International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Implementation of Machine Learning Techniques for Risks Evaluation in Cloud and Cybersecurity

Saif Aamer Fadhil[1], Lubna Emad Kadhim[1], Mahmood Ahmed Hamdi[1], Amjed Abbas Ahmed[1,2], Mohammad Kamrul Hasan[2], Shayla Islam[2], Azana Hafizah Mohd Aman[2] and Nurhizam Safie[2]

[1]Department of Computer Techniques Engineering, Imam Al-Kadhum College (IKC), Baghdad 10011, Iraq
[2]Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Malaysia

E-mail address: saifaamer@alkadhum-col.edu.iq, lubnaemad@alkadhum-col.edu.iq, Mahmood@alkadhum-col.edu.iq, amjedabbas@alkadhum-col.edu.iq, mkhasan@ukm.edu.my, shayla@ucsiuniversity.edu.my, azana@ukm.edu.my, nurhizam@ukm.edu.my

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Cloud computing has emerged as an essential element of the modern and future industry. Various corporates utilizes the capabilities of cloud computing services. There is skeptism and fear regarding application of cloud services that remains an open challenge, as it gaining and growing popularity amongst many business entities around globally. Many challenges are determined and found out research; majorly pertaining to security or protection. Since its inceptions Security risks concerning to cloud computing has invited large portion of attention. Cloud computing services and their providers are always on the lookout for new and improved security methods and solutions. Many service providers in field of cloud computing exist with their services in cloud domains. Such services comes with many features, specifications, along with techniques of acquiring security measures. Methodologies acquired and adopted by many service players to attain security is different in nature. Depending on his need and quality of security received from service providers. A user can choose a particular service. For studying a specific service that depends on its many security features is a dominant issue. In order to build a comprehensive risk assessment methodology, an extensive literature review was conducted to identify all risk factors that can affect cloud computing adoption. In this context various risk factors were identified. After feature selection methods and identification of risk factors, utilized to select most effective features. Then machine learning techniques are used as an efficient technique to analyze hazard in an environment of cloud computing. From all the partitioning strategies tested, the results showed that dividing the dataset into 96% and 4% yielded the best results. The Decision Tree Classifier method also performed the best across all the datasets.

Keywords: Cloud computing, Cybersecurity, Risk Evaluation, and Machine Learning.

1. INTRODUCTION

Artificial intelligence (AI) cloud risk assessment tools are a state-of-the-art way to improve cloud cybersecurity [1]. By harnessing the potential of artificial intelligence (AI), these techniques provide organizations with a thorough instrument for identifying, evaluating, and reducing the risks associated with cloud computing [2]. Machine learning algorithms allow AI-based approaches to constantly learn and adjust to new dangers, as opposed to traditional systems that rely on static rules and human analysis. This results in a security policy that is both proactive and dynamic, with the help of these measures.
Modern cybersecurity methods [3], [4] increasingly rely on cloud risk assessment methodologies that are built on machine learning. Businesses can take a proactive stance in identifying and addressing potential threats inside cloud systems by using these strategies. In order to better understand their security posture and vulnerabilities, businesses can use these tools, which include data analytics and artificial intelligence, to sift through the mountain of data produced by cloud platforms.
The capacity of cloud risk assessment based on machine learning [5] to detect patterns and irregularities in real time is a major advantage of this method. Because they keep an eye on user actions, network traffic, and system settings, these methods can spot suspicious activity or hacking attempts in no time. By using this preventative

E-mail: saifaamer@alkadhum-col.edu.iq, lubnaemad@alkadhum-col.edu.iq, Mahmood@alkadhum-col.edu.iq, amjedabbas@alkadhum-col.edu.iq, mkhasan@ukm.edu.my, shayla@ucsiuniversity.edu.my, azana@ukm.edu.my, nurhizam@ukm.edu.my

measure, businesses can identify potential security breaches early on, before they escalate, reducing the impact on operations and ensuring data integrity. Machine learning algorithms can also evolve and adapt over time, learning from past mistakes and enhancing their detection abilities to stay ahead of emerging dangers. Companies can stay ahead of the constantly evolving threat landscape and maintain a solid defensive position against both known and unknown threats by using this dynamic approach to risk assessment.

Machine learning-based cloud risk assessment [6] also has the added benefits of being both effective and scalable. Due to the increasing quantity and complexity of data generated within cloud settings, traditional methods of risk assessment that are carried out manually are now insufficient. Fast risk identification and prioritization based on severity and likelihood of materialization is made feasible by machine learning algorithms' capacity to process and evaluate large amounts of data at scale. In addition, businesses can enhance their ability to meet compliance and regulatory standards by integrating machine learning-based risk assessment methods into their comprehensive cybersecurity plan. Compliance with industry regulations and standards, such as GDPR [7], HIPAA [8], and PCI DSS [9], can be shown by businesses using these approaches. Cloud risk assessment methods powered by machine learning not only identify threats proactively, but also provide predictive capabilities. This lets companies prepare for and deal with any problems even before they happen. Organisations can improve their cloud security posture by proactively addressing identified vulnerabilities and weaknesses using these methods, which include reviewing historical data and spotting patterns.

When it comes to prioritizing risk mitigation activities, machine learning algorithms can help businesses by calculating the probability and possible effect of different risks [10]. Organizations can better allocate resources and prioritize the most essential threats with the aid of these tools, which analyze historical data and correlate it with current security incidents. Cloud risk assessment methods based on machine learning can also aid with incident response and remediation, potentially making both processes more efficient. By reducing the amount of time attackers spend within their cloud systems, these solutions help firms uncover and respond to security vulnerabilities more rapidly. The real-time alerting and automated analysis of security incidents makes this possible.

