Evaluating Vulnerability Prioritization Strategies Based on CVSS and EPSS

Felipe Curty do Rego Pinto[1]

[1] Banco Nacional do Desenvolvimento Econômico e Social
Rio de Janeiro, RJ, Brazil

Abstract. *The number of disclosed vulnerabilities should surpass 50,000 in 2026, making prioritization essential for allocating scarce remediation capacity to the vulnerabilities that most require attention. This paper empirically evaluates CVSS-based, EPSS-based, and combined CVSS+EPSS vulnerability prioritization strategies in terms of effectiveness and remediation effort, providing guidance on how to select among them. We built a corpus of* $31,779$ *vulnerabilities disclosed in 2024 from the National Vulnerability Database and used the CISA Known Exploited Vulnerabilities (KEV) catalog as ground truth to identify the vulnerabilities that warrant prioritization. Our results show that CVSS-based strategies provide broader coverage but produce substantially larger prioritized sets, whereas EPSS-based strategies are more selective and concentrate a higher share of known-exploited vulnerabilities in the prioritized set. Combined strategies based on OR conditions can provide balanced trade-offs by increasing coverage with a modest increase in remediation effort, while AND combinations offer limited advantage over EPSS alone. Finally, our results provide evidence to enable organizations to choose the best prioritization strategy considering their remediation capacity and tolerance for missing relevant vulnerabilities.*

1. Introduction

The number of disclosed vulnerabilities has increased substantially in recent years. FIRST [2026] reported that 2026 is projected to be the first year to exceed 50,000 published vulnerabilities, setting a new record for vulnerability disclosure volume. This growing volume forces organizations to triage vulnerabilities and prioritize the remediation of those that pose the greatest risk [Alomar et al., 2020].

Sharma et al. [2021] explain that many organizations rely on scoring systems to prioritize vulnerabilities, with the Common Vulnerability Scoring System (CVSS) being one of the most commonly used criteria. The CVSS evaluates the characteristics of a vulnerability and assigns a score that maps the vulnerability to a severity category [FIRST, 2024a]. Based on this severity score, organizations decide which vulnerabilities to prioritize for remediation. However, Spring et al. [2021] argue that using the CVSS base score alone to prioritize vulnerabilities is not appropriate because it does not fully capture the security risk. Moreover, Scarfone and Mell [2009]; Bozorgi et al. [2010] identified limitations in CVSS scores, including difficulties in accurately measuring vulnerability characteristics, limited score diversity, and the inability of a fixed equation to model the severity of all vulnerabilities accurately.

These limitations highlight the importance of exploitability assessment as an input for vulnerability prioritization [Elder et al., 2024]. The Exploit Prediction Scoring

System (EPSS) uses information about vulnerabilities and evidence of exploitation to train a machine learning model that estimates the probability that a vulnerability will be exploited [Jacobs et al., 2021; FIRST, 2024b]. Unlike CVSS, EPSS incorporates threat-related evidence, addressing the limitation that severity scores do not indicate whether a vulnerability is being actively exploited or is likely to be exploited [Jacobs et al., 2020]. This distinction is particularly important because most cataloged vulnerabilities have no evidence of real-world exploitation [Sabottke et al., 2015].

Although prior studies have shown the limitations of CVSS for vulnerability prioritization and the potential of EPSS and other data-driven approaches to improve exploitation assessment [Allodi and Massacci, 2014; Bozorgi et al., 2010; Howland, 2022; Jacobs et al., 2020, 2021, 2023; Sharma et al., 2021], there is still limited evidence on how these scores should be operationalized as concrete threshold-based prioritization strategies. Prior studies also recognize that organizations face limited remediation capacity and may benefit from combining severity, exploitation likelihood, and evidence of known exploitation [De Smale et al., 2023; Pinto, 2024; Shimizu and Hashimoto, 2026]. It is not clear whether EPSS should be used as an alternative to CVSS or in combination with CVSS to add exploitation-likelihood information to severity-based prioritization.

To address this gap, this paper empirically evaluates CVSS-based, EPSS-based, and combined CVSS+EPSS prioritization strategies using fixed thresholds that can be directly translated into vulnerability remediation policies. The contributions of this paper are: (i) an evaluation of threshold-based prioritization strategies over a corpus of $31,779$ vulnerabilities disclosed in 2024, using the CISA Known Exploited Vulnerabilities (KEV) catalog as ground truth for vulnerabilities that warrant prioritization; (ii) an analysis of the precision–recall trade-off and remediation effort implied by each strategy; and (iii) practitioner-oriented recommendations for selecting a strategy according to remediation capacity and tolerance for missing relevant vulnerabilities.

Our results show a clear trade-off between precision, recall, and remediation effort. The choice of an appropriate strategy depends on the organization's remediation capacity and tolerance for missing relevant vulnerabilities. CVSS-based strategies provide broader coverage but generate larger remediation sets. EPSS-based strategies are more selective. Combined strategies reveal additional trade-offs: OR-based combinations can increase coverage with a modest increase in remediation effort, whereas AND-based combinations offer limited advantage over EPSS alone.

