Machine Learning Approach for Detecting and Combating Bring Your Own Device (BYOD) Security Threats and Attacks: A systematic Mapping Review

Christopher Ifeanyi Eke (eke.christopher@science.fulafia.edu.ng)
Federal University of Lafia
Azah Anir Norman
University of Malaya
Mwenge Mulenga
Mulungushi University

Research Article

Keywords: BYOD environment, Security threats and attacks, Machine learning, Datasets, Evaluation metrics

Posted Date: October 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2124645/v1

Abstract

Bring your own device (BYOD) paradigm that permits employees to come with their own mobile devices to join the organizational network is rapidly changing the organizational operation method by enhancing flexibility, productivity, and efficiency. Despite these benefits, security issues remain a concern in organizational settings. A considerable number of studies have been conducted and published in this domain without a detailed review of the security solution mechanisms. Moreover, some reviews conducted focused more on the conventional approaches such as mobile content management, and application content management. Hence, the implementation of security in BYOD using the conventional method is ineffective. Thus, machine learning approaches seem to be the promising approach, which provides a solution to the security problem in the BYOD environment. This study presents a comprehensive systematic mapping review that focused on the application of the machine learning approach for the mitigation of security threats and attacks in the BYOD environment by highlighting the current trends in the existing studies. Five academic databases were searched and a total of 753 of the primary studies published between 2012 and 2021 were initially retrieved. These studies were screened based on their title, abstract and full text to check their eligibility and relevance for the study. However, forty primary studies were included and analyzed in the systematic mapping review (SMR). Based on the analysis and bubble plot mapping, significant research trends were identified on security threats and attacks, machine learning approaches, datasets usage, and evaluation metrics. The SMR result demonstrates the rise in the number of investigations regarding malware and unauthorized access to existing security threats and attacks. The SMR study indicates that supervised learning approaches such as SVM, DT, and RF are the most employed learning model by the previous research. Thus, there is an open research issue in the application of unsupervised learning approaches such as clustering and deep learning approaches. Therefore, the SMR has set the pace for creating new ground research in the machine learning implementation in the BYOD environment, which will offer invaluable insight into the study field, and researchers can employ it to find a research gap in the research domain.

1. Introduction

The unending advancement in mobile computing has shifted interest in the innovative ubiquitous paradigms. Among these paradigms is bring your own device (BYOD) (Ballagas et al., 2004), which is separating attention and investments (Costa et al., 2018). BYOD paradigm permits employees to come with their own mobile devices, including tablets, laptops, and smartphones, to work and join them in an organization to access its resources instead of utilizing the organization's own devices (Samarathunge et al., 2018). Various organizations and corporations such as Citrix system, Unisys, Intel, Apple, and the White House are competing in the adoption of BYOD, certainly, the BYOD paradigm will rapidly receive attention in adoption by the leading organization worldwide (French et al., 2014). A survey conducted by CISCO reveals that 95% of companies permit the utilization of their owned devices to some degree, 36% provide complete support for their own devices, whereas 48% support selected devices (Barbier et al., 2012). Since the organization aims to attain high productivity and satisfy employees, they decided to permit employees to bring their own mobile devices into the process due to its benefit to the organization.

The benefits of BYOD cannot be overemphasized. One, there is a cost reduction in the organization that adopts the BYOD policy. This is because employers save a lot of money in acquiring high-cost devices and resources, including service agreements, hardware, licensing, software, insurance, and purchasing of data plans (Caldwell et al., 2012; Wang et al., 2014). Two, there is an increase in the flexibility, productivity, and mobility on the side of the employees in a sense that they perform more jobs with their devices anywhere and at any time since they are very familiar and satisfied as they use their devices (Rivera et al., 2013). Three, the durability of the device(s). This is possible because the employees usually handle their own devices more carefully than their provided counterparts (Ghosh et al., 2013). Four, online education. BYOD learning experience helps educational organizations such as Harvard edX, Khan Academy, and MIT edX to offer excellent quality online tutoring at a minimal cost (Miller et al., 2012). Thus, the organization's purpose of adopting the BYOD policy is to facilitate the convenience, flexibility, and device portability to take care of their employees' workflow, enhancing their efficiency and confidence (French et al., 2014). Besides, it also improves communication within an organization, online transactions, employees' interaction, and remote access to corporate data outside the organization is made easier.

Even though BYOD offers numerous benefits to both employees and organizations, security issues remain the major challenge in the BYOD environment. For instance, when data is copied from or into a mobile device, the data will be kept on that device even after the device has been disconnected from the corporate network. In such cases, the disclosure of confidential data to

unauthorized users will be much easier. Also, users exhibit confidence that their data is secured in all circumstances, but when there is a malfunction within the corporate network, it will alter and expose the confidential data saved in the mobile device. Thus, there may be a need for an employee's device that is permitted at the workplace to be configured in both software and hardware at the maximum level due to the organization's security concerns (Samarathunge et al., 2018).

In order to secure BYOD in an organization network, the machine learning technique aspect of artificial intelligence are promising alternative (Kamal et al., 2022). Machine learning is one of the advanced artificial intelligence techniques that can perform well in a dynamic network without being explicitly programmed. Machine learning can be employed to train a model to detect different attacks and offer corresponding preventive policies. In this context, the attack can be identified at the initial stage. Furthermore, machine learning approaches seem to be promising results in identifying new attacks by employing learning ability and handling them intelligently. Thus, the machine learning model can offer a potential security protocol for BYOD, which offer more reliability and accessibility than other models.

Currently, few reviews and survey research have been conducted on BYOD security implementation (Akin-Adetoro & Kabanda, 2015; Oktavia et al., 2016; Wang et al., 2014). For instance, Garba et al. (2015) conducted a review study on BYOD with a focus on information security and privacy challenges. The study looked into the current organizational practice that reveals information on BYOD and the drawbacks of adopting it. The study's findings indicate that the failure to achieve security and privacy on data will lead to the futility of BYOD adoption. They further added that if users' experience is not established, the solution will become unsuccessful. Similarly, Oktavia et al. (2016) conducted a systematic study of the security and privacy challenges in the BYOD environment. The study critically investigated the components based on security and privacy issues on BYOD. The study's findings maintained that the organization is required to amend its security policy and embrace the enhanced ones based on the identified threats. The authors added that the best solution should be able to split personal data and cooperate space, which will result in the protection of corporate data (Wang et al., 2014). In addition, (Jamal et al., 2020) presented a systematic study on the authentication technique employed for BYOD security implementation. The study identified the existing BYOD authentication methods and classified them based on BYOD threats, and further analyzed them by identifying their limitations. In another study, Akin-Adetoro and Kabanda (2015) conducted a review on BYOD with a focus on small and medium enterprises (SMEs). The study highlighted the contextual issues that SMEs in developing countries are required to know before the adoption of BYOD. The findings from the review indicated that the organization is required to plan for a change as a result of the ubiquitous and strategic nature of Potential IT changes that may be introduced by the employees. The authors also considered that due to the employees' appealing nature to utilize their devices in the workplace, SMEs in developing countries must equip themselves to harvest the possible benefits of BYOD adoption. In a related study, Palanisamy et al. (2020) conducted a systematic review study on compliance with BYOD security policies in organizations. The author utilized a total of 21 articles published between 2012 and 2019. The findings from the review provides an overview of the theory employed to describe and analyze security behavior and the factors that affect the BYOD security policies compliance behavior. Thus, the review focused on the features and factors influencing compliance behavior in BYOD environment.

Despite the reviews and surveys that have been conducted and reported in the literature, several limitations have been identified and summarized below.

1. The aforementioned studies followed the formal literature review approach and did not include any research questions, search strategies, data extraction processes, and data analysis. Hence, there is a need for a more systematic approach to reviewing the existing knowledge in BYOD security implementation.
2. The existing review provided a systematic study only on the security threats, authentication, mitigation mechanism, and privacy of BYOD security implementation.
3. A few of the studies concentrated on the conventional BYOD security models for mitigation strategy and none of the reviews provided a comprehensive review of BYOD security mechanisms currently in place, such as the machine learning methods for secure BYOD solutions.

However, the drawbacks mentioned above, identified in the current reviews have motivated the authors and necessitate the proposed study, which aimed to conduct a systematic mapping review to analyze the existing research literature that focuses on the application of the machine learning approach as a mitigation mechanisms implementation in BYOD security threats and

attacks. The purpose of conducting this systematic mapping review (SMR) is to provide an adaptable and reliable evaluation of a BYOD implementation based on the artificial intelligent research domain (Juárez & Cedillo, 2017). In this investigation, the SMR study was conducted by answering four formulated research questions. A total of 753 articles published from 2012 to 2021 were initially retrieved from 4 major academic databases. However, by following the screening process with the consideration of inclusion and exclusion criteria, 40 primary studies were selected.

The major contributions of this study are outlined below:

1. Detailed background study of the existing security threats and attacks in the BYOD environment
2. A comprehensive review of the machine learning techniques for addressing the security threats and attacks in BYOD environments in the aspects of the dataset usage, the machine learning approaches, and the performance metrics.
3. Analysis of the security threats and attacks, machine learning approaches, datasets, and evaluation metrics based on the selected primary studies for the systematic mapping review
4. Illustration of the percentage distributions and bubble plots to demonstrate the research trends on the BYOD threats and attacks, machine learning approaches, datasets, and performance metrics that are utilized in BYOD security implementation.

The rest of the article is divided into 7 sections. Section 2 presents the review method. Section 3 describes the machine learning approaches for BYOD security implementation. In section 4, the systematic mapping result is provided. Section 5 presents the discussions of findings from the systematic mapping review. In section 6, the threats to validity are presented whereas section 7 gives the concluding remarks and future direction.

2. Review Method

Systematic mapping is a process of exploring the existing studies to obtain overview results and the type of research that has been conducted in a particular research area. Thus, it identifies the type of research, the quality of research, and the available output. It displays the publication trends by mapping publication frequencies with time. Besides, it provides a summary of the research domain. According to Petersen et al. (2008), " systematic mapping provides a structure of the types of reports and results that have been published by classifying them, and it often provides a visual summary through the mapping of its results."

Conversely, a systematic literature review investigates relevant existing literature in a particular research field and carries out an in-depth review, evaluation, interpretation, and description of the methodology and results (Keele, 2007). Over the years, several researchers have embraced and followed the systematic review guideline provided by Petersen for conducting the review. Though a systematic mapping study is most applicable in the software engineering domain, it is not limited to the software engineering field. It has been acknowledged in other research fields by employing the same guideline provided by Peterson. This guideline has provided a lot of benefits in describing a study domain or sub-domain (Abdelmaboud et al., 2015; Cavalcante et al., 2016; Fernandez et al., 2015).

In addition, the adoption of Peterson guidelines with a Kitchenham systematic literature review has become common in conducting a systematic mapping review. Adopting this new approach will further enrich the previous approach by making systematic mapping studies more comprehensive and obtaining a profound conclusion. Therefore, this study combines the Peterson systematic mapping review guideline (Petersen et al., 2008) and Kitchenham systematic literature review guideline (Kitchenham & Brereton, 2013). The systematic mapping process is usually performed in five different major phases, where the output of each phase provides the input for the next phase. The illustration of the phases is depicted in Fig. 1 as demonstrated by Petersen et al. (2008)

Phase 1: Definition of research questions and the corresponding research objectives

Phase 2: Definition of the search strategy and relevant studies selection process.

Phase 3: Performing the screening and selection criteria (inclusion and exclusion)

Phase 4: Performing the classification scheme, which is the core structure of the systematic mapping.

Phase 5: Data extraction and mapping process based on the classification scheme to show the research trends

2.1 Research Questions and the Corresponding Research Objectives

Kitchenham and Brereton (2013) maintained that research questions are expected to determine the problems being addressed and the aim of the research method. This study aims to explore the existing primary studies focusing on the security threats and attacks on BYOD, and the application of the machine learning approach as mitigation mechanisms on the threats and attacks on the BYOD environment to identify the research trends, open issues, and further research in the domain. With regards to the necessity of conducting a systematic mapping study and to achieve the aim of this research, the following research questions and the corresponding research objectives have been defined as depicted in Table 1

Table 1
Research questions and objectives

	Research Questions	Objectives
RQ1	What are the potential security threats and attacks that are found in the BYOD environment?	To examine the potential security threats and attacks that are found in the BYOD environment.
RQ2	What are the different machine learning algorithms employed for detecting and combating security threats and attacks in BYOD environments?	To identify various machine learning techniques employed for detecting and combating the security threats and attacks in the BYOD environment
RQ3	What datasets do researchers employ in machine learning algorithms for detecting and combating security threats and attacks in the BYOD environment?	To identify various datasets that have been employed by researchers in the machine learning approach for detecting and combating security threats and attacks in the BYOD environment
RQ4	What are the evaluation metrics that are employed to evaluate the performance of machine learning algorithms for detecting and combating security threats and attacks in the BYOD environment?	To investigate the evaluation metrics that are employed to evaluate the performance of machine learning algorithms for detecting and combating security threats in the BYOD environment.

2.2 Search Strategy and Relevant Studies Selection Process

Our search for suitable articles involves three sequence activities, which include keyword identification, search strategy formulation, and data source selection. The keywords were identified and the necessary search strategy was created based on the research question's content. However, the keywords were enhanced after the initial search. Some keywords were merged and searching was conducted in various iterations. Published articles covering 2012–2021 were included in the current study. The time frame was chosen because it was the period that the enterprise permitted consumer devices on the enterprise network (Micro, 2012).

To obtain suitable articles, we considered relevant articles published in five academic databases, which include ACM Digital Library, IEEE Xplore, Springer, and Science Direct. The choice of the database selection is based on Petersen's suggestion (Petersen et al., 2008). Nonetheless, the combination of the overarching nature of the above databases also provides access to BYOD literature from the related disciplines. In addition, the Google Scholar database was also employed to complement the databases mentioned above to facilitate the broad search of the article that may have been skipped while using the proposed databases.

The search strategy began with wide coverage by using keywords to query for articles on “Application of Machine Learning Approaches for Mitigating Bring Your Own Device (BYOD) Security threats and attacks ". In the search terms and synonyms were formulated based on the related studies using bring your own device, security threats and attacks, and machine learning techniques. All these terms and synonyms were included in the Search Query (SQ), which is shown below:

Query_1 = "bring your own device" OR "b.y.o.d" OR "mobile device" OR “tablets” OR "personal device" OR "personal smartphone" OR "personal mobile" OR “notebooks” OR "personal laptop" OR "personal tablet".

Query_2 = "spoofing attack” OR “intrusion attack” OR “malware” OR “dos attack”, “eavesdropping” OR “man-in-middle attack” OR “advanced persistent threat” OR “phishing attack” OR “lost device” OR “viruses".

Query_3 = "machine learning" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "clustering" OR "semi-supervised learning" OR "reinforcement learning".

SQ = Query_1 AND Query_2 AND Query_3

The queries using the search string were employed on the selected database to retrieve the academic literature.

Selected articles were used for the snowballing process. A search strategy known as "snowballing" leverages the currently obtained articles to find new ones. There are two ways to achieve this: either by looking at the publication's reference list (also known as "backward snowballing") or by seeking articles that cited the identified publications (a.k.a., forward snowballing). During this SLR study, both forward and backward snowballing have been done. The study selection criteria were also applied to publications that were found through a snowball search. The list of chosen articles now included the publications that passed the check. This returned a total of 753 articles. Table 4 shows the description of the search process and the final obtained number of studies.

