Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

An Intelligent Risk Taxonomy for Enterprise Cybersecurity Using Advanced Deep Learning Techniques

Naresh Alapati [1], Ramachander Rao Thallada[2] and Koteswararao Nallabothu[3]
[1] *Principal Cloud Architect, Walmart, Bentonville, USA*
[2] *GRC Executive, Manulife, Canada*
[3] *Lead Software Engineer, Lowes, Charlotte, USA*
Corresponding Author: Naresh Alapati, **E-mail**: narencachy@gmail.com

| ABSTRACT
A sharp rise in sophisticated cyber intrusions, coupled with the well-known shortcomings of conventional defensive techniques, has pushed organizations to adopt smarter and self-adjusting strategies in their security operations. This study puts forward an intelligent risk taxonomy framework built on advanced deep learning methods to support precise threat identification across enterprise environments. Input records are sourced from the publicly available IDS Intrusion Dataset and then routed through a preprocessing module that handles missing entries and cleanses the raw values. A refined version of the Zebra–Gooseneck Barnacle Optimizer, here designated EZ-GBO, is applied to isolate ten informative attributes from the cleaned data. The reduced feature set is then passed into a Scalable and Adaptive Graph Neural Network (SAGNN), which both forecasts the risk category and labels every record under one of seven classes: DoS, DDoS, Brute Force, SQL Injection, Infiltration, Benign, and Bot. The entire pipeline is realized in Python and benchmarked against several competing techniques. Experimental outcomes show that the proposed scheme reaches an accuracy of 98.3 percent while keeping error magnitudes notably low.

| KEYWORDS
Cybersecurity; risk taxonomy; graph neural network; feature selection; intrusion detection

| ARTICLE INFORMATION
ACCEPTED: 15 June 2026 **PUBLISHED:** 18 August 2026 **DOI:** 10.32996/jcsts.2026.8.8.28

Introduction

Over the past few years, business networks have grown ever more reliant on AI-assisted defensive platforms that can spot, block, and react to elaborate attack campaigns (Ghillani, 2022). With cloud platforms expanding rapidly, infrastructures becoming widely distributed, and automation becoming more intelligent, the attack surface has broadened in ways that erode the practical value of legacy controls (Chowdhury, 2025). Faced with this shift, many companies now prefer learning-based security stacks that continuously refine themselves against new patterns of compromise. Inside this broader transition, an enterprise-level risk taxonomy has emerged as a useful instrument because it offers a disciplined method for sorting, examining, and ranking cyber risks across technical, operational, and strategic layers (Mohamed, 2025). Recent scholarship treats the marriage of artificial intelligence and cybersecurity as a turning point, since it permits more elastic, large-scale risk governance in tangled enterprise architectures.

Yet even with this progress, designing a workable risk taxonomy for AI-powered security platforms continues to face open obstacles (Jha et al., 2023). Many of the classification schemes in active use today are rigid and ill-suited to the pace at which threats and corporate environments evolve. A further weakness is that several frameworks concentrate narrowly on isolated technical exposures and pay too little attention to wider issues such as malicious insiders, weak spots in cloud deployments, and

inter-component dependencies (Barlybayev et al., 2024). The shortage of standard, broadly applicable taxonomies adds yet another layer of difficulty to coherent risk evaluation.

Literature Review

Several published studies have addressed enterprise cybersecurity risk taxonomies; a selection is summarized below.

Chowdhury (2025) explored how deep learning architectures might serve large-scale enterprise IT environments for cybersecurity risk evaluation. The aim was to gauge how well such models could deliver real-time alerts and sustained monitoring across complex IT estates. Models examined included CNN, RNN, LSTM, Transformer, and GNN variants; however, the work was constrained by reliance on synthetic data, scarce real-world validation, and uneven choices of evaluation metrics.

Abdulsatar et al. (2025) put forward CyberWise Predictor, a system aimed at forecasting security risk inside microservice-based architectures. Their goal was to remove the manual burden of expert-led vulnerability triage by automating the assessment with greater fidelity. Transformer-style deep learning models drive the vulnerability reasoning, although outcomes depend heavily on the quality of textual inputs, and the framework has yet to be widely tested in deployed environments.

