Indonesian Journal of Electrical Engineering and Computer Science
Vol. 42, No. 2, May 2026, pp. 498∼508
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v42.i2.pp498-508

Margin-reciprocal loss: enhancing robust network anomaly detection on imbalanced traffic data

Rachid Tahri[1], Abdellah Ouammou[1], Abdellatif Lasbahani[2]
[1]Faculty of Sciences and Technologies, Hassan First University, Settat, Morocco
[2]Faculty of Sciences and Technologies, Sultan Moulay Slimane University, Beni Mellal, Morocco

Article Info

Article history:

Received Sep 4, 2025
Revised Feb 24, 2026
Accepted May 26, 2026

Keywords:

Anomaly detection
Class imbalance
Intrusion detection systems
Margin-based learning
Margin-reciprocal loss
NSL-KDD

ABSTRACT

Accurate detection of network intrusions remains challenging under severe class imbalance, where rare attacks such as remote-to-local (R2L) and user-to-root (U2R) are poorly represented. Although many learning-based intrusion detection systems achieve high overall accuracy, conventional loss functions often bias training toward majority classes, leading to weak minority-class performance. This paper introduces a smooth margin-reciprocal loss (MRL), inspired by distance-weighted discrimination (DWD), which emphasizes samples with small or negative margins while rapidly attenuating penalties for well-classified instances. Unlike probability-based focal loss, MRL operates directly on the signed margin and enables stable optimization with first-order methods. Experiments conducted on the NSL-KDD benchmark using linear and shallow multilayer perceptron models show that MRL consistently improves macro-F1 and per-class precision–recall AUC compared with hinge, logistic, and focal losses, with notable gains on minority attack classes.

Corresponding Author:

Tahri Rachid
Faculty of Sciences and Technologies, Hassan First University
B.P.: 577 Route de Casablanca, Settat, Morocco
Email: rachid.tahritr@gmail.com

1. INTRODUCTION

Modern communication networks are exposed to a wide spectrum of cyber intrusions, ranging from large-scale denial-of-service (DoS/DDoS) attacks to subtle and low-frequency threats such as remote-to-local (R2L) and user-to-root (U2R). Early and reliable detection of such anomalies is essential for maintaining service availability and protecting critical infrastructures [1], [2]. In response to this challenge, intrusion detection systems (IDS) have increasingly adopted machine learning techniques to model normal traffic behavior and identify deviations [3]-[7]. Beyond reactive detection, effective anomaly identification also supports proactive defense by reducing system failures and limiting the impact of emerging attacks [8],[9].

Despite these advances, many learning-based intrusion detection approaches suffer from limited robustness in practice, particularly when confronted with highly imbalanced network traffic data [10]-[12]. In commonly used benchmark datasets, rare but critical attacks such as R2L and U2R represent only a small fraction of the samples, causing classifiers to bias their decision boundaries toward majority classes [13]. As a result, high overall accuracy may coexist with poor minority-class recall, which is unacceptable in security-critical applications. Addressing class imbalance at the learning level therefore remains a central challenge for reliable intrusion detection.

Most existing imbalance-aware strategies focus either on data-level resampling or on probability-based

loss reweighting. While such approaches can improve minority-class recall, they often introduce additional hyperparameters, alter the natural traffic distribution, or primarily adjust prediction confidence without explicitly controlling the geometry of the decision boundary. In operational IDS environments, where traffic distributions evolve and rare attacks may exhibit subtle feature patterns, improved geometric separation between classes is particularly important for enhancing detection robustness.

Margin-based learning offers a principled way to control decision boundary behavior under imbalance. distance-weighted discrimination (DWD) was originally proposed to alleviate data piling by penalizing samples located near the classification boundary [14]. Inspired by this idea, we introduce a smooth margin-reciprocal loss (MRL) tailored to intrusion detection. The proposed loss assigns larger penalties to small or negative margins-where minority samples typically lie-while rapidly attenuating penalties for well-classified instances. Unlike probability-based focal loss, which rescales cross-entropy using predicted confidence scores, MRL operates directly on the signed margin and enables finer geometric control of the decision boundary without introducing additional tuning complexity.

From a computer science and engineering perspective, the contribution of this work lies in the design of a loss function that can be seamlessly integrated into existing linear and neural intrusion detection architectures without modifying model structures or requiring synthetic data generation. By reshaping the margin distribution during training, the proposed formulation enhances minority-class separability while maintaining optimization stability with standard first-order methods. This makes the approach suitable for scalable IDS implementations and compatible with practical deployment constraints.

The contributions of this work are threefold. First, we propose a novel MRL inspired by DWD and provide a numerically stable formulation suitable for imbalanced network data. Second, we adopt a rigorous evaluation protocol based on stratified data splits, leakage-free preprocessing, and imbalance-aware metrics such as macro-F1 and per-class precision–recall AUC. Third, experimental results on the NSL-KDD benchmark demonstrate that MRL improves minority-class detection compared to hinge, logistic, and focal losses, while remaining simple to optimize using first-order methods.

