

Enhancing cybersecurity vulnerability detection using different machine learning severity prediction models

Fawaz Alanazi ¹, Ahmed Badi Alshammari ², Chams Sallami ¹, Asma A. Alhashmi ¹, Rachid Effghi ³, Anil Kumar KM ⁴, Abdulbasit Darem ^{5*}

¹ Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia

² Department of Computer Science, College of Computing and Information Technology, Northern Border University, Saudi Arabia

³ Department of Big Data Analytics and Management, Bahcesehir University, Türkiye

⁴ JSS Science and Technology University, Department of Computer Science and Engineering, Mysuru, India

⁵ Center for Scientific Research and Entrepreneurship, Northern Border University, Arar, Saudi Arabia

ABSTRACT

In today's highly connected digital environment, effectively managing cybersecurity vulnerabilities is essential to protecting organizational systems. This research examines the use of machine learning models to predict the severity of vulnerabilities, utilizing data from the 2022, Cybersecurity and Infrastructure Security Agency (CISA) known exploited vulnerabilities catalogue. The study evaluates five machine learning models—Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Support Vector Machine—based on their performance in terms of accuracy, precision, recall, and computational efficiency. The results show that tree-based models, especially Decision Tree, Random Forest, and Gradient Boosting, achieved perfect accuracy (100%) in categorizing vulnerabilities by severity, outperforming Logistic Regression and Support Vector Machine, which faced difficulties with critical vulnerabilities. Additionally, tree-based models demonstrated superior computational efficiency, with Decision Tree standing out in terms of both speed and accuracy, making it ideal for real-time use. The study emphasizes the potential of machine learning to automate and improve vulnerability management, allowing security teams to prioritize significant threats and better allocate resources. Future work should focus on incorporating real-time data and exploring deep learning methods to enhance model adaptability and performance. Overall, the research highlights the importance of machine learning in bolstering cybersecurity defenses.

Keywords: Cybersecurity, Machine learning models, Threat prioritization, Vulnerability management, Vulnerability severity prediction.

OPEN ACCESS


Received: January 21, 2025

Revised: February 21, 2025

Accepted: March 9, 2025

Corresponding Author:

Abdulbasit A. Darem
basit.darem@nbu.edu.sa

 **Copyright:** The Author(s). This is an open access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted distribution provided the original author and source are cited.

Publisher:

[Chaoyang University of Technology](https://www.chaoyang.edu.cn/)

ISSN: 1727-2394 (Print)

ISSN: 1727-7841 (Online)

1. INTRODUCTION

In the digital age, cybersecurity is a paramount concern, as most industries heavily rely on information technology to support their operations. With the increased interconnectivity of systems, the threat of cybersecurity breaches has become more pervasive, exposing organizations to significant risks (Solms and Niekerk, 2013). The surge in cyberattacks highlights the critical importance of robust vulnerability management processes to safeguard these systems. A vulnerability is defined as a flaw or weakness in a system that can be exploited by attackers to gain unauthorized access or cause damage. The identification, assessment, and mitigation of vulnerabilities form the cornerstone of cybersecurity risk management, enabling organizations to proactively protect their infrastructures (Smadi et al., 2018). However, vulnerability management is

not without its challenges. Modern systems are complex, with the new vulnerabilities being discovered daily (Zhang et al., 2011). As organizations continue to expand their digital footprints, the number and variety of vulnerabilities they face grows, making manual management methods inefficient and impractical. This ever-increasing volume of threats has spurred the need for automated, scalable approaches to vulnerability assessment and prioritization (Nayak et al., 2014). While we build upon existing research, we also identify limitations in these approaches to justify the need for our proposed machine learning-based solution for vulnerability severity prediction. We aim to address these limitations by developing an optimized model that improves accuracy, efficiency, and scalability in assessing cybersecurity threats. Such approaches should provide organizations with the ability to quickly assess potential threats and implement timely mitigations to protect critical assets.

A fundamental aspect of vulnerability management is determining the severity of vulnerability. The severity level indicates the potential damage an exploit can inflict, and it guides security professionals in allocating resources to address most critical vulnerabilities first (Mell et al., 2007). Accurate prediction of vulnerability severity is therefore vital for maintaining an effective security posture. Security teams use this information to prioritize efforts based on the threat level, thereby enhancing their organization's overall resilience against cyberattacks (Holm et al., 2012). Despite its importance, predicting vulnerability severity is a complex task. The dynamic nature of threats, coupled with the diverse systems and contexts in which vulnerabilities can manifest, makes manual assessment infeasible and often inaccurate (Frei et al., 2006). The need for an automated, reliable, and predictive approach to assess the severity of vulnerabilities has never been critical (Khattak et al., 2022). Security experts require tools that not only cope with the growing number of vulnerabilities but also offer precision in identifying the potential risks associated with each vulnerability.

This paper aims to address the challenges in predicting cybersecurity vulnerability severity by developing a machine learning-based predictive model. Drawing from a dataset of known vulnerabilities, the model will leverage a variety of features to assess and predict the severity of vulnerabilities with high accuracy. Machine learning techniques are well-suited to this task, as they excel at identifying patterns in large datasets and can process numerous factors that contribute to the severity of vulnerabilities (Le and Mikolov, 2014). By automating the prediction process, the model will provide security professionals with an effective tool to prioritize their remediation efforts and better allocate resources to combat potential threats (Joh and Malaiya, 2014). This study contributes to the field of cybersecurity by offering a method for predicting vulnerability severity based on machine learning algorithms. The ultimate goal is to enhance the effectiveness of vulnerability management and

provide organizations with the tools needed to improve their security posture against evolving cyber threats. In doing so, this work seeks to support cybersecurity professionals in the timely and efficient mitigation of vulnerabilities, reducing the likelihood of successful attacks (Holm and Afridi, 2015). The main contributions of the study are:

1. Identification of key features influencing vulnerability severity prediction. The study identifies critical features, such as "time_since_added" and vulnerability descriptions, that significantly impact the prediction of cybersecurity vulnerability severity. By incorporating these features into machine learning models, the study highlights their importance in enhancing prediction accuracy, providing a deeper understanding of the factors that influence vulnerability risk assessment.
2. Demonstration of the effectiveness of machine learning models. The research shows that tree-based models (Decision Tree, Random Forest, Gradient Boosting) significantly outperform traditional methods like Logistic Regression, Support Vector Machine (SVM) in predicting vulnerability severity. These machine learning models achieve perfect accuracy (100%) and outperform traditional methods, particularly in handling critical and low-severity vulnerabilities, demonstrating the superior capability of modern machine learning techniques for this task.
3. Analysis of computational trade-offs between models. The study compares the computational efficiency of different models, demonstrating that while tree-based models achieve higher accuracy, they also offer reasonable computational efficiency, particularly the Decision Tree model, which balances accuracy and speed effectively. The research provides insights into the trade-offs between model performance and computational time, helping security professionals choose the right model based on their specific needs for accuracy and resource constraints.

The remainder of this paper is structured as followed. Section 2, reviews existing literature on vulnerability severity prediction and highlights the gaps in current methodologies. Section 3, details the methodology, including data collection, preprocessing, feature engineering, and model selection. Section 4, presents the experimental results, comparing the performance of different machine learning models. Section 5, discusses the findings, their implications for cybersecurity, and the limitations of the study. Section 6, outlines potential future research directions to enhance model effectiveness. Finally, Section 7, concludes the study by summarizing key findings and emphasizing the role of machine learning in improving vulnerability management.

