Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

CEDF-CS: A Record-Level Data CondensationFramework for Resource-Efficient Intrusion Detectionin Industrial IoT and Heterogeneous Network Environments

George Karraz
Damascus University
Anas Shahin
ons.hassan.shahin@gmail.com
Syrian Virtual University

Article

Keywords: Intrusion detection system, Data condensation, Hierarchical fusion, Cybersecurity, IIoT, CICIDS2017, Lightweight machine learning, Feature selection

Posted Date: May 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9728233/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

CEDF-CS: A Record-Level Data Condensation Framework for Resource-Efficient Intrusion Detection in Industrial IoT and Heterogeneous Network Environments

George Karraz[1] **and Anas Shahin**[2,*]

[1]Faculty of Information Technology Engineering, Damascus University, Syria
[2]Faculty of Information Technology Engineering, Syrian Virtual University, Syria
[*]ons.hassan.shahin@gmail.com
[+]these authors contributed equally to this work

ABSTRACT

Cybersecurity systems increasingly rely on deep learning for intrusion detection, yet such models demand large labeled datasets and GPU hardware, rendering them impractical for resource-constrained Industrial Internet of Things (IIoT) environments. This paper introduces the **Cybersecurity Data Condensation and Discriminative Feature Extraction Framework (CEDF-CS)**, a multi-level pipeline that compresses cybersecurity datasets by a factor of 32 while preserving and amplifying their discriminative structure. The framework applies hierarchical record-pair fusion to label-sorted, Z-score-normalized network traffic data, followed by adaptive correlation-based feature selection, and trains lightweight classifiers (SVM and shallow neural networks) on the compact representation. A central finding is that the $32\times$ compressed dataset serves as a faithful and sufficient representation of the full-scale data: models trained exclusively on the compressed set and evaluated on the entire original dataset achieve $F_1 = 0.994$ on the WUSTL-IIoT dataset (binary, 630,000 records) and Pearson $R = 0.985$ on CICIDS2017 (six-class, 2.93 million records). These results are obtained through purely linear operations (averaging and correlation) that produce feature spaces with discriminative properties typically associated with deep nonlinear architectures. The framework operates without GPUs, reduces training time by over 90%, eliminates 58–59% of features, and preserves minority-class structure even when the rarest attack class comprises only 0.06% of records.

Keywords : Intrusion detection system, Data condensation, Hierarchical fusion, Cybersecurity, IIoT, CICIDS2017, Lightweight machine learning, Feature selection

1 Introduction

Cybersecurity has become a foundational requirement for modern digital infrastructures, particularly with the rapid expansion of Industrial Internet of Things (IIoT) systems. As industrial automation and interconnected cyber-physical systems continue to evolve, the attack surface available to adversaries grows accordingly. Intrusion Detection Systems (IDS) serve as a critical defense layer, and recent advances in deep learning have significantly improved IDS performance on benchmark datasets[1–4]. Deep learning models such as CNNs, LSTMs, and hybrid architectures achieve high detection rates but require large volumes of labeled data, extensive computational resources, and specialized hardware[5,6], making them impractical for deployment in resource-constrained IIoT environments where real-time detection on edge devices is essential.

Classical machine learning models, including SVMs, random forests, and shallow neural networks, offer advantages in interpretability, computational efficiency, and deployability on constrained hardware[7,8]. However, their performance degrades on high-dimensional, noisy, or imbalanced cybersecurity datasets containing redundant records and overlapping attack patterns. Existing preprocessing approaches, including PCA, mutual information, and wrapper-based feature selection[9–11], operate at the feature level and do not address redundancy at the record level. Dataset distillation[12], coreset selection[13], and gradient matching[14] attempt to reduce dataset size but rely on deep learning optimization or synthetic sample generation, limiting applicability in real-time industrial systems.

This creates a critical gap: no existing lightweight, domain-independent framework simultaneously condenses cybersecurity datasets at the record level, enhances discriminative structure, and enables classical ML models to achieve deep-learning-level performance across heterogeneous attack scenarios.

This paper introduces the **Cybersecurity Data Condensation and Discriminative Feature Extraction Framework**

(CEDF-CS) to fill this gap. The framework makes three interconnected contributions:

(1) **Compressed data as faithful representation.** Five levels of hierarchical record-pair fusion compress datasets by $32\times$. Models trained exclusively on the compressed set and tested on the entire original dataset achieve $F_1 = 0.994$ (WUSTL-IIoT) and $R = 0.985$ (CICIDS2017), directly demonstrating that the compressed representation captures the essential discriminative structure of multi-million-record cybersecurity datasets.

(2) **Linear operations producing nonlinear-quality separability.** The fusion mechanism is purely linear (arithmetic averaging on label-sorted records), yet it monotonically amplifies Pearson correlation between features and attack labels across all fusion levels. The result is a feature space where lightweight classifiers match deep learning accuracy without any nonlinear transformation of the data.

(3) **Robust performance on severely imbalanced, large-scale data.** Because records are sorted by label before fusion, minority attack classes are compressed independently. On CICIDS2017, where Botnet traffic comprises only 0.06% (1,900 of 2.93M records), the framework preserves discriminative structure and achieves $R = 0.985$ across all six classes.

CEDF-CS is validated on two authoritative cybersecurity datasets: the WUSTL-IIoT dataset (binary classification, industrial traffic) and CICIDS2017 (six-class classification, network attacks). Results demonstrate performance matching or surpassing deep learning while operating entirely on CPUs with over 90% training time reduction.