2.3 Screening and Selection Criteria

This study adopted PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) guidelines for the outcome report of the search results to elucidate suitably, excluded, or included primary studies in the analysis. The final selected literature after the screening process is depicted in Fig. 3 based on the guideline provided by PRISMA. As stated in the previous section, five databases were chosen to search the relevant studies, which include ACM Digital Library, IEEE Xplore, Springer, Google Scholar, and Science Direct. After the keywords were searched in the aforementioned online databases, a total of 753 articles were extracted. The search output was comparatively large, but it is a normal characteristic of this form of research (Kitchenham et al., 2009). Endnote was employed as a software reference manager software to manage the articles. At the end of queries, duplicate articles were recognized and eliminated accordingly, leaving a total of 450 articles in which the further refined method of the articles selection process was introduced.

This paper selection stage was performed in two different stages. In the first stage, the selection was done by checking the titles and abstracts of the articles based on the inclusion and exclusion criteria (Table 3) and eliminating the unrelated articles. At this stage, 450 studies were retained. In the second stage, the selection was made by reading the full texts of the included papers and this stage returned 65 articles. Therefore, an in-depth reading of each article and an analysis of the 65 studies were performed to verify if they indeed contribute to the aim of the study. Finally, 40 articles were used in the present systematic mapping review as depicted in Fig. 3.

Table 2
Inclusion and exclusion criteria

Inc/Exc	Inclusion Criteria
Inc1	Articles must have been written in the English language
Inc2	Articles publication must be between 2012–2021
Inc3	Articles must be a research article related to BYOD security implementation using machine learning/deep learning approach such as Journal, conference paper, or published thesis
Inc4	Articles must have mention security threats/attacks in the BYOD environment
Inc5	Articles must have mention machine/deep learning performance evaluation in the BYOD environment
	Exclusion Criteria
Exc1	Articles that did not study BYOD/personal device/mobile device/ smart device security
Exc2	Articles that consist of reviews, abstracts, presentations
Exc3	Articles that do not meet up with any of the inclusion criteria
Exc4	Articles that are not available/accessible in electronic format

Table 3
Selection and screening process from 5 academic databases

Database	Initial Search Results	Results after Duplicate removal	Screened results based on inclusion/exclusion criteria	Screened results based on full-text
ACM Digital Library	150	90	10	4
IEEE Xplore	175	152	30	21
Science Direct	105	78	5	2
Springer	198	110	10	7
Google Scholar	125	20	10	6
Total	**753**	**450**	**65**	**40**

The distribution of the forty (40) selected primary studies for analysis is depicted in Fig. 2. Out of the 40 selected studies, as can be seen in Fig. 2, 21 of the studies were selected from IEEE Xplore, 4 studies from ACM digital library, 6 studies from the Google Scholars, 2 studies from Science Direct, and 7 studies from the Springer.

2.4 Classification Scheme

The classification scheme of this study was established based on the Petersen (Petersen et al., 2008) guideline, which comprises the activity of studying the abstracts of the articles, searching for keywords and ideas that reveal the contribution provided by the primary study (Petersen et al., 2008). The classification scheme aims to help in developing a categorization scheme that represents the primary population study, building a sophisticated knowledge of the nature and contribution made by each selected study, and guaranteeing that the expected results are included in the SMR (Fatima & Colomo-Palacios, 2018; Petersen et al., 2008). The classification scheme is illustrated using the keywording. The classification scheme is shown in Fig. 3 whereas the classification process is shown in Fig. 4. The process involved in the classification scheme are itemized below:

√ Keywording: The keywording involves the activity of reading the abstract and looking for keywords that characterize the context associated with the SMR objectives (Fatima & Colomo-Palacios, 2018)

√ Article sorting into the classification scheme: This is the activity of sorting the classification scheme after adding articles to it (Fatima & Colomo-Palacios, 2018). However, the production of the classification scheme is carried out by reading the introduction and conclusion parts of each selected primary study.

√ Classification scheme update: This is the activity of modifying the scheme after the addition of a primary study in the classification scheme. (Fatima & Colomo-Palacios, 2018). The classification process is depicted in Fig. 4.

The research question of this study was addressed by extracting information from the selected articles using the following categorization.

1. Security threats and attacks
2. Machine learning approaches
3. Datasets.
4. Performance metrics

1. Threats and Attacks

√ Threats: Threats are potential security disruptions that occur where there is an entity, action, or occasion that can break up security and cause damage. Thus, it is something that brings about vulnerability to security (Stallings, 2006).

√ Attacks: A violation of information security that is caused by an intelligent threat. It is an effort to get illegal access to data or resources in a malicious form with the aim to harm the information systems (Stallings, 2006).

2. Machine learning approaches: These are the learning algorithm employed to mitigate the security threats and attacks in BYOD environments.
3. Datasets: These are various datasets employed by researchers in machine learning security implementation in the BYOD environment for mitigating security threats and attacks.
4. Performance metrics: Performance metrics are the performance parameters employed to measure the machine learning security implementation in the BYOD environments.

2.5 Data Extraction and Mapping Process

In the data extraction phase, we adopted the established SMR approach provided by (Petersen et al., 2008) for data collection. Moreover, the classification scheme on the machine learning approach already formulated was utilized to sort the actual data extracted from the relevant literature into the scheme. The documentation of the data extraction process was carried out using an excel spreadsheet. Next, the analysis of the publication frequencies in each classification was performed using the table (Petersen et al., 2008). To examine the trends in each category, the publication frequencies were the main focus. This is to recognize the categories that have been emphasized in the existing research and consequently, to find the gaps and possible research directions. The analysis and results presentation are performed by following the process itemized below.

a. Illustrate the statistical summary of the data using tables by indicating the publication frequencies in each classification scheme (Petersen et al., 2008).
b. Report the publication frequency using a bubble plot (which is two X-Y scatter plots with the bubble in category insertions). The bubble size is proportional to the total number of articles that belong to the pair of categories, which matches the bubble coordinates(Petersen et al., 2008).

3. Machine Learning Techniques For Detecting And Combating Security Threats And Attacks In Bring Your Own Device (Byod) Environment.

Machine learning is a branch of artificial intelligence that models the extracted data to produce the expected future. Additionally, the computer algorithm should receive a set of instructions to understand the nature of the data. The core concept of machine

learning is algorithm design that offers the machine to identify the set of data and classify it based on the attributes of the data. The learning process occurs by using data extracted by the algorithm after removing some noise (Conway & White, 2012). The classification techniques help the learning algorithm to make an effective decision. Machine learning is capable of evaluating past and existing risks to obtain improved future performance (Blum & Langley, 1997). There are four major types of machine learning algorithms that are usually employed in the BYOD security implementation, including supervised, unsupervised, reinforcement, and deep learning, which are briefly described below:

Supervised learning approaches can be utilized for detecting threats and attacks in a BYOD environment, and to create a countermeasure. Supervised learning is the most useful learning algorithm in ML where the output is classified according to the input by employing trained data for the algorithm to learn. Supervised learning is of two categories which include classification and regression learning (Tahsien et al., 2020). Classification is a type of machine learning algorithm, whereby the output is a fixed or categorical value, which could be represented as [yes or No], or [True or False]. Examples of supervised classification learning algorithms include support vector machines, decision trees, random forest, k-nearest neighbor, association rule, and Bayesian theorem. Regression learning on the other hand is a type of supervised learning whereby the learning output is a continuous value depending on the input variables. Some examples of the regression-learning algorithm include neural networks, Decision Trees, Ensemble Learning, etc.

Unsupervised learning is a type of learning algorithm employed in complex data analysis and categorization. In Unsupervised learning, there is no target data for a given input value. This type of learning does not require labeled data and can examine the unlabeled data and categorizes the data into different groups as clusters. Various unsupervised learning techniques have been employed for BYOD security for privacy protection using the infinite Gaussian mixture model (IGMM) to detect DoS attacks using multivariate correlation analysis (Tan et al., 2013). Some examples of unsupervised learning algorithms include k-means clustering and principle component analysis.

Semi-supervised learning on the other hand comprised the combination of both supervised and unsupervised learning algorithms (Shah & Shankarappa, 2018). Thus, the semi-supervised learning algorithm sits between supervised and unsupervised learning, having the ability to deal with the labelled datasets and unlabeled datasets for all the observations. In some practical circumstances, the labelling of the dataset is quite high since it needs human expert opinion to perform the labelling. Thus, when the majority of the observation does not require labelling of data but a few of them, semi-supervised learning deems to be the suitable algorithm for model construction (Hussain et al., 2020).

Reinforcement machine learning is a type of learning that is usually employed in the gaming environment. In this form of learning, the algorithm learns based on the interaction with its environment (similar to human interaction) by executing an action that increases the overall feedback (Mnih et al., 2015). However, the feedback might be a return that relies on the performing task output. In reinforcement learning, there is no initial action for any task to be performed while the algorithm utilizes trial and error methods. Thus, the learning agent can recognize and implement the best method from its experience to obtain the best reward based on trial and error.

A subset of machine learning algorithms also referred to as deep learning is another learning model usually employed in the implementation of BYOD security. Deep learning is a machine learning approach that comprises an architecture, which is centered on artificial neural networks (ANNs). Artificial neural networks are supervised deep learning algorithm that is stimulated by the brain. However, it does not imply that the ANNs work basically as the biological brain. The neural network consists of neurons (referred to as variables) connected via weighted connections (usually regarded as parameters). The network is connected with either a supervised or unsupervised learning approach to attain the desired performance results. The learning itself is performed by employing the labelled and unlabeled data respectively from the supervised and unsupervised learning approaches followed by the iteration modification of the weights among every pair of neurons. Thus, while describing deep learning, we refer to a larger neural network where the term deep denotes the number of that network layers (Yang et al., 2014). In the early times of artificial neural networks, it was hard to train the network because of the constraints in computational powers, even for relatively networks. However, the advancement of technology has brought about more effective methods such as graphical user interfaces (GPUs) for estimating the optimal network weights, which permits the construction of larger networks containing more hidden layers. Even though it is not a severe rule, artificial neural networks that contain more than one hidden layer are regarded as deep learning

models. Some deep learning models used in the BYOD implementation are the convolutional neural networks, recurrent neural networks, autoencoders, etc.

3.1 Review of Machine Learning Techniques for Detecting and Combating Security Threats and Attacks in Bring Your Own Device (BYOD) Environment

This section provides a review of machine learning techniques for security threats and attack implementation in the BYOD environment. The review is based on some aspects of BYOD machine learning implementations such as the dataset used, the machine learning algorithm employed, and the performance measures adopted. Specifically, this section is categorized into three major subheadings. Section 3.1.1 looks into the different datasets used in machine learning-based approach implementation for BYOD security threats and attacks. Section 3.1.2 discusses the various machine learning algorithms employed for the implementation of BYOD security threats and attacks while section 3.1.3 reassesses the different performance evaluation metrics considered by different authors to assess the performance of machine learning implementation of BYOD security threats and attacks. This section followed the same concept used in (Christopher Ifeanyi Eke et al., 2019). The summary of the review is shown in Table 1.

3.1.1 Review of Datasets employed in Machine Learning Algorithm for Detecting and Combating Security Threats and Attacks in the BYOD Environment.

Learning models generally are based on the past occurrences or experiences of an event or scenario. This scenario is referred to a dataset which is a key element used to train, test, and implement BYOD security models. Thus, the first step in an attempt to implement machine learning techniques for detecting and combating security threats and attacks in the BYOD environment is dataset gathering. The findings as summarized in Table 4 demonstrate different datasets utilized to implement the machine learning approach for detecting and combating security threats and attacks in the BYOD environment. The selected study's analysis shows that datasets can be generally classified into homogeneous and heterogeneous data. When the author uses one type of dataset, it is termed a homogeneous dataset. On the other hand, when more than one type of dataset is used to perform machine learning for detecting and combating security threats and attacks in the BYOD environment, it is called heterogeneous data. Thus, the review of the datasets utilized according to the nature of the dataset used is explained as follows.

a Homogeneous Datasets

In a homogeneous dataset, the authors employed only one type of dataset. For instance, Shah and Shankarappa (2018) utilized a homogenous data source called the MDM events log. The MDM here is a scheme implemented in a BYOD environment to control and monitor the role of smartphones including their data operations. In a separate study, Chizoba et al. (2020), used homogenous data generated from network traffic logs when packets are transferred between networks. The network logs were used to implement the machine learning approach for BYOD security threats and attacks. Muhammad et al. (2017), leveraged the IAT packet data gotten from the local network of the Institute of Technology Georgia. The packet IAT data of 27 mobile devices were collected using UDP, TCP, and ICMP protocols. The dataset is homogeneous in nature as it contains only the inter-arrival time of packets sent in a BYOD environment. In a related study Muhammad et al. (2019), employed a test-bed dataset carefully gathered via mobile devices without meddling. The dataset contains Inter-Arrival Times of 27 mobile devices such as tablets, laptops, and smartphones to evaluate device type profiling. The data is homogeneous in nature as it contains only the inter-arrival time of 2 successive packets. In another study, Petrov and Znati (2018), utilized the MIT dataset that is made up of 84 issues of phone event records such as call start time, incoming / outgoing direction including the type of calls (phone, data, or message call), etc. Eslahi et al. (2016), in their study on botnet detection, utilized a network traffic dataset generated from a mobile botnet. However, a data sieving approach is employed in the model to gather only the HTTP traffic records only during HTTP and server communication. In another research conducted by RIASAT et al. (2017), a publicly available android malware dataset gathered from the Contagio mobile was utilized. The data contain 600 samples composed of two segments (reptiles crawling and the malicious applications of the contagio library.

b Heterogeneous Datasets

In a heterogeneous dataset, the data are obtained from various sources. For instance, the study conducted by (Arora & Bhatia, 2019) utilized several different datasets such as FVC2006, ATVSFFpDB, Spoofing-Attack Finger Vein Database, and LivDet 2013 fingerprint Datasets, to experiment with the model. However, datasets from LivDet 2015 were used to test the model. These datasets are heterogeneous as they consist of different fingerprint biometrics, which originated from different sources. In another study, Yerima et al. (2013), used 2000 samples of malware and benign dataset out of which malware consists of 1000 and benign takes the other half of 1000 samples. The authors asserted that there is high variability in the malware samples than in the benign samples when 20 features were chosen. In a related study, Chen et al. (2016), employed two distinct dataset sources (benign and malware) which amount to 7,970 samples. The benign sample is made up of 4,350 while the malware sample constitutes 3,620 samples. The analysis stating which dataset achieved a better result is not considered in the study. Similarly, Lashkari et al. (2017) in their proposed framework for android malware characterization and detection utilized both benign and malware datasets for machine learning classification to train the model. The author collected 1527 benign apps from the google play market between 2015 and 2016. The collection of these apps relies on how popular they are for each category present in the market. However, the author stated that 27 of the apps were eliminated before the modeling phase because they were classified as suspicious by the two different anti-virus products. On the other hand, 400 malware apps were collected based on two classes (adware, containing 250 apps, and general malware consisting of 150 apps). However, the adware category is consists of different families, including Airpush, Dowgin, Kemoge, Mobidash, and Shuanet. Finally, the author utilized Droidkin, which is a lightweight android apps similarity detector to find the relationship based on the category of each apps dataset (general malware, adware, and benign).

