Int. J. Adv. Eng. Comput. Sci. 2026; 2(4); 81-92

International Journal of Advances in Engineering and Computer Science

E-ISSN: 3121-6382

Volume 2, Issue 4, October-December, 2026
(An Academians Publishers of Research & Academic Resources)



Research Article

Development of an Intelligent Security Risk Assessment Model in Cloud Computing Environments Using Supervised Machine Learning Algorithms

Mohammad Rasoul Popal[1]*, Habibullah Andishah[1], and Abdul Aziz Rahimi[2]

[1] Department of Software Engineering, Faculty of Computer Science, Hariwa University, Herat, Afghanistan
[2] Department of Database, Faculty of Computer Science, Hariwa University, Herat, Afghanistan

Abstract: Cloud computing has emerged as the key infrastructure for digital transformation, providing scalable and flexible access to computing resources, storage, and software services. Despite its major advantages such as scalability, flexibility, lower infrastructure costs, and on-demand access to computing resources, widespread adoption of cloud computing has introduced significant security concerns regarding the confidentiality, integrity, and availability of information. As a result, the development of intelligent approaches to assigning security risk has become an important requirement for cloud computing environments. The objective of this study is to develop an intelligent model for the assessment of security risk in cloud computing environments, using supervised machine learning algorithms. First, a systematic literature review was conducted to identify and classify the major factors affecting safety risk. The identified factors were evaluated and prepared for model development, resulting in a set of 18 safety risk factors. Three supervised machine learning algorithms—Decision Tree, K-Star (K*), and Random Filtered Classifier—were then evaluated and compared to predict the level of safety risk. To identify relevant variables and to support the model development, three methods for selecting features, namely Best First Search, Ranker, and Random Search, were applied using the WEKA and Orange Data Mining environments. The performance of the model was evaluated using the Root Mean Square Error (RMSE) and correlation coefficient (R) under different training–testing data-split scenarios. The results showed that the 95% training and 5% testing split produced the lowest RMSE among the evaluated data-split scenarios. The Decision Tree algorithm achieved the lowest RMSE across the evaluated scenarios, reaching 0.0018 under the 95% training and 5% testing split. The Random Filtered Classifier and K-Star algorithms produced identical RMSE values across the evaluated scenarios. These findings demonstrate the applicability of supervised machine learning algorithms to intelligent security risk assessment and provide a basis for developing data-driven approaches to security risk analysis in cloud computing environments.

Keywords: *Cloud Computing; Cloud Security; Security Risk Assessment; Supervised Machine Learning; Feature Selection; Decision Tree; K-Star; Risk Prediction.*

1. Introduction

One of the most significant advantages in information technology over the past 20 years is how it has transformed the delivery of computing resources, data storage, and software as a service (Buyya et al., 2009), referred to today as cloud computing. Government agencies, private companies, educational institutions, and digital businesses are increasingly adopting the use of clouds for operational efficiency, cost savings, and organizational agility.

Although these advantages have been achieved, the explosive growth of cloud computing has also posed a set of new and complex security challenges to cloud environments, such as unauthorized access, data loss or disclosure, distributed denial-of-service (DDoS) attacks, malware, identity-related threats, confidentiality breaches, and service-availability disruptions (Miller, 2008). As a result, information security and security risk assessment have become important issues for cloud service providers and organizations that rely on cloud services.

Received: 10 August 2026; Accepted: 20 September 2026; Published: 10 October 2026

© 2026 by the authors. This work is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Cloud computing is characterized by a number of fundamental properties, including that computing resources, storage, and software applications are delivered as network-based services, which can be accessed without large capital investments in local hardware and software infrastructure (Savu, 2011). Examples of this model are cloud services offered by major technology companies that allow users to access applications and data without needing to install them on local hardware (Subashini & Kavitha, 2011). Cloud computing also differs from traditional information-processing infrastructures by allowing virtualization, dynamic resource allocation, and automated scalability (Butt et al., 2020; Samani et al., 2014).

From a service-delivery standpoint, cloud computing is typically categorized into three main service models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) (Hubbard & Sutton, 2010; Sharma & Singh, 2022). Because each service model assigns security responsibilities differently between the cloud service provider and the user, each service model also presents different security requirements and risk profiles.

In the SaaS model, software applications are delivered to users through network-based services, while the cloud provider is primarily responsible for maintaining the underlying infrastructure, updating the service, and implementing major security controls. Common SaaS environments provide users with access to applications and data without requiring them to manage the underlying computing infrastructure (Attar et al., 2021).

In the PaaS model, the provider provides a development, test, and deployment environment for applications, thus allowing developers to concentrate on application development without having to manage the physical infrastructure directly. The responsibilities for security lie with the service provider and the user, depending on the service architecture and deployment configuration (Zekri et al., 2017).

In the IaaS model, computing resources, storage, networking capabilities, and virtual machines are provided as services. Although IaaS offers users a high degree of flexibility and control, users also assume greater responsibility for securing operating systems, virtual machines, applications, data, and relevant security configurations (Nguyen et al., 2018).

With the volume of data stored in cloud environments and the complexity of new cyberattacks and security threats, conventional security risk assessment approaches have become increasingly limited. Over the last few years, however, supervised machine learning has increasingly been used for security analytics, threat-pattern identification, risk-factor analysis, and predictive security assessment (Vinayakumar et al., 2019). Machine-learning models learn relationships from historical or labeled data and can recognize complex patterns among risk factors to assist in systematic risk prediction (Vinayakumar et al., 2019).

