Journal of Computer Science and Technology Studies
ISSN: 2978-8048
DOI: 10.32996/jcsts.2026.5.8.1
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

An Explainable Machine-Learning Framework for Prioritizing Cybersecurity Vulnerabilities in the Post-Mythos Era

Nandita Das
New Jersey, USA
Corresponding Author: Nandita Das, **E-mail:** nandita.das@ey.com

| ABSTRACT

AI-enabled vulnerability discovery is increasing both the volume and speed of newly identified security flaws, while organizations continue to face limited remediation capacity. This paper presents EVPD, an explainable machine-learning framework that ranks newly disclosed CVEs by their likelihood of entering the CISA Known Exploited Vulnerabilities Catalog within 180 days. The dataset combines 156,444 CVEs from NVD and MITRE records with historically aligned EPSS observations and CISA KEV dates. After temporal eligibility filtering, 134,875 vulnerabilities were divided chronologically into training, validation, and 2025 test partitions. A stacked ensemble combining a random-forest score, EPSS, and CVSS achieved a test PR-AUC of 0.0600, compared with 0.0337 for a tuned CVSS–EPSS baseline and a no-skill value of 0.0016. A paired bootstrap analysis produced a mean PR-AUC improvement of 0.0242, with a 95% confidence interval of 0.0032–0.0511. When remediation capacity was limited to the top 1% of the test backlog, EVPD identified 27.0% of the vulnerabilities that later entered KEV, compared with 17.6% for CVSS–EPSS and 12.2% for CVSS alone. SHAP analysis showed that historical EPSS, affected-configuration breadth, and CVSS impact measures con-tributed most strongly to the predictions. The results indicate that temporally aligned and explainable ranking can improve early vulnerability selection when remediation capacity is highly constrained.

| KEYWORDS

Vulnerability prioritization, Exploit prediction, Explainable Machine learning, CVSS, EPSS, CISA KEV, SHAP

| ARTICLE INFORMATION

ACCEPTED: 01 June 2026 **PUBLISHED:** 26 July 2026 **DOI:** 10.32996/jcsts.2026.8.8.14

1. Introduction

Artificial intelligence is changing the speed at which soft-ware vulnerabilities can be found and validated. Tasks that once required prolonged manual analysis can increasingly be supported by autonomous systems. Anthropic's April 2026 disclosure of Claude Mythos Preview brought this shift into sharper focus by describing a model capable of identifying and validating complex vulnerabilities across major software platforms [1]. The term *post-Mythos era* is used here to describe an environment in which AI-enabled discovery can increase vulnerability volume faster than most organizations can expand remediation capacity.

The operational problem is therefore not limited to discovery. Security teams must decide which findings require attention first. Many programs continue to rely on the Common Vulnerability Scoring System (CVSS) as their primary prioritization mechanism. CVSS summarizes technical characteristics such as attack vector, attack complexity, required privileges, user interaction, and potential impact [2]. It does not estimate whether exploitation is likely within a defined period. NVD consequently distinguishes severity from broader risk [3]. A critical vulnerability may never be used in an observed attack, whereas a medium-severity vulnerability may be rapidly weaponized because it is remotely reachable, affects a widely deployed product, or provides reliable initial access.

Previous studies have shown that exploit availability and observed exploitation depend on combinations of technical, temporal, and contextual factors [4], [5]. EPSS addresses part of this problem by estimating the probability that exploitation activity will be observed during the following 30 days [7]. CISA KEV supplies a complementary signal by identifying vulnerabilities for which exploitation evidence has been con-firmed [8]. These resources represent different dimensions of risk: CVSS describes technical severity, EPSS estimates short-term exploitation probability, and KEV records confirmed exploitation evidence.

Several methodological problems remain. Historical experiments may use present-day EPSS scores or subsequently enriched CVE records, allowing information unavailable at the original decision point to enter the model. Random train–test splitting can produce similarly optimistic estimates because vulnerabilities from related products or disclosure periods may appear in both partitions. Existing scoring rules also have difficulty representing nonlinear interactions among exploitation probability, attack characteristics, affected configurations, and disclosure evidence. Finally, a model is difficult to use operationally when analysts cannot explain why one vulnerability was placed above another.