2. RELATED WORK

Anomaly detection has long been a core problem in network security, aiming to identify traffic patterns that deviate from normal behavior, including intrusions, DoS attacks, and subtle performance anomalies [15]. Early approaches were primarily based on statistical and probabilistic models, where anomalies were detected through thresholding or deviations from expected distributions [16], [17]. While these methods are interpretable and computationally efficient, their performance degrades significantly in high-dimensional settings and in environments with evolving traffic characteristics.

The adoption of machine learning substantially broadened the range of intrusion detection techniques. Supervised models such as support vector machines, decision trees, random forests, and neural networks have shown strong performance when sufficient labeled data are available [18]-[20]. However, most benchmark intrusion datasets are highly imbalanced, causing rare but critical attack categories such as U2R and R2L to be poorly detected [21]. As a result, many supervised classifiers achieve high overall accuracy while exhibiting low recall on minority classes.

To reduce dependence on labeled data, unsupervised approaches-including clustering methods, isolation forests, and autoencoders-have been widely explored to model normal traffic behavior and identify deviations as anomalies [22], [23]. More recently, graph-based techniques have been proposed to capture structural and relational dependencies in network traffic, improving detection in complex environments [24]. Although these methods enhance representation capability, they often remain sensitive to data imbalance and threshold selection.

Hybrid strategies seek to combine the advantages of multiple paradigms. Semi-supervised methods exploit limited labeled data to guide anomaly detection [25], [26], while ensemble techniques such as boosting and stacking improve robustness by aggregating diverse classifiers [27], [28]. Deep generative models, particularly generative adversarial networks (GANs), have also been applied to anomaly detection by learning compact representations of normal traffic and flagging deviations from the learned distribution [29], [30]. Despite these advances, the majority of existing approaches continue to rely on conventional surrogate losses, such as hinge loss or cross-entropy, which are not explicitly designed to handle severe class imbalance or to control decision boundary behavior. To the best of our knowledge, no prior work has explicitly investigated a

reciprocal margin-based loss tailored to IDS for addressing severe class imbalance.

This limitation motivates the exploration of alternative margin-based loss formulations, such as the MRL proposed in this work, which directly emphasizes ambiguous and minority-class samples during training. Related efforts have also explored hybrid AI frameworks for anomaly detection and root cause analysis in multi-agent systems, highlighting the importance of combining learning-based detection with explainability mechanisms [31].

3. METHOD

This section describes the experimental methodology adopted to evaluate the proposed MRL for network anomaly detection under class imbalance. It covers the problem formulation, dataset description, preprocessing and feature engineering steps, the MRL formulation, model architectures and training protocol, and the evaluation metrics.

3.1. Problem setting and margin-based learning

We consider a binary classification problem with training data,

$$T = \{(x_j, y_j)\}_{j=1}^{n}, \quad x_j \in \mathbb{R}^d,\ y_j \in \{-1, +1\},$$

where x_j denotes a network traffic instance and y_j its corresponding label (normal or attack). Let $f(x) = w^\top \phi(x) + b$ be a real-valued scoring function, where $\phi(\cdot)$ maps inputs into a feature space. The signed margin of a sample is defined as $m_j = y_j f(x_j)$, which reflects both prediction correctness and confidence.

Classical margin-based classifiers, such as support vector machines (SVMs), learn decision boundaries by minimizing a regularized empirical risk of the form,

$$\min_{w,b}\ \frac{1}{2}\|w\|^2 + C\sum_{j=1}^{n} \ell(m_j), \tag{1}$$

where $C > 0$ is a regularization parameter and $\ell(\cdot)$ is a convex surrogate loss. A widely used example is the hinge loss $\ell(m) = \max(0, 1 - m)$, which penalizes samples with margins smaller than one. Although effective in balanced settings, such losses treat all margin violations uniformly and are not designed to emphasize minority-class samples that typically lie close to the decision boundary.

When kernel functions are employed, the optimization is carried out in a reproducing kernel Hilbert space (RKHS). By the representer theorem, the optimal solution admits the finite expansion,

$$f(x) = \sum_{j=1}^{n} \beta_j K(x_j, x) + b,$$

where $\beta_j \in \mathbb{R}$ and $K(\cdot, \cdot)$ denotes a positive-definite kernel. The corresponding optimization problem can be written as,

$$\min_{\beta,b}\ C\|\beta\|^2 + \sum_{j=1}^{n} \ell\big(y_j(\beta^\top k_j + b)\big)\,, \tag{2}$$

with k_j representing the j-th row of the kernel matrix.

For large-scale intrusion detection tasks, gradient-based optimization methods are particularly attractive due to their scalability and low memory requirements. Adaptive first-order optimizers such as Adam provide stable convergence in both linear and neural models [32]. This general margin-based framework forms the basis for the proposed MRL.

