Hindawi
Security and Communication Networks
Volume 2022, Article ID 9663052, 23 pages
https://doi.org/10.1155/2022/9663052

Review Article

A Systematic Review on Hybrid Intrusion Detection System

Elijah M. Maseno,[1,2] **Zenghui Wang**,[2] **and Hongyan Xing**[3]

[1]*School of Information Technology for Defence Systems, Defence Forces Technical College, Nairobi 19120-00501, Kenya*
[2]*College of Science, Engineering and Technology, University of South Africa, Pretoria 1709, South Africa*
[3]*Collaborative Innovation Center for Meteorological Disaster Prediction and Evaluation, Nanjing University of Information Science and Technology, Nanjing 210044, China*

Correspondence should be addressed to Zenghui Wang; wangzengh@gmail.com

Received 23 December 2021; Revised 5 March 2022; Accepted 29 March 2022; Published 10 May 2022

Academic Editor: Leandros Maglaras

As computer networks keep growing at a high rate, achieving confidentiality, integrity, and availability of the information system is essential. Intrusion detection systems (IDSs) have been widely used to monitor and secure networks. The two major limitations facing existing intrusion detection systems are high rates of false-positive alerts and low detection rates on zero-day attacks. To overcome these problems, we need intrusion detection techniques that can learn and effectively detect intrusions. Hybrid methods based on machine learning techniques have been proposed by different researchers. These methods take advantage of the single detection methods and leverage their weakness. Therefore, this paper reviews 111 related studies in the period between 2012 and 2022 focusing on hybrid detection systems. The review points out the existing gaps in the development of hybrid intrusion detection systems and the need for further research in this area.

1. Introduction

The Internet has thrived, hence an increase in information sharing, making network security a problem of concern. Attackers around the globe have their eyes on computer systems with the motive of deploying attacks. The security of an electronic device is breached when a successful attack occurs. Intrusion is defined as "any set of actions that attempt to compromise the integrity, confidentiality, or availability of a resource" [1]. The *Integrity* aspect of a given infrastructure serves to ensure information remains unaltered by unauthorized users. *Availability* incorporates all aspects of the infrastructure that makes information readily available to users in the system. *Confidentiality* implies that the information in a given system is protected from unauthorized access and viewing by external parties. Therefore, a computer network is considered to be fully secured when the core objectives of these three attributes are sufficiently met. To help achieve these objectives, intrusion detection systems have been developed with the primary intent of monitoring incoming traffic in computer networks for any potential malicious intrusions.

An intrusion detection system (IDS) scans information system resources and reports any malicious activities in the system. More advanced IDSs have the capability of acting against the attacks. The action taken by this advanced IDS is to block the malicious users or activities from accessing the computer resources. We have two major categories of intrusion detection systems, which include misuse based and anomaly based. Misuse-based IDSs are developed to flag known attacks using patterns of the known attacks [2]. Misuse detection systems use patterns of well-known attacks or weak spots of the system to match and identify known intrusions. The positive side of misuse IDS is the ability to detect known attacks with great precision. The major challenge facing this type of IDS is their inability to flag new forms of attacks [3]. Misuse intrusion detection systems stand out because of their ability to flag many or all known attack patterns. The main problem facing misuse-based systems is the inability to flag emerging attacks or zero-day

attacks. In general, they have a high rate of detection and low rate of false alarms compared to anomaly-based systems. The anomaly-based technique stores the normal behavior of a user in a database and compares it with the current behavior of the user [4]. If there is a substantial difference, then there is something wrong or abnormal. The major advantage of anomaly detection is that it does not require information of known attacks, and thus they can detect new forms of attacks. It has a high rate of false alarm compared to misuse-based IDS.

Hybrid intelligent systems have been developed to solve the challenges of the existing intrusion detection systems, such as high rate of false-positive alerts and low detection rate of novel attacks. Hybrid is a technique that combines misuse-based and anomaly-based techniques [5]. The hybrid technique resolves the disadvantages of the two legacy IDSs. Research shows that hybrid detection systems have better performance compared to single IDS.

Despite their proven performance, hybrid intrusion detection systems remain largely unexplored as seen from the few number of existing systematic literature reviews on the topic. This work, therefore, attempts to perform a comprehensive systematic literature review on hybrid intrusion detection systems between 2012 and 2022 with the objective of pointing out existing gaps in the development of these systems.

This study is arranged as follows. Section 2 introduces and discusses IDS. Section 3 provides a discussion on hybrid detection techniques. Section 4 discusses the methodology adopted in this paper. Section 5 discusses the findings. Section 6 points out the existing gaps in the reviewed literature and insights for future research. Table 1 summarizes all hybrid intrusion detection systems between the periods of 2012 and 2022. Finally, Table 2 lists all abbreviations in this study.

2. Intrusion Detection Systems

Denning introduced the technique of detecting intrusion, and since then researchers have worked hard to automatically detect intrusions in network systems [6]. Intrusion detection systems have been defined as the technique of using artificial intelligence, machine learning, and database systems to uncover malicious patterns in large datasets [2]. IDS can be broadly classified into two major categories, anomaly-based IDS and misuse-based IDS. Recently, other methods have emerged through the integration of anomaly and misuse intrusion IDSs to yield more categorizes.

2.1. Anomaly-Based Intrusion Detection Systems. Anomaly intrusion detection systems profile the normal behavior of a system. They monitor the normal operations of the system, and if they detect an anomaly, a flag is raised. Instead of keeping all patterns of well-known malicious dataset and updating as new patterns emerge, anomaly detection systems outline "normal" operations of a system and flag anything that deviates from the outline [2]. According to [7], anomaly IDS contains three stages: parameterization, training stage, and detection stage. In the parametrization stage, the data are formatted to capture the normal behavior of the device. After parameterization, the model is trained to represent the normal behavior. The detection stage is where the model detects and flags any deviation from the normal behavior based on the parameterized data [7].

Different intrusion detection mechanisms have been used in the development of the anomaly IDS. Mishra and Yadav [8] outlined the following techniques: data mining techniques, machine learning-based techniques, and statistical approaches. In these techniques, some researchers have used single algorithms while others have opted to integrate algorithms to improve the performance of the IDS [8].

Atefi et al. [9] developed anomaly detection based on profile signature using genetic algorithm and support vector machine algorithms. SVM outperformed GA in terms of precision rate. The researchers combined the two algorithms to form a hybrid IDS. The evaluation of the hybrid IDS produced better performance compared to the single algorithms.

Khoei et al. [10] investigated the application of three types of ensemble learning techniques for anomaly IDS. The three techniques applied were bagging, boosting, and stacking. The performance of the three techniques was compared with that of decision tree (DT), Naïve Bayes (NB), and K-nearest neighbor (KNN). The results showed that stacking-based ensemble learning techniques outperformed the traditional learning techniques in terms of detection rate, false alarm rate, miss detection rate, and accuracy rate.

Rakshe and Gonjari [11] developed an intrusion detection model based on SVM and random forest algorithms. The two algorithms were used for classification purposes. The models were evaluated using NSL-KDD. The models recorded detection accuracy of more than 95%. The performance of the two models was compared, and the random forest algorithm performed better than SVM in the classification of traffic.

Kumar et al. [12] developed an anomaly intrusion detection system based on four algorithms, namely, Naïve Bayes, ID3, MLP, and ensemble learning. The models were evaluated using CICIDS2017 dataset. The ensemble model was developed by combining NB, ID3, and MLP. The metrics used in the evaluation of the models were precision, recall, accuracy, and F1 score. ID3 (decision tree) performed better compared to the other models.

Once anomaly intrusion detection systems have been developed, they do not need regular updates unless a major user or system change has been done. Anomaly IDS can flag new forms of attacks, unlike the misuse IDS. Due to the above-mentioned characteristic of anomaly intrusion detection systems, they are considered to be more effective compared to their counterpart misuse intrusion detection system whose performance highly depends on stored patterns that require regular updates.

Profile creation is the main issue in anomaly intrusion detection because there is no fixed normal action or behavior of the user, and different users use computer systems

Table 1: A summary of all hybrid intrusion detection systems between the periods of 2012 and 2022.

ID	Reference	Dataset	Strength/weakness
C1	[34]	Flow-based dataset	The model demonstrated high-speed intrusion detection in large network infrastructure through data reduction and processing time.
C2	[35]	KDD′99 dataset	According to the evaluation results, when choosing the optimum parameter like feature size reduction, the overall performance of the intrusion detection system improves.
C3	[36]	KDD dataset	The proposed algorithms detect intrusion simultaneously and their output is combined using the rule-based method. The model is tested using the KDD dataset and records an outstanding performance.
C4	[37]	Kyoto 2006 + dataset	The researchers proposed the use of K-medoids instead of K-means in data clustering. K-medoids performed better compared to K-means clustering. Naïve Bayes was used for classification. The researchers observed that with the use of K-medoids clustering, processing time increases as the data grow. How to manage time forms a future research gap.
C5	[38]	Kyoto 2006 + datasets	This model was developed to tackle the problems of a previous work in which the researchers combined K-medoids clustering and Naïve Bayes classification. To further improve the performance of intrusion detection on this model, the researchers combined support vector machine classification with K-medoids clustering. The model recorded better performance. Still, time management is an issue in this model as the previous one.
C6	[39]	KDD Cup99 dataset	The model uses a double classifier based on AdaBoost with J48 base learner and Bayesian network classifier. The model performed better than J48 and Bayesian cascaded classifier.
C7	[40]	KDD99 dataset	The model performed better compared to pure SVM in terms of detection rate, training time, and false negative and false positive. In addition, it performed better than pure CSOACN in terms of less training time with comparable detection rate and false alarm rates. In the future, further performance analysis can be conducted using other algorithms.
C8	[41]	KDD99 dataset	In this model, GA and SOFM were used for feature extraction on the dataset. The goal was feature reduction on the dataset to be used in training SVM. In this study, SVM was deployed as a classification algorithm. The model performed better compared to SVM.
C9	[42]	NSL-KDD	At the first stage of the model, PCA is used for feature reduction. The second stage of the model deploys genetic algorithms for the anomaly detection process by labeling the dataset as either normal or anomaly. The final stage uses different types of classifiers for confirmation if the datasets are labeled properly and give detailed information of the attacks. The model was able to demonstrate the importance of combining different machine learning algorithms for intrusion detection.
C10	[43]	CAIDA UCSD 2007 dataset	Proposed a model to detect DDoS attacks, and the model combines GA and multilayer perceptron (MLP) of ANN. GA is used in feature selection while MLP is used for classification.
C11	[44]	KDD Cup 1999	The model recorded a true-positive value of 0.973 and false-positive value of 0.017, which was an outstanding performance. The model needs further evaluation using current datasets.
C12	[45]	KDDcup'99	The researchers report that the model achieved a high detection rate and low false-positive rate, but at the same time, they acknowledge the need for further evaluation of the model using different datasets in the future.
C13	[46]	KDD CUP 99	Particle swarm optimization has been used in this model to select optimal parameters for multiple criteria linear programming. The model recorded a better performance in terms of accuracy and running time compared with the MCLP model. The model should be further investigated using KDD CUP 99 for its capabilities of detecting different attacks simultaneously.
C14	[47]	KDDCUP′99 dataset	The research demonstrated the importance of feature selection in intrusion detection. With reduced features, the model was able to improve accuracy rate and detection rate; at the same time, false alarm rate decreased. Evaluation of the model using only one dataset is not enough, and the model needs to be evaluated using another dataset in the future.
C15	[48]	KDD CUP99 data	The system provides advantages such as feature reduction on the training dataset which improves the performance of intrusion detection systems.

Table 1: Continued.