3.1.2 Review of Machine Learning Algorithm for Detecting and Combating security Threats and Attacks in the BYOD Environment.

Based on these research findings, different machine learning approaches have been employed in the implementation of BYOD security. A comprehensive summary of the findings is presented in Table 1, illustrating the different algorithms used by different researchers for implementing security threats and attacks in a BYOD environment. It is observed that certain authors use several algorithms, to determine which algorithm performs better. Consequently, Shah and Shankarappa (2018), employed multiple algorithms which include SVM, MLP, BN, and RF out of which the SVM algorithm outperformed the other three algorithms returning low false positive, low true negative, and highest performance accuracy. Thus, SVM stands tall in terms of BYOD security threats and attack implementation based on their findings. Chizoba et al. (2020) utilized SVM, DT, RF, and ensemble algorithms. Ensemble learning is used to combine the performance of the other three individual algorithms. However, the RF algorithm put up the best vote using the ensemble combination model. Similarly, Naive Bayes, RF, and SVM algorithms were adopted by Sokolova et al. (2017), for anomaly detection in BYOD environments. The authors reported only the results achieved using the NB model because it performed far better than the other schemes. Muhammad et al. (2017) modelled an intelligent filtering approach for BYOD security using K-means to isolate incidents towards uncovering different clusters of normal behaviours from abnormal behaviours in a BYOD environment. In a related study, by Muhammad et al. (2019), the author leveraged the Clustering-based Multivariate Gaussian Outlier Score (CMGOS) to identify irregular device behaviours. CMGOS constitute clustering and density approximation schemes. In clustering, the K-means algorithm was employed the while the density approximation used a multivariate Gaussian algorithm. The K-means scheme recorded some inconsistencies in the result. Hence, k-means was used to organize the limits (Centroid 1 and 2) and the outcome serves as input to the density approximation to implement the model. To control unauthorized access in the BYOD environment, Petrov and Znati (2018), laid hold on the artificial neural networks and decision tree algorithms to detect any un-authorize effort to get into delicate information by adversaries. In addition, the model further perplex or confuse their access to secure the data. Eslahi et al. (2016), leveraged the J48 form of Decision Tree (DT) to categorize the data and hence analyze the network behaviour. The J48 DT can proficiently detect recurring events in a mobile HTTP Botnet. Yerima et al. (2013), employed the Naïve Bayes classifier to identify malware in android devices. The authors noted that the Bayesian model is capable of performing both expert and learning schemes much better than other learning algorithms. In another study, Chen et al. (2016), used multiple learning algorithms including SVM, DT, ANN, NB, K-NN, and Bagging predictor to detect malware in an android environment. The performance result indicates that the KNN algorithm outperformed the other learning algorithms. RIASAT et al. (2017), adopted the use of SVM and random forest learning models to detect the behaviour of android malware. The authors noted with experimental facts that the RF algorithm produced a better result as compared to the SVM algorithm given the same processing time interval. The K-Nearest Neighbors algorithm was employed in the (Gangwal & Conti, 2019) study for categorizing time series to detect crypto converts in mobile environments. The model operates with or without access rights to the suspicious gadget.

3.1.3 Review of Evaluation Metrics that are employed to evaluate the performance of Machine Learning Algorithm for Detecting and Combating Security Threats and Attacks in the BYOD Environment.

Performance measures are the metrics that are used to evaluate the performance of machine learning classification on the BYOD security threats and attacks. The authors employed several evaluation metrics to ascertain the performance of BYOD models. Performance metrics such as accuracy, precision, recall, F-score, etc. were employed by researchers to evaluate the performance of the machine learning model on the BYOD security implementation. These metrics can be calculated by employing the values of false positive (FP), false negative (FN), true positive (TP), and true negative (TN), which constitute the components of the confusion matrix. The option of evaluation metric to be selected depends on the researcher's aim and expertise. In this regard, Yerima et al. (2013), utilized several metrics including false negative, true positive, false positive, true negative, precision as well as accuracy and error rate. These metrics were used to measure the performance of the model at different folds and the authors maintained that 15 to 20 features can provide good performance. In another study, Eslahi et al. (2016), used the accuracy, detection rate, and false alarm evaluation metrics to assess the performance of the model. The metrics yielded 98.60, 96.35, and 1.25 percent results respectively. In a separate study, Chen et al. (2016), leveraged the true positives, false positives, ROC, precision, recall, and accuracy metrics to ascertain how well the model can detect malware in an android environment to assess the performance of the Android malware detection model. Aneja et al. (2018), utilized the accuracy evaluation metric to assess the performance of the model, which showed an overall accuracy of 86.7 percent. The study by Daniel et al, 2018, utilized recall, precision, and accuracy for evaluating the performance of the model. The model returns a reliable accuracy/precision performance result of over 99 percent at each run time. Shah and Shankarappa (2018) used TP, TN, FP, FN, and Accuracy metrics to assess the performance of the BYOD security model developed. In a separate study, Sokolova et al. (2017) relied on the true positives, false positives, false negatives and true negatives evaluation metrics to assess the performance of the BYOD security model built. Muhammad et al. (2019), leveraged outlier secure accuracy to ascertain the performance of its BYOD scheme. The performance result shows that for 9; 100; and 324 IAT points, 99.3%, and 0.7% outlier secure accuracy was achieved in normal and abnormal profiling respectively. In the same year, Arora and Bhatia (2019), employed the use of performance metrics such as false acceptance rate, false rejection rate, accuracy, and average classification error. For each evaluation metric and dataset, a corresponding performance result was achieved. Similarly, Standard evaluation schemes such as accuracy, precision, recall, and f1-score were adopted in the study conducted by (Gangwal & Conti, 2019) to assess the performance of a BYOD model. Precision and f-measure metrics yielded an average of 88 and 87 percent accuracy respectively. Accordingly, Chizoba et al. (2020), in their study to identify advanced persistent threats using ensemble classifiers, employed several evaluation metrics such as true positive, false positive, precision and recall, others include f1-score, MCC, ROC and PRC. The authors used all these aforementioned metrics to carefully assess the performance of the developed BYOD security scheme.

Based on some of the reviewed studies, it shows that most of the related studies employed accuracy, recall, precision, and f-measure to evaluate the performance of the machine learning model. However, employing only such metrics may not be enough due to the imbalance in the dataset in some cases. Thus, the best metric to evaluate the model in such an instance is AUC.

Table 4
The review summary of the machine learning approaches implementation for BYOD security threats and attacks

S/N	Author	Year	Datasets	ML Approach	Threats /Attacks Detected	Performance Measure	Area of Attack	Database
1	(Shah & Shankarappa, 2018)	2018	MDM event log	SVM, MLP,BN, RF	Privacy breaches, Data leakage	ACC	Device	IEEE
2	(Chizoba et al., 2020)	2020	Network traffic	SVM, DT, MV ensemble	Persistent threat	PRE, REC, F-M, AUC	Network	Google scholar
3	(Sokolova et al., 2017)	2016	9512 application and set of malware	RF, SVM, NB	Malware	REC, F-M, AUC, ACC	Network	Science direct
4	(Muhammad et al., 2019)	2019	Test-bed data containing 94 files of packets IAT	K-means clustering	unauthorized access	ACC	Network	ACM
5	(Muhammad et al., 2017)	2017	IAT packets data	K-means clustering	Unauthorized access, data leakage, data theft	ACC	Network	ACM
6	(Mora et al., 2014)	2014	URL Session dataset	RF, J48, PART/NNge, Reduce and Pruning tree	Unauthorized access	ACC	Network	Google Scholar
7	(Ho, 2014)	2014	Smartphone sensor data	Manhattan distance classifier, RF, Gaussian Discriminant Analysis (GDA) SVM	Data Theft	FRR	Device	Google Scholar
8	(Shabtai et al., 2012)	2012	Event log	K-means, RR, DT, BN, NB	Malware	TPR, FPR, ACC, AUC	Device	Springer
9	(Kumar et al., 2020)	2020	Social network IoT nodes	DNN	Untrusted network	ACC	Device	Springer
10	(Arora & Bhatia, 2019)	2019	fingerprint benchmarks	DCNN	Spoofing attack	FAR, FRR, ACE and ACC.	Apps	Springer
11	(Samarathunge et al., 2018)	2018	Email dataset	KNN	Malware	ACC	Email application	IEEE
12	(Petrov & Znati, 2018)	2018	Phone records of subjects	ANN, DT	Malware	ACC, PRE, REC	Data	IEEE
13	(Eslahi et al., 2016)	2016	Mobile botnet dataset	DT	Botnet	ACC, false alarm	Device	IEEE

S/N	Author	Year	Datasets	ML Approach	Threats /Attacks Detected	Performance Measure	Area of Attack	Database
14	(Gangwal & Conti, 2019)	2019	Profiled Crypto Currency mining sample magnetic field data	KNN	Un-authorized access	ACC, REC, PRE, F-M	Device	IEEE
15	(Aneja et al., 2018)	2018	Device finger print packed	CNN	Spoofing attack	ACC	Device	IEEE
16	(Joshi et al., 2016)	2016	NSL-KDD dataset	SVM	DOS attack, intrusion attack	ACC	Device	IEEE
17	(Yerima et al., 2013)	2013	APKs made up of Benign apps and Malware samples	BN	Malware	ERR, ACC, TNR, FPR, TPR, FNR, PRE, AUC	Device	IEEE
18	(Chen et al., 2016)	2016	Malicious Apk files	SVM, C4.5, MLP, NB, K-NN, IBK) and Bagging	Malware	ACC	Apps	ACM
19	(RIASAT et al., 2017)	2017	Apk files, android malware public datasets	SVM, RF	Malware	ACC	Apps	Google scholars
20	(Sahs & Khan, 2012)	2012	Benign and malicious android applications	SVM	Malware	ACC, PRE, REC, F-M	Apps	IEEE
21	(Akhuseyinoglu & Akhuseyinoglu, 2016)	2016	Traffic data log	NB	Malware	ACC and kappa statistics	Mobile devices	IEEE
22	(Tan et al., 2020)	2020	Network traffic data logs	MLPs	Malware	ACC	Network	IEEE
23	(Kyriazis, 2018)	2018	Apache Spark dataset	K-means clustering	Malware	NIL	Cloud environments	IEEE
24	(Tout et al., 2019)	2019	Real-time generated dataset	LR, SVR, NN and DNN	Device overheads	RMSE	Device resources	Science Direct
25	(San Miguel et al., 2018)	2018	Drebin[2] and Androzoo[3] repositories	DT, SVM, KNN, and NB	Malware	ACC, PRE, REC, F-M	Network	ACM
26	(Temper et al., 2015)	2017	Biometric sample dataset	Fuzzy Rough Nearest neighbor	User privacy breaches	Equal Error Rate (EER)	Mobile devices	IEEE

S/N	Author	Year	Datasets	ML Approach	Threats /Attacks Detected	Performance Measure	Area of Attack	Database
27	(Wang et al., 2017)	2018	Network traffic data	SVM	Malware	F-M	Devices	IEEE
28	(Chukka, 2020)	2020	APK files	MNB, RF, SVM	Malware	PRE, REC, F-M	Device, Application	Google scholar
29	(Kotak & Elovici, 2019)	2021	Network traffic data	Neural Network	Unauthorized access	ACC, PRE, REC, F-M	Network	Springer
30	(Bai et al., 2021)	2021	Network traffic data	CNN	Malware	ACC, PRE, REC, F-M	Mobile network devices	Google scholars
31	(Narayanan et al., 2018)	2018	Malware, Benign and Wild datasets	SVM, SMO	Malware	ACC, PRE, REC, F-M	Network	Springer
32	(Saracino et al., 2016)	2016	Genome, Contagio-Mobile, and VirusShare Datasets.	LDC, K-NN, MLP, PARZC and RBF.	Malware	NIL	Mobile Network devices	IEEE
33	(Li et al., 2018)	2018	Benign dataset	SVM, PART, Random Forest	Malware	ACC, PRE, REC, F-M	Network, devices	IEEE
34	(Narayanan et al., 2017)	2017	Malware and Benign datasets	SVM and RF	Malware	ACC, PRE, REC, F-M	Network, Devices	IEEE
35	(Zhu et al., 2017)	2017	Benign, Malware And VirusShare datasets	CNN, RBM, DBN and RNN including Bayesian, SVM and MLP	Unauthorized access	PRE, REC, F-M	Networks	IEEE
36	(Pajouh et al., 2018)	2017	Malware and Benign samples	C5.0 and RF	Malware	PRE, REC	Network devices	Springer
37	(Malhotra & Bajaj, 2016)	2016	Malware sample data	ANN and K-Means	Malware	ACC, PRE, REC	Devices	Springer
38	(Das et al., 2015)	2016	Malware sample	J48, NB, LR, SVM, SMO, JRIP, MLP	Malware	AUC	network	IEEE
39	(Lashkari et al., 2017)	2017	127 benign app datasets and 400 malware datasets	KNN, RF, DT, RANDOM TREE and LR	Malware	ACC, PRE, FPR	Application	IEEE
40	(Anwar et al., 2016)	2016	UNBISCX public datasets	SVM, KNN, J48, BAGGING, NB, RF	Botnet	TPR, FPR, ACC	Application	IEEE

4. Results

The systematic mapping results and discussion on the machine learning-based technique for detecting and combating security threats and attacks in the BYOD environment is provided in this section. Table 4 depicts the list of the included primary studies in this research. However, a total of 40 articles were finally selected in this research, by considering published articles in the year, ranging between 2012 and 2021. Five academic databases were used to produce the primary studies, including ACM Digital Library (4), IEEE Xplore (21), Springer (7), Google Scholar (6), and Science Direct (2). Based on the analysis of the paper, it can be seen that IEEE produced the majority of the articles. Thus, the results of the mapping, which are grouped according to the formulated research questions (RQ1 to RQ4) are presented below.

RQ1: What are the potential security threats and attacks that are found in the BYOD environment?

To answer this question, Table 5 provides a summary of the existing security threats and attacks that have been implemented using the machine learning approaches in a BYOD environment. Based on the table, this study identified 12 threats and attacks in the BYOD environment that have been implemented using machine learning approaches. However, the percentage distribution of the selected studies is depicted in Fig. 5. The analysis and the demonstration in Fig. 5 show that the most security threats and attacks that have been implemented in the existing studies are malware with 23 studies (57.5%) and unauthorized access with 6 studies (15%) out of the 40 selected studies. On the other hand, spoofing attacks, botnet, and user privacy breaches are represented in 2 studies with 5% each. Conversely, other threats and attacks such as persistent threats, data leakage, data theft, DOS attack, intrusion attack, and untrusted network are all shown in one study each with 2.5%. Moreover, Fig. 6 shows the bubble plot of the security threats and attacks implemented using the machine learning approaches in the BYOD environment. In the plot, X-axis represents the security threats and attacks while Y-axis represents the year. It can be seen in the plot that malware and unauthorized access are gaining attention and dominance in the research domain. Research in malware attacks ascended in the year 2016 but later descended in the year 2020. However, much work has not been implemented using the machine learning approaches on the other threats and attacks such as persistent threats, data leakage, data theft, DOS attack, intrusion attack, and untrusted network in the BYOD environment.

