis of various ransomware detection and mitigation approaches; however, none have comprehensively addressed the aspect of the application of Federated Learning (FL) to ransomware detection, common challenges facing FL application in ransomware detection domain and current Ransomware trend landscape. This study has made the following key contributions:

- Conducting a detailed analysis of the ransomware trend landscape.
- Summarizing and tabulating various detection approaches, highlighting research gaps
- Analyzing recent applications of FL for Ransomware detection and Mitigation.
- Providing common challenges limiting FL application to the Ransomware detection domain and suggestions for future directions.

The rest of this study is organized as follows: Section 2 is about the Related Survey, and Section 3 analyzes the trend landscape. Section 4 discusses various ransomware detection techniques and the associated deployment challenges. Section 5 dives into the federated learning application for ransomware detection, Section 6 outlines the review methodology, and Section 7 presents the limitations and conclusion. Figure 3 represents a comprehensive and high-level overview of the structure of this study, detailing the sections and demography, subject area, research questions, and methodology.

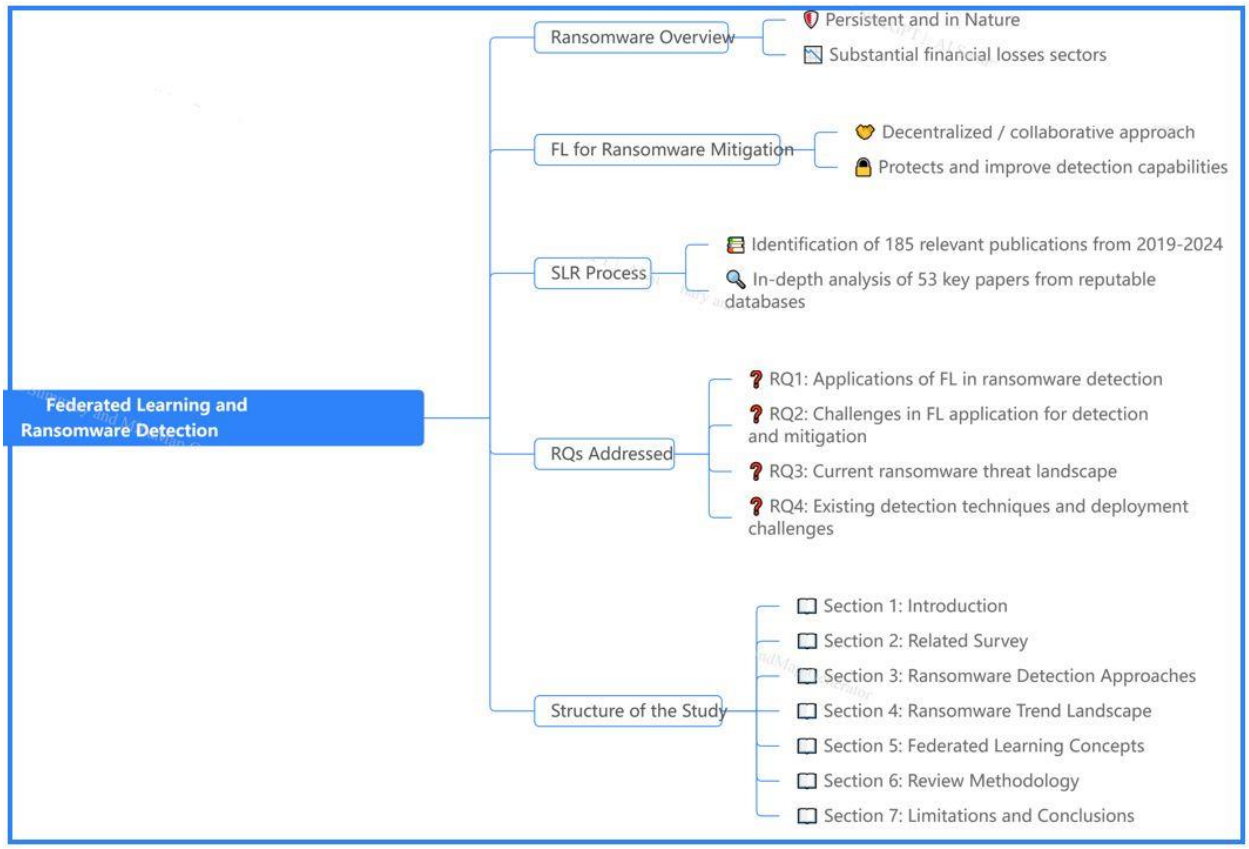

Fig 3 High-level overview of this SLR, highlighting the areas of interest, procedures, research questions, and the structure of the study

2. Related Surveys

The study conducted by authors in [9]made a notable contribution to raising awareness among the general public, especially those with limited technical knowledge, regarding the threats posed by cybercriminals using ransomware for illicit activities. They outlined several critical risks, including system shutdowns caused by ransomware infections, loss or theft of data, financial implications, and severe but rare cases, even loss of life, particularly in sectors like healthcare, where data integrity is crucial. The authors proposed several key mitigation strategies, such as implementing robust email security protocols, utilizing intrusion prevention systems, employing download insights to identify and block malicious content, ensuring browser and exploit protection, and adopting best practices in cybersecurity to fortify defenses against such attacks. The research study [10]outlined the issues facing malware detection using data mining methods; the study conducted a comparative analysis of these methods based on the classification types, dataset size, and analytical methodology.

Research by [11] surveyed the ransomware problem concerning the Internet of Things (IoT) domain application. The authors examined the varying sophistication of ransomware attacks, emphasizing the susceptibility of the IoT and the consequent need to act decisively. Another related work [12] examined the methods used for Android security. Specifically, they addressed the approaches attackers employ to exploit vulnerabilities in malware detection systems. The authors presented a thorough analysis of the merits and limitations of well-established research approaches within the domain of Android security. The survey also furnishes scholars and practitioners with a basis for suggesting novel methodologies to examine and counteract these attacks. A taxonomy of ransomware research efforts only utilizing clever Machine

Learning (ML) methods was surveyed by research work in [13]. Their study involved thorough investigations of papers published from 2016 to 2020, subsequently highlighting several potential future paths and obstacles in the implementation of Deep Learning (DL) methods for ransomware protection.

In addition to revealing the business side of ransomware and its actors, research work [14] conducted a comprehensive study looking at the economic consequences of ransomware, specifically from the standpoint of Bitcoin transactions. The study also presented an overview of each investigated ransomware origin, evolution, and method of ransomware attack operation. Authors in [15] focused on memory forensics, explicitly targeting the widespread impact of the Wannarcy ransomware on the global computer network. Authors in[16] and [12] analyzed the application of ML to malware classification using the Windows operating system as a benchmark, particularly for Portable Executables. They selected papers according to their objectives and the application of ML methods, then rounded off by identifying issues, particularly those about datasets and prospects for progress. The study work [17] employed a unique approach to examining research papers on ransomware, mainly focusing on the variety of platforms that are most targeted; they also addressed relevant literature about mobile devices as well as the IoT.

On the FL concept,[18] [19]explored the various models of data partitioning in FL, with a focus on the neural network topologies used in federated models. In the same vain, the study in [11] provided a thorough examination of existing research on FL from five dimensions: basic FL understanding, privacy and security considerations in FL, overhead associated with communication concerns, FL heterogeneity challenges, and several FL implementations in different domains while the study in [20] focused on diverse areas of FL application comprising health related attributions, a network of vehicles, smart cities, recommender systems. The study also highlighted developmental parameters, such as application programmers' interface, system designs, and communication efficiency.

The study in [21] investigated the progress of FL in the field of healthcare informatics. With a comprehensive overview, they analyzed the statistical obstacles and their corresponding resolutions, system obstacles, and privacy concerns in this context. Their target was to offer valuable resources for computational research on ML methods for handling extensively dispersed data while considering its privacy and health informatics. A similar work in [22] reported how FL can be used in different fields and businesses in multiple ways, such as enhancing distributed data collection and improving neural network architectures for better collaboration in an FL environment. They also examined the problems that FL causes, such as differences in statistics and systems, uneven data distribution, problems with allocating resources, and privacy issues. A recent approach by [23] evaluated the existing challenges in implementing federated learning in energy computing, including their corresponding remedies. Additionally, the authors presented portable edge improvements and identified the key obstacles and issues that need addressing in future FL studies.

The comparison of studies in Table 1 reveals that many existing works within ransomware detection and FL are often involved in superficial or one-sided surveys in their interest instead of presenting a round view. They are conducting more thorough investigations focused solely on ransomware. FL will enhance our understanding of the roadmap to integrating FL into ransomware detection and mitigation and guide us through the complexities involved, identifying best practices and addressing existing gaps in methodologies.

Table 1 Summary of related surveys

Work	Background		Evolution	Discussed Challenges	Future Perspectives	Taxonomy
	Ransomware	FL				
	✓	☒	☒	✓	✓	☒
[9]	✓	☒	✓	☒	✓	☒
[10]	☒	✓	☒	✓	☒	✓
[4	✓	☒	✓	✓	✓	☒
[14]	✓	☒	☒	✓	☒	☒
[15]	✓	☒	☒	✓	✓	✓
[17]	☒	✓	☒	✓	✓	☒
[11]	✓	☒	☒	✓	✓	✓
[20]	✓	☒	☒	☒	✓	✓
[23]	☒	✓	☒	✓	☒	☒
Our	✓	✓	☒	✓	✓	☒

☒: **Denotes what the survey did not cover.**

✓: **Denotes that the survey has addressed the criterion.**

3. Ransomware Detection and Mitigation Approaches

As ransomware attacks continue to increase, researchers and security specialists have developed numerous detection techniques. Several studies [24][25][26] have dedicated their prowess to devising innovative mitigation strategies to combat the ransomware menace, encompassing the complete process to ensure secured information across three key areas, namely detection, prediction, and mitigation. These techniques predominantly leverage ML/DL technologies and employ static, dynamic, or hybrid analysis approaches. As illustrated in Figure 4, static methods focus on signature attributes, dynamic methods on behavioral attributes, and hybrid methods combine both techniques. The primary aim of these processes is to extract pertinent information for the detection phase. These parameters are usually collated and analyzed to make informed predictions based on specific metrics, ultimately guiding actions to mitigate these attacks.

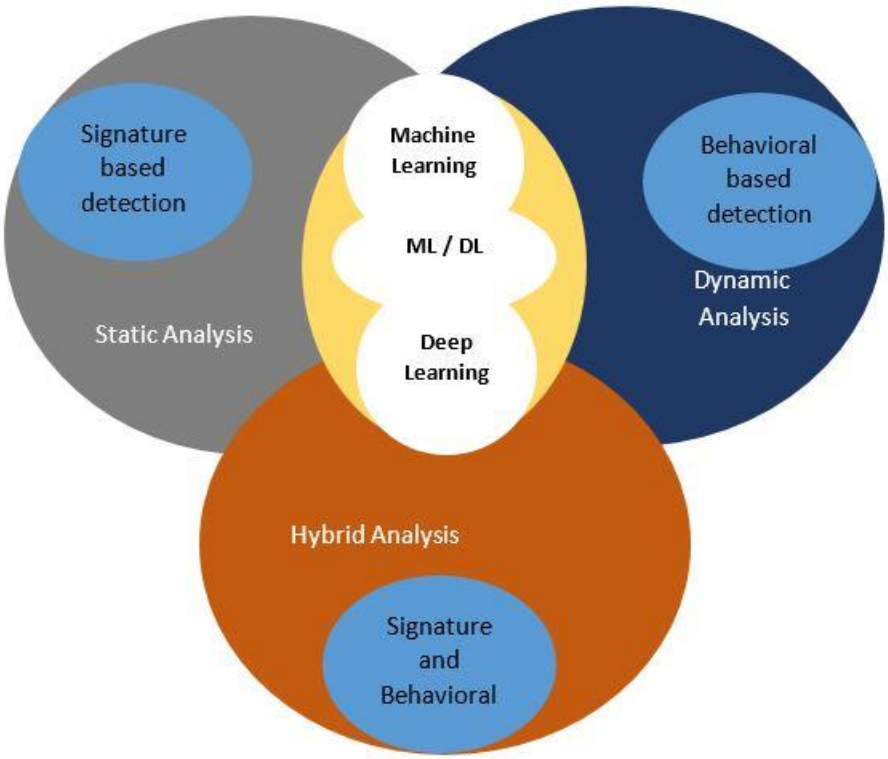

Fig 4 Annotation of Ransomware detection techniques and the enabling technologies involved in various approaches

3.1 Static Analysis

The static analysis method employs the tacting of searching for a known signature for detection by comparing the digital footprint of a given sample to that stored in the database of known malicious signatures. This approach involves scrutinizing the code and structure of suspicious software without executing it. It relies on detailed information about ransomware activities, typically stored in the META-INF directory, which stores metadata about the application and ensures its integrity and authenticity. For example, [27] clarifies that the Android operating system's permission management mechanism uses information within the APK file to identify and detect malicious applications. In their review of Windows malware detection, [28] also emphasize that the signatures-based detection paradigm gives researchers critical insights into the content and structure of samples, enabling the early identification of potentially harmful instructions. This mechanism is essential for minimizing impact and reducing the likelihood of successful attacks. Notable studies, such as [29] and [30], have effectively integrated this approach. This methodology's parameters for the detection process include hashes, API calls, executables, entropy change, and opcode frequencies. The prominent application of the static approach was proposed by [31], who utilized assembly language to perform reverse engineering on PE information and then applied dynamic linkable libraries (DLLs) and function call extraction to the header file for feature extraction. Using CNN,[32] extracted headers to create a grayscale image with a zigzag pattern and trained the model using the images to predict patterns. At the same time, [33] focused on using entropy parameters to analyze and capture the decisive characteristics of a typical Ransomware; however, [34] argues that static-related approaches are relatively limited because they are unable to capture the information necessary for precise

ransomware detection; therefore, the authors combined feature approach in their study to acquire a wide range of features which enabled them to classify ransomware behavior and characteristics accurately.