Machine learning algorithms can help businesses find new vulnerabilities [11] and threats by comparing internal security data with global threat information sources. Examining streams of external threat information is one way to do this. This foresight allows businesses to protect their cloud infrastructure from the constantly shifting cyber dangers by staying one step ahead of the curve.

A. Research Contributions

- Through our research, ultimately businesses now have an effective and thorough tool at their disposal in cloud risk assessment methodologies powered by machine learning. These tools allow enterprises to pinpoint, analyze, and reduce security concerns in their cloud systems.
- In this research work, companies can stay in compliance with regulatory requirements, get greater insights into their security posture, and spot dangerous threats in real time with the use of artificial intelligence and data analytics. Organizations can make advantage of these features.
- Our research work shows machine learning-based risk assessment is going to be crucial in protecting sensitive data and making sure that cloud-based services are available and secure as the threat environment keeps changing.

2. LITERATURE REVIEW

Gert-Jan de Vreede and Gerald M. Masson [12] have done extensive research on cloud-based risk assessment frameworks and decision support systems. Their work often highlights the use of machine learning algorithms to enhance decision-making, especially when it comes to risk assessment. In order to analyze the complex datasets generated by cloud platforms, their research is focused on developing new methods and algorithms. To improve the efficacy and precision of cloud risk assessment, they stress the need of using machine learning techniques. The output of machine learning algorithms isn't always easy to understand because of how complex and cryptic they could be. One drawback of their approach is that it makes it hard for consumers to understand the logic behind risk evaluations.

Yiwei Gong and Qiang Cheng have collaborated to make significant strides in the field of anomaly identification in cloud computing systems using machine learning techniques. Their research looks at many machine learning methods, including neural networks and clustering methodologies, to find out-of-the-ordinary actions that could indicate cloud security issues. They developed novel strategies for identifying and addressing security vulnerabilities in cloud systems by using state-of-the-art machine learning techniques. Problems in efficiently detecting anomalies in cloud systems that generate massive volumes of data are addressed in this work. A potential drawback is the high likelihood of false positive rates, which happen when the system incorrectly labels normal behavior as aberrant. As a consequence, security personnel have more work to complete and get unnecessary alerts.

Building cloud-based intrusion detection systems using deep learning algorithms is the primary emphasis of Chang and Yan's research [13]. In order to detect and

respond to security breaches in real time, their research centers on the usage of complex neural network designs. Their efforts have aided in the development of sophisticated deep learning models that can detect and reduce security risk in cloud computing systems. They emphasize the significance of using deep learning techniques to strengthen cloud security systems. One of the many drawbacks of deep learning models is the enormous amount of labelled data that is required for training. The models' utility is diminished since this data could be hard to get for some types of cyber attacks.

When it comes to cloud risk assessment, Xiong and Li [14] have both done extensive research on using reinforcement learning techniques. The researchers are looking at how reinforcement learning algorithms can make decision-making and security policies more efficient in dynamic cloud environments. In order to tackle cloud security concerns in a flexible way, they have developed new frameworks that leverage reinforcement learning. Their research delves into the topic of autonomous and flexible security systems, which are necessary to combat the dynamic threats found in cloud environments. One limitation is the computational complexity of techniques for reinforcement learning. These methods could be too demanding on system resources to be practical for large-scale cloud risk assessment in real-time. The high initial investment required to put them into action is another drawback.

Ko and Choo [15] have written extensively on the topic of cloud security and machine learning coming together, with a focus on threat intelligence and risk assessment. They tackle the challenges of developing prediction models to detect new dangers and analyzing large-scale information in their work. Important new information about the use of machine learning algorithms for cloud-based threat intelligence and risk assessment has emerged from their research. They stress the need of using machine learning algorithms to detect and respond appropriately to evolving cyber threats. Due to reliance on historical data, which could not accurately portray the dynamic nature of cyber threats, there are holes in risk assessment and detection. Among the constraints, this is one.

Throughout their research, Qiao and Lu [16] are focusing on building machine learning-based frameworks for cloud computing risk assessment and continuous monitoring. They highlight the need of real-time analytics and automated response systems as a way to minimize security problems. Their efforts have aided in the development of cloud-based risk assessment frameworks that are adaptable and scalable. They stress the need of proactive risk management strategies that use machine learning-based methodologies to identify and counteract security threats as they occur. One of the constraints is the potential for the model to wander. Model drift occurs when the underlying data distribution changes, leading to a gradual decline in the performance of machine learning models. This occurrence need regular retraining and further upgrades.

Cloud risk assessment using adversarial machine learning techniques has been studied by Zhou and Fu [17]. Their work focuses on identifying potential vulnerabilities in machine learning algorithms and studying how attackers can use them to penetrate cloud systems undetected. Their research highlights the need of high-quality security in cloud environments against hostile attacks. The importance of building secure and resilient machine learning models that can resist advanced attacks from enemies is heavily emphasized. One of the drawbacks of machine learning is that it can be easily attacked by adversaries. This kind of assault can undermine the effectiveness of risk assessment systems since the perpetrators can alter the supplied data to evade detection.

Aledhari and Yousif's [18] research is on cloud-based anomaly detection and threat assessment using machine learning. They are looking at various machine learning techniques to better detect and respond to security threats. Their study's findings have aided efforts to develop and deploy advanced anomaly detection algorithms for use in cloud environments. The need of using machine learning technologies to identify and address security vulnerabilities in varied and dynamic cloud systems is highlighted by them. One limitation is the potential for issues related to data privacy. Ethical and legal concerns arise, in particular, when sensitive data is used to train machine learning algorithms.

