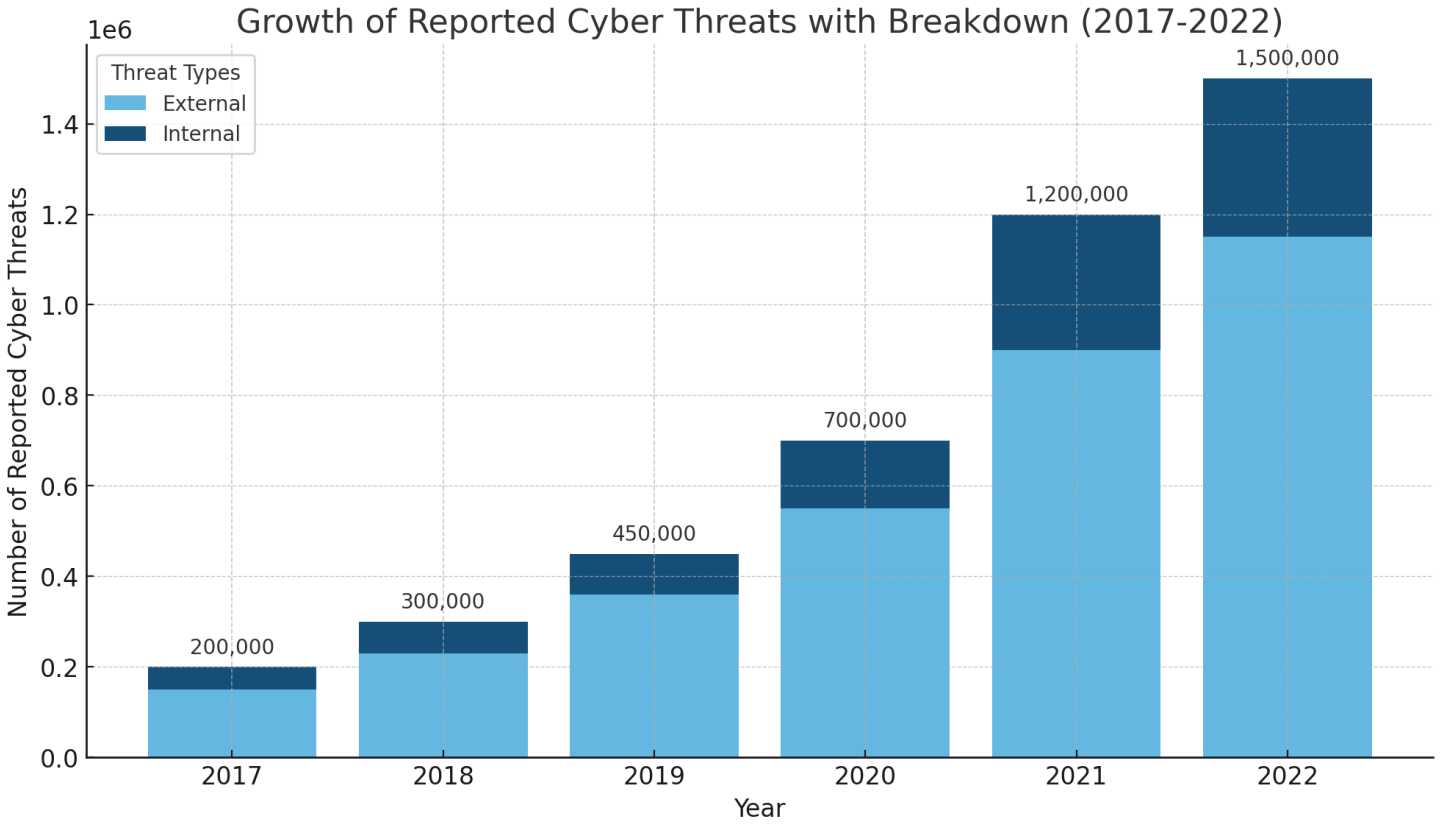
DRCN 2024

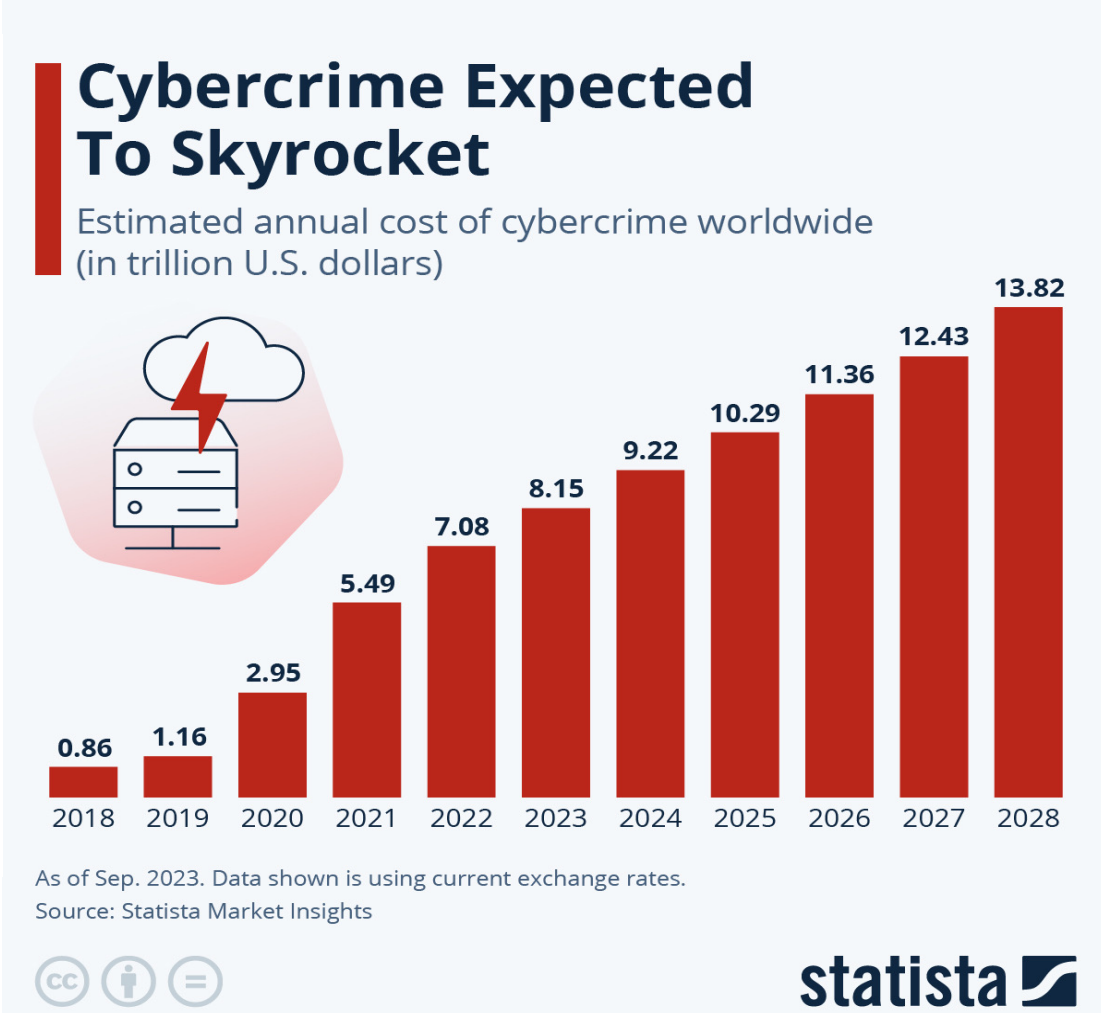
DRCN 2024

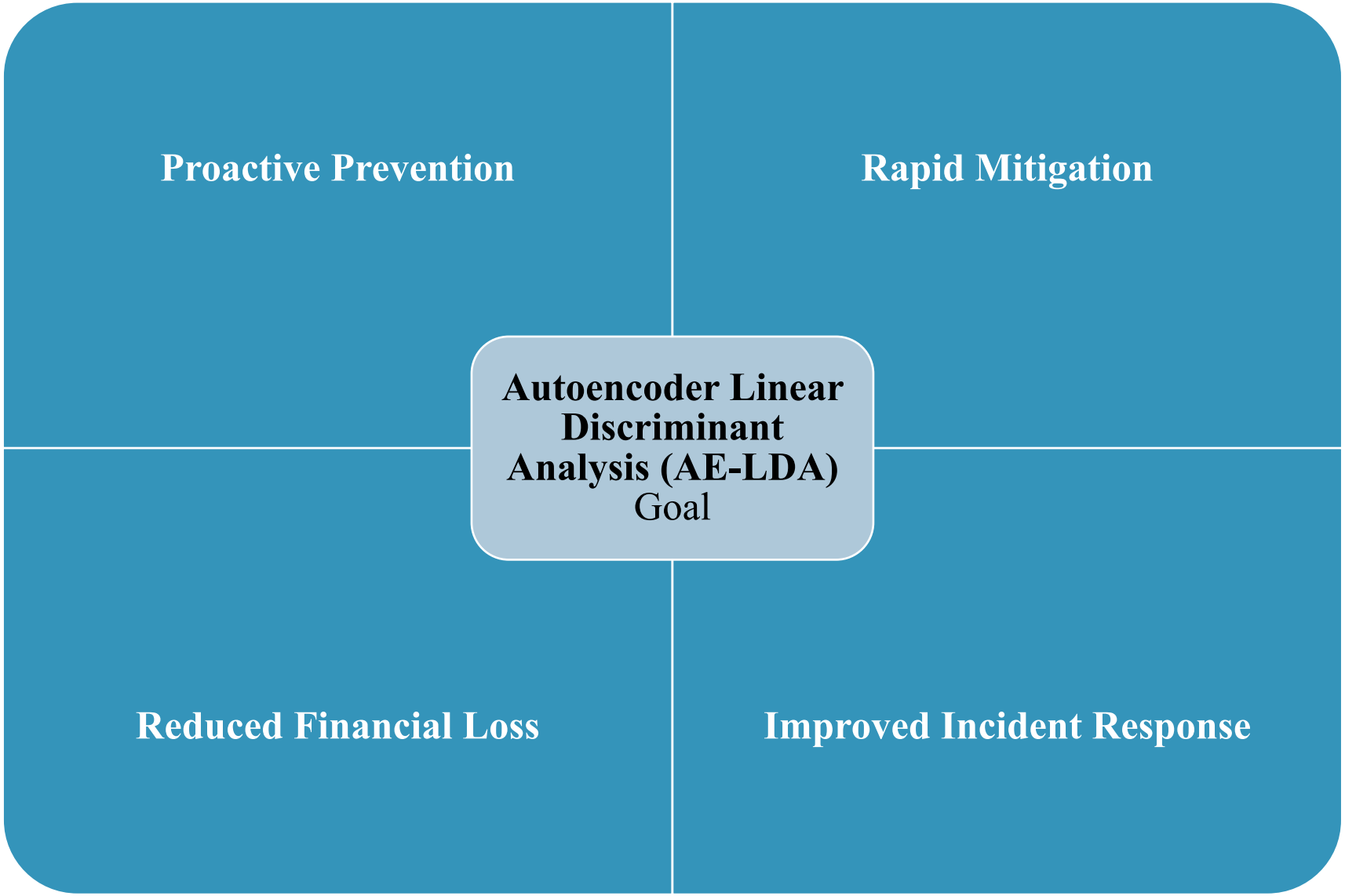
ADVANCING NETWORK SURVIVABILITY AND RELIABILITY: INTEGRATING XAI-ENHANCED AUTOENCODERS AND LDA FOR EFFECTIVE DETECTION OF UNKNOWN ATTACKS

Fatemeh Stodt[1,2], Christoph Reich[1], Fabrice Theoleyre[2]

[1] IDACUS/Furtwangen University, 78120 Furtwangen, Germany
[2] ICube/University of Strasbourg, Strasbourg, France


Growth of Reported Cyber Threats with Breakdown (2017-2022)
1e6
Threat Types
External
Internal
Number of Reported Cyber Threats
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
200,000
300,000
450,000
700,000
1,200,000
1,500,000
2017
2018
2019
2020
2021
2022
Year


Cybercrime Expected To Skyrocket