3.2 Dynamic Analysis

Dynamic analysis is considered more accurate and comprehensive than static analysis because it involves deep behavior analysis, assessing software behavior in real-time by executing and observing it during runtime. Authors in the study [35] point out that this behavioral-based analysis takes place in the Region of Interest (ROI), which constitutes a segment of the computing environment where file encryption occurs; this is to improve the security of the process. The extraction algorithm is applied to dynamically sample the trace files into smaller segments known as sliding windows. These sliding windows enable continuous monitoring of file behavior and interactions with the environment. By assessing the frequency of these interactions against predefined thresholds, the system identifies potential malware instances, delivering the capacity to combat obfuscation, polymorphism, encryption, and anti-disassembler; it can compensate for the static analysis limitations. To demonstrate the capacity of a behavior-based approach to zero-day detection, [36] focused on crypto-ransomware early detection models that safeguard users from being victimized in any attack.

3.3 Hybrid Analysis

Combines static and dynamic analysis for a more comprehensive outcome; it uses static analysis to identify malicious patterns, then uses dynamic analysis to confirm suspicions or uncover hidden malicious activities. You may see hybrid as combining the strength of static with the strength of dynamic to develop a robust and complete detection system. An automated ransomware detection method for email filtering, named R-Killer, was introduced by [37]; this approach utilizes a DL technique to implement a robust model, which tracks processes generated through email attachments to analyze the potential ransomware threat. The study, called the “R-Killer” system, gathered threat intelligence while preserving user data privacy and used the meta-analysis for email protection against cyber-attacks. The study [38] proposed the combination of signatures before dynamic analysis for malware detection in cloud environments. This research work was designed to enhance the volume of data generated within network environments, enabling the rapid classification of potential malware files.

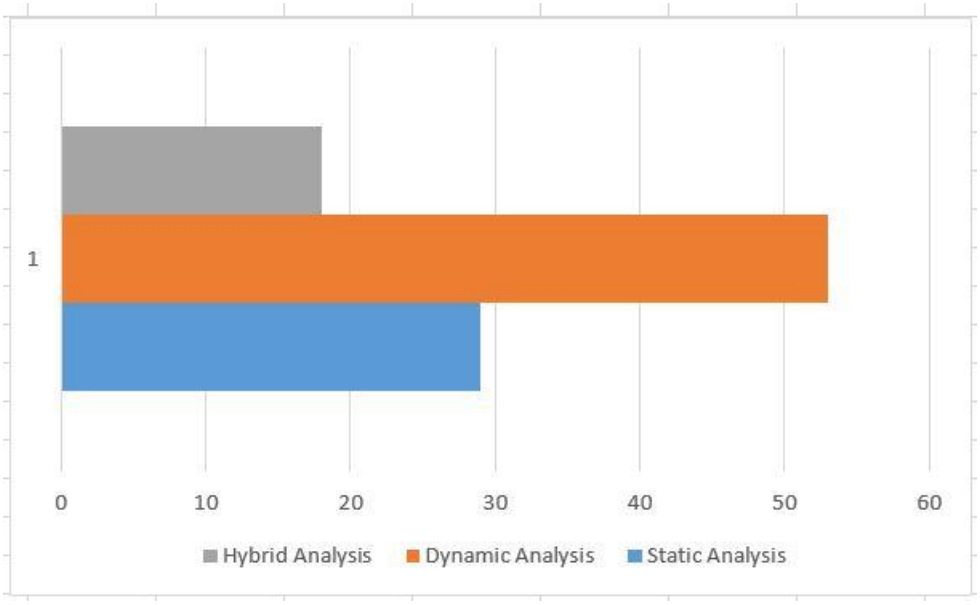

Fig 5 Distribution of detection and mitigation approaches in the literature study

Table 2 Comparative analysis of various ransomware detection techniques

Work	Mode of Analysis	Technique	Result	Research Gap
[39]	(PE) header files via the Xception (CNN) model through Static Analysis	DL	Was able to achieve 90% accuracy and recall of 100%	Only the PE header was used to extract features for ransomware analysis. Other detection indicators should be utilized.
[26]	Applied Dynamic analysis using Ai-based Sandboxing	ML	On-premise cross-validation using different ML classifiers. Recorded an F-measure of 93.7	It can be improved to detect Zero-day vulnerabilities + on a larger dataset.
[31]	Dynamic analysis using a pre-encryption algorithm	ML	RF and NB classifiers achieved an impressive 0.0156 FPR	Only used for crypto-ransomware, it may be prone to code obfuscation.
[40]	Static analysis using API calls	ML	Achieved a FNR of 1.3% with 99% accuracy	Prone to signature-based evasion techniques.
[41]	Dynamic analysis using Cuckoo sandbox	ML	Using memory dump analysis, they recorded an improved accuracy of 97.85% while maintaining a low % false positive rate of 2%.	Memory dumps analysis is not suitable for real-time detection.
[42]	Hybrid Analysis In sandbox	ML	Used the binary versus fixed parameters paradigm through sandboxing to detect a variant known as the W-32 dropper.	The Cerber ransomware group has many types but can only detect the W-32 dropper.
[43]	Static Analysis	FL	Using FL with API calls to train an ML model resulted in a 93.1%	It is only Windows-based. Future directions can enable cross-

			accuracy rate in identifying ransomware.	platform compatibility and diverse datasets.
[44]	Hybrid Analysis Using context-aware entropy Feature extraction	DL	accurately classified ransomware via context-aware analysis-based, high-precision recorded	the approach can be improved by using the kernel I/O analysis to Implement file filtering features in place of context-aware
[45]	Dynamic Analysis	DL	Used AI sandboxing environment to interact with binaries and analyze their behavior, obtaining zero false positives	The artificial sandboxing environment may not detect DLL hijacking.
[46]	Dynamic Analysis	ML	Applied Network Security principles for their analysis, recording a precision of 92.32% and recall of 99.97%, then an accuracy of 99.99% in detecting ransomware	The traffic patterns recorded may conceal exploits using stenographic methods.
[47]	Dynamic Analysis	ML	They Used network traffic analysis with a decision tree (J48) classifier and achieved a TPR of 97.1%.	Unable to identify different types of ransomware because it wasn't trained on enough data (limited dataset)

Table *2* summarizes various detection techniques, highlighting the modes of analysis, applied techniques, results obtained, and potential research gaps. At the same time, Figure 5 details the current trend in applying these approaches to malware analysis, detection, and prevention. Our findings review that about 53% of studies recently adopted the dynamic analysis approach to detect, while a little above half of that percentage, at 29%, still believes in the prowess of the signature-based static analysis method. Meanwhile, 18% of recent publications have combined dynamic and static approaches for more effective, robust, and complete results. The above summary shows that ransomware detection methods have significantly improved their effectiveness against significant attacks. State-of-the-art techniques now apply hybrid methodologies, often incorporating artificial intelligence (AI)-based strategies to enhance their efficacy. However, [48] maintains that despite these advancements, the new malware variant continues to evade detection because most existing techniques may be unable to address evolving ransomware strains

simultaneously. Most solutions are tailored to detect specific strains or types of ransomware, resulting in a lack of generic solutions due to the inherent challenges in developing such comprehensive systems. Therefore, future research should focus on developing robust and secure detection systems through the collaboration of cyber threat intelligence sharing and federated learning (FL)

4 Ransomware Trend Landscape

The threat of ransomware has consistently been ranked top by the European Union Agency for Cybersecurity (ENISA), as reported in [50]. Meanwhile, the Cybercrime magazine [49] predicts a disturbing increase in the global cybercrime index. They estimate a 15 annual growth over the next five years, reaching a staggering $10.5 trillion by 2025. The team also believes this dimension would be the most significant transfer of wealth ever seen, and it could discourage both innovation and investment because the projected damage is worse than the yearly cost of natural disasters and even more profitable than all the world's illegal drug trade combined. To understand the latest scope and tactics implored by cyber criminals, we have classified the current trend landscape into four main categories.

4.1.1 Ransom – Economy

The ransom economy era showcases how ransomware groups have progressed beyond mere execution of attacks to offering packaged software products and tools for fellow cybercriminals. Although this trajectory isn't entirely novel, the degree to which criminal entities have adopted it tells us the need to be concerned. This trend has intensified with ransomware as a service (RaaS), a model that streamlines the distribution and execution of ransomware attacks. It presents cybercriminals, including those lacking technical expertise, an opportunity to access sophisticated ransomware tools and infrastructure through subscription-based services or partnerships with experienced hackers. Figure 6 shows some commercially exploited ransomware and how they affected the globe, including how long they stayed, the victims, and the average number of victims recorded per 30 days before being discovered and resolved. RaaS providers offer a range of services, from customizable ransomware variants to user-friendly dashboards for managing attacks, significantly lowering the barrier to entry for aspiring cybercriminals, where anyone with financial motivation can engage in extortion activities with relative ease.

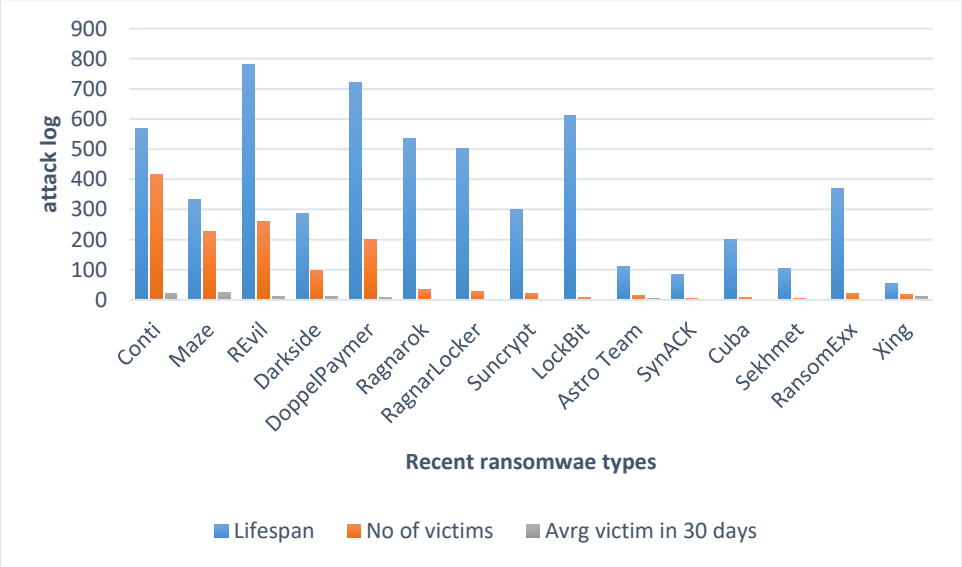

Fig 6 Notable commercially exploited ransomware with respective effects globally

Ransomware groups have adopted a more organized approach, employing sophisticated and structured strategies that mirror the cooperation and practices found in the conventional business world; for example, according to a Microsoft Threat Intelligence Report[50], one affiliate gang first used ransomware from Ryuk, Conti, and Hive before moving on to use malware from BlackCat. Another affiliate worked with Ryuk, REvil, and Conti and later worked with BlackCat. Authors in [51][52] out that this trend shows how hackers made more than $18 million from April 2014 to June 2015 when the CryptoWall ransomware was released, not also forgetting the noble Wannarcy2017, which exploited the vulnerability in the healthcare system by affecting hundreds of hospitals across the United Kingdom, creating mass distribution that led to over 19,000 appointments being canceled.

4.1.2 Data Monetization

Generally, the motive behind stealing or encrypting data is to extort the victim, but in current trends, stolen data is not just valuable to its rightful owners. Upon compromising a system and gaining access to a company's secrets and sensitive documents, hackers may monetize such information by selling it to the highest bidder. According to [53], this technique shows a shift in the trend and dynamics of cybercrime, wherein stolen data becomes a commodity with substantial market value, highlighting the multifaceted risks posed by such breaches to the targeted entities.

The study [54] stresses that this approach is often classified as “Double Dipping,” a trend illustrating the combination of ransomware attack techniques with other social engineering tactics to suppress the victim into falling prey. Usually, this scenario occurs when cybercriminals gain unauthorized access to an organization's data, subsequently encrypting it. In cases where the targeted firm hesitates to comply with ransom demands, often due to having backup systems in place for file restoration, the attackers resort to coercive measures. They may threaten to publicly disclose the pilfered data on the dark web or sell it to a third party; such actions can expose personally identifiable information (PII) and the organization's proprietary intellectual property, causing significant damage to its reputation, so the company may have no option than to pay even if they have a backup option available. A notable instance of this tactic was witnessed in 2019 at Allied Universal security staffing company, where the attackers demanded a ransom of $2.3 million. When the firm refused to pay, the cybercriminals threatened to use sensitive information extracted from the company’s system for spam impersonation.

4.1.3 Automated Approach

These days, even cybercriminals utilize automation’s ability to save time and resources. Like professional companies, cyber criminal gangs are trying to improve their efficiency by automating operations and minimizing human error. System penetration, the most time- and resource-consuming part of a ransomware attack, can now be streamlined through automated processes. This capability empowers groups with limited workforce or resources to execute attacks more efficiently and effectively. The study [38]emphasized that hackers increasingly leverage blockchain technology for attack methodologies. Also worthy of note is that the advancement and popularity of IoT devices have boosted this automation approach. [6]contends that the Internet of Things (IoT) device population is increasing continuously and is estimated to reach around 64 billion by 2025.

The past few years have seen an upsurge in cybercrime groups targeting these devices, especially in the industrial and agricultural sectors. [55] notes that these attacks range from relatively straightforward tactics, such as utilizing IoT devices as entry points to propagate ransomware across interconnected devices within networks, to more intricate strategies. Figure 7 shows the Top ten(10) ransomware attack targets in

2023 by country, with the United States being the most targeted, followed by the UK and Canada among the top 3. Meanwhile, Spain, Brazil, and India attracted minimal attention from the Darkweb.