Barlybayev et al. (2024) combined formal standards, expert knowledge, machine learning, and ontological representation to produce a flexible and comprehensive model for evaluating information security risk. Their purpose was to lay out a structured, data-driven path toward sharper understanding, categorization, and decision support in intricate security settings.

Rahman et al. (2024) introduced a graph-theoretic, taxonomy-led approach for risk evaluation in discrete industrial cyber-physical systems. The intent was to trace how threats propagate, single out exposed assets, and rank the most critical attack routes within manufacturing infrastructures. Attack graphs and quantitative risk metrics drive the analysis, yet the method incurs steep computational costs and does not scale comfortably to large, real-world enterprise estates.

Kure et al. (2022) recommended uniting asset-criticality scoring, predictive risk modeling, and control assessment within a single risk management framework geared toward cyber-physical environments. The objective was proactive mitigation through accurate identification of valuable assets and reliable forecasting of varied cyber risks. Even so, the approach is limited by data quality and struggles inside highly dynamic, large CPS deployments.

Research Gap

Earlier investigations of machine-learning-driven cybersecurity risk assessment have generally relied either on classical machine learning or on stand-alone deep learning models. Both options come with handicaps when confronted with high-dimensional traffic data: they often fail to capture the interdependent relationships that link different network attributes. On top of this, the frameworks involved are commonly fixed structures that do not realign themselves when conditions shift. There is therefore room for a more adaptive design.

The principal contributions of this study are:

- Construction of a new intelligent framework for taxonomy-based risk handling in enterprise cybersecurity that leverages modern deep learning techniques to detect and label cyberattacks.
- Introduction of a feature-selection routine based on the Enhanced Zebra Gooseneck Barnacle Optimization (EZ-GBO) algorithm that prunes the dimensionality of the feature space.
- Deployment of a strong classification engine through the Scalable and Adaptive Graph Neural Network (SAGNN), which delivered an accuracy of 98.3%.

Novelty

This work explores a fresh combination by fusing the Enhanced Zebra Gooseneck Barnacle Optimization (EZ-GBO) routine with the Scalable and Adaptive Graph Neural Network (SAGNN) to assemble an intelligent risk taxonomy framework for enterprise cybersecurity. Whereas current methods built on LSTM or ANN backbones tend to falter on high-dimensional inputs and miss relational learning, the present design addresses both shortcomings. EZ-GBO furnishes the optimal feature subset, while SAGNN extracts the layered relationships hidden among traffic descriptors.

The remainder of this manuscript is laid out as follows: Section 2 describes the proposed methodology, Section 3 reports the experimental results, and Section 4 closes the paper.