The remainder of this paper is organized as follows. Section 2 describes our study design, corpus construction, and data analysis procedures. Section 3 presents our results. Section 4 discusses the results and provides insights into how organizations can choose an appropriate vulnerability prioritization strategy according to their remediation capacity and tolerance for risk. Section 5 positions our study in relation to prior work. Finally, Section 6 concludes the paper and presents future work.

2. Materials and Methods

In this section, we describe our vulnerability *corpus* (Section 2.1), our study design (Section 2.2), the ground truth used to identify vulnerabilities that warrant prioritization (Section 2.3), the analysis process (Section 2.4), and the threats to the validity of our study (Section 2.5). To support reproducibility, we also provide a replication package contain-

ing the datasets and scripts used in this study.

2.1. Vulnerability *corpus*

We constructed a vulnerability *corpus* suitable for evaluating vulnerability prioritization strategies. We started from all vulnerabilities published in 2024 and then applied filtering criteria based on vulnerability status, CVSS availability, and EPSS availability. We used three main data sources to compose our *corpus*:

1. The National Vulnerability Database (NVD) to fetch the vulnerabilities, which is the repository for vulnerabilities used by the U.S. government and is widely used by researchers and practitioners;
2. The Known Exploited Vulnerabilities (KEV) catalog, which is maintained by the Cybersecurity and Infrastructure Security Agency (CISA) and provides an authoritative list of vulnerabilities that have been exploited in the wild; and
3. The Forum of Incident Response and Security Teams (FIRST) EPSS Database, which holds the EPSS score and percentile for most published vulnerabilities.

We queried the NVD in April 2026 and selected all vulnerabilities published in 2024. We found 40,704 vulnerabilities published in 2024. We intentionally excluded vulnerabilities published in 2025 and 2026 because the inclusion of some vulnerabilities in the KEV may be delayed until evidence of exploitation is observed. By selecting only vulnerabilities published in 2024, we allow sufficient time for KEV inclusion and avoid incomplete labeling of recently published vulnerabilities. As discussed later in this section, some vulnerabilities published in 2024 were added to the KEV catalog only in 2025 and 2026, confirming the need to allow sufficient time for KEV inclusion.

The NVD entries include a status field indicating the current state of each vulnerability record in the database. We excluded 745 vulnerabilities with status *Rejected*, as they do not represent actual threats, and 7778 with status *Awaiting Analysis* or *Undergoing Analysis*, since their analysis is still incomplete and may therefore be unreliable.

The NVD entries also provide severity information based on the Common Vulnerability Scoring System (CVSS). We observed that vulnerabilities were scored using different CVSS versions, including 2.0, 3.0, 3.1, and 4.0. As newer CVSS versions are likely to provide more accurate severity assessments, we selected the most recent version available for each vulnerability. We found 27 vulnerabilities without a CVSS score, which we excluded because they could not be evaluated by the CVSS-based prioritization strategy used in our study.

We queried the EPSS Database using the official dataset published on GitHub. This dataset contains all the EPSS values from 14th April, 2021, when the first EPSS score was released. We excluded 375 vulnerabilities that did not have an EPSS score in the dataset.

Because the EPSS scores vary over time as new information becomes available for a given vulnerability, we had to define a reference date for the analysis. To minimize the risk of incorporating information influenced by later events, such as KEV inclusion and exploit disclosure, we used the first EPSS score available after the vulnerability's publication date.

Finally, we used the NVD data to determine KEV inclusion, rather than directly querying the CISA KEV catalog. The NVD provides the date the vulnerability was added to KEV. Based on this information, we identified 156 KEV-listed vulnerabilities in the corpus. Among them, 27 were added to the KEV catalog only in 2025 and 4 only in 2026, which reinforces the decision to restrict the corpus to vulnerabilities published in 2024.

After applying these filters sequentially, the final *corpus* comprises $31,779$ vulnerabilities, all published in 2024, of which 156 were included in the KEV catalog. Each vulnerability includes the CVE identifier, publication date, CVSS score (using the most recent available version), the first available EPSS score and its date, and the KEV inclusion date.

2.2. Study Design

We designed the study to compare CVSS-based, EPSS-based, and combined prioritization strategies. We aim to understand whether these strategies identify the vulnerabilities that warrant prioritization and how much remediation effort they require.

The CVSS-based prioritization strategy relies solely on CVSS scores to prioritize vulnerabilities for remediation. It is widely used in practice and is commonly observed in industry. We defined four CVSS-based strategies:

1. CVSS 5.0 (CVSS score ≥ 5): a permissive scenario that prioritizes medium, high, and critical vulnerabilities;
2. CVSS 7.0 (CVSS score ≥ 7): a common scenario that aims to prioritize high and critical severity vulnerabilities;
3. CVSS 9.0 (CVSS score ≥ 9): a more conservative scenario that prioritizes only critical severity vulnerabilities; and
4. CVSS 10.0 (CVSS score $= 10$): a very conservative strategy that prioritizes only vulnerabilities with the maximum severity.