EVPD addresses these limitations through a temporally controlled experimental design. NVD, MITRE CVE List V5, MITRE CWE, historical EPSS observations, and CISA KEV dates are joined at the CVE level [9], [10]. Predictors are restricted to information available at an early-life prediction point, and the outcome is defined as subsequent KEV inclusion within 180 days. Models are trained on vulnerabilities published from 2021 through 2023, tuned on 2024, and evaluated on an untouched 2025 test set.

The study addresses four questions:

- **RQ1:** How accurately can future KEV inclusion be predicted from early-life vulnerability information?
- **RQ2:** Does EVPD outperform CVSS-, EPSS-, and CVSS–EPSS-based ranking when remediation capacity is limited?
- **RQ3:** Which vulnerability attributes contribute most strongly to the model's rankings?
- **RQ4:** Are the learned feature relationships stable across later disclosure periods?

The primary contribution is a temporally controlled, capacity-aware, and explainable framework for early vulnerability prioritization. Unlike retrospective studies that may apply current scores to historical vulnerabilities or use random data partitions, EVPD reconstructs an explicit early-life prediction point using archived EPSS observations, excludes vulnerabilities already known to be exploited, and evaluates future KEV inclusion on an untouched chronological test year. The study further measures operational value under fixed remediation budgets, statistically compares ranking performance through paired bootstrap analysis, and examines the temporal stability of model explanations. The contribution therefore lies in combining temporal reconstruction, future-outcome prediction, constrained-capacity evaluation, statistical comparison, and explanation stability within one operational decision framework, rather than in proposing a new classifier alone.

2. Related Work

2.1 Severity and Exploitability

CVSS is widely used to summarize vulnerability severity, but it does not estimate the probability of exploitation [2]. Empirical studies have also shown that technical severity and observed exploitation are not equivalent [11]. Bozorgi *et al.* demonstrated that combinations of technical and contextual attributes can predict exploit availability more effectively than severity heuristics alone [4]. Householder *et al.* further showed that exploit availability develops over time and varies across vulnerability populations [12].

EPSS provides a daily estimate of the probability that exploitation activity will be observed for a published CVE during the following 30 days [7]. Its daily score is useful for prioritization, but the score changes as new information becomes available. Historical evaluation therefore requires the EPSS value that existed at the intended decision point rather than the latest value. CISA KEV provides a narrower but high-confidence outcome source by recording vulnerabilities confirmed to have been exploited in the wild [8]. Absence from KEV, however, cannot be interpreted as proof that exploitation never occurred.

2.2 Predictive Vulnerability Prioritization

Machine-learning studies have used vulnerability descrip-tions, technical metadata, exploit references, and external activity to estimate exploitability. Suciu *et al.* emphasized that exploitability evolves after disclosure and that negative labels may contain unknown positives [5]. Early-prediction research has also shown that the timing and availability of public information materially affect reported performance.

Linear models remain useful as interpretable baselines, while tree ensembles can represent nonlinear interactions among heterogeneous features. Adjacent research has explored graph-based risk scoring, privacy-preserving transaction-risk intelligence, and compliance-aware operational resilience. These studies illustrate how technical evidence, risk signals, and governed response processes can be combined to support operational decisions, although they do not directly address future KEV prediction [6], [13], [14]. Evaluations based on random data splits may nevertheless exaggerate performance when related vendors, products, or disclosure periods occur in both training and testing.

2.3 Explainable Decision Support

Security prioritization may lead to emergency patching, service interruption, or accelerated change approval. Analysts therefore need more than a numerical rank. SHAP provides global and local explanations by estimating each feature's con-tribution to a model output [15]. Evidence-grounded incident-response research likewise argues that automated recommen-dations should remain connected to observable technical evi-dence [16]. This principle is relevant to vulnerability prioriti-zation because a ranking should be accompanied by evidence that an analyst can review rather than treated as an automatic patch order.
EVPD differs from earlier work in three respects. First, it aligns EPSS and KEV information with an explicit prediction date. Second, it uses chronological partitions rather than ran-dom splitting. Third, it evaluates both statistical performance and the proportion of future KEV records captured under a fixed remediation budget.