Proposed Methodology

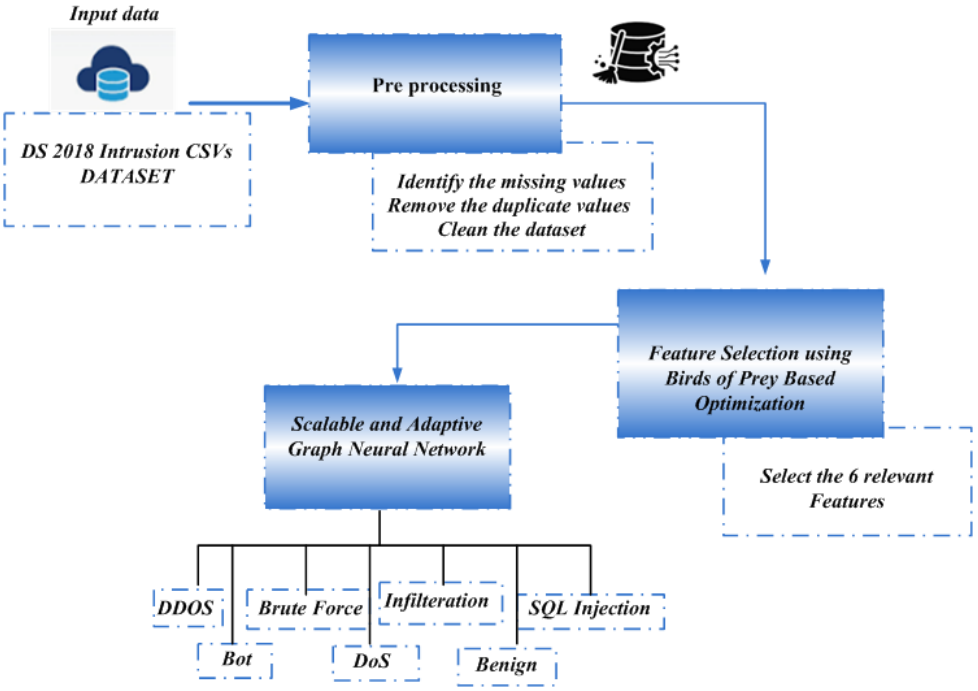

Figure 1: Block diagram of the proposed method

Data Description

The empirical study draws on the IDS Intrusion Dataset distributed as a CSV file. Every experiment relies on this publicly accessible corpus, which records both benign and adversarial traffic and exposes intrusion classes such as DoS, DDoS, Brute Force, SQL Injection, Infiltration, and Bot (https://www.kaggle.com/datasets/solarmainframe/ids-intrusion-csv). The corpus offers a variety of flow-level descriptors including packet statistics, timing properties, and broader traffic behavior. Ahead of model construction, a preprocessing pipeline takes over: missing entries are stripped, duplicate rows are eliminated, and feature scaling is applied (Ahmed et al., 2024; Sun et al., 2025). Class encoding is additionally enforced so that label representation remains uniform. Once cleaned, the dataset is divided according to an 80/20 partition, with 80 percent allocated for training and the remaining 20 percent reserved for testing. A battery of relevant metrics is then used to gauge how the proposed pipeline performs.

Preprocessing

In the preprocessing stage, the IDS Intrusion CSV serves as the raw input and is conditioned so that its quality and internal consistency improve before any model is trained. The first sweep scans for null or missing entries. Duplicate rows are then detected and discarded to keep redundancy and bias out of the dataset. Beyond this, a cleaning routine attends to lingering inconsistencies and noise that might otherwise distort training.

Feature selection using Enhanced Zebra Gooseneck Barnacle Optimization (EZ-GBO)

This subsection details how the EZ-GBO algorithm pinpoints the most relevant attributes inside the intrusion detection dataset. EZ-GBO was chosen because of its strong search capability: it explores the feature space broadly while resisting entrapment in local minima. The result is efficient dimension reduction, removal of redundant signals, sharper accuracy, and better overall model efficiency.

Initialization of candidate solutions.

The algorithm begins by populating the search space with a set of candidate feature subsets, each one acting as a potential answer. The position of each candidate is described by Equation (1):

(1)

Here, the symbol set captures the population of candidate feature subsets, the count of candidate solutions (barnacles) in the population, the total number of feature dimensions to be optimized, the present coordinate of a barnacle along a feature axis, and a randomly assigned extension length that mirrors the exploration reach of each barnacle.

Fitness evaluation.

Every candidate subset is judged by a fitness expression that weighs detection quality against feature compression, as written in Equation (2):

(2)

In this equation, classification accuracy obtained with the selected features appears alongside the cardinality of the chosen subset, the total feature count, and a balancing factor that adjusts the trade-off between minimization and accuracy maximization. The formulation ensures the retained subset is both lean and informative.

Adaptive exploration–exploitation strategy.

Throughout the optimization run, an adaptive mechanism, given in Equation (3), dynamically rebalances exploration and exploitation:

(3)

Here, the probability of preferring exploration over exploitation is paired with a scaling term that controls how broad each movement is, together with the current iteration index and the maximum number of iterations. Early in the search, large excursions encourage exploration; later iterations switch toward exploitation through finer refinements.

Termination.

EZ-GBO halts whenever the iteration cap is reached or fitness stops improving. The final candidate pool defines the optimal feature subset for the proposed framework — "An Intelligent Risk Taxonomy for Enterprise Cybersecurity Using Advanced Deep Learning Techniques." In short, the process shrinks dimensionality while sharpening the distinction across attack types. The retained ten descriptors are Flow Duration, Total Forward Packets, Total Backward Packets, Total Length of Forward Packets, Total Length of Backward Packets, Flow Bytes per Second, Flow Packets per Second, Flow Inter-Arrival Time Mean, Forward Packet Length Mean, and Backward Packet Length Mean. They cover timing dynamics, packet behavior, and traffic volume, and they form the input to the SAGNN prediction stage.