Much existing work has been done applying machine learning to cloud security, but most of it has either focused on detecting a particular type of cyberattack or evaluating a single machine-learning algorithm; comparatively fewer studies have developed an integrated framework that combines risk-factor identification, feature selection, machine-learning model development, and comparative performance evaluation within a unified security risk assessment framework. This gap indicates a need for approaches that can systematically integrate multiple security-risk dimensions with data-driven predictive techniques.

Therefore, this study develops a smart security risk assessment model for cloud computing environments using supervised machine-learning algorithms. Specifically, the major security risk factors were identified through a systematic review of relevant scientific literature and security-related frameworks, classified, and prepared as input variables for model development, using feature-selection techniques to identify relevant variables and reduce unnecessary dimensionality, and the predictive performance of Decision Tree, K-Star (K*), and Random Filtered Classifier was evaluated and compared under different training–testing data-split scenarios.

2. Literature Review and Related Studies

Cloud computing has emerged as an integral part of the modern information technology infrastructure, offering scalable, flexible, and on-demand computing resources, storage, applications, and digital services (Subashini & Kavitha, 2011). However, the distributed and dynamic nature of cloud environments, coupled with virtualization, resource sharing, multi-tenancy, and reliance on network-based services (Bhadauria et al., 2011; Masdari & Jalali, 2016), has given rise to a myriad of security challenges (Subashini & Kavitha, 2011), and thus, security risk assessment has become a significant element of cloud security management (Subashini & Kavitha, 2011; Bhadauria et al., 2011).

Cloud security includes a number of different dimensions such as network and infrastructure security, virtualization security, identity and access management, authentication, application security, data protection, service availability, regulatory compliance, and the three fundamental pillars of confidentiality, integrity, and availability (Subashini & Kavitha, 2011; Bhadauria et al., 2011; Srivastava & Kumar, 2014). The interdependence of these dimensions makes security risk assessment in cloud environments a multidimensional issue rather than a one-dimensional threat-detection task.

2.1. Security Threats and Risks in Cloud Computing Environments

These security risks are due to the technological characteristics of cloud infrastructures and operational and managerial processes of providing cloud services, as compared with traditional computing environments (Subashini & Kavitha, 2011; Bhadauria et al., 2011; Masdari & Jalali, 2016). Traditional computing environments use dedicated hardware and local access, while cloud infrastructures use virtualization, sharing, dynamic allocation, remote access, and multi-tenant architectures that can make identifying vulnerabilities and potential consequences more complex (Subashini & Kavitha, 2011; Bhadauria et al., 2011; Masdari & Jalali, 2016).

Another key area of concern in cloud security risk involves identity and access control. Poor authentication methods, incorrect access permissions, compromised accounts, and poor identity management can enable unauthorized users to gain access to protected resources; therefore, robust authentication and access-control mechanisms are critical to protecting cloud resources and limiting unauthorized activities (Subashini & Kavitha, 2011; Srivastava & Kumar, 2014).

Another important aspect of cloud security is data security, which refers to the confidentiality and integrity of data stored or processed in the cloud (Subashini & Kavitha, 2011; Bhadauria et al., 2011; Parte & Dehariya, 2015). Since organizations can store sensitive and business-critical information on shared cloud infrastructures, they need proper data protection, segregation, recovery, and investigation mechanisms.

Infrastructure and virtualization security also constitute important areas of concern. The sharing of physical resources among multiple tenants can introduce risks associated with isolation, virtualization vulnerabilities, and resource exhaustion. Vulnerabilities in virtual machines, hypervisors, or shared infrastructure may potentially affect multiple services or tenants within the same environment (Bhadauria et al., 2011; Masdari & Jalali, 2016; Srivastava & Kumar, 2014).

Additional risks include insecure application programming interfaces (APIs), application vulnerabilities, inappropriate configurations, and weaknesses in service architectures that may serve as attack vectors for an adversary to gain unauthorized access or disrupt cloud services (Subashini & Kavitha, 2011; Bhadauria et al., 2011). Issues related to service interfaces, shared technologies, data protection, and account-related security have been identified as critical areas for attention based on previous surveys of cloud-security issues (Subashini & Kavitha, 2011; Bhadauria et al., 2011).

Another dimension is that of service continuity and availability, as cloud services are usually assumed to be available continuously and reliably, and service availability can be impacted by distributed denial-of-service (DDoS) attacks, resource exhaustion, infrastructure failures, and other disruptions, resulting in operational consequences (Masdari & Jalali, 2016). Denial-of-service and DDoS attacks have been specifically identified in the literature as significant security issues for cloud environments (Masdari & Jalali, 2016).

Aside from technical threats, governance, regulatory, and organizational factors also impact the security of cloud computing (Subashini & Kavitha, 2011; Samani et al., 2014). Thus, security risk assessment should include both technical vulnerabilities and organizational and governance-related factors.

In addition, the complexity of cloud-security threats has led to the use of machine-learning techniques for security analysis, with investigations of machine-learning methods for various cloud-security problems such as attack detection, intrusion detection, malware detection, anomaly detection, and security-risk analysis (Butt et al., 2020; Zekri et al., 2017; Nguyen et al., 2018; Vinayakumar et al., 2019; Diro & Chilamkurti, 2018; Berman et al., 2019). In their comprehensive review, Butt et al. (2020) found that supervised, unsupervised, semi-supervised, and reinforcement-learning approaches have all been studied for various cloud-security challenges.

2.2. Review of Related Studies on Security Risk Assessment

A number of previous studies have used different methodologies for cloud security risk assessment, such as traditional risk analysis and security taxonomies, machine-learning-based detection and prediction models, and other approaches (Subashini & Kavitha, 2011; Butt et al., 2020; Bhadauria et al., 2011).