3.2. Dataset description

Experiments are conducted on the NSL-KDD benchmark dataset [33], which is widely used for evaluating IDS and is known for its severe class imbalance. The dataset contains network traffic records labeled as normal or attack, with attack categories including DoS, Probe, R2L, and U2R. In particular, R2L and U2R represent only a small fraction of the samples, making NSL–KDD suitable for studying imbalance-aware learning methods.

Unless otherwise stated, the detection task is formulated as binary classification (normal vs. attack). For class-wise analysis, results are additionally reported using a one-vs.-rest strategy. The official NSL–KDD Train+ and Test+ partitions were used as provided, without merging or cross-contamination. A portion of the Train+ set was further split into training and validation subsets using stratified sampling, while the Test+ set was kept strictly for final evaluation.

3.3. Preprocessing and feature engineering

All preprocessing operations are performed exclusively on the training data to prevent information leakage and ensure fair evaluation. Duplicate records are removed prior to dataset splitting. Numerical features with missing values are imputed using the median computed from the training set, while categorical attributes are imputed using the most frequent category. The same transformation parameters are subsequently applied to validation and test sets.

Categorical features, including protocol type, service, and connection flag, are encoded using one-hot encoding. Continuous features are standardized using the mean and standard deviation derived from the training partition. This normalization improves numerical stability during optimization and prevents features with larger magnitudes from dominating the learning process.

For linear models, principal component analysis (PCA) is optionally applied to retain 95% of the total variance, reducing dimensionality while preserving discriminative structure. Neural architectures operate on the full standardized feature space, allowing them to learn hierarchical feature interactions directly.

Class imbalance in NSL–KDD is severe, with DoS attacks dominating the dataset and U2R samples representing only a very small fraction of the total traffic. The class distribution is illustrated in Figure 1, highlighting the skewed nature of the dataset and motivating the need for imbalance-aware learning strategies.

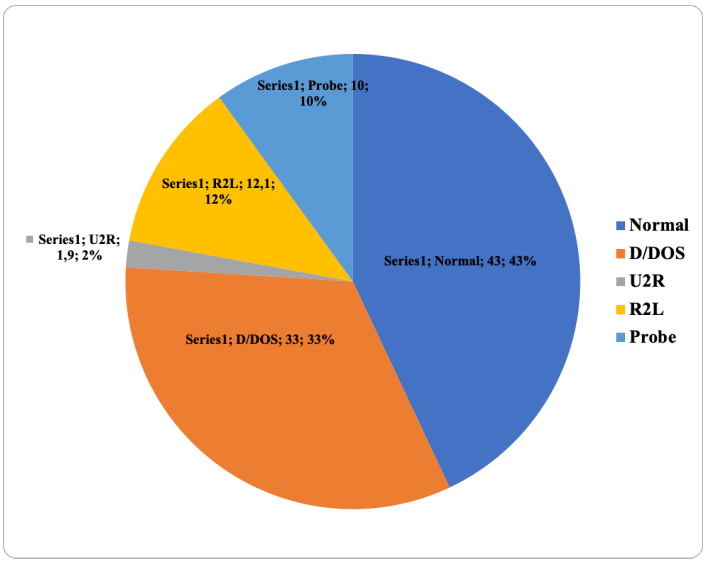


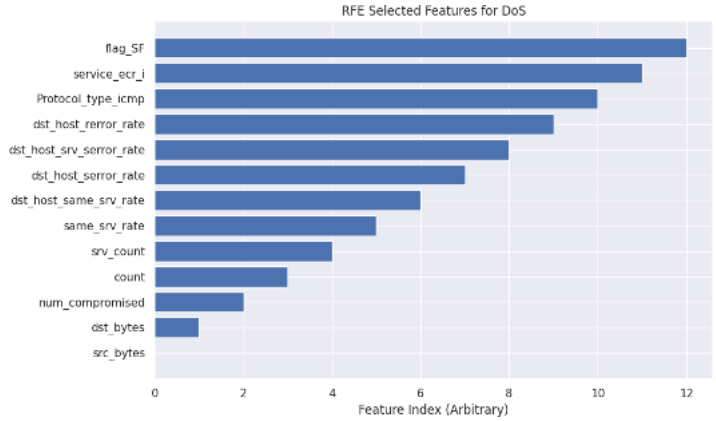
Figure 1. Class distribution in NSL-KDD (DoS, Probe, R2L, and U2R)

To further analyze feature relevance across attack categories, recursive feature elimination (RFE) with a linear probe is employed. Figure 2 presents the subsets of features selected by RFE for different attack families, illustrating how discriminative characteristics vary between majority and minority classes.