Table 5
Existing security threats and attacks implemented using machine learning approaches

S/N	Threats and attacks	References
1	Malware	(Samarathunge et al., 2018), (Petrov & Znati, 2018), (RIASAT et al., 2017), (Yerima et al., 2013), (Chen et al., 2016), (Lashkari et al., 2017), (Sokolova et al., 2017), (Shabtai et al., 2012), (Sahs & Khan, 2012), (Akhuseyinoglu & Akhuseyinoglu, 2016), (Tan et al., 2020), (Kyriazis, 2018), (San Miguel et al., 2018), (Wang et al., 2017), (Chukka, 2020), (Bai et al., 2021), (Narayanan et al., 2018), (Saracino et al., 2016), (Li et al., 2018), (Narayanan et al., 2017), (Pajouh et al., 2018), (Malhotra & Bajaj, 2016), (Das et al., 2015)
2	Unauthorized access	(Muhammad et al., 2017), (Muhammad et al., 2019), (Gangwal & Conti, 2019), (Mora et al., 2014), (Kotak & Elovici, 2019), (Zhu et al., 2017)
3	Spoofing attack	(Arora & Bhatia, 2019), (Aneja et al., 2018)
4	Botnet	(Eslahi et al., 2016), (Anwar et al., 2016)
5	User privacy breaches	(Shah & Shankarappa, 2018), (Temper et al., 2015)
6	Persistent threat	(Chizoba et al., 2020)
7	Data leakage	(Shah & Shankarappa, 2018)
8	Data theft	(Ho, 2014)
9	DOS attack	(Joshi et al., 2016)
10	Intrusion attack	(Joshi et al., 2016)
11	Untrusted network	(Kumar et al., 2020)
12	Computation overhead	(Tout et al., 2019)

RQ2: What are the different machine learning algorithms employed for detecting and combating security threats and attacks in BYOD environments?

Due to the limitation of the conventional approach to security threats and attacks in the BYOD environment, machine learning aspects of artificial intelligence have been employed by previous researchers to mitigate the security threats and attacks in the BYOD environment. This study identified 14 machine learning approaches that have been employed to mitigate security threats and attacks, which include SVM, DT, RF, ANN, KNN, DNN, NB, BN, LR, GDA, LDC, RT, Clustering, and Ensemble. Figure 7 illustrated the percentage distribution of the machine learning approaches that have been implemented by previous researchers for security threats and attacks. However, it can be observed from Fig. 7 that the most employed machine learning approaches supervised machine learning algorithms that comprised of SVM, DT, and RF, with a percentage distribution of 20%, 14%, and 15% of the selected studies respectively. On the other hand, supervised neural networks, such as artificial neural networks, KNN, and deep neural networks (DNN) also attained a percentage distribution of 11%, 9%, and 7% respectively. It can also be noticed that NB and clustering approaches possess similar percentage distributions of 6% each out of the selected studies. Similarly, GDA, LCD, and RT possess similar percentage distributions of 1% in each of the selected studies.

Moreover, the bubble plot of the machine learning approaches that shows the research trend based on the analysis of the selected studies is depicted in Fig. 8. In the plot, Y-axis depicts the machine learning approaches whereas X-axis shows the year of the studies. However, it is obvious from the plot that the research trend in machine learning approaches such as SVM, DT, and RF started ascending in the year 2016 as it gained more attention in the years 2016, 2017, and 2018, but began to descend in the year 2019. Conversely, DNN and clustering approach maintained a consistent trend between 2016 and 2019, and both approaches were proposed (2021). Thus, the research domain is currently active and will thrive in the years to come.

Table 6
Machine learning approaches implementation for BYOD security threats and attacks

S/N	Machine Learning Technique	References
1	SVM	(Shah & Shankarappa, 2018), (Chizoba et al., 2020), (RIASAT et al., 2017), (Chen et al., 2016), (Sokolova et al., 2017), (Ho, 2014), (Joshi et al., 2016), (Sahs & Khan, 2012), (Tout et al., 2019), (Wang et al., 2017), (Chukka, 2020), (Narayanan et al., 2018), (Li et al., 2018), (Narayanan et al., 2017), (Zhu et al., 2017), (Das et al., 2015), (Anwar et al., 2016)
2	DT	(Petrov & Znati, 2018), (Chizoba et al., 2020), (Eslahi et al., 2016), (Chen et al., 2016), (Lashkari et al., 2017), (Mora et al., 2014), (Shabtai et al., 2012), (Saracino et al., 2016), (Li et al., 2018), (Pajouh et al., 2018), (Das et al., 2015), (Anwar et al., 2016)
3	RF	(Shah & Shankarappa, 2018), (RIASAT et al., 2017), (Lashkari et al., 2017), (Sokolova et al., 2017), (Mora et al., 2014), (Ho, 2014), (Chukka, 2020), (Li et al., 2018), (Narayanan et al., 2017), (Pajouh et al., 2018), (Anwar et al., 2016)
4	ANN	(Petrov & Znati, 2018), (Shah & Shankarappa, 2018), (Chen et al., 2016), (Tan et al., 2020), (Kotak & Elovici, 2019), (Saracino et al., 2016), (Zhu et al., 2017), (Malhotra & Bajaj, 2016), (Das et al., 2015)
5	KNN	(Samarathunge et al., 2018), (Chen et al., 2016), (Lashkari et al., 2017), (Gangwal & Conti, 2019), (Tout et al., 2019), (Temper et al., 2015), (Saracino et al., 2016), (Anwar et al., 2016)
6	DNN	(Arora & Bhatia, 2019), (Aneja et al., 2018), (Kumar et al., 2020), (Bai et al., 2021), (Saracino et al., 2016), (Zhu et al., 2017)
7	NB	(Shabtai et al., 2012), (Akhuseyinoglu & Akhuseyinoglu, 2016), (Tout et al., 2019), (Chukka, 2020), (Das et al., 2015)
8	Clustering	(Muhammad et al., 2017), (Muhammad et al., 2019), (Shabtai et al., 2012), (Kyriazis, 2018), (Malhotra & Bajaj, 2016)
9	BN	(Shah & Shankarappa, 2018), (Yerima et al., 2013), (Sokolova et al., 2017), (Shabtai et al., 2012)
10	LR	(Lashkari et al., 2017), (Tout et al., 2019), (Das et al., 2015)
11	ENSEMBLE	(Chizoba et al., 2020), (Chen et al., 2016)
12	GDA	(Ho, 2014)
13	LDC	(Saracino et al., 2016)
14	RT	(Lashkari et al., 2017)

RQ3: What datasets do researchers employ in machine learning algorithms for detecting and combating security threats and attacks in the BYOD environment?

To answer the RQ3, this study identified and classified the datasets utilized in the selected studies into 14. Table 7 illustrates the identified datasets and the corresponding studies that used them. Figure 9 depicts the percentage distributions of various datasets used in the selected studies. It is obvious from Fig. 9 that malware samples and benign apps with 27% and network traffic data with 15% are the most used datasets in the selected studies. In addition, the second most used datasets are APK files with 10% and publicly available datasets with 10% on the selected studies. However, the presence of the publically available datasets shows that some research in the research domain did not collect their own datasets by themselves for the experiments but rather, utilized the publicly available datasets such as the NSL-KDD dataset, UNBISSCX, etc. (Table 1 provides the details). Figure 9 also shows that datasets such as Apache Spark, phone records, email data, IoT node data, smartphone sensor data, URL session, and mobile botnet data are the least used datasets in the selected studies with 3% each.

Moreover, Fig. 10 illustrates the bubble plot of the used datasets in the selected studies. In the plot, X-axis represents the year of the studies whereas the Y-axis represents the used datasets. However, the plot shows that there is an increase in the number of studies on the utilization of malware sample and benign apps, network traffics, APK files, publically available datasets, and

biometric sample data as they gained researchers' attention between the year 2016 and 2020. In contrast, there are fewer studies on the utilization of email data, phone record data, and Apache Spark data.

Table 7
Datasets used in the selected studies

S/N	Datasets	References
1	Malware sample and Benign apps	(Yerima et al., 2013), (Sokolova et al., 2017), (Lashkari et al., 2017), (Sahs & Khan, 2012), (Narayanan et al., 2018), (Li et al., 2018), (Narayanan et al., 2017), (Zhu et al., 2017), (Pajouh et al., 2018), (Malhotra & Bajaj, 2016), (Das et al., 2015)
2	Network traffic	(Chizoba et al., 2020), (Akhuseyinoglu & Akhuseyinoglu, 2016), (Tan et al., 2020), (Wang et al., 2017), (Kotak & Elovici, 2019), (Bai et al., 2021)
3	APKs files	(RIASAT et al., 2017), (Chen et al., 2016), (Sokolova et al., 2017), (Chukka, 2020)
4	Publicly available dataset	(Joshi et al., 2016), (Narayanan et al., 2018), (Saracino et al., 2016), (Anwar et al., 2016)
5	Biometric samples	(Arora & Bhatia, 2019), (Aneja et al., 2018), (Temper et al., 2015)
6	Event log data	(Shah & Shankarappa, 2018), (Shabtai et al., 2012)
7	IAT packet	(Muhammad et al., 2017), (Muhammad et al., 2019)
8	Mobile botnet data	(Eslahi et al., 2016)
9	URL session data	(Mora et al., 2014)
10	Smartphone sensor data	(Ho, 2014)
11	IoT node	(Kumar et al., 2020)
12	Email data	(Samarathunge et al., 2018)
13	Phone record data	(Petrov & Znati, 2018)
14	Apache spark	(Kyriazis, 2018)

RQ4: What are the evaluation metrics that are employed to evaluate the performance of machine learning algorithms for detecting and combating security threats and attacks in the BYOD environment?

Evaluation metrics are the performance measure used to evaluate the performance of the machine learning approaches implementation of the BYOD security threats and attacks. As can be seen in Table 8 of this study, 16 evaluation metrics were identified, which include ACC, PRE, REC, F-M, AUC, FPR, TPR, FAR, FRR, ACE, FA, TNR, FNR, ERR, RMSE, and Kappa statistics. Figure 11 demonstrates the percentage distributions of the research on the evaluation metrics based on the selected primary studies. As can be seen from Fig. 11, the studies focused more on the metrics such as ACC, PRE, REC, F-M, and AUC with a percentage of 31%, 17%, 16%, 13%, and 7% respectively. Out of the five metrics, ACC stands out as it has the highest percentage distribution. These metrics are gaining researchers' attention in the domain. It should also be noted that ACE, FA, TNR, FNR, ERR, RMSE, AND Kappa statistics have a very low percentage score of 1% each, which shows that those metrics have equal attention according to the selected studies.

Figure 12 also represents the bubble plot of evaluation metrics for evaluating the machine learning approaches implementation of the selected primary studies in the BYOD environment. In the plot, Y-axis represents the evaluation metrics and X-axis represents

the years. As demonstrated in Fig. 12, there is an increase in attention towards the ACC, PRE, REC, F-M; AUC beginning in the year 2016 as the trend keeps increasing with the increase in the number of publications. However, the research trend in the domain started losing attention in the year 2020, as the publication trend began to decline.

Table 8
Evaluation metrics

S/N	Performance metrics	References
1	ACC	(Samarathunge et al., 2018), (Petrov & Znati, 2018), (Shah & Shankarappa, 2018), (Muhammad et al., 2017), (Muhammad et al., 2019), (Eslahi et al., 2016), (RIASAT et al., 2017), (Arora & Bhatia, 2019), (Yerima et al., 2013), (Chen et al., 2016), (Lashkari et al., 2017), (Sokolova et al., 2017), (Gangwal & Conti, 2019), (Aneja et al., 2018), (Mora et al., 2014), (Shabtai et al., 2012), (Kumar et al., 2020), (Joshi et al., 2016), (Sahs & Khan, 2012), (Akhuseyinoglu & Akhuseyinoglu, 2016), (Tan et al., 2020), (Kotak & Elovici, 2019), (Bai et al., 2021), (Narayanan et al., 2018),(Li et al., 2018), (Narayanan et al., 2017), (Malhotra & Bajaj, 2016), (Anwar et al., 2016)
2	PRE	(Petrov & Znati, 2018), (Chizoba et al., 2020), (Yerima et al., 2013), (Lashkari et al., 2017), (Gangwal & Conti, 2019), (Sahs & Khan, 2012), (Chukka, 2020), (Kotak & Elovici, 2019), (Bai et al., 2021), (Narayanan et al., 2018),(Li et al., 2018), (Narayanan et al., 2017), (Zhu et al., 2017), (Pajouh et al., 2018) (Malhotra & Bajaj, 2016)
3	REC	(Petrov & Znati, 2018), (Chizoba et al., 2020), (Sokolova et al., 2017), (Gangwal & Conti, 2019), (Sahs & Khan, 2012), (Chukka, 2020), (Kotak & Elovici, 2019), (Bai et al., 2021), (Narayanan et al., 2018),(Li et al., 2018), (Narayanan et al., 2017), (Zhu et al., 2017), (Pajouh et al., 2018) (Malhotra & Bajaj, 2016)
4	F-M	(Chizoba et al., 2020), (Sokolova et al., 2017), (Gangwal & Conti, 2019), (Sahs & Khan, 2012), (Wang et al., 2017), (Chukka, 2020), (Kotak & Elovici, 2019), (Bai et al., 2021), (Narayanan et al., 2018),(Li et al., 2018), (Narayanan et al., 2017), (Zhu et al., 2017)
5	AUC	(Chizoba et al., 2020), (Yerima et al., 2013), (Sokolova et al., 2017), (Shabtai et al., 2012), (Sahs & Khan, 2012), (Das et al., 2015)
6	FPR	(Lashkari et al., 2017), (Shabtai et al., 2012), (Anwar et al., 2016)
7	TPR	(Shabtai et al., 2012), (Anwar et al., 2016)
8	FAR	(Arora & Bhatia, 2019), (Ho, 2014)
9	FRR	(Arora & Bhatia, 2019)
10	ACE	(Arora & Bhatia, 2019)
11	FA	(Eslahi et al., 2016)
12	TNR	(Yerima et al., 2013)
13	FNR	(Yerima et al., 2013)
14	ERR	(Yerima et al., 2013), (Temper et al., 2015)
15	RMSE	(Tout et al., 2019)
16	KAPPA STATISTICS	(Akhuseyinoglu & Akhuseyinoglu, 2016)

5. Discussions Of Findings From The Systematic Mapping Review

This SMR adopted both Peterson and Kitchenham guidelines for comprehensive results in providing an overview of the study on the machine learning approach for mitigating the security threats and attacks in the BYOD environment. In this study, a sum of 753 articles published from 2012 to 2021 was initially obtained. However, after the screening process, a total of 40 primary studies were finally selected for SMR. The SMR was conducted by answering four research questions under four classification schemes, which include threats and attacks, machine learning approaches, datasets used, and evaluation metrics.