Subashini and Kavitha (2011) examined security issues associated with different cloud service-delivery models and demonstrated that security concerns vary according to the characteristics and responsibilities associated with SaaS, PaaS, and IaaS environments. Their study provides an important conceptual basis for understanding security risks according to different cloud service models.

A general overview of security challenges in cloud computing was presented by Bhadauria et al. (2011), who highlighted the various types of security threats arising from the architecture and characteristics of cloud environments. Likewise, Masdari and Jalali (2016) surveyed denial-of-service attacks in cloud computing and categorized various types of DoS attacks, which also highlighted the importance of service availability as a major security concern.

With the development of machine learning, researchers have increasingly applied supervised and other machine-learning techniques to cybersecurity problems. Decision Tree, Random Forest, Support Vector Machine, Artificial Neural Network, and other algorithms have

been investigated for detecting cyberattacks, classifying security events, and identifying anomalous behavior (Butt et al., 2020; Zekri et al., 2017; Nguyen et al., 2018; Vinayakumar et al., 2019; Diro & Chilamkurti, 2018; Berman et al., 2019).

Some examples of using machine-learning techniques for DDoS attack detection include the works of Zekri et al. (2017), who developed machine-learning methods for DDoS attack detection in cloud computing environments, and Nguyen et al. (2018), who developed a deep-learning approach for cyberattack detection in mobile cloud computing. Vinayakumar et al. (2019) applied deep-learning techniques for intelligent intrusion detection, and Diro and Chilamkurti (2018) applied distributed deep-learning approaches to attack detection in Internet of Things environments.

Other studies have explored machine learning from a more general cybersecurity perspective, such as a review of deep-learning methods for cybersecurity and their applications to various security tasks (Berman et al., 2019), and investigations into risk prediction for imbalanced cybersecurity data using a deep-learning classification framework (Sun et al., 2019).

In the general context of risk assessment, Hegde and Rokseth (2020) reviewed the use of machine-learning methods for engineering risk assessment and noted the growing trend toward data-driven approaches for risk-related decision-making; Lin et al. (2021) also applied fuzzy-set theory and machine-learning methods to risk assessment and management, demonstrating the potential to apply machine learning to risk assessment beyond traditional attack-detection tasks; and Radanliev et al. (2020) outlined a method combining artificial intelligence, machine learning, and real-time intelligence for predictive cyber-risk analytics, which suggests a trend toward more dynamic security assessment to data-driven and predictive risk analysis.

A review indicates substantial progress in cloud threat detection (Hubbard & Sutton, 2010), cybersecurity analytics (Attar et al., 2021) and machine-learning based security assessment (Zekri et al., 2017). The reviewed literature highlights that while there are significant developments made on cloud threat detection (Hubbard & Sutton, 2010), cybersecurity analytics (Nguyen et al., 2018) as well as machine-learning based security assessment (Vinayakumar et al., 2019); the differences exist across risk factors, datasets, algorithms used to develop supervised models, feature selection procedures followed, and evaluation strategies. This research aims at integrating identification of risk factor(s), selecting features, developing a supervised machine learning model for predicting outcome labels, constructing an ensemble or consensus system that can combine results from multiple sources into one prediction with the comparative performance assessment into a holistic security risk assessment framework.

2.3. Identification and Classification of Security Risk Factors

Because security incidents can be caused by technical vulnerabilities, infrastructure characteristics, data-management practices, service-management processes, governance requirements, or interactions between these factors (Subashini & Kavitha, 2011; Butt et al., 2020; Bhadauria et al., 2011), the identification and classification of relevant risk factors are a crucial step in constructing an intelligent risk assessment model. Based on the literature reviewed in this study, the security risk factors were organized into six major dimensions:

1. Identity and Access Control Security
2. Data Security
3. Infrastructure and Virtualization Security
4. Software and Service Security
5. Security Management and Governance
6. Interoperability and Portability

These dimensions provide the conceptual structure for organizing the 18 security risk factors used in the proposed model. The classification establishes a basis for subsequent data preparation, feature selection, and supervised machine-learning model development.