ID	Reference	Dataset	Strength/weakness
C16	[49]	NSL-KDD	Proposed K-means and Naïve Bayes classifier for hybrid intrusion detection. K-means was used for data clustering to reduce dataset features, while Naïve Bayes classifier was deployed for classification of the features as normal or attack. The model recorded a better performance in the detection of probe, R2L, and U2R attacks.
C17	[50]	NSL-KDD dataset	In this model, AGAAR is used for feature learning and reduction. The model uses GPLS for the classification of the dataset as normal or attack. The researchers used only a presentation of the dataset to train the model.
C18	[51]	NSL-KDD	The researchers demonstrated that combining classifier algorithms using the sum rule approach has the potential of providing good results compared to single classifiers. The model outperformed a single classifier.
C19	[52]	KDD CUP 99 dataset	In this research, the K-means algorithm was used for classification while J48 was used for feature selection. SOM increased the accuracy rate of the system. The system registered high computation and longer processing time which affected its performance.
C20	[53]	KDDCUP 99 dataset	The dataset will be fed simultaneously into the parallel algorithms; if both algorithms define the data as normal, the data will be classified as normal. If both classifiers classify the data as an attack, ACO will further classify the data into a class of intrusion. The model outperforms individual SVM and ACO algorithms in intrusion detection.
C21	[54]	ISCX 2012	The model combined K-means clustering with discretization technique and Naïve Bayes classifier to create a hybrid intrusion detection system. KMC-D is used for clustering and NBC is used for classification. The model significantly reduced the false alarm rate.
C22	[55]	NSL KDD	The important characteristic of active learning SVM is the ability to develop an intrusion detection system with small samples of datasets, hence reducing the training time and increasing the efficiency of the model.
C23	[56]	KDD′99	To reduce the number of features, the model uses the spatial correlation-based dimension reduction method. The new feature set is used to train the SVM classifier for intrusion detection. The model achieves high performance in training the classifier algorithm.
C24	[57]	KDD′99 dataset	The mode consists of two levels. In level one, K-means is used for data dimensionality reduction, and in level two, RF is used for classification. The model was evaluated using an outdated dataset.
C25	[58]	NSL-KDD dataset	Proposed combination of PCA and LSTM-RNN. PCA is deployed for feature reduction; on the other hand, LSTM-RNN is used for classification. The proposed model performs better compared to a single algorithm.
C26	[59]	NSL-KDD dataset	Proposed K-means for clustering the dataset and SMO for feature classification in the second stage. The model outperforms individual algorithm K-mean clustering and sequential minimal optimization (SMO). In the future, the model can be evaluated using other datasets.
C27	[60]	NSL_KDD	The model has two stages of classification using SOM and BPNN. The model uses SOM in the first stage for classification. The dataset flagged as the attack in the first stage is further classified in the second stage using BPNN into different forms of attacks. The model can be verified using other types of datasets in the future.
C28	[61]	Wormhole dataset	The combination of K-means and decision tree has a high detection rate compared to single algorithms of K-means and decision tree. But the hybrid significantly reduces false positives suffered by the two single algorithms. In the future, research can be done on how to improve the detection rate of the hybrid.
C29	[62]	NSL-KDD dataset	The model deployed two feature selection algorithms in a cascaded manner. The model outperformed the RNN-based deep neural network in terms of accuracy. In the future, the model can be evaluated using different datasets.
C30	[63]	NSL-KDD dataset	Plant growth optimization in this model is used for feature reduction and selection. SVM is used for classification. In the future, further investigation of the mode can be done using a different dataset.
C31	[64]	Simulated attacks	The model implemented GMM, OCSVM, isolation forest, and SOM in parallel to improve the classification. In addition to this, they added a decision module to provide the final classification. The model reported low CPU and RAM usage with high accuracy.
C32	[65]	NSL-KDD	K-means clustering with random forest classifiers outperformed the Gaussian mixture clustering with random forest classifiers.

Table 1: Continued.

ID	Reference	Dataset	Strength/weakness
C33	[66]	DARPA-KDD99	In this model, fuzzy rules are generated and then inputted as a particle in PSO. In the future, more compact and intelligent fuzzy logic can be generated to enhance the detection of more attacks.
C34	[67]	KDD CUP	The model proposed the optimization of ANN using MOA-PSO. The model performed better compared to other models. The model was evaluated using an old dataset.
C35	[68]	NSL-KDD	The model recorded high accuracy and low FAR, but the model was tested using only one dataset. The model should be tested using other datasets for verification of its performance.
C36	[69]	NSL-KDD	The model recorded high detection rates of DoS, R2L, and probe attacks. According to the researchers, the model performed poorly in the detection of the user to root (U2R) attacks.
C37	[70]	NSL-KDD	This was the first model to combine rough set theory and random forest for intrusion detection. The model outperformed other models in terms of accuracy. The model was tested using only one dataset.
C38	[71]	NSL-KDD and UNSW-NB15 datasets	The model outperformed other models in detection rate and false-positive rates. The researchers proposed an investigation of the model using different datasets.
C39	[72]	A novel dataset	Proposed a hybrid detection model based on K-means clustering and support vector machine (SVM) classification. The model was evaluated using a novel dataset retrieved from a wireless network packet traffic flow. The model recorded a low false-positive rate with an improved detection rate.
C40	[73]	KDD′99 dataset	Proposed optimization of FCM using hybrid rice optimization algorithm. The model was evaluated using the KDD99 dataset. The model recorded better clustering performance. In the future, the model can be evaluated using modern datasets.
C41	[74]	KDD 99	K-means and K-nearest neighbors were used to reduce the time complexity of the system with great accuracy.
C42	[75]	KDD 99	The model training time of K-means and random tree algorithm-based intrusion detection system is more suitable than using a single random tree algorithm both in 10-fold cross-validation and 66–34 percent validation.
C43	[76]	ISCX dataset	Proposed a hybrid model based on KM and RF. KM performs clustering of the best features in the first stage. RF performs the classification of the clusters. The accuracy of the model can be further improved by improving the clustering operation of the KM
C44	[77]	KDD99	The model deployed hybrid rice algorithm to optimize the extreme learning machine. HRO-ELM improves the accuracy of network intrusion detection. How to improve the structure of ELM using hybrid rice algorithm was proposed for future research.
C45	[78]	Real large-scale dataset	The model deployed PCA, CCA, and ICA for dimensionality reduction, and the goal is to learn and maintain the key features. What follows is a classification of the key features using a Bloom filter to categorize the dataset as either abnormal or normal. The normal dataset is further classified using the KNN algorithm. In the future, the model can be evaluated using a different type of SCADA dataset.
C46	[79]	UNSW-NB15 and ISCX2012	The researchers used EBat to optimize MLP neural networks to increase the accuracy of classification. EBat algorithm was used in the selection of suitable weights and biases.
		KDD Cup 99 and NSL-KDD	The researchers used both current and classical datasets for the evaluation of the model.
C47	[80]	MSU and ORNL	The researchers used WOA to obtain the optimal weights and biases for training ANN. The model recorded superior performance compared with other models.
C48	[81]	NSL-KDD	The main purpose of the model was to increase the rate of precision in the detection of malicious activities in information systems through a selection of appropriate features. The researchers used NSL-KDD which does not capture new attacks.
C49	[82]	NSL-KDD and CICIDS2017	The researchers used double particle swarm optimization (PSO)-based algorithm for feature selection and hyperparameter selection. PSO was used to set the hyperparameters of a deep learning model automatically. The model was evaluated using NSL-KDD and CICIDS2017. CICIDS2017 is considered to be an up-to-date reliable NIDS dataset.

Table 1: Continued.

ID	Reference	Dataset	Strength/weakness
C50	[83]	NSL-KDD dataset	The research proposes an incremental learning model for DDoS attack detection. When the divergence test fails to detect an attack, the output forms the input into the classifiers. The classifiers are arranged in parallel to speed the detection process and the cost of computation. Finally, the determiner flags the attack if any. The advantage of using more than one classifier is that algorithms select a different category of features.
C51	[84]	NSL-KDD dataset	The model combines LSTM and decision tree; at the first level, LSTM is used to cluster data as normal or attack. On the second level of detection, the normal data from the first level are fed into the decision tree for further inspection. The model recorded a low detection rate to some attacks like U2R due to small samples. In the future, research can be done on how to balance the dataset.
C52	[85]	CICIDS2017, UNSW-NB15, and NSL-KDD	The researchers developed a model known as MS-DHPN. The model combines multimode deep autoencoder and LSTM. Multimode deep autoencoder forms the first layer of the model. The goal of MDAE is to learn and process multifeature groups. At the second layer, LSTM is used for temporal feature extraction automatically. In the future, the model can be tested in a real-world environment.
C53	[86]	NSL-KDD	SMOTE-ENN is used for data balancing to increase the minority classes. The model uses CNN for feature selection. The model recorded low accuracy of 83.31%; in future, research can focus on how to improve this accuracy.
C54	[87]	UNSW-NB15	The model deploys PCA for feature reduction, to select only the relevant features. K-means is used for clustering and SVM is used for classification. The model reports a high false alarm rate compared with other models, which makes it risky to deploy the model in a production environment.
C55	[88]	KDD Cup 99 dataset	The model integrates a genetic algorithm with improved feature selection with SVM. GA performs the initial feature selection. SVM classifies the selected features into either normal or abnormal (DOS, probe, R2L, and U2R). The model can be further evaluated using updated datasets.
C56	[89]	CICIDS2017	In this model, DBN is used to reduce the number of features in the dataset and keep only the important features. SVM on the other hand is used for the classification of attacks. In the future, research can be done on how to improve the accuracy and the efficiency of intrusion detection.
C57	[90]	ISCX2012 and UNSW-NB15 and KDD Cup 99 and NSL-KDD	The model combines ABC and DA to form an optimization algorithm known as HAD. HAD is used in this model for optimizing the MLP neural network. The model was evaluated using two modern datasets, i.e., ISCX2012 and UNSW-NB15 and two old datasets, that is, KDD Cup 99 and NSL-KDD. In future, research can be done on how to reduce the features in the dataset.
C58	[91]	NSL-KDD	The model applied XGBoost for feature selection and deep neural network (DNN) for the classification of network intrusion. The researchers used only one dataset for the validation of the model.
C59	[92]	ISCX 2012, NSL-KDD and CIC-IDS2017	In this research, ECAGOA is used to optimize SVM by selecting key SVM parameters to eliminate overfitting issues of SVM. When the model is evaluated using three types of datasets, it records superior performance compared with other models.
C60	[93]	NSL-KDD, AWID, and CIC-IDS 2017	In this model, MSAP-GOBA, a variety of GOA, is used to select relevant features in the dataset to improve the detection rate and reduce overfitting problems. Three forms of the dataset were used to validate the model and the result was outstanding compared to other models.
C61	[94]	ADFA-LD dataset	The model uses VED on the first stage to reconstruct the dataset and the RNN is used for capacity memorization. The model recorded a low false-positive rate compared with other models.
C62	[95]	NSL-KDD dataset, UNSW-NB15 dataset, and CIC-IDS2017	The main objective was to reduce the number of features in the dataset to improve classification performance and computation time. NSGA-II id is used as a search strategy and LR is used as a learning algorithm. The model succeeded in reducing the number of features, hence increasing detection accuracy, but this reduced the detection rate of some of the attacks like U2R, backdoor, analysis, exploits, DoS, and web-attack-XSS. This was due to the underrepresentation of the attacks or missing information.

Table 1: Continued.