The EPSS-based prioritization strategy relies solely on EPSS scores to prioritize vulnerabilities for remediation. It is a more recent approach that has been gaining attention in the industry. Unlike CVSS, EPSS does not define severity classes such as high or critical. Thus, we defined five EPSS-based scenarios based on thresholds that represent increasing levels of exploitation likelihood. We expressed these thresholds as percentages of the EPSS score, so that, for example, 0.05% corresponds to EPSS score > 0.0005. The evaluated thresholds were 0.05%, 0.10%, 0.50%, 1.00%, and 5.00%.

The combined prioritization strategies rely on integrating CVSS and EPSS scores to prioritize vulnerabilities for remediation. Their goal is to combine vulnerability severity and exploitation likelihood to potentially improve prioritization effectiveness while controlling remediation effort. We defined combined strategies using two logical operators:

1. AND operator: a more restrictive strategy that prioritizes vulnerabilities that meet both the CVSS and the EPSS thresholds; and
2. OR operator: a more inclusive strategy that prioritizes vulnerabilities that meet at least one of the two thresholds.

We used EPSS 0.50% as the EPSS threshold in all combined strategies because, among the standalone EPSS-based strategies evaluated in this study, it provided the best balance between precision and recall. For CVSS, we selected the 7.0, 9.0, and 10.0 thresholds. The combinations were designed to evaluate two complementary uses of EPSS: as a filtering condition for broad CVSS thresholds and as an expansion condition for highly selective CVSS thresholds:

1. CVSS 7.0 alone is relatively broad, so we paired it only with the AND operator to restrict coverage by requiring the EPSS condition to also be met;
2. CVSS 9.0 represents an intermediate severity threshold and was therefore evaluated under both AND and OR combinations; and
3. CVSS 10.0 alone is highly selective, so we paired it only with the OR operator to extend coverage to vulnerabilities with higher predicted exploitation likelihood.

2.3. Ground Truth for Evaluation

To assess whether a strategy correctly identifies vulnerabilities that warrant prioritization, we need to define a ground truth. Thus, we can compare the strategies' results with the ground truth to determine which strategy performs better.

We used the KEV catalog as our ground truth for known exploited vulnerabilities, as it provides an authoritative list of vulnerabilities that are known to have been exploited in the wild. The KEV catalog is a solid ground truth because it represents vulnerabilities with evidence of real-world exploitation, regardless of their severity or likelihood of exploitation, rather than prioritization solely through model-derived proxies. This choice is aligned with prior work [Jacobs et al., 2020, 2021, 2023; Shimizu and Hashimoto, 2026] that used evidence of real-world exploitation as the outcome of interest. Using KEV as our ground truth also improves reproducibility, since it is a public and authoritative catalog that is easy to access and widely used in both practice and research.

It is worth noting that EPSS is time-dependent and may incorporate information related to the KEV catalog. However, this dependency should be interpreted with caution. First, both EPSS and the KEV catalog are linked to real-world events and exploitation evidence. Second, the most urgent vulnerabilities (e.g., zero-day vulnerabilities) may be exploited very early and included in the KEV catalog before being published in the NVD and before an EPSS score becomes available. In fact, in our *corpus*, this occurred for 40 of the 156 vulnerabilities included in the KEV catalog. This raises an important methodological decision: whether to exclude vulnerabilities whose KEV catalog inclusion predates the first available EPSS score. We chose to retain these vulnerabilities, because excluding them could remove some of the most critical and time-sensitive cases, thereby introducing selection bias.

As our goal is to evaluate whether prioritization strategies can identify vulnerabilities that warrant prioritization based on the information available to practitioners, retaining vulnerabilities whose KEV inclusion predates the first available EPSS score does not compromise our study. Moreover, as we do not aim to demonstrate causal independence between EPSS and KEV, nor to predict KEV using EPSS, the KEV catalog remains an appropriate ground truth for evaluating EPSS-based strategies.

2.4. Data Analysis

We assessed the vulnerability prioritization strategies from two perspectives: effectiveness and remediation effort. Effectiveness refers to whether a strategy prioritizes relevant vulnerabilities while avoiding missing vulnerabilities that warrant prioritization, whereas remediation effort refers to the amount of work generated by the set of vulnerabilities selected by the strategy.

We used precision and recall to measure the effectiveness of the strategies. In a high-precision strategy, most of the prioritized vulnerabilities are relevant, while in a high-recall strategy, most of the relevant vulnerabilities are prioritized. In this paper, a true positive (TP) corresponds to a vulnerability that is prioritized by the strategy and included in the KEV catalog, a false positive (FP) corresponds to a vulnerability that is prioritized by the strategy but not included in the KEV catalog, and a false negative (FN) corresponds to a vulnerability that is not prioritized by the strategy but included in the KEV catalog.

As more restrictive thresholds generally increase precision and decrease recall, we need a measure that combines both metrics. The F1-score, defined as the harmonic mean of precision and recall, is commonly used for this purpose. However, assigning equal weight to precision and recall favor very selective strategies that missed many KEV vulnerabilities. The F2-score is a weighted harmonic mean of precision and recall that gives more weight to recall. Thus, we used the F2-score as the primary measure to compare strategies, and the F1-score as a secondary criterion when strategies presented similar F2-scores.