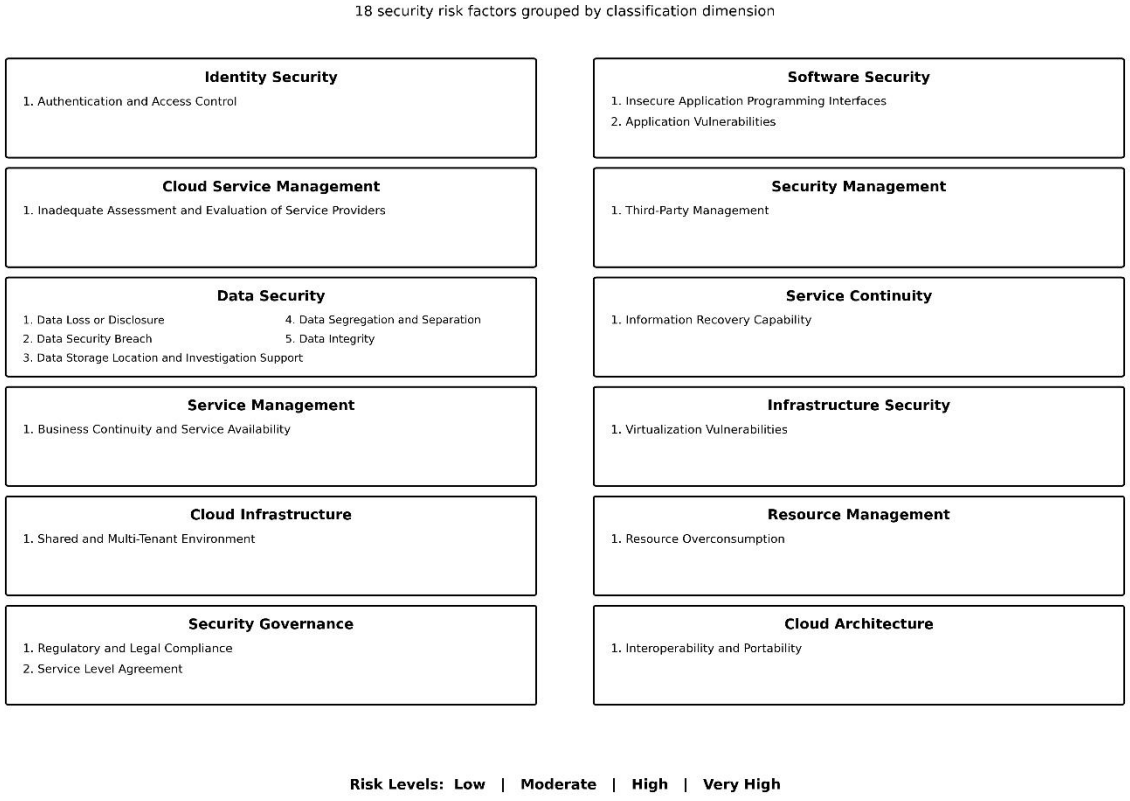


Figure 1. Classification of Security Risk Factors in Cloud Computing Environments

The classification presented in Figure 1 encompasses the principal security dimensions considered in the present study. The Identity and Access Control Security dimension covers risks associated with authentication, authorization, identity management, and access to cloud resources (Subashini & Kavitha, 2011; Srivastava & Kumar, 2014). The Data Security dimension encompasses risks related to data loss, disclosure, data breaches, data integrity, data segregation, storage location, investigation support, and recovery (Subashini & Kavitha, 2011; Bhadauria et al., 2011).

The Infrastructure and Virtualization Security dimension addresses risks arising from shared and multi-tenant environments, virtualization vulnerabilities, and excessive resource consumption (Bhadauria et al., 2011; Masdari & Jalali, 2016). The Software and Service Security dimension covers insecure application programming interfaces, application vulnerabilities, and business continuity and service availability (Subashini & Kavitha, 2011; Butt et al., 2020).

The Security Management and Governance dimension includes regulatory and legal compliance, third-party management, and service-level agreements. Finally, Interoperability and Portability addresses risks associated with the ability of cloud systems and services to interact with or be transferred across different environments.

The identified factors were subsequently operationalized and prepared as model input variables. Their detailed operational definitions, codes, dimensions, and risk-level classifications are presented in Table 2 in the Results and Discussion section.

3. Methodology

The current study is purposeful and descriptive–analytical with a focus on designing a smart security risk assessment model via supervised machine-learning algorithms and methodology that combines security-risk-factor identification, data preparation, feature selection, supervised machine-learning model development, ensemble construction, and model performance evaluation. There has been a growing body of research exploring the use of machine-learning techniques for cloud-security analysis, such as security-threat detection, classification, and risk analysis (Butt et al., 2020; Zekri et al., 2017; Nguyen et al., 2018; Vinayakumar et al., 2019; Diro & Chilamkurti, 2018; Berman et al., 2019).

3.1. Proposed Model Framework

This paper proposed a framework that integrates major stages needed for intelligent security risk assessment in cloud computing environments, including security risk factor identification and classification, data preparation, feature selection, supervised machine-learning model development, ensemble construction, and performance evaluation. The main stages of the proposed framework are as follows:

1. Identification and classification of security risk factors;

2. Data preparation and preprocessing;
3. Selection of relevant features;
4. Development and training of supervised machine-learning models;
5. Construction of the ensemble model; and
6. Evaluation and comparison of model performance.

This integrated structure follows the general direction of recent cloud-security research in which machine-learning techniques are applied to security-related data to identify patterns, classify security conditions, and support risk analysis (Butt et al., 2020). The framework establishes a logical relationship between the conceptual identification of security risks and their empirical assessment through machine-learning techniques. The identified risk factors constitute the input space of the model, while the resulting security-risk level represents the target variable.

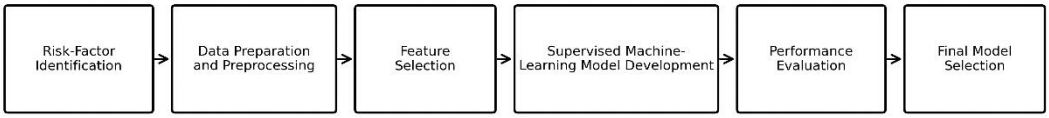


Figure 2. Overall Framework for Developing the Proposed Intelligent Security Risk Assessment Model in Cloud Computing Environments

As illustrated in Figure 2, the proposed framework follows a sequential process in which security-risk factors are first identified and classified, followed by data preparation and feature selection. The selected variables are then used to train the supervised machine-learning algorithms. Their outputs are subsequently evaluated using the defined performance criteria to compare the predictive behavior of the models.

3.2. Risk-Factor Identification and Data Preparation

First, security risk factors applicable to cloud computing environments were identified by conducting a systematic review of scientific literature, studies, security reports, and security-related frameworks with respect to technical, operational, managerial, governance, and service-related factors that may impact the security of the cloud. This systematic identification and classification of security factors is in line with other cloud-security studies that have emphasized the importance of taking into account multiple types of threats and vulnerabilities instead of singular types of attacks (Subashini & Kavitha, 2011; Butt et al., 2020; Bhadauria et al., 2011).