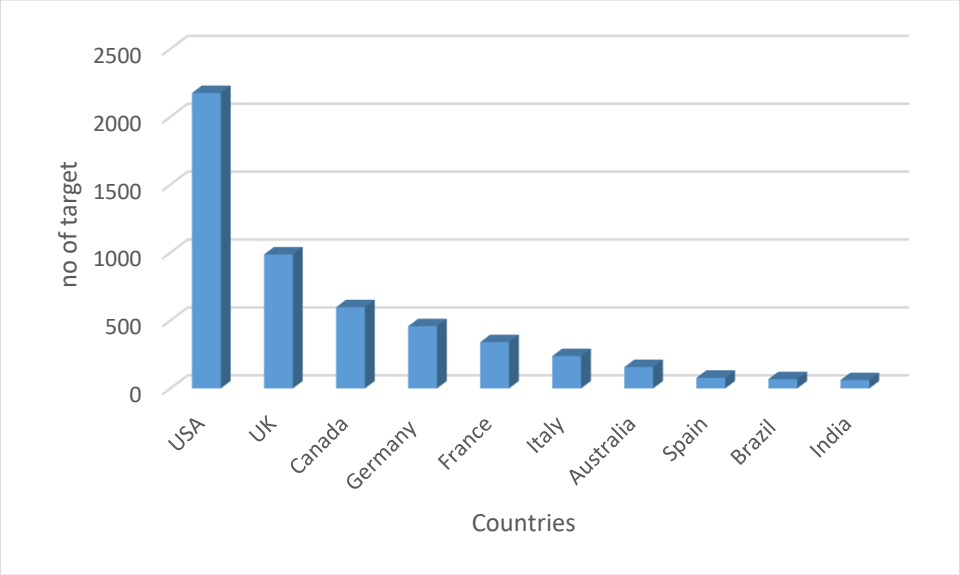

Fig 7 Top 10 ransomware attack targets in 2023 by country.

4.1.4 Exploiting Cloud

Ransomware in the cloud operates within a distinct paradigm. Unlike traditional ransomware attacks that rely on tricking users into running some malicious software to encrypt stored data, cloud-based ransomware often involves threat actors' exfiltration of data that has not been adequately secured, deleting the original files, and subsequently demanding a ransom for their restoration. If organizations do not have adequate backups, paying the ransom may be the only solution to regain access to the compromised data.

The growing reliance on cloud services has provided fraudulent individuals ample opportunities to engage in this malicious activity. As businesses increasingly store sensitive data in the cloud, attackers can target a larger pool of valuable traits. This trend, coupled with the ease of deploying attacks and the potential for substantial financial gain, has made cloud ransomware the next big deal in the industry.

More so, the anonymity facilitated by cryptocurrencies has also simplified the process of demanding and receiving ransoms, to some extent offering attackers a means to operate without fear of being traced. Google's Cloud Team [56] identifies that 86% of hacked cloud instances are used for cryptocurrency mining. Individuals already engaged in "crypto-jacking" might easily switch to using ransomware on infected systems or making money by selling access to more established groups.[57] maintains that the present cloud systems are often accidentally left exposed to the internet and are less secure than systems in a traditional IT system.

4.2 Challenges Facing Deployment of Detection Approaches to Industry

Previous sections have highlighted the integral role of several features in understanding the complex landscape of malware detection, epitomizing the synergy of modern analytical methodologies. However, a significant problem facing malware detection methods is the difficulty in industry deployment. The study [28] emphasizes that numerous impressive ransomware approaches have undergone proper validation. However, many of these have not been deployed to industry because of limited real-world validation. This

lack of deployment is a critical challenge, as deploying these techniques is crucial to ensuring their robustness. The reasons for this inadequacy can be categorized into three aspects.

4.2.2 Diverse and Evolving Threats

The nature and complexity of ransomware are constantly changing with real-world scenarios; cybercriminals are continuously developing new techniques to make their attacks more sophisticated and efficient and also to evade detection approaches; as a result, a detection technique that performs well in a research environment may face difficulties in keeping up with the rapidly evolving malware landscape. Therefore, if existing malware detection approaches are not thoroughly validated in real-world applications, then it can not be certain that such techniques will detect sophisticated malware and fail industry validation.

4.2.3 Cost of Deployment

Deploying new technology in an organizational environment generally involves significant resource constraints, including computing power, storage, and physical space. Additionally, the organization may need to modify its architecture or framework, including changing or upgrading its network firewalls, introducing a new Intrusion detection or prevention apparatus, or deploying an overall Security information and event management(SIEM) to accommodate the intended new deployment. [29] affirms that deploying the malware model as a new technology in an organization will always be time-consuming and resource-intensive, necessitating substantial computing power and storage. This challenge justifies why many researchers are unable to deploy their proposed methods.

4.2.4 scalability

Integrating malware detection methods into real-world scenarios presents a significant scalability challenge. Many detection approaches rely on small datasets generated in simulated environments, which may not accurately reflect the characteristics of malware in actual settings. This challenge is critical because scalability involves processing larger volumes of data and adapting to new threats and evolving attack techniques. Inadequate datasets hinder the system's ability to generalize effectively, limiting its ability to detect novel and unknown malware variants. [28] explains that a contributing factor is that security researchers cannot access most real-world datasets used for developing malware detection systems. Future directions should focus on collaborating with industry partners, threat intelligence providers, and open-source communities to facilitate access to diverse and representative datasets for malware detection.

5 Federated Learning Concept

FL technology uses ML to preserve the model's privacy by enabling multiple devices to collaborate on a central project without revealing individual data. Each client has its own training and test datasets. Figure 5 depicts a basic FL model consisting of an aggregator, or central server, that distributes the global model to a group of selected clients. These clients then train their respective models locally using their domain dataset in an iterative process. After these steps, the training outcomes are subsequently sent back to the server for aggregation and refinement by the global model

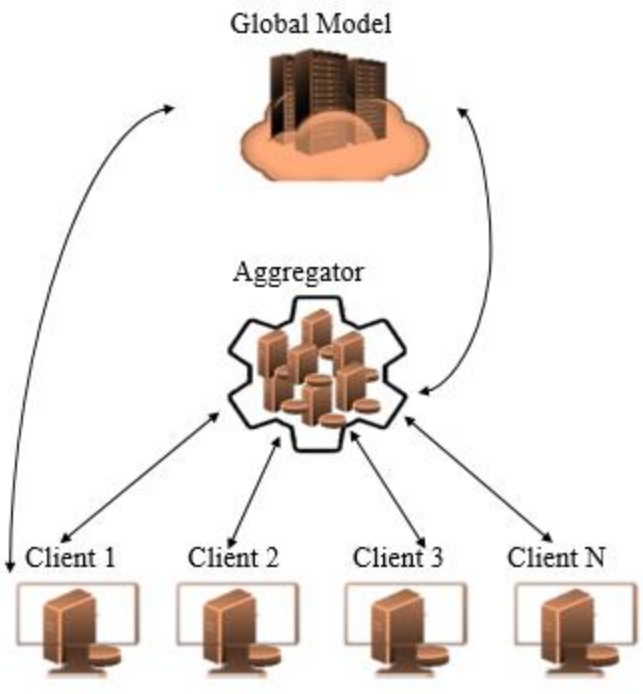

Fig 8 A Federated Learning Model

This collaborative approach has become increasingly crucial as ransomware attacks continue to pose significant challenges to the cybersecurity community, with cybercriminals continuously advancing their techniques, leading to more sophisticated and diverse ransomware attacks. The FL concept has been applied to overcome these sophisticated attacks by improving the detection and mitigation of ransomware. It offers the benefit of cooperation, sharing, and utilizing relevant threat information. The in [43] proposed FLDetect proposed FLDetect, a unique FL-based method for identifying ransomware on Windows machines using API calls with an open-source dataset known as ransomwaredataset2016; they achieved an accuracy of 93.1%. By using feature extraction to classify malware, [58] applied a federated malware classification approach, which focused more on classifying new malware variants with an impressive area under the curve (AUC) of 0.9270 on the dataset provided by VirusTotal. To identify malware IoT-enabled devices, [4] introduced the SIM-FED method, which attained an impressive accuracy rate of 99.52%. Ultimately, the quest for a permanent solution to ransomware has driven a substantial surge in academic literature in this area. However, despite these efforts, a permanent and comprehensive solution that can curtail the menace of ransomware remains a dream.

5.1 FL Aggregation Models

The effectiveness of an FL model relies on the proficiency of its update aggregation, which leads to a global model that securely combines the data from all participants. This essential function is critical for the system's performance and operational efficiency. The statutory objective of the FL aggregation mechanism remains the improvement of decision-making for a group of participants, denoted as β, by reducing the attribute of its loss function β_k of each participating client. These local clients, built with weights W_k and trained on private datasets D_k containing n_k samples, become integral to the aggregation process.

The First version of federated learning aggregation, known as **FedAvg**, was introduced by Google's team in 2016 in conjunction with the launch of Federated Learning technology [59]. This framework

randomly selects several micro-class members for aggregation in a training round. Then, in the aggregation process, the parameters of each client's model are weighted and averaged to create a global model, with the weighting factor based on the client's data volume. Essentially, the participants are trained using batch gradient descent over a single local epoch; when the training phase is complete, the model transfers their gradients to the central server, collates the results, and averages them before finally updating accumulated global weights using the gradient descent algorithm. Despite its advantages, FedAvg has its challenges; specifically, it requires the client models to train their dataset using just one local epoch and a batch, affecting its speed and requiring hundreds of communication rounds before the desired accuracy can be attained.

In addressing the challenges with FL aggregation, several models that seek to improve the functionalities have continued to evolve. In 2018, [60] proposed **FedMeta.** This model significantly addresses the limitations of FedAvg by utilizing a meta-learning approach on a collection of client tasks, enabling it to solve new tasks with minimal samples. FedMeta operates through a two-step process in each federated learning communication round, which takes place in two phases: the inner update and the outer update. During the inner update, class members train their respective models using their domain-owned dataset and the global model's weight vectors. Within this process and period, the outer paradigm connects with the central server to distribute all parameters to the participants to complete the communication cycle.

Subsequently, in 2021, [61] introduced the **FedDist.** This innovative FL aggregation method applies Human Activity Recognition (HAR) to resolve divergence issues experienced in both heterogeneous and non-independent and identically distributed (non-IID) datasets. This approach usually aggregates the clients using FedAvg before applying pairwise dissimilarity between neurons in class members' local and global layers. When the dissimilarity exceeds a specific threshold score, an activation neuron feature is added to the main model; otherwise, the request is rejected. Clients then conduct layer-wise training, freezing updated layers and continuing with subsequent ones. Evaluation of the framework showed significant improvement as it outperformed Fedmeta and other state-of-the-art approaches in handling non-IID data. Nevertheless, a critical disadvantage of FedDist is its considerable communication cost, as including neurons in every FL round raises the workload on both clients and the server.

To resolve data and model heterogeneity [62] introduced the **FedGA**, in which, unlike the conventional FL where participants share part of their private data with the main model, the framework of FedGat only permits sending the base layer weights to the central server before applying a genetic combinatorial algorithm to collate the accumulated weights across all class members. With this principle, it minimizes the model's loss function. FedGAt demonstrated faster convergence and enhanced accuracy compared to other methods while reducing communication costs. However, critics believe genetic algorithms are more complex than methods such as direct weight averaging.

The **FMTDA** approach, proposed by [63], tackles the domain shift problem caused by the uneven distribution of local datasets. In this framework, the local devices of participants serve as target domains, which are usually made up of datasets that are not labeled. At the same time, a central server manages the labeled dataset. The framework aims to improve the accuracy of class members by aggregating their locally trained convolutional neural network (CNN) models. This method strategically focuses on balancing local adaptation and global model consistency, ensuring improved recognition performance across heterogeneous datasets while mitigating the impact of domain shifts. Combining the maximum classifier discrepancy (MCD) technique and the Gaussian mixture model (GMM) allows for efficiently handling statistical discrepancies among diverse local datasets, facilitating robust and accurate model aggregation.

To mitigate against Byzantine poisoning attacks [64], Introduced the **Split Aggregation**: This framework was designed to optimize FL aggregation efficiency and accuracy by encrypting and dividing

user gradients, employing an adaptive weighting strategy for aggregation, and utilizing Randomized Singular Value Decomposition (RSVD) to balance computational efficiency and accuracy. The model's efficacy is demonstrated through experimental results, which showed the ability of the framework to prevent erroneous discarding of honest user gradients, outperforming existing robust and privacy-preserving FL methods in both computational complexity and communication overhead. However, the approach relies on a dual-server architecture, which entails that the servers must not collude; otherwise, data privacy and security will be compromised.

The overview of the various aggregation methods we have highlighted demonstrates that each strategy was initially established to address the numerous issues encountered in the aggregation paradigm. These strategies are specifically developed to improve one specific aspect while upholding the overall principles of the entire system.

5.2 Application of Federated Learning to Ransomware Detection and Mitigation

The rapid growth of digitization and electronic transactions has made sensitive data of individuals and organizations perpetually vulnerable to hackers and intruders, particularly with the rise of IoT-based applications. Therefore, adopting an FL approach has become crucial to collectively train and manage the many aspects of threats arising from general malware and ransomware across our devices, applications, and networks. FL facilitates collaboration among security stakeholders towards a shared objective and guarantees data privacy, security, decreased latency, reduced power consumption, and on-device training. Additionally, FL enables the delivery of personalized ML models to users, who then learn collectively to enhance the user experience. This section highlights the application of FL to general malware and ransomware detection, prevention, and mitigation.

An innovative DeepFed algorithm was proposed by [65] to address the challenges associated with heterogeneous and large-scale related cyber-physical systems due to the scarcity of attack data. This algorithm FL enhances IDS capabilities by combining a neural network (NN). This architecture facilitates the development of a collective IDS by consolidating knowledge from numerous industrial CPS environments. A secure communication protocol based on the Paillier cryptosystem was implemented to guarantee the security and privacy of the training process. The experimental results demonstrated the proposed model's effectiveness and robustness in detecting and mitigating intrusions conducted on a real-time industrial CPS dataset.

[11] proposed an FL system for ransomware botnet detection utilizing an autoencoder model. The approach involves collecting IoT network traffic at edge devices, which host local models and virtual workers. These local class members were trained on the data of the edge device before combining the updates. The global model then transmits the modified parameters back to the edge devices, which train new local models. This decentralized technique made it possible to provide data privacy while harnessing collaborative learning. The system displayed a high efficacy, obtaining 99% accuracy in categorizing IoT traffic as benign or malicious, making it a scalable and safe solution for ransomware detection in IoT networks

Following the concept, FLDetect was introduced by [43], leveraging FL to detect ransomware on Windows machines. This system employs a distributed architecture, where models are trained locally on devices, allowing for enhanced privacy and data security while analyzing API calls from Windows systems. FLDetect utilized an open-source dataset, ransomwaredataset2016**,** which contains a collection of ransomware-related API call logs. The system involved advanced feature extraction techniques to classify malware, focusing on differentiating between benign and malicious activities based on API behavior. With this method, FLDetect achieved a commendable accuracy of 93.1% in identifying ransomware attacks, demonstrating its potential for real-world applications, although the dataset contained about 600 samples,

which is not enough. However, this FL-based system offered the dual benefit of high accuracy in detection while preserving user privacy, as sensitive data never leaves local devices during training.