Moreover, we used the total number of prioritized vulnerabilities to measure the remediation effort of the strategies. We used this measure as a proxy for overall effort, and we acknowledge that in the real world, different vulnerabilities may demand different remediation efforts. In this context, strategies that prioritize fewer vulnerabilities require less remediation effort and are therefore likely to be less costly than those that prioritize more vulnerabilities.

2.5. Threats to validity

We discuss threats to validity following the classification proposed by Wohlin [2014]:

Internal validity: The EPSS scores are generated by a statistical model, and we only have access to its historical values for each vulnerability. We do not control the data that are incorporated into the model, and, as discussed earlier, the model may use the KEV catalog. We used the first available EPSS score to reduce the likelihood of incorporating information that may have been influenced by later events (e.g., exploitation evidence).

Construct validity: We used the KEV catalog as a ground truth to determine whether a vulnerability should be prioritized. Although some vulnerabilities not included in KEV may still warrant prioritization, the KEV catalog provides a reasonable proxy for high-priority vulnerabilities because it captures observed exploitation in the wild. Moreover, we approximated remediation effort by the number of prioritized vulnerabilities. This proxy does not capture differences in remediation complexity, patch availability, validation effort, or coordination cost across vulnerabilities.

Conclusion validity: The results of our study depend on the selected CVSS and EPSS

Table 1. Performance of the CVSS-based prioritization strategies

	Effectiveness				Quantity			
Strategy	Precision	Recall	F1-score	F2-score	TP	FP	FN	Total
CVSS 5.0	0.56%	99.36%	1.11%	2.73%	155	27,560	1	27,715
CVSS 7.0	0.96%	83.33%	1.90%	4.59%	130	13,402	26	13,532
CVSS 9.0	2.36%	41.03%	4.46%	9.59%	64	2,650	92	2,714
CVSS 10.0	5.30%	4.49%	4.86%	4.63%	7	125	149	132

thresholds. While different threshold configurations may lead to different prioritization outcomes, we chose commonly used benchmarks for CVSS and EPSS to define our prioritization strategies.

External validity: Our study focuses on vulnerabilities published in 2024, which may limit the generalization of the results to other time periods, as vulnerability characteristics and exploitation patterns may evolve over time.

3. Results

We ran each strategy against all the vulnerabilities in our *corpus*. Tables 1, 2, and 3 summarize the performance of the CVSS-based, EPSS-based, and combined strategies, including precision, recall, F1-score, and F2-score, true positives, false positives, false negatives, and total number of prioritized vulnerabilities.

Table 1 shows the performance of the CVSS-based strategies. In general, the strategies with lower CVSS thresholds prioritized more vulnerabilities, which increased recall but reduced precision. The strategies based on the CVSS 5.0, CVSS 7.0, and CVSS 9.0 thresholds followed this pattern. The CVSS 10.0 threshold stands apart from the other thresholds because it is considerably more restrictive, resulting in the lowest recall among the CVSS thresholds.

The CVSS 9.0 strategy was the most balanced according to the F2-score and provided a more conservative trade-off between precision and recall than the other CVSS-based strategies. The CVSS 10.0 strategy had the highest precision, but was also the most restrictive and had the lowest recall among the CVSS-based strategies. Although the CVSS 5.0 strategy achieved the highest recall, it prioritized 86.20% of the vulnerabilities, which is almost the entire *corpus*. These results suggest that the CVSS 5.0 strategy is too broad to support selective prioritization.

Overall, increasing the CVSS threshold reduces the total number of prioritized vulnerabilities. However, the reduction in prioritized vulnerabilities is more pronounced than the reduction in KEV-listed vulnerabilities captured. For instance, the CVSS 7.0 strategy prioritized 42.58% of the vulnerabilities in the *corpus* while capturing 83.33% of the vulnerabilities in the KEV catalog. The CVSS 9.0 strategy prioritized only 8.54% of the vulnerabilities in the *corpus* but captured 41.03% of the vulnerabilities in the KEV catalog. This indicates that CVSS thresholds can substantially reduce the prioritization set while still preserving broad coverage of KEV-listed vulnerabilities.

Table 2 shows the performance of the EPSS-based strategies. As with the CVSS-based strategies, lower EPSS thresholds prioritized more vulnerabilities. Overall, the

Table 2. Performance of the EPSS-based prioritization strategies

	Effectiveness				Quantity			
Strategy	Precision	Recall	F1-score	F2-score	TP	FP	FN	Total
EPSS 0.05%	1.66%	39.74%	3.18%	7.10%	62	3,678	94	3,740
EPSS 0.10%	13.27%	26.28%	17.63%	21.97%	41	268	115	309
EPSS 0.50%	63.64%	17.95%	28.00%	20.96%	28	16	128	44
EPSS 1.00%	80.77%	13.46%	23.08%	16.15%	21	5	135	26
EPSS 5.00%	80.00%	2.56%	4.97%	3.18%	4	1	152	5