The authors Xu and Xue [19] look at how multi-cloud security management and risk assessment could be tackled using machine learning techniques. In order to identify and prioritize security risks, they are investigating the challenges of combining data from different cloud platforms and creating predictive models. The study's findings have been crucial in informing the creation of machine learning-based approaches to risk assessment in multi-cloud systems. They emphasize the requirement of using thorough risk management strategies that include the unique characteristics of different kinds of multi-cloud deployments. One drawback is that different cloud platforms aren't always compatible with one other, which could make it harder to combine and evaluate all the data needed for a comprehensive risk assessment across different clouds.

Anomaly detection and threat intelligence fusion using machine learning in cloud systems is the subject of Reeta and Gautam's research [20]. They look into methods to effectively detect and reduce security vulnerabilities by using advanced analytics techniques. Their efforts have improved the state of the art in cloud security risk assessment by integrating data from several sources and using machine learning methods. Their main point on how to strengthen cloud security is the need of proactive threat

detection and response systems. Model bias is one potential drawback; it happens when machine learning algorithms provide biased results due to irregular training data or built-in biases in the algorithm's architecture. Erroneous risk evaluations could emerge from this model.

The authors John et. al., [21], whose research work is covered in this article, provide an overview of the various machine learning techniques that are used for evaluating security concerns in cloud systems. Finding vulnerabilities, anomalies, and cyber threat predictions is the goal of this research, which explores supervised, unsupervised, and semi-supervised learning techniques. This paper aims to analyze several machine learning algorithms and discuss their pros, cons, and practical considerations for solving cloud security issues. The most up-to-date machine learning techniques for evaluating cloud security risks may have slipped the survey's radar. One of the research work's limitations is this. Also, in cloud environments that mimic the real world of cloud computing, the quality and representativeness of the training data could be limited, which might affect the performance of machine learning models.

In their joint research work, Lee et. al., [22] present an intrusion detection system (IDS) based on deep learning that is tailored to environments using cloud computing. Since it employs deep neural networks, the intrusion detection system (IDS) can decipher intricate patterns and behaviors that point to cyberattacks and threats in cloud infrastructure. Included in the article are experimental results that show how effective the proposed method is at accurately and efficiently detecting both known and zero-day attacks with minimal false-positive rates. A research paper is used to present the findings. The deep learning models in this research work need a large amount of labelled training data, which may be challenging to get, especially for unusual or novel cyber threats. This is one of the research work's limitations. In addition, scalability and performance could be impacted by the computational overhead of training and deploying deep learning models in cloud systems.

For the objective of enhancing cloud systems' capacity to identify cyber threats, their research work investigates ensemble learning methodologies. The authors of this research are Wang et. al., [23]. The system takes use of diversity and resilience by combining many basis classifiers into an ensemble model, which improves detection accuracy and makes it more resistant to evasion attacks. Decision trees, SVMs, and neural networks are examples of these foundational classifiers. In this work, we provide an assessment of ensemble learning algorithms' performance based on real-world cloud datasets and benchmarks. Ensemble learning methods may be unsuitable for cloud environments with limited resources due to their potential to increase computational complexity and resource requirements. The research work that has been done has this as one of its limitations. The upkeep and revision of ensemble models that account for evolving threat environments and system dynamics could also provide logistical challenges.

For the exchange of threat intelligence in cloud security while safeguarding user privacy, Kim et. al., [24] describe a federated learning technique in their research work. Collaborating on threat detection without revealing sensitive data is now feasible using federated learning. This is achieved by integrating local updates from scattered edge devices or cloud nodes and decentralizing model training. Within the framework of cloud security, this research aims to examine the privacy assurances and performance of federated learning. Furthermore, an implementation prototype is shown. The federated learning system relies on safe aggregation techniques and effective communication protocols, which may both add complexity and overhead; this is one of the limitations of this research work. In addition, it is critical to ensure that the training data across all federated nodes is varied and representative of the population in order to maintain detection accuracy and generalization.

Transfer learning methodologies for vulnerability assessment in cloud architecture are the subject of research work by authors Liu et. al., [25]. By drawing on knowledge from previously trained models or domains that are comparable, transfer learning enables efficient adaptation to new cloud environments and new hazards. The research suggests a transfer learning framework work as a possible implementation for discovering vulnerabilities in cloud environments, software dependencies, and network work protocols. One of the research work's limitations is that transfer learning necessitates altering or selecting source models or datasets with great care to fit the destination domain. Very varied and ever-changing cloud environments can make this a challenge. In addition, the generalizability and robustness of vulnerability assessment models may be affected by differences in the transferability of obtained information and features across different cloud platforms and architectures.

In order to avoid detection and beat security measures in cloud environments, cybercriminals use adversarial machine learning techniques, which their research work examines. This work was authored by Zhang et. al., [26]. By capitalizing on weaknesses and vulnerabilities in machine learning models, threat actors provide malicious instances or disturbances. The purpose of these instances or disturbances is to influence the predictions of the model while evading detection. This article examines several adversarial attacks and countermeasures, shedding insight on the usefulness of adversarial resilience in cloud security. Certain limitations exist in this research work since opponents are always evolving and improving their techniques. Combating complex evasion attacks is particularly difficult with adversarial machine learning, which is one of these limitations. In addition, it may be

necessary to spend more computational resources and incur substantial cost in order to implement defensive measures that are both robust and resilient against adversarial attacks.

In order to assess the risks associated with cloud computing environments, Wilson et. al., [27] use probabilistic graphical models like Bayesian networks and Markov random fields. Full probability inference and risk analysis are both made feasible by probabilistic graphical models. All interdependencies and interactions between risk factors are included in these models. with order to aid with cloud security decision-making and risk management techniques, this research aims to provide a framework for developing and reasoning with graphical models. Scaling to large-scale cloud settings and complex interdependencies among risk factors may be challenging for the probabilistic graphical models. The research work that has been done has this as one of its limitations. Furthermore, both objective and subjective assessments may be required to elicit and quantify appropriate probabilistic assessments of risks and uncertainties.