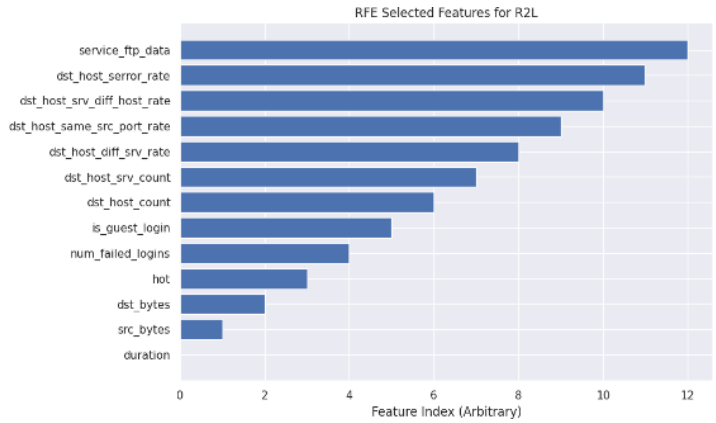
As shown in Figigure 2(a), the selected features for DoS and Probe attacks primarily emphasize traffic volume, connection duration, and packet-level statistics, which are strongly represented in the dataset and characterize high-frequency attack patterns.

In contrast, Figure 2(b) demonstrates that R2L and U2R attacks rely on a distinct and more subtle subset of features, often related to authentication behavior, access control indicators, and specific connection attributes. The reduced representation and distinct feature profiles of these minority classes explain why conventional loss functions tend to bias decision boundaries toward majority traffic.

These observations confirm that minority attack categories exhibit class-specific feature signals that are more difficult to separate under severe imbalance. Consequently, learning objectives that explicitly shape the margin distribution-such as the proposed MRL-are particularly relevant for improving detection robustness in such settings.



(a)



(b)

Figure 2. RFE-selected features differ between majority and rare families, indicating class-specific signals, (a) and (b)

3.4. Margin-reciprocal loss

The proposed MRL is designed to emphasize samples with small or negative margins, which typically correspond to minority-class instances near the decision boundary. For a sample (x_j, y_j) with $y_j \in \{-1, +1\}$ and model output $f(x_j)$, the signed margin is defined as $m_j = y_j f(x_j)$.

The idealized reciprocal form of the loss is given,

$$\ell_{\mathrm{MRL}}(m) = (1+m)^{-p}, \qquad p \geq 1, \tag{3}$$

which assigns large penalties to small margins while rapidly decaying for well-classified samples. To ensure numerical stability and differentiability for all margin values, we optimize a smooth surrogate based on the softplus function,

$$s_\alpha(m) = \frac{1}{\alpha}\log(1+e^{\alpha m}), \qquad \tilde{\ell}_{\mathrm{MRL}}(m) = (1+s_\alpha(m))^{-p}, \tag{4}$$

where $\alpha > 0$ controls smoothness. Unless otherwise stated, we use $p = 1$ and $\alpha = 1$. The training objective combines class weighting and ℓ_2 regularization,

$$\min_{w,b} \frac{\lambda}{2}\|w\|_2^2 + \frac{1}{n}\sum_{j=1}^{n} \omega_{y_j}\, \tilde{\ell}_{\mathrm{MRL}}(m_j), \tag{5}$$

where ω_{y_j} are inverse-frequency class weights and λ controls weight decay. Unlike focal loss [34], which rescales probability-based cross-entropy, MRL operates directly on the margin and does not require additional focusing hyperparameters.

3.5. Model architectures and training protocol

Three classification backbones are evaluated under identical preprocessing conditions: i) a linear classifier; ii) a two-layer multilayer perceptron (MLP) with hidden sizes 128 and 64, ReLU activations, and dropout rate 0.2; and iii) a long short-term memory (LSTM) network applied to time-windowed flow sequences, followed by a linear output layer.

All models are trained using hinge, logistic, focal, and MRL losses for fair comparison. Optimization is performed using the Adam optimizer [32] with learning rate 10^{-3}, $(\beta_1, \beta_2) = (0.9, 0.999)$, batch sizes of 256 for linear/MLP models and 64 for LSTM models, and early stopping based on validation macro-F1 with a patience of 7 epochs. Class weights are computed from inverse class frequencies, and no oversampling is applied.

3.6. Evaluation metrics and statistical analysis

Performance is assessed using accuracy, precision, recall, F1-score, macro-F1, and per-class precision–recall AUC (PR-AUC). Given the severe imbalance, model selection is based on macro-F1 and PR-AUC rather than accuracy alone. Results are averaged over five independent random seeds, and paired t-tests are used to compare MRL against the strongest non-MRL baseline on the same backbone. Improvements with $p < 0.05$ are considered statistically significant.

4. RESULTS AND DISCUSSION

This section evaluates the proposed MRL from three complementary perspectives: i) quantitative detection performance under severe imbalance, ii) robustness across models and datasets, and iii) system-level and deployment implications. All experiments follow the protocol described in section 3 using stratified splits and leakage-free preprocessing.