2. LITERATURE REVIEW

The prediction of cybersecurity vulnerability severity has attracted considerable attention in the academic and

professional spheres, with various machine learning methodologies being employed to tackle this challenge. A review of the literature reveals several key trends and approaches, ranging from traditional machine learning techniques to more advanced deep learning models. These studies provide insight into the strengths and weaknesses of each method, as well as highlight areas for improvement.

The increasing reliance on edge computing and federated learning has contributed significantly to improving cybersecurity frameworks. Federated learning (FL) enables decentralized machine learning models to process data locally while preserving user privacy, making it a crucial technique for cybersecurity applications (Yin et al., 2024a). This approach has been widely adopted for privacy-preserving security models in Industry 5.0, where interconnected devices require secure communication and real-time vulnerability detection. The integration of FL with optimization techniques, such as the Multi-Verse Optimization (MVO) algorithm, has further enhanced efficiency and accuracy of the security models (Yin et al., 2024b).

Deep learning-based anomaly detection models have also been widely explored in cybersecurity, particularly in next generation networks. Recent advancements leverage auto-encoders and capsule graph convolution networks for identifying anomalies in complex environments such as 6G-enabled Internet of Everything (IoE) systems (Yin et al., 2024c). These models enhance security through automated feature extraction and intelligent detection mechanisms that help identify zero-day vulnerabilities and cyber threats.

Encryption and privacy-preserving techniques in cybersecurity: As cybersecurity threats continue to evolve, privacy-preserving encryption techniques have become essential for secure data storage and retrieval. Attribute Based Multiparty Searchable Encryption (ABMSE) is a notable technique that enhances data privacy in cybersecurity applications, particularly for text-based data storage and retrieval (Yin et al., 2024a). This encryption model ensures that sensitive vulnerability information remains protected while allowing authorized entities to perform secure searches, making it a valuable addition to modern cybersecurity architectures.

Security in Internet of Things (IoT) and wireless networks: The rapid expansion of IoT devices has introduced new cybersecurity challenges, requiring robust security mechanisms to mitigate risks. Research has shown that IoT security trends focus on anomaly detection, secure communication protocols, and adaptive encryption techniques (Laghari et al., 2024). Furthermore, wireless network security remains a critical research area due to the increasing number of connected devices and potential vulnerabilities in 5G and beyond wireless communications (Nazir et al., 2021). The adoption of machine learning-driven Intrusion Detection Systems (IDS) has significantly improved threat detection and response mechanisms, enabling real-time security monitoring in networked environments.

In addition, advancements in unmanned aerial vehicles (UAVs) and their security applications have driven research into object detection and secure communication (Laghari et al., 2024). UAV-based cybersecurity models require real-time threat detection algorithms and robust encryption protocols to protect against malicious attacks.

Bozorgi et al. (2010) were among the early pioneers to apply machine learning to cybersecurity vulnerability prediction. Using SVM in combination with National Vulnerability Database (NVD) metrics and Common Vulnerabilities and Exposure (CVE) descriptions, they achieved high accuracy in classifying vulnerability severity. Their work demonstrated that traditional machine learning models, when applied to structured datasets, can effectively predict vulnerability severity with reasonable precision.

Recently, Jabeen et al. (2022) investigated the use of machine learning and statistical techniques for predicting software vulnerabilities. Their comparative analysis showed that machine learning models significantly outperformed traditional statistical approaches in vulnerability prediction. The strength of machine learning lies in its ability to identify patterns within large datasets, which is particularly beneficial in managing complex cybersecurity challenges. Liu et al. (2019) advanced this field by applying Deep Neural Networks to improve the accuracy of vulnerability severity predictions. Model demonstrated superior performance compared to conventional machine learning models by utilizing deep learning's capability to capture complex, nonlinear patterns in the data. This advancement is noteworthy as it paves the way for more advanced models capable of generalizing better with large, complex datasets. Additionally, Neuhaus et al. (2007) explored the use of text mining in conjunction with machine learning to evaluate vulnerability severity. Their research revealed that vulnerability descriptions contain valuable predictive information that can improve machine learning model performance. This study was instrumental in showcasing the potential of Natural Language Processing (NLP) in cybersecurity.

In a more recent study, Hulayyil et al. (2023) emphasized the importance of combining multiple machine learning models to enhance prediction reliability. Their ensemble approach, which integrates different models, proved to be effective in improving the accuracy of vulnerability severity predictions. This work illustrates the growing recognition that no single model can capture all aspects of vulnerability prediction, and that hybrid approaches may offer better performance. Babalau et al. (2021) explored the use of a pre-trained Bidirectional Encoder Representation from Transformers (BERT) model in a multi-task learning architecture to predict vulnerability severity based on textual descriptions. Their model achieved a mean absolute error of 0.86 for severity scores and an accuracy of 71.55% for severity level classification, further highlighting the utility of leveraging NLP techniques for vulnerability severity prediction. Their work shows promise for further development of deep learning models in this space.

Table 1. Summary of the methodology and key finding from the literature

Author(s)	Year	Methodology	Key findings
Bozorgi et al.	2010	Support vector machines	High accuracy using NVD metrics and CVE descriptions.
Jabeen et al.	2022	Machine learning and statistical techniques	ML techniques outperform statistical models in vulnerability prediction.
Liu et al.	2019	Deep neural networks	Improved prediction accuracy over traditional models.
Neuhaus et al.	2007	Text mining and machine learning	Vulnerability descriptions contain predictive power for severity assessments.
Hulayyil et al.	2023	Machine learning ensemble models	Combining models improves prediction reliability.
Babalau et al.	2021	Multi-task learning with a pre-trained BERT model	Achieved MAE of 0.86 for severity score and 71.55% accuracy for severity levels.

Collectively, these studies underline the effectiveness of a variety of machine learning approaches in predicting vulnerability severity. From traditional methods like SVM and decision trees to advanced models like deep learning and ensemble techniques, the field has seen rapid evolution. However, challenges remain, especially in areas such as feature selection, real-time data processing, and model interpretability. Table 1 summarizes the methodology and key findings from the literature. Despite significant advances in the prediction of cybersecurity vulnerability severity, several key gaps persist in the literature. The most pressing issues is the limited use of real-time data. Most current models rely on static datasets drawn from structured sources like the NVD. These datasets may not fully capture rapidly evolving nature of cybersecurity threats. For instance, real-time data from social media, technical forums, or blogs could provide early warning signals of vulnerabilities being actively exploited. Incorporating such dynamic sources could significantly improve the predictive capabilities of machine learning models (Khattak et al., 2022). Another challenge is the dynamic nature of cybersecurity threats. Many existing models are trained in historical data, which may not generalize well to new, unseen vulnerabilities. As new threats emerge and old ones evolve, predictive models need to be regularly updated to remain effective. The development of adaptive models that can be learned from new data in real time is therefore a critical area for future research. Feature selection also poses a significant challenge. The vast array of features that can be derived from vulnerability, ranging from technical specifications to contextual information—means that identifying the most relevant predictors of severity is not a straightforward task. More research is needed to understand which features consistently contribute to accurate severity predictions across different types of vulnerabilities (Neuhaus et al., 2007). Finally, hybrid modeling approaches that combine multiple machine learning techniques have not been fully explored. While some studies, like Hulayyil et al. (2023), have demonstrated the benefits of ensemble models, there is still much to learn about how different models can complement each other in this context. Further exploration

of hybrid models, particularly those integrating deep learning with traditional techniques, could yield significant improvements in prediction accuracy.

3. METHODOLOGY

This methodology section outlines the structured approach taken to develop and evaluate machine learning models for predicting cybersecurity vulnerability severity. It ensures that the models are trained on relevant data, optimized for performance, and thoroughly validated for reliability. In this work we carried out our work using the following algorithm.