Cloud log and telemetry anomaly detection is the goal of the research work of Andrew et. al., [28], who provide an unsupervised learning approach. Through the modeling of regular behavior patterns and the detection of deviations or outliers, unsupervised learning algorithms are able to identify aberrant acts that may indicate security breaches or system malfunctions. Applying several anomaly detection approaches to real-world cloud datasets and benchmarks is the goal of this work. It is possible that unsupervised anomaly detection algorithms may overlook significant security threats or have a high false-positive rate since they can't tell the difference between harmless abnormalities. The research work that has been done has this as one of its limitations. In addition, having domain expertise and contextual understanding of cloud settings is essential for analyzing and investigating observed irregularities.

Chen et. al., [29] are the authors of their research work, which focuses on building understandable machine learning models for auditing and analyzing security concerns in cloud systems. Security auditing processes are made more trustworthy, accountable, and understandable with the use of explainable AI methods. To do this, these methods provide clear and interpretable explanations of model predictions and judgments. The basic goal of this research is to provide a framework for building and assessing cloud security explainable machine learning models. There is a potential impact on the accuracy and utility of explainable machine learning models from the process of balancing the model's complexity with its interpretability, which is one of the limitations of this research work. Also, making sure that model explanations are strong and fair across all cloud settings and stakeholders is crucial for user acceptance and compliance.

As a means of adaptive risk management in cloud security, dynamic Bayesian networks (DBNs) are proposed by researchers Rachel Smith and Eric Wang [30]. As threat landscapes and system conditions change, DBNs may provide dynamic risk assessment and decision-making. This is achieved by creating risk factors and modeling temporal relationships. The paper presents a system for designing and updating distributed block networks (DBNs) that may help with real-time risk assessment and mitigation in cloud environments. Parameter estimation and inference in dynamic Bayesian networks need accurate and up-to-date data, which is one of the limitations of this research work. It could be challenging to gather this kind of data in distributed and dynamic cloud environments. Scalability and real-time performance in large-scale cloud deployments may be hindered by the computational complexity of DBNs.

In their research work, Wong et. al., [31] look at how to react to cybersecurity risks in cloud environments using strong deep reinforcement learning (DRL) algorithms. By exposing agents to simulated attack scenarios and helping them develop appropriate response strategies, DRL offers autonomous and adaptive threat mitigation operations. Training in this area is carried out via contact. The research presents a technique for constructing cybersecurity protective mechanisms based on DRL and evaluates their performance against different attack vectors. In order for deep reinforcement learning models to develop effective threat response techniques, a lot of time and computer power may be required during training. The research work in this area has several limitations. For real-world cybersecurity scenarios, it is also important to ensure the reliability and security of autonomous decision-making to prevent vulnerabilities and unexpected consequences.

According to the research work of Johnson et. al., [32], cloud service reliability is assessed using probabilistic graphical models that take into account a number of risk factors and trust indicators. Assessing the reliability and security posture of a service may be done with the use of probabilistic graphical models, which enable probabilistic inference and decision-making. In order to construct these models, the probabilistic linkages among service characteristics, security measures, and performance data from the past are modelled. To make trustworthiness assessment in cloud settings easier, the research suggests a method for creating and reasoning using graphical models. In the abstract, the technique is laid forth. Probabilistic graphical models may struggle to capture the intricate and dynamic relationships between trust factors and cloud service characteristics, which is one of the research work's limitations. In addition, systems that are open and collaborative may be required to acquire reliable probabilistic assessments of trustworthiness from several stakeholders and incorporate them into the process.

Integrating anomaly detection based on machine learning with detection based on signatures, this research work develops a hybrid intrusion detection system (IDS). Written by Jessica et. al., [33] is this piece of work. Through the combination of both systems' capabilities, a hybrid intrusion detection system (IDS) improves detection accuracy and resistance against both known and new cyber threats. This research presents a framework for a unified detection system that combines machine learning models with signature-based heuristics. Furthermore, the study presents cloud-based intrusion detection scenarios to assess the efficacy of this system. A hybrid intrusion detection system (IDS) requires meticulous calibration and integration of machine learning algorithms with signature-based criteria for design and optimization. To find a happy medium between computational overhead, detection accuracy, and false-positive rates, this is essential. The research work in this area has several limitations. In addition, keeping up with the ever-expanding variety of threat settings could make it challenging to maintain and update signature databases.

Safeguarding sensitive data while monitoring security events and abnormalities in cloud settings is the focus of their research work on privacy-preserving machine learning techniques. David et. al., wrote this part of work [34]. Machine learning that protects users' privacy enables collaborative threat detection and response without exposing sensitive data by using cryptographic protocols, homomorphic encryption, or differential privacy approaches. A privacy-preserving machine learning model deployment architecture is introduced in this research, tailored to cloud security monitoring. We also test these models' privacy assurances and performance. Concerns that privacy-preserving machine learning techniques may lead to higher computing costs and communication delays are one of the research work's limitations. The reason for this is because secure computing protocols and encryption are used. Furthermore, to ensure the precision and utility of machine learning models while simultaneously safeguarding data privacy and secrecy, rigorous validation and design is necessary.