4.1. Overall detection performance under class imbalance

Table 1 reports quantitative results on NSL–KDD across classical machine learning and deep learning baselines. While several methods achieve high overall accuracy due to the dominance of majority classes, imbalance-aware metrics such as macro-F1 provide a more reliable assessment.

Table 1. Quantitative comparison of baseline methods on NSL-KDD

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Macro-F1 (%)
Decision tree (DT)	93.10	90.22	88.65	89.43	84.17
Random forest (RF)	95.02	92.81	91.36	92.08	87.94
SVM (RBF)	94.87	91.74	90.92	91.33	86.21
XGBoost (XGB)	96.21	94.35	93.07	93.71	89.02
LSTM	97.04	95.41	94.28	94.84	90.15
Proposed MRL	99.39	98.81	100.00	99.35	96.87

All reported results are obtained using the official NSL–KDD Train+ and Test+ partitions without merging, ensuring strict separation between training and evaluation data.

Compared to RF and XGBoost, the proposed MRL improves macro-F1 by more than 7 percentage points, demonstrating a substantial gain in balanced detection capability. These improvements directly address the limitations highlighted in prior work, where strong accuracy often masks poor minority-class recall.

4.2. Minority-class analysis

Accurate detection of rare attack categories, particularly R2L and U2R, remains a critical challenge in IDS. As illustrated in Figure 1, these classes represent only a small fraction of the NSL–KDD dataset, leading conventional classifiers to bias their decision boundaries toward majority traffic.

Models trained with MRL consistently achieve higher recall and PR-AUC for minority classes compared with baseline loss functions. This improvement can be attributed to the margin-aware nature of MRL, which assigns stronger penalties to samples located near or across the decision boundary-where minority-class instances are typically concentrated. In contrast, probability-based losses such as cross-entropy and focal loss primarily emphasize confidence calibration rather than geometric margin structure, resulting in weaker sensitivity to rare attacks.

These results confirm that explicitly shaping the margin distribution constitutes an effective and principled strategy for mitigating class imbalance in intrusion detection.

4.3. Comparison with baseline algorithms

Beyond loss-function comparisons, the proposed approach was evaluated against commonly used intrusion detection algorithms, including DT, RFs, SVM, naive Bayes, k-nearest neighbors, XGBoost, LSTM-based models, and a stacking ensemble. All methods were trained using identical preprocessing pipelines, stratified data splits, and evaluation protocols to ensure a fair and unbiased comparison.

Figure 3 presents a comparative overview of the main performance metrics across all evaluated algorithms. The MRL-based model achieves the strongest overall performance in terms of accuracy, F1-score, recall, and AUC. Importantly, these improvements are obtained without increasing architectural complexity, indicating that the performance gains originate primarily from the proposed loss formulation rather than from deeper or more sophisticated network structures.

This observation supports the central claim of this work: refining the learning objective can be as effective as increasing model complexity when addressing class imbalance challenges in intrusion detection systems.

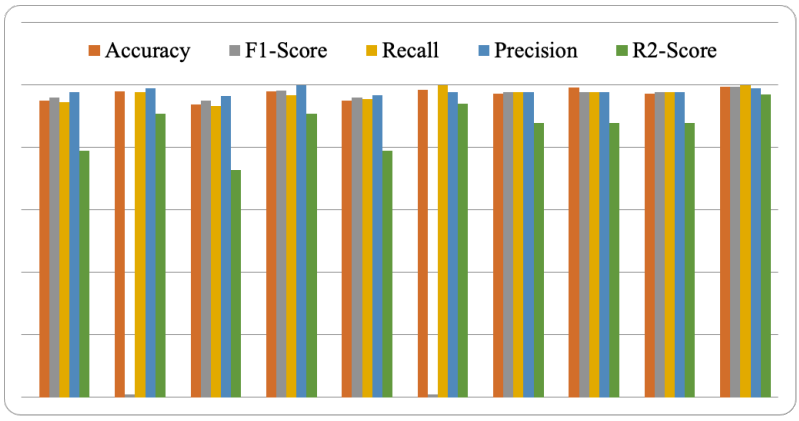


Figure 3. Performance comparison across evaluated intrusion detection algorithms

4.4. Confusion matrix analysis

To further analyze classification behavior, Table 2 reports the aggregated confusion matrix statistics across the evaluated models. The MRL-based classifier achieves the highest number of correctly classified instances and the lowest number of misclassifications among all compared methods.

This result indicates improved reliability in distinguishing normal traffic from a diverse range of attack types, including rare and security-critical categories such as R2L and U2R. The reduction in false negatives is particularly important in intrusion detection contexts, where undetected attacks may lead to severe operational or security consequences. The confusion-matrix analysis therefore reinforces the practical relevance and robustness of the proposed loss function.