After identifying and screening the relevant factors, overlapping or closely related concepts were merged, resulting in the identification of 18 major security risk factors that represent various dimensions of cloud security, such as identity and access control, data security, infrastructure security, software and application security, service management, security governance, and cloud architecture. The identified factors were used as the independent variables of the proposed model, and each factor was evaluated with four security-risk levels (Low, Moderate, High, and Very High). These qualitative categories were converted into numerical values through the coding procedure used in the dataset so that they could be processed by the machine-learning algorithms.

The resulting dataset had 1,940 instances and 18 input features, where each instance represented a particular security-risk condition and the overall security-risk level was treated as the dependent variable. Table 2 in Section 4 provides a more detailed description of the 18 input variables, including their categories, symbols, and assessment ranges.

3.3. Feature Selection

In this study, feature selection was performed to identify the variables contributing to security-risk prediction, reduce unnecessary dimensionality, and prepare for model training before machine-learning applications, particularly in the case of model training, since the inclusion of irrelevant or redundant variables can increase computational complexity and may affect model performance. Feature selection was performed using Best First Search, Ranker, and Random Search using the WEKA and Orange Data Mining environments.

Feature-selection procedures are often integrated into machine-learning workflows to extract relevant attributes prior to classification or prediction. In the context of cloud security, prior research (Butt et al., 2020) has investigated various machine-learning algorithms and their respective feature sets to detect security threats and vulnerabilities; the feature-selection procedures were applied to the variables that represent security risks prior to developing the supervised machine-learning models to identify the relevant variables and prepare an appropriate feature set for model training and evaluation.

3.4. Development of the Supervised Machine-Learning Model

After preparing data and selecting features, three supervised machine-learning algorithms were chosen for the development of models: Decision Tree, Random Filtered Classifier, and K-Star (K*). This is in line with

prior studies in cloud security where supervised methods have been used on labeled security data for classification, detection, and prediction (Butt et al., 2020; Zekri et al., 2017; Nguyen et al., 2018; Vinayakumar et al., 2019; Diro & Chilamkurti, 2018; Berman et al., 2019).

The three algorithms were a tree-based decision structure that classifies variables and target classes, a filtered classification approach using a Random Filtered Classifier, and an instance-based learning algorithm using K-Star with an entropy-based distance function. The WEKA data-mining environment was used for implementing and evaluating the three algorithms, which includes a collection of machine-learning and data-mining tools that can be used for data preprocessing, attribute selection, classification, and model evaluation (Hall et al., 2009).

After the preprocessing and feature-selection stages, the prepared dataset was supplied to the algorithms, and each algorithm was trained to predict the security-risk level from the selected risk factors. Several supervised algorithms were employed to examine the predictive behavior of the models under the same experimental conditions, which allowed for the comparison of their prediction errors and how the selected algorithms performed under various training–testing data-split scenarios.

3.5. Ensemble Model Construction

Besides the evaluation of the individual machine-learning algorithms, the proposed framework included a Voting Ensemble approach, which is used to combine multiple individual classifiers and provide an aggregated prediction. Ensemble learning takes the outputs of multiple predictive models and can be an alternative to improving the robustness of a prediction system. The Voting Ensemble was added to the proposed modelling framework in this study to aggregate the predictions from the individual classifiers.

However, the experimental results do not present quantitative performance measures for the Voting Ensemble; thus, its performance is considered as a methodological aspect of the proposed framework and is not compared separately with the three individual algorithms in the results, as there are numerical RMSE values for the three individual algorithms, but not for the Voting Ensemble.

3.6. Model Performance Evaluation

The predictive performance of the machine-learning algorithms was evaluated using the Percentage Split approach. Four training–testing configurations were examined:

- 70% training – 30% testing
- 80% training – 20% testing
- 90% training – 10% testing
- 95% training – 5% testing

The goal of testing multiple data-splitting configurations was to investigate the impact of the training and testing data proportions on the prediction error of the evaluated algorithms. Experiments were conducted within the WEKA environment (Hall et al., 2009), which offers classification, attribute selection, and evaluation of machine-learning models.

The experimental results reported the Root Mean Square Error (RMSE), a common performance measure that quantifies the size of the prediction error, with smaller values indicating smaller prediction error. The study design also considered the Correlation Coefficient (R) as a performance measure for investigating the relationship between predicted and observed values; however, the available comparative results only provide numerical values for RMSE. Hence, the main algorithm comparison presented in Section 4 is based on the reported RMSE values.

The four percentage-split scenarios were applied consistently to the three supervised machine-learning algorithms, allowing their prediction errors to be compared under the same experimental conditions. The resulting values are presented in Table 4 and further discussed in Sections 4.1 and 4.2.

4. Results and Discussion

The results of the study were obtained through the systematic identification and assessment of security risk factors, followed by data preparation, feature selection, and the application of supervised machine-learning algorithms. The analysis focused on evaluating the predictive performance of the developed models under different training–testing data-split scenarios.

4.1. Evaluation of Machine-Learning Model Performance

The 18 major security risk factors identified as input features of the proposed model based on the assessment conducted in the study (Table 2) are classified into eight categories of cloud-computing security dimensions (identity and access management, data security, infrastructure security, software security, service management, legal and regulatory compliance, virtualization, and service continuity) and were evaluated using four risk levels (Low, Moderate, High, and Very High). The qualitative categories of risk levels were converted into numerical values using the coding scheme in the dataset to make them suitable for processing by the machine-learning algorithms.

The final dataset consisted of 1,940 instances and 18 input features. Each instance represented a specific security-risk condition, while the corresponding risk level was used as the target variable. The characteristics of the input features and their classification are presented in Table 2.