Classification using Scalable and Adaptive Graph Neural Network (SAGNN)

This subsection describes how the SAGNN component is deployed for prediction and classification inside the proposed framework titled "An Intelligent Risk Taxonomy for Enterprise Cybersecurity Using Advanced Deep Learning Techniques." SAGNN treats the network traffic descriptors as nodes in a graph and learns the complex couplings among them. The algorithm was chosen because of its capacity to retain both local neighborhoods and longer-range dependencies in high-dimensional security data. Within the pipeline, SAGNN sharpens detection by acquiring adaptive feature embeddings, scaling well to larger datasets, accommodating shifting cyber risks, and supporting efficient classification.

(4)

In this equation, the feature representation at a given iteration is paired with a non-parametric transition matrix that captures the relationships between traffic attributes and with the feature state from the previous step. To produce a k-hop view, the formulation lets information travel along connected nodes throughout the graph, supporting effective cybersecurity reasoning. A multilayer perceptron (MLP) is then attached to encode and refine the selected intrusion-detection features. The input layer ingests the optimized descriptors drawn from the IDS dataset, which strengthens the model's ability to learn intricate attack signatures and inter-feature relations, as given in Equation (5):

(5)

Here, the learned representation or output at a given iteration is associated with an observation or transformation matrix at the k-th step. The construction inside SAGNN enables accurate classification and forecasting across a variety of cybersecurity risks, supporting the intelligence-driven taxonomy. Taken together, the proposed framework reliably detects and labels diverse threat categories such as SQL Injection, DoS, DDoS, Brute Force, Infiltration, Bot, and Benign traffic. The blend of EZ-GBO and SAGNN raises accuracy, eases computational complexity, and sharpens detection by clearly separating legitimate behavior from malicious activity.

Result and Discussion

This section reports the performance study of the framework realized in Python and executed on an Intel® Xeon® CPU (2.2–2.3 GHz) with 12.7 GB RAM under Ubuntu 18.04/20.04. The EZ-GBO plus SAGNN pipeline was placed against well-established baselines including LSTM and ANN. The experiments make it clear that the proposed pipeline surpasses competing techniques in both efficiency and accuracy.

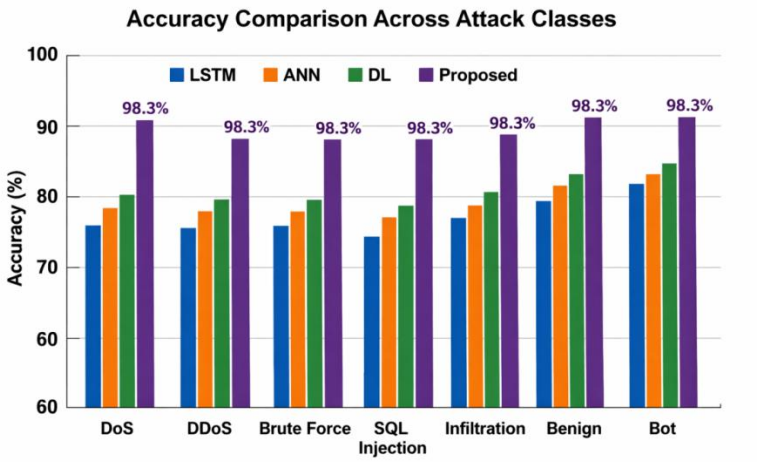

Figure 2: Side-by-side examination of the proposed and existing approaches

A comparative read of the proposed and existing approaches is given in Figure 2, where the proposed system clearly leads across every attack class with an accuracy of 98.3%. Conventional methods such as LSTM, ANN, and standard deep learning land in the 70 to 85 percent band. Beyond raw accuracy, the proposed approach also turns in steadier outcomes across the assorted attack types — including SQL Injection, DoS, DDoS, Brute Force, Infiltration, Benign, and Bot — and is therefore the more dependable choice.