In threats and attacks, the results of this SMR indicate that the malware attack is gaining considerable attention in the BYOD environment with 65% of the selected primary studies. Malware is a terminology used to identify an application that executes intentional malicious payloads on a target device or network (Aslan & Samet, 2020). Malware attack denotes malicious applications that can attack corporate applications as well as mobile devices. It consists of applications that contain code that breaches the mobile device or data security. There has been a rapid increase in mobile malware since 2011(Chang et al., 2014). This shows that malware is regarded as the most hazardous threat targeted at corporate data. In a BYOD environment, malware manipulates employees' mobile devices and uses them to steal confidential data (de las Cuevas et al., 2015) and direct banking transactions (Romer, 2014). When malware affects a device, it will lead to the exposure of confidential data, granting the attacker an opportunity to gain corporate identity. Besides the compromise on the employees' devices, mobile malware also attacks corporate applications by making them useless. It usually behaves as a potential corporate application that has been embedded with malicious code. There are several types of malware, namely, worms, viruses, ransomware, rootkit, and Trojan (Aslan & Samet, 2020). Malware can cause damage to an institution through activities such as theft of confidential information and impersonating the corporation (Olalere et al., 2015). Malware installed on BYOD devices can bypass conventional security mechanisms while communicating with external nodes. The work in (Olalere et al., 2015) observed that malware is disguised as normal applications that have hidden malicious code which infects devices when users visit compromised sites. In addition, the result of our mapping shows that those issues are relatively dominant in the research domain. On the other hand, the persistent threats, data leakage, data theft, DOS attack, intrusion attack, untrusted networks, and computational overheads are having a 2% distribution each in the selected studies, which shows less attention in the study field.

The study also indicated the machine learning algorithm implementation for security threats and attacks in the BYOD environment. However, it was found that SVM, DT, RF, ANN, AND KNN are the most considered machine learning algorithm for security threats and attacks in the BYOD environment. SVM, DT, RF, ANN, AND KNN are all supervised learning algorithms. Thus, the result shows that supervised learning algorithms are mostly used for detecting and combating security threats and attacks in the BYOD environment. The most effective machine learning algorithm is supervised learning, which uses trained data to help the system learn and allows the output to be categorized according to the input. In supervised learning, the model receives labelled training data and uses that information to learn how to classify incoming observations (C. I. Eke et al., 2019). The algorithm learns from the available training data and uses its application on real data. SVM is a supervised learning model that constructs a classification model by employing the learning theory of statistics. The classification task needs the splitting of the data into the training and test set. In SVM, a hyper-plane, which also refers to as a support vector is employed for separating two-class data points by using the training sets to reduce the space between them (Cristianini & Shawe-Taylor, 2000). A DT is an instance-focused induced learning algorithm. Instance classification is trained using decision trees, which may categorize instances based on the specific attribute occurrence of the value sets. One of a decision tree algorithm's flaws is the over-fitting problem. Utilizing several classifiers, such as the random forest, which divides the training set into different trees to be created and trained, then combining the final predictions over the tree, can, however, solve this issue (Eke et al., 2021). RF on the other hand is an ensemble learning model that uses sub-training sets to construct the decision tree algorithm. Consequently, the decision tree classifies every input vector in a forest and the most predicted model is chosen. RF addresses the overfitting issue and it attains better prediction over a single decision tree (Fernández-Delgado et al., 2014). A neural network is a learning model that has the same characteristic that performs the same nerve system function found in the human brain. An ANN is made up of three different layers; the input layer, the hidden layer, and the output layer. The input and the hidden layers are made up of various nodes, whereas the output layer contains just one node (Christopher Ifeanyi Eke et al., 2019). In the neural network, there is a connection between each unit and other units of the network, which possesses a summation function that integrates all the input values. K-NN is another supervised learning model that is used to solve both regression and classification tasks. The learning model relies on the similarity measure for data classification by employing majority neighbor vote. Thus, the allocation of data to the class is established by the highest nearest neighbor and the increment in the nearest neighbor enhances the classification accuracy. The analysis of the SMR showed that these learning algorithms attained a total of 68% of the selected primary studies, which shows an increase in the attention of those machine learning approach for security mitigation mechanisms in the field. Conversely, GDA, LDC, Ensemble, LR, BN, and RT are having a percentage distribution of less than 20% of the selected studies, which shows less research and a decrease in the consideration in the domain.

In datasets utilization for the machine learning approaches in detecting and combating the security threats and attacks in the BYOD environment, the systematic mapping result identified and classified 14 categories of datasets in the selected primary studies. The review findings show that the datasets for security threats and attacks based on machine learning implementation in the BYOD environment could be homogeneous data or heterogeneous data. The researchers have the option of collecting their own dataset or utilizing the publicly available datasets. However, many studies utilized the publicly available dataset. The review also indicated that the most used dataset from the selected studies is Malware and Benign samples dataset. The analysis of the data shows that there is an increase in the number of studies that utilized malware samples and benign apps, network traffics, APK files, publically available datasets, and biometric sample data as they obtained the highest percentage distribution of 69% of the selected studies. However, this shows that those datasets gained researchers' attention in the domain, more especially from the year 2016 to 2020 as the analysis showed the rise in the trend in those years. In contrast, there are fewer studies on the utilization of email data, phone record data, and Apache Spark data.

The study also provided the results of the evaluation metrics employed to evaluate the performance of the machine learning approaches in detecting and combating security threats and attacks in BYOD environments. The result of the analysis shows that ACC, REC, PRE, and F-M were mostly used in the selected primary studies. However, these measures might not be sufficient to accurately assess the performance of the classifier. This is due to the class imbalance that is mostly seen across different datasets in the chosen research. Due to its applicability in measuring the classification performance related to a specific class, AUC would be the ideal choice in this circumstance (Provost & Fawcett, 1997; Provost et al., 1998). Additionally, compared to F-Measure, AUC has great resilience to the skewness in datasets by applying TPR. Thus, the findings from the study analysis indicated that ACC, REC, PRE, and F-M were utilized in 77% of the selected primary studies, showing that those metrics are standard metrics that most researchers use for a reliable performance evaluation of machine learning approaches implementation for security threats and attacks in the BYOD environment. In contrast, less than 30% of the selected studies utilized AUC, FPR, TPR, FAR, FRR, ACE, FA, TNR, FNR, ERR, RMSE, and Kappa statistics as evaluation metrics in the domain, which shows that they are not commonly used for machine learning approaches evaluation of security threats and attacks in the BYOD environment.

6. Threats To Validity

This systematic mapping study process is not reliable just like other secondary study methods but is also subjected to some validity threats. Several risks are required to be considered to ensure the validity of this systematic mapping review. This research segment describes threats and their mitigation mechanisms, which include the search criteria, online databases, and the selection criteria (inclusion and exclusion) (O'donovan et al., 2015).

6.1 The search criteria

To search for academic databases, the key attention is focused on the definition of valid search strings. Thus, the formulation of the search string serves as a threat to the validity of this research [100]. To alleviate the threat, this study's search string was created by employing the PICO criteria [30], which is the standard and is mostly used in systematic mapping studies. Thus, it helped in retrieving the required articles in the search result and to alleviate the threat.

6.2 Digital databases

The chosen databases that consist of ACM Digital Library, IEEE Xplore, Springer, Google Scholar, and Science Direct stand as a threat to this study's validity because the related studies would have been absent in those databases. To alleviate these threats as noted by Kitchenham and Brereton (Kitchenham & Brereton, 2013) and asserted by (Dyba et al., 2007), the selection of ACM digital library, IEEE, and the addition of any other databases is adequate, which saves time and effort, instead of searching several databases (Dyba et al., 2007; Kitchenham & Brereton, 2013). Thus, in this study, the author selected five databases that included ACM digital library and IEEE that will alleviate the threat

6.3 Inclusion and exclusion criteria

In this study, the instruction and the requirement of selection criteria are demonstrated based on the scope of this study. The criteria were formulated based on the research team discussion. However, formulating a rule to identify the primary literature to

review signifies the presence of a threat that may be overlooked by the relevance search when different terms to that criterion are employed. Nevertheless, the search terms utilized in this study, consisting of bring your own device, machine learning approach, security threat, and attacks, are conventional, well-accepted, and defined terms, which should reduce the count of unrelated studies. Furthermore, as the research focused on the machine learning approach for BYOD security threats and attacks, there is less concern with including studies that are loosely related to the field. Thus, this form of threat can be avoided by ensuring that our search strategy produces relevant studies by refining the query iteratively.

7. Concluding Remarks And Future Direction.

The advancement in mobile computing has made researchers have an interest in one of the innovative, ubiquitous paradigms called bring your own device (BYOD) (Ballagas et al., 2004). BYOD paradigm enables employees to come with their own mobile devices and join the organization networks to enhance flexibility, productivity, and mobility on the side of the employees. Despite the numerous benefits that BYOD offers to both employees and organizations, security issues remain the major challenge in organizational settings. A considerable number of studies have been conducted and published on BYOD with great interest in security threats and mitigation mechanisms. However, the detailed review of the security solution mechanisms is not emphasized. Moreover, none of the existing reviews focused on the application of the machine learning approach, as a mitigation mechanism for the security threats and attacks in the BYOD environment. Besides, none of the existing studies demonstrated the ongoing trends in the domain, the datasets utilized for the implementation, and the evaluation metrics employed for the performance evaluations of the approaches in the existing solution.

This study presents a comprehensive systematic mapping review by highlighting the current trends in the implementation of machine learning approaches for mitigating security threats and attacks in the BYOD environment based on the existing primary studies published between 2012 and 2021. The SMR study was carried out by addressing four research questions. Out of the 753 studies that were initially retrieved, 40 primary studies were selected from 5 different academic databases after undergoing the selection criteria process. The SMR was conducted on the primary studies by exploring the existing security threats and attacks on the BYOD environment. Moreover, the machine learning approach implementation and the evaluation metrics employed to evaluate the performance of the machine learning approaches were demonstrated. In addition, the datasets that are used to implement the machine learning models were identified and reviewed.

The SMR result demonstrates the rise in the number of investigations into malware and unauthorized access in relation to the existing security threats and attacks in the BYOD environment. Moreover, concerning the machine learning approaches for mitigation mechanisms for security threats and attacks in the BYOD environment, the mapping result shows that supervised learning approaches such as SVM, DT, and RF gained many researchers' attention. However, the investigation also shows that there is a research gap in other machine learning approaches such as ensemble learning, clustering, LR, NB, BN, and DNN since they have not received much recognition in the domain. In addition, the study also indicates that there is a need for comprehensive publicly available datasets for the implementation of machine learning-based solutions since most researchers collect their own datasets, which can be a tedious task. Moreover, four standard performance evaluation metrics have been extensively used by researchers to evaluate models. However, researchers in the domain can also consider other metrics to evaluate models. Thus, the SMR has set the pace for creating new ground research in the machine learning-based approach for implementing the BYOD environment, which will offer invaluable insight into the study field. Therefore, researchers can employ this SMR to find a research gap in the research domain. In the future, the authors will consider the following research.

√ Malware classification for Apple, Android, and Unix platforms, and integrating it with mail services like Gmail, using malware publicly available datasets

√ Employing more enriched meta-data features, combining graph kernel with a set of kernel and application of semi-supervised machine learning approach for implementation.

√ Training the neural network with large datasets to separate normal and abnormal devices by employing ping, Skype, and iperf applications with TCP, UDP, and ICMP protocols.

√ Investigation of other classification approaches together with the implementation of the Genetic programming approach to address the problem of imbalance by using a modification of the cost-related misclassification

Declarations

Acknowledgments

The authors would like to extend our heartfelt gratitude and appreciation to the Ministry of Education (MOE), Malaysia for funding this project with the Fundamental Research Grant Scheme (FRGS) (FP056-2019A)

References

1. Abdelmaboud, A., Jawawi, D. N., Ghani, I., Elsafi, A., & Kitchenham, B. (2015). Quality of service approaches in cloud computing: A systematic mapping study. *Journal of Systems and Software, 101*, 159-179.
2. Akhuseyinoglu, N. B., & Akhuseyinoglu, K. (2016). *AntiWare: An automated Android malware detection tool based on machine learning approach and official market metadata.* Paper presented at the 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON).
3. Akin-Adetoro, A., & Kabanda, S. (2015). *Contextualizing BYOD in SMEs in developing countries.* Paper presented at the Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists.
4. Aneja, S., Aneja, N., & Islam, M. S. (2018). *Iot device fingerprint using deep learning.* Paper presented at the 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS).
5. Anwar, S., Zain, J. M., Inayat, Z., Haq, R. U., Karim, A., & Jabir, A. N. (2016). *A static approach towards mobile botnet detection.* Paper presented at the 2016 3rd International Conference on Electronic Design (ICED).
6. Arora, S., & Bhatia, M. S. (2019). Fingerprint Spoofing Detection to Improve Customer Security in Mobile Financial Applications Using Deep Learning. *Arabian Journal for Science and Engineering*, 1-17.
7. Aslan, Ö. A., & Samet, R. J. I. A. (2020). A comprehensive review on malware detection approaches. *8*, 6249-6271.
8. Bai, H., Liu, G., Liu, W., Quan, Y., Huang, S. J. S., & Networks, C. (2021). N-gram, semantic-based neural network for mobile malware network traffic detection. *2021*.
9. Ballagas, R., Rohs, M., Sheridan, J. G., & Borchers, J. (2004). *Byod: Bring your own device.* Paper presented at the Proceedings of the Workshop on Ubiquitous Display Environments, Ubicomp.
10. Barbier, J., Bradley, J., Macaulay, J., Medcalf, R., & Reberger, C. (2012). BYOD and Virtualization Top 10 Insights from Cisco IBSG Horizons Study. *Cisco IBSG Horizons Study*, 1-5.
11. Blum, A. L., & Langley, P. J. A. i. (1997). Selection of relevant features and examples in machine learning. *97*(1-2), 245-271.
12. Caldwell, C., Zeltmann, S., & Griffin, K. (2012). *BYOD (bring your own device).* Paper presented at the Competition forum.
13. Cavalcante, E., Pereira, J., Alves, M. P., Maia, P., Moura, R., Batista, T., . . . Pires, P. F. (2016). On the interplay of Internet of Things and Cloud Computing: A systematic mapping study. *Computer Communications, 89*, 17-33.
14. Chang, J. M., Ho, P.-C., & Chang, T.-C. (2014). Securing byod. *It Professional, 16*(5), 9-11.
15. Chen, S., Xue, M., Tang, Z., Xu, L., & Zhu, H. (2016). *Stormdroid: A streaminglized machine learning-based system for detecting android malware.* Paper presented at the Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security.
16. Chizoba, O. J., Kyari, B. A. J. G. J. o. E., & Advances, T. (2020). Ensemble classifiers for detection of advanced persistent threats. *2*(2), 001-010.
17. Chukka, H. V. (2020). *Detection of Malware using Machine Learning in Android Devices/Applications.* Dublin, National College of Ireland,