The research work that studies semi-supervised learning techniques to improve threat intelligence and situational awareness in cloud security was written by Sarah et. al., [35]. Using a combination of labelled and unlabeled data, semi-supervised learning approaches make threat detection models more resilient and able to generalize. An important improvement over more conventional approaches to learning is the fact that this lessens the need for massive labelled datasets. This paper presents a method for training and deploying semi-supervised learning models to identify new potential threats and anomalies in cloud systems. Unlabeled data may need meticulous selection and preparation for semi-supervised learning to ensure it is high-quality and representative of the population; this is one of the limitations of this research work. Finding a happy medium between the model's complexity, performance, and labeling efforts is another challenge during real deployment and operation of semi-supervised learning.

For the objective of enhancing risk management techniques in cloud security, authors Brown et. al., [36] conducted research on evolutionary computing approaches such genetic algorithms and evolutionary strategies. When it comes to optimizing risk in light of changing threat environments and company objectives, evolutionary computing provides dynamic and adaptable solutions. This is achieved by gradually refining approaches to risk reduction and decisions about the distribution of available resources. We provide a paradigm for using evolutionary computing to assess and optimize cloud security risks as part of this research. When dealing with complex and high-dimensional optimization issues, evolutionary computing methods may need a lot of processing power and iterations before they converge to optimal solutions. The research work that has been done has this as one of its limitations. It is also critical to provide an effective and scalable implementation of evolutionary algorithms to guarantee their practical deployment and operation in large-scale cloud environments.

In their research work, Smith et. al., [37] look at multi-objective optimization techniques for effectively allocating security resources in cloud computing environments. Through the use of multi-objective optimization techniques, it is feasible to discover Pareto-optimal solutions that reconcile conflicting objectives. Several competing goals are considered by these algorithms, including minimizing security concerns, cutting expenses, and optimizing performance. In order to help with the creation and solving of multi-objective optimization problems related to cloud security resource allocation, this research aims to provide a methodology. Finding optimal solutions may be a challenge when dealing with decision spaces that are complex and high-dimensional, as might be the case with multi-objective optimization. The research work in this area has several limitations. To further aid in decision-making and risk management, effective visualization and decision-support tools are required for understanding and navigating the Pareto-optimal bounds.

Daniel et. al., evaluated and understood the causal links that exist between security incidents, vulnerabilities, and system settings in cloud environments using techniques of causal inference in their research work [38]. Using causal inference, which entails determining the origins and routes of security breaches or occurrences, preventive steps may be taken to reduce risks and fix them. The paper lays forth a framework for doing causal analysis and inference in the context of managing and responding to incidents involving cloud security. It may be challenging to use causal inference methods in complex and dynamic cloud environments since doing so

may need significant assumptions and domain knowledge. The research work in this area has several limitations. Furthermore, when used in real-world contexts, addressing biases and confounding variables in causal inference poses challenges to both technique and computers.

3. RISK ASSESMENT MODEL

In order to find, analyze, and mitigate any security vulnerabilities that can exist within cloud systems, a machine learning-based cloud risk assessment must be conducted. Fig. 1 shows the risk assessment steps.

Figure 1. Risk assessment steps

To do a risk assessment using ML techniques, one must adhere to the following steps:

- **Risk Identification**: To begin the process of risk assessment, relevant data from various sources inside the cloud environment, including logs, network traffic, system settings, and user behaviors, must be obtained. The collected data could include details on past incident reports, resource use, access control rules, and security issues. After data acquisition, preprocessing is required to clean the data and prepare it for analysis. Data cleansing, data normalization, feature extraction, and dimensionality reduction are all possible components of this. Data quality and database preparation are two of several critical factors that can significantly affect the efficacy and precision of machine learning models. By taking the required measures to guarantee that the data is clean, consistent, and formatted correctly, firms can enhance the reliability of risk assessments.
- **Risk Analysis**: In order to foresee future security threats and vulnerabilities inside a cloud environment, feature selection is the process of identifying which characteristics or variables are most crucial. To achieve this goal, it can be required to analyze the interrelationships of different components, determine which characteristics are relevant, and choose important features by drawing on domain knowledge. Machine learning models can be enhanced by feature engineering, which comprises enhancing and creating new features from raw data. The development of reliable risk assessment models relies heavily on engineering and feature selection. Focusing on the most relevant properties and building meaningful representations of the stored data can help firms increase the prediction ability of their machine learning models.
- **Risk Assessment**: Model selection refers to the process of choosing the machine learning algorithm or combination of algorithms that is best suited for the present risk assessment task. This could need doing tests with different algorithms, tweaking their hyperparameters, and then evaluating their performance using cross-validation techniques. Training the chosen model on the labeled training data is the next step after model selection for understanding the relationships between input features and security results. Careful model selection and training are necessary to build efficient risk assessment models. Businesses can improve their machine learning models' accuracy, robustness, and generalizability by selecting the right algorithms and adjusting their parameters appropriately.
- **Risk Evaluation**: Model evaluation is the process of evaluating the performance of a trained machine learning model using validation datasets or cross-validation processes. This procedure lends itself to a variety of evaluation tools, such as recall, accuracy, precision, F1-score, ROC curve, and area under the curve (AUC). To ensure the model successfully generalizes to novel, previously unseen scenarios, it is necessary to assess its performance on previously unseen data. A company's risk assessment models' efficacy in practical situations can be better understood with the help of model validation and evaluation. Once organizations have a good grasp of their models' strengths and weaknesses, they can work to enhance them, eliminate biases and limitations, and make informed decisions on deploying them in production settings.

A. A Cloud Risk Assessment Method Using Machine Learning

1) Decision Trees: Decision trees are a kind of supervised learning method that is used for classification and regression issues. They used a recursive technique to

partition the dataset into subsets according to the most significant property that could efficiently group comparable data sets. Until a certain condition is met, such a maximum tree depth or a specific quantity of samples per leaf node, the method will keep running until it finishes. To find out how risky cloud computing is, decision trees can be used to examine a lot of different things, such system settings, user behavior, and network traffic patterns. As an example, decision trees can classify security occurrences into many groups based on criteria retrieved from security logs or statistics about network traffic. They can also determine which critical risk factors are the main causes of security incidents in cloud environments.