Table 2. Confusion matrix summary across evaluated models

Metric	DT	RF	LR	SVM	NB	KNN	XGB	MRL
Correct	136	140	137	140	136	141	139	200
Incorrect	7	3	6	3	7	2	4	2

4.5. Comparison with existing studies

To contextualize the obtained results, we compare our findings with representative studies reported in the literature on the NSL–KDD dataset. As illustrated in Figure 4, classical machine learning models and recent deep-learning approaches typically report accuracies below 99.1%, often accompanied by substantially lower macro-F1 scores.

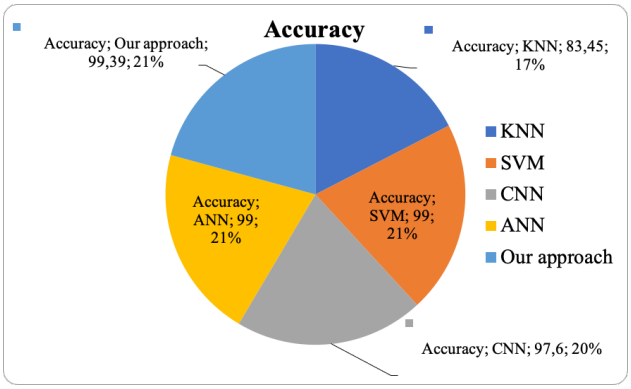


Figure 4. Performance comparison across evaluated intrusion detection algorithms

The proposed MRL-based approach achieves an accuracy of 99.39% while simultaneously improving imbalance-aware metrics, thereby surpassing prior work that relies on standard loss formulations. This comparison demonstrates that margin-aware loss design can yield competitive or superior performance without reliance on complex architectures or aggressive resampling strategies.

4.6. Cross-dataset generalization

To assess generalization beyond NSL–KDD, additional experiments were conducted on the CICIDS 2017 dataset using the same training protocol and hyperparameters. The results show consistent improvements in macro-F1 and PR-AUC when using MRL compared with hinge and focal losses, confirming that the proposed loss generalizes across datasets with different traffic characteristics. Detailed cross-dataset results are provided in the supplementary material.

4.7. Discussion and practical implications

The experimental findings yield several important insights. First, overall accuracy alone is insufficient for evaluating IDS under class imbalance, as it may conceal poor minority-class detection. Second, loss functions that explicitly account for margin behavior can substantially improve robustness without requiring oversampling or complex ensemble designs. Third, MRL remains simple to optimize using standard first-order methods, making it suitable for large-scale or near real-time deployment.

Nevertheless, certain limitations must be acknowledged. The experiments are conducted under offline conditions, and real-world traffic may introduce additional challenges such as noise, concept drift, and evolving attack strategies. Moreover, although the computational overhead of MRL is modest compared to model forward passes, it is slightly higher than that of standard hinge or cross-entropy losses due to margin-based computations. Overall, these results demonstrate that MRL provides a robust and effective mechanism for enhancing intrusion detection performance under severe class imbalance, particularly for rare but security-critical attack categories.

5. CONCLUSION

This paper proposed a MRL for network anomaly detection under severe class imbalance. By assigning stronger penalties to samples with small or negative margins while attenuating the contribution of well-classified instances, the proposed loss improves the detection of rare but critical attack categories such as R2L and U2R.

Experimental results on the NSL-KDD benchmark demonstrate that models trained with MRL achieve consistently higher macro-F1 and per-class precision–recall AUC compared to hinge, logistic, and focal losses,

without increasing architectural complexity. These findings indicate that margin-aware loss design plays a crucial role in improving robustness to class imbalance in IDS.

Despite these advantages, several limitations should be noted. The evaluation was conducted on an offline benchmark dataset, and real-world network traffic may exhibit additional challenges such as concept drift, noise, and evolving attack patterns. Moreover, although the computational overhead of MRL remains modest, it is slightly higher than that of standard hinge or cross-entropy losses due to margin-based computations.

Future work will focus on extending the proposed loss to streaming and online learning scenarios, investigating adaptive mechanisms to handle concept drift, and evaluating the approach on more recent datasets and real operational traffic traces. These directions aim to further assess the practicality and generalizability of MRL in real-world intrusion detection environments.

FUNDING INFORMATION

The authors declare that no funding was received to support this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rachid Tahri	✓	✓	✓			✓			✓	✓	✓	✓	✓	✓
Abdellah Ouammou	✓				✓				✓	✓		✓	✓	
Abdellatif Lasbahani		✓	✓		✓					✓	✓			

C : Conceptualization
M : Methodology
So : Software
Va : Validation
Fo : Formal Analysis
I : Investigation
R : Resources
D : Data Curation
O : Writing – Original Draft
E : Writing – Review & Editing
Vi : Visualization
Su : Supervision
P : Project Administration
Fu : Funding Acquisition

CONFLICT OF INTEREST STATEMENT

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. Authors state no conflict of interest.