3. Dataset and Experimental Design

3.1 Data Construction and Temporal Target

The merged snapshot contains CVEs published between April 14, 2021 and December 31, 2025. NVD contributes publication dates, CVSS metrics, descriptions, configurations, references, and CWE mappings [3]. CVE List V5 supplements CNA and affected-product records [9]; CWE supplies stan-dardized weakness categories [10]; historical EPSS supplies time-specific exploitation scores; and CISA KEV supplies future exploitation outcomes.
Historical EPSS values were obtained from archived weekly snapshots. For each CVE, the first Wednesday on or after publication was identified, and the EPSS snapshot from the fol-lowing Wednesday was selected. This produced a reproducible publication-to-prediction lag of 7–13 days. The rule approxi-mates an early operational decision point while avoiding the use of present-day EPSS values for historical vulnerabilities.

Let t^*_i denote the prediction date and t^k_i the CISA KEV date-added value. The target is

$$y_i = \begin{cases} 1, & 0 < t_i^k - t_i^* \leq 180, \\ 0, & \text{no KEV inclusion within the mature 180-day window.} \end{cases} \qquad (1)$$

CVEs already listed in KEV on or before t^*_i are excluded from the predictive cohort because they are already known urgent-remediation cases. A zero label therefore means that the CVE was not added to KEV during the observation window; it does not establish that the vulnerability was never exploited. The predictors include CVSS severity and attack character-istics, historical EPSS score and percentile, CWE information, affected-configuration and product counts, description length, and reference indicators. KEV fields, ransomware indica-tors, later EPSS observations, modification timestamps, and outcome-window variables are excluded. Numerical values are imputed using training-set medians, while missing categorical values are represented by an explicit Unknown category.

3.2 Chronological Partitions and Baselines

The data are divided by publication date: April 2021–2023 for training, 2024 for validation, and 2025 for testing. The validation period is used for feature processing, hyperparam-eter selection, calibration, threshold selection, and ensemble construction. The 2025 test set is evaluated once.
The experiment compares logistic regression, random forest, XGBoost, and LightGBM with CVSS, EPSS, and a weighted baseline,

$$S_i = \alpha \left(\frac{\text{CVSS}_i}{10} \right) + (1 - \alpha)\text{EPSS}_i, \qquad (2)$$

where α is selected on validation data. Class-weighted learning is used because future KEV inclusion is rare. PR-AUC is the primary metric; ROC-AUC, precision, recall, F1, Brier score, and expected calibration error are also reported. Ranking utility is measured using Recall@K at 1%, 5%, 10%, and 20% of the test backlog.

3.3 Proposed EVPD Ensemble

EVPD uses temporal stacking. The candidate machine-learning models are trained on CVEs published from April 2021 through December 2023 and compared using the 2024 temporal validation set. Random forest is selected as the learned component of the final ensemble based on its validation-period ranking performance. Its validation-period risk score is combined with historical EPSS and normalized CVSS, and a logistic meta-model learns the final ranking from these three inputs. Five-fold cross-validation within the 2024 validation period is used to select the meta-model reg-ularization parameter. The complete procedure is fixed before evaluation on the untouched 2025 test set.

This arrangement separates three complementary roles. Ran-dom forest captures nonlinear interactions among vulnerability attributes, EPSS supplies an external exploitation-probability signal, and CVSS preserves technical-severity information. The ensemble therefore tests whether these signals provide complementary ranking information rather than treating any one of them as a complete measure of risk.

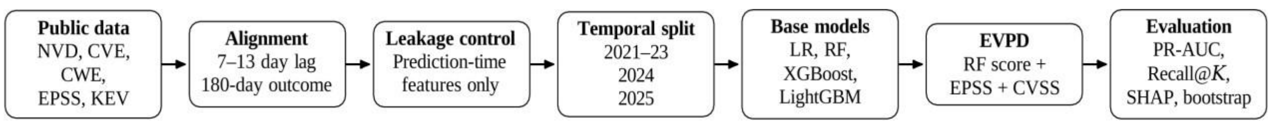

Figure 1. EVPD workflow from temporally aligned vulnerability records to explainable constrained-capacity prioritization.