[66] introduced FedA-GRU, a novel detection mechanism utilizing the neural network technique to model the FL environment for malicious intrusion. The approach is particularly tailored to enhance the security of wireless edge networks and safeguard them against malware attacks. In contrast to traditional centralized approaches where data across the devices are broadcast to a global server, FedA-GRU updates the global model by sharing just the parameters learned from each edge device, thus ensuring data privacy. To further boost the system's efficiency, the researchers devised an attention mechanism that prioritizes updates from critical devices while filtering out less significant ones; this concept was a game changer for reducing wasteful transmission. Finally, the approach minimized the communication overhead and promoted faster convergence of the learning process.

In the research work [58], the authors employed the Artificial Neural Network (ANN) technology for intrusion detection in the IoT-based healthcare paradigm. The approach utilized the distributed nature of FL to improve security without centralizing sensitive medical data, thus preserving patient privacy while enabling robust malware detection and prevention capabilities. The authors demonstrated significant improvements in handling the heterogeneous nature of IoT data, which often varies across devices and environments in healthcare settings. Moreover, the system effectively mitigated poisoning attacks, a common security threat in controlling malware within large databases, where adversaries attempt to corrupt the process by injecting malicious data. By enhancing performance and resilience against such attacks, their ANN-based approach ensured a promising solution for securing IoT healthcare networks, where data diversity and security are critical concerns.

The application of FL is extended to agricultural IoT environments by [67], where the authors implemented three distinct global models, namely Convolutional Neural Network (CNN)**,** Recurrent Neural Network (RNN)**,** and Deep Neural Network (DNN) for malware intrusion detection and mitigation within hierarchical FL (HFL) architectures. Through this research work, the authors analyzed and reported the performance of these models across three diverse IoT datasets, addressing the unique challenges presented by agricultural IoT systems, such as the heterogeneity of data sources and devices. They leveraged the hierarchical FL model to improve scalability and efficiency by organizing devices into layers, thereby reducing communication overhead and enhancing model performance. The use of multiple neural network architectures ensured a comparative analysis of their effectiveness in identifying malicious activities across different datasets. The experimental results proved the versatility and robustness of these models in securing IoT networks in agriculture, a sector increasingly reliant on IoT technology for precision farming and resource management.

The authors in [68] pioneered using a privacy-preserving FL framework to identify Android malware using API calls and permissions. The proposed system employed secure multi-party computation methods and a Support Vector Machine (SVM) to improve malware identification. Importantly, this study was acknowledged as the first implementation of an FL-based system for Android malware detection, representing an essential advancement in mobile security.

Following the experience of [68], the research work in [69] came up with a framework called Less is More (LiM)**,** an improved FL-based malware detection and classification system. An essential factor of this paradigm is that it can securely maintain the privacy of other installed programs on a mobile device. The researchers applied a semi-supervised ensemble technique to increase the performance of the detection and classification tasks. Their empirical findings indicated the framework's efficacy, as the approach attained an impressive F1 Score of 95%, and local class members maintained a minimal false positive rate of 1% across a dataset of over 50000 malicious and benign samples across 100 federated applications.

In their study, [58] developed a federated malware classification approach targeting identifying new malware variants in Internet of Things (IoT) devices. This method leverages FL, allowing IoT devices to train models locally without sharing sensitive data, thereby preserving privacy and enhancing collective network detection capabilities. Using the VirusTotal dataset, the approach achieved a robust 0.9270 AUC score, indicating high performance in distinguishing malicious and benign software. The emphasis on classifying novel malware variants is particularly significant, given the fast-evolving nature of Ransomware threats in IoT environments. Application of FL, this method also contributed to addressing the privacy issues, scalability, and the need for decentralized data handling, making it a valuable contribution to securing IoT ecosystems against malware attacks.

Similarly, the study by [4] introduced the SIM-FED framework, an FL-based approach that achieved an outstanding accuracy of 99.52% in detecting ransomware. This high-performance model further illustrated the potential of FL to significantly improve ransomware detection by enabling distributed systems to collaboratively learn from diverse datasets without compromising privacy. The continuous prevalence and sophistication nature of ransomware remain a motivation for the substantial surge in academic literature as efforts continue to intensify to find comprehensive solutions. However, despite the impressive progress in FL, DL, and encryption-based defenses, a permanent and all-encompassing solution to the ransomware threat remains elusive. The continuous evolution of ransomware tactics and the complexity of securing heterogeneous systems such as IoT networks and cloud infrastructures present ongoing challenges for researchers and practitioners. As a result, the quest for a long-lasting solution remains an active area of exploration. Although the use of FL offers numerous benefits, as we have summarized, with more annotation in Table 3, it also presents some challenges; we shall discuss these in a subsequent section and provide probable future directions related to malware detection and mitigation

Table 3 Application of FL to Malware/Ransomware detection and mitigation

Authors	Features	Model	Domain	Dataset	Evaluation
[11]	Network Logs	Autoencoders	IoT	N-BaIoT	99% accuracy
[68]	Sequence API calls	SVM	Andriod	OperaStore	95,92% F-score
[43]	API call System logs	ML	Windows	ransomwaredataset2016	93.01% accuracy
[67]	System calls	ANN	IoT_Health	Ember	96.00% accuracy
[70]	Permissions Broadcast receivers	LiM	Andriod	Androzoo	95% F-score
[4]	traffic logs	CNN	IoT	IoT-23 dataset	99.52% accuracy
[71]	Static API sequences	CNN	Andriod	Androzoo	97.89% and 94.39% accuracy

[12]	Behavioral features	DW-FedAvg	Andriod	Kronodroid, Drebin, Melgenome	99.18% F-score

5.3 Common Challenge Facing FL application to Malware Detection, prevention, and Mitigation

Data and transaction security continues to be a crucial issue in FL applications. While existing centralized ML frameworks have created significant concerns regarding the secure transfer and storage of user data, FL offers a decentralized solution by storing user data on local devices, decreasing the danger of disclosing sensitive information. However, this strategy creates new issues, particularly in safeguarding the model updates, such as weights and gradients, which are shared between devices and central servers. These shared parameters, although not direct data, can nonetheless be vulnerable to attacks through inference, data leakage, data poisoning, and model inversion. Cybercriminals can reconstruct sensitive information from the model updates from these loopholes.

5.3.1 Membership Inference Attack(MIA)

According to [72][73], inference attacks involve an adversary attempting to determine if a particular sample data or class member participated in the training. This attack exploits the information embedded in the model's parameters, such as gradients or weights, shared between the local members and the global server in an FL environment. Given the above analysis, an adversary actor may analyze the model's output or the sensitivity of gradients to compromise the training. Mathematically, given a model $f_\theta(x)$, the adversary tries to assess if the information on point x' which represents the training set by observing the output at $f_\theta(x')$ and comparing it against predefined thresholds. These attacks pose a significant privacy threat, particularly when the model is overfitted or highly confident in its predictions.

5.3.2 Communication Overhead

Communication overhead in a federated environment is particularly critical in large-scale applications such as malware detection, prevention, and mitigation. In a typical FL setting, each device (or participant) computes local model updates and sends these updates (e.g., gradients or model weights) to a global aggregation model. This to-and-fro movement between the clients and the central server can create significant overhead, particularly when operating in a distributed network of devices like IoT systems, where bandwidth is often limited. The size of the model being trained significantly influences communication costs. Large models with millions of parameters result in substantial data transmission per round of training, slowing down the system and increasing bandwidth consumption. This issue is particularly prevalent in malware detection tasks, where DL models, such as CNNs or transformers, are regularly used due to their high performance, thereby increasing rounds of communication afront between clients and the server to achieve convergence if each round requires transmitting extensive model updates, the total communication cost becomes a major concern.

5.3.3 Data Leakage

Data leakage occurs when an attacker intercepts communicable parameters, like shared weights or combinatorial gradients, to infer and reconstruct the original training data. Authors in [74][10] have demonstrated that FL is vulnerable to gradient leakage attacks, where adversaries exploit the information embedded in gradients to reverse-engineer sensitive data. These can be expressed mathematically using $f_\theta(x)$ to represent the global model with parameters θ

While $\nabla_\theta L(x)$ denotes the loss function gradient in relation to the input x of the participant. The adversary can observe and interfere in the shared gradients $\nabla\theta Lx$ using some optimization techniques to iteratively minimize the difference between the actual gradients and those computed from the reconstructed data $\hat{x}$. This process is formalized as:

$$\hat{x} = \arg\min_{x'} \|\nabla_\theta L(x') - \nabla_\theta L(x)\|$$

where $\hat{x}$ represents the reconstructed data that closely approximates the original input x. As highlighted in research work [75], this attack severely compromises participant privacy in FL environments, significantly when models are updated frequently and the gradients contain rich information about the underlying data.

5.3.4 Data Labeling and Quality

In general malware analysis and classification, completely labeled data is considered quality and crucial for modeling effective detection models. Using the FL environment as a focus, devices often contain varying proportions of labeled and unlabeled data, which can negatively impact model performance. In recent times, for cybersecurity researchers, analyzing and labeling malware data has been particularly challenging due to the constantly evolving nature of malware and the complexity of distinguishing between benign and malicious behaviors. Recent studies have explored Semi-Supervised Learning (SSL) and Self-Supervised Learning (Self-SL) approaches to mitigate this problem. For instance, authors in [76]demonstrated how combining SSL techniques with FL improves malware detection accuracy by harmonizing a well-labeled dataset.

5.3.5 Data poisoning

Data poisoning occurs when a participating model injects falsified or corrupted data into the local training process, intentionally or inadvertently, to manipulate the global model's performance and inadvertently pose a danger in dealing with ransomware detection. This concept is explained meticulously in the study by ([19],[75]). In FL, where local member updates are collated to form a central update, an adversary can corrupt the local dataset x_i or alter the gradient updates $\nabla_\theta L(x_i)$, leading to skewed global parameters. This behavior can degrade the model's accuracy, introduce biases, or even insert backdoor vulnerabilities., by corrupting the local gradient $\nabla_\theta L(x_i)$ that contributes to the global update($\theta_{t+1} = \theta_t - \eta \sum_{i=1}^{n} w_i \nabla_\theta L(x_i)$. Techniques such as Byzantine fault tolerance**,** robust aggregation methods (e.g., Krum or Trimmed Mean), and outlier detection are essential in mitigating these attacks, ensuring the global model remains robust against poisoned data contributions.

5.3.6 Model inversion attacks:

In this scenario, an adversary exploits the shared model updates, such as gradients or weights submitted by participating clients to the central aggregator, to reverse-engineer sensitive information about the original training data[71] points out that, Unlike traditional data breaches, model inversion targets the inherent relationships captured in the learned parameters, allowing the adversary to approximate or reconstruct features of the original data. Mathematically, let $f_\theta(x)$ represent the global model with parameters θ, and let $\nabla_\theta L(x)$ denote the loss function gradient as it relates to a class member at point x. In a model inversion attack, the adversary, having access to the aggregated gradients $\nabla_\theta L(x)$, can apply optimization techniques to reconstruct x iteratively; this concept is represented by the entire equation

$$\hat{x} = \arg\min_{x'} \|\nabla_\theta L(x') - \nabla_\theta L(x)\|$$

where $\hat{x}$ represents the reconstructed approximation of the original input x. This attack is particularly concerning in privacy-sensitive environments, as it allows the extraction of private data without direct

access to the member's information. Mitigation techniques like the secure multi-party computation, discussed in the sections below, are often used to obfuscate gradient information and reduce the risk of model inversion, ensuring the security and privacy of participants' data in FL systems.

5.4 Solutions for Future Perspectives

5.4.1 Secure Multi-Party Computation (SMPC)

SMPC is an innovative method that enables multiple entities to collaboratively compute a function over their private inputs without revealing those inputs to one another. This concept is mainly carried out by partitioning individual tasks into independent entities before they are executed and before securely aggregating the results. In malware detection, prevention, and mitigation, Future directions must consider adopting and improving SMPC as it plays an important role in safe-keeping relevant information of the model while facilitating collaborative training. It allows class members to carry out training on their domain data. In addition, leveraging SMPC techniques like secret sharing, where class member updates, including gradients, are transformed into distributed secrets among all members, offers even more comfort and trust.

5.4.2 Homomorphic Encryption (HE)

Homomorphic encryption refers to an encryption apparatus uniquely built to leverage diverse computations on encrypted data without decryption, thus preserving data privacy throughout the process. Malware detection, prevention, and mitigation ensure that sensitive data such as logs, static or dynamic behavioral parameters, network traffic, or system metrics remain encrypted while allowing collaborative model training across distributed devices. This approach becomes especially valuable when privacy and data security are paramount, such as when dealing with proprietary malware signatures or sensitive user behavior data. HE enables mathematical permutations to be encrypted so that the information will be reviewed without explicitly decrypting the data, ensuring that sensitive data is never exposed during the computation. This concept guarantees robust privacy and prevents adversaries from inferring private information from the model updates, even if they gain access to intermediary results.

Two common variations of HE exist; they are Fully Homomorphic Encryption (FHE), which mainly indicates arbitrary permutations with encrypted data as discussed by the research work [77], and we also have Semi-Homomorphic Encryption (SHE), which enables specific operations, like addition and multiplication, on ciphertexts [78]. Future research needs to build upon the contemporary approaches by [79],[80],[81] to enhance the multi-party computational capability since there are still challenges associated with the current, such as the computational and communication overheads associated with cryptographic operations.

5.4.3 Differential Privacy (DP)

Differential Privacy is a powerful technique for preserving individual information while analyzing or sharing data, especially on a large scale. The principle of DP involves adding randomized noise global parameters to mask sensitive information while maintaining the statistical integrity of the aggregated data. Techniques like the Laplace mechanism [82] and the Gaussian mechanism [83] are commonly used to achieve this. Consider an FL system involving n class members, with each contributing its quota to the updates on local environment Δw_i computed on their domain data D_i. To incorporate DP, each participant i generates a noisy model update Δw_i^{DP} using the formula:

$$\Delta w_i^{DP} = \Delta w_i + \text{Lap}\left(\frac{\Delta f}{\epsilon}\right)$$

Where $\text{Lap}\left(\frac{\Delta f}{\epsilon}\right)$ represents the Laplace noise, Δf is the loss function's sensitivity, and ϵ is the privacy inclination budget, representing a positive integer that controls noise addition. The parameter sensitivity Δf reflects how much a single data point can affect the model's output, and its precise determination is crucial to balance privacy and utility. Once all participants compute these DP-protected model updates, they are sent to the central combinatory for updates:

$$W_{\text{new}} = W_{\text{old}} + \frac{1}{n}\sum_{i=1}^{n} \Delta w_i^{DP}$$

W_{new} denotes the update while W_{old} refers to the pre-update state. By incorporating noise at the participant level, Local Differential Privacy (LDP) ensures that sensitive details from individual datasets are obscured, even from the central server, making it especially useful in untrusted environments. In general contrast, DP applies noise at the server level after aggregation, which assumes a trusted aggregator.