ID	Reference	Dataset	Strength/weakness
C63	[96]	KDD CUP99	The research showed that in dealing with data redundancy and class imbalance, we can solve the problem of high false-positive rate (FPR) for minority samples and improve F1.
C64	[97]	NSL-KDD and UNSW-NB15	The integration of the two algorithms enabled the learning of spatial and temporal features. The researchers recommended the optimization of the model to detect U2R and worm attacks.
C65	[98]	CICIDS2017 and NSL-KDD	The researchers used both current and classical dataset to validate the model. The model recorded a very low false-positive rate and high accuracy, above 99% on each dataset.
C66	[99]	KDD cup, database1, and database2	The model outperformed other state-of-the-art algorithms in terms of accuracy, detection accuracy, precision, and recall. The major limitation of the model is that it registered high computational costs. Studies can be done on how to reduce the computational cost.
C67	[100]	NSL-KDD and CICIDS2017	The hybrid model of binary classification outperformed other models in precision and recall. In addition, the model significantly reduced the processing time compared to the KNN algorithm. The model focused on the first level of the two-stage detection method; in the future, research can be done on attack detection at the second level of detection.
C68	[101]	NSL-KDD dataset	The method outperformed all the methods, such as GS-ANN, DT, GD-ANN, GA- ANN, PSO-ANN, and GSPSO-ANN. Further validation is needed in the future using different datasets.
C69	[102]	Bot-IoT dataset	The researchers proposed a two-stage hybrid intrusion detection system. The deep autoencoders (DAEs) were deployed on the first stage for anomaly intrusion detection. In the second stage, the researchers deployed machine learning-based attack classifiers. The model performed better in the detection of both known and unknown attacks.
C70	[103]	CICIDS2017	The hybrid method was effective in the classification of anomaly detection compared to other classifications of DNN. The model was evaluated with an updated dataset named "CICIDS2017" which captures current intrusion.
C71	[104]	Alibaba Tianchi dataset	The model combined CNN and LSTM to develop a hybrid detection model. The model outperformed other models in terms of accuracy and MSE. The model was evaluated using Alibaba Tianchi dataset which represents real-life malicious behaviors. In the future, the model needs to be evaluated using different datasets.
C72	[105]	CICIDS2017 datasets	The model outperformed other intrusion detection models in terms of detection rate, accuracy, and false-positive-rate. The model was evaluated using an up-to-date dataset. In the future, the model can be tested using other datasets.
C73	[106]	KDD Cup 99 dataset	The model consists of three stages; in the first stage, the model deploys U-Net and LSTM for feature extraction. In stage two, global attention mechanism is used to learn and select critical information in the features. Finally, SVM is used to classify the information. The model was evaluated using an old dataset that does not capture current intrusion.
C74	[107]	KDD CUP	The model had high performance compared to single algorithms, but the researchers used an old dataset that does not capture modern attacks.
C75	[108]	KDDCUP99 and UNSW-NB15	HNGFA outperforms other techniques in exploration, detection, and evolving rules for all small forms of intrusion with high accuracy and low FAR in different settings of the datasets.
C76	[109]	CICIDS2017	The model takes advantage of two classifiers, that is, long short-term memory and convolutional neural network. The model possesses the capability of detecting evolving cyber threats.
C77	[110]	KDD 99	CSO was used for feature reduction and RNN was used for classification. The researchers used an old dataset to test the model and they proposed the use of a modern dataset for future research.
C78	[111]	KDD Cup99	The research proposed a three-stage hybrid intrusion detection model. Snort was used to detect signature-based attacks in the first stage. In the second stage, three feature reduction techniques were applied for feature reduction. The techniques used were univariate, principal component analysis, and linear discriminant analysis. Finally, the model deployed four supervised machine learning algorithms for classification. The model was evaluated using KDD CUP99, and it was observed that RF outperformed other models in terms of accuracy. Currently, we have new forms of datasets that are up to date in terms of attacks. The model can be evaluated using this dataset.
C79	[112]	UNSW-NB15	The researchers compared the CART algorithm with other decision tree classifiers, namely, the J48 decision tree, fast decision tree, random tree, fine tree, medium tree, and coarse tree. CART recorded superior performance.

Table 1: Continued.

ID	Reference	Dataset	Strength/weakness
C80	[113]	NSL-KDD	The model yielded better results compared with other models, but the researchers observed that the increase of neurons caused an increase in complexity and run times.
C81	[114]	NSL-KDD, UNSW NB15, and Kyoto2006	Information gain and principal component analysis are used for feature extraction and reduction. DBSCAN is used for clustering the dataset. WGAN-DIV was applied in the final stage for data generation. The researchers proposed stability improvement of the model in future work.
C82	[115]	Microsoft Windows server event logs	The research showed the importance of user profile creation in the performance of misuse intrusion detection systems.
C83	[116]	Real dataset	The model uses online incremental SVM for the detection of intrusion on IoT platforms. To make sure that new forms of attacks are detected, MLP is deployed as the second layer of IDS to filter any undetected attacks by the SVM module. The advantage of the model is that it evolves with new forms of attacks due to regular updates from the Internet.
C84	[117]	NSL-KDD and UNSW-NB15	The model combined CNN and BiLSTM for feature extraction. The model extracts spatial and temporal features of the dataset simultaneously. Application of the two algorithms in construction of a balanced dataset improves the learning capabilities of the model, hence reducing the time required to train the model.
C85	[118]	NSL-KDD and CICIDS2017	The researchers observed that deep learning models' performance is highly dependent on the amount of data used for training. If a lot of data is used for training, the model will perform better.
C86	[119]	NSL-KDD	The method aimed at feature reduction to improve the classification. NSL-KDD was used for evaluation which does not reflect new forms of attacks.
C87	[120]	NSL-KDD	The model was superior compared to other clustering algorithms for unsupervised detection but with high computation time.
C88	[121]	CAN-intrusion-dataset and CICIDS2017 dataset	Decision tree (DT), random forest (RF), extra trees (ET), and extreme gradient boosting (XGBoost) algorithms are applied in this model to develop a signature-based IDS. The second phase of the model deploys a cluster labeling (CL) K-means (CL-K-means) algorithm to develop anomaly-based IDS, and this will detect unknown attacks. For future work, model performance can be improved by investigating other unsupervised learning and online learning methods to be used in the anomaly-based IDS framework.
C89	[122]	CICIDS2017 and UNSW-NB15	Proposed a two-stage intrusion detection system based on DT and RF to improve detection. The first stage extracts some selected features for classification. The second stage deals with the extraction of features that were not classified in the first stage. The model was evaluated using two modern datasets.
C90	[123]	NSL-KDD and CIC-IDS2017	The model combined K-means, deep learning algorithm, and RF. K-means and RF are deployed to classify the event as either normal or attacks, while deep learning algorithms are used to learn the hidden features of attack events. The model recorded better processing speed and less training time.
C91	[124]	ICS dataset	The model combines PSO and ANN to create a hybrid detection model. In this model, the PSO search method is used in the ICS dataset to enhance the classification performance of the ANN model. In the future, further investigation can be done on various optimization techniques to increase detection accuracy.
C92	[125]	KDD CUP 99	The outcome shows that PSO + KNN outperformed PSO + DT in network anomaly detection. In the future, the model should be evaluated with current datasets.
C93	[126]	NSL-KDD dataset	Proposed a two-level intrusion detection model; at level one, the model used Naïve Bayes classifier to detect DoS and probe, and at level two, the model deployed SVM to distinguish R2L and U2R from normal instances. In the future, the model can be tested against other forms of intrusion.
C94	[127]	DARPA 1998	The model proposed the optimization of ANN using GWO. The goal was to solve the limitations of using the backpropagation algorithm, which include local minimum limitations. The model took a longer time to train; in the future, research can be done on how to reduce the training time. In addition, the model can be evaluated using the current dataset.
C95	[128]	NSL-KDD	The researchers used a new algorithm known as slime mould optimization algorithm to optimize the weighted extreme learning machine. The model reduces training time, and the real-time performance of intrusion detection is improved.

Table 1: Continued.

ID	Reference	Dataset	Strength/weakness
C96	[129]	NSL-KDD and CICIDS2017 dataset	In this model, WOA is used to optimize the kernel parameters of the RVM and the weight coefficients of the hybrid kernel. The model was evaluated using two types of datasets, NSL-KDD representing the old dataset and CICIDS2017 dataset representing an updated dataset.
C97	[130]	CICIDS2017 and ISCXIDS2012	The model consists of two major components, packet-based early detectors (PEDs) and hybrid lazy detectors (HLDs). PED performs fist level classification by classifying every packet received. Any packet that does not meet the minimum score in the first level is further classified in the second level using the HLD classifier. PED can be constructed by selecting between RF and DT while HLD can be constructed by selecting between RF and ADT.
C98	[131]	NSL KDD	The model has a high implementation complexity and cost than existing IDSs composed of a simple classifier. Research can be done on how to improve implementation complexity and cost. Proposed hybrid model by combining supervised algorithm Light GBM and unsupervised algorithm K-means. The model recorded superior performance than other models, but it requires higher training time.
C99	[132]	CSE-CIC-IDS2018	The model performed multiclass classification based on multilayer perceptron and LSTM.
C100	[133]	KDD′99 dataset	Proposed a model combining serial-based IDS (SIDS) and parallel-based IDS (PIDS). PIDS is deployed to detect known intrusion and SIDS is deployed to detect unknown intrusion. The model can be evaluated using an updated dataset.
C101	[134]	UNSWNB15 dataset and the CICIDS2018 dataset	The researchers combined OCSVM and LOF to develop a hybrid detection model. OCSVM was deployed on the first phase of the model to detect the flow of the traffic. If the flow of the traffic is abnormal, it forms the input to the second phase of the model which consists of the LOF component. The model proved that it can effectively detect intrusion while maintaining privacy. In the future, the model can be further investigated using real data.
C102	[135]	KDD Cup 99 and NSL-KDD	In this study, the researchers combined SCAE and SVM to develop a hybrid intrusion detection model. In this model, SCAE is used in the first phase for feature extraction and SVM is deployed in the second phase for classification purposes. The hybrid model recorded a better accuracy in detection compared to single SVM and long short-term memory (LSTM). In the future, research can be done on how to reduce the computational time for the hybrid model.
C103	[136]	NSL-KDD	The model uses GA for feature selection. The output of GA forms the input to SVM, DT, and ensemble for classification. The combination of GA and ensemble classifier records the highest detection accuracy. The model can be evaluated using different datasets.
C104	[137]	NSL-KDD and UNSW-NB15	The researchers developed a hybrid semantic deep learning (HSDL) intrusion detection model in the cloud. To develop the model, the study integrated LSTM, CNN, and SVM. LSTM and CNN in this model were used for feature extraction while SVM was used for classification. The model was evaluated using NSL-KDD and UNSW-NB15 datasets. The model recorded high accuracy on both datasets: 99.98% for the NSL-KDD dataset and 98.47% for the UNSW-NB15 dataset.
C105	[138]	Laboratory setup dataset	The researchers proposed a hybrid detection model based on LSTM and transformer for detection of malware system calls. The model consists of a component of LSTM on the first stage and the second phase is made up of two layers of transformers. The model recorded 92.6% precision and a recall of 93.8%. In the future, this performance can be improved.
C106	[139]	CICIDS2017 dataset	The researchers developed a hybrid detection model based on convolutional neural network and network. The goal was to solve the issue of feature selection. To achieve this, the researchers applied the forward feature selection technique. The model was tested using the CICIDS2017 dataset and the result showed that forward feature selection is a promising technique in feature selection. The model can be tested using a different dataset in the future.
C107	[140]	NSL KDDCup 99	The researchers used dragonfly algorithm to optimize multilayer perceptron. OSVM was used for the classification of data as either normal or intrusive.
C108	[141]	KDD′99 and UNSW-NB15 datasets	In the first stage, the researchers improved on feature selection. In the second stage of the model, the researchers combined signature and anomaly-based attack detection techniques. The model recorded a very high rate of accuracy of 99.69%.

Table 1: Continued.

ID	Reference	Dataset	Strength/weakness
C109	[142]	NSLKDD dataset	The model consists of two main components. The first component is made of machine learning algorithms. The second component is made of the inference detection model. The model was evaluated using NSLKDD dataset. The results showed that the inference detection model improves the accuracy of the ML algorithms.
C110	[122]	CICIDS2017	The first component of the model is a signature-based model. In this stage, RF is used to detect known attacks. The second component of the model uses an anomaly intrusion detection mechanism. Weighted K-means is used in this stage for the detection of unknown attacks. The model was evaluated using up-to-date dataset.
C111	[143]	Controller area network (CAN) data	The model combines rule-based technique and machine learning algorithm. The rule-based intrusion detection technique was used in the first stage to increase the speed of detection and reduce false alarms. DNN was deployed in the second stage of the model to increase the overall accuracy of the model by lagging any undetected intrusion from the first phase. The evaluation results showed that the model can detect intrusions in different vehicle models.

Table 2: List of abbreviations.