3.4 Reproducibility and Statistical Protocol

The data snapshot was collected on July 14, 2026. Pre-processing statistics were estimated from the training parti-tion only. Hyperparameters, score thresholds, and calibration. Mappings were selected using the 2024 validation period. The meta-model regularization parameter was selected from $C \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$, with $C = 10^{-4}$ selected.

A 5,000-sample paired bootstrap of the test set was used to estimate uncertainty in the PR-AUC difference between EVPD and the tuned CVSS–EPSS baseline. Random procedures used seed 42.

4. Results

4.1 Dataset and Predictive Performance

The merged snapshot contained 156,444 CVEs. After re-moving records without a mature outcome window and CVEs already in KEV at prediction, 134,875 remained. Only 216 records were positive. Table 1 reports the chronological par-titions. Test prevalence, and therefore no-skill PR-AUC, was approximately 0.0016.

Table 2 reports test performance. EVPD achieved the highest PR-AUC, 0.0600. Random forest achieved the highest ROC-AUC, 0.8300, and LightGBM achieved the strongest threshold-based F_1. CVSS produced comparatively high recall but extremely low precision.

EVPD improved PR-AUC by 78.1% over CVSS–EPSS and by approximately 9.4 times over CVSS. The paired bootstrap mean improvement over CVSS–EPSS was 0.0242, with a 95% confidence interval of 0.0032–0.0511. Because the interval ex-cluded zero, the test data support a positive PR-AUC difference between EVPD and CVSS–EPSS.

Table 1. Chronological Experimental Cohort

Partition	Period	CVEs	KEV positives
Training	2021–2023	50,228	88
Validation	2024	38,358	54
Testing	2025	46,289	74
Total	2021–2025	134,875	216

Table 2. Performance on the 2025 Chronological Test Set

Method	PR-AUC	ROC-AUC	Prec.	Rec.	F1
EVPD ensemble	**.0600**	.8262	.0570	.2027	.0890
Random forest	.0471	**.8300**	.0443	.2568	.0755
LightGBM	.0374	.8299	.0833	**.1216**	.0989
CVSS–EPSS	.0337	.8038	.0635	.1081	.0800
EPSS	.0326	.7778	.0467	.1622	.0725
XGBoost	.0108	.7132	.0325	.1351	.0524
Logistic regression	.0071	.7941	.0122	.1216	.0221
CVSS	.0064	.7543	.0047	.3919	.0093

4.2 Performance Under Limited Remediation Capacity

At a 1% remediation budget, EVPD captured 20 of 74 future KEV vulnerabilities, or 27.0%. Random forest captured 25.7%, CVSS–EPSS and EPSS each captured 17.6%, and CVSS captured 12.2%. At a 5% budget, EVPD captured 45.9%, compared with 44.6% for random forest and CVSS–EPSS, 41.9% for EPSS, and 25.7% for CVSS.

The advantage was concentrated at the smallest budgets. At 10%, CVSS–EPSS captured 56.8%, while EVPD captured 48.6%. EVPD is therefore most useful as a scarce-capacity pri-oritizer rather than a method that dominates at every capacity level.