5.4.4 Model Regularization and Pruning

Reducing model overfitting through techniques like dropout, weight regularization, or model pruning is a way forward for enthusiasts in this field, as it can help mitigate the risk of Membership Inference Attacks (MIA). By preventing the model from learning exact data patterns specific to individual participants, it becomes harder for adversaries to make inferences about specific data samples. Moreover, MIAs are often exacerbated when models overfit individual participants' data, making it easier for adversaries to infer participation. Future research should prioritize advanced regularization techniques (e.g., dropout, weight decay) to reduce model overfitting. This adaption will help reduce the model's sensitivity to specific data points, limiting the effectiveness of MIAs.

5.4.5 Adversarial Training

Integrating adversarial training techniques where the model is trained against simulated inference attacks could improve its robustness. By adversarially augmenting the training process, the model can learn to resist attempts at inference. Future work should explore adaptive adversarial defenses tailored specifically for FL systems focused on malware detection and prevention.

5.4.6 Benchmarking and Standardization

The field would benefit from standardized benchmarks and datasets designed to test the effectiveness of privacy-preserving techniques against MIAs in cybersecurity-focused FL applications. The MITRE ATT@CK framework is an excellent example of a standardized framework that can be used for enhanced ransomware detection, primarily through collaborative initiatives. Future works should also consider establishing standardized evaluation metrics to help researchers compare approaches and optimize defenses against MIAs more effectively.

5.4.7 Byzantine fault tolerance (BFT)

This approach ensures that the system can withstand adversarial updates by limiting the influence of malicious participants. In addition, aggregation methods like Krum and Trimmed Mean are designed to filter out extreme or deviant updates that deviate significantly from the majority, reducing the likelihood that poisoned data will skew the global model. Krum selects the local update closest to most other updates, while Trimmed Mean discards a fixed number of outlier updates before averaging. These methods help

maintain the model's integrity by ensuring that outliers, whether malicious or erroneous, do not disproportionately affect the final global update.

5.4.8 Combining Privacy-Preserving Techniques

Rather than relying on a single privacy-preserving technique, future research should explore combining methods like SMPC, DP, and HE to create a layered privacy approach. Each technique provides different levels of protection, and a combination can address different types of attacks while balancing efficiency and privacy. For instance, DP can protect against MIAs, while HE can secure data during aggregation, and SMPC can prevent direct access to sensitive updates.

5.4.9 Adapting Cryptographic Protocols for Real-Time Applications

One of the significant challenges in cybersecurity applications like malware detection is the need for real-time response. Cryptographic protocols such as SMPC and HE tend to introduce delays, which can impede the system's ability to respond quickly to threats. Future research should prioritize the optimization of cryptographic protocols for speed and scalability, ensuring that FL systems can operate effectively in real-time scenarios. Techniques such as batch processing of encrypted data or parallel processing of homomorphic operations could help improve the efficiency of these cryptographic protocols in time-sensitive applications.

5.4.10 Outlier Detection

The Outlier concept is crucial in identifying and excluding poisoned contributions. By flagging anomalous updates based on statistical deviations or unexpected patterns in gradient behavior, the system can reject potentially harmful inputs before aggregating them into the global model. As we advance, research must focus on improving the scalability and efficiency of these defense mechanisms, particularly in terms of malware detection in large-scale, distributed IoT networks. Combining these techniques with other security-related concepts like SMPC or DP may offer a more comprehensive solution, ensuring the model is secure from poisoning attacks and protects participants' sensitive data.

6 Method of SLR

This review followed the established guidelines for SLR, namely Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), also inspired by the review techniques presented in[84][85], which comprised:

- Formulating clear questions: we defined several research questions to guide our search.
- Systematic searching: applied a methodical approach to finding relevant studies across various sources.
- Extracting critical information: Once we found the studies, we carefully extracted essential data from them.
- Following our SLR design in Figure 9, we synthesized, analyzed, and extracted data using seven(7) steps: Step 1 involves formulating some research questions that connect with our topic, steps 2 and 3 are concurrent, and they incorporate all the searching strategies and refining processes using keywords, search string, search process, and source selection, in step 4 we started the selections criteria, and that continued with the data extraction in step 5 and snowballing in step 6, then the final step is the qualitative assessment in which we summarized the overall findings of the reviewed studies.

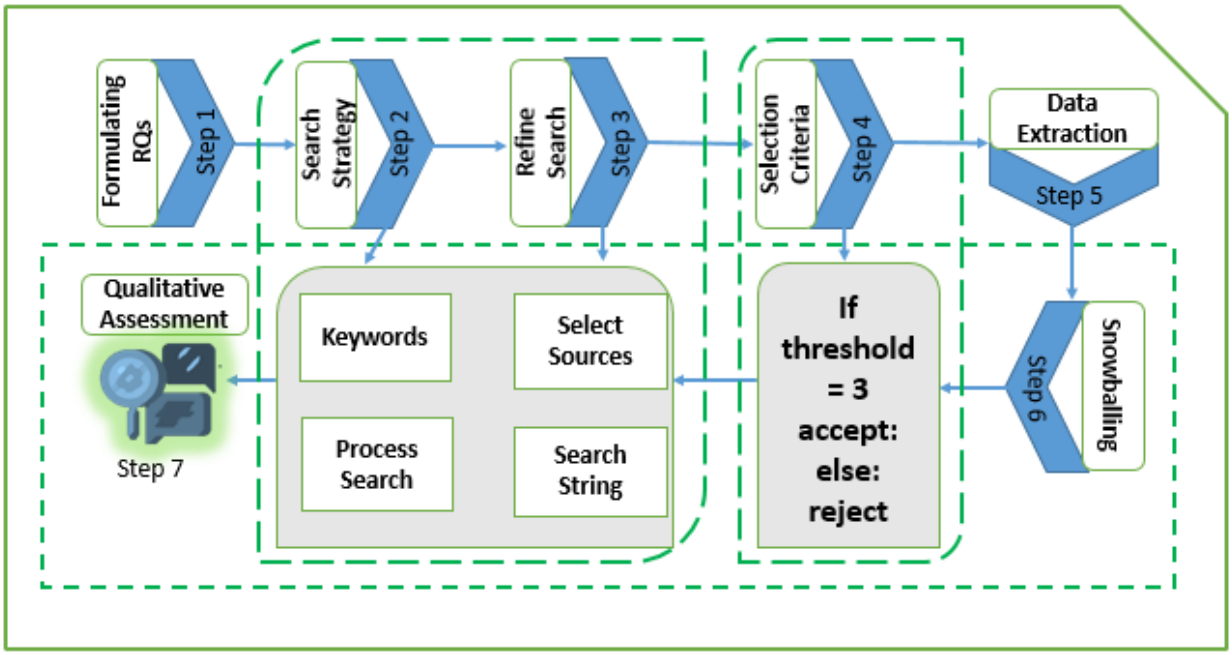

Fig 9 SLR Design showin the various stages of the research process

4.1 Formulating Research Questions(RQ)

To present a comprehensive review of the current state of research in our focus area, we formulated some critical research questions; details are in **Table *4***. These questions form a guide to the literature of information we seek.

Table 4 Information on RQs

No	Research Question	Objective and Justification
RQ1	How has FL been applied to ransomware detection?	This Research question allowed us to navigate the various approaches that have utilized FL to improve or enhance ransomware detection and mitigation.
RQ2	What are the common challenges facing FL applications in Ransomware Detection and Mitigation?	To properly guide future enthusiasts who may wish to apply federated concepts into the domain of ransomware detection, this research question ensured that the most common likely challenges they should be worried about are discussed.
RQ3	What is the current ransomware threat landscape?	With this research question, we try to analyze and highlight the current trend in ransomware's business, including its distribution and propagation.
RQ4	What are the existing ransomware detection techniques, and what	RQ4 seeks to investigate the recent performance of various approaches applied in ransomware detection and mitigation; it

are the challenges of deploying them across industries?	will also highlight why these approaches are barely deployed in the industry setting.

4.2 Search Strategy

The structured search approach comprises four key stages: keywords, Search string, Selecting source, and Processing. **Table *5*** shows our search keywords, initiated with a broad, meticulous, and extensive exploration of literature across online databases such as Taylor & Francis, Scopus, ScienceDirect, et cetera. Additionally, our search extended to Google Scholar to incorporate literature not covered by the databases, given its popularity as a supplementary database. We coined the search keywords from the research question we formulated earlier to retrieve accurate results relevant to our target topics.

Table 5 Keyword Strings

From RQ	Search keyword Coined
RQ1	"Application," "Federated," AND "Learning" AND "Ransomware," OR "Malware" OR "FL" OR "Collaborative," OR "Distributive" OR "Decentralized," AND "Detection"
RQ2	"Challenges" OR "Facing" AND "Federated," AND "Learning" AND "Application" "Ransomware" AND "Detection" OR "Malware," "Techniques" OR "Method" OR "Approach" OR "Framework"
RQ3	"Ransomware" AND "Trend" OR "Threat" AND "Landscape" OR "Current" OR "Latest"
RQ3	"Ransomware" OR "Malware" AND "Detection" OR "Classification" OR "Mitigation" OR "Prevention" OR "Approaches" OR "Methods"

4.3 Refining Search(RS)

Refining search involves optimizing our search strings and keywords, annotated in stages 2 and 3 of our SLR protocol design, and more information in Table 4. For example, our protocol recognizes that using keywords alone is insufficient for effective searching and recommends combining them into a search string. However, some databases do not support this feature, so we remove the quotation marks and rephrase the keyword when searching in such databases.

Table 6 Results of Refining Search(RS)

S/N	Database	Articles before RS	Articles after RS
1	Scopus	2345	758
2	IEEE explore	1602	341

3	Wiley online library	340	169
4	Springer	237	97
5	Science Direct	118	78

4.4 Selection

This stage involves the application of specific criteria in including or excluding an article, as shown in **Table *7***, which are carried out to ensure that we achieve our review objective. The purpose of these selection criteria is to ensure that they remain up-to-date with the most recent advancements in the area.

Table 7 Inclusion-Exclusion Criteria

Inclusion	Exclusion
Research articles published in the English language	Book chapters, magazines, et cetera that are not a primary study or main research work were excluded.
Primary studies have been included.	Articles published in other languages aside from English were excluded.
Research papers from 2019 to 2024 have been included	Research publications before 2019 were excluded.
Open-access papers were included.	Publications not in full open access were excluded.
Only journal and conference studies are included	Survey and review articles are excluded

The threshold scoring system of this systematic literature review is adopted to ensure the completeness and relevance of the studies to be included. A study is eligible if it accumulates a minimum score of 3; it should cover specific aspects in each of the three research domains below:

- Algorithm Analysis: Discusses the ML algorithm (e.g., SVM, Random Forest) for ransomware classification, highlighting the DL algorithm (e.g., CNN, RNN) for ransomware classification. Analysis of trends in algorithm usage (e.g., comparison over time, emerging techniques).
- It describes the concept of FL and its application to malware detection and classification, identifies factors influencing the choice of FL frameworks, and discusses challenges facing FL application ransomware detection, mitigation, and tradeoffs.

- The paper should refer to insight and challenges facing the FL application, as well as problems halting the deployment of the ransomware system to industries. Compares different techniques' performance for the detection of ransomware using multiple measures.

Inclusion Criteria Formula:

$$\text{Inclusion} = \begin{cases} \sum_{i=1}^{n} A_{\mathrm{RQ}x} \text{ is TRUE}, & \text{if } \sum_{i=1}^{n} A_{\mathrm{RQ1}} + \sum_{i=1}^{n} A_{\mathrm{RQ2}} + \sum_{i=1}^{n} A_{\mathrm{RQ3}} \geqslant 3 \quad \text{(Accept)} \\ Satisfy\ if & \end{cases}$$

Otherwise, $then$ (Reject)

Where:

$\sum_{i=1}^{n} A_{\mathrm{RQ}x}$ represents the sum of scores based on the aspects within each respective research question (RQ1, RQ2, RQ3, RQ4).

Apparently, a paper is considered for selection if it satisfactorily answers at least one aspect from among the three categories listed. This rigorous selection process assures that the studies meet a multidimensional perspective of ransomware detection: algorithmic approaches, ransomware trend landscape, FL frameworks, and performance evaluation methodologies.

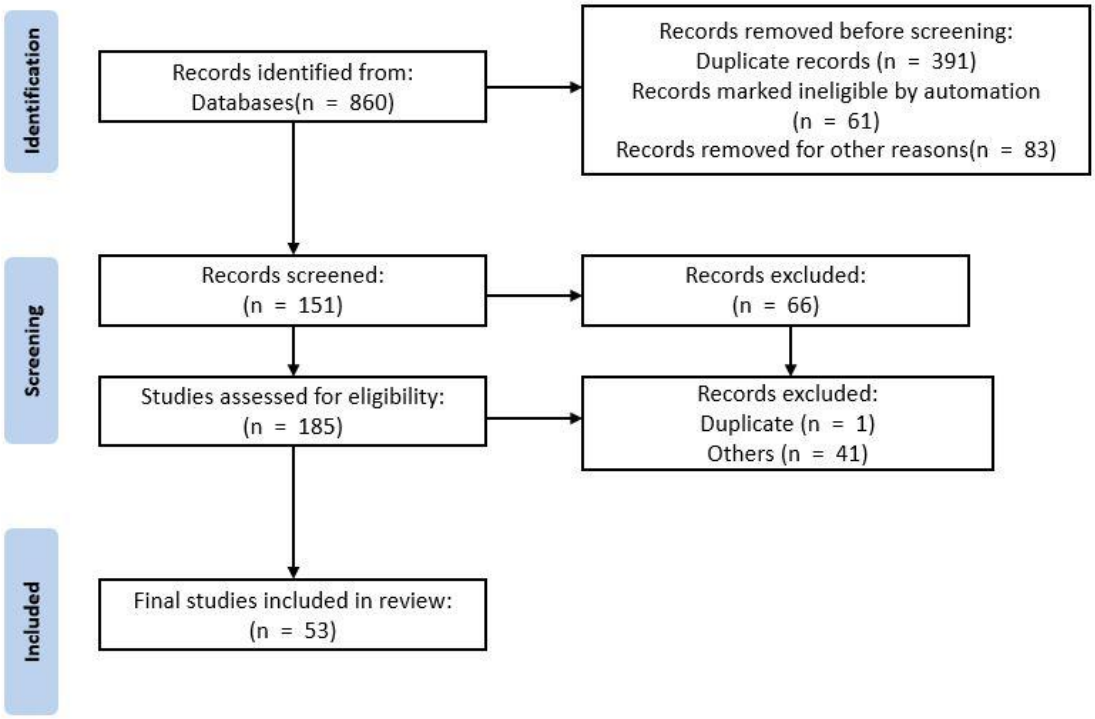

Fig 10 Systematic review outcome indicating the data records we have otantained using the research and review methodology before undergoing analysis

Figure 10 shows our systematic review outcome; after implementing our inclusion and exclusion criteria, 185 articles were eligible. Redundant papers were removed, and each article was evaluated against the defined keywords and formulated research questions. Papers were initially considered based on their titles and abstracts.