Table 3. Performance of the combined CVSS and EPSS prioritization strategies

	Effectiveness				Quantity			
Strategy	Precision	Recall	F1-score	F2-score	TP	FP	FN	Total
CVSS 7.0 AND EPSS 0.50%	58.33%	13.46%	21.88%	15.91%	21	15	135	36
CVSS 9.0 AND EPSS 0.50%	42.11%	5.13%	9.14%	6.22%	8	11	148	19
CVSS 9.0 OR EPSS 0.50%	3.07%	53.85%	5.80%	12.49%	84	2,655	72	2,739
CVSS 10.0 OR EPSS 0.50%	20.00%	22.44%	21.15%	21.90%	35	140	121	175

EPSS-based strategies were more selective and prioritized fewer vulnerabilities than the CVSS-based strategies. Moreover, as the EPSS threshold increased, precision increased markedly, while recall decreased only moderately.

The EPSS 0.10% strategy performed best according to the F2-score. However, the EPSS 0.50% strategy achieved a similar F2-score, along with substantially higher precision and F1-score. This suggests that the EPSS 0.50% strategy may be preferable when greater selectivity is desired without a substantial loss in F2-score. The EPSS 1.00% strategy achieved the highest precision. In contrast, the EPSS 0.05% strategy achieved the highest recall, but at the cost of the lowest precision, resulting in a less balanced strategy than the other EPSS-based strategies under both the F1-score and the F2-score. Finally, the EPSS 5.00% strategy offered no advantage over the EPSS 1.00% strategy, as it drastically reduced recall without increasing precision.

Overall, as observed with the CVSS thresholds, increasing the EPSS threshold reduced the number of prioritized vulnerabilities more sharply than the number of KEV-listed vulnerabilities captured. In addition, the EPSS-based strategies yielded a higher proportion of KEV-listed vulnerabilities than the CVSS-based strategies. For instance, the CVSS 10.0 strategy, the most restrictive CVSS-based strategy, included one KEV-listed vulnerability for every 19 prioritized vulnerabilities, whereas the EPSS 0.50% strategy included two KEV-listed vulnerabilities for every three prioritized vulnerabilities. Moreover, the CVSS 10.0 strategy prioritized 132 vulnerabilities and the EPSS 0.50% strategy prioritized 44, or nearly one third as many. Therefore, the EPSS-based strategies not only prioritized fewer vulnerabilities, but also concentrated a higher share of KEV-listed vulnerabilities within the prioritized set.

Table 3 shows the performance of the combined strategies. The AND strategies were more restrictive and tended to increase precision and reduce recall, while the OR strategies were more inclusive and tended to reduce precision and increase recall. Overall,

the OR-based combinations provided a better balance between coverage and selectivity than the AND-based combinations.

Among the AND-based combined strategies, the CVSS 7.0 AND EPSS 0.50% strategy achieved the best F2-score. This strategy outperformed the standalone CVSS-based strategies, but still underperformed the EPSS 0.50% strategy in terms of F2-score. Moreover, the EPSS strategy captured more KEV-listed vulnerabilities while prioritizing only a few more vulnerabilities. These results suggest that adding CVSS as an additional restrictive condition did not provide an advantage over using EPSS alone.

Among the OR-based combined strategies, the CVSS 10.0 OR EPSS 0.50% strategy achieved the best F2-score. This strategy outperformed the standalone CVSS-based strategies in terms of F2-score. The strategy matched the EPSS 0.50% strategy while capturing more KEV-listed vulnerabilities at the cost of prioritizing more vulnerabilities. Even so, both strategies prioritized relatively small sets of vulnerabilities.

Overall, the CVSS-based strategies provided broad coverage with low selectivity, whereas the EPSS-based strategies provided a more selective approach. The combined strategies were designed to assess whether combining severity and exploit likelihood would improve prioritization over standalone strategies. The OR combinations appeared to be the most promising among the combined strategies.

4. Discussion

The main question addressed in this paper is which strategy practitioners should use to prioritize vulnerabilities. In an ideal setting, organizations would remediate all vulnerabilities promptly. However, in a real setting, this would result in the remediation of less relevant vulnerabilities. Additionally, in practice, organizations often lack the capacity to remediate all vulnerabilities.

Therefore, it is essential to define a prioritization strategy and the best choice depends on the organizational context, capacity to remediate vulnerabilities, and tolerance for missing relevant vulnerabilities (Section 4.1). Moreover, organizations may consider additional factors when prioritizing vulnerabilities, such as the asset criticality and patching programs, to define the best vulnerability prioritization strategy (Section 4.2).

4.1. Choosing a strategy

Our results show that there is no single clear winner among the evaluated strategies. Instead, the evaluated strategies reflect a clear trade-off between precision and recall. Some strategies favor precision and are more suitable for contexts in which remediation capacity is limited (Section 4.1.1). Other strategies favor recall and are more suitable for contexts in which missing relevant vulnerabilities is a greater concern than reducing remediation effort (Section 4.1.2). Between these extremes, balanced strategies provide a compromise between these two objectives and may be useful when organizations seek to preserve coverage without adopting very broad thresholds (Section 4.1.3). Finally, some strategies either perform poorly in precision and recall or provide limited discrimination between vulnerabilities, suggesting limited practical value for prioritization (Section 4.1.4).