Table 2. Input Features of the Model and Classification of Security Risk Factors

No.	Security Risk Factor	Main Category	Variable Symbol	Assessment Range
1	Authentication and Access Control	Identity Security	A&AC	Low, Moderate, High, Very High
2	Inadequate Assessment and Evaluation of Service Providers	Cloud Service Management	IDD	Low, Moderate, High, Very High
3	Data Loss or Disclosure	Data Security	DL	Low, Moderate, High, Very High
4	Insecure Application Programming Interfaces	Software Security	IAP	Low, Moderate, High, Very High
5	Application Vulnerabilities	Software Security	AV	Low, Moderate, High, Very High
6	Business Continuity and Service Availability	Service Management	BC&SA	Low, Moderate, High, Very High
7	Shared and Multi-Tenant Environment	Cloud Infrastructure	ShE	Low, Moderate, High, Very High
8	Regulatory and Legal Compliance	Security Governance	RC	Low, Moderate, High, Very High
9	Data Security Breach	Data Security	DB	Low, Moderate, High, Very High
10	Data Storage Location and Investigation Support	Data Security	DL&IS	Low, Moderate, High, Very High
11	Third-Party Management	Security Management	TPM	Low, Moderate, High, Very High
12	Data Segregation and Separation	Data Security	DS	Low, Moderate, High, Very High
13	Information Recovery Capability	Service Continuity	R	Low, Moderate, High, Very High
14	Data Integrity	Data Security	DI	Low, Moderate, High, Very High
15	Virtualization Vulnerabilities	Infrastructure Security	VV	Low, Moderate, High, Very High
16	Resource Overconsumption	Resource Management	RE	Low, Moderate, High, Very High
17	Service Level Agreement	Security Governance	SLA	Low, Moderate, High, Very High
18	Interoperability and Portability	Cloud Architecture	I&P	Low, Moderate, High, Very High

Following data preparation and feature-selection procedures, these variables were used as inputs for training and evaluating the machine-learning models.

The experiments were conducted using WEKA, which provides tools for data preprocessing, attribute selection, classification, and model evaluation (Hall et al., 2009). The use of WEKA is consistent with its established application as a machine-learning and data-mining workbench (Hall et al., 2009).

In this study, three supervised machine-learning algorithms—Decision Tree, Random Filtered Classifier, and K-Star (K*)—were evaluated. The application of

these algorithms to cloud-security risk assessment is also consistent with previous research, including the closely related study by Sharma and Singh (2022), which examined Decision Tree, Randomizable Filter Classifier, and K-Star for cloud-security risk assessment using RMSE.

To investigate the effect of different training and testing proportions, four Percentage Split scenarios were evaluated:

- 70% training – 30% testing
- 80% training – 20% testing
- 90% training – 10% testing
- 95% training – 5% testing

The predictive performance of the three algorithms was assessed using the Root Mean Square Error (RMSE). The results are presented in Table 4.

Table 4. Comparison of Machine-Learning Algorithms Based on RMSE under Different Data-Splitting Scenarios

Algorithm	70% Training – 30% Testing	80% Training – 20% Testing	90% Training – 10% Testing	95% Training – 5% Testing
Decision Tree	0.0022	0.0020	0.0019	0.0018
Random Filtered Classifier	0.0046	0.0047	0.0038	0.0035
K-Star	0.0046	0.0047	0.0038	0.0035

4.2. Comparative Performance of the Algorithms

As shown in Table 4, the Decision Tree algorithm produced the lowest reported RMSE across all four data-splitting scenarios. Its RMSE decreased from 0.0022 under the 70% training–30% testing configuration to 0.0018 under the 95% training–5% testing configuration.

The Random Filtered Classifier and K-Star produced identical RMSE values across the four evaluated scenarios, namely 0.0046, 0.0047, 0.0038, and 0.0035, respectively.

The observed performance pattern is consistent with the findings reported by Sharma and Singh (2022), who evaluated the same general group of algorithms for cloud-computing security risk assessment and reported that the Decision Tree produced the lowest RMSE among the evaluated algorithms, while the 95%–5% partition produced the lowest reported error among the tested partitions.

The results also show that the reported RMSE decreased as the proportion of training data increased. However, the 95% training–5% testing configuration should be interpreted cautiously. Although it produced the lowest reported RMSE in the present experiment, only 5% of the observations were used for testing. Therefore, this result alone cannot establish superior generalizability to independent or unseen datasets.