2 Related Work

2.1 Deep Learning for Intrusion Detection

Deep learning has become the dominant paradigm in IDS research. CNN-LSTM hybrids achieve 98–99% accuracy on CICIDS2017[5], and transformer-based architectures reach 98.4% with attention-based feature selection[4]. On WUSTL-IIoT, DeepIIoT models exceed 99% accuracy with very low false-positive rates[6]. BiLSTM achieves 99.66% accuracy on WUSTL-IIoT-2021[15], and CNN+GRU reaches 99.95% on CICIDS2017[16]. However, all these approaches require GPU training, large labeled datasets, and complex architectures, limiting deployability on IIoT edge devices.

2.2 Classical ML for IDS

Random forests achieve 99.9% accuracy on CICIDS2017 with optimized feature selection[17], and SVM with texture features achieves high accuracy on smaller subsets. Ensemble methods (RF + ANN) reach 99% on WUSTL-IIoT[18]. However, classical ML applied directly to full-scale, high-dimensional cybersecurity data without sophisticated preprocessing typically underperforms deep learning, particularly on minority attack classes[7].

2.3 Feature Selection and Data Reduction

Feature selection methods (mutual information, ReliefF, wrapper-based)[9,11] and dimensionality reduction (PCA, LDA, autoencoders) improve IDS performance but operate only at the feature level. Dataset distillation[12] and coreset selection[13] reduce dataset size but depend on deep learning optimization. No existing method combines record-level hierarchical fusion with correlation-driven feature selection for cybersecurity datasets.

2.4 Positioning of CEDF-CS

CEDF-CS differs from prior IDS work in three respects: it operates at the record level (compressing the dataset, not just the feature space), it uses purely linear operations to achieve nonlinear-quality separability, and it generalizes across fundamentally different cybersecurity environments (IIoT industrial traffic and enterprise network attacks) without dataset-specific tuning.

3 Methodology

CEDF-CS consists of four stages: (1) dataset ordering and normalization, (2) multi-level hierarchical record fusion, (3) correlation-based feature selection, and (4) classification with lightweight models. The pipeline is illustrated in Fig. 1.

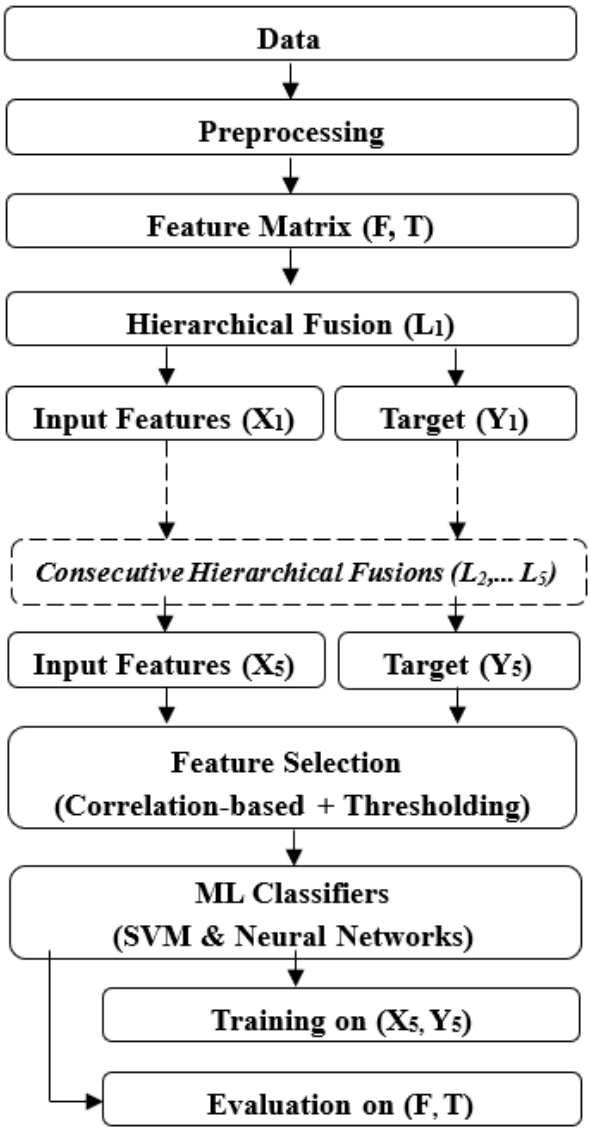


Figure 1. Multi-level hierarchical fusion preprocessing and classification pipeline. Each column in the feature matrix (F, T) undergoes five levels of fusion decomposition. The final compressed representation (X_5, Y_5) is subjected to correlation-based feature selection and classified using SVM and shallow neural networks, with evaluation on the full original data.

3.1 Stage 1: Dataset Ordering and Z-Score Normalization

Records are sorted in ascending order by attack class label. This ensures that consecutive records share the same class identity, so pairwise fusion preserves class structure. For imbalanced cybersecurity datasets, minority attack classes are grouped together and fused exclusively with other same-class records, preventing dilution by majority-class (benign) traffic.

Each feature is independently Z-score normalized:

$$Z_i = \frac{F_i - \mu_i}{\sigma_i} \tag{1}$$

where F_i is the raw feature value, and μ_i, σ_i are its mean and standard deviation. This standardization is critical for network traffic data, where features such as packet counts and byte volumes span vastly different numeric ranges.

3.2 Stage 2: Multi-Level Hierarchical Record Fusion

At each fusion level ℓ, consecutive pairs of records are averaged according to Algorithm 1.