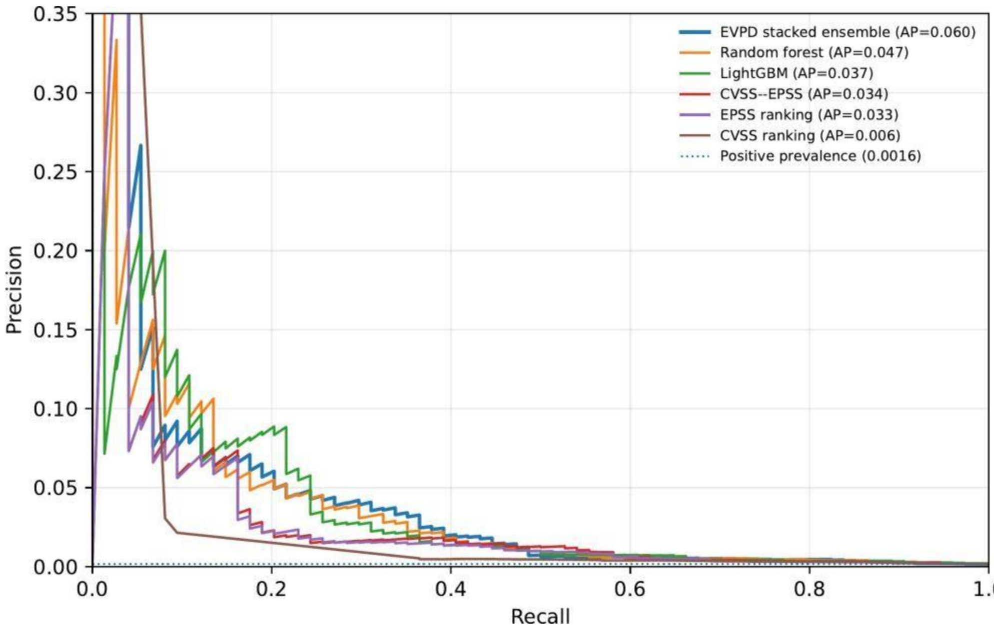

Figure 2. Precision–recall performance on the chronological 2025 test set. The dotted line represents the positive-class prevalence.

Table 3. Recall of Future KEV Records by Remediation Budget

Method	1%	5%	10%	20%
EVPD ensemble	**27.0**	**45.9**	48.6	64.9
Random forest	25.7	44.6	51.4	63.5
CVSS–EPSS	17.6	44.6	**56.8**	63.5
EPSS	17.6	41.9	55.4	64.9
CVSS	12.2	25.7	35.1	55.4

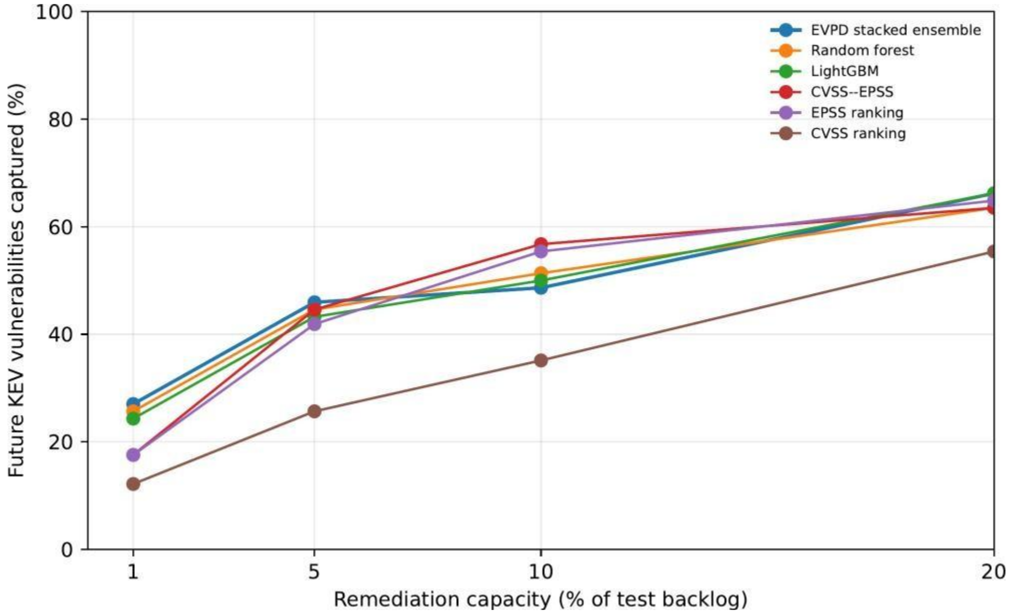

Figure 3. Percentage of future KEV records captured as remediation capacity increases.

4.3 Explainability and Calibration

Historical EPSS percentile and score were the most influen-tial random-forest features, followed by affected-configuration count, CVSS base score, CVSS impact score, description length, and CVSS exploitability score. Feature ordering was highly stable between validation and test periods, with Spear-man ρ = 0.990 and p < 0.001.