4.1.1. Selective strategies

Selective strategies are useful when remediation capacity is limited and the organization needs to focus on a smaller set of vulnerabilities with higher expected relevance. These strategies favor precision and tend to reduce remediation effort, but at the cost of missing some relevant vulnerabilities.

The CVSS-based strategies were not effective selective strategies. While raising the CVSS threshold leads to a more selective strategy, our results show that the CVSS 10.0 strategy captured only a few KEV-listed vulnerabilities and prioritized vulnerabilities that may not warrant immediate prioritization. This suggests that severity alone is insufficient to support selective prioritization.

The EPSS-based strategies were the most suitable candidates for this type of approach. The EPSS 0.50% and EPSS 1.00% thresholds achieved high precision while still prioritizing a reasonable number of vulnerabilities. Thus, they may be useful for teams that need to reduce remediation workload while still focusing on vulnerabilities with higher likelihood of exploitation. Therefore, these thresholds represent a practical compromise for organizations that need a selective strategy without becoming overly restrictive.

The combined strategy CVSS 7.0 AND EPSS 0.50% was the most selective combined strategy. However, it performed worse than the EPSS-based strategies in both precision and recall. This suggests that adding CVSS 7.0 as an additional condition did not improve prioritization in our *corpus*, i.e., the combination reduced coverage without providing a compensating gain in precision. In this case, using EPSS alone appears to be a simpler and more effective selective strategy.

4.1.2. Broad coverage strategies

Some strategies favor recall. They tend to include more vulnerabilities that may not require immediate action. These strategies are more suitable for teams with high capacity to remediate vulnerabilities or organizations with low tolerance for missing relevant vulnerabilities.

The CVSS-based strategies were more suitable for broad-coverage. Among the evaluated strategies, the CVSS 7.0 captured a larger number of KEV-listed vulnerabilities than more restrictive CVSS thresholds. However, this higher coverage comes at the cost of a larger remediation workload. Therefore, CVSS 7.0 may be useful when the goal is to maximize coverage and the organization has enough capacity to handle a larger remediation workload.

The EPSS-based strategies were less suitable for broad coverage. Lower EPSS thresholds increased recall, but also significantly reduced precision. Thus, EPSS alone appears more useful for selective prioritization than for broad-coverage strategies.

The combined strategy CVSS 9.0 OR EPSS 0.50% achieved a favorable trade-off between coverage and remediation effort. It prioritized 79.7% fewer vulnerabilities than the CVSS 7.0 strategy, with only a small reduction in recall. This suggests that OR-based combinations may be useful when organizations seek broad coverage but cannot absorb

the remediation workload imposed by lower CVSS thresholds.

4.1.3. Balanced strategies

Balanced strategies aim to provide a compromise between selective and broad-coverage approaches. They are useful when organizations want to preserve coverage of relevant vulnerabilities, but cannot absorb the remediation workload imposed by broad strategies. In this context, a balanced strategy should increase recall without excessively reducing precision or substantially increasing the number of prioritized vulnerabilities.

The CVSS 10.0 OR EPSS 0.50% strategy achieved higher recall than either the EPSS 0.50% strategy and the CVSS 10.0 strategy alone, without a substantial increase in the number of prioritized vulnerabilities. This result suggests that this strategy provided a favorable trade-off between recall and remediation effort in our *corpus*. From a practical perspective, this strategy may be useful for organizations that want to reduce the risk of missing relevant vulnerabilities without relying on low CVSS thresholds, which could substantially increase the remediation workload.

4.1.4. Unworthy strategies

The purpose of a prioritization strategy is to distinguish the vulnerabilities that warrant immediate attention from those that can be addressed later. Some strategies are too broad and are not suitable for practical vulnerability prioritization because they prioritize many irrelevant vulnerabilities. Others may be too selective and miss many relevant vulnerabilities.

At first glance, CVSS 5.0 is the strategy with the highest recall among the evaluated strategies. However, on closer inspection, this result is misleading because it prioritized 86.20% of the vulnerabilities in our *corpus*, i.e., almost all vulnerabilities. Therefore, low CVSS thresholds may achieve high recall simply by prioritizing an excessively large number of vulnerabilities, which limits their practical value for prioritization. A similar issue happened with low EPSS score thresholds. As we reduce the threshold, we include more vulnerabilities and the prioritization strategy becomes less useful in practice.

The high EPSS thresholds present another caveat. The strategy may become so selective that it prioritizes very few vulnerabilities. As shown in Table 2, the EPSS 5.00% achieved the same precision as the EPSS 1.00%, but a lower recall. This suggests that increasing the EPSS threshold beyond this point reduces recall without providing a compensating gain in precision.