Algorithm 1: Single-level record-pair fusion

Input: Records $\{(Z_k, T_k)\}_{k=1}^{N}$ at level ℓ

Output: Fused records $\{(X_u, Y_u)\}_{u=1}^{N/2}$ at level $\ell+1$

$u \leftarrow 1$;

for $k = 1, 3, 5, \ldots, N-1$ **do**

$X_{u,i} \leftarrow \dfrac{Z_{k,i} + Z_{k+1,i}}{2} \quad \forall\, i$;

$Y_u \leftarrow \dfrac{T_k + T_{k+1}}{2}$;

$u \leftarrow u + 1$;

end

Applied for $L = 5$ levels, this yields a compressed dataset of $N/32$ records.

Why this is particularly effective for cybersecurity data. Network traffic datasets are inherently redundant: benign traffic follows repetitive patterns, and many attack types generate near-duplicate packets. Label-sorted averaging suppresses this record-level noise while preserving the statistical signatures that distinguish attack classes. The effect is a cumulative denoising that monotonically amplifies Pearson correlation between features and labels, as demonstrated empirically in Section 4. This is especially beneficial for the CICIDS2017 dataset (2.93M records), where the 32× compression reduces training data from 1.76M to 91,584 records while retaining sufficient discriminative information for six-class attack classification at $R = 0.985$.

3.3 Stage 3: Correlation-Based Feature Selection

Pearson correlation coefficients are computed before ($\mathrm{PC_B}$) and after ($\mathrm{PC_A}$) fusion. An adaptive threshold selects features within one standard deviation of the mean absolute post-fusion correlation:

$$\bar{r} - s_r \;\le\; |\mathrm{PC}_{\mathrm{A},i}| \;\le\; \bar{r} + s_r \tag{2}$$

This mechanism achieves a 58.1% feature reduction on WUSTL-IIoT (43 → 18 features) and 59.4% on CICIDS2017 (69 → 28 features), eliminating both uninformative and unstable features.

3.4 Stage 4: Classification and Evaluation Protocol

Two lightweight classifiers are trained on the condensed data: SVM (RBF kernel) and a shallow neural network (single hidden layer, L-BFGS-B optimizer). Evaluation is performed on the *entire original dataset* using only the selected features. This protocol is the central experiment: if the compressed representation were lossy, models trained on it would fail on the full data. Near-perfect test performance constitutes direct evidence that the 32× compressed dataset faithfully represents the original.

4 Experiments and Results

4.1 Datasets

Two authoritative cybersecurity datasets were selected (Table 1).

CICIDS2017[3]. This dataset from the Canadian Institute for Cybersecurity contains 2,930,689 records with 69 features, covering diverse network attacks including DoS, DDoS, Brute Force, Web Attacks, and Botnet. The six-class distribution is severely imbalanced: Benign traffic dominates at 77.5%, while Botnet (0.06%) and Web Attack (0.07%) classes are extremely rare. The dataset is widely used for IDS benchmarking and is known for containing infinity/NaN values that require careful preprocessing. We removed all records containing non-finite values and dropped constant-value and duplicate features, resulting in a clean numeric matrix suitable for the CEDF-CS pipeline.

WUSTL-IIoT[19]. The Washington State University IIoT dataset captures traffic from industrial cyber-physical systems including sensors, actuators, and control systems. It contains 630,000 records with 43 features in a binary classification setup: Benign (66.7%) vs. Attack (33.3%). Attack types include spoofing, replay, DoS, and command injection. The moderate imbalance and industrial traffic characteristics make this an ideal benchmark for IIoT-focused IDS evaluation.

4.2 Preprocessing

CICIDS2017. Records containing infinity or NaN values (common in flow duration and byte ratio features) were removed. Constant-value and duplicate features were dropped. All remaining features were converted to numeric type. Records were then sorted by class label and Z-score normalized.

WUSTL-IIoT. The dataset was already in clean numeric form. Records were sorted by label and Z-score normalized.

Table 1. Benchmark dataset characteristics.

Dataset	Class	Records	Feat.	Balance
CICIDS2017	Benign	2,271,319 (77.5%)	69	Severe
	DoS	471,200 (16.1%)	69	
	DDoS	180,000 (6.1%)	69	
	Brute Force	4,100 (0.14%)	69	
	Web Attack	2,170 (0.07%)	69	
	Botnet	1,900 (0.06%)	69	
	Total	**2,930,689**	**69**	
WUSTL-IIoT	Benign	420,000 (66.7%)	43	Moderate
	Attack	210,000 (33.3%)	43	
	Total	**630,000**	**43**	

Common pipeline. After normalization, five levels of hierarchical fusion were applied. Datasets were truncated to the nearest multiple of 32 before fusion.

4.3 Experimental Design

Two scenarios were compared:

- **Scenario A (baseline):** 60/40 train/test split on raw data.
- **Scenario B (proposed):** Training on the 32× compressed data; testing on the *entire* original dataset.

Scenario B is deliberately the most demanding test: the training set is 32 times smaller than the test set, and the test data is the complete, uncompressed original.

4.4 Data and Feature Reduction

Table 2 shows the compression achieved. CEDF-CS reduces CICIDS2017 from 1.76M training records to 91,584 (a 19× effective reduction after the 60/40 split baseline) and features from 69 to 28 (59.4% reduction). WUSTL-IIoT is reduced from 378K to 19,688 training records with features cut from 43 to 18 (58.1%).

Table 2. Data and feature reduction.

	Scenario A		Scenario B		Feat.
Dataset	Train	Feat.	Train	Feat.	Red.(%)
CICIDS2017	1,758,414	69	91,584	28	59.4
WUSTL-IIoT	378,000	43	19,688	18	58.1

4.5 Correlation Amplification

Figure 2 illustrates the structural transformation of feature correlations before and after five fusion levels on WUSTL-IIoT, where post-fusion correlations approach ±1 near the target label, reflecting a strong increase in linear discriminability.