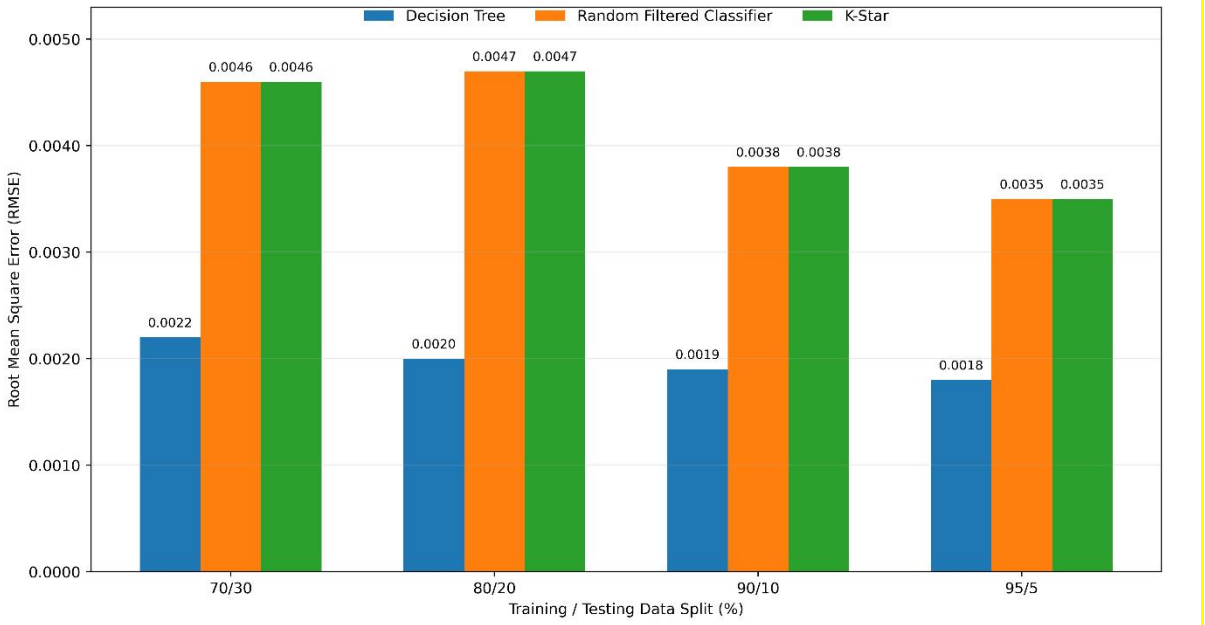


Figure 3. Comparison of Machine-Learning Algorithms Based on RMSE under Different Data-Splitting Scenarios

Figure 3 presents a visual comparison of the RMSE values reported in Table 4. The figure should illustrate the lower reported RMSE values of the Decision Tree across the four experimental configurations and the identical RMSE pattern observed for the Random Filtered Classifier and K-Star.

The Voting Ensemble was incorporated into the proposed methodology as a model-combination component. However, because separate quantitative RMSE results for the ensemble are not reported in the available experimental results, no independent performance claim is made for the Voting Ensemble.

Overall, the results provide empirical evidence that supervised machine-learning algorithms can be applied to the assessment and prediction of security-risk levels in cloud-computing environments. The comparative results also demonstrate the importance of evaluating algorithms under consistent experimental conditions and across different training–testing configurations.

5. Conclusion

The proliferation of cloud computing has brought tremendous opportunities to organizations through elastic and agile access to computing resources, reducing infrastructure needs, and facilitating the provision of digital services on demand. However, the complexity of cloud infrastructures and the variety of security threats have made security risk assessment a significant challenge, and conventional approaches that rely heavily on static analysis and expert judgment may not fully reflect the complexity and evolving nature of cloud environments.

The objective of this research was to design an intelligent security risk assessment model for cloud computing environments based on supervised machine-learning algorithms, which included identifying and categorizing security-risk factors through a literature review and expert validation, employing feature-selection techniques, and training and comparing supervised machine-learning algorithms for predicting security-risk levels.

In the expert-validation stage, 35 international experts evaluated the identified security-risk factors in terms of their importance, likelihood of occurrence, and severity of impact, resulting in 18 major security-risk factors being retained as input variables, with 1,940 instances and 18 input features, represented with security-risk conditions through four levels: Low, Moderate, High, and Very High.

Three supervised machine-learning algorithms—Decision Tree, Random Filtered Classifier, and K-Star—were evaluated under four training–testing configurations: 70%–30%, 80%–20%, 90%–10%, and 95%–5%. The reported results show that the Decision Tree achieved the lowest RMSE in all four scenarios, with its RMSE decreasing from 0.0022 under the 70%–30% split to 0.0018 under the 95%–5% split. The Random Filtered Classifier and K-Star produced identical RMSE values across the reported scenarios. These results are consistent with the general applicability of supervised machine-learning techniques to cloud-security risk analysis reported in previous research (Butt et al., 2020; Sharma & Singh, 2022).

The results therefore demonstrate that, within the experimental conditions and dataset used in this study, the Decision Tree generated lower prediction error than the other evaluated algorithms. The 95%–5% configuration also produced the lowest reported RMSE values. However, because this configuration uses only 5% of the dataset for testing, its lower error should be interpreted specifically as an experimental result rather than as independent evidence of superior generalizability.

The main contribution of the study is the formulation of an integrated framework for identifying security-risk factors, validating the identified factors by experts, selecting features, applying supervised machine learning, and evaluating the models comparatively, which offers a solid foundation for analyzing cloud security risks and could facilitate the design of intelligent decision-support systems for organizations, cloud-service providers, and information-security managers.

5.1. Limitations and Directions for Future Research

However, the findings of this study should be interpreted with caution due to the following limitations: (1) the experimental dataset contained 1,940 instances and 18 input features, thus the reported performance is dependent on the characteristics of the dataset and the risk-assessment framework used in this study; (2) model evaluation was based on percentage-based training–testing splits, particularly 95%–5% which led to a relatively small testing subset, and thus the lower RMSE value should not be used to conclude higher performance on unseen real-world data.

Third, although a Voting Ensemble approach was incorporated into the methodological framework, the available experimental results do not report separate quantitative performance measures for the ensemble model. Therefore, no independent conclusion can be drawn regarding whether the ensemble approach improves prediction performance relative to the individual algorithms.

These can be derived from real cloud-service environments and by using stronger validation techniques, such as k-fold cross-validation and independent external test datasets. Additional supervised and ensemble-learning algorithms may also be explored and compared using multiple evaluation measures (rather than just RMSE), and continuously updated security data could enable the model to adapt more quickly to shifts in cloud-security conditions and new threats.

In general, the proposed framework offers a systematic approach to incorporating security-risk-factor analysis,

feature selection, supervised machine learning, and model evaluation into an intelligent cloud-security risk assessment process, and further validation with larger, independently collected, and dynamically updated datasets would establish the applicability of the framework in various cloud environments and operational contexts.

References

Attar, R. A., Al-Nemri, J., Homsi, A., & Qusef, A. (2021). Risk assessment for emerging domains (IoT, cloud computing, and AI). In *2021 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 120–127). IEEE. https://doi.org/10.1109/JEEIT53412.2021.9634156

Beimborn, D., Miletzki, T., & Wenzel, S. (2011). Platform as a service (PaaS). *Business & Information Systems Engineering*, *3*(6), 381–384.

Berman, D., Buczak, A., Chavis, J., & Corbett, C. (2019). A survey of deep learning methods for cyber security. *Information*, *10*(4), 122. https://doi.org/10.3390/info10040122

Bhadauria, R., Chaki, N., Chaki, R., & Sanyal, S. (2011). A survey on security issues in cloud computing. *arXiv preprint arXiv:1109.5388*, 1–15.

Butt, U. A., Mehmood, M., Shah, S. B. H., Amin, R., Shaukat, M. W., Raza, S. M., Suh, D. Y., & Piran, M. J. (2020). A review of machine learning algorithms for cloud computing security. *Electronics*, *9*(9), 1379. https://doi.org/10.3390/electronics9091379

Buyya, R., Ranjan, R., & Calheiros, R. N. (2009). Modeling and simulation of scalable cloud computing environments and the CloudSim toolkit: Challenges and opportunities. In *2009 International Conference on High Performance Computing & Simulation* (pp. 1–11). IEEE.

Dawoud, W., Takouna, I., & Meinel, C. (2010). Infrastructure as a service security: Challenges and solutions. In *2010 7th International Conference on Informatics and Systems (INFOS)* (pp. 1–8). IEEE.

Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. *Future Generation Computer Systems*, *82*, 761–768. https://doi.org/10.1016/j.future.2017.08.043

Hall, M. A., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18. https://doi.org/10.1145/1656274.1656278

Hegde, J., & Rokseth, B. (2020). Applications of machine learning methods for engineering risk assessment – A review. *Safety Science*, *122*, 104492. https://doi.org/10.1016/j.ssci.2019.09.015

Hubbard, D., & Sutton, M. (2010). *Top threats to cloud computing V1.0*. Cloud Security Alliance.

Jain, A., & Kumar, R. (2014). A taxonomy of cloud computing. *International Journal of Scientific Research Publications*, *4*(7), 1–5.

Lin, S. S., Shen, S. L., Zhou, A., & Xu, Y. S. (2021). Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods. *Automation in Construction*, *122*, 103490. https://doi.org/10.1016/j.autcon.2020.103490

Masdari, M., & Jalali, M. (2016). A survey and taxonomy of DoS attacks in cloud computing. *Security and Communication Networks*, *9*(16), 3724–3751.

Miller, M. (2008). *Cloud computing: Web-based applications that change the way you work and collaborate online*. Que Publishing.

Nguyen, K. K., Hoang, D. T., Niyato, D., Wang, P., Nguyen, D., & Dutkiewicz, E. (2018). Cyberattack detection in mobile cloud computing: A deep learning approach. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1–6). IEEE. https://doi.org/10.1109/WCNC.2018.8376973

Parte, S., & Dehariya, N. (2015). Cloud computing: Issues regarding security, applications and mobile cloud computing. *International Journal of Advanced Research in Computer Science and Software Engineering*, *5*(3), 403–406.

Radanliev, P., De Roure, D., Page, K., et al. (2020). Design of a dynamic and self-adapting system, supported with artificial intelligence, machine learning and real-time intelligence for predictive cyber risk analytics in extreme environments – cyber risk in the colonisation of Mars. *Safety in Extreme Environments*, *2*, 219–230. https://doi.org/10.1007/s42797-021-00025-1

Samani, R., Reavis, J., & Honan, B. (2014). *CSA guide to cloud computing: Implementing cloud privacy and security*. Syngress.

Savu, L. (2011). Cloud computing: Deployment models, delivery models, risks and research challenges. In *2011 International Conference on Computer and Management (CAMAN)* (pp. 1–4). IEEE.

Sharma, A., & Singh, U. K. (2022). Modelling of smart risk assessment approach for cloud computing environment using AI & supervised machine learning algorithms. *Global Transitions Proceedings*, *3*(1), 243–250. https://doi.org/10.1016/j.gltp.2022.03.030

Srivastava, H., & Kumar, S. A. (2014). Control framework for secure cloud computing. *Journal of Information Security*, *6*(1), 12–21.

Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud

computing. *Journal of Network and Computer Applications*, *34*(1), 1–11.

Sun, D., Wu, Z., Wang, Y., Lv, Q., & Hu, B. (2019). Risk prediction for imbalanced data in cyber security: A Siamese network-based deep learning classification framework. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). https://doi.org/10.1109/IJCNN.2019.8852030

Tan, L., Yu, K., Ming, F., Chen, X., & Srivastava, G. (2021). Secure and resilient artificial intelligence of things: A HoneyNet approach for threat detection and situational awareness. *IEEE Consumer Electronics Magazine*.

Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *IEEE Access*, *7*, 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334

Vu, D. L., Nguyen, T. K., Nguyen, T. V., Nguyen, T. N., Massacci, F., & Phung, P. H. (2020). HIT4Mal: Hybrid image transformation for malware classification. *Transactions on Emerging Telecommunications Technologies*, *31*(11), e3789.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann. https://doi.org/10.1016/C2009-0-19715-5

Zekri, M., El Kafhali, S., Aboutabit, N., & Saadi, Y. (2017). DDoS attack detection using machine learning techniques in cloud computing environments. In *2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech)* (pp. 1–7). https://doi.org/10.1109/CloudTech.2017.8284731