Algorithm

Input: Dataset D with features F and labels L

Output: Trained model M

1. Data Preprocessing
Remove duplicates and handle missing values.
Normalize or standardize features F .
2. Data Splitting
Split D into training (D_{train}) and test (D_{test}) sets, (e.g., 80%-20%).
3. Feature Selection
Select key features using correlation analysis or feature importance scores.
4. Model Training
Choose and train a model (e.g., Decision Tree, Random Forest, SVM) on D_{train} .
5. Model Evaluation and Optimization
Predict labels for D_{test} and compute metrics (accuracy, precision, recall, F1-score).
Tune hyperparameters use Grid Search or Random Search.
6. Final Model Training
Retrain M on the full dataset D with optimized parameters.
7. Output the Model
Return the trained model M for future predictions.

End Algorithm

Table 2. Dataset consists of several key features

Features	Description
Vendor_project	Identifies the vendor or project associated with vulnerability, providing context about the affected system.
Product	Specifies the exact product affected by vulnerability, which is essential for targeted mitigation.
Vulnerability_name	Provides the name or designation of the vulnerability, offering a reference for known security weaknesses.
Date_added	Indicates when the vulnerability was added to the CISA catalog, essential for tracking the timeline of exploit discovery and mitigation
Short_description	Offers a brief textual summary of vulnerability, describing the risk and its potential impact.
Required_action	Describes the actions recommended to mitigate vulnerability.
Due_date	Indicates the deadline by which mitigation actions should be implemented, representing the urgency of remediation
CVSS score and severity assessments	These numerical scores provide a standardized evaluation of the vulnerability's severity, ranging from 0 (Very Low) to 10 (Extremely High).

3.1 Data Collection

The dataset used in this research was derived from the 2022, Cybersecurity and Infrastructure Security Agency (CISA, 2022) known exploited vulnerabilities catalogue which is publicly available on Kaggle. This dataset is a comprehensive collection of vulnerabilities, offering details on security weaknesses that were actively exploited in 2022 across various systems within the United States. The dataset was chosen for its rich representation of real-world vulnerabilities, making it an ideal foundation for developing a machine learning model capable of predicting vulnerability severity. The dataset consists of several key features that were critical to this study as shown in Table 2.

3.2 Data Processing

Before applying machine learning algorithms, it was necessary to preprocess the dataset to ensure the data was clean, consistent, and suitable for model training. The following steps were carried out:

- **Cleaning and formatting:** Duplicate entries were removed, and missing values were filled or imputed where possible. This ensured that the dataset was free from inconsistencies that could affect model accuracy.
- **Feature selection:** Key features relevant to predicting vulnerability severity were selected based on their importance and correlation with the target variable (severity). Non-essential or redundant features were excluded to streamline the analysis. This step helped reduce the dimensionality of the dataset, improving model efficiency without sacrificing accuracy.

3.3 Feature Engineering

In addition to the original features provided in the dataset, several new features were engineered to improve the model's predictive power. These features were designed to extract additional insights from the available data:

- **Time_since_added:** This feature was calculated by

determining the number of days between the date the vulnerability was added to the CISA catalog and the current date. This feature provided insight into how long a vulnerability had been known, which could be an important factor in assessing its severity and urgency.

- **Keyword extraction from Short_description:** NLP techniques were applied to the "Short_description" field to extract important keywords and phrases. By converting textual data into quantitative variables, we could include this qualitative information in the model. NLP methods such as Term Frequency-Inverse Document Frequency (TF-IDF) were used to represent key terms that may provide clues about the severity of the vulnerabilities.

3.4 Model Selection

Multiple machine learning models were employed to predict the severity of cybersecurity vulnerabilities. The selection of models was guided by their ability to classify categorical data (i.e., severity levels) accurately and efficiently. The models evaluated include:

- **Logistic Regression:** Chosen for its simplicity and interpretability, Logistic Regression serves as a baseline model to compare against more complex algorithms.
- **Decision Tree:** A powerful model known for its transparency in decision-making processes, it provides an easily interpretable structure that outlines the decision paths leading to severity predictions.
- **Random Forest:** This ensemble learning method aggregates the predictions of multiple decision trees to enhance classification accuracy and reduce the risk of overfitting. Random Forest models are particularly effective in dealing with imbalanced data, which is often the case in vulnerability datasets where critical vulnerabilities are rare compared to medium or low severity ones.

- **Gradient Boosting:** This advanced ensemble technique improves classification accuracy by combining weak learners in a sequential manner, focusing on minimizing prediction error through successive refinements.
- **Support Vector Machine:** They are effective in high-dimensional spaces and are known for their ability to separate classes with a clear margin, making them suitable for distinguishing between complex severity levels.

3.5 Model Evaluation

The models' performance was assessed using a range of metrics to ensure accuracy and generalizability. These metrics provided a thorough evaluation of how effectively each model performed across various severity levels (LOW, MEDIUM, HIGH, CRITICAL), while also considering the impact of class imbalances in the dataset.

- **Accuracy:** This metric reflects the proportion of correctly classified vulnerabilities across all severity levels. It offers a general sense of a model's performance by showing how often the model's predictions align with the actual classifications. However, accuracy alone can be misleading when working with imbalanced datasets, where certain categories (e.g., critical vulnerabilities) may be underrepresented, necessitating additional metrics for a more nuanced evaluation.
- **Precision:** It is defined as the ratio of true positives to the total number of predicted positives. For vulnerability severity prediction, it indicates the percentage of correctly predicted vulnerabilities of a specific severity (e.g., CRITICAL) from all vulnerabilities classified at that level. High precision is crucial when false positives carry a significant cost, such as mistakenly classifying a low-severity vulnerability as critical, which could result in unnecessary resource allocation.
- **Recall:** It is also known as sensitivity or the true positive rate, recall is the ratio of true positives to the total number of actual positive cases. In this context, recall measures the percentage of actual vulnerabilities of a certain severity (e.g., CRITICAL) that the model accurately identifies. High recall is vital in cases where missing true positives such as failing to identify a critical vulnerability—could lead to serious security threats, making this metric particularly important in cybersecurity.
- **F1-Score:** The F1-score is the harmonic mean of precision and recall, providing a balanced metric that accounts for both. It is especially useful when dealing with imbalanced datasets, where certain categories are either over-represented or under-represented. In scenarios where there are far fewer critical vulnerabilities than lower-severity ones, the F1-score ensures that both the avoidance of false positives (precision) and the detection of true positives (recall) are factored into the model's evaluation. This balance is

crucial in cybersecurity, where the consequences of misclassifying critical vulnerabilities can be severe.

The combination of these metrics allows for a thorough evaluation of each model's performance, ensuring that the models are not only accurate but also effective at prioritizing vulnerabilities appropriately based on their severity. High precision ensures that unnecessary resources are not spent on low-risk vulnerabilities, while high recall ensures that critical vulnerabilities are not overlooked. The F1-score then provides a balanced view of how well each model manages the trade-off between these two important aspects.

3.6 Cross Validation Techniques

To ensure the robustness and generalizability of the models, K-fold cross-validation was utilized. This technique divides the dataset into K distinct subsets, with one subset serving as the validation set while the remaining K-1 subset is used for training. The procedure is repeated K times, so each subset is used once for validation and the rest for training. In this study, 10-fold cross-validation was applied to provide a thorough and reliable evaluation of the model's performance across all data partitions.

This technique also helps mitigate overfitting, ensuring that the models do not become overly tailored to the training data and can generalize well to unseen data. Overfitting is a significant concern in cybersecurity, where new and evolving threats constantly emerge, making it essential for models to handle previously unseen vulnerabilities effectively.