18. Conway, D., & White, J. (2012). *Machine learning for hackers*: " O'Reilly Media, Inc.".
19. Costa, G., Merlo, A., Verderame, L., & Armando, A. (2018). Automatic security verification of mobile app configurations. *Future Generation Computer Systems, 80*, 519-536.
20. Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*: Cambridge university press.
21. Das, S., Liu, Y., Zhang, W., Chandramohan, M. J. I. t. o. i. f., & security. (2015). Semantics-based online malware detection: Towards efficient real-time protection against malware. *11*(2), 289-302.
22. Dyba, T., Dingsoyr, T., & Hanssen, G. K. (2007). *Applying systematic reviews to diverse study types: An experience report.* Paper presented at the First international symposium on empirical software engineering and measurement (ESEM 2007).
23. Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2019). Sarcasm identification in textual data: systematic review, research challenges and open directions. *Artificial Intelligence Review*, 1-44.
24. Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2019). A Survey of User Profiling: State-of-the-Art, Challenges, and Solutions. *IEEE Access, 7*, 144907-144924. doi:10.1109/ACCESS.2019.2944243
25. Eke, C. I., Norman, A. A., & Shuib, L. J. P. o. (2021). Multi-feature fusion framework for sarcasm identification on twitter data: A machine learning based approach. *16*(6), e0252918.
26. Eslahi, M., Yousefi, M., Naseri, M. V., Yussof, Y., Tahir, N., & Hashim, H. (2016). *Cooperative network behaviour analysis model for mobile Botnet detection.* Paper presented at the 2016 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE).
27. Fatima, A., & Colomo-Palacios, R. (2018). Security aspects in healthcare information systems: A systematic mapping. *Procedia computer science, 138*, 12-19.
28. Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research, 15*(1), 3133-3181.
29. Fernandez, A., Black, J., Jones, M., Wilson, L., Salvador-Carulla, L., Astell-Burt, T., & Black, D. (2015). Flooding and mental health: a systematic mapping review. *PloS one, 10*(4), e0119929.
30. French, A. M., Guo, C., & Shim, J. P. (2014). Current status, issues, and future of bring your own device (BYOD). *Communications of the Association for Information Systems, 35*(1), 10.
31. Gangwal, A., & Conti, M. (2019). Cryptomining Cannot Change Its Spots: Detecting Covert Cryptomining Using Magnetic Side-Channel. *IEEE Transactions on Information Forensics and Security, 15*, 1630-1639.
32. Garba, A. B., Armarego, J., Murray, D., & Kenworthy, W. (2015). Review of the information security and privacy challenges in Bring Your Own Device (BYOD) environments. *Journal of Information privacy and security, 11*(1), 38-54.
33. Ghosh, A., Gajar, P. K., & Rai, S. (2013). Bring your own device (BYOD): Security risks and mitigating strategies. *Journal of Global Research in Computer Science, 4*(4), 62-70.
34. Ho, G. (2014). Tapdynamics: strengthening user authentication on mobile phones with keystroke dynamics. *Technicalreport, StanfordUniversity*.
35. Hussain, F., Hussain, R., Hassan, S. A., Hossain, E. J. I. C. S., & Tutorials. (2020). Machine learning in IoT security: Current solutions and future challenges. *22*(3), 1686-1721.
36. Jamal, F., Taufik, M., Abdullah, A. A., & Hanapi, Z. M. (2020). *A Systematic Review Of Bring Your Own Device (BYOD) Authentication Technique.* Paper presented at the Journal of Physics: Conference Series.
37. Joshi, P., Jindal, C., Chowkwale, M., Shethia, R., Shaikh, S. A., & Ved, D. (2016). *Protego: A passive intrusion detection system for Android smartphones.* Paper presented at the 2016 International Conference on Computing, Analytics and Security Trends (CAST).
38. Juárez, D. X. J., & Cedillo, P. (2017). *Security of mobile cloud computing: A systematic mapping study.* Paper presented at the 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM).
39. Kamal, M. F. A. H., Hamid, I. R. A., Abdullah, N., Abdullah, Z., Ahmad, M., & Shah, W. M. (2022). *Android Botnet Detection Based on Network Analysis Using Machine Learning Algorithm.* Paper presented at the International Conference on Soft Computing and Data Mining.

40. Keele, S. (2007). *Guidelines for performing systematic literature reviews in software engineering*. Retrieved from
41. Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., Linkman, S. J. I., & technology, s. (2009). Systematic literature reviews in software engineering–a systematic literature review. *51*(1), 7-15.
42. Kitchenham, B., & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. *Information and software technology, 55*(12), 2049-2075.
43. Kotak, J., & Elovici, Y. (2019). *Iot device identification using deep learning.* Paper presented at the Computational Intelligence in Security for Information Systems Conference.
44. Kumar, J. S., Sivasankar, G., & Nidhyananthan, S. S. (2020). An artificial intelligence approach for enhancing trust between social IoT devices in a network. In *Toward Social Internet of Things (SIoT): Enabling Technologies, Architectures and Applications* (pp. 183-196): Springer.
45. Kyriazis, D. (2018). *BYOS: Bring Your Own Security in Clouds and Service Oriented Infrastructures.* Paper presented at the 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA).
46. Lashkari, A. H., Kadir, A. F. A., Gonzalez, H., Mbah, K. F., & Ghorbani, A. A. (2017). *Towards a network-based framework for android malware detection and characterization.* Paper presented at the 2017 15th Annual conference on privacy, security and trust (PST).
47. Li, J., Sun, L., Yan, Q., Li, Z., Srisa-An, W., & Ye, H. J. I. T. o. I. I. (2018). Significant permission identification for machine-learning-based android malware detection. *14*(7), 3216-3225.
48. Malhotra, A., & Bajaj, K. J. C. t. o. I. (2016). A hybrid pattern based text mining approach for malware detection using DBScan. *4*(2), 141-149.
49. Micro, T. J. R. J. (2012). Enterprise readiness of consumer mobile platforms. *12*, 2012.
50. Miller, K. W., Voas, J., & Hurlburt, G. F. (2012). BYOD: Security and privacy considerations. *It Professional, 14*(5), 53-55.
51. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., . . . Ostrovski, G. J. n. (2015). Human-level control through deep reinforcement learning. *518*(7540), 529-533.
52. Mora, A. M., de las Cuevas, P., & Guervós, J. J. M. (2014). *Going a Step Beyond the Black and White Lists for URL Accesses in the Enterprise by Means of Categorical Classifiers.* Paper presented at the IJCCI (ECTA).
53. Muhammad, M. A., Ayesh, A., & Wagner, I. (2019). *Behavior-Based Outlier Detection for Network Access Control Systems.* Paper presented at the Proceedings of the 3rd International Conference on Future Networks and Distributed Systems.
54. Muhammad, M. A., Ayesh, A., & Zadeh, P. B. (2017). *Developing an intelligent filtering technique for bring your own device network access control.* Paper presented at the Proceedings of the International Conference on Future Networks and Distributed Systems.
55. Narayanan, A., Chandramohan, M., Chen, L., & Liu, Y. J. E. S. E. (2018). A multi-view context-aware approach to Android malware detection and malicious code localization. *23*(3), 1222-1274.
56. Narayanan, A., Chandramohan, M., Chen, L., & Liu, Y. J. I. T. o. E. T. i. C. I. (2017). Context-aware, adaptive, and scalable android malware detection through online learning. *1*(3), 157-175.
57. O'donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. J. J. o. B. D. (2015). Big data in manufacturing: a systematic mapping study. *2*(1), 1-22.
58. Oktavia, T., Tjong, Y., & Prabowo, H. (2016). *Security and privacy challenge in Bring Your Own Device environment: A Systematic Literature Review.* Paper presented at the 2016 International Conference on Information Management and Technology (ICIMTech).
59. Olalere, M., Abdullah, M. T., Mahmod, R., & Abdullah, A. J. S. O. (2015). A review of bring your own device on security issues. *5*(2), 2158244015580372.
60. Pajouh, H. H., Dehghantanha, A., Khayami, R., Choo, K.-K. R. J. J. o. C. V., & Techniques, H. (2018). Intelligent OS X malware threat detection with code inspection. *14*(3), 213-223.
61. Palanisamy, R., Norman, A. A., Kiah, M. L. M. J. C., & Security. (2020). Compliance with bring your own device security policies in organizations: A systematic literature review. *98*, 101998.

62. Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). *Systematic mapping studies in software engineering.* Paper presented at the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12.
63. Petrov, D., & Znati, T. (2018). *Context-Aware Deep Learning-Driven Framework for Mitigation of Security Risks in BYOD-Enabled Environments.* Paper presented at the 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC).
64. Provost, F. J., & Fawcett, T. (1997). *Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions.* Paper presented at the KDD.
65. Provost, F. J., Fawcett, T., & Kohavi, R. (1998). *The case against accuracy estimation for comparing induction algorithms.* Paper presented at the ICML.
66. RIASAT, R., SAKEENA, M., SADIQ, A. H., WANG, C., ZHANG, C.-y., WANG, Y.-j. J. D. T. o. C. S., & Engineering. (2017). Machine Learning Approach for Malware Detection by Using APKs. (cnsce).
67. Rivera, D., George, G., Peter, P., Muralidharan, S., & Khanum, S. (2013). Analysis of security controls for BYOD (bring your own device).
68. Romer, H. (2014). Best practices for BYOD security. *Computer Fraud & Security, 2014*(1), 13-15.
69. Sahs, J., & Khan, L. (2012). *A machine learning approach to android malware detection.* Paper presented at the 2012 European Intelligence and Security Informatics Conference.
70. Samarathunge, R., Perera, W., Ranasinghe, R., Kahaduwa, K., Senarathne, A., & Abeywardena, K. (2018). *Intelligent Enterprise Security Enhanced COPE (Intelligent ESECOPE).* Paper presented at the 2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS).
71. San Miguel, J. M., Kline, M. E., Hallman, R. A., Slayback, S. M., Rogers, A., & Chang, S. S. (2018). *Aggregated Machine Learning on Indicators of Compromise in Android Devices.* Paper presented at the Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security.
72. Saracino, A., Sgandurra, D., Dini, G., Martinelli, F. J. I. T. o. D., & Computing, S. (2016). Madam: Effective and efficient behavior-based android malware detection and prevention. *15*(1), 83-97.
73. Shabtai, A., Kanonov, U., Elovici, Y., Glezer, C., & Weiss, Y. (2012). "Andromaly": a behavioral malware detection framework for android devices. *Journal of Intelligent Information Systems, 38*(1), 161-190.
74. Shah, N., & Shankarappa, A. (2018). *Intelligent Risk management framework for BYOD.* Paper presented at the 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE).
75. Sokolova, K., Perez, C., & Lemercier, M. (2017). Android application classification and anomaly detection with graph-based permission patterns. *Decision Support Systems, 93*, 62-76.
76. Stallings, W. (2006). *Cryptography and network security, 4/E*: Pearson Education India.
77. Tahsien, S. M., Karimipour, H., Spachos, P. J. J. o. N., & Applications, C. (2020). Machine learning based solutions for security of Internet of Things (IoT): A survey. *161*, 102630.
78. Tan, X., Li, H., Wang, L., & Xu, Z. (2020). *End-Edge Coordinated Inference for Real-Time BYOD Malware Detection using Deep Learning.* Paper presented at the 2020 IEEE Wireless Communications and Networking Conference (WCNC).
79. Tan, Z., Jamdagni, A., He, X., Nanda, P., Liu, R. P. J. I. t. o. p., & systems, d. (2013). A system for denial-of-service attack detection based on multivariate correlation analysis. *25*(2), 447-456.
80. Temper, M., Tjoa, S., & Kaiser, M. (2015). *Touch to authenticate—Continuous biometric authentication on mobile devices.* Paper presented at the 2015 1st International Conference on Software Security and Assurance (ICSSA).
81. Tout, H., Kara, N., Talhi, C., & Mourad, A. (2019). Proactive machine learning-based solution for advanced manageability of multi-persona mobile computing. *Computers & Electrical Engineering, 80*, 106497.
82. Wang, S., Yan, Q., Chen, Z., Yang, B., Zhao, C., Conti, M. J. I. T. o. I. F., & Security. (2017). Detecting android malware leveraging text semantics of network flows. *13*(5), 1096-1109.
83. Wang, Y., Wei, J., & Vangury, K. (2014). *Bring your own device security issues and challenges.* Paper presented at the 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC).
84. Yang, L., Chen, Y., Li, X.-Y., Xiao, C., Li, M., & Liu, Y. (2014). *Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices.* Paper presented at the Proceedings of the 20th annual international conference on Mobile computing

and networking.

85. Yerima, S. Y., Sezer, S., McWilliams, G., & Muttik, I. (2013). *A new android malware detection approach using bayesian classification.* Paper presented at the 2013 IEEE 27th international conference on advanced information networking and applications (AINA).
86. Zhu, D., Jin, H., Yang, Y., Wu, D., & Chen, W. (2017). *DeepFlow: Deep learning-based malware detection by mining Android application for abnormal usage of sensitive data.* Paper presented at the 2017 IEEE symposium on computers and communications (ISCC).