Abbreviation	Explanation
NB	Naïve Bayes
KNN	K-nearest neighbor
NBNS	NetBIOS name service
MDNS	Multicast DNS
IPFIX	Internet Protocol Flow Information Export
CNN	Convolutional neural network
RNN	Recurrent neural network
PSO	Particle swarm optimization
IRELM	Regularized extreme learning machine
LSTM	Long short-term memory
CSOACN	Self-organized ant colony network
GA-SOFM	Genetic algorithm-self-organized feature map
MCLP	Multiple criteria linear programming
AGAAR	Accelerated genetic algorithm and rough set theory
GPLS	Genetic programming with local search
ACO	Ant colony optimization
KMC-D	K-means clustering with discretization
PCA	Principal component analysis
SMO	Sequential minimal optimization
SOM	Self-organization map
BPNN	Backpropagation neural network
FCM	Fuzzy C-means
KM	K-means
CCA	Canonical correlation analysis
ICA	Independent component
WOA	Whale optimization algorithm
MDAE	Multimode deep autoencoder
DBN	Deep belief network
ABC	Artificial bee colony
SMOTE-ENN	Synthetic minority oversampling technique combined with edited nearest neighbors
CSO	Crow swarm optimization
CART	Classification and regression trees
DBSCAN	Density-based algorithm for discovering clusters in large spatial databases with noise
WGAN-DIV	Wasserstein GAN divergence
BiLSTM	Bi-directional long short-term memory

differently. Capturing the profile of different users as normal has proven to be difficult, hence creating the main limitation of anomaly IDS. With the limitation arises the issue of high false-positive alerts because any abnormal action by the user is considered an attack. Research in this area is focused on how to profile normal action and how to reduce high false-positive rates.

2.2. Misuse/Signature Intrusion Detection System. Misuse intrusion detection systems depend on well-known attack signatures to capture attacks and to flag intrusions using well-known patterns. The well-known signatures are captured and labeled to assist in intrusion detection. The labeled patterns are stored in a database that needs regular updates when new patterns are captured. For detection of attacks, misuse-based IDS compares the received traffic with the stored signatures in the database; if the patterns are similar, the traffic is marked as an intrusion; else, the traffic will be marked as normal.

Unlike anomaly-based IDSs, misuse IDSs are easy to create as the pattern of malicious code is known. The code of the malicious malware is analyzed for a unique pattern, and this pattern is used to create the baseline signature to be used for detection. This makes misuse-based IDSs have a high positive detection rate as they depend on well-known information. Users must keep updating the corresponding databases for new signatures.

Over the years, research has been done on this area of misuse intrusion detection. Zhang et al. [13] proposed a misuse intrusion detection system for defending LAN users using the XGBoost algorithm. To develop and evaluate the model, the researchers used real-time data collected from LAN of 10 different Asian countries. The model was evaluated using collected data from 45 networks. The model recorded 97.5% in overall precision and 97.5% in the overall recall. In addition, the researchers observed that LAN intrusion detection is affected by ARP, MDNS, and NBNS protocols. The main advantage of this model is that it was evaluated using real-time network data which means that the model can be deployed in the existing LANs as it is or with minor changes.

Taher et al. [14] used the artificial neural network (ANN) and support vector machine (SVM) technique to develop a signature-based intrusion detection model. The two algorithms were to find the algorithm with the best performance in terms of classification. NSL-KDD dataset was used for the

evaluation of the models. According to the researchers, the ANN-based model outperformed the SVM model in classification. The ANN-based model recorded a detection rate of 94.02%. The model can be further investigated using an updated dataset.

Erlacher and Dressler [15] proposed Internet Protocol Flow Information Export (IPFIX) signature-based intrusion detection known as FIXIDS. The model uses the newly added HTTP-related flow information elements (IEs) to detect intrusion in high-speed networks. The model outperformed Snort in general. This technique can be investigated further in future for standard flow.

Tug et al. [16], using blockchain technology, proposed collaborative signature-based intrusion detection system referred to as CBSigIDS. The model uses blockchain technology to incrementally update and distribute secure signatures database in a collaborative network. Evaluation of the model shows that blockchain technology can be used to improve the performance of signature-based IDS in secure manure. In future, research can be done on the application of blockchain technology in anomaly IDS.

The main limitation with misuse intrusion detection systems is that they cannot detect zero-day attacks or new forms of attacks. At the point of realization of a new form of attack and the creation of the signature of the attack, most of the computer systems are already left vulnerable. Misuse intrusion detection systems also require large storage memory to store the signature library.

The focus area of research on this type of intrusion detection system is on how to reduce the volume consumed by the database. Another potential area of research is how to make this IDS able to detect zero-day attacks.

3. Hybrid Intrusion Detection System

With the evolving variety of attacks, the two classical IDSs mentioned above cannot protect our information systems effectively. New methods of combining different intrusion detection systems to improve their effectiveness have been proposed. Research has shown that combined algorithms perform better than single algorithms [17].

The goal of hybrid intrusion detection systems is to combine several detection models to achieve better results. A hybrid intrusion detection system consists of two components. The first component processes the unclassified data. The second component takes the processed data and scans it to flag out intrusion activities [18].

Hybrid intrusion detection systems are based on combining two learning algorithms. Each learning algorithm possesses unique features, which assist in improving the performance of the hybrid [19]. Hybrid IDSs can be broadly categorized into cascaded hybrid, integrated-based hybrid, and cluster + single hybrid.

In [5], Kim et al. proposed a hybrid intrusion detection system based on signature-based and anomaly detection components. In the first stage of the model, a misuse detection component was applied to detect known attacks based on the captured patterns. This component was based on the C4.5 decision tree algorithm. The second stage consisted of an anomaly detection component to leverage the shortcomings of the misuse detection component. To develop the second component of the model, multiple one-class SVM algorithms were used. The performance of the model was tested using the KDD Cup 99 dataset. The model performed better than the single traditional IDS.

In [20], the researchers combined feature extraction techniques and classification techniques to increase detection rate while at the same time reducing false alarm rate. In the first stage of the hybrid, chi-square was used for feature selection. The goal of this stage was to reduce the number of features in the dataset but maintaining the important features that capture the attacks. In the second stage, a multiclass support vector machine (SVM) algorithm was used for classification. Multiclass support vector machine was used in this model to improve classification rate. The model was evaluated using the NSL-KDD dataset, with the results showing that the model recorded a high detection rate with a low false alarm rate.

In [21], Khraisat et al. developed a hybrid detection model based on a C5 decision tree classifier and one-class support vector machine (OC-SVM). The model consisted of two major components. A C5.0 decision tree classifier was used to develop the first component of the model for misuse detection. The second component was developed using OC-SVM for anomaly detection. The researchers tested the performance of the model using the NSL-KDD and Australian Defence Force Academy (ADFA) datasets, and the results showed that the hybrid model was superior to single-based models.

Khan proposed a hybrid intrusion detection model based on convolutional neural network (CNN) and recurrent neural network (RNN). The research aimed to improve feature extraction, which is fundamental in the performance of intrusion detection systems. CNN was used in the first phase to extract local features in the dataset, with the RNN being used in the second phase to extract temporal features in the dataset. This technique resolved the issue of data imbalance on the available dataset. To test the performance of the model, the CSE-CIC-DS2018 dataset was used, which is the updated dataset. The model outperformed other intrusion detection models, with an intrusion detection accuracy of 97.75% [22].

In [23], the researchers proposed a hybrid model intrusion detection model for smart home security. The model consisted of two components. The first component applied machine learning algorithms to real-time intrusion detection. Algorithms used in this component included random forest, XGBoost, decision tree, and K-nearest neighbors. The second component applied the misuse intrusion detection technique for detection of known attacks. To test the performance of the model, the CSE-CIC-IDS2018 and NSL-KDD datasets were used. The model recorded an outstanding performance for detection of both network intrusion and user-based anomalies in smart homes.

In [24], the authors proposed a hybrid intrusion detection system for online network intrusion detection. The researchers integrated improved particle swarm optimization and regularized extreme learning machine (IPSO-IRELM). In this study, IPSO was used to optimize IRELM.

The model was tested using UCI balance dataset, NSL-KDD dataset, and UNSWNB15 dataset. The model recorded a high accuracy rate as well as capabilities to classify the minority features.

In [25], a hybrid detection model based on Spark ML and the convolutional-LSTM (Conv-LSTM) network was proposed. The model consists of two components: the first component uses Spark ML to detect anomaly intrusion while the second component deploys Conv-LSTM for misuse detection. To investigate the performance of the model, the researchers used ISCX-UNB dataset. The model recorded an outstanding performance of 97.29% accuracy in detection. The researchers proposed that the model can be evaluated further using a different dataset as a way of attempting to reproduce the results.

In [26], the authors developed an intrusion detection system by combining firefly and Hopfield neural network (HNN) algorithms. The researchers used Firefly algorithm to detect denial-of-sleep attacks through node clustering and authentication.

In [27], the researchers proposed a hybrid detection system for VANET (vehicular ad hoc network). The model consisted of two components. The researchers deployed a classification algorithm on the first component and a clustering algorithm on the second component. In the first stage, they used random forest to detect known attacks through classification. For the second stage, they deployed weighted K-means algorithm for the detection of anomaly intrusion. The model was evaluated using the current dataset, CICIDS2017 dataset. The researchers proposed further evaluation of the model in real-world environments. In another work [28], the researchers integrated random forest algorithm with unsupervised clustering algorithm based on coresets. This model was used for detection of real-time intrusions in VANET. Compared with other models, the model recorded better performance in terms of accuracy, computational time, and detection rate.

Barani [29] proposed a hybrid detection model based on genetic algorithm and artificial immune system (AIS) (GAAIS) for intrusion detection on ad hoc on-demand distance vector-based mobile ad hoc network (AODV-based MANET). The model was evaluated using different routing attacks. Compared with other models, the model improved detection rate and decreased the false alarm rate.

In [30], the researchers used integrated firefly algorithm with a genetic algorithm for feature selection MANET. To classify the selected features in the first stage of the model as either intrusion or normal, the researchers used replicator neural network for classification. The model performance was compared to that of fuzzy-based IDS. The model outperformed fuzzy-based IDS in accuracy as well as precision and recall.

4. Methodology

The methodology used consists of three primary phases: planning, conducting, and reporting as outlined by Kitchenham and Charters [31]. The three steps can be explained as follows:

(a) Planning: the main goal of this phase is to define the research goals and the review protocol. Review protocol defines how the review will be done. It consists of all the elements of review.

(b) Conducting: once the protocol has been defined, the review process can start. The main stages in this phase include identifying relevant research, selecting primary studies, and extracting required data and synthesis data.

(c) Report: finally, in reporting the review, data extraction strategies are defined and the steps to be used in data synthesis are outlined.

5. Review Process

5.1. Research Questions (RQs). The main objective of this paper was to analyze the hybrid intrusion detection system techniques that were developed from 2012 to 2022. The following research questions were developed in line with the main objective:

(a) RQ1: which hybrid techniques have been used in intrusion detection systems? Objective: to identify techniques used in the development of hybrid IDS.

(b) RQ2: which classical algorithms were used in the integration of the hybrid? Objective: to identify commonly used algorithms in hybrid IDS.

(c) RQ3: which evaluation metrics are used in the hybrid intrusion detection systems? Objective: to identify commonly used metrics in the evaluation of IDS.

(d) RQ4: which datasets are used in hybrid intrusion detection system research? Objective: to identify commonly used datasets in hybrid IDS.

5.2. Search Strategy. Research shows that it is important to be guided by a search strategy in the systematic review [31]. In defining our search strategy, we were guided by the steps outlined by Thyago et al. [32]. The main two steps in this process are defining keywords and the sources of the study. The keywords were derived from the research questions. The keywords and synonyms used are as follows:

(1) Hybrid OR Integrated OR Cascaded.

(2) Intrusion detection System OR IDS

(3) Artificial Intelligence OR Machine Learning

We used the Boolean operators (OR) and (AND) to define the search string. The operator (OR) was used between synonyms, while (AND) was used between the keywords. The following search strings were defined:

(1) “Hybrid” OR “Integrated” OR “Cascaded”

(2) “Intrusion detection System” OR “IDS”

(3) “Artificial Intelligence” OR “Machine Learning”

Finally, the search strings were combined as follows: ((1) AND (2) OR (1) AND (2) AND (3)).

The researchers used the following digital libraries that are recognized in publishing research in the area of intrusion detection systems [33].