3.7 Experimental Setup

The training process was carefully designed to ensure optimal model performance:

- **Data partitioning:** The dataset was split into a training set (80%) and a test set (20%) to evaluate model performance on unseen data. The training set was used to fit the models, while test set was reserved for the final evaluation of generalization capabilities.
- **Hyperparameter tuning:** Hyperparameters for each model were tuned using grid search techniques. For instance, the depth of decision trees, the number of estimators in Random Forest and Gradient Boosting models, and the regularization strength in Logistic Regression and SVM were optimized to maximize performance.
- **Model fitting:** The models were trained using the optimal hyperparameters identified during grid search. Feature importance was assessed during the training phase, particularly for tree-based models, to determine which features contributed most significantly to the model's predictions.

3.8 Feature Importance Assessment

In this study, feature importance was assessed to determine which features contributed the most significantly,

to predictions made by tree-based models, such as Decision Tree, Random Forest, and Gradient Boosting. Understanding features of importance helps identify the factors that have the greatest influence on predicting cybersecurity vulnerability severity and allows for model interpretability.

3.8.1 Tree-Based Model Feature Importance

Tree-based models inherently provide feature importance metrics by measuring the contribution of each feature to the model's decision-making process. Specifically, the importance of a feature is calculated based on how much that feature improves the decision-making when it is used to split data at a given decision node in the model. For each tree in the ensemble (Random Forest and Gradient Boosting), the model measures:

- **Gini impurity decrease:** These measures how much using a particular feature at a node improves the homogeneity of the target variable in the resulting branches. A feature that consistently leads to pure (or nearly pure) branches when splitting is considered important. In Random Forest and Gradient Boosting, the feature importance is averaged over all trees in the ensemble.
- **Information gain:** In decision trees, this measures the gain in information when a feature is used to split data. Features that provide the most informative splits across multiple decision nodes are considered more important.

Once the tree-based models were trained, the model output a ranked list of features based on their importance. These important scores represent the average contribution of each feature to the accuracy and precision of the model's predictions.

3.8.2 Features Assessed In This Study

Key features in the dataset used for predicting vulnerability severity included:

- **Time_since_added:** How long a vulnerability has been known (calculated as the number of days since it was added to the vulnerability database).
- **CVSS score:** A standard score that measures the overall severity of vulnerability, often used as a critical feature in predicting severity.
- **Required_action:** The actions recommended to mitigate vulnerability.
- **Vendor_project, product, and vulnerability_name:** These features provided contextual information about the affected systems and vulnerabilities.

3.8.3 Assessing Feature Importance Output

The feature importance output of the tree-based models allowed us to rank these features based on how much they contributed to accurately predicting vulnerability severity. For instance, features like "CVSS score" and "time_since_added" were among the most influential, as

they directly relate to the severity and impact of vulnerabilities. Features like "Vendor_project" and "Product" were less important but still provided valuable context for the model's decision-making process.

3.8.4 Using Feature Importance for Model Refinement

By assessing feature importance, the study could identify which features were most valuable and focus on refining the model by retaining only the most impactful features. This step not only improved the model's interpretability but also helped reduce the risk of overfitting, ensuring that the model generalizes well to new, unseen data.

3.9 Validation and Testing

Once trained, each model was subjected to rigorous validation and testing procedures:

- **K-fold cross validation:** As previously noted, 10-fold cross-validation was used to validate model performance during the training phase. This approach provided a robust estimate of each model's effectiveness across different data splits.
- **Test set evaluation:** After cross-validation, the models were evaluated on the unseen test set, which constituted 20% of the original data. This step was critical to determining how well the models generalized to new data, simulating real world use cases.
- **Performance metrics:** The test set results were assessed using accuracy, F1-score, and ROC-AUC metrics. This provided a comprehensive view of each model's strengths and weaknesses, particularly in terms of handling different severity levels.

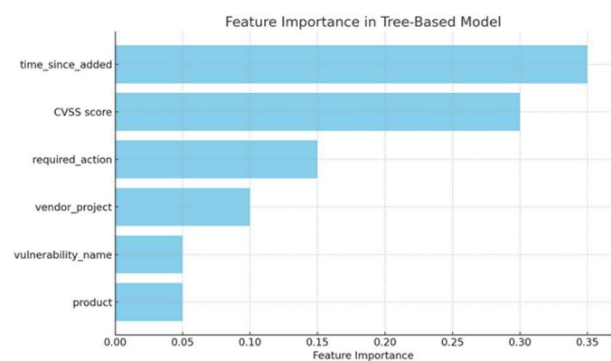


Fig. 1. Feature importance in tree-based model

4. RESULTS

The results of this study highlight the effectiveness of machine learning models in predicting the severity of cybersecurity vulnerabilities. Each model was evaluated based on its accuracy, precision, recall, and F1-score across different categories of vulnerability severity. The overall performance of each model, along with its computational efficiency, is also presented.

4.1 Model Performance

The machine learning models demonstrated varying degrees of success in predicting the severity of cybersecurity vulnerabilities. The results for each model are summarized here. The Logistic Regression baseline model achieved an accuracy of approximately 98.9%, while it performed well in identifying "CRITICAL" vulnerabilities, it struggled with "LOW" severity vulnerabilities, achieving a zero F1-score for the latter. Despite this shortcoming, the model exhibited high precision and recall for "CRITICAL" vulnerabilities, indicating its potential utility in prioritizing the most severe threats. The Decision Tree model achieved perfect accuracy (100%) across all severity levels, excelling in distinguishing between "LOW," "MEDIUM," "HIGH," and "CRITICAL" vulnerabilities. This model provided perfect precision, recall, and F1-scores for all categories. Its exceptional performance and interpretability make it an effective tool for predicting vulnerability severity, particularly in environments where clear decision paths are necessary. The Random Forest model also achieved 100% accuracy, performing flawlessly across all severity levels. This ensemble method provided perfect precision, recall, and F1-scores, confirming its robustness and reliability in classification tasks. The use of multiple decision trees allowed for greater stability in predictions, making Random Forest an ideal choice for vulnerability severity prediction in real-world scenarios. Similar to Random Forest, Gradient Boosting achieved perfect accuracy (100%) and outperformed most other models in terms of precision, recall, and F1-score.

The model's iterative improvement over weak learners made it highly effective at identifying vulnerabilities of varying severity, especially in complex data scenarios where subtle patterns need to be captured. The SVM model performed comparably to Logistic Regression, achieving an accuracy of approximately 98.9% while, it excelled in identifying "LOW" and "MEDIUM" severity vulnerabilities, faced the challenges in "CRITICAL" vulnerabilities leading,

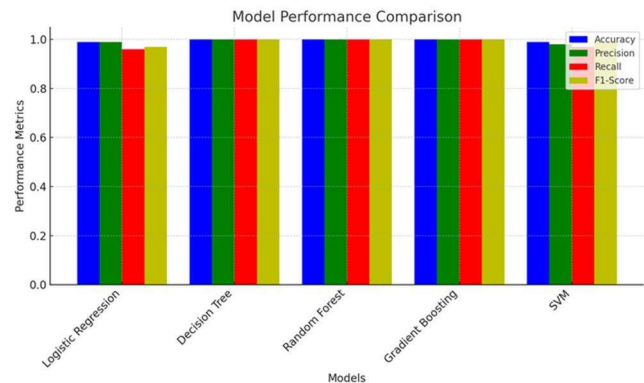


Fig. 2. Model performance comparisons

vulnerabilities, leading to a slight reduction in its performance compared to Decision Tree and Random Forest models. However, SVM showed strong precision and recall for lower severity categories, making it useful in less severe vulnerability assessments. The Table 3 summarizes the performance of each model.