Figures

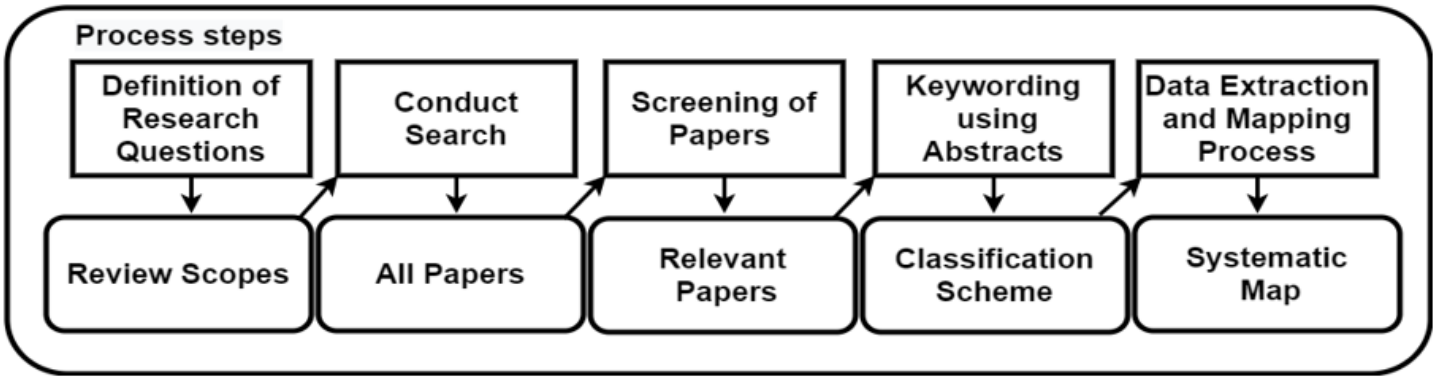

Figure 1

Systematic mapping process

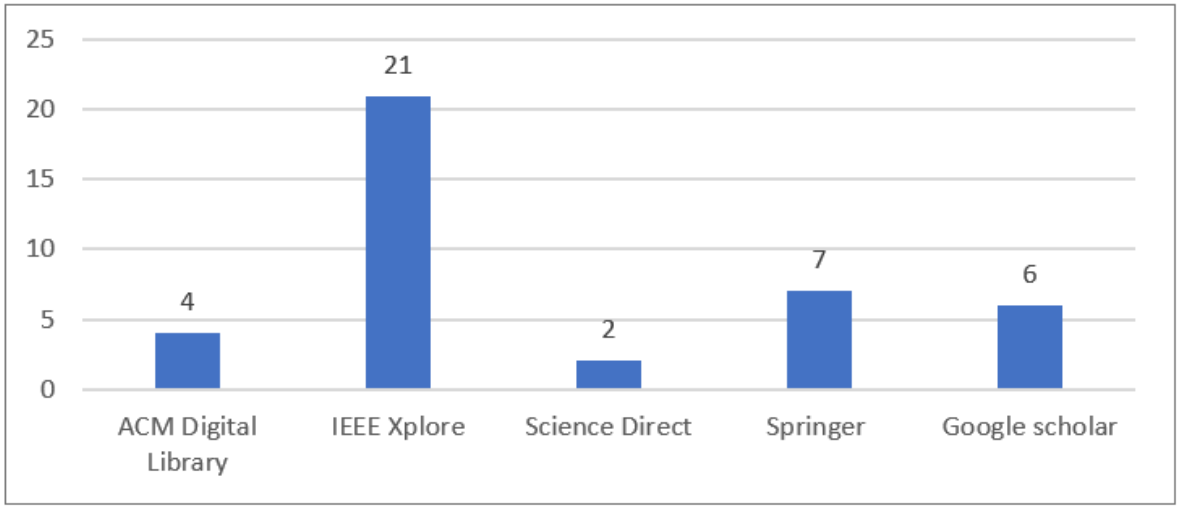

Figure 2

Selected articles distributions

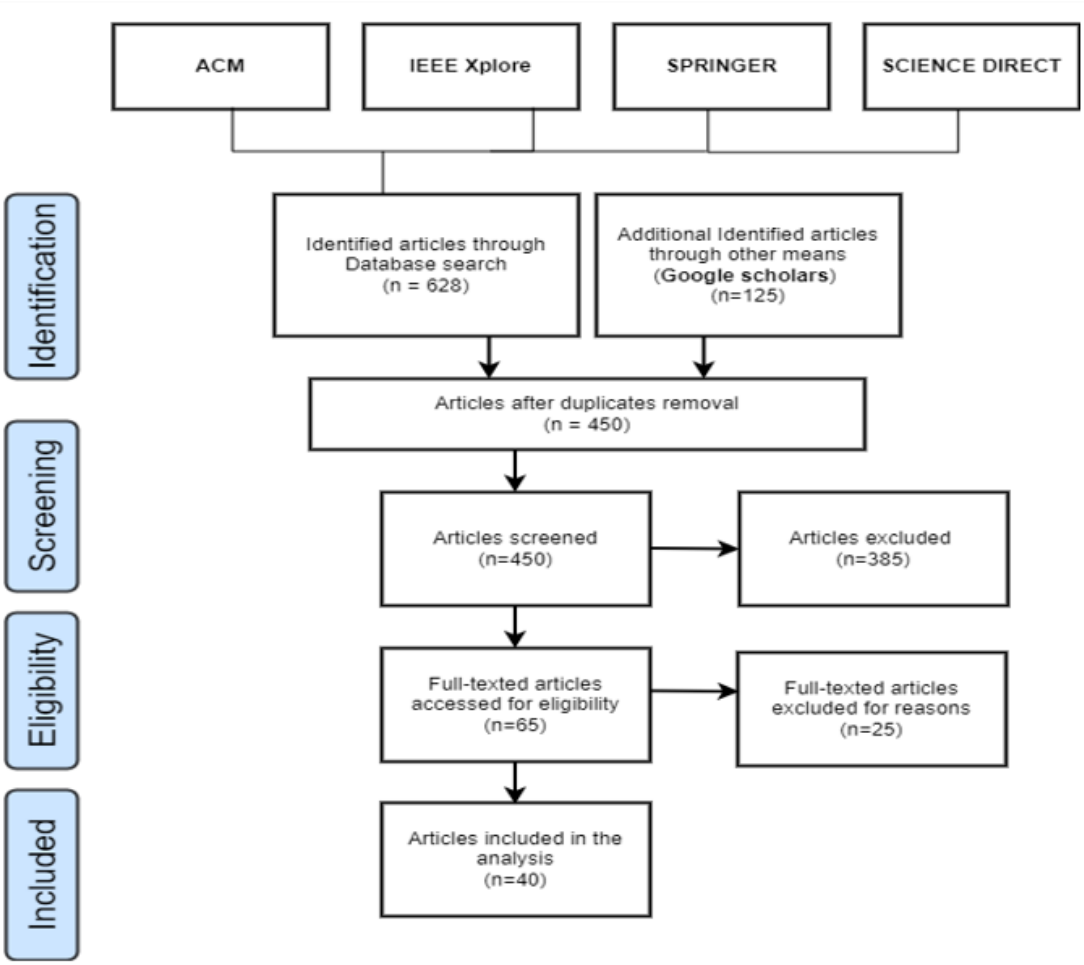

Figure 3

PRISMA process flow

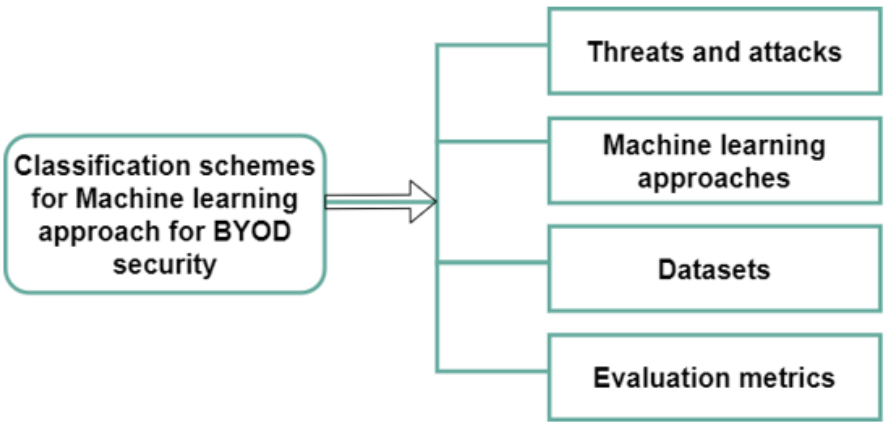

Figure 4

Classification schemes for machine learning approaches for BYOD security threats and attacks.