(i) The Institute of Electrical and Electronics Engineers (IEEE) Library (https://ieeexplore.ieee.org/)

(ii) The Association for Computing Machinery (ACM) Digital Library (https://dl.acm.org/)

(iii) Springer Link (https://link.springer.com/)

(iv) Science Direct (https://www.sciencedirect.com)

Several searches were done on the above listed libraries but the search strings that yielded better result on each database are as follows:

(i) IEEE: ((“Hybrid” OR “Integrated” OR “Cascaded”) AND (“Intrusion detection System” OR “IDS”) AND (Artificial Intelligence OR Machine Learning))

(ii) ACM: [All: [[all: “hybrid”]] OR [All: [all: “integrated”]] OR [[All: [all: “cascaded”]]] AND [[All: [[all:] OR [All: “intrusion detection system”] OR [All:]]]] OR [[All: [all: “ids”]]] AND [[All: [publication] OR [All:]]]]

(iii) Springer Link: ((“Hybrid” OR “Integrated” OR “Cascaded”) AND (“Intrusion detection System” OR “IDS”) AND (Artificial Intelligence OR Machine Learning))

(iv) Science Direct: ((“Hybrid” OR “Integrated” OR “Cascaded”) AND (“Intrusion detection System” OR “IDS”))

The initial search obtained 2,084 articles. Table 3 shows the number of articles obtained from the digital database.

5.3. Publication Selection Criteria. For inclusion criteria, all primary studies that have reviewed hybrid intrusion detection systems and articles published between January 2012 and February 2022 were included in the study. Single algorithm studies, secondary studies, short papers, duplicated studies, non-English studies, and incomplete papers were excluded. In addition, all studies that were not relevant to the research questions were excluded from the research. Table 4 summarizes the inclusion/exclusion criteria.

5.4. Study Selection Process. To conduct the selection process, the papers were selected according to the established strings, and papers were also selected based on the title, abstract, and keywords on this stage. The selected papers from the first selection process were subjected through the second selection process, which was based on reading the entire text of the paper.

The primary reviewer conducted the selection process. The secondary reviewer conducted an inter-rater reliability test on the selected papers. This was done to make sure that there was no bias in the selection process from the primary reviewer. In the first step, 1875 studies were excluded by the reviewers as they did not satisfy the inclusion criteria. Of those excluded, 1786 were out of scope, 8 were grey studies, 27 were single algorithm studies, 53 were short papers, and 1 was duplicate paper. In the second step, 98 studies were excluded by the reviewers as they did not satisfy the inclusion criteria. Of those excluded, 78 were out of scope, 2 were single algorithm studies, 17 were short papers, 1 was non-English paper, and 1 was incomplete paper. In this research 111 papers were selected for the review as shown in Table 5.

Table 3: Number of articles obtained from each digital database.

Digital database	Number of articles
IEEE	563
ACM	337
Springer Link	123
Science Direct	1,025

5.5. Data Extraction Process. The objective of this step is to provide an answer to the research questions for each paper in a semi-structured way. To avoid bias in the data extraction process, a data extraction form was developed. The data extraction form captured key elements to answer the research questions as shown in Table 6.

6. Results and Discussion

6.1. Year of Publication. Figure 1 shows the number of publications per year. The year with the most publication is 2020. The graph indicates a continuous increase in research in the field of hybrid IDS. This can be attributed to the desire of improving the efficiency and effectiveness of IDS.

6.2. Research Questions (RQs). In this section, the outcome of the literature review will be analyzed and discussed as per the research questions.

RQ1: which hybrid techniques have been used in intrusion detection systems?

In this question, the research sought to understand which techniques were used in the development of the hybrid IDS. Research shows that hybrid approaches can be broadly categorized into three: cascaded hybrid, integrated-based hybrid, and cluster + single hybrid.

As shown in Table 7, the most used hybrid technique was the cascaded hybrid technique (72 papers), the integrated-based hybrid technique (36 papers), and the cluster + single technique (3 papers).

RQ2: which classical algorithms were used in the integration of the hybrid?

In this question, the researcher sought to understand the classical algorithms applied to hybrid techniques. It was established that the most used algorithms in hybrid detection systems were SVM, DT, K-means, Naïve Bayes, KNN, GA, and PSO as shown in Table 8. The rest of algorithms appeared less than 5 times in the selected papers.

RQ3: which are the evaluation metrics used in the hybrid intrusion detection system?

Table 4: Inclusion/exclusion criteria.

Number	Type	Description
1	Inclusion	Primary studies
2.	Inclusion	Studies published between January 2012 and February 2022
3	Inclusion	Studies that propose hybrid/integrated/cascaded intrusion detection system
4	Exclusion	Studies that propose single algorithm studies
5.	Exclusion	Grey studies
6.	Exclusion	Short papers (less than 5 pages)
7.	Exclusion	Duplicated studies
8.	Exclusion	Non-English studies
9.	Exclusion	Incomplete papers

Table 5: Selected papers per digital database.

Digital database	Number of articles
IEEE	81
ACM	06
Springer Link	11
Science Direct	13

Table 6: Data extraction form.

Element	Description
ID	Unique identifier for the study
Extractor	
Authors, year, title, and country	
Journal/conference name	
Research question 1	
Research question 2	
Research question 3	
Research question 4	

Metric is the measure of the performance of ML algorithm on a given dataset. Metrics are used mostly to compare the performance of different models and determine the most effective one.

Accuracy is a frequently applied metric. The purpose of this metric is to compare the correctly detected outcomes against the total detected outcomes.

True-positive rate (TPR), also known as either recall, sensitivity, or detection rate, is the fraction of correctly detected positive outcomes compared to positive observation.

False-positive rate (FPR), which is referred to as false alarm rate (FAR) or fall-out, is the fraction of wrongly predicted positive outcomes compared to actual negative observations.

True-negative rate (TNR) is also called specificity. This metric is the ratio of correctly predicted negative outcomes compared to actually negative observations.

False-negative rate (FNR) is also called miss rate. This metric is the ratio of wrongly predicted negative outcomes compared to positive observations.

F-score/F-measure is a measure that combines a model's precision and recall into an overall accuracy figure. F1 scores range from 0 to 1 with 1 being perfect and 0 indicating poor performance.

Precision is the ratio of correctly predicted positive outcome compared to positive prediction.

Time is a metric used to measure the efficiency of a model. This can be done either during the training stage or during the evaluation stage.

This study found that three metrics were used in more than 50% of the research as shown in Figure 2. These are accuracy, detection rate, and false alarm rate. Accuracy tests the performance of a model in terms of the number of correctly predicted results. The higher the accuracy, the better the model. This explains why the metric has been used in most of the studies. TPR or detection rate measures the capabilities of a model to flag attacks. This is a very important metric as the objective of any intrusion detection system is to flag attacks. Lastly, false alarm rate (FAR) is the measure of false alarms produced by the model. The more the false alarms, the poor the model. The metric can be used by the designers to improve the performance of the model by reducing or eliminating false alarms.

The above three metrics form the key evaluation metrics for any detection model. With the three metrics, it is possible to determine the overall performance of a model.

RQ4: which datasets were used in hybrid intrusion detection system research?

Figure 3 depicts datasets used in hybrid intrusion detection system research. Dataset used is one of the most important elements in the development of anomaly-based intrusion detection systems. Despite that, the conducted

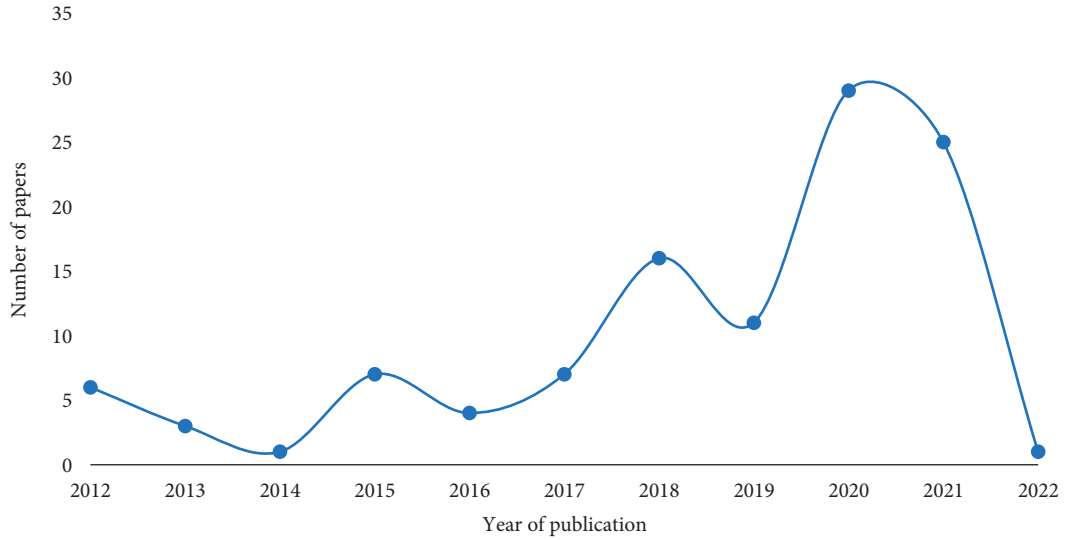

Figure 1: Number of publications per year.

Table 7: Hybrid techniques used in the intrusion detection system.

Hybrid technique	No. of articles
Cascaded hybrid technique	72
Integrated-based hybrid technique	36
Cluster + single technique	3

Table 8: Classical algorithms used in hybrid IDS.

Algorithm	No. of papers	Paper ID
ANN	47	C1, C2, C3, C9, C10, C14, C25, C27, C34, C46, C47, C49, C50, C51, C52, C53, C56, C57, C58, C61, C64, C66, C67, C68, C69, C70, C71, C72, C73, C76, C77, C78, C80, C83, C84, C85, C86, C90, C91, C94, C99, C100, C104, C105, C106, C107, C111
SVM	33	C3, C5, C7, C8, C11, C12, C14, C20, C22, C23, C30, C31, C39, C54, C55, C56, C59, C60, C61, C68, C73, C74, C78, C82, C83, C87, C93, C101, C102, C103, C104, C107, C109
DT	33	C6, C9, C18, C19, C24, C28, C29, C31, C32, C35, C36, C37, C42, C43, C50, C51, C58, C61, C76, C65, C69, C74, C78, C79, C88, C89, C90, C92, C97, C98, C103, C109, C110
K-means	20	C4, C5, C14, C16, C19, C21, C24, C26, C28, C29, C32, C39, C41, C42, C43, C54, C88, C90, C98, C110
GA	12	C8, C9, C10, C11, C15, C17, C55, C61, C62, C74, C80, C103
Naïve Bayes	11	C4, C6, C9, C16, C21, C50, C61, C69, C82, C93, C109
KNN	10	C15, C36, C41, C45, C50, C53, C65, C67, C92, C109
PSO	8	C13, C33, C34, C48, C49, C68, C91, C92

review indicates that researchers are using old datasets in developing hybrid intrusion detection systems. The two most commonly used datasets are KDDCup99 and NSL-KDD. Research shows that these two datasets were developed in 1999. With the ever-changing digital landscape, these datasets cannot be used to develop effective models to combat current cyber threats. The analysis of SLR has shown that we have very few updated datasets to be used in the existing network infrastructure. The pie chart is the representation of our results.

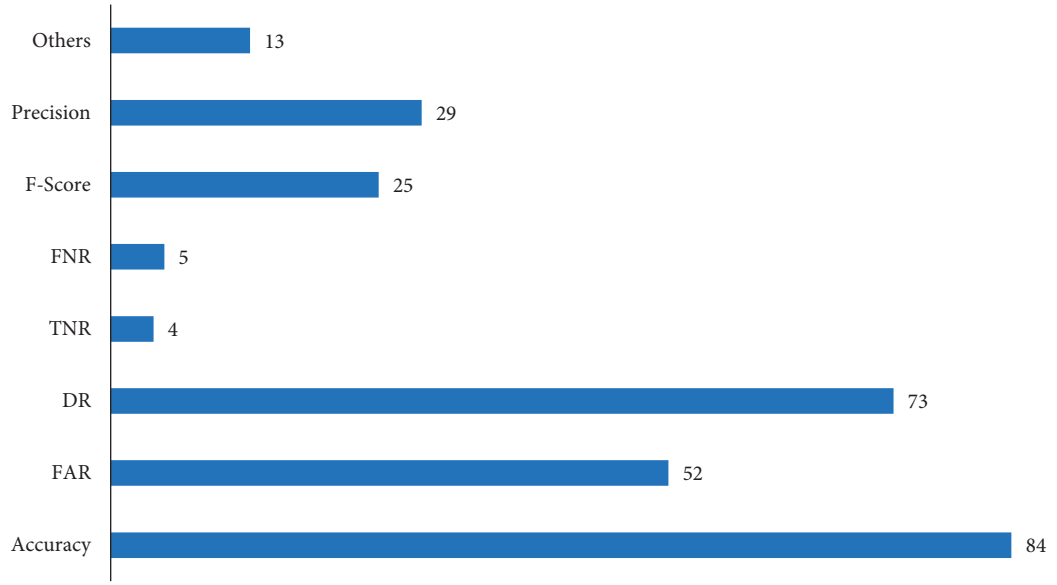

Figure 2: Metrics used for evaluation.