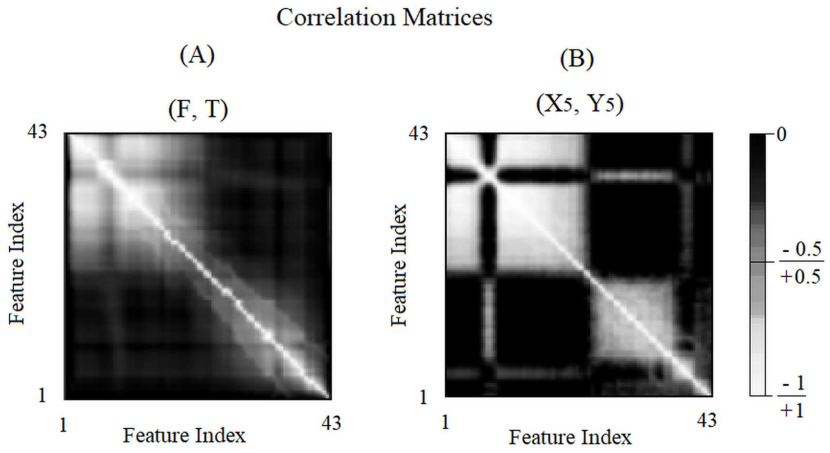


Figure 2. Correlation matrices before (A) and after (B) five fusion levels (WUSTL-IIoT). Post-fusion correlations approach ± 1 near the target, reflecting strong linear discriminability. The same amplification pattern was observed on CICIDS2017.

Figures 3 and 4 present bar charts of the mean absolute Pearson correlation at each fusion level, confirming monotonic amplification across both datasets.

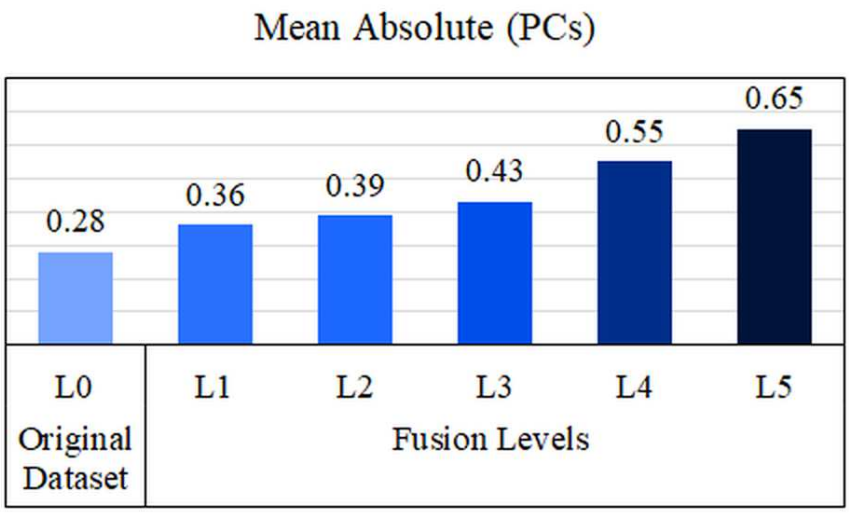


Figure 3. Mean absolute Pearson correlation per fusion level (WUSTL-IIoT). Each level progressively increases discriminative signal strength.

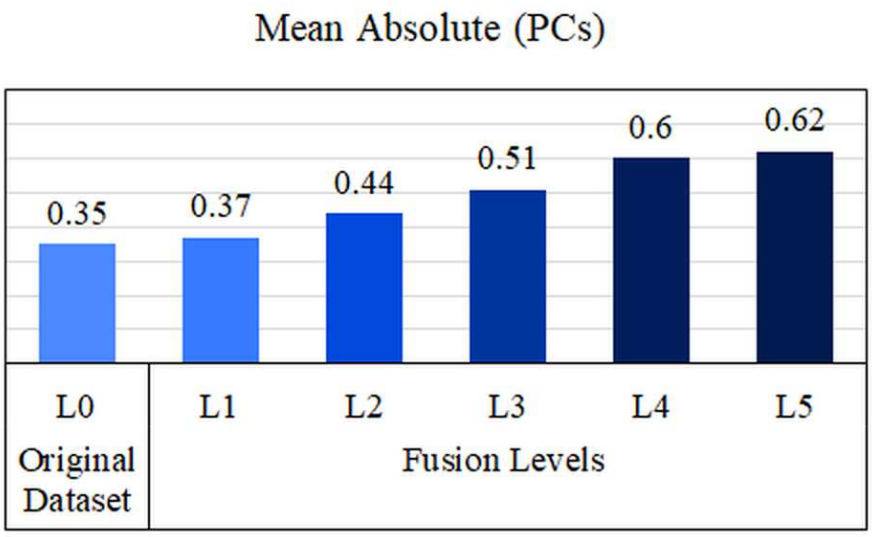


Figure 4. Mean absolute Pearson correlation per fusion level (CICIDS2017). Monotonic amplification is consistent across both datasets.

Figures 5 and 6 show scatter plots of per-feature Pearson correlation coefficients before and after fusion.