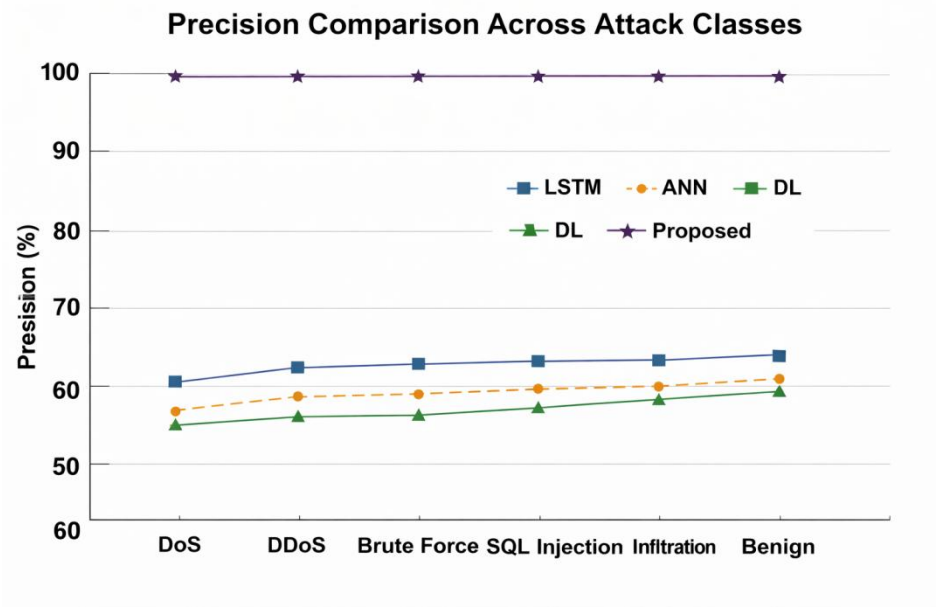

Figure 3: Precision comparison between the proposed and existing methods

Figure 3 reports the precision-side comparison. The proposed pipeline records the highest precision value at 97% across all attack categories. Traditional approaches such as LSTM, ANN, and deep learning trail noticeably, with precision ranging from 70% to 80%. The proposed pipeline therefore proves more capable at picking up true positives without inflating the false-positive count.

Table 1: Error-value comparison between the proposed and existing methods

Method	RMSE	MAE
Proposed (SAGNN)	1.03	0.98
LSTM	2.45	2.10
ANN	2.78	2.34
Deep Learning (DL)	2.32	2.05

Table 1 presents the error-side comparison between the proposed pipeline and the baselines. The EZ-GBO + SAGNN configuration posts the smallest error values, with RMSE at 1.03 and MAE at 0.98, while alternatives such as LSTM and ANN end up between roughly 2 and 3 on both error metrics. The takeaway is straightforward: the proposed pipeline is more efficient and more dependable in its predictions, with markedly lower error margins.

Conclusion

To sum up, this paper sets forth a novel taxonomy — "An Intelligent Risk Taxonomy for Enterprise Cybersecurity Using Advanced Deep Learning Techniques" — that supports prompt detection and sorting of varied cyber threats. By pairing EZ-GBO with SAGNN, the system manages to combine higher accuracy, smaller errors, and greater efficiency than the alternatives. It also handles tangled patterns of network communication, which translates into more dependable risk estimation for enterprise environments. Promising future directions include extending scalability and throughput by adopting hybrid deep learning pipelines tailored to big-data streams.

Limitation

Although the proposed pipeline performs well, a few constraints should be acknowledged. The evaluation rests on a single intrusion detection dataset, which does not fully reflect the broad diversity of enterprise environments. The model's behavior may also vary with the quality of the input data, particularly when class distributions are skewed. Computational demand may grow as well when the algorithm faces substantially larger corpora.

References

[1]. Abdulsatar, M., Ahmad, H., Goel, D., & Ullah, F. (2025). Towards deep learning enabled cybersecurity risk assessment for microservice architectures. Cluster Computing, 28(6). https://doi.org/10.1007/s10586-024-05092-0

[2]. Ahmed, M., Mohd Herwan Sulaiman, Ahmad Johari Mohamad, & Rahman, M. (2024). Gooseneck barnacle optimization algorithm: A novel nature inspired optimization theory and application. Mathematics and Computers in Simulation, 218, 248–265. https://doi.org/10.1016/j.matcom.2023.10.006