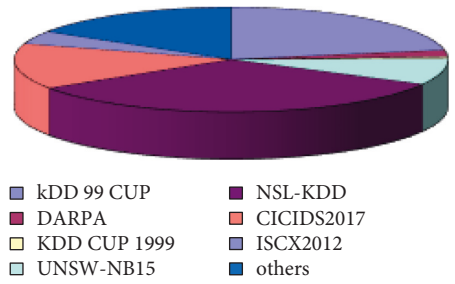

Figure 3: Dataset being used in IDS research from 2012 to 2022.

7. Conclusion

This study has filled the gap that exists in the current body of knowledge on systematic literature review on hybrid intrusion detection systems. This systematic analysis on hybrid IDS points out the existing gaps in the development of hybrid intrusion detection systems and the need for further research on this area. The analysis of SLR indicates that the field of hybrid intrusion detection techniques is an area of focus for many researchers due to its potential of solving the issue of intrusion because this technique increases the performance and efficiency of intrusion detection systems compared to a single algorithm. Investigation on how well to integrate the existing algorithms is of the essence in this field. Most of the hybrid intrusion detection systems are based on three major categories: cascaded hybrid technique, integrated-based hybrid technique, and cluster + single technique. Based on this work, most of the studies focused on cascaded hybrid technique (65%) This method combines the classical algorithms either parallel or in serial format. The second most widely used technique according to the conducted analysis is the integrated-based hybrid technique (35%). This technique aims at optimizing the classical algorithms. Integrated-based hybrids are more efficient and give better results compared to other forms of hybrid techniques. Thus, to develop an efficient and effective IDS, integrated-based hybrid should be adopted in developing the IDS. Lastly, cluster + single technique was the least used technique (3%). The literature review has shown that the existing algorithms have the potential to solve the problem of intrusion but cannot still evolve with the ever-changing digital environment. Most of the models rely on human intervention to update them. There is a need for models which can learn their environment and update themselves without human input.

According to the conducted study, researchers have deployed different types of algorithms in the development of hybrid intrusion detection. The commonly used algorithm includes ANN, SVM, DT, K-means, Naïve Bayes, KNN, GA, and PSO.

For evaluation of the models, fifteen types of datasets were used in the analyzed studies. The datasets that recorded high utilization in the analyzed studies include KDDCup99 and NSL-KDD. Despite their high recorded popularity, these datasets have received criticism from researchers. Most researchers point out that these datasets were developed years ago, and hence they are outdated and ineffective in developing modern intrusion detection systems. In addition, researchers have observed that these datasets do not capture the current forms of detection, and hence they lack the capabilities of defending modern network infrastructure. To resolve this challenge, the analyzed literature review observed emerging datasets which capture current intrusions. These include CICIDS2017, UNSW-NB15, CSE-CIC-IDS2018, and Bot-IoT datasets. The problem is that most of the studies are still using old datasets. For effective IDS, researchers in this field of intrusion detection systems need to embrace the updated datasets.

The three most commonly used metrics for performance evaluation of IDS are accuracy, TPR, and FPR. Future studies should consider also including CPU utilization and detection time as performance metrics. The detection of

intrusion should be done on a real-time basis before any damage is caused, and hence the detection time should be as low as possible. In the development of intrusion detection systems, resource utilization should be considered. In this review, only a few papers included CPU utilization as a performance metric.

Data Availability

The secondary data supporting this systematic review are from previously reported studies and datasets, which have been cited. The processed data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the National Key Research and Development Program of China (grant nos. 2019YFE012990 and 2018YFC1506102), National Natural Science Foundation of China (grant no. 41605121), South African National Research Foundation (grant nos. 114911, 137951, and 132797), and Tertiary Education Support Programme (TESP) of South African ESKOM.

References

[1] D. Tsai, W. Tai, and C. Chang, "A hybrid intelligent intrusion detection system to recognize novel attacks," in *Proceedings of the IEEE 37th Annual International Carnahan Conference on Security Technology*, pp. 428–434, Taipei, Taiwan, October 2003.

[2] H. K. Shaikha and W. M. Abduallah, "Review of intrusion detection systems," *Academic Journal of Nawroz University*, vol. 6, no. 3, pp. 106–111, 2017.

[3] H. J. Highland, "A pattern matching model for misuse intrusion detection," *Computers & Security*, vol. 14, no. 1, p. 28, 1995.

[4] H. Om and A. Kundu, "A hybrid system for reducing the false alarm rate of anomaly intrusion detection system," in *Proceedings of the 2012 1st International Conference on Recent Advances in Information Technology (RAIT)*, pp. 131–136, Dhanbad, India, March 2012.

[5] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1690–1700, 2014.

[6] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. SE-13, no. 2, pp. 222–232, 1987.

[7] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," *Procedia Computer Science*, vol. 60, no. 1, pp. 708–713, 2015.

[8] A. Mishra and P. Yadav, "Anomaly-based IDS to detect attack using various artificial intelligence machine learning algorithms: a review," in *Proceedings of the 2nd International Conference on Data, Engineering and Applications, IDEA*, Bhopal, India, February 2020.

[9] K. Atefi, S. Yahya, A. Rezaei, and S. H. M. H. Binti, "Anomaly detection based on profile signature in network using machine learning technique," in *Proceedings of the 2016 IEEE Region 10 Symposium (TENSYMP)*, pp. 71–76, Sanur, Bali island, Indonesia, May 2016.

[10] T. T. Khoei, G. Aissou, W. C. Hu, and N. Kaabouch, "Ensemble learning methods for anomaly intrusion detection system in smart grid," in *Proceedings of the 2021 IEEE International Conference on Electro Information Technology (EIT)*, pp. 129–135, Mt. Pleasant, MI, USA, May 2021.

[11] T. Rakshe and V. Gonjari, "Anomaly based network intrusion detection using machine learning techniques," *International Journal of Engineering Research and Technology*, vol. 6, no. 5, pp. 216–220, 2017.

[12] V. Kumar, V. Choudhary, V. Sahrawat, and V. Kumar, "Detecting intrusions and attacks in the network traffic using anomaly based techniques," in *Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 554–560, Coimbatore, India, June 2020.

[13] Z. Zhang, P. Chirupphapa, H. Esaki, and H. Ochiai, "XGBoosted misuse detection in LAN-internal traffic dataset," in *Proceedings of the 2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 1–6, Arlington, VA, USA, November 2020.

[14] K. A. Taher, B. Mohammed Yasin Jisan, and M. M. Rahman, "Network intrusion detection using supervised machine learning technique with feature selection," in *Proceedings of the 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, pp. 643–646, Dhaka, Bangladesh, January 2019.

[15] F. Erlacher and F. Dressler, "FIXIDS: a high-speed signature-based flow intrusion detection system," in *Proceedings of the NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, pp. 1–8, Taipei, Taiwan, April 2018.

[16] S. Tug, W. Meng, and Y. Wang, "CBSigIDS: towards collaborative blockchained signature-based intrusion detection," in *Proceedings of the 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 1228–1235, Halifax, NS, Canada, July 2018.

[17] U. S. Musa, S. Chakraborty, M. M. Abdullahi, and T. Maini, "A review on intrusion detection system using machine learning techniques," in *Proceedings of the 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 541–549, Greater Noida, India, February 2021.

[18] M. Khari and A. Karar, "Analysis on intrusion detection by machine learning techniques: a review," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 4, 2013.

[19] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: a review," *Expert Systems with Applications*, vol. 36, no. 10, Article ID 12000, 2009.

[20] I. Sumaiya Thaseen and C. Aswani Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 4, pp. 462–472, 2017.

[21] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, and A. Alazab, "Hybrid intrusion detection system based on the stacking ensemble of C5 decision tree classifier and one class

support vector machine," *Electronics*, vol. 9, no. 1, p. 173, 2020.

[22] M. A. Khan, "HCRNNIDS: hybrid convolutional recurrent neural network-based network intrusion detection system," *Processes*, vol. 9, no. 5, p. 834, 2021.

[23] F. Alghayadh and D. Debnath, "A hybrid intrusion detection system for smart home security based on machine learning and user behavior," *Advances in Internet of Things*, vol. 11, no. 01, pp. 10–25, 2021.

[24] Y. Tang and C. Li, "An online network intrusion detection model based on improved Regularized Extreme learning machine," *IEEE Access*, vol. 9, pp. 94826–94844, 2021.

[25] M. Khan, M. Karim, and Y. Kim, "A scalable and hybrid intrusion detection system based on the convolutional-LSTM network," *Symmetry*, vol. 11, no. 4, p. 583, 2019.

[26] R. Fotohi and B. Firoozi, "A novel countermeasure technique to protect WSN against denial-of-sleep attacks using firefly and Hopfield neural network (HNN) algorithms," *The Journal of Supercomputing*, vol. 76, no. 9, pp. 6860–6886, 2020.

[27] H. Bangui, M. Ge, and B. Buhnova, "A hybrid data-driven model for intrusion detection in VANET," *Procedia Computer Science*, vol. 184, pp. 516–523, 2021.

[28] H. Bangui, M. Ge, and B. Buhnova, "A hybrid machine learning model for intrusion detection in VANET," *Computing*, vol. 104, no. 3, pp. 503–531, 2021.

[29] F. Barani, "A hybrid approach for dynamic intrusion detection in ad hoc networks using genetic algorithm and artificial immune system," in *Proceedings of the 2014 Iranian Conference on Intelligent Systems (ICIS)*, Auckland, New Zealand, February 2014.

[30] D. Shona and M. S. Kumar, "Efficient IDs for MANET using hybrid firefly with a genetic algorithm," in *Proceedings of the 2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 191–194, Coimbatore, India, July 2018.

[31] K. Barbara and C. Stuart, "Guideline for Performing Systematic Literature Reviews in Software Engineering," EBSE Technical Report, Elsevier, Amsterdam, Netherland, 2007.

[32] T. Thyago, I. I. Bittencourt, S. Isotani, and A. Silva, "Does peer assessment in on-line learning environments work? A systematic review of the literature," *Computers in Human Behavior*, vol. 64, pp. 94–107, 2016.

[33] F. Salo, M. Injadat, A. B. Nassif, A. Shami, and A. Essex, "Data mining techniques in intrusion detection systems: a systematic literature review," *IEEE Access*, vol. 6, p. 56046, Article ID 56058, 2018.

[34] S. Mansour and Z. Jadidi, "Flow-based anomaly detection in high-speed links using modified GSA-optimized neural network," *Neural Computing & Applications*, vol. 24, 2012.

[35] S. Mansour and S. R. Maryam, "Gravitational search algorithm–optimized neural misuse detector with selected features by fuzzy grids–based association rules mining," *Neural Computing & Applications*, vol. 23, no. 7-8, pp. 2451–2463, 2012.

[36] K. Qazanfari, M. S. Mirpouryan, and H. Gharaee, "A novel hybrid anomaly-based intrusion detection method," in *Proceedings of the 6th International Symposium on Telecommunications (IST)*, pp. 942–947, Tehran, Iran, November 2012.

[37] R. Chitrakar and C. Huang, "Anomaly based intrusion detection using hybrid learning approach of combining k-medoids clustering and naïve bayes classification," in *Proceedings of the 2012 8th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–5, Limassol, Cyprus, August 2012.

[38] R. Chitrakar and H. Chuanhe, "Anomaly detection using Support Vector Machine classification with k-Medoids clustering," in *Proceedings of the 2012 Third Asian Himalayas International Conference on Internet*, pp. 1–5, Kathmundu, Nepal, November 2012.