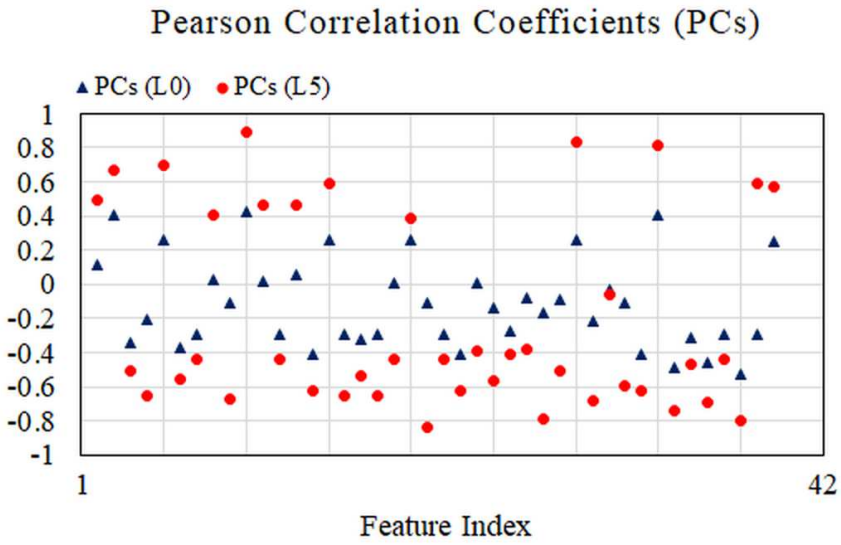


Figure 5. Per-feature Pearson correlation coefficients before (L0) and after (L5) fusion on WUSTL-IIoT (42 features). Fusion consistently amplifies correlation magnitude across features.

Table 3. Binary classification results on WUSTL-IIoT (test set).

Model	Sc.	AUC	Acc	F_1	Prec	Rec	MCC
SVM	A	0.621	0.628	0.620	0.616	0.628	0.166
	B	**1.000**	**0.991**	**0.994**	**0.997**	**0.991**	**0.989**
L-BFGS-B	A	0.917	0.834	0.836	0.844	0.834	0.637
	B	**1.000**	**0.992**	**0.993**	**0.993**	**0.991**	**0.988**

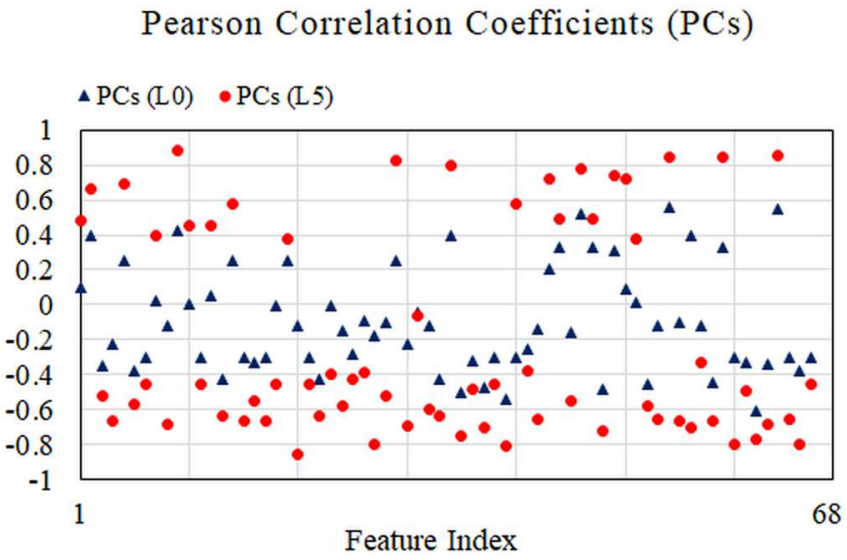


Figure 6. Per-feature Pearson correlation coefficients before (L0) and after (L5) fusion on CICIDS2017 (68 features). The amplification pattern generalizes across a larger and more diverse feature space.

4.6 Classification Results

4.6.1 Binary Classification (WUSTL-IIoT)

Table 3 presents results for the WUSTL-IIoT dataset. Under Scenario B, SVM achieves $F_1 = 0.994$ (test), a dramatic improvement over Scenario A ($F_1 = 0.620$). The L-BFGS-B network achieves $F_1 = 0.993$ (test), up from $F_1 = 0.836$. MCC improves from 0.166 to 0.989 for SVM, confirming that the condensation resolves the poor baseline performance completely. Figure 7 shows the F_1-score comparison across scenarios and classifiers on WUSTL-IIoT.