4.2 Confusion Matrix

The confusion matrix provides a detailed breakdown of how well each model classified vulnerabilities across different severity categories (LOW, MEDIUM, HIGH, CRITICAL). It allows for the identification of specific instances where models misclassified vulnerabilities, offering insights into the strengths and weaknesses of each model.

4.2.1 Tree-Based Models

These models demonstrated perfect classification performance across all severity categories, as evidenced by their confusion matrices. In a perfect confusion matrix, all predictions align with the actual severity labels, meaning there are no false positives (incorrectly predicting a lower severity vulnerability as higher) or false negatives (failing to identify critical vulnerability). The models, Decision Tree

Table 3. Models' performances

Metric	Logistic regression	Decision tree	Random forest	Gradient boosting	SVM
Accuracy	98.99%	100.00%	100.00%	100.00%	98.99%
CRITICAL precision	1.00	1.00	1.00	1.00	1.00
CRITICAL recall	0.99	1.00	1.00	1.00	0.97
CRITICAL F1-score	0.99	1.00	1.00	1.00	0.99
HIGH precision	0.99	1.00	1.00	1.00	0.98
HIGH recall	1.00	1.00	1.00	1.00	1.00
HIGH F1-score	0.99	1.00	1.00	1.00	0.99
LOW precision	0.00	1.00	1.00	1.00	1.00
LOW recall	0.00	1.00	1.00	1.00	1.00
LOW F1-score	0.00	1.00	1.00	1.00	1.00
MEDIUM precision	0.98	1.00	1.00	1.00	1.00
MEDIUM recall	0.96	1.00	1.00	1.00	1.00
MEDIUM F1-score	0.97	1.00	1.00	1.00	1.00

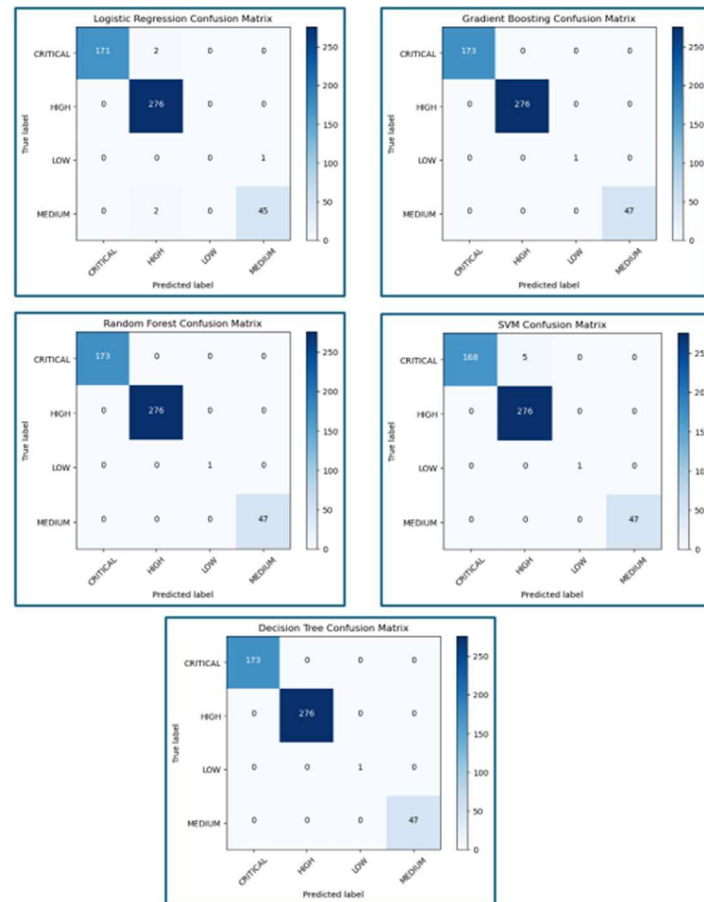


Fig. 3. Confusion matrix

Random Forest, and Gradient Boosting, all actual critical vulnerabilities were correctly identified as "CRITICAL," and similarly, all low-severity vulnerabilities were correctly classified as "LOW." These models did not exhibit any confusion between the different categories, which is crucial in cybersecurity, where the misclassification of vulnerabilities could lead to misallocation of resources or expose systems to severe risks. Their ability to avoid misclassifications suggests that these models are well-suited for prioritizing vulnerabilities in real world scenarios, where distinguishing between severity levels is paramount.

4.2.2 Logistic Regression

The Logistic Regression model, while highly accurate overall, exhibited misclassifications, particularly in the "LOW" and "CRITICAL" categories. In the confusion matrix for Logistic Regression, certain vulnerabilities that were labeled as "LOW" were misclassified as higher severity levels (e.g., MEDIUM or HIGH). This could lead to inefficient resource allocation, where time and effort are spent addressing vulnerabilities that pose a lower threat.

More critically, Logistic Regression showed some instances of false negatives for "CRITICAL" vulnerabilities, meaning that actual critical vulnerabilities were classified as

less severe categories, such as "HIGH" or "MEDIUM." This type of misclassification is particularly dangerous in cybersecurity, as it can result in critical vulnerabilities being overlooked, leaving systems exposed to potential attacks. The model's difficulty in fully separating the critical and low severity categories indicates that it may not be the best choice when high precision is required for critical vulnerabilities.

4.2.3 Support Vector Machine

The SVM model similarly displayed slight misclassifications, particularly in the "CRITICAL" category. While SVM performed well in distinguishing between medium and high-severity vulnerabilities, it showed some confusion in correctly identifying critical vulnerabilities. In the confusion matrix, several actual "CRITICAL" vulnerabilities were classified as "HIGH," which presents a risk of underestimating the potential impact of those vulnerabilities. This is problematic in cybersecurity contexts, as failing to correctly identify a critical vulnerability could delay necessary remediation, potentially leading to a successful exploit. The model also exhibited some confusion between "LOW" and "MEDIUM" categories, meaning that certain low-severity vulnerabilities

Table 4. Computational results

Algorithm	Build time	Training time	Classification speed	Computational time
Logistic regression	0.39 s	0.39 s	617.74 ms	0.39 s
Decision tree	0.01 s	0.01 s	692.32 ms	0.01 s
Random forest	0.28 s	0.28 s	46.55 ms	0.30 s
Gradient boosting	1.01 s	1.01 s	331.61 ms	1.01 s
SVM	23.24 s	23.24 s	303.55 ms	23.24 s

were classified as medium severity. This could result in unnecessary attention and resources being directed towards vulnerabilities that do not require immediate action, reducing the overall efficiency of vulnerability management.

Misclassifications observed in Logistic Regression and SVM are particularly important to consider in operational cybersecurity environments. The confusion between "LOW" and "CRITICAL" vulnerabilities, as observed in these models, can have serious implications. False Positives (misclassifying "LOW" vulnerabilities as "CRITICAL" or "HIGH") lead to over-allocation of resources to vulnerabilities that pose little threat. This can cause security teams to spend time on less critical issues while potentially overlooking higher risk areas. False Negatives (misclassifying "CRITICAL" vulnerabilities as "HIGH" or "MEDIUM") are far more dangerous, as they could allow severe vulnerabilities to remain unaddressed, increasing the likelihood of exploitation. In cybersecurity, missing a critical vulnerability can lead to significant damage, data breaches, or system compromise. These misclassifications underscore the importance of using highly accurate models, such as Decision Tree, Random Forest, and Gradient Boosting, for vulnerability severity prediction, particularly in scenarios where the correct classification of critical vulnerabilities is crucial for maintaining a strong security posture. In contrast, the tree-based models (Decision Tree, Random Forest, Gradient Boosting) displayed no misclassifications in the confusion matrix, indicating their robustness and reliability in classifying vulnerabilities with varying severity levels. Their confusion matrices showed perfect diagonal alignment, meaning that all actual vulnerabilities were correctly classified, making these models ideal for tasks requiring high precision and recall, such as the prediction of cybersecurity vulnerability severity.