4.2. Additional factors for prioritization

Organizations may use different prioritization strategies according to asset criticality. For instance, an organization may favor recall-oriented strategies for high-criticality assets and precision-oriented strategies for lower-criticality assets. For high-criticality assets, a broader strategy will select more vulnerabilities, but applying this strategy only to a restricted subset of assets may reduce the number of applicable vulnerabilities. Thus, the resulting remediation effort may remain manageable. Moreover, by adopting a selective

strategy for lower-criticality assets, the organization may still remediate relevant vulnerabilities without substantially increasing the overall remediation effort.

In addition to choosing different prioritization strategies, organizations may define different deadlines for vulnerability remediation based on the asset criticality, vulnerability severity, and likelihood of exploitation. This approach may offer greater flexibility for teams to manage the remediation workload according to their capacity, aligning remediation urgency without compromising other important organizational initiatives.

Finally, by defining a regular patching program, the organization may still remediate non-prioritized vulnerabilities under a less urgent schedule. This approach helps reduce the risk and align expectations with senior management, as the relevant vulnerabilities will be prioritized and the remaining will be handled in traditional patch programs that happen quarterly or semiannually.

5. Related Work

In this section, we review prior work on vulnerability prioritization, focusing on three related lines of research: studies that discuss the limitations of CVSS as a prioritization criterion, studies that develop or evaluate data-driven approaches for vulnerability assessment and exploit prediction, including EPSS, and studies that examine how organizations prioritize vulnerabilities in practice.

Allodi and Massacci [2014] and Howland [2022] both highlight important limitations of CVSS as a basis for vulnerability prioritization. Although the two studies rely on distinct datasets and analytical methods, they reach a similar conclusion that CVSS scores are weak indicators of real-world exploitation and provide limited support for risk-based remediation decisions. Neither study evaluates fixed threshold-based strategies, which is the focus of our work.

Bozorgi et al. [2010] used machine learning over vulnerability descriptions, timestamps, cross-references, and other metadata to predict whether and how soon vulnerabilities would be exploited. Sharma et al. [2021] used a convolutional neural network to classify vulnerabilities into severity categories based on their textual descriptions. Although these studies show the value of data-driven vulnerability assessment, they focus on building or evaluating predictive models rather than comparing prioritization strategies based on fixed CVSS and EPSS thresholds.

Jacobs et al. [2020] analyzed historical vulnerability data from MITRE's Common Vulnerabilities and Exposures (CVE) program, together with NVD records and other exploit databases, covering the period from 2009 to 2018. In later work, Jacobs et al. [2021, 2023] extended this research and developed the EPSS model. Jacobs et al. [2023] evaluated EPSS for vulnerability prioritization against CVSS-based strategies, using observed exploitation activity in the following 30 days as ground truth. Their results showed that EPSS outperformed CVSS.

De Smale et al. [2023] highlighted that in the real world, organizations are unable to remediate all vulnerabilities and need to adopt prioritization strategies. Based on interviews with 22 organizations in critical infrastructure and government services, the study suggests that organizations adopt selective vulnerability prioritization strategies due to limited resources. Pinto [2024] compared CVSS, EPSS, and other threat analysis ap-

proaches in a conceptual analysis. The study argued that, although EPSS addresses some limitations of CVSS, both scoring approaches should be complemented with broader risk analysis to support effective vulnerability prioritization. These studies highlight both the practical need for vulnerability prioritization and the limitations of relying on CVSS or EPSS alone, whereas our work empirically evaluates vulnerability prioritization strategies based on CVSS and EPSS scores.

Shimizu and Hashimoto [2026] devised a decision framework based on CVSS score, EPSS score, and the CISA KEV catalog. The authors evaluated the framework using real-world CVEs and vendor-reported exploitation data, showing that the combined approach improves efficiency over CVSS-only prioritization while maintaining high coverage. While their work defines specific thresholds and a decision tree for prioritization, our work compares different threshold-based prioritization strategies using CVSS and EPSS scores.

Overall, prior work shows that CVSS alone provides limited support for risk-based prioritization, while EPSS adds exploitation-likelihood information that can improve prioritization decisions. However, less attention has been given to how CVSS and EPSS can be operationalized as concrete threshold-based vulnerability prioritization strategies and how these strategies affect remediation effort. Our study addresses this gap by comparing CVSS-based, EPSS-based, and combined CVSS+EPSS strategies in terms of effectiveness and remediation effort.

6. Conclusion

In this paper, we empirically evaluated four CVSS-based, five EPSS-based, and four combined prioritization strategies over a corpus of $31,779$ vulnerabilities disclosed in 2024, using the CISA KEV catalog as ground truth. Our results show that EPSS-based strategies are more selective than CVSS-based strategies while still capturing a relevant share of known-exploited vulnerabilities. Combining CVSS and EPSS scores with OR extends coverage with a modest increase in remediation effort, whereas combining them with AND yields little advantage over EPSS alone. Moreover, some threshold configurations should be avoided in practice, such as CVSS 5.0 and EPSS 5.00%, because they either prioritize too many vulnerabilities or become overly restrictive without providing a clear compensating benefit.