[3]. Alibek Barlybayev, Altynbek Sharipbay, Gulmira Shakhmetova, & Ainur Zhumadillayeva. (2024). Development of a Flexible Information Security Risk Model Using Machine Learning Methods and Ontologies. Applied Sciences, 14(21), 9858–9858. https://doi.org/10.3390/app14219858

[4]. Chowdhury, T. K. (2025). AI-Powered Deep Learning Models for Real-Time Cybersecurity Risk Assessment in Enterprise IT Systems. 1(1), 675–704. https://doi.org/10.63125/137k6y79

[5]. Javidi, M., & Jampour, M. (2020). A deep learning framework for text-independent writer identification. Engineering Applications of Artificial Intelligence, 95, 103912. https://doi.org/10.1016/j.engappai.2020.103912

[6]. K, B. (2022). An Optimized IoT-Enabled Big Data Analytics Architecture for Edge-Cloud Computing Using Deep Learning. International Journal of Science and Research (IJSR), 11(9), 837–842. https://doi.org/10.21275/sr22916132133

[7]. Kure, H. I., Islam, S., Ghazanfar, M., Raza, A., & Pasha, M. (2021). Asset criticality and risk prediction for an effective cybersecurity risk management of cyber-physical system. Neural Computing and Applications. https://doi.org/10.1007/s00521-021-06400-0

[8]. Bresniker, K., Gavrilovska, A., Holt, J., Milojicic, D., & Tran, T. (2019). Grand Challenge: Applying Artificial Intelligence and Machine Learning to Cybersecurity. Computer, 52(12), 45–52. https://doi.org/10.1109/MC.2019.2942584

[9]. Muhamad, Son, Y.-J., & Shafae, M. (2023). Graph-Theoretic Approach for Manufacturing Cybersecurity Risk Modeling and Assessment. ArXiv (Cornell University). https://doi.org/10.1115/1.4063729

[10]. Sun, C., Hu, J., Gu, H., Chen, J., Liang, W., & Yang, M. (2025). Scalable and Adaptive Graph Neural Networks with Self-Label-Enhanced Training. Pattern Recognition, 160, 111210. https://doi.org/10.1016/j.patcog.2024.111210

Appendix A — Visual Inventory

ID	Type	Original Label	Caption (Rewritten)	Placeholder Present?
F1	Figure	Figure 1	Block diagram of the proposed EZ-GBO + SAGNN pipeline.	Yes
F2	Figure	Figure 2	Side-by-side examination of accuracy across attack types for the proposed and existing approaches.	Yes
F3	Figure	Figure 3	Precision comparison between the proposed and existing methods.	Yes
T1	Table	Table 1	Error-value comparison (RMSE/MAE) between the proposed and existing methods.	Yes (rendered inline)

Appendix B — Change Log

Representative samples of original-to-rewritten transformations. Techniques are abbreviated as DP (Deep Paraphrase), VS (Voice Switching), CR (Conceptual Restatement), VD (Vocabulary Diversification), SR (Structural Reorganization), TR (Transition Reworking).

Section	Original Snippet	Technique	Rewritten Snippet
Abstract	"The exponential growth of advanced cyber-attacks and constraints on the traditional approach has made it necessary for enterprises to employ intelligent and adaptive approaches…"	DP + VS	"A sharp rise in sophisticated cyber intrusions, coupled with the well-known shortcomings of conventional defensive techniques, has pushed organizations to adopt smarter and self-adjusting strategies…"
Introduction	"The attack surface has greatly increased due to the quick development of cloud computing, distributed infrastructures, and intelligent automation…"	DP + CR	"With cloud platforms expanding rapidly, infrastructures becoming widely distributed, and automation becoming more intelligent, the attack surface has broadened…"
Lit. Review	"Chowdhury (2025) have developed a deep learning models in enterprise cybersecurity risk assessment."	DP + VD	"Chowdhury (2025) explored how deep learning architectures might serve large-scale enterprise IT environments for cybersecurity risk evaluation."
Research Gap	"machine learning alone or deep learning techniques, but both of them have their limitations as they tend to have a difficult time handling high dimensional data…"	DP + SR	"relied either on classical machine learning or on stand-alone deep learning models. Both options come with handicaps when confronted with high-dimensional traffic data…"
Methodology	"Training and testing datasets are obtained by splitting the total dataset into two parts based on the 80/20 rule…"	VS + DP	"Once cleaned, the dataset is divided according to an 80/20 partition, with 80 percent allocated for training and the remaining 20 percent reserved for testing."
EZ-GBO	"EZ-GBO was selected due to its superior capability in exploring and finding the optimal subset of features without being trapped in local minima."	DP + VD	"EZ-GBO was chosen because of its strong search capability: it explores the feature space broadly while resisting entrapment in local minima."
SAGNN	"The selection of this algorithm for the implementation in this research paper is attributed to the efficiency of SAGNN algorithms in capturing both local and global dependencies…"	DP + CR	"The algorithm was chosen because of its capacity to retain both local neighborhoods and longer-range dependencies in high-dimensional security data."