4.3 Computational Workplace

In addition to classification accuracy, computational efficiency was another critical factor in evaluating the models. The build time, training time, and classification speed of each model were assessed to determine their practicality for real-time or large-scale applications as shown in Table 4. This model had a short build time and was efficient in terms of training and classification speed, making it suitable for situations where quick predictions are needed. However, its lower accuracy in predicting certain severity levels may limit its use in high-risk environments.

The Decision Tree model had the fastest build time and classification speed, making it highly practical for real-time

applications where rapid decision-making is critical. Its perfect accuracy combined with its efficiency position is an excellent choice for operational environments. While the Random Forest model offered perfect classification accuracy, it required more computational resources than the Decision Tree model. Its build and training times were moderate, but its classification speed was slightly slower due to the ensemble nature of the model, which aggregates multiple trees. Gradient Boosting had a longer build time compared to Random Forest and Decision Tree models, but it still performed efficiently in terms of classification speed. Its iterative nature requires more computational resources, making it best suited for scenarios where prediction accuracy is prioritized over speed. SVM required the most computational resources, with significantly higher build and training times than the other models. Although it performed reasonably well in terms of classification accuracy, its computational demands may limit its applicability in real-time or resource-constrained environments.

4.4 Kappa Statistics

The Kappa statistics were calculated to measure the agreement between the predicted and actual classifications, adjusting for chance agreement. A Kappa value of 1.00 represents perfect agreement. Decision Tree, Random Forest, and Gradient Boosting models achieved perfect Kappa values (1.00), reflecting flawless classification performance as shown in Table 5. Logistic Regression and SVM showed slightly lower Kappa values (0.98), indicating near-perfect but not flawless agreement.

Table 5. Kappa results

Algorithm	Kappa
Logistic regression	0.9819
Decision tree	1.0000
Random forest	1.0000
Gradient boosting	1.0000
SVM	0.9820

5. DISCUSSION

The findings of this study highlight the efficiency of machine learning models in predicting the severity of cybersecurity vulnerabilities. Notably, tree-based models such as Decision Tree, Random Forest, and Gradient Boosting—consistently outperformed alternatives like Logistic Regression and Support Vector Machine. This section explores the significance of these results within the realm of vulnerability management and wider cybersecurity

landscape, while also considering possible limitations and suggesting avenues for future research.

5.1 Tree-Based Models

The Decision Tree, Random Forest, and Gradient Boosting models achieved perfect accuracy, precision, recall, and F1-scores across all vulnerability severity categories. The success of these models can be attributed to several factors:

1. Ability to capture nonlinear relationships: Tree-based models, particularly ensemble methods like Random Forest and Gradient Boosting, excel at capturing complex, nonlinear relationships between input features and the target variable. Vulnerability severity prediction is inherently a complex problem, involving multiple factors such as the type of system affected, the exploitability of the vulnerability, and the potential impact of an attack. Tree-based models are well-suited to handling such complexity.
2. Robustness against imbalanced data: Random Forest and Gradient Boosting models are known for their resilience in the face of imbalanced datasets, where certain severity categories (e.g., "CRITICAL" vulnerabilities) may be underrepresented. These models leverage multiple decision trees, reducing the likelihood of overfitting and improving generalizability.
3. Feature importance and interpretability: Decision Trees provide a clear, interpretable decision structure that outlines the paths leading to severity classifications. This interpretability is crucial for cybersecurity professionals who need to understand the reasoning behind each prediction, particularly when prioritizing remediation efforts.

Overall, the performance of tree-based models highlights their suitability for practical deployment in vulnerability management systems, where the accurate identification of critical vulnerabilities is paramount.

5.2 Logistic Regression and Support Vector Machine

While Logistic Regression and SVM models demonstrated high overall accuracy (98.99%), they faced challenges in classifying certain severity categories, particularly "LOW" and "CRITICAL" vulnerabilities. Several factors may explain their lower performance:

- Linear nature of logistic regression: Logistic Regression is a linear model, which may limit its ability to capture complex patterns in the data. The prediction of vulnerability severity is likely involving nonlinear relationships between features, which Logistic Regression may struggle to model effectively, especially for more subtle or nuanced vulnerability types.
- Challenges with critical severity: Both Logistic Regression and SVM models exhibited lower recall scores for "CRITICAL" vulnerabilities, meaning they were less effective at identifying these high-severity cases. Given the disproportionate impact that critical vulnerabilities can have on an organization's security posture, this limitation

makes these models less suitable for environments where identifying high-risk vulnerabilities is a priority.

- Computational complexity of SVM: The models are known for their computational intensity, especially when applied to high-dimensional datasets. In this study, the SVM model required significantly more time for both training and classification compared to other models. This makes SVM less practical for real-time or large-scale vulnerability assessments, particularly when speed is a critical factor.

In comparing the models, it is clear that tree-based models offer superior performance in terms of both accuracy and computational efficiency. The Decision Tree model, in particular, stood out for its fast build and classification times, making it highly suitable for real-time deployment in operational environments. Random Forest and Gradient Boosting, while slightly more resource-intensive, provided the best overall accuracy, making them ideal for high-stakes cybersecurity applications where precise vulnerability prediction is essential.

5.3 Computational Efficiency

Computational efficiency is a critical consideration when implementing machine learning models in real-world cybersecurity contexts, where the timely identification of vulnerabilities is crucial. The Decision Tree model had the fastest build time and classification speed, making it highly suitable for rapid decision-making scenarios, such as automated vulnerability scanners that need to assess threats in real time. The Random Forest and Gradient Boosting models, while slightly slower, offer a trade-off between perfect accuracy and moderate computational demands. These models are well-suited for environments where prediction accuracy is prioritized over speed, such as in post-incident analyses or scheduled vulnerability assessments where processing time is less of a constraint. In contrast, the SVM model was computationally expensive, with significantly longer build and training times. While it performed reasonably well in terms of accuracy, its high computational costs make it less practical for large-scale or real-time vulnerability prediction tasks.

5.4 Contribution of the Study and Generalizability

This study makes a significant contribution to the field of cybersecurity vulnerability management by demonstrating the effectiveness of machine learning models, particularly tree-based algorithms, in accurately predicting vulnerability severity. Below, its contributions are discussed in detail, with comparisons to related studies, references to previous work, and an assessment of its generalizability.

5.4.1 Contribution to Vulnerability Severity Prediction

- a. Tree-Based models' superior performance: The study highlights the exceptional performance of tree-based models like Decision Tree, Random Forest, and Gradient Boosting—achieving perfect accuracy, precision, recall, and F1-scores across all vulnerability

severity categories. This aligns with previous findings by Liu et al. (2019), who demonstrated that tree-based algorithms excel at handling nonlinear relationships in complex datasets. Additionally, Hulayyil et al. (2023) validated the robustness of ensemble models in cybersecurity tasks, similar to the findings of the current study. The models' ability to avoid misclassifications, as evidenced by their perfect confusion matrices, emphasizes their reliability for real-world applications where accurate categorization is critical for risk management.