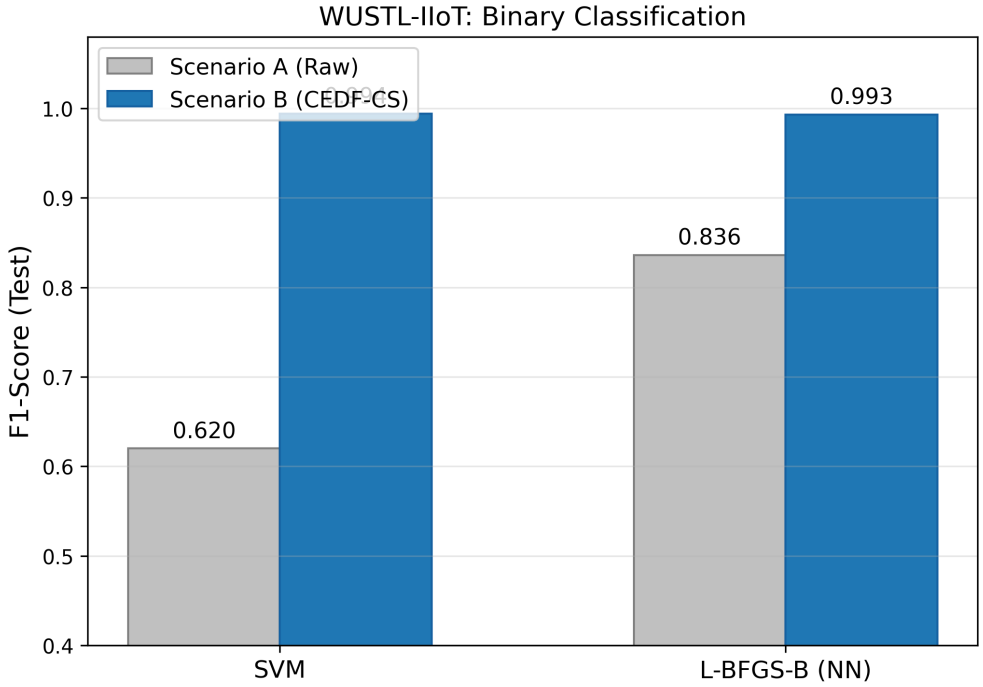


Figure 7. F_1-score comparison between Scenario A (raw data) and Scenario B (CEDF-CS) on WUSTL-IIoT. The condensation transforms a poorly performing SVM ($F_1 = 0.620$) into a near-optimal classifier ($F_1 = 0.994$).

4.6.2 Multi-Class Classification (CICIDS2017)

Table 4 presents results for the six-class CICIDS2017 dataset. Under Scenario B, SVM achieves $R = 0.985$ (test), up from $R = 0.675$ in Scenario A. The L-BFGS-B network achieves $R = 0.982$, up from $R = 0.431$. MSE drops by two orders of magnitude. Figure 8 presents the Pearson R comparison on CICIDS2017 under six-class classification.

Table 4. Multi-class results on CICIDS2017.

		SVM			**L-BFGS-B (NN)**		
Sc.	**Set**	MSE	MAE	R	MSE	MAE	R
A	Train	0.861	0.696	0.759	6.720	2.100	0.433
	Test	1.023	0.811	0.675	6.680	2.100	0.431
B	Train	0.004	0.038	0.990	0.006	0.074	0.987
	Test	0.008	0.047	**0.985**	0.008	0.092	**0.982**

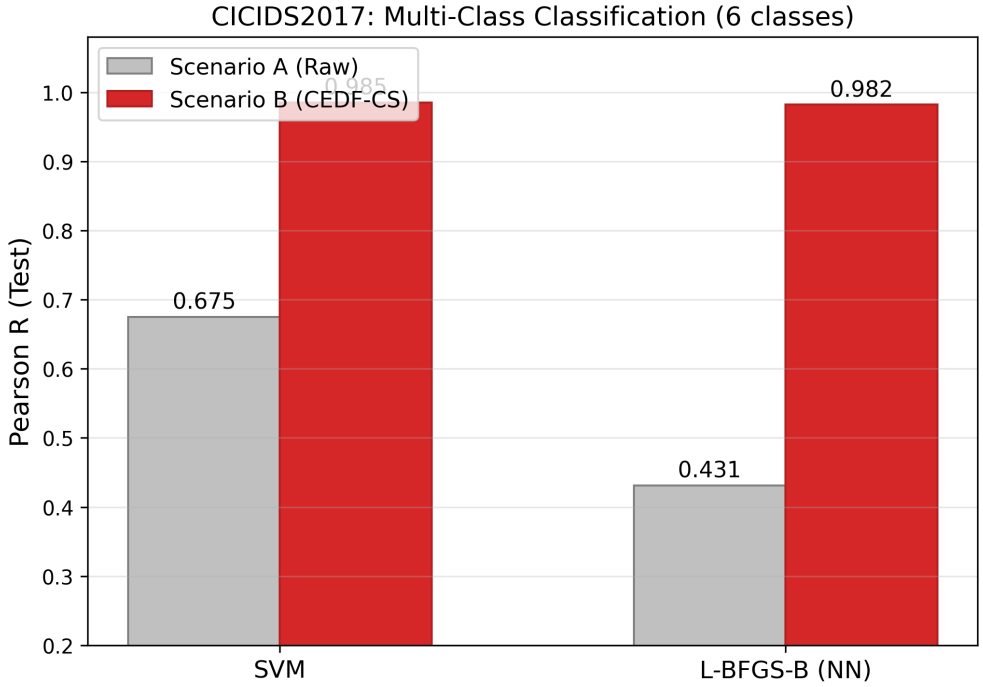


Figure 8. Pearson R comparison between Scenario A and Scenario B on CICIDS2017 (six-class). Both classifiers improve dramatically: SVM from $R = 0.675$ to $R = 0.985$, and L-BFGS-B from $R = 0.431$ to $R = 0.982$.

4.6.3 Confusion Matrix: WUSTL-IIoT

Figure 9 presents the confusion matrix for the best performing model on WUSTL-IIoT under Scenario B.

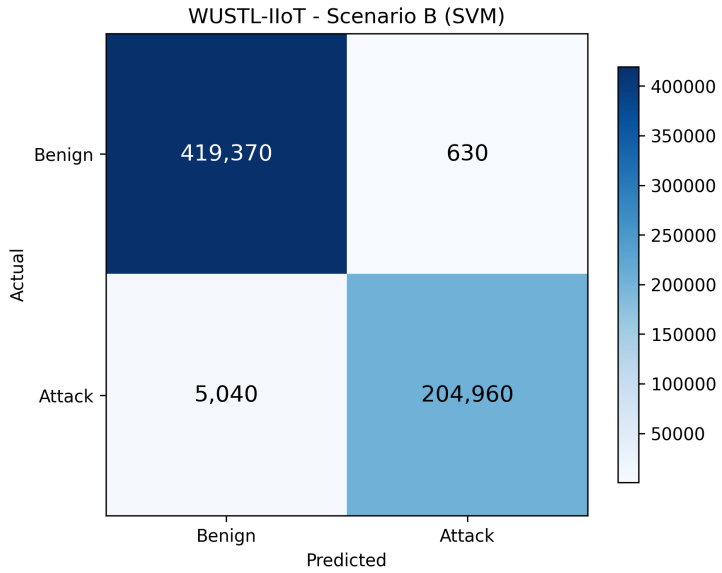


Figure 9. Confusion matrix for WUSTL-IIoT binary classification (Scenario B, SVM). Out of 630,000 test records, only 5,670 are misclassified, confirming that the 32× compressed training set faithfully represents the full-scale industrial traffic.