These findings provide concrete, threshold-level guidance for practitioners who need to select a vulnerability prioritization strategy. Organizations with limited remediation capacity should adopt more selective strategies, whereas more risk-averse organizations may prefer broader, coverage-oriented strategies. These strategies should also be adapted to organizational context: broader strategies may be appropriate for high-criticality assets, while more selective strategies may be sufficient for lower-criticality assets. In addition, prioritization strategies should be combined with remediation deadlines and regular patching programs so that non-prioritized vulnerabilities are still addressed under less urgent schedules.

Future work can extend this study in four main directions. First, our analysis covers vulnerabilities disclosed in 2024. Replicating this study over a longer window, for instance starting in 2021, when EPSS became publicly available, would assess the temporal robustness of our findings. Second, we evaluated a discrete set of CVSS and EPSS

thresholds aligned with current practice. Future work should explore the full continuous threshold space to identify thresholds that maximize precision and recall trade-offs. Third, our ground truth relies on the CISA KEV catalog, which captures only part of real-world exploitation. Future work may evaluate prioritization strategies against other sources, such as exploit databases and vendor exploitation reports. Fourth, a study could validate our recommendations through case studies to assess prioritization strategies under specific scenarios or product sets.

Data availability: The data that support the findings of this study and the scripts we used to run our experiments are available in our replication package: [redacted].

Competing Interests: The authors declare that they have no competing interests. Furthermore, this publication is the sole initiative of the authors and does not represent the position of any company or organization.

Use of Generative AI: The authors used ChatGPT and Claude to support language revision, clarity improvements, and critical review of the manuscript structure and argumentation. The authors reviewed and are responsible for all content of the paper.

References

Allodi, L. and Massacci, F. (2014). Comparing Vulnerability Severity and Exploits Using Case-Control Studies. *ACM Trans. Inf. Syst. Secur.*, 17(1):1:1–1:20.

Alomar, N., Wijesekera, P., Qiu, E., and Egelman, S. (2020). "You've Got Your Nice List of Bugs, Now What?" Vulnerability Discovery and Management Processes in the Wild. pages 319–339.

Bozorgi, M., Saul, L. K., Savage, S., and Voelker, G. M. (2010). Beyond heuristics: learning to classify vulnerabilities and predict exploits. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 105–114, New York, NY, USA. Association for Computing Machinery.

De Smale, S., Van Dijk, R., Bouwman, X., Van Der Ham, J., and Van Eeten, M. (2023). No one drinks from the firehose: How organizations filter and prioritize vulnerability information. In *2023 IEEE symposium on security and privacy (SP)*, pages 1980–1996. IEEE.

Elder, S., Rahman, M. R., Fringer, G., Kapoor, K., and Williams, L. (2024). A Survey on Software Vulnerability Exploitability Assessment. *ACM Comput. Surv.*, 56(8):205:1–205:41.

FIRST (2024a). CVSS v4.0 Specification - 2024-06-18. Technical report.

FIRST (2024b). The EPSS Model.

FIRST (2026). FIRST Releases 2026 Vulnerability Report, Projecting Record-Breaking Common Vulnerabilities and Exposures.

Howland, H. (2022). CVSS: Ubiquitous and Broken. *Digital Threats*, 4(1):1:1–1:12.

Jacobs, J., Romanosky, S., Adjerid, I., and Baker, W. (2020). Improving vulnerability remediation through better exploit prediction. *Journal of Cybersecurity*, 6(1):tyaa015.

Jacobs, J., Romanosky, S., Edwards, B., Adjerid, I., and Roytman, M. (2021). Exploit Prediction Scoring System (EPSS). *Digital Threats*, 2(3):20:1–20:17.

Jacobs, J., Romanosky, S., Suciu, O., Edwards, B., and Sarabi, A. (2023). Enhancing Vulnerability Prioritization: Data-Driven Exploit Predictions with Community-Driven Insights. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 194–206.

Pinto, F. C. d. R. (2024). *O Impacto da Análise de Ameaças Cibernéticas na Classificação e Tratamento de Vulnerabilidades*. Trabalho de Conclusão do Curso Superior de Segurança de Defesa Cibernética (CSSDC), Escola Superior de Guerra (ESG), Brasil.

Sabottke, C., Suciu, O., and Dumitraş, T. (2015). Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits. In *Proceedings of the 24th USENIX Security Symposium*.

Scarfone, K. and Mell, P. (2009). An analysis of CVSS version 2 vulnerability scoring. In *2009 3rd International Symposium on Empirical Software Engineering and Measurement*, pages 516–525.

Sharma, R., Sibal, R., and Sabharwal, S. (2021). Software vulnerability prioritization using vulnerability description. *International Journal of System Assurance Engineering and Management*, 12(1):58–64.

Shimizu, N. and Hashimoto, M. (2026). Vulnerability Management Chaining: An Integrated Framework for Efficient Cybersecurity Risk Prioritization. *IEEE Access*, 14:31407–31424.

Spring, J., Hatleback, E., Householder, A., Manion, A., and Shick, D. (2021). Time to Change the CVSS? *IEEE Security & Privacy*, 19(2):74–78. Conference Name: IEEE Security & Privacy.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10.