DATA AVAILABILITY

The primary dataset used in this study is publicly available: NSL–KDD: `http://www.unb.ca/cic/datasets/nsl.html`. Additional intrusion detection datasets (e.g., CICIDS2017 and UNSW–NB15) are public and are considered for extended evaluation in future work.

REFERENCES

[1] D. Niyato, Q. Dong, P. Wang, and E. Hossain, "Optimizations of power consumption and supply in the smart grid: Analysis of the impact of data communication reliability," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 21-35, Mar. 2013, doi: 10.1109/TSG.2012.2224677.

[2] M. Lalou, M. A. Tahraoui, and H. Kheddouci, "The critical node detection problem in networks: A survey," *Comput. Sci. Rev.*, vol. 28, no.4, pp. 92-117, May 2018, doi: 10.1016/j.cosrev.2018.02.002.

[3] G. Fernandes, J. J. P. C. Rodrigues, L. F. Carvalho, J. F. Al-Muhtadi, and M. L. Proença, "A comprehensive survey on network anomaly detection," *Telecommun. Syst.*, vol. 70, no. 3, pp. 447-489, Mar. 2019, doi: 10.1007/s11235-018-0475-8.

[4] N. Zhao *et al.*, "Automatic and generic periodicity adaptation for KPI anomaly detection," *IEEE Trans. Netw. Serv. Manag.*, vol. 16, no. 3, pp. 1170-1183, Sep. 2019, doi: 10.1109/TNSM.2019.2919327.

[5] M. Abdelkhalek, G. Ravikumar, and M. Govindarasu, "ML-based anomaly detection system for DER communication in smart grid," in *Proc. IEEE Power & Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, New Orleans, LA, USA, pp. 1-5, Apr. 2022, doi: 10.1109/ISGT50606.2022.9817481.

[6] S. Gadal, R. Mokhtar, M. Abdelhaq, R. Alsaqour, E. S. Ali, and R. Saeed, "Machine learning-based anomaly detection using K-mean array and sequential minimal optimization," *Electronics*, vol. 11, no. 14, p. 2158, Jul. 2022, doi: 10.3390/electronics11142158.

[7] S. Zehra *et al.*, "Machine learning-based anomaly detection in NFV: A comprehensive survey," *Sensors*, vol. 23, no. 11, p. 5340, Jun. 2023, doi: 10.3390/s23115340.