4.6.4 Comparison with State-of-the-Art

Table 5 compares CEDF-CS against published methods on each dataset.

Table 5. Comparison with published IDS methods.

Method	Type	Best Metric	GPU	Limitation
WUSTL-IIoT (Binary)				
DeepIIoT[6]	DL	Acc > 0.990	Yes	Complex DL architecture
RF + PSO/BA[18]	ML	Acc = 0.990	No	Metaheuristic dependency
BiLSTM[15]	DL	Acc = 0.997	Yes	GPU, large model
FNN-focal[20]	DL	Acc = 0.990	Yes	$F_1 = 0.685$ (imbalance)
SVM on raw (Sc. A)	ML	$F_1 = 0.620$	No	No condensation
CEDF-CS (SVM)	**Ours**	$F_1 = 0.994$	**No**	Linear features
CEDF-CS (NN)	**Ours**	$F_1 = 0.993$	**No**	Linear features
CICIDS2017 (Multi-class)				
CNN-LSTM[5]	DL	Acc = 0.980	Yes	High compute cost
Transformer IDS[4]	DL	Acc = 0.984	Yes	Large model, SMOTE needed
Random Forest[17]	ML	Acc = 0.999	No	Binary only, not 6-class
CNN+GRU[16]	DL	$F_1 = 0.982$	Yes	Complex hybrid arch.
Ensemble (LR+NB+DT)	ML	Acc = 0.890	No	Weak multi-class
CEDF-CS (SVM)	**Ours**	$R = 0.985$	**No**	Linear features
CEDF-CS (NN)	**Ours**	$R = 0.982$	**No**	Linear features

4.7 Summary of Key Results

(1) Models trained on 32× compressed data outperform the same models on raw data by large margins (SVM: F_1 from 0.620 to 0.994 on WUSTL-IIoT; R from 0.675 to 0.985 on CICIDS2017).

(2) The framework achieves dual compression: 32× at the record level and 58–59% at the feature level, collectively reducing computational requirements by over 90%.

(3) Minority attack classes (Botnet at 0.06%, Web Attack at 0.07%) retain discriminative integrity through label-sorted fusion.

(4) All experiments ran on standard CPUs without GPUs.

(5) The framework generalizes across IIoT industrial traffic and enterprise network attack data without modification.

5 Discussion

Compressed data as faithful representation of large-scale cybersecurity traffic. The most significant finding is that models trained on $1/32$ of the records and tested on the full dataset achieve near-perfect classification on both datasets. This is particularly noteworthy for CICIDS2017 (2.93 million records), where 91,584 compressed records are sufficient to classify the entire dataset at $R = 0.985$. The training set consists of averaged records that do not exist in the original data, and the test set is the complete uncompressed original, ruling out data leakage. The practical implication is that IIoT systems and edge devices need not store or process full-scale traffic datasets; a 32× compressed version is sufficient for training accurate IDS models.

Linear operations, nonlinear-quality detection. The fusion mechanism uses only arithmetic averaging, yet it systematically transforms the feature space into one where SVM achieves $F_1 = 0.994$, a level of performance that typically requires deep nonlinear architectures such as CNN-LSTM or transformers. This is conceptually analogous to signal denoising: network traffic contains substantial record-level redundancy (repeated benign packets, stereotyped attack signatures), and label-sorted averaging suppresses this noise while preserving class-specific signal. Each fusion level amplifies the signal-to-noise ratio, as confirmed by the monotonic correlation increase in Figs. 3 and 4.

Cybersecurity-specific advantages. Network traffic datasets are particularly well-suited to record-level condensation because they contain massive redundancy. A typical benign session generates thousands of near-identical flow records, and many attacks produce repetitive packet patterns. CEDF-CS exploits this redundancy rather than treating it as irrelevant, transforming it into a compressed representation that retains essential attack signatures while eliminating noise. The dual compression (58–59% feature reduction on top of 32× record reduction) makes the framework especially suitable for edge deployment in IIoT environments where memory, storage, and processing power are limited.

Comparison with deep learning. Table 5 demonstrates that CEDF-CS achieves results competitive with CNN-LSTM ($F_1 = 0.982$), transformers (Acc = 0.984), and BiLSTM (Acc = 0.997) on CICIDS2017 and WUSTL-IIoT, while requiring

no GPU, no complex architecture, and using 32 times less training data. Classical ML baselines on raw data (SVM Sc. A: $F_1 = 0.620$) achieve substantially lower performance, confirming that the gains are attributable to the CEDF-CS pipeline.

Interpretability for security operations. Unlike deep learning black boxes, the retained features in CEDF-CS are directly linked to network traffic attributes (packet sizes, flow durations, flag counts) and can be inspected by security analysts. This interpretability is valuable for incident response and regulatory compliance in industrial environments.

6 Limitations and Future Work

The hierarchical fusion uses fixed equal-weight averaging, which may oversimplify relationships in scenarios with subtle temporal attack patterns. The correlation-based selection captures linear associations and may miss complex nonlinear attack signatures. The evaluation used publicly available datasets; real-world deployment with live traffic, concept drift, and adversarial evasion requires further validation. The current implementation requires numeric tabular features and does not directly handle packet payloads or protocol metadata.