4.5 Data Extraction

Following the selection process, the review proceeded to the data extraction stage, where the attributes and findings of the selected articles were systematically documented and presented. In this stage, we examined the detection approach applied in the paper, the aim and objective, and the technologies used. After reviewing and making an appropriate selection, the relevant papers were exported as a Comma-Separated Values (CSV) file and uploaded to Rayyan[1] online for the quantitative analysis as the final stage; Rayyan and VOSviewer[2] software were selected for these tasks. Rayyan is an online platform that offers a collaborative screening environment for literature review, while VOSviewer is appropriate for visualization of the screened result. Figure 11**Fig 11** displays the association of relevant keywords from the studies we have selected; the areas with yellow and green indicate the current actively researched areas, while the

[1] https://new.rayyan.ai/reviews/1007035/screening
[2]https://www.vosviewer.com/

purple shows areas yet to be exploited enough. The visualizations make finding a research gap in ransomware detection and mitigation easy.

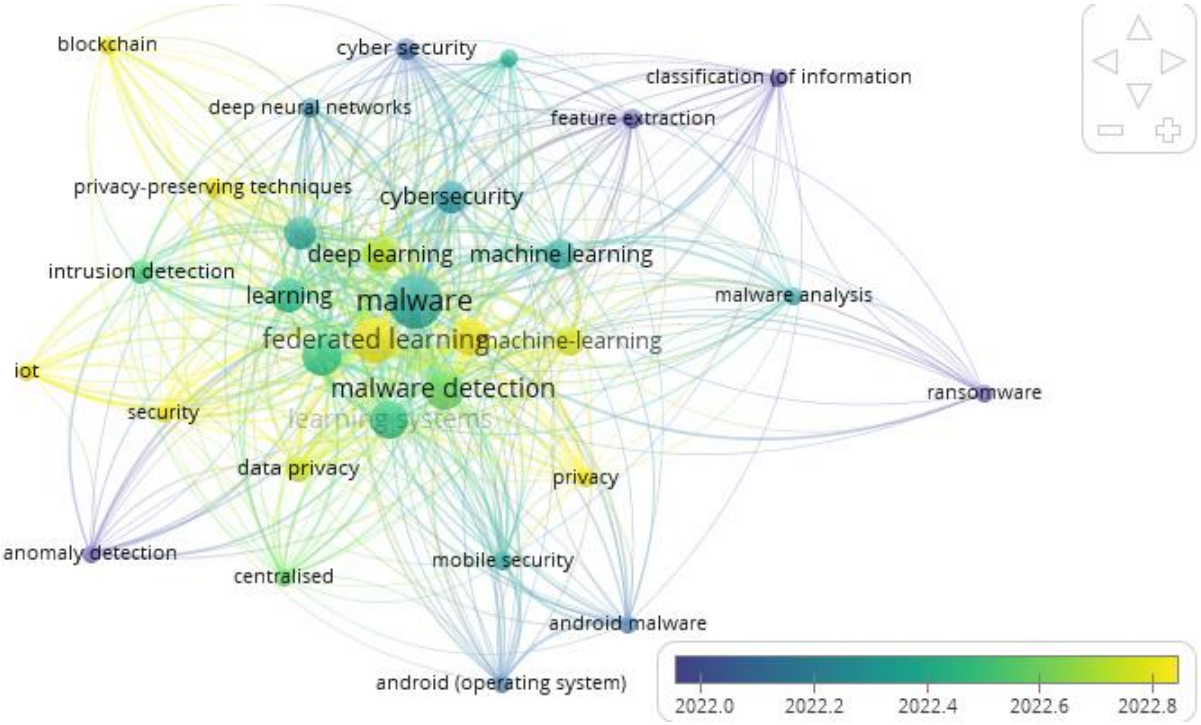

Fig 11 Visualization of keyword by association using VOSviewer[3]

4.5 Snowballing

This mechanism is critical for including papers, as it involves extracting the most pertinent literature within the research domain [86]. Initially, we applied this concept to the 185 selected papers during the data extraction stage. After thoroughly reviewing the abstracts and methodologies, 85 papers were retained. Subsequently, a careful examination of the result analysis and full text narrowed the list to 53 papers used for this study.

4.6 Results and Discussion on Quantitative Analysis

Quantitative analysis is the heart of our (SLR), as it statistically evaluates and discusses the data from research studies. This study's objective will be realized by analyzing the data from various perspectives and parameters, achieving a comprehensive understanding of the nature and trends within our target area, as outlined in our introduction. Our findings highlight that ransomware detection remains a highly active and evolving research area; however, the FL application is still very new and under-explored.

[3] https://www.vosviewer.com/

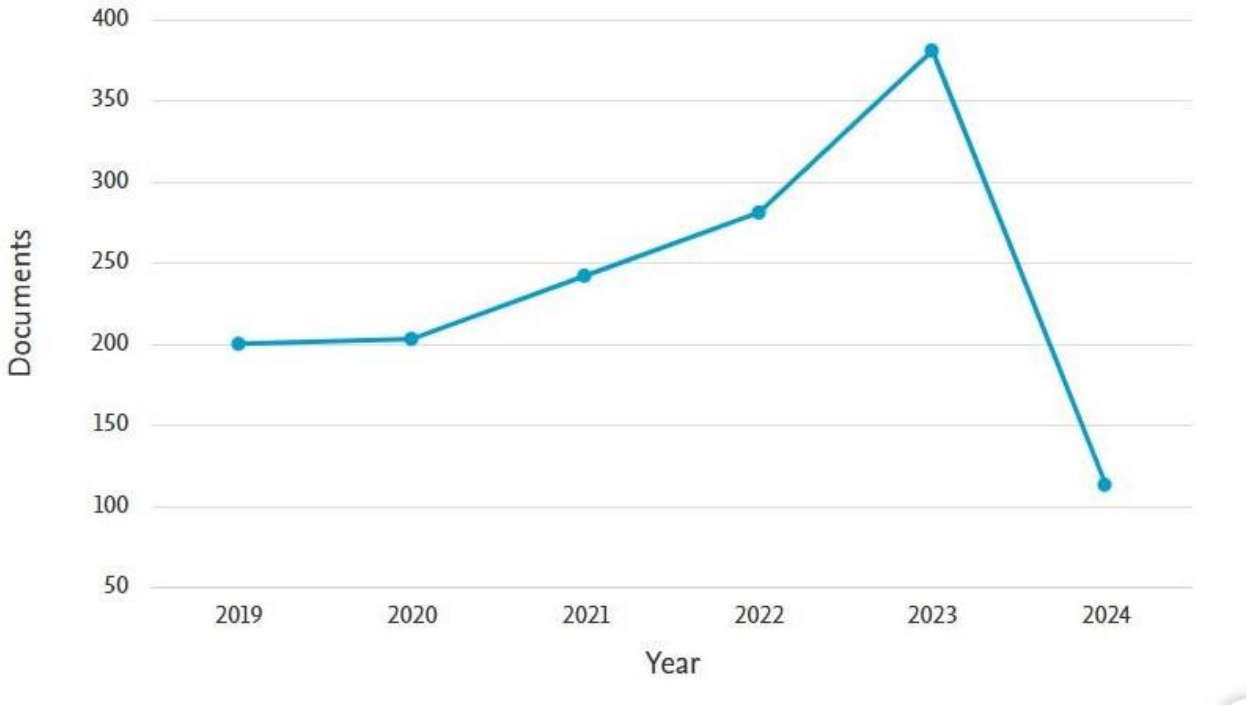

Fig 12 Publication in the research domain in 5 years (2019 – 2024), which indicates active research in the detection, response, analysis, and mitigation

As illustrated in Figure 12, the number of studies published from 2019 to 2024 has shown a consistent upward trend, reflecting the growing importance of this field in the cybersecurity landscape. Notably, approximately 45% of the surveyed studies advocate for collaborative threat information-sharing approaches, such as FL, as a critical strategy for enhancing cyberspace security. However, this collaborative approach is underutilized in ransomware detection and mitigation strategies despite its potential. Integrating FL in ransomware detection could significantly enhance the collective defense mechanisms, allowing for faster adaptation to emerging threats by leveraging data from multiple sources. Therefore, we believe that applying more research efforts, especially in a synergic manner, is needed to harness the full potential of collaborative approaches in combating the dynamic and rapidly evolving nature of ransomware attacks. This gap presents an opportunity for future work to explore the integration of FL and other collaborative methods to build more resilient ransomware detection and response frameworks.

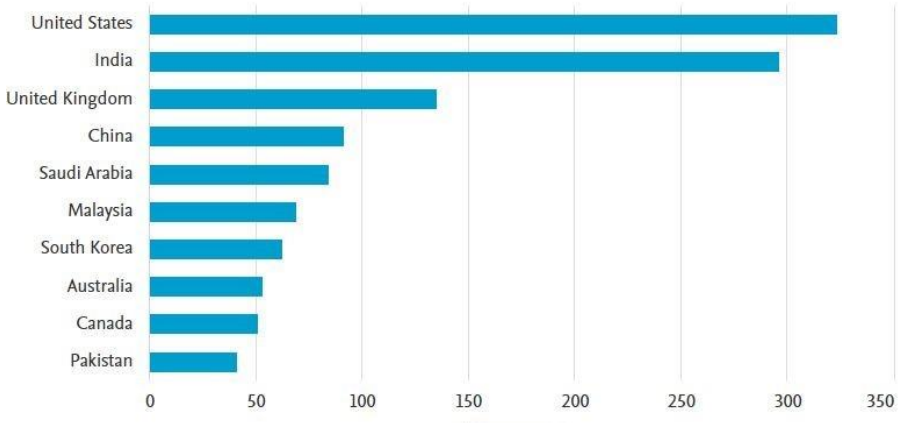

Fig 13 Study published by country, with the US, India, UK, and China ranking as top destinations

Figure 13 presents the distribution of publications related to malware detection, classification, and prevention across different countries. The data reveal some intriguing insights: despite being among the top five most targeted countries for malware attacks in 2023, highlighted in Figure 5 of section 3, specifically the United States, Canada, United Kingdom, Germany, and France, only the United States and the United Kingdom demonstrate significant engagement in malware research among these nations. This disparity highlights a gap between being a primary target of cyber threats and actively contributing to the body of research to combat such threats. Interestingly, China is not listed among the top ten most targeted countries for malware attacks; however, they firmly commit to research in this field. This proactive stance likely reflects strategic priorities in cybersecurity, emphasizing preemptive measures and technological leadership rather than merely reacting to direct threats. The active involvement of the Chinese research community suggests a focused effort on developing innovative solutions and advancing the scientific understanding of malware behaviors, detection techniques, and preventive measures. Practically, the uneven global distribution of research efforts indicates that countries most affected by malware do not always contribute to research output proportionally. International collaboration and investment in research, particularly from highly targeted regions, should be encouraged to strengthen global cybersecurity resilience.

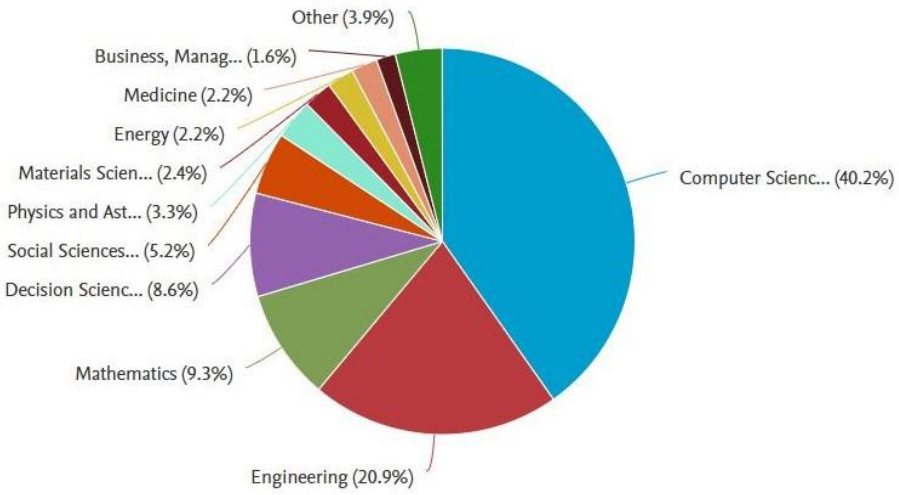

Fig 14 Distribution of subject areas for malware detection, response, and prediction

Figure 14 shows the critical knowledge domains essential for malware detection and mitigation expertise. This chart provides valuable insights into the field's interdisciplinary nature, showcasing the primary and complementary areas of expertise that contribute to advancements in this domain. The core subject areas include computer science, engineering, and mathematics, which form the foundational pillars, equipping researchers with the necessary technical skills, such as algorithm development, data analysis, and system design, essential for developing sophisticated detection and mitigation techniques. Computer science, in particular, plays a pivotal role, offering programming, ML, cybersecurity, and software engineering knowledge, which are integral to understanding and combating malware threats. Engineering complements this by focusing on designing, implementing, and optimizing hardware and software systems that are resilient to malicious attacks. Mathematics contributes through its analytical frameworks and statistical methods, enabling the modeling and prediction of malware behaviors, which are critical for practical detection algorithms.