2) Random Forest Classifier: A collection of decision trees is what makes up the Random Forest technique, an ensemble learning methodology. Every tree in the forest has its own training procedure, using a randomly selected portion of the training data and attributes. A final forecast is formed by combining the findings of many trees in the prediction process. A majority vote (for classification) or an average (for regression) is a typical way to do this. Using Random Forest for cloud risk assessment can help improve prediction models' accuracy and resilience. Because it integrates the forecasts of many decision trees, Random Forest effectively handles noisy data while also reducing the risk of overfitting. Among the many potential uses for this technology are the categorization of security events, the estimation of the likelihood of security breaches, and the discovery of key risk factors that lead to incidents involving cloud security.

3) K- Nearest Neighbour (K-NN or K):* Classification and regression analyses make use of K-Nearest Neighbors, an instance-based learning approach that is both simple and effective. For classification, KNN finds the neighbors that are geographically closest to the training dataset and uses a distance measure like Euclidean distance to identify which K neighbors are closest. For regression, it finds the average value among these neighbors. Every time fresh information is given, this is executed. One possible use of KNN in cloud risk assessment is the classification of security events and the identification of anomalous behavior by way of comparisons to surrounding data points. One way to do this is to look at how similar they are. For example, KNN can classify user behaviors or network traffic patterns as normal or suspicious based on their similarity to known dangerous patterns. When the data under consideration is well-structured and shows signs of local patterns or clusters, the KNN algorithm shows better performance.

4. EXPERIMENTATION & RESULTS

To determine the final risk factors, we conducted a survey. In this, we ask participants to categorise risk variables into three different groups according to their likelihood of occurring and impact on cloud computing. Outcome of these classes are as provided: Insignificant, significant, and impartial. In this survey, 35 foreign experts from different countries responded, and all of them agreed that the previously recommended factors are significant, indicating that they have huge influence across cloud-based software. Subsequently, we assigned each risk factor a numerical range of values. Finally, we built expert methods and standards, utilising statistical tools to generate data based on those guidelines. The dataset had approximately 1940 occurrences and contained 18 input attributes. 18 qualities are labelled. were IDD, DI, BC&SA, RC, I&P, TPM, IAP, DL, R, DL&IS, SLA, RE, ShE, A&AC, DS, DB, DI and VV. Table I lists the risk factors along with the ranges of values for each. To assess and experiment with algorithms, we used percentage split. The dataset is randomly dividing testing and training data with percentage split as follows:

- 50–50% (A)
- 65–35% (B)
- 85–15% (C)
- 96–4% (D)

TABLE I. RISK FACTORS WITH CORRESPONDING ASSOCIATED RANGE VALUES

Risk factor	Range value	Risk error	Range value
DT	0-3	R	1-3
IDD	1-3	RE	0-2
RC	0-1	SLA	0-3
BC & SA	1-3	A&AC	0-3
TPM	0-2	ShE	1-3
I&P	0-1	DB	0-2
DL	0-3	DS	0-1
LAP	0-1	VV	1-3
DL & IS	0-3	DI	0-2

With every algorithm, Table II briefs best test outcomes of (RMSE) from each data percentage. RMSE is shown in Fig. 2 for the percentages of the whole dataset and shows that, for the best results, an additional training data percentage of roughly 4% testing and 96% training is generated, indicating greater learning.

TABLE II. EACH DATASET'S RMSE PERFORMANCE

Algorithms	A	B	C	D
DT	0.0023	0.0022	0.0021	0.0020
RC	0.0048	0.0049	0.0040	0.0037
KNN	0.0048	0.0049	0.0040	0.0037

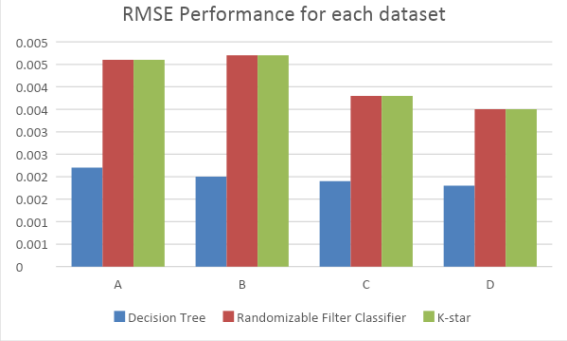


Figure 2. RMSE for test data using different datasets

5. CONCLUSION

Big data systems have emerged as one of the major engines of innovation that provide a route for information management, as the use of data grows over time. Big data resolutions are greatly regulated by cloud domain, which provides big data systems with altered domains. While big data in cloud computing is a sturdy and powerful system that aids in the development and enhancement of organisations as well as further study, there are some conjectures surrounding risk assessment that need to be discussed and really investigated. More effort needs to be placed into creating and developing risk assessment mechanisms related to security in the big data cloud computing space. Additional should be added later, but as soon as possible, to address the security concerns raised by the risk assessment. The main goals of this experimental challenge are to reduce characteristics, locate the best precision on the testing data set, and identify the best current schemes that can be used to our dataset. We examined the behaviour of several machine learning algorithms to illustrate the cloud computing risk element. This research examines the impact of subsets of information testing and training by arbitrarily dividing a subset of the dataset into four different groups. The experimental results show that the Decision Tree Results from using a classifier approach outperform those from using any other partitioning method in the cloud computing environment, and the optimal partitioning is 96% - 4%.