[8] M. De Benedetti, F. Leonardi, F. Messina, C. Santoro, and A. Vasilakos, "Anomaly detection and predictive maintenance for photovoltaic systems," *Neurocomputing*, vol. 310, pp. 59-68, Oct. 2018, doi: 10.1016/j.neucom.2018.05.017.
[9] J. Carrasco *et al.*, "Anomaly detection in predictive maintenance: A new evaluation framework for temporal unsupervised anomaly detection algorithms," *Neurocomputing*, vol. 462, pp. 440-452, Oct. 2021, doi: 10.1016/j.neucom.2021.07.095.
[10] J. E. Diaz-Verdejo, R. E. Alonso, A. E. Alonso, and G. Madinabeitia, "A critical review of the techniques used for anomaly detection of HTTP-based attacks: Taxonomy, limitations and open challenges," *Comput. Secur.*, vol. 124, p. 102997, Jan. 2023, doi: 10.1016/j.cose.2022.102997.
[11] M. Ahmed, A. Naser Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *J. Netw. Comput. Appl.*, vol. 60, pp. 19-31, Jan. 2016, doi: 10.1016/j.jnca.2015.11.016.
[12] S. A. Ajila, C. H. Lung, and A. Das, "Analysis of error-based machine learning algorithms in network anomaly detection and categorization," *Ann. Telecommun.*, vol. 77, no. 4, pp. 359-370, Jun. 2022, doi: 10.1007/s12243-021-00836-0.
[13] S. Russo *et al.*, "The value of human data annotation for machine learning based anomaly detection in environmental systems," *Water Res.*, vol. 206, p. 117695, Nov. 2021, doi: 10.1016/j.watres.2021.117695.
[14] J. S. Marron, M. J. Todd, and J. Ahn, "Distance-weighted discrimination," *J. Amer. Statist. Assoc.*, vol. 102, no. Dec., pp. 1267-1271, 2007, doi: 10.2307/27639976.
[15] C. C. Aggarwal, "An introduction to outlier analysis," in *Outlier Analysis*. Cham, Switzerland: Springer, pp. 1-34, 2017, doi: 10.1007/978-3-319-47578-3_1.
[16] D. M. Hawkins, "Identification of Outliers," *Springer Netherlands*, vol. 11, pp. 1-188, 1980.
[17] M. L. Shyu, S. C. Chen, K. Sarinnapakorn, and L. Chang, "A novel anomaly detection scheme based on principal component classifier," in *Proc. IEEE Foundations New Directions Data Mining Workshop*, pp. 172-179, Nov. 2003.
[18] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, Sep. 1995, doi: 10.1023/A:1022627411411.
[19] S. Wang, J. F. Balarezo, S. Kandeepan, A. Al-Hourani, K. G. Chavez, and B. Rubinstein, "Machine learning in network anomaly detection: A survey," *IEEE Access*, vol. 9, pp. 152379-152396, 2021, doi: 10.1109/ACCESS.2021.3126834.
[20] G. Pang, C. Shen, C. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1-38, Mar. 2021, doi: 10.1145/3439950.
[21] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1-58, Jul. 2009, doi: 10.1145/1541880.1541882.
[22] J. Chen, S. Sathe, C. Aggarwal, and D. Turaga, "Outlier detection with autoencoder ensembles," in *Proc. SIAM Int. Conf. Data Mining (SDM)*, pp. 90-98, Apr. 2017, doi: 10.1137/1.9781611974973.11.
[23] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proc. 9th EAI Int. Conf. Bio-inspired Inf. Commun. Technol.*, pp. 21-26, Dec. 2016, doi: 10.4108/eai.3-12-2015.2262516.
[24] J. Tang, J. Li, Z. Gao, and J. Li, "Rethinking graph neural networks for anomaly detection," *arXiv preprint*, 2022, doi: 10.48550/arXiv.2205.15508.
[25] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," *Inf. Sci.*, vol. 177, no. 18, pp. 3799-3821, Sep. 2007, doi: 10.1016/j.ins.2007.03.025.
[26] H. Song, Z. Jiang, A. Men, and B. Yang, "A hybrid semi-supervised anomaly detection model for high-dimensional data," *Comput. Intell. Neurosci.*, vol. 2017, no. 1, p. 8501683, 2017, doi: 10.1155/2017/8501683.
[27] G. Pu, L. Wang, J. Shen, and F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," *Tsinghua Sci. Technol.*, vol. 26, no. 2, pp. 146-153, Apr. 2021, doi: 10.26599/TST.2019.9010051.
[28] Z. Ghrib, R. Jaziri, and R. Romdhane, "Hybrid approach for anomaly detection in time series data," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Glasgow, UK, pp. 1-7, Jul. 2020, doi: 10.1109/IJCNN48605.2020.9207013.
[29] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "GANomaly: Semi-supervised anomaly detection via adversarial training," in *Proc. Comput. Vision (ACCV)*, vol. 11363, pp. 622-637, 2019, doi: 10.1007/978-3-030-20893-6_39.
[30] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Efficient GAN-based anomaly detection," *arXiv preprint*, 2018, doi: 10.1109/ICDSCNC62492.2024.10939527.
[31] R. Tahri, A. Ouammou, A. Lasbahani, A. Jarrar, and Y. Balouki, "Hybrid ai framework for anomaly detection and root cause analysis in multi-agent systems," in *International Journal of Artificial Intelligence*, vol. 2252, no. 8938, p. 5291, 2024, doi: 10.11591/ijai.v16.i6.pp5290-5302.
[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint*, 2014, doi: 10.48550/arXiv.1412.6980.
[33] M. L. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "NSL-KDD: A revised benchmark dataset for intrusion de tection," in *Proceedings of the 2009 International Conference on Computational Intelligence for Security and Defense Applications (CISDA)*, Ottawa, ON, Canada, 2009, doi: 10.1109/cisda.2009.5356528.
[34] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Doll´ar, "Focal loss for dense object detection," in *IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.324.

BIOGRAPHIES OF AUTHORS

Rachid Tahri was born in Zagora, Morocco, in 1990. He received the Master's degree in Networks and Informatics Systems from the Faculty of Sciences and Techniques, Settat, in 2014. He is currently pursuing a Ph.D. at the University of Hassan 1st, Morocco. His research interests include security integration in Artificial Intelligence and Machine Learning systems. He is the corresponding author of this article. He can be contacted at: rachid.tahritr@gmail.com.

Abdellah Ouammou holds a Ph.D. in Applied Mathematics from the Computer, Networks, Mobility and Modeling Laboratory at the Faculty of Sciences and Techniques, Hassan First University of Settat, Morocco. He currently serves as a professor at the Ministry of National Education, Preschool and Sports in Morocco, and is affiliated with the Faculty of Sciences and Techniques at Hassan First University. His research interests include probability theory, stochastic optimization, discrete stochastic processes, discrete optimization, and applications in cloud computing environments. He can be contacted at: a.ouammou@uhp.ac.ma.

Abdellatif Lasbahani is a professor at University Sultan Moulay Slimane and holds a PhD from University Hassan First. His research focuses on formal methods and artificial intelligence, particularly modeling complex systems. He is committed to advancing knowledge in his field through publication and collaboration. He can be contacted at: abbdellatif.lasbahani@gmail.com.