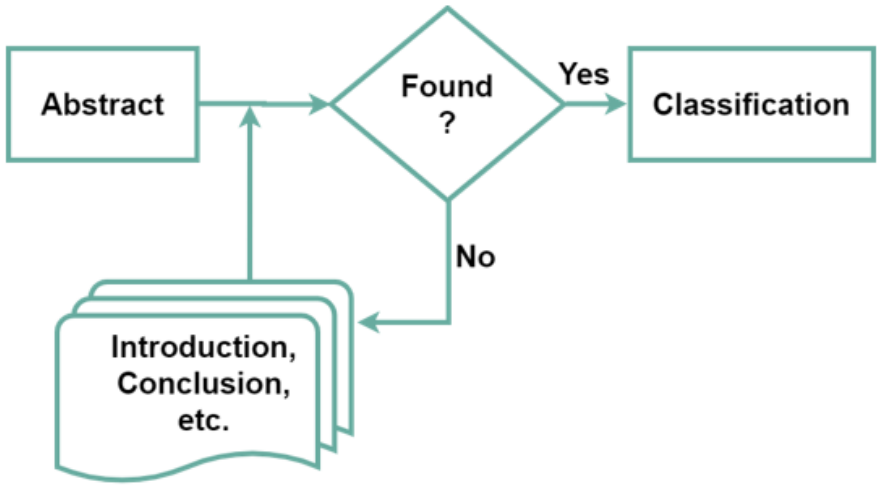

Figure 5

Classification process

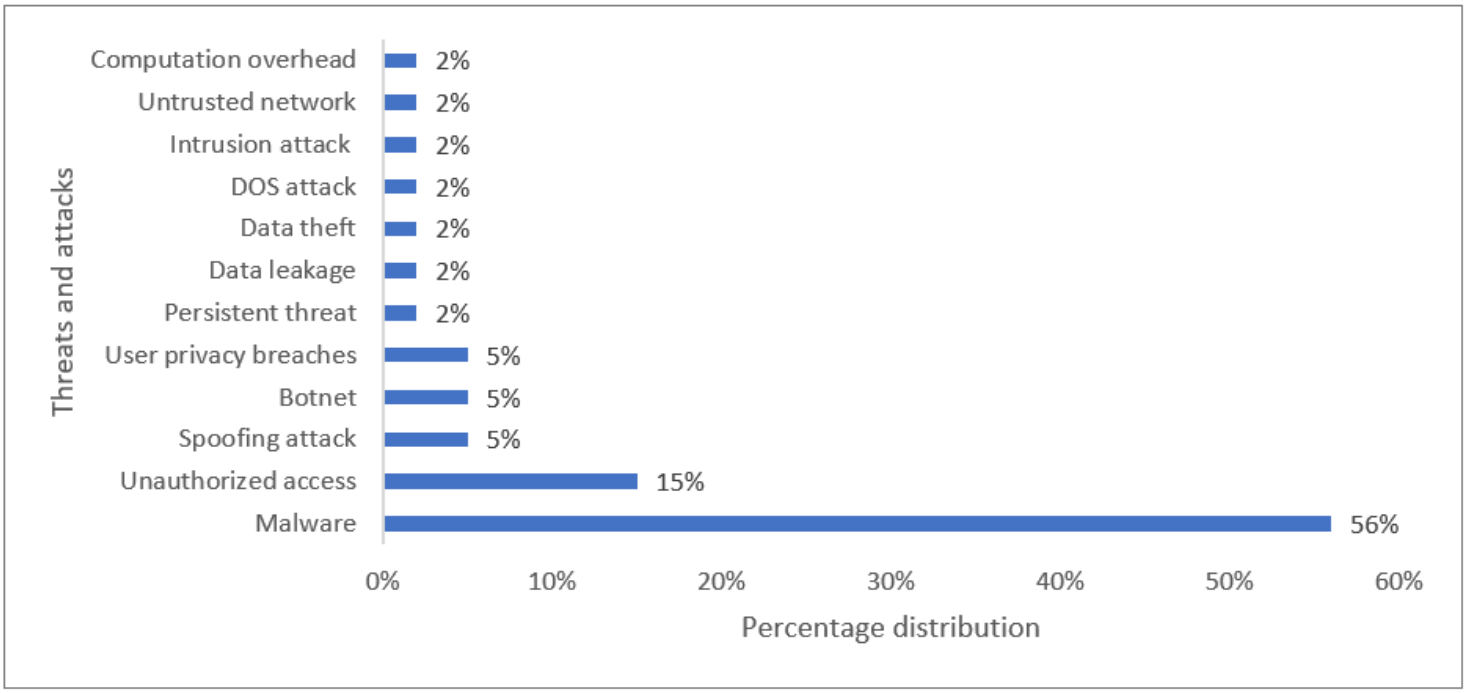

Figure 6

Threats and attacks percentage distributions using Machine Learning

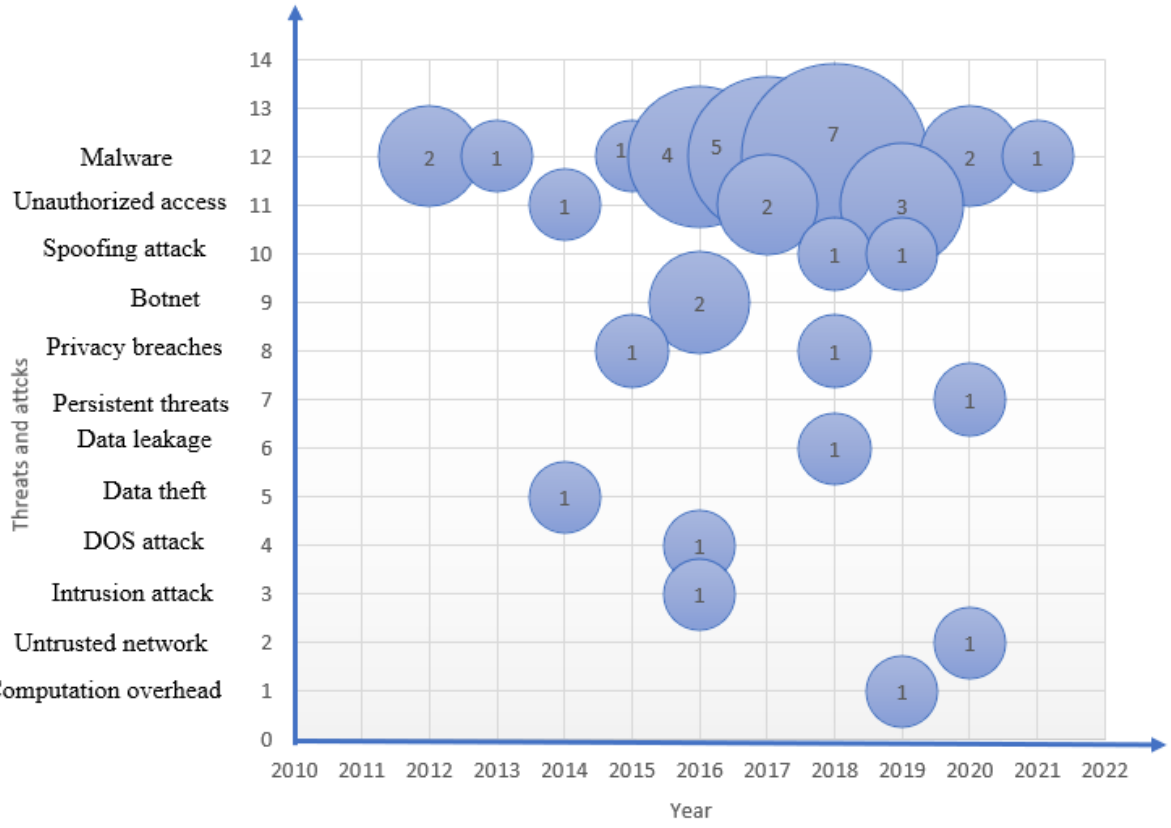

Figure 7

Bubble plot of the security threats and attacks using machine learning

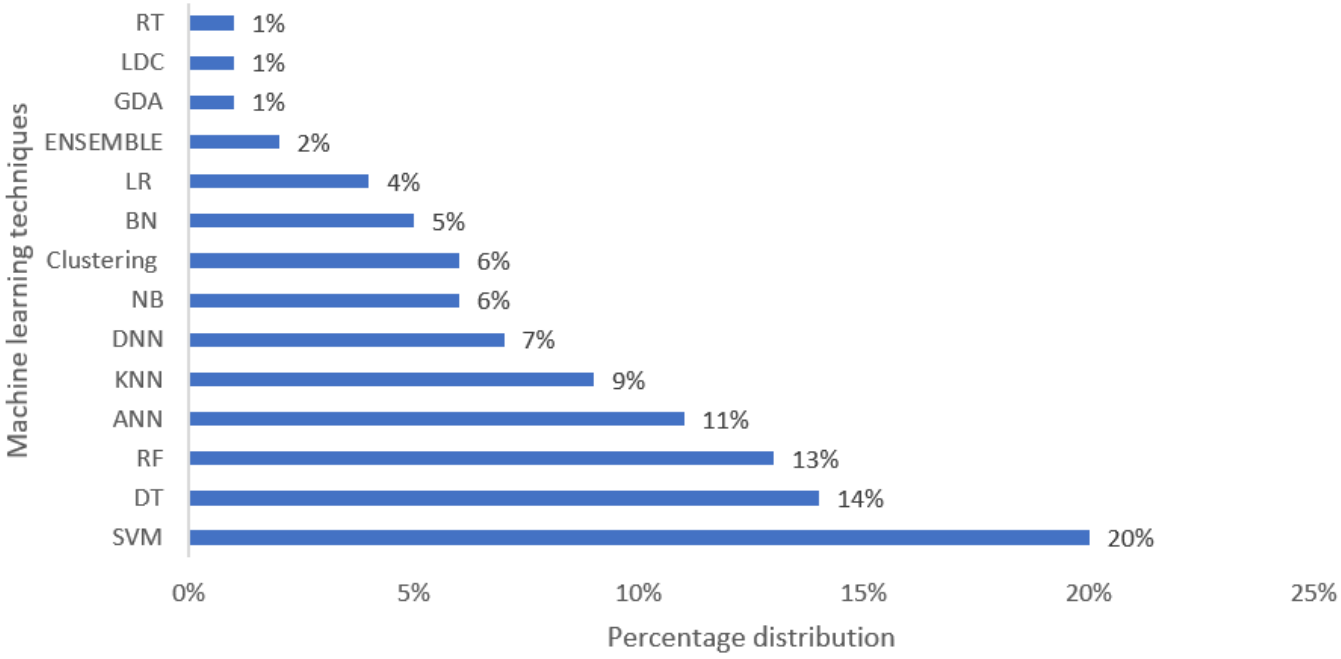

Figure 8

Machine learning approaches Percentage distribution for security mitigation

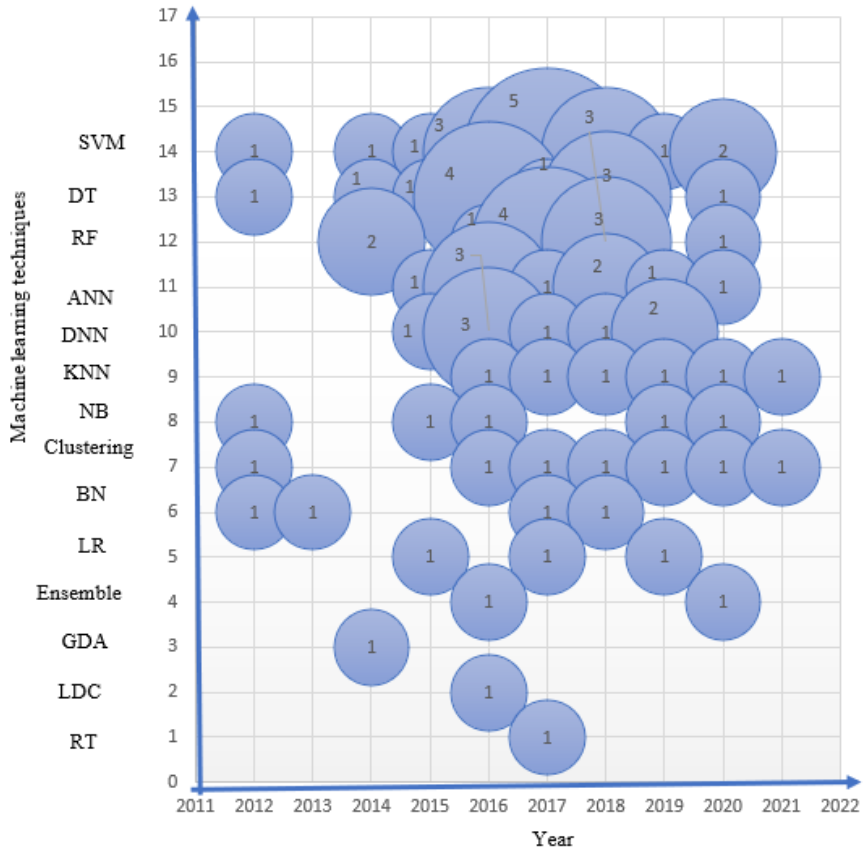

Figure 9

The bubble plot of the machine learning approaches implementation for security mitigation

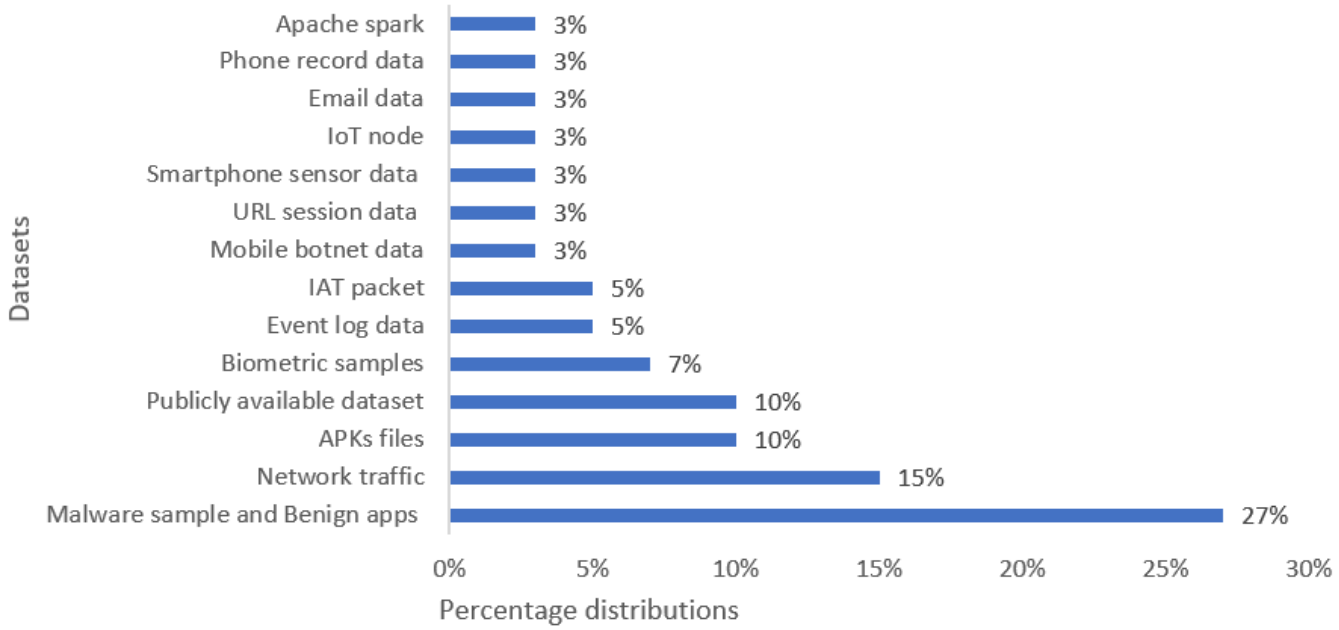

Figure 10

Datasets percentage distributions

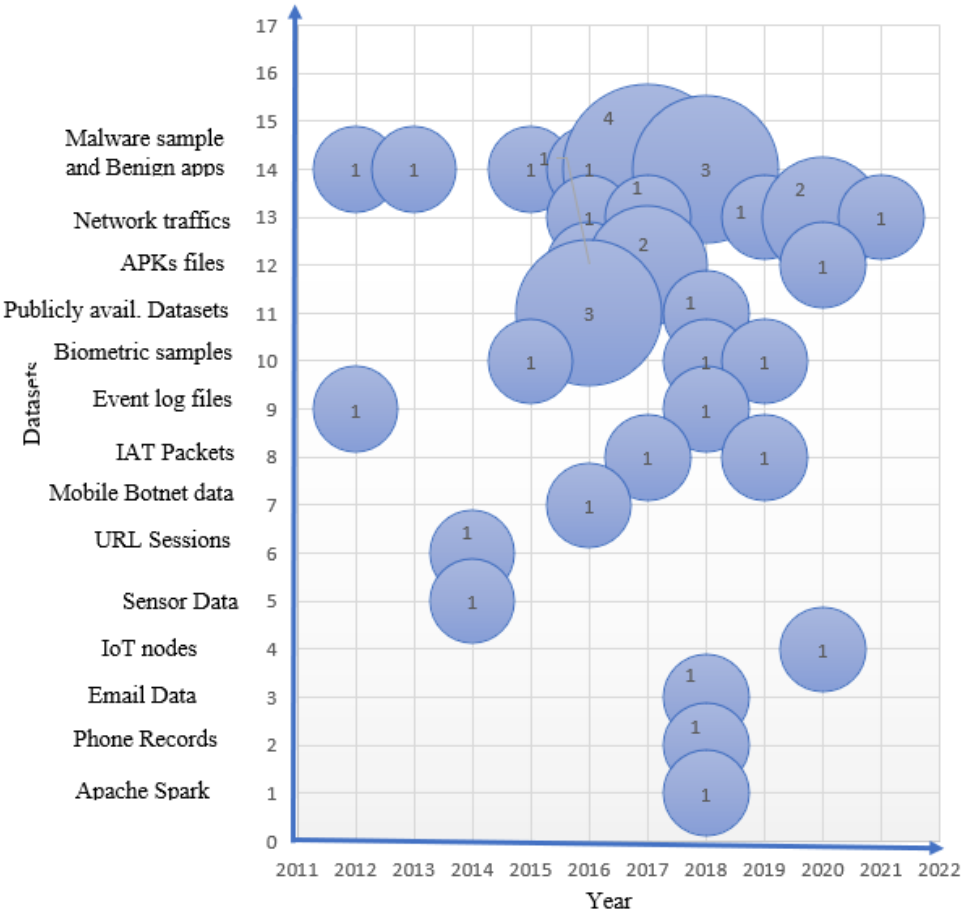

Figure 11

Bubble plot of the used datasets

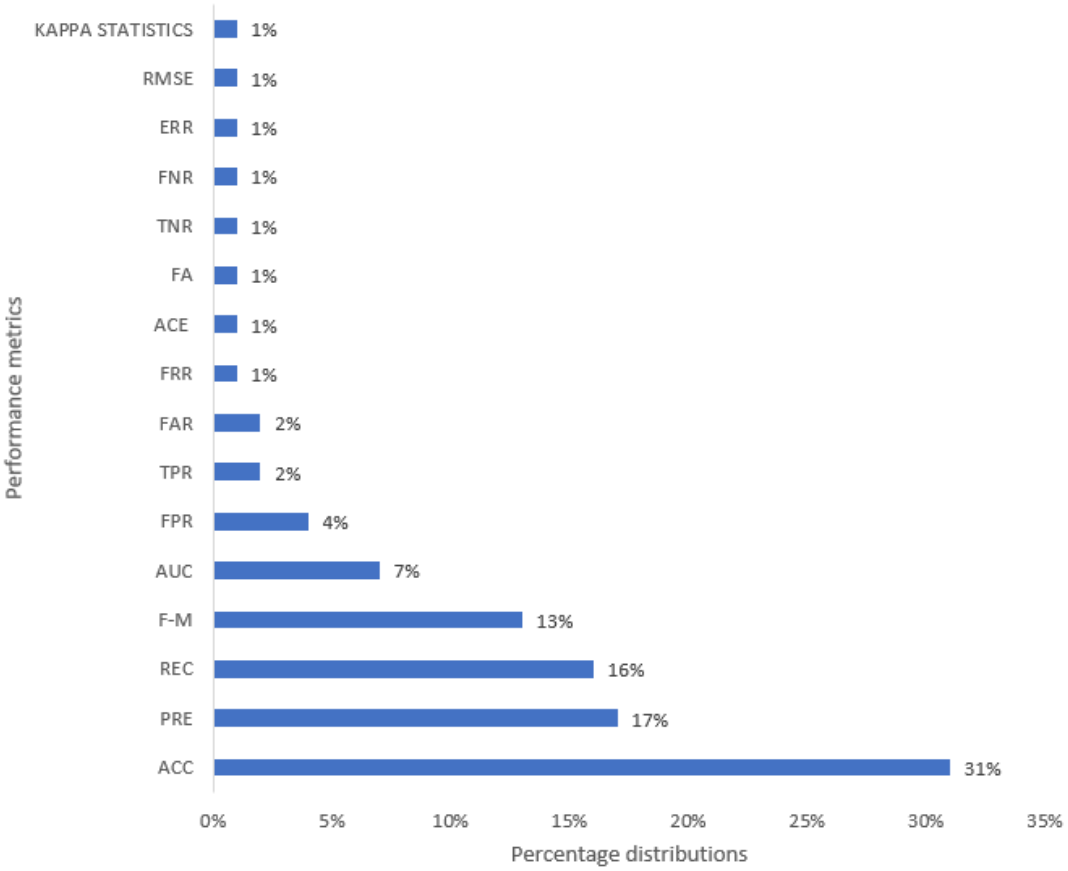

Figure 12

Evaluation metrics percentage distributions

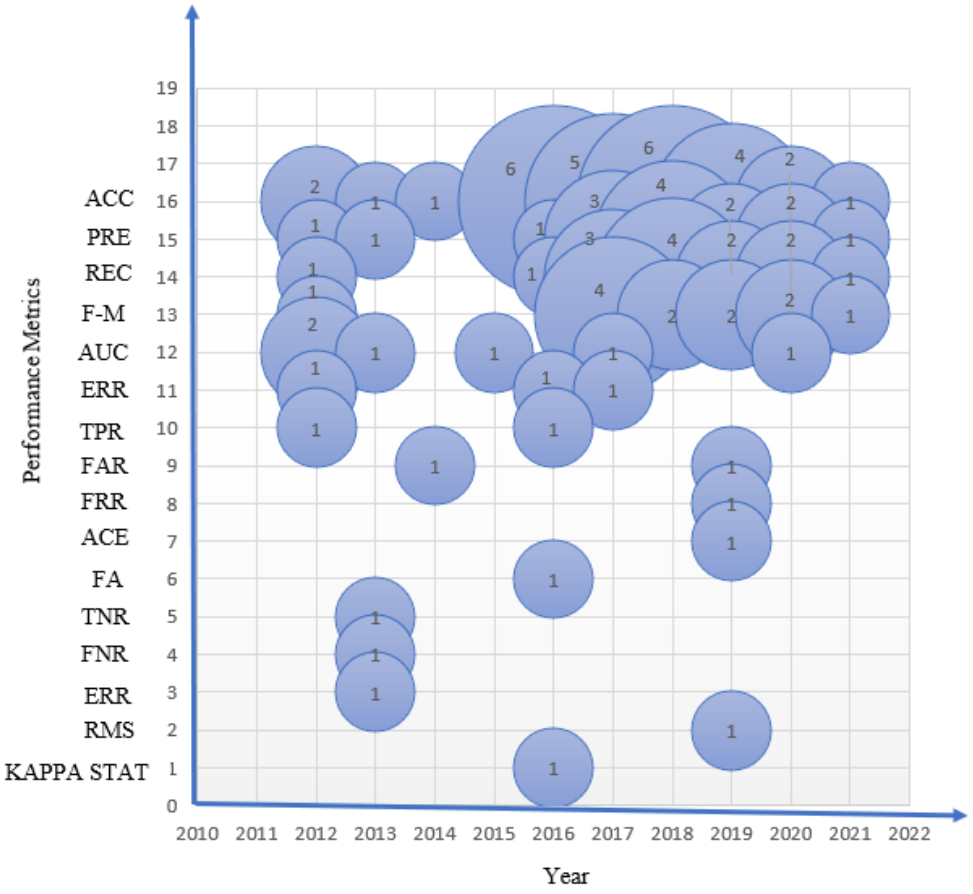

Figure 13

Evaluation metrics bubble plot