The calibrated ensemble obtained a Brier score of 0.00166 and an expected calibration error of 0.00062. These aggregate values are influenced by the very low positive-class prevalence. Inspection of the reliability curve showed overprediction in the highest-score bins, where comparatively few records were available. EVPD is therefore interpreted primarily as a ranking mechanism. Any organization using its output as an absolute probability should recalibrate the model against its own vul-nerability population and observation process.

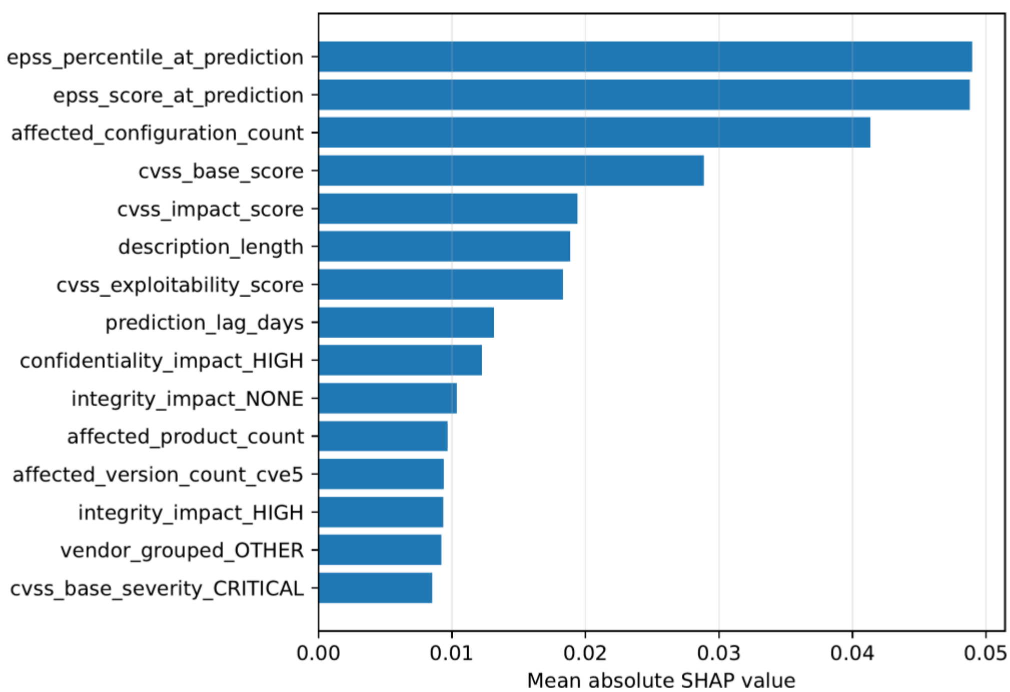

Figure 4. Mean absolute SHAP importance for the random-forest component of EVPD.

4.4 Decision-Threshold Error Profile

At the validation-selected threshold, EVPD identified 15 of the 74 future KEV records and produced 248 false positives. Random forest identified four additional positives but gen-erated 410 false positives. LightGBM produced the highest precision, 0.0833, with 9 true positives and 99 false positives. These differences show why a single threshold metric is insufficient. EVPD was strongest across the full precision–recall curve and at the smallest ranking budgets, while other models offered different precision–recall tradeoffs.

5. Discussion

The main finding is not simply that one ensemble achieved a higher aggregate score. The experiment shows that temporal alignment, future-outcome labeling, and capacity-aware evalu-ation materially affect how vulnerability-prioritization systems should be assessed. EVPD was evaluated using information available near disclosure, tested on a later calendar year, and judged by the number of future KEV records recovered within realistic remediation limits. This design provides a more operationally meaningful evaluation than random splitting or severity-only comparison.

For RQ1, future KEV inclusion was predictable above the no-skill baseline despite a test prevalence of only 0.16%. EVPD achieved a PR-AUC of 0.0600, approximately 37.5 times the no-skill value of 0.0016. The absolute PR-AUC nevertheless remained modest because future KEV inclusion is rare and KEV does not observe every exploitation event. The result should therefore be interpreted as evidence of useful ranking ability rather than complete exploit prediction.