REFERENCES

[1] A. Jamil and Z. M. Yusof, "Information security governance framework of Malaysia public sector," Asia-Pacific Journal of Information Technology Multimedia, vol. 7, no. 2, pp. 85-98, 2018.

[2] A. A. Ahmed and M. K. Hasan, "Design and Implementation of Side Channel Attack Based on Deep Learning LSTM," in 2023 IEEE Region 10 Symposium (TENSYMP), Canberra, Australia, 2023, pp. 1-6: IEEE.

[3] N. Musa, "A conceptual framework of IT security governance and internal controls," in 2018 Cyber Resilience Conference (CRC), Putrajaya, Malaysia, 2018, pp. 1-4: IEEE.

[4] A. A. Ahmed, M. K. Hasan, N. S. Nafi, A. H. Aman, S. Islam, and S. A. Fadhil, "Design of Lightweight Cryptography based Deep Learning Model for Side Channel Attacks," in 33rd International Telecommunication Networks and Applications Conference, Melbourne, Australia, 2023, pp. 325-328: IEEE.

[5] A. A. Ahmed, M. K. Hasan, N. S. Nafi, A. H. Aman, S. Islam, and M. S. Nahi, "Optimization Technique for Deep Learning Methodology on Power Side Channel Attacks," in 2023 33rd International Telecommunication Networks and Applications Conference, Melbourne, Australia, 2023, pp. 80-83: IEEE.

[6] A. A. Ahmed et al., "Detection of Crucial Power Side Channel Data Leakage in Neural Networks," in 2023 33rd International Telecommunication Networks and Applications Conference, Melbourne, Australia, 2023, pp. 57-62: IEEE.

[7] U. A. Butt et al., "A review of machine learning algorithms for cloud computing security," Electronics, vol. 9, no. 9, p. 1379, 2020.

[8] A. B. Nassif, M. A. Talib, Q. Nasir, H. Albadani, and F. M. Dakalbab, "Machine learning for cloud security: a systematic review," IEEE Access, vol. 9, pp. 20717-20735, 2021.

[9] Reddy P, Adetuwo Y, Jakkani AK. implementation of machine learning techniques for cloud security in detection of ddos attacks.

[10] S. Tuli, S. Tuli, R. Tuli, and S. S. Gill, "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing," Internet of things, vol. 11, p. 100222, 2020.

[11] Jakkani AK, Reddy P, Jhurani J. Design of a Novel Deep Learning Methodology for IOT Botnet based Attack Detection. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(9):4922-7.

[12] Q. McGrath, A. R. Hevner, and G.-J. de Vreede, "Managing Ethical Risks of Artificial Intelligence in Business Applications," Authorea Preprints, 2024.

[13] V. Chang et al., "A survey on intrusion detection systems for fog and cloud computing," Future Internet, vol. 14, no. 3, p. 89, 2022.

[14] S. Liu, J. Wu, Z. Lu, and H. Xiong, "Vmras: A novel virtual machine risk assessment scheme in the cloud environment," in 2013 IEEE International Conference on Services Computing, Santa Clara, CA, USA, 2013, pp. 384-391: IEEE.

[15] S. Iqbal et al., "On cloud security attacks: A taxonomy and intrusion detection and prevention as a service," Journal of Network Computer Applications, vol. 74, pp. 98-120, 2016.

[16] W. Qiao et al., "A Novel Method for Resource Efficient Security Service Chain Embedding Oriented to Cloud Datacenter Networks," IEEE Access, vol. 9, pp. 77307-77324, 2021.

[17] Z. Fu, L. Xia, X. Sun, A. X. Liu, and G. Xie, "Semantic-aware searching over encrypted data for cloud computing," IEEE Transactions on Information Forensics Security, vol. 13, no. 9, pp. 2359-2371, 2018.

[18] A. J. Ouda, A. N. Yousif, A. S. Hasan, H. M. Ibrahim, and M. A. Shyaa, "The impact of cloud computing on network security and the risk for organization behaviors," Webology, vol. 19, no. 1, pp. 195-206, 2022.

[19] G. Xu, S. Xu, J. Ma, J. Ning, and X. Huang, "An Adaptively Secure and Efficient Data Sharing System for Dynamic User Groups in Cloud," IEEE Transactions on Information Forensics Security, vol. 18, pp. 5171 - 5185, 2023.

[20] D. Gautam and R. Shivhare, "Cloud security aspects using homomorphic encryption: a review," Research Journal of Engineering Technology and Medical Sciences, vol. 5, no. 04, 2022.

[21] John, Jomina, and Jasmine Norman. "Major vulnerabilities and their prevention methods in cloud computing." Advances in Big Data and Cloud Computing: Proceedings of ICBDCC18. Springer Singapore, 2019.

[22] Lee, Sangwoong, et al. "Security enhancement through comparison of domestic and overseas cloud security policies." Proceedings of the Korean Institute of Information and

Commucation Sciences Conference. The Korea Institute of Information and Commucation Engineering, 2021.

[23] Li, Xiang, et al. "Enhancing cloud-based IoT security through trustworthy cloud service: An integration of security and reputation approach." IEEE access 7 (2019): 9368-9383.

[24] Kim J, Kim E, Yang J, Jeong J, Kim H, Hyun S, Yang H, Oh J, Kim Y, Hares S, Dunbar L. Ibcs: Intent-based cloud services for security applications. IEEE Communications Magazine. 2020 Apr 20;58(4):45-51.

[25] Liu, Y., Zhou, T., Yue, Z., Liu, W., Han, Y., Li, Q. and Yang, X., 2021. Secure and efficient online fingerprint authentication scheme based on cloud computing. IEEE Transactions on Cloud Computing, 11(1), pp.564-578.