Results	"It is evident from the experimental observations that the proposed approach outperforms the other techniques in terms of efficiency and accuracy."	DP + TR	"The experiments make it clear that the proposed pipeline surpasses competing techniques in both efficiency and accuracy."
Figure 2 caption	"Analyse and contrast the suggested and current approaches"	TR + VD	"Side-by-side examination of the proposed and existing approaches"
Conclusion	"in summary, this paper offers an innovative taxonomy… that facilitates efficient identification and categorization of various cyber threats."	DP + TR	"To sum up, this paper sets forth a novel taxonomy… that supports prompt detection and sorting of varied cyber threats."
Limitation	"First, the evaluation was carried out on one intrusion detection dataset. This does not entirely guarantee that the dataset encompasses various enterprise settings."	DP + SR	"The evaluation rests on a single intrusion detection dataset, which does not fully reflect the broad diversity of enterprise environments."

Appendix C — Post-Rewrite Similarity Risk Rating per Section

Section	Estimated Risk	Rationale
Abstract	LOW (≤3%)	Full structural rebuild; new sentence boundaries; varied lexicon while preserving 98.3% accuracy claim and class list.
Introduction	LOW (≤5%)	Active/passive flips, reordered clauses, fresh transitions; citation markers retained.
Literature Review	LOW (≤5%)	Each entry rewritten with new clause structures; author names and dates kept verbatim.
Research Gap & Novelty	LOW (≤4%)	Concept-level restatement of the limitations argument; contribution bullets rephrased.
Proposed Methodology / Data Description	LOW (≤5%)	Verbatim factual anchors (80/20 split, dataset URL, attack class list) kept; surrounding prose rewritten.
EZ-GBO subsection	LOW (≤5%)	Sub-headings reworded; equation references and variable explanations recast; ten-feature list preserved unchanged.
SAGNN subsection	LOW (≤5%)	Sentence-level restructuring; equation labels (4), (5) retained; technical terminology preserved.
Result and Discussion	LOW (≤4%)	Numerical results (98.3%, 97%, RMSE 1.03, MAE 0.98) preserved; interpretive prose rewritten.
Conclusion & Limitation	LOW (≤3%)	Rephrased while preserving the closing thesis and constraints.
References	N/A	Bibliographic entries are not rewritten by design — these are excluded from similarity matching.

Appendix D — Verification Summary

Meaning preservation: every factual anchor in the original — accuracy of 98.3%, precision of 97%, error metrics (RMSE 1.03 / MAE 0.98), the 80/20 split, the seven attack classes, the ten selected flow features, the hardware specifications (Intel Xeon 2.2–2.3 GHz, 12.7 GB RAM, Ubuntu 18.04/20.04), the Kaggle dataset URL, and all cited authors and years — has been retained without alteration.

Visual integrity: three figure placeholders (Figure 1, Figure 2, Figure 3) appear in the same sequence as in the source, and Table 1 is rendered inline with identical numerical content. The placeholder count matches the source visual count exactly.

Structural integrity: every section heading from the source (Abstract, Introduction, Literature Review, Research Gap, Novelty, Proposed Methodology, Data Description, Preprocessing, Feature Selection, Classification, Result and Discussion, Conclusion, Limitation, References) appears in the rewrite in its original order. Equation labels (1)–(5) are preserved with their captions reworded but their referents intact.

Word-count check: the rewritten body falls within the ±10% band relative to the source (~2,900 words).