For RQ2, EVPD achieved the highest test PR-AUC and significantly outperformed the tuned CVSS–EPSS baseline in the paired bootstrap analysis. Its practical advantage was most evident when remediation was restricted to the top 1% of the backlog. Under that constraint, EVPD captured 27.0% of future KEV records, compared with 17.6% for CVSS–EPSS. The ensemble did not dominate at every budget. At 10% capacity, CVSS–EPSS captured more positives.

An important contribution is the separation between model quality and operational usefulness. EVPD achieved the high-est overall PR-AUC and performed best at the 1% and 5% remediation budgets, while CVSS–EPSS recovered more positives at the 10% budget. The preferred ranking method therefore depends on available remediation capacity. Rather than proposing a universal replacement for existing scores, the findings support a layered policy in which EVPD drives a nar-row emergency-review queue and broader score combinations support secondary review.

For RQ3, historical EPSS was the strongest predictor, but it was not sufficient by itself. Configuration breadth, CVSS impact, exploitability, and disclosure features added useful information. The model therefore distinguished between a technically severe flaw and a vulnerability that combines exploitation probability with broad exposure.

For RQ4, feature rankings were highly consistent between the validation and test periods, with Spearman ρ = 0.990. This stability is encouraging, but it should not be interpreted as permanent. EPSS models, vendor reporting practices, and attacker behavior change over time. Continued monitoring, recalibration, and periodic retraining are therefore necessary. The post-Mythos framing does not imply that proprietary Mythos findings were used as training data. Instead, Mythos motivates the operational problem addressed by the paper: defenders may receive more findings without a proportional increase in patching capacity. EVPD provides a transparent process for ranking that larger stream, but prospective evalua-tion on post-April 2026 disclosures will only be possible once their observation windows have matured.

5.1 Operational Use

Known KEV vulnerabilities should bypass predictive scor-ing and enter an urgent-remediation process. EVPD is intended for the remaining backlog, where exploitation has not yet been confirmed. Each prioritized record should be presented with its score, the major SHAP contributors, affected-product evidence, source timestamps, and the applicable CVSS and EPSS values. Analysts should retain authority to change the ranking based on asset criticality, internet exposure, compensating controls, patch availability, and business impact. These organization-specific factors are intentionally outside the public-data model and must be added before deployment.

5.2 Threats to Validity

CISA KEV is authoritative for the vulnerabilities it con-tains, but it is not a complete inventory of every exploitation event. Some records labeled negative may therefore have been exploited without entering KEV. The 180-day window also excludes exploitation occurring later.

Only 74 positive records occurred in the test year. The paired bootstrap interval supported the improvement over CVSS–EPSS, but estimates for individual thresholds and bud-get points remain sensitive to a small number of records.

Historical EPSS and KEV dates were temporally aligned, whereas NVD and CVE attributes were collected from a fixed snapshot.

Some descriptions, references, or configurations may have been added after the original prediction date. This residual temporal contamination is explicitly acknowledged. A stronger replication would use archived CVE and NVD records for every prediction date.

Finally, the model does not include local asset inventory, deployment prevalence, business criticality, control effective-ness, or patch cost. The results establish public-data ranking performance, not complete enterprise risk.

6. Conclusion

This paper introduced EVPD, a temporally aligned and explainable framework for ranking newly disclosed vulnera-bilities by future CISA KEV inclusion. On the chronological 2025 test set, the stacked ensemble achieved a PR-AUC of 0.0600, compared with 0.0337 for a tuned CVSS–EPSS baseline. When only 1% of the backlog could be selected, EVPD captured 27.0% of the vulnerabilities that later entered KEV.

The experiment also showed that no single input was sufficient. Historical EPSS provided the strongest signal, while affected-configuration breadth and CVSS impact added complementary information. The learned feature ordering re-mained stable across the validation and test periods.

EVPD should be used as an analyst-facing ranking system, not as an automatic patch order or a universally calibrated exploitation probability. Future work should use archived CVE snapshots, incorporate organization-specific exposure data, test alternative outcome windows, and evaluate matured post-Mythos disclosures.