In addition to these core fields, complementary domains like decision science, social science, energy, and business management provide broader perspectives that enhance malware research. Decision Science contributes to optimizing response strategies and improving decision-making processes in cybersecurity contexts. Social Science offers insights into the human factors associated with cyber threats, including user behavior and threat actor motivations. The relevance of Energy and Business Management emerges from the increasing targeting of critical infrastructure and enterprise environments, highlighting the need for specialized knowledge to protect these sectors. These interdisciplinary knowledge areas underscore the complexity of malware detection and mitigation, highlighting the need for a broad and diverse set of skills and expertise that fosters innovative solutions and a more holistic approach to addressing the challenges posed by ever-evolving cyber threats.

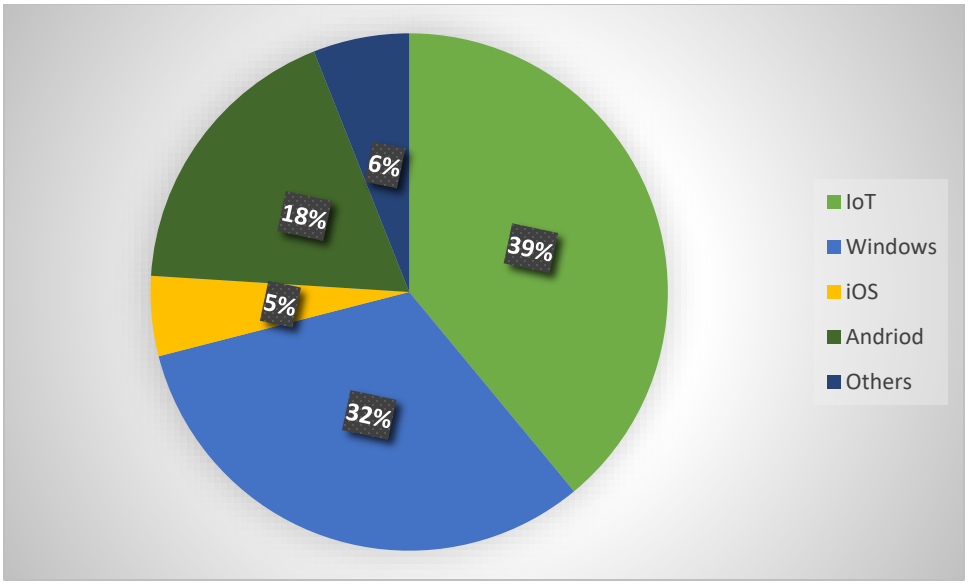

Fig 15 Cross-platform distribution on the application of FL in malware detection and mitigation, indicating a high percentage in IoT devices, Windows and Andriod platform

Figure 15 illustrates the cross-platform distribution of FL applications in malware detection and mitigation, highlighting significant adoption across IoT devices, Windows, and Android platforms. The high percentage of applications in IoT devices is particularly noteworthy, as the proliferation of IoT-enabled appliances, often powered by artificial intelligence and automation, has significantly expanded the attack surface for cybercriminals. The surge in FL research within this domain is consistent with the rapid growth of IoT devices, which are increasingly targeted due to their often-limited security measures and widespread use across households, offices, and firms. The Windows platform also exhibits similar usage of FL, particularly within the financial sector, which has recently been a major target for cybercriminals. The adaptability of FL to continuously improve detection models by learning from diverse, real-world threat data without exposing sensitive information makes it an attractive solution for this sector, where security and compliance requirements are paramount.

The Android platform ranks among the top three due to its global prevalence and widespread use among mobile users. The platform's popularity makes it a frequent target for malware, and the application of FL helps enhance detection mechanisms by training on decentralized datasets gathered from numerous Android devices worldwide. The application of FL across these platforms demonstrates its potential as a pivotal tool for future research in malware detection and mitigation. FL's ability to harness distributed data while safeguarding privacy positions it as a critical enabler in the ongoing battle against cyber threats. As technology evolves, further exploration and implementation of FL will be essential in developing robust, adaptive defenses capable of countering increasingly sophisticated malware attacks.

7 Limitations and Conclusions

7.1 Limitations

This study employed a rigorous, comprehensive and straight-to-the-point methodology, leveraging prominent online databases renowned for hosting high-quality, peer-reviewed research publications. While this approach enhances the credibility and reliability of our findings, it also presents inherent limitations. Firstly, our inclusion criteria were restricted to studies published in English, which inherently excludes relevant research available in other languages. This language bias may limit the comprehensiveness of our review, particularly in regions where significant research on Ransomware and FL applications for malware and ransomware detection might be published in non-English journals. Similarly, our review's scope was confined to publications from 2019 - 2024, potentially omitting earlier foundational works that could provide valuable historical context or insights into the evolution of FL in ransomware detection and broader cybersecurity trends. This temporal constraint may have led to excluding relevant studies that predate the chosen timeframe but still hold significance in the field.

Our study incorporated an analysis of FL applications for ransomware detection, common challenges faced in these applications, and an exploration of the techniques and trends within this landscape. While these inclusions substantiate our analysis and align with our research objectives, it is essential to acknowledge that the observations presented may reflect conservative estimates. Given the rapidly evolving nature of malware and ransomware threats, additional techniques and emerging trends may exist beyond those identified within the constraints of this SLR.

Furthermore, the keyword-driven search strategy employed, although carefully designed, might have inadvertently missed some relevant studies, particularly those that may not explicitly mention our targeted keywords but still contribute valuable insights to the fields of malware detection, prevention, and FL applications. This dynamic and interdisciplinary domain means that relevant research could be published under varied terminologies or within adjacent fields not immediately captured by our search terms.

7.2 Conclusion

Our SLR comprehensively examined the landscape and evolving ransomware detection and mitigation trends, explicitly focusing on the FL application and other advanced techniques. By meticulously reviewing studies published between 2019 and 2024, we synthesized critical insights into various algorithms for ransomware detection, the application of FL in malware detection, and related challenges encountered in the field. Our analysis highlighted key findings and emerging trends, providing a deeper understanding of the current state of ransomware detection technologies.

The review identified that FL is pivotal in advancing ransomware detection. This technique offers innovative approaches that enhance detection accuracy and enable real-time, distributed threat analysis without compromising data privacy. Since its introduction by Google in 2016, FL stands out as a transformative approach, enabling collaborative learning across decentralized networks while protecting sensitive data. This feature is crucial given the increasing sophistication of ransomware attacks targeting diverse platforms like IoT-enabled, Windows, and Android devices.

We suggest leveraging FL and graph-based dynamic analysis to bolster ransomware detection capabilities significantly. Graph Neural Networks (GNNs) and related graph analysis methods have shown excellent efficiency in capturing complex relationships within malware behaviors, making them highly

effective in identifying and mitigating ransomware threats. As ransomware continues to pose a significant threat to organizations and individuals worldwide, the insights and recommendations provided herein can serve as a foundation for future research and development of next-generation ransomware defense mechanisms.

Author Contributions All authors have contributed equally to this study.

Funding No funding was received to conduct this study.

Data Availability Not applicable.

Code Availability Not applicable.

Declarations

Conflict of interest There is no potential conflict of interest.

Human Participants and/or Animals This work does not contain any studies with human participants and/or animals.

Informed Consent Not applicable.

Reference

[1] R. Gunaratna, "The Global Ransomware Marketplace: Victim Profiling and Strategic Targeting Crafting a Corporate Veil: The Dichotomy of Ransomware PR," *The Global Ransomware Marketplace*, May 2024.

[2] A. Kulkarni, Y. Wang, M. Gopinath, D. Sobien, A. Rahman, and F. A. Batarseh, "A Review of Cybersecurity Incidents in the Food and Agriculture Sector," Mar. 2024.

[3] B. A. S. Al-rimy, M. A. Maarof, and S. Z. M. Shaid, "Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions," May 01, 2018, *Elsevier Ltd*. doi: 10.1016/j.cose.2018.01.001.

[4] M. Nobakht, R. Javidan, and A. Pourebrahimi, "SIM-FED: Secure IoT malware detection model with federated learning," *Computers and Electrical Engineering*, vol. 116, May 2024, doi: 10.1016/j.compeleceng.2024.109139.

[5] SOPHOS, "The State of Ransomware 2024 – Sophos News." Accessed: May 16, 2024. [Online]. Available: https://news.sophos.com/en-us/2024/04/30/the-state-of-ransomware-2024/

[6] J. L. "Jamey" Worrell, "A SURVEY OF THE CURRENT AND EMERGING RANSOMWARE THREAT LANDSCAPE," *EDPACS*, vol. 69, no. 2, pp. 1–11, 2024, doi: 10.1080/07366981.2024.2315639.

[7] S. A. Wadho, A. Yichiet, M. L. Gan, L. C. Kang, R. Akbar, and R. Kumar, “Emerging Ransomware Attacks: Improvement and Remedies-A Systematic Literature Review,” in *2023 4th International Conference on Artificial Intelligence and Data Sciences: Discovering Technological Advancement in Artificial Intelligence and Data Science, AiDAS 2023 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 148–153. doi: 10.1109/AiDAS60501.2023.10284647.

[8] B. S. Guendouzi, S. Ouchani, H. EL Assaad, and M. EL Zaher, “A systematic review of federated learning: Challenges, aggregation methods, and development tools,” Nov. 01, 2023, *Academic Press*. doi: 10.1016/j.jnca.2023.103714.

[9] A. Kapoor, A. Gupta, R. Gupta, S. Tanwar, G. Sharma, and I. E. Davidson, “Ransomware detection, avoidance, and mitigation scheme: A review and future directions,” Jan. 01, 2022, *MDPI*. doi: 10.3390/su14010008.

[10] Y. Huang, G. Yang, H. Zhou, H. Dai, D. Yuan, and S. Yu, “VPPFL: A verifiable privacy-preserving federated learning scheme against poisoning attacks,” *Comput Secur*, vol. 136, no. October 2023, p. 103562, 2024, doi: 10.1016/j.cose.2023.103562.

[11] V. Rey, P. M. Sánchez Sánchez, A. Huertas Celdrán, and G. Bovet, “Federated learning for malware detection in IoT devices,” *Computer Networks*, vol. 204, no. October 2021, p. 108693, 2022, doi: 10.1016/j.comnet.2021.108693.

[12] A. Chaudhuri, A. Nandi, and B. Pradhan, *for Android Malware Classification*. Springer Nature Singapore, 2023. doi: 10.1007/978-981-19-9858-4.

[13] “Applications of Federated Learning ; Taxonomy , Challenges , and Research Trends,” 2022.

[14] M. Conti, A. Gangwal, and S. Ruj, “On the economic significance of ransomware campaigns: A Bitcoin transactions perspective,” Nov. 01, 2018, *Elsevier Ltd*. doi: 10.1016/j.cose.2018.08.008.

[15] D. Y. Kao, S. C. Hsiao, and R. Tso, “Analyzing WannaCry Ransomware Considering the Weapons and Exploits,” *International Conference on Advanced Communication Technology, ICACT*, vol. 2019-February, pp. 1098–1107, Apr. 2019, doi: 10.23919/ICACT.2019.8702049.

[16] A. Ferrante, M. Malek, F. Martinelli, F. Mercaldo, and J. Milosevic, “Extinguishing Ransomware - A Hybrid Approach to Android Ransomware Detection,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10723 LNCS, pp. 242–258, 2018, doi: 10.1007/978-3-319-75650-9_16.

[17] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak, and M. K. Khan, “Ransomware: Recent advances, analysis, challenges and future research directions,” *Comput Secur*, vol. 111, p. 102490, 2021, doi: 10.1016/j.cose.2021.102490.

[18] E. T. Martinez Beltran *et al.*, “Decentralized Federated Learning: Fundamentals, State of the Art, Frameworks, Trends, and Challenges,” *IEEE Communications Surveys and Tutorials*, vol. 25, no. 4, pp. 2983–3013, 2023, doi: 10.1109/COMST.2023.3315746.

[19] P. Chen and T. Chen, “CAFE : Catastrophic Data Leakage in Vertical Federated Learning,” no. NeurIPS, 2021.

[20] W. Yang *et al.*, “Federated Learning in Mobile Edge Networks : A Comprehensive Survey,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 2031–2063, 2020, doi: 10.1109/COMST.2020.2986024.

[21] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, “Federated Learning for Healthcare Informatics,” pp. 1–19, 2021.

[22] M. Alazab, S. P. Rm, M. Parimala, P. K. R. Maddikunta, T. R. Gadekallu, and Q. V. Pham, “Federated Learning for Cybersecurity: Concepts, Challenges, and Future Directions,” *IEEE Trans Industr Inform*, vol. 18, no. 5, pp. 3501–3509, 2022, doi: 10.1109/TII.2021.3119038.

[23] B. Ghimire and D. B. Rawat, “Recent Advances on Federated Learning for Cybersecurity and Cybersecurity for Federated Learning for Internet of Things,” *IEEE Internet Things J*, vol. 9, no. 11, pp. 8229–8249, 2022, doi: 10.1109/JIOT.2022.3150363.

[24] V. S. Nafade, A. B. Patil, Y. M. Khandelwal, R. R. More, and M. Deshmukh, “A Survey on Image and Text Encryption Using ECC and Steganography,” pp. 929–936, 2023, doi: 10.1007/978-981-99-3485-0_74.

[25] M. A. Ayub, A. Siraj, B. Filar, and M. Gupta, “RWArmor: a static-informed dynamic analysis approach for early detection of cryptographic windows ransomware,” *Int J Inf Secur*, vol. 23, no. 1, pp. 533–556, Feb. 2024, doi: 10.1007/s10207-023-00758-z.

[26] J. A. Herrera-Silva and M. Hernández-Álvarez, “Dynamic Feature Dataset for Ransomware Detection Using Machine Learning Algorithms,” *Sensors*, vol. 23, no. 3, Feb. 2023, doi: 10.3390/s23031053.

[27] A. M. Maigida, S. M. Abdulhamid, M. Olalere, J. K. Alhassan, H. Chiroma, and E. G. Dada, “Systematic literature review and metadata analysis of ransomware attacks and detection mechanisms,” Jul. 01, 2019, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1007/s40860-019-00080-3.