[26] Zhang, Yinghui, et al. "Attribute-based encryption for cloud computing access control: A survey." ACM Computing Surveys (CSUR) 53.4 (2020): 1-41.

[27] Wilson S, Choppali U. Collaborative edge computing for smart villages [energy and security]. IEEE Consumer Electronics Magazine. 2021 Jan 22;10(3):68-71.

[28] Andrew AU, Joshua JT. Design Framework of Cyber Security Solutions to Threats and Attacks on Critical Infrastructure of Electricity Power Systems of Nigeria Companies.

[29] Yang Y, Chen Y, Chen F. A compressive integrity auditing protocol for secure cloud storage. IEEE/ACM Transactions on Networking. 2021 Feb 15;29(3):1197-209.

[30] Smith, Adam. Examining the Security and Privacy Barriers of Third-Party Public Cloud Services A Case Study of Diverse Norwegian Municipalities. MS thesis. University of Agder, 2023.

[31] Wong DS, Li Y. Security of federated learning for cloud‐edge intelligence collaborative computing. International Journal of Intelligent Systems. 2022 Nov;37(11):9290-308.

[32] Johnson M. et. al., Design and implementation of a research and education cybersecurity operations center. Cybersecurity and Secure Information Systems: Challenges and Solutions in Smart Environments. 2019:287-310.

[33] Jessica JY, Khan SU. QuantCloud: big data infrastructure for quantitative finance on the cloud. IEEE Transactions on Big Data. 2017 Jan 9;4(3):368-80.

[34] David DS, Anam M, Kaliappan C, Selvi S, Sharma DK, Dadheech P, Sengan S. Cloud Security Service for Identifying Unauthorized User Behaviour. Computers, Materials & Continua. 2022 Feb 1;70(2).

[35] Sarah, and Jianbiao Zhang. "Comparative analysis of cloud security classifications, taxonomies, and ontologies." In Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science, pp. 666-672. 2019.

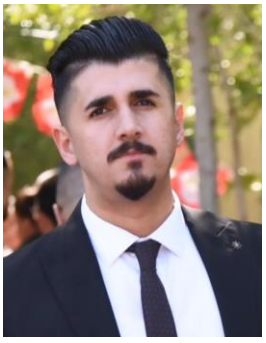
[36] Brown, Adam J., et al. "Cloud forecasting: Legal visibility issues in saturated environments." Computer Law & Security Review 34.6 (2018): 1278-1290.

[37] Smith E. Ramgovind S, Eloff MM, The management of security in cloud computing. In2010 Information Security for South Africa 2010 Aug 2 (pp. 1-7). IEEE.

[38] Daniel WK. Challenges on privacy and reliability in cloud computing security. In2014 international conference on information science, electronics and electrical engineering 2014 Apr 26 (Vol. 2, pp. 1181-1187). IEEE.

Saif Aamer Fadhil received the M.Sc. degree in Computer science, 2012, Iraq Commission for Computers & Informatics, Informatics Institute for Postgraduate Studies, Iraq and BSc. degree in Software Engineering, 2007, Al- Mansour University College, Iraq. TOT, Jordan Summit House for Training and Development JSH, 2013. CCNA1 certification from CISCO. MikroTik ROS Principles certification. Areas of interest, Computer and Data Security and Network Security.

Lubna Emad Kadhim received the M.Sc. degree in Computer science, 2008, University of Technology, Department of Computer Science, Iraq and BSc. degree in Computer science, 2006, University of Technology, Department of Computer Science, Iraq. TOT, Jordan Summit House for Training and Development JSH, 2013. CCNA1 certification from CISCO and MikroTik ROS Principles certification.



Mahmood Ahmed Hamdi received the M.Sc. degree in (Computer and Communication) 2024 Islamic University of Lebanon Faculty of Engineering Department Of Computer And Communications Engineering, received the BSc. Computer Technology Engineering ,2017 Imam Al-Kadhum College (IKC), Baghdad, Iraq.
Areas Of Interest: Network security, cloud computing system, UNIX systems

Amjed A. Ahmed received the B.Sc. degree in computer science from the University of Baghdad, and the M.Sc. degree in computer science from Binary University, Kuala Lumpur, Malaysia, in 2012. He is currently pursuing the Ph.D. degree with University Kebangsaan Malaysia, Malaysia, with a focus on artificial intelligence. From July 2013 to July 2022, he was a Lecturer with the Imam Al-Kadhum College, Baghdad, Iraq.

Mohammad K. Hasan received the Doctor of Philosophy (Ph.D.) degree in electrical and communication engineering from the Faculty of Engineering, International Islamic University, Malaysia, in 2016. He is currently with the Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), as a Senior Lecturer.

Shayla Islam (Senior Member, IEEE) received the B.Sc. degree in computer science and engineering from International Islamic University Chittagong, Bangladesh, and the M.Sc. and Ph.D. degrees in engineering from the Electrical and Computer Engineering (ECE) Department, International Islamic University Malaysia (IIUM), in 2012 and 2016, respectively. She is currently an Assistant Professor with UCSI University, Malaysia. She has awarded a silver medal for her research work at International Islamic University Malaysia. Her research interest: Wireless communications, Mobile networks, and Cyber security.

Azana H. M. Aman received the B.Eng., M.Sc., and Ph.D. degrees in computer and information engineering from International Islamic University Malaysia, Malaysia. She is currently working as Senior Lecturer at the Research Center for Cyber Security, Faculty of Information Science and Technology (FTSM), The National University of Malaysia, Malaysia. Her research interests include computer system and networking, computer information and network security, the Internet of Things (IoT), cloud computing, and big data.