Funding: This research received no external funding

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

1. Anthropic Frontier Red Team, "Assessing Claude Mythos Preview's Cybersecurity Capabilities," Apr. 7, 2026. [Online]. Available: https://red.anthropic.com/2026/mythos-preview/. Accessed: Jul. 18, 2026.
2. Forum of Incident Response and Security Teams, "Common Vul-nerability Scoring System Version 3.1: Specification Document," FIRST, 2019. [Online]. Available: https://www.first.org/cvss/v3.1/specification-document. Accessed: Jul. 14, 2026.
3. National Institute of Standards and Technology, "National Vulnerability Database: Vulnerability Data Feeds," 2026. [Online]. Available: https://nvd.nist.gov/vuln/data-feeds. Accessed: Jul. 14, 2026.
4. M. Bozorgi, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond Heuris-tics: Learning to Classify Vulnerabilities and Predict Exploits," in *Proc. 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 105–114, 2010, doi: 10.1145/1835804.1835821.
5. O. Suciu, C. Nelson, Z. Lyu, T. Bao, and T. Dumitraș, "Expected Exploitability: Predicting the Development of Functional Vulnerability Exploits," in *Proc. 31st USENIX Security Symposium*, pp. 377–394, 2022.
6. V. T. Manne, "Privacy-Preserving Chargeback Intelligence for Tokenized Payment Systems," *Journal of Computer Science and Technology Studies*, vol. 5, no. 2, pp. 54–65, 2023, doi: 10.32996/jcsts.2023.5.2.5.
7. J. Jacobs, S. Romanosky, B. Edwards, M. Roytman, and I. Adjerid, "Exploit Prediction Scoring System," *Digital Threats: Research and Practice*, vol. 2, no. 3, pp. 1–17, 2021, doi: 10.1145/3436242.
8. Cybersecurity and Infrastructure Security Agency, "Known Exploited Vulnerabilities Catalog," 2026. [Online]. Available: https://www.cisa.gov/known-exploited-vulnerabilities-catalog. Accessed: Jul. 14, 2026.
9. CVE Program, "CVE List Downloads in CVE JSON 5 Format," Mar. 29, 2023. [Online]. Available: https://www.cve.org/Media/News/item/blog/2023/03/29/CVE-Downloads-in-JSON-5-Format. Accessed: Jul. 14, 2026.
10. The MITRE Corporation, "Common Weakness Enumeration: CWE List and Downloads," 2026. [Online]. Available: https://cwe.mitre.org/data/downloads.html. Accessed: Jul. 14, 2026.
11. L. Allodi and F. Massacci, "Comparing Vulnerability Severity and Exploits Using Case-Control Studies," *ACM Transactions on Infor-mation and System Security*, vol. 17, no. 1, pp. 1–20, 2014, doi: 10.1145/2630069.
12. A. D. Householder, J. Chrabaszcz, T. Novelly, D. Warren, and J. M. Spring, "Historical Analysis of Exploit Availability Timelines," in *Proc. Cyber Security Experimentation and Test Workshop*, 2020.
13. V. T. Manne, "Adaptive Graph-Based Risk Scoring for Real-Time Instant Payment Systems," in *Proc. 2026 Second International Conference on Intelligent Systems for Communication, IoT and Security*, pp. 1–7, 2026, doi: 10.1109/ICISCoIS62701.2026.11447558.
14. V. T. Manne and S. B. Venkata, "Compliance-Aware Payment Gateway Operations for Healthcare and Medical Device Ecosystems: A DevOps Framework for Resilient Digital Transaction Systems," *Journal of Com-puter Science and Technology*

Studies, vol. 6, no. 1, pp. 351–371, 2024, doi: 10.32996/jcsts.2024.6.1.37.
15. S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.
16. S. B. Venkata, "Evidence-Gated Search: Controlling Operational Search Explosion in LLM-Driven Incident Response," *Journal of Computer Science and Technology Studies*, vol. 8, no. 5, pp. 106–120, 2026, doi: 10.32996/jcsts.2026.8.5.8.