Estimated annual cost of cybercrime worldwide (in trillion U.S. dollars)
0.86
1.16
2.95
5.49
7.08
8.15
9.22
10.29
11.36
12.43
13.82
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
As of Sep. 2023. Data shown is using current exchange rates.
Source: Statista Market Insights
statista


Proactive Prevention
Rapid Mitigation
Autoencoder Linear Discriminant Analysis (AE-LDA) Goal
Reduced Financial Loss
Improved Incident Response

Problem Statement

Challenges of detecting zero-day attacks

Limitations of existing solutions

Need for an adaptive, resilient, and real-time anomaly detection system

Related Work

Knowledge-Based Techniques: Effective for known threats but often bypassed due to rigid patterns.

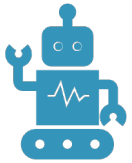
Rely on predefined rules or signatures to detect threats.

Limitations: Ineffective against new or unknown attacks due to rigidity.

Statistical-Based Techniques: Detect deviations but miss subtle and sophisticated attacks.

Detect deviations from normal network behaviour using statistical baselines.

Limitations: Can miss subtle or sophisticated attack patterns.



Machine Learning-Based Techniques: Adaptive but may suffer from generalization issues.

Use clustering or classification algorithms to detect abnormal traffic.

Limitations: May be overfit to training data or struggle with non-linear patterns.

Autoencoder-based methods are promising but require further refinement.

Deep learning models like autoencoders can learn normal behaviour and identify anomalies.