[39] P. Natesan and P. Rajesh, "Cascaded classifier approach based on Adaboost to increase detection rate of rare network attack categories," in *Proceedings of the 2012 International Conference on Recent Trends in Information Technology*, pp. 417–422, Chennai, Tamil Nadu, India, April 2012.

[40] W. Feng, Q. Zhang, G. Hu, and J. X. Huang, "Mining network data for intrusion detection through combining SVMs with ant colony networks," *Future Generation Computer Systems*, vol. 37, pp. 127–140, 2014.

[41] S. Anil and R. Remya, "A hybrid method based on genetic algorithm, self-organised feature map, and support vector machine for better network anomaly detection," in *Proceedings of the 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pp. 1–5, Tiruchengode, India, July 2013.

[42] A. S. A. Aziz, A. E. Hassanien, S. E.-O. Hanaf, and M. F. Tolba, "Multi-layer hybrid machine learning techniques for anomalies detection and classification approach," in *Proceedings of the 13th International Conference on Hybrid Intelligent Systems (HIS 2013)*, pp. 215–220, Gammarth, Tunisia, December 2013.

[43] M. Barati, A. Abdullah, N. I. Udzir, R. Mahmod, and N. Mustapha, "Distributed Denial of Service detection using hybrid machine learning technique," in *Proceedings of the 2014 International Symposium on Biometrics and Security Technologies (ISBAST)*, pp. 268–273, Kuala Lumpur, Malaysia, August 2014.

[44] R. Rasoul, C. Milad, E. Mohammad et al., "A hybrid method consisting of GA and SVM for intrusion detection system," *Neural Computing & Applications*, vol. 27, no. 6, pp. 1669–1676, 2015.

[45] Y. Maleh, A. Ezzati, Y. Qasmaoui, and M. Mbida, "A global hybrid intrusion detection system for wireless sensor networks," *Procedia Computer Science*, vol. 52, pp. 1047–1052, 2015.

[46] S. M. H. Bamakan, B. Amiri, M. Mirzabagheri, and Y. Shi, "A new intrusion detection approach using PSO based multiple criteria linear programming," *Procedia Computer Science*, vol. 55, pp. 231–237, 2015.

[47] R. Ujwala, M. Nilesh, and P. Puja, "Feature selection based hybrid anomaly intrusion detection system using K means and RBF kernel function," *Procedia Computer Science*, vol. 45, pp. 428–435, 2015.

[48] Y. Canbay and S. Sagiroglu, "A hybrid method for intrusion detection," in *Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pp. 156–161, Miami, Florida, USA, December 2015.

[49] S. Varuna and P. Natesan, "An integration of k-means clustering and naïve bayes classifier for Intrusion Detection," in *Proceedings of the 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN)*, pp. 1–5, Chennai, Tamil Nadu, India, March 2015.

[50] A.-R. Hedar, M. A. Omer, A. F. Al-Sadek, and A. A. Sewisy, "Hybrid evolutionary algorithms for data classification in intrusion detection systems," in *Proceedings of the 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/*

Distributed Computing (SNPD), pp. 1–7, Takamatsu, June 2015.

[51] K. Jasmin, J. Samed, and S. Abdulhamit, "An effective combining classifier approach using tree algorithms for network intrusion detection," *Neural Computing & Applications*, vol. 28, pp. 1051–1058, 2016.

[52] A. D. Landress, "A hybrid approach to reducing the false positive rate in unsupervised machine learning intrusion detection," *2016*, pp. 1–6, 2016.

[53] A. Wankhade and K. Chandrasekaran, "Distributed-intrusion detection system using combination of ant colony optimization (ACO) and support vector machine (SVM)," in *Proceedings of the 2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE)*, pp. 646–651, Ghaziabad, India, September 2016.

[54] H. M. Tahir, A. M. Said, N. H. Osman, N. H. Zakaria, P. N. M. Sabri, and N. Katuk, "Oving K-means clustering using discretization technique in network intrusion detection system," in *Proceedings of the 2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*, pp. 248–252, Kuala Lumpur, Malaysia, August 2016.

[55] V. V. Kumari and P. R. K. Varma, "A semi-supervised intrusion detection system using active learning SVM and fuzzy c-means clustering," in *Proceedings of the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 481–485, Palladam, Tamil Nadu, India, February 2017.

[56] J. Zhu, J. Mu, D. Wei, B. Feng, Y. Wang, and K. Yin, "A spatial correlation-based hybrid method for intrusion detection," in *Proceedings of the 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*, pp. 1097–1102, 2017.

[57] Y. Y. Aung and M. M. Min, "An analysis of random forest algorithm-based network intrusion detection system," in *Proceedings of the 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pp. 127–132, Kanazawa, Japan, June 2017.

[58] F. Meng, Y. Fu, F. Lou, and Z. Chen, "An effective network attack detection method based on kernel PCA and LSTM-RNN," in *Proceedings of the 2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*, pp. 568–572, Dalian, China, December 2017.

[59] S. M. A. M. Gadal and R. A. Mokhtar, "Anomaly detection approach using hybrid algorithm of data mining technique," in *Proceedings of the 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*, pp. 1–6, Khartoum, Sudan, January 2017.

[60] S. AlHamouz and A. Abu-Shareha, "Hybrid classification approach using self-organizing map and back propagation artificial neural networks for intrusion detection," in *Proceedings of the 2017 10th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 83–87, Paris, France, June 2017.

[61] P. Shukla, "ML-IDS: a machine learning approach to detect wormhole attacks in Internet of Things," in *Proceedings of the 2017 Intelligent Systems Conference (IntelliSys)*, pp. 234–240, London, United Kingdom, September 2017.

[62] A. Pattawaro and C. Polprasert, "Anomaly-based network intrusion detection system through feature selection and hybrid machine learning technique," in *Proceedings of the 2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE)*, pp. 1–6, Bangkok, Thailand, November 2018.

[63] S. Sagar, A. Shrivastava, and C. Gupta, "Feature Reduction and selection based optimization for hybrid intrusion detection system using PGO followed by SVM," in *Proceedings of the 2018 International Conference on Advanced Computation and Telecommunication (ICACAT)*, pp. 1–7, Bhopal, India, December 2018.

[64] P. Bhatt and A. Morais, "HADS: hybrid anomaly detection system for IoT environments," in *Proceedings of the 2018 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*, pp. 191–196, Hammamet, Tunisia, December 2018.

[65] P. Singh and M. Venkatesan, "Hybrid approach for intrusion detection system," in *Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pp. 1–5, Coimbatore, India, March 2018.

[66] S. Rani and S. Jain, "Hybrid approach to detect network based intrusion," in *Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1–5, Pune, India, August 2018.

[67] A. S. Sadiq, B. Alkazemi, S. Mirjalili, N. Ahmed, I. Ali, and G. K. Zrar, "An efficient IDS using hybrid magnetic Swarm optimization in WANETs," *IEEE Access*, vol. 6, Article ID 29053, 2018.

[68] N. T. Pham, E. Foo, S. Suriadi, H. Jeffrey, and H. F. M. Lahza, "Improving performance of intrusion detection system using ensemble methods and feature selection," in *Proceedings of the Australasian Computer Science Week Multiconference (ACSW '18)*, pp. 1–6, Melbourne, VIC, Australia, February 2018.

[69] F. Zohreh and L. Yue, "Intrusion detection system by using hybrid algorithm of data mining technique," in *Proceedings of the ICSCA 2018: Proceedings of the 2018 7th International Conference on Software and Computer Applications*, pp. 119–123, Kuantan, Malaysia, February 2018.

[70] J. Jianguo, W. Qiwen, S. Zhixin, L. Bin, and Q. Biao, "RST-RF: a hybrid model based on Rough set theory and random forest for network intrusion detection," in *Proceedings of the ICCSP 2018: Proceedings of the 2nd International Conference on Cryptography, Security and Privacy*, pp. 77–81, Guiyang, China, March 2018.

[71] V. Hajisalem and S. Babaie, "A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection," *Computer Networks*, vol. 136, pp. 37–50, 2018.

[72] M. Nivaashini and P. Thangaraj, "A framework of novel feature set extraction based intrusion detection system for Internet of things using hybrid machine learning algorithms," in *Proceedings of the 2018 International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 44–49, Greater Noida, India, September 2018.

[73] C. Jin, Z. Ye, C. Wang, L. Yan, and R. Wang, "A network intrusion detection method based on hybrid Rice optimization algorithm improved fuzzy C-means," in *Proceedings of the 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)*, pp. 47–52, Lviv, Ukraine, September 2018.

[74] Y. Y. Aung and M. Myat Min, "Hybrid intrusion detection system using K-means and K-nearest neighbors algorithms," in *Proceedings of the 2018 IEEE/ACIS 17th International*

Conference on Computer and Information Science (ICIS), pp. 34–38, Singapore, June 2018.

[75] Y. Y. Aung and M. M. Min, "Hybrid intrusion detection system using K-means and random tree algorithms," in *Proceedings of the 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pp. 218–223, Busan, Korea (South), June 2018.

[76] S. Soheily-Khah, P. Marteau, and N. Béchet, "Intrusion detection in network systems through hybrid supervised and unsupervised machine learning process: a case study on the iscx dataset," in *Proceedings of the 2018 1st International Conference on Data Intelligence and Security (ICDIS)*, pp. 219–226, TX, USA, April 2018.

[77] X. Zheng, Z. Ye, J. Su, H. Chen, and R. Wang, "Network intrusion detection based on hybrid Rice algorithm optimized Extreme learning machine," in *Proceedings of the 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)*, pp. 149–153, Lviv, Ukraine, September 2018.

[78] I. A. Khan, D. Pi, Z. U. Khan, Y. Hussain, and A. Nawaz, "HML-IDS: a hybrid-multilevel anomaly prediction approach for intrusion detection in scada systems," *IEEE Access*, vol. 7, Article ID 89521, 2019.

[79] G. Waheed and J. Aman, "A new approach for intrusion detection system based on training multilayer perceptron by using enhanced Bat algorithm," *Neural Computing and Applications*, vol. 32, no. 15, Article ID 11698, 2019.

[80] L. Haghnegahdar and Y. Wang, "A whale optimization algorithm-trained artificial neural network for smart grid cyber intrusion detection," *Neural Computing & Applications*, vol. 32, no. 13, pp. 9427–9441, 2019.

[81] S. Velliangiri and P. Karthikeyan, "Hybrid optimization scheme for intrusion detection using considerable feature selection," *Neural Computing & Applications*, vol. 32, 2019.

[82] E. Wisam, A. Akhan, and Z. Abdul, "Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic," *Computer Networks*, vol. 168, Article ID 107042, 2019.

[83] H. Soodeh and A. Mehrdad, "The hybrid technique for DDoS detection with supervised learning algorithms," *Computer Networks*, vol. 158, pp. 35–45, 2019.

[84] S. Wang, C. Xia, and T. Wang, "A novel intrusion detector based on deep learning hybrid methods," in *Proceedings of the 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pp. 300–305, Washington, DC, USA, May 2019.

[85] H. He, X. Sun, H. He, G. Zhao, L. He, and J. Ren, "A novel multimodal-sequential approach based on multi-view features for network intrusion detection," *IEEE Access*, vol. 7, Article ID 183221, 2019.

[86] X. Zhang, J. Ran, and J. Mi, "An intrusion detection system based on convolutional neural network for imbalanced network traffic," in *Proceedings of the 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, pp. 456–460, Dalian, China, October 2019.

[87] I. Aljamal, A. Tekeoğlu, K. Bekiroglu, and S. Sengupta, "Hybrid intrusion detection system using machine learning techniques in cloud computing environments," in *Proceedings of the 2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 84–89, HI, USA, May 2019.

[88] E. C. Matel, A. M. Sison, and R. P. Medina, "Optimization of network intrusion detection system using genetic algorithm with improved feature selection technique," in *Proceedings of the 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pp. 1–6, Laoag, Philippines, November 2019.

[89] H. Zhang, Y. Li, Z. Lv, A. K. Sangaiah, and T. Huang, "A real-time and ubiquitous network attack detection based on deep belief network and support vector machine," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 3, pp. 790–799, 2020.

[90] W. A. H. M. Ghanem, A. Jantan, S. A. A. Ghaleb, and A. B. Nasser, "An efficient intrusion detection model based on hybridization of artificial bee colony and dragonfly algorithms for training multilayer perceptrons," *IEEE Access*, vol. 8, Article ID 130475, 2020.