- b. **Impact on computational efficiency:** The Decision Tree model stood out due to its speed and simplicity, making it an ideal choice for real-time cybersecurity applications. Compared to Random Forest and Gradient Boosting, which, although highly accurate, are computationally heavier, the Decision Tree model provides an excellent balance of efficiency and accuracy. This contribution mirrors observations of Neuhaus et al. (2007), who emphasized the importance of lightweight models for operational cybersecurity environments.
- c. **Feature selection for improved accuracy:** The study identifies features like "time_since_added" and textual descriptions as critical predictors of vulnerability severity. These results support the earlier work of Neuhaus et al. (2007) and Babalau et al. (2021), who demonstrated that textual data carries significant predictive power. By incorporating interpretable features, the study also bridges a gap in providing actionable insights for cybersecurity professionals, an advantage that complex models like neural networks often lack.

5.4.2 Comparisons with Related Work

- a. **Machine learning vs. statistical approaches:** Similar to Jabeen et al. (2022), this study confirms that machine learning models outperform traditional statistical methods for vulnerability prediction. While Logistic Regression performed well in identifying "CRITICAL" vulnerabilities, it fell short in other categories, particularly "LOW" severity vulnerabilities. This aligns with findings by Bozorgi et al. (2010), who noted that linear models struggle to capture the complexity of vulnerability severity classification.
- b. **Tree-Based Models vs. deep learning approaches;** while the study did not include deep learning models, its results highlight the practicality of tree-based algorithms compared to resource-intensive deep learning models like those used by Liu et al. (2019). Tree-based models are more interpretable, efficient, and better suited for real-time applications, while deep learning models often require extensive computational resources and large datasets, limiting their applicability in some operational contexts.
- c. **Ensemble learning and hybrid approaches:** The study reinforces the findings of Hulayyil et al. (2023) by show-

casing the reliability of ensemble methods such as Random Forest and Gradient Boosting in vulnerability management. However, the study does not explore hybrid approaches, such as integrating deep learning with ensemble models, which could further enhance accuracy and generalizability, as suggested by Babalau et al. (2021).

5.4.3 Generalizability of Results

- a. **Dependence on structured data:** The study relies on the CISA known exploited vulnerabilities catalog, a structured and curated dataset. While this ensures high quality results, the reliance on structured data may limit the generalizability of the models to unstructured or real-time data sources, such as social media or technical forums. Neuhaus et al. (2007) emphasized the importance of text mining for handling unstructured data, suggesting a potential direction for future work.
- b. **Static dataset vs. dynamic threats:** The use of a static dataset restricts the applicability of the findings to rapidly evolving cybersecurity landscapes. Incorporating adaptive learning techniques, as proposed by Khattak et al. (2022), could improve the models' ability to generalize to new, unseen vulnerabilities.
- c. **Feature-specific generalizability:** Features like "time_since_added" may not be universally applicable to all vulnerability datasets, particularly those lacking temporal information. This limitation aligns with challenges highlighted by Nayak et al. (2014), who noted that feature engineering is critical for ensuring model transferability across datasets.
- d. **Model robustness in real-world scenarios:** The perfect accuracy reported for tree-based models raises concerns about potential overfitting to the dataset. Future studies could validate these models on external datasets to confirm their robustness and ensure generalizability, as suggested by Liu et al. (2019).

5.5 Implications for Future Research and Practice

- a. **Real-time data integration:** Future research should incorporate real-time data from diverse sources, such as vulnerability feeds and exploit databases, to enhance the models' adaptability. This would address a key limitation of the current study and align with recommendations from Neuhaus et al. (2007).
- b. **Exploration of hybrid approaches:** Combining tree-based algorithms with deep learning models may offer the best of both worlds; high interpretability and the ability to handle complex, unstructured data. This hybrid approach was advocated by Babalau et al. (2021) and represents a promising direction for future work.
- c. **Continuous model updates:** Adaptive learning techniques that allow models to update dynamically with new data are essential for maintaining relevance in the face of evolving threats. The importance of such adaptability was underscored by Khattak et al. (2022).

5.6 Implications for Cybersecurity Vulnerability Management

The findings of this study have several important implications for cybersecurity vulnerability management. Machine learning models, particularly tree-based algorithms, provide a powerful tool for automating the assessment of vulnerability severity. By leveraging these models, organizations can prioritize remediation efforts with accurate predictions of vulnerability severity, security teams can focus their efforts on addressing the most critical vulnerabilities first. This ensures that limited resources are allocated efficiently, reducing the risk of successful cyberattacks. It can also enhance Incident Response. Automated vulnerability severity predictions allow security teams to respond more quickly to emerging threats. When integrated with threat intelligence systems, machine learning models can help identify vulnerabilities that are likely to be exploited, enabling proactive defense measures. In addition, organizations can reduce human error. Manual vulnerability assessment is prone to human error, particularly when dealing with large volumes of vulnerabilities. Automated models provide consistent, objective predictions, minimizing the risk of misclassification and ensuring a more reliable vulnerability management process.

5.7 Limitations

Despite the promising results, several limitations of this study should be acknowledged. The first limitation is the Dependence on Structured Data. The data set used in this study was derived from the CISA known exploited vulnerabilities catalog, which provides structured data. However, many vulnerabilities in the wild are discussed in unstructured formats, such as blog posts, forums, or social media. Incorporating real-time data from these sources could improve the predictive power of the models. The second limitation is featuring selection. Although feature engineering was performed to improve model performance, further research is needed to validate and refine the features used in vulnerability severity prediction. Additional features, such as network topology or exploitability trends, could enhance the models' accuracy. The last limitation is generalizability to new vulnerabilities. The models in this study were trained on historical data, which may not fully generalize to future vulnerabilities, particularly those involving new technologies or attack vectors. Regular model updates and the incorporation of adaptive learning techniques will be crucial for maintaining the relevance of the predictions.

6. FUTURE RESEARCH DIRECTIONS

Based on the findings and limitations of this study, several avenues for future research are proposed to enhance the effectiveness of machine learning models for cybersecurity vulnerability severity prediction.

- **Integration of Real-time and unstructured data:** Future research must focus on incorporating a broader range of data sources into predictive models. Unstructured text from social media platforms, blogs, and forums offers a wealth of information that could provide early indicators of vulnerability exploitation (Khattak et al., 2022). These real-time data streams can capture emerging threats as they unfold, offering valuable insights that structured datasets, such as those from vulnerability databases, may not fully represent. By integrating these data sources with existing structured datasets, researchers can develop models that are more responsive to emerging threats, improving predictive accuracy and timeliness in identifying potential cybersecurity risks.
- **Exploration of deep learning approaches:** While this study focused on traditional machine learning models, future research could investigate the use of deep learning models, particularly those capable of handling sequential or temporal data. Models such as Recurrent Neural Networks (RNNs) and transformers could capture the temporal patterns and trends in vulnerability evolution, potentially uncovering more complex relationships in the data. These models are particularly well-suited to capturing how vulnerabilities progress and are exploited over time, offering richer predictive insights. Moreover, Natural Language Processing (NLP) techniques combined with deep learning approaches could unlock deeper insights from textual data, such as vulnerability descriptions, leading to more accurate severity assessments (Babalau et al., 2021).
- **Adaptive and reinforcement learning:** With the fast-evolving nature of cybersecurity threats, adaptive learning methods such as reinforcement learning, and online learning algorithms should be explored. These methods allow models to evolve with new data, ensuring that the predictive models remain relevant as new vulnerabilities emerge and the threat landscape changes. Adaptive models can learn from newly available information in real time, which would help maintain the accuracy and applicability of predictions in rapidly changing environments. This would prevent models from becoming obsolete and improve their resilience against novel or previously unseen threats.
- **Hybrid model approaches:** Future research should also explore hybrid modeling techniques that combine the strengths of different machine learning approaches. For instance, integrating tree-based models (known for their interpretability and strong performance on structured data) with deep learning models (capable of handling unstructured, sequential, or complex data) could lead to improved prediction accuracy and robustness. Hybrid models may also help balance the trade-offs between accuracy and computational efficiency, optimizing performance in real-world cybersecurity settings where resources are often limited.