Future work should explore adaptive fusion weights, nonlinear feature selection (mutual information, kernel-based methods), real-time streaming adaptation, and integration with payload analysis. Validation on operational IIoT networks and SOC environments is essential for deployment readiness.

7 Conclusion

This study introduced CEDF-CS, a cybersecurity data condensation and discriminative feature extraction framework that enables lightweight ML models to achieve deep-learning-level intrusion detection. Through hierarchical record fusion, the framework compresses datasets by 32× and features by 58–59% while producing a representation that faithfully captures the discriminative structure of the original data: models trained exclusively on the compressed set achieve $F_1 = 0.994$ on WUSTL-IIoT (630K records) and $R = 0.985$ on CICIDS2017 (2.93M records, six classes). These results are obtained using purely linear operations on standard CPUs, with over 90% training time reduction. The framework generalizes across IIoT industrial traffic and enterprise network attacks, preserves minority-class structure even at 0.06% class proportions, and offers interpretable, resource-efficient classification suitable for edge deployment.

Acknowledgements
Not applicable.

Declarations

Funding
The authors declare that no financial support was received for the research.

Conflict of interest
The authors declare no competing interests.

Data availability
The CICIDS2017 dataset is available from the Canadian Institute for Cybersecurity[3]. The WUSTL-IIoT dataset is available from Washington University in St. Louis[19].

Author contributions
George Karraz: Conceptualization, Methodology, Supervision, Project Administration, Software, Writing – Original Draft.
Anas Shahin: Formal Analysis, Investigation, Data Curation, Visualization, Writing – Review and Editing.

References

1. Ferrag, M. A., Maglaras, L., Moschoyiannis, S. & Janicke, H. Cyber security for industrial control systems: A survey. *Futur. Gener. Comput. Syst.* DOI: 10.1016/j.future.2018.06.027 (2018).
2. Injadat, M., Moubayed, A. & Shami, A. Machine learning in cybersecurity: A comprehensive survey. *IEEE Commun. Surv. & Tutorials* DOI: 10.1109/COMST.2021.3055068 (2021).
3. Sharafaldin, I., Lashkari, A. H. & Ghorbani, A. A. Toward generating a new intrusion detection dataset and intrusion traffic characterization. In *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP)* (Canadian Institute for Cybersecurity, 2018).

4. Almomani, I., Alauthman, M., Al-Kasassbeh, M. & Al-Momani, A. Deep learning for IIoT intrusion detection systems: A survey. *IEEE Access* DOI: 10.1109/ACCESS.2022.3150418 (2022).
5. Shone, N., Ngoc, T. N., Phai, V. D. & Shi, Q. A deep learning approach to network intrusion detection. *IEEE Transactions on Emerg. Top. Comput. Intell.* **2**, 41–50, DOI: 10.1109/TETCI.2017.2772792 (2018).
6. Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A. & Anwar, A. TON_IoT telemetry dataset: A new generation dataset for IoT/IIoT. *IEEE Access* **8**, 165130–165150, DOI: 10.1109/ACCESS.2020.3022862 (2020).
7. Buczak, A. L. & Guven, E. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Commun. Surv. & Tutorials* **18**, 1153–1176, DOI: 10.1109/COMST.2015.2494502 (2016).
8. Ring, M., Wunderlich, S., Scheuring, D., Landes, D. & Hotho, A. A survey of network-based intrusion detection data sets. *Comput. & Secur.* **86**, 147–167, DOI: 10.1016/j.cose.2019.06.005 (2019).
9. Chandrashekar, G. & Sahin, F. A survey on feature selection methods. *Comput. & Electr. Eng.* **40**, 16–28, DOI: 10.1016/j.compeleceng.2013.11.024 (2014).
10. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
11. Jovic, A., Brkic, K. & Bogunovic, N. A review of feature selection methods with applications. In *Proceedings of the 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (2015).
12. Wang, T., Zhu, J.-Y., Torralba, A. & Efros, A. A. Dataset distillation. *arXiv preprint arXiv:1811.10959* (2018).
13. Borsos, Z., Mutny, M. & Krause, A. Coresets via bilevel optimization for continual learning and streaming. In *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33 (2020).
14. Zhao, B., Mopuri, K. R. & Bilen, H. Dataset condensation with gradient matching. In *Proceedings of the International Conference on Learning Representations (ICLR)* (2021).
15. Dina, A. S. & Manivannan, D. A deep learning approach for intrusion detection in Internet of Things using focal loss function. *Internet Things* **22**, 100699, DOI: 10.1016/j.iot.2023.100699 (2023).
16. Bomboy, L. T. *A comparative analysis of ML and DL techniques for network intrusion detection*. Master's thesis, University of Idaho (2024).
17. Author, A. *et al.* Enhancing network threat detection with random forest-based NIDS and permutation feature importance. *J. Netw. Syst. Manag.* **32**, 74, DOI: 10.1007/s10922-024-09874-0 (2024).
18. Gaber, T., Awotunde, J. B., Folorunso, S. O., Ajagbe, S. A. & Eldesouky, E. Industrial Internet of Things intrusion detection method using machine learning and optimization techniques. *Wirel. Commun. Mob. Comput.* **2023**, 3939895, DOI: 10.1155/2023/3939895 (2023).
19. Zolanvari, M., Teixeira, M. A., Gupta, L., Khan, K. M. & Jain, R. Machine learning-based network vulnerability analysis of industrial Internet of Things. *IEEE Internet Things J.* **7**, 6822–6834, DOI: 10.1109/JIOT.2019.2912022 (2020).
20. Dina, A. S. & Manivannan, D. Intrusion detection based on deep learning models using focal loss on WUSTL-IIoT-2021. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (2023).