[91] D. Preethi and K. Neel, "An efficient XGBoost–DNN-based classification model for network intrusion detection system," *Neural Computing & Applications*, vol. 32, Article ID 12514, 2020.

[92] D. Shubhra, V. Manu, and T. Sarsij, "An effect of chaos grasshopper optimization algorithm for protection of network infrastructure," *Computer Networks*, vol. 176, 2020.

[93] A. K. Shukla, "Detection of anomaly intrusion utilizing self-adaptive grasshopper optimization algorithm," *Neural Computing & Applications*, vol. 33, no. 13, pp. 7541–7561, 2020.

[94] L. Bouzar-Benlabiod, S. H. Rubin, K. Belaidi, and N. E. Haddar, "RNN-VED for reducing false positive alerts in host-based anomaly detection systems," in *Proceedings of the 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 17–24, Las Vegas, NV, USA, August 2020.

[95] K. Chaouki and K. Saoussen, "A NSGA2-LR wrapper approach for feature selection in network IntrusionDetection," *Computer Networks*, vol. 172, 2020.

[96] P. Leilei and X. Xiaolan, "Network intrusion detection model based on PCA + ADASYN and XGBoost," pp. 44–48, 2020.

[97] S. Jay and M. Manollas, "Efficient Deep CNN-BiLSTM Model for Network Intrusion Detection," pp. 223–231, 2020.

[98] A. Hanane and A. Z. E. Boukhamla, "Two-Stage Intrusion Detection System Based on Hybrid Methods," 2020.

[99] S. Velliangiri and H. M. Pandey, "Fuzzy-Taylor-elephant herd optimization inspired Deep Belief Network for DDoS attack detection and comparison with state-of-the-arts algorithms," *Future Generation Computer Systems*, vol. 110, pp. 80–90, 2020.

[100] C. A. de Souza, C. B. Westphall, R. B. Machado, J. B. M. Sobral, and G. d. S. Vieira, "Hybrid approach to intrusion detection in fog-based IoT environments," *Computer Networks*, vol. 180, Article ID 107417, 2020.

[101] H. Soodeh and M. H. Z. Behnam, "New hybrid method for attack detection using combination of evolutionary algorithms, SVM, and ANN," *Computer Networks*, vol. 173, Article ID 107168, 2020.

[102] G. Bovenzi, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapé, "A hierarchical hybrid intrusion detection approach in IoT scenarios," in *Proceedings of the GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–7, Taipei, Taiwan, December 2020.

[103] K. Atefi, H. Hashim, and T. Khodadadi, "A hybrid anomaly classification with deep learning (DL) and binary algorithms

(BA) as optimizer in the intrusion detection system (IDS)," in *Proceedings of the 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, pp. 29–34, Langkawi Island, Malaysia, February 2020.

[104] A. Xu, L. Cheng, X. Kuang, H. Lv, Y. Jiang, and B. Li, "A hybrid deep learning model for malicious behavior detection," in *Proceedings of the 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pp. 55–59, Baltimore, MD, USA, May 2020.

[105] H. Mennour and S. Mostefai, "A hybrid deep learning strategy for an anomaly based N-ids," in *Proceedings of the 2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pp. 1–6, Fez, Morocco, June 2020.

[106] W. Chen, H. Cao, X. Lv, and Y. Cao, "A hybrid feature extraction network for intrusion detection based on global attention mechanism," in *Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, pp. 481–485, Guiyang, China, April 2020.

[107] A. Kumari and A. K. Mehta, "A hybrid intrusion detection system based on decision tree and support vector machine," in *Proceedings of the 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, pp. 396–400, Greater Noida, India, October 2020.

[108] R. Elhefnawy, H. Abounaser, and A. Badr, "A hybrid nested genetic-fuzzy algorithm framework for intrusion detection and attacks," *IEEE Access*, vol. 8, pp. 98218–98233, 2020.

[109] J. Malik, A. Akhunzada, I. Bibi, M. Imran, A. Musaddiq, and S. W. Kim, "Hybrid deep learning: an efficient Reconnaissance and surveillance detection mechanism in SDN," *IEEE Access*, vol. 8, pp. 134695–134706, 2020.

[110] A. A. Abdul Lateef, S. T. Faraj Al-Janabi, and B. Al-Khateeb, "Hybrid intrusion detection system based on deep learning," in *Proceedings of the 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, pp. 1–5, Sakheer, Bahrain, October 2020.

[111] V. S. F. Enigo, K. T. Ganesh, N. N. V. Raj, and D. Sandeep, "Hybrid intrusion detection system for detecting new attacks using machine learning," in *Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 567–572, Coimbatore, India, June 2020.

[112] Z. Chkirbene, S. Eltanbouly, M. Bashendy, N. AlNaimi, and A. Erbad, "Hybrid machine learning for network anomaly intrusion detection," in *Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, pp. 163–170, Doha, Qatar, February 2020.

[113] M. H. Ali, K. Al-Jawaheri, M. M. Adnan, A. Aasi, and A. H. Radie, "Improved intrusion detection accuracy based on optimization fast learning network model," in *Proceedings of the 2020 3rd International Conference on Engineering Technology and its Applications (IICETA)*, pp. 198–202, Najaf, Iraq, September 2020.

[114] D. Li, D. Kotani, and Y. Okabe, "Improving attack detection performance in nids using gan," in *Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 817–825, Madrid, Spain, July 2020.

[115] P. Pokharel, R. Pokhrel, and S. Sigdel, "Intrusion detection system based on hybrid classifier and user profile enhancement techniques," in *Proceedings of the 2020 International Workshop on Big Data and Information Security (IWBIS)*, pp. 137–144, Depok, Indonesia, October 2020.

[116] A. E. Ghazi and A. Moulay Rachid, "Machine learning and data mining methods for hybrid IoT intrusion detection," in *Proceedings of the 2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech)*, pp. 1–6, Marrakesh, Morocco, May 2020.

[117] K. Jiang, W. Wang, A. Wang, and H. Wu, "Network intrusion detection combined hybrid sampling with deep hierarchical network," *IEEE Access*, vol. 8, pp. 32464–32476, 2020.

[118] S. N. Pakanzad and H. Monkaresi, "Providing a hybrid approach for detecting malicious traffic on the computer networks using convolutional neural networks," in *Proceedings of the 2020 28th Iranian Conference on Electrical Engineering (ICEE)*, pp. 1–6, Tabriz, Iran, May 2020.

[119] D. Mehanović, D. Kečo, J. Kevrić, S. Jukić, A. Miljković, and Z. Mašetić, "Feature selection using cloud-based parallel genetic algorithm for intrusion detection data classification," *Neural Computing & Applications*, vol. 33, no. 18, pp. 11861–11873, 2021.

[120] G. Pu, L. Wang, J. Shen, and F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," *Tsinghua Science and Technology*, vol. 26, no. 2, pp. 146–153, 2021.

[121] L. Yang, A. Moubayed, and A. Shami, "MTH-IDS: a multitiered hybrid intrusion detection system for Internet of vehicles," *IEEE Internet of Things Journal*, vol. 9, no. 1, 2021.

[122] W. Seo and W. Pak, "Real-time network intrusion prevention system based on hybrid machine learning," *IEEE Access*, vol. 9, Article ID 46397, 2021.

[123] C. Liu, Z. Gu, and J. Wang, "A hybrid intrusion detection system based on scalable K-Means+ random forest and deep learning," *IEEE Access*, vol. 9, Article ID 75740, 2021.

[124] S. Bebortta and S. K. Singh, "An adaptive machine learning-based threat detection framework for industrial communication networks," in *Proceedings of the 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 527–532, Bhopal, India, June 2021.

[125] R. O. Ogundokun, J. B. Awotunde, P. Sadiku, E. A. Adeniyi, M. Abiodun, and O. I. Dauda, "An enhanced intrusion detection system using Particle Swarm optimization feature extraction technique," *Procedia Computer Science*, vol. 193, pp. 504–512, 2021.

[126] T. Wisanwanichthan and M. Thammawichai, "A double-layered hybrid approach for network intrusion detection system using combined naive bayes and SVM," *IEEE Access*, vol. 9, Article ID 138450, 2021.

[127] A. Sharma and U. Tyagi, "A hybrid approach of ANN-GWO technique for intrusion detection," in *Proceedings of the 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, pp. 467–472, Bangalore, India, August 2021.

[128] T. Xiong, G. Lina, Z. Guifen, and Q. Donghong, "A intrusion detection algorithm based on improved slime mould algorithm and weighted Extreme learning machine," in *Proceedings of the 2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 157–161, Chengdu, China, May 2021.

[129] P. Gao, M. Yue, and Z. Wu, "A novel intrusion detection method based on WOA optimized hybrid kernel RVM," in

Proceedings of the 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS), pp. 1063–1069, Chengdu, China, April 2021.

[130] T. Kim and W. Pak, "Hybrid classification for high-speed and high-accuracy network intrusion detection system," *IEEE Access*, vol. 9, Article ID 83817, 2021.

[131] A. K. M. Mashuqur Rahman Mazumder, N. Mohammed Kamruzzaman, N. Akter, N. Arbe, and M. M. Rahman, "Network intrusion detection using hybrid machine learning model," in *Proceedings of the 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, pp. 1–8, Bhilai, Chhattisgarh, February 2021.

[132] B. B. Borisenko, S. D. Erokhin, A. S. Fadeev, and I. D. Martishin, "Intrusion detection using multilayer perceptron and neural networks with long short-term memory," in *Proceedings of the 2021 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO)*, pp. 1–6, Kaliningrad, Russia, June 2021.

[133] I. Das, S. Singh, and A. Sarkar, "Serial and parallel based intrusion detection system using machine learning," in *Proceedings of the 2021 Devices for Integrated Circuit (DevIC)*, pp. 340–344, Kalyani, Nadia, India, March 2021.

[134] J. Shi, Y. Lin, Z. Zhang, and S. Yu, "A hybrid intrusion detection system based on machine learning under differential privacy protection," in *Proceedings of the 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pp. 1–6, Article ID Norman,OK,USA, September 2021.

[135] W. Wang, X. Du, D. Shan, R. Qin, and N. Wang, "Cloud intrusion detection method based on stacked contractive auto-encoder and support vector machine," *IEEE Transactions on Cloud Computing*, 2021.

[136] T. Saba, T. Sadad, A. Rehman, Z. Mehmood, and Q. Javaid, "Intrusion detection system through advance machine learning for the Internet of things networks," *IT Professional*, vol. 23, no. 2, pp. 58–64, 2021.

[137] V. Prabhakaran and A. Kulandasamy, "Hybrid semantic deep learning architecture and optimal advanced encryption standard key management scheme for secure cloud storage and intrusion detection," *Neural Computing & Applications*, vol. 5, 2021.

[138] Y. Guan and N. Ezzati-Jivan, "Malware system calls detection using hybrid system," in *Proceedings of the 2021 IEEE International Systems Conference (SysCon)*, pp. 1–8, Vancouver, Canada, March 2021.

[139] L. S. Matsa, P. G.-A. Zodi-Lusilao, and P. F. Bhunu-Shava, "Forward feature selection for DDoS detection on cross-plane of software defined network using hybrid deep learning," in *Proceedings of the 2021 3rd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, pp. 1–7, Windhoek, Namibia, November 2021.

[140] S. Amaran and R. Madhan Mohan, "An optimal multilayer perceptron with dragonfly algorithm for intrusion detection in wireless sensor networks," in *Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1–5, Erode, India, April 2021.

[141] M. Ozkan-Okay, O. Aslan, R. Eryigit, and R. Samet, "SABADT: hybrid intrusion detection approach for cyber attacks identification in WLAN," *IEEE Access*, vol. 9, Article ID 157653, 2021.

[142] A. Singhal, A. Maan, D. Chaudhary, and D. Vishwakarma, "A hybrid machine learning and data mining based approach to network intrusion detection," in *Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 312–318, Pichanur, Tamil Nadu, March 2021.

[143] L. Zhang and D. Ma, "A hybrid approach toward efficient and accurate intrusion detection for in-vehicle networks," *IEEE Access*, vol. 10, Article ID 10866, 2022.