[28] P. Maniriho, A. N. Mahmood, and M. J. M. Chowdhury, “A systematic literature review on Windows malware detection: Techniques, research issues, and future directions,” *Journal of Systems and Software*, vol. 209, Mar. 2024, doi: 10.1016/j.jss.2023.111921.

[29] S. Song, B. Kim, and S. Lee, “The Effective Ransomware Prevention Technique Using Process Monitoring on Android Platform,” *Mobile Information Systems*, vol. 2016, 2016, doi: 10.1155/2016/2946735.

[30] A. Continella *et al.*, “ShieldFS: A self-healing, ransomware-aware file system,” *ACM International Conference Proceeding Series*, vol. 5-9-December-2016, pp. 336–347, Dec. 2016, doi: 10.1145/2991079.2991110.

[31] S. H. Kok, A. Abdullah, N. Z. Jhanjhi, and M. Supramaniam, “Prevention of crypto-ransomware using a pre-encryption detection algorithm,” *Computers*, vol. 8, no. 4, Dec. 2019, doi: 10.3390/computers8040079.

[32] F. Manavi, “A novel approach for ransomware detection based on PE header using graph embedding,” *Journal of Computer Virology and Hacking Techniques*, vol. 18, no. 4, pp. 285–296, Dec. 2022, doi: 10.1007/S11416-021-00414-X/METRICS.

[33] J. Zhu, J. Jang-Jaccard, A. Singh, I. Welch, H. AL-Sahaf, and S. Camtepe, “A few-shot meta-learning based siamese neural network using entropy features for ransomware classification,” *Comput Secur*, vol. 117, Jun. 2022, doi: 10.1016/j.cose.2022.102691.

[34] C. C. Moreira, D. C. Moreira, and C. Sales, “A comprehensive analysis combining structural features for detection of new ransomware families,” *Journal of Information Security and Applications*, vol. 81, Mar. 2024, doi: 10.1016/j.jisa.2024.103716.

[35] N. Hampton, Z. Baig, and S. Zeadally, “Ransomware behavioural analysis on windows platforms,” *Journal of Information Security and Applications*, vol. 40, pp. 44–51, Jun. 2018, doi: 10.1016/J.JISA.2018.02.008.

[36] B. A. S. Al-rimy, M. A. Maarof, and S. Z. M. Shaid, “A 0-Day Aware Crypto-Ransomware Early Behavioral Detection Framework,” *Lecture Notes on Data Engineering and Communications Technologies*, vol. 5, pp. 758–766, 2018, doi: 10.1007/978-3-319-59427-9_78.

[37] B. Lokuketagoda, M. P. Weerakoon, U. M. Kuruppu, A. N. Senarathne, and K. Yapa Abeywardena, “R - Killer: An email based ransomware protection tool,” *13th International Conference on Computer Science and Education, ICCSE 2018*, pp. 735–741, Sep. 2018, doi: 10.1109/ICCSE.2018.8468807.

[38] A. De Paola, S. Gaglio, G. Lo Re, and M. Morana, “A hybrid system for malware detection on big data,” *INFOCOM 2018 - IEEE Conference on Computer Communications Workshops*, pp. 45–50, Jul. 2018, doi: 10.1109/INFCOMW.2018.8406963.

[39] C. C. Moreira, D. C. Moreira, and C. de S. de Sales, “Improving ransomware detection based on portable executable header using xception convolutional neural network,” *Comput Secur*, vol. 130, Jul. 2023, doi: 10.1016/j.cose.2023.103265.

[40] C. Zheng, N. Dellarocca, N. Andronio, S. Zanero, and F. Maggi, “GreatEatlon: Fast, Static Detection of Mobile Ransomware,” *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 198 LNICST, pp. 617–636, 2017, doi: 10.1007/978-3-319-59608-2_34.

[41] M. Aljabri *et al.*, “Ransomware detection based on machine learning using memory features,” *Egyptian Informatics Journal*, vol. 25, Mar. 2024, doi: 10.1016/j.eij.2024.100445.

[42] I. Kara and M. Aydos, “Static and Dynamic Analysis of Third Generation Cerber Ransomware,” *International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism, IBIGDELFT 2018 - Proceedings*, pp. 12–17, Jan. 2019, doi: 10.1109/IBIGDELFT.2018.8625353.

[43] T. Petros, H. Ghirmay, S. Otoum, R. Salem, and M. Debbah, “FLDetect: An API-Based Ransomware Detection Using Federated Learning,” in *Proceedings - IEEE Global Communications Conference, GLOBECOM*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 4449–4454. doi: 10.1109/GLOBECOM54140.2023.10437540.

[44] S. Jung and Y. Won, “Ransomware detection method based on context-aware entropy analysis,” *Soft comput*, vol. 22, no. 20, pp. 6731–6740, Oct. 2018, doi: 10.1007/s00500-018-3257-z.

[45] W. Song *et al.*, *{UNVEIL}: A {Large-Scale}, Automated Approach to Detecting Ransomware*. 2016.

[46] L. F. Maimó, A. H. Celdrán, Á. L. Perales Gómez, F. J. García Clemente, J. Weimer, and I. Lee, “Intelligent and dynamic ransomware spread detection and mitigation in integrated clinical environments,” *Sensors (Switzerland)*, vol. 19, no. 5, Mar. 2019, doi: 10.3390/s19051114.

[47] O. M. K. Alhawi, J. Baldwin, and A. Dehghantanha, “Leveraging machine learning techniques for windows ransomware network traffic detection,” in *Advances in Information Security*, vol. 70, Springer New York LLC, 2018, pp. 93–106. doi: 10.1007/978-3-319-73951-9_5.

[48] Z. Boussouf, H. Amrani, M. Zerhouni Khal, and F. Daidai, *Artificial Intelligence in Education: a Systematic Literature Review*, vol. 3. Springer Nature Singapore, 2024. doi: 10.56294/dm2024288.

[49] M. , S. Steve, “PREDICT World $10.5 Trillion Annually By 2025.” Accessed: Jul. 08, 2024. [Online]. Available: https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/

[50] MTI, “The many lives of BlackCat ransomware | Microsoft Security Blog,” Online. Accessed: May 15, 2024. [Online]. Available: https://www.microsoft.com/en-us/security/blog/2022/06/13/the-many-lives-of-blackcat-ransomware/

[51] N. Polatidis, S. Kapetanakis, M. Trovati, I. Korkontzelos, and Y. Manolopoulos, “FSSDroid: Feature subset selection for Android malware detection,” *World Wide Web*, vol. 27, no. 5, 2024, doi: 10.1007/s11280-024-01287-y.

[52] S. Das, A. Garg, and S. Kumar, “Stacking Ensemble-Based Approach for Malware Detection,” *SN Comput Sci*, vol. 5, no. 1, 2024, doi: 10.1007/s42979-023-02513-6.

[53] D. Shackleford, “Who ’ s Using Cyberthreat Intelligence and How ?,” *Sans Institute Survey*, no. February, p. 23, 2015.

[54] S. Paul, “A Comprehensive Review on Machine Learning-based Approaches for Next Generation Wireless Network,” *SN Comput Sci*, vol. 5, no. 5, 2024, doi: 10.1007/s42979-024-02831-3.

[55] H. N. Nguyen, N. N. Tran, T. H. Hoang, and V. L. Cao, “Denoising Latent Representation with SOMs for Unsupervised IoT Malware Detection,” *SN Comput Sci*, vol. 3, no. 6, pp. 1–15, 2022, doi: 10.1007/s42979-022-01344-1.

[56] CISO, “Providing threat intelligence to those in the Cloud.”

[57] F. Hacquebord, S. Hilt, and D. Sancho, “The Near and Far Future of Ransomware Business Models.”

[58] K. Y. Lin and W. R. Huang, “Using Federated Learning on Malware Classification,” *International Conference on Advanced Communication Technology, ICACT*, vol. 2020, pp. 585–589, Feb. 2020, doi: 10.23919/ICACT48636.2020.9061261.

[59] H. Brendan McMahan Eider Moore Daniel Ramage Seth Hampson Blaise AgüeraAg and A. Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” 2017.

[60] F. Chen, M. Luo, Z. Dong, Z. Li, and X. He, “Federated Meta-Learning with Fast Convergence and Efficient Communication,” Feb. 2018.

[61] S. Ek, F. Portet, P. Lalanda, and G. Vega, “A Federated Learning Aggregation Algorithm for Pervasive Computing: Evaluation and Comparison,” in *2021 IEEE International Conference on Pervasive Computing and Communications, PerCom 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021. doi: 10.1109/PERCOM50583.2021.9439129.

[62] S. B. Guendouzi, S. Ouchani, and M. Malki, “Enhancing the Aggregation of the Federated Learning for the Industrial Cyber Physical Systems,” in *Proceedings of the 2022 IEEE International Conference on Cyber Security and Resilience, CSR 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 197–202. doi: 10.1109/CSR54599.2022.9850301.

[63] C.-H. Yao, B. Gong, H. Qi, Y. Cui, Y. Zhu, and M.-H. Yang, “Federated Multi-Target Domain Adaptation.”

[64] Z. Lu, S. F. Lu, Y. Q. Cui, X. M. Tang, and J. J. Wu, “Split Aggregation: Lightweight Privacy-Preserving Federated Learning Resistant to Byzantine Attacks,” *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 5575–5590, 2024, doi: 10.1109/TIFS.2024.3402993.

[65] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, “DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber-Physical Systems,” *IEEE Trans Industr Inform*, vol. 17, no. 8, pp. 5615–5624, 2021, doi: 10.1109/TII.2020.3023430.

[66] Z. Chen, N. Lv, P. Liu, Y. Fang, K. Chen, and W. Pan, “Intrusion Detection for Wireless Edge Networks Based on Federated Learning,” *IEEE Access*, vol. 8, pp. 217463–217472, 2020, doi: 10.1109/ACCESS.2020.3041793.

[67] O. J. Falana, A. S. Sodiya, S. A. Onashoga, and B. S. Badmus, “Mal-Detect: An intelligent visualization approach for malware detection,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1968–1983, May 2022, doi: 10.1016/J.JKSUCI.2022.02.026.

[68] R. H. Hsu *et al.*, “A Privacy-Preserving Federated Learning System for Android Malware Detection Based on Edge Computing,” *Proceedings - 2020 15th Asia Joint Conference on Information Security, AsiaJCIS 2020*, pp. 128–136, 2020, doi: 10.1109/AsiaJCIS50894.2020.00031.

[69] V. Moonsamy and C. Diaz, “Less is More : A privacy-respecting Android malware classifier using federated learning,” pp. 1–21.

[70] M. Khari, A. K. Garg, A. H. Gandomi, R. Gupta, R. Patan, and B. Balusamy, “Securıng data ın Internet Of Thıngs (Iot) usıng cryptography and steganography technıques,” *IEEE Trans Syst Man Cybern Syst*, vol. 50, no. 1, pp. 73–80, Jan. 2020, doi: 10.1109/tsmc.2019.2903785.

[71] W. Han, J. Xue, Y. Wang, L. Huang, Z. Kong, and L. Mao, “MalDAE: Detecting and explaining malware based on correlation and fusion of static and dynamic characteristics
,” 2019, doi: 10.1016/j.cose.2019.02.007.

[72] I. Yaqoob *et al.*, "The rise of ransomware and emerging security challenges in the Internet of Things," *Computer Networks*, vol. 129, pp. 444–458, Dec. 2017, doi: 10.1016/J.COMNET.2017.09.003.

[73] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive Privacy Analysis of Deep Learning," *2019 IEEE Symposium on Security and Privacy (SP)*, pp. 739–753, doi: 10.1109/SP.2019.00065.

[74] J. Lee, J. Oh, S. Lim, S.-Y. Yun, and J.-G. Lee, "TornadoAggregate: Accurate and Scalable Federated Learning via the Ring-Based Architecture," Dec. 2020.

[75] F. Wang, Y. He, Y. Guo, P. Li, and X. Wei, "Privacy-Preserving Robust Federated Learning with Distributed Differential Privacy," *Proceedings - 2022 IEEE 21st International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2022*, pp. 598–605, 2022, doi: 10.1109/TrustCom56396.2022.00087.

[76] E. Fini, J. Mairal, K. A. X. Alameda-pineda, M. Nabi, and E. Ricci, "Semi-supervised learning made simple with self-supervised clustering," pp. 3187–3197.

[77] R. Bendlin, I. Damg, C. Orlandi, and S. Zakarias, "Semi-homomorphic Encryption and Multiparty Computation," pp. 169–188, 2011.

[78] M. Van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully Homomorphic Encryption over the Integers," pp. 24–43, 2010.

[79] K. Bonawitz, P. Kairouz, B. Mcmahan, and D. Ramage, "Federated Learning and Privacy," no. october, pp. 87–114, 2021, doi: 10.1145/3494834.3500240.

[80] P. Mohassel and Y. Zhang, "SecureML : A System for Scalable Privacy-Preserving Machine Learning," *2017 IEEE Symposium on Security and Privacy (SP)*, pp. 19–38, 2017, doi: 10.1109/SP.2017.12.

[81] S. Saha, A. Chakraborty, A. Chatterjee, S. Dhargupta, S. K. Ghosal, and R. Sarkar, "Extended exploiting modifcation direction based steganography using hashed-weightage array," *Multimed Tools Appl*, vol. 79, no. 29–30, pp. 20973–20993, Aug. 2020, doi: 10.1007/s11042-020-08951-1.

[82] N. Holohan, S. Braghin, and M. Aonghusa, "The Bounded Laplace Mechanism in Differential Privacy," pp. 0–5.

[83] J. Dong and A. Roth, "Gaussian differential privacy," no. May 2019, pp. 3–54, 2022, doi: 10.1111/rssb.12454.

[84] Y. Ali, H. U. Khan, and M. Khalid, "Engineering the advances of the artificial neural networks (ANNs) for the security requirements of Internet of Things: a systematic review," Dec. 01, 2023, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1186/s40537-023-00805-5.

[85] B. Liao, Y. Ali, S. Nazir, L. He, and H. U. Khan, "Security Analysis of IoT Devices by Using Mobile Computing: A Systematic Literature Review," 2020, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2020.3006358.

[86] C. Wohlin, “Guidelines for snowballing in systematic literature studies and a replication in software engineering,” *ACM International Conference Proceeding Series*, 2014, doi: 10.1145/2601248.2601268.