- Rigorous feature validation and selection: Another critical avenue for future research involves the validation of features used in vulnerability prediction. A rigorous validation process would ensure that the features selected for training the models are truly representative of the underlying threat dynamics. This would prevent overfitting and improve the generalizability of the models across different types of vulnerabilities and threat scenarios (Frei et al., 2006). Additionally, future research could investigate advanced feature engineering techniques that better capture the nuances of cybersecurity threats, leading to even more precise predictions.
- Improving model interpretability: While tree-based models are generally interpretable and easy to understand, more complex models like Gradient Boosting and deep learning models tend to be seen as "black boxes." Further research should focus on improving the interpretability of these complex models, making it easier for cybersecurity professionals to understand the rationale behind predictions. This would enhance their usability in operational settings, allowing security teams to trust the predictions and apply them more effectively in decision-making processes.

7. CONCLUSION

This study demonstrated the effectiveness of machine learning models, particularly tree-based algorithms like Decision Tree, Random Forest, and Gradient Boosting, in predicting the severity of cybersecurity vulnerabilities. These models achieved perfect accuracy, surpassing baseline models such as Logistic Regression and Support Vector Machine, which struggled with vulnerabilities of critical and low severity. Tree-based models excelled in both accuracy and computational efficiency, with Decision Tree being the fastest, making it ideal for real-time applications. These results underscore the potential of machine learning to automate vulnerability prioritization, allowing organizations to allocate resources more efficiently to address the most critical threats. The study also highlighted the significance of feature engineering, where time-based features and keyword extraction played a key role in improving model performance. However, some limitations were noted, including the dependence on structured datasets and the lack of real-time data integration, which can hinder the models' ability to respond to emerging threats. Future research should focus on incorporating unstructured and real-time data sources, such as social media, blogs, and technical forums, to enhance prediction accuracy. Moreover, advanced models like deep learning and adaptive learning techniques should be explored to better address the ever-changing nature of cybersecurity threats. Overall, this research emphasizes the transformative potential of machine learning in vulnerability management, offering an accurate, scalable, and efficient solution for strengthening organizational defenses against evolving cyber risks.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENTS

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2025-2903-16."

REFERENCES

- Babalau, I., Corlatescu, D., Grigorescu, O., Sandescu, C., Dascalu, M. 2021. Severity prediction of software vulnerabilities based on their text description, 171–177
- Bozorgi, M., Saul, L.K., Savage, S., Voelker, G.M. 2010. Beyond heuristics: learning to classify vulnerabilities and predict exploits. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 105–114.
- Cybersecurity and Infrastructure Security Agency (CISA). (2022). *Known Exploited Vulnerabilities Catalog*. Retrieved from <https://www.cisa.gov/known-exploited-vulnerabilities-catalog>.
- Frei, S., May, M., Fiedler, U., Plattner, B. 2006. Large-scale vulnerability analysis. Proceedings of the 2006 SIGCOMM Workshop on Large-Scale Attack Defense, 131–138.
- Holm, H., Afridi, K.K. 2015. An expert-based investigation of the common vulnerability scoring system. Computers and Security, 53, 18–30.
- Holm, H., Ekstedt, M., Andersson, D. 2012. Empirical analysis of system-level vulnerability metrics through actual attacks. IEEE Transactions on Dependable and Secure Computing, 9, 825–837.
- Hulayyil, S.B., Li, S., Xu, L. 2023. Machine-learning-based vulnerability detection and classification in Internet of Things device security. Electronics, 12, 3927.
- Jabeen, G., Rahim, S., Afzal, W., Khan, D., Khan, A.A., Hussain, Z., Bibi, T. 2022. Machine learning techniques for software vulnerability prediction: A comparative study. Applied Intelligence, 52, 17614–17635.
- Joh, H., Malaiya, Y.K. 2014. Defining and assessing quantitative security risk measures using vulnerability lifecycle and CVSS metrics. The International Conference on Security and Management (SAM), 10–16.
- Khattak, A., Almujiab, H., Elamary, A., Matara, C.M. 2022. Interpretable dynamic ensemble selection approach for the prediction of road traffic injury severity: A case study of Pakistan's National Highway N-5. Sustainability, 14, 12340.
- Laghari, A.A., Jumani, A.K., Laghari, R.A., Li, H., Karim, S., Khan, A.A. 2024. Unmanned aerial vehicles advances in object detection and communication security

- review. Cognitive Robotics.
- Laghari, A.A., Li, H., Khan, A.A., Shoulin, Y., Karim, S., Khani, M.A.K. 2024. Internet of Things (IoT) applications security trends and challenges. *Discover Internet of Things*, 4, 36.
- Le, Q., Mikolov, T. 2014. Distributed representations of sentences and documents. *Proceedings of the 31st International Conference on Machine Learning*, 1188–1196.
- Liu, K., Zhou, Y., Wang, Q., Zhu, X. 2019. Vulnerability severity prediction with deep neural network. In *2019 5th international conference on big data and information analytics (BigDIA)*, 114–119.
- Mell, P., Scarfone, K., Romanosky, S. 2007. A complete guide to the common vulnerability scoring system version 2.0. FIRST–Forum of Incident Response and Security Teams, 23.
- Nayak, K., Marino, D., Efstathiopoulos, P., Dumitras, T. 2014. Some vulnerabilities are different than others - studying vulnerabilities and attack surfaces in the wild. In *International Workshop on Recent Advances in Intrusion Detection*. Cham: Springer International Publishing, 426–446.
- Nazir, R., Laghari, A.A., Kumar, K., David, S., Ali, M. 2021. Survey on wireless network security. *Archives of Computational Methods in Engineering*, 1–20.
- Neuhaus, S., Zimmermann, T., Holler, C., Zeller, A. 2007. Predicting vulnerable software components. *Proceedings of the 14th ACM Conference on Computer and Communications Security*, 529–540.
- Smadi, S., Aslam, N., Zhang, L. 2018. Detection of online phishing email using dynamic evolving neural network based on reinforcement learning. *Decision Support Systems*, 107, 88–102.
- Smaha, S.E. 1988. Haystack: An intrusion detection system. *IEEE Aerospace Computer Security Applications Conference*, 37–44.
- Solms, R. von, Niekerk, J. van. 2013. From information security to cyber security. *Computers and Security*, 38, 97–102.
- Yin, S., Li, H., Laghari, A.A., Teng, L., Gadekallu, T.R., Almadhor, A. 2024a. FLSN-MVO : Edge computing and privacy protection based on federated learning siamese network with multi-verse optimization algorithm for industry 5.0. *IEEE Open Journal of the Communications Society*.
- Yin, S., Li, H., Laghari, A.A., Gadekallu, T.R., Sampedro, G.A., Almadhor, A. 2024b. An anomaly detection model based on deep auto-encoder and capsule graph convolution via sparrow search algorithm in 6G internet-of-everything. *IEEE Internet of Things Journal*, 11, 29402–29411.
- Yin, S., Li, H., Teng, L., Laghari, A.A., Estrela, V.V. 2024c. Attribute-based multiparty searchable encryption model for privacy protection of text data. *Multimedia Tools and Applications*, 83, 45881–45902.
- Zhang, S., Caragea, D., Ou, X. 2011. An empirical study on using the national vulnerability database to predict software vulnerabilities. *Lecture Notes in Computer Science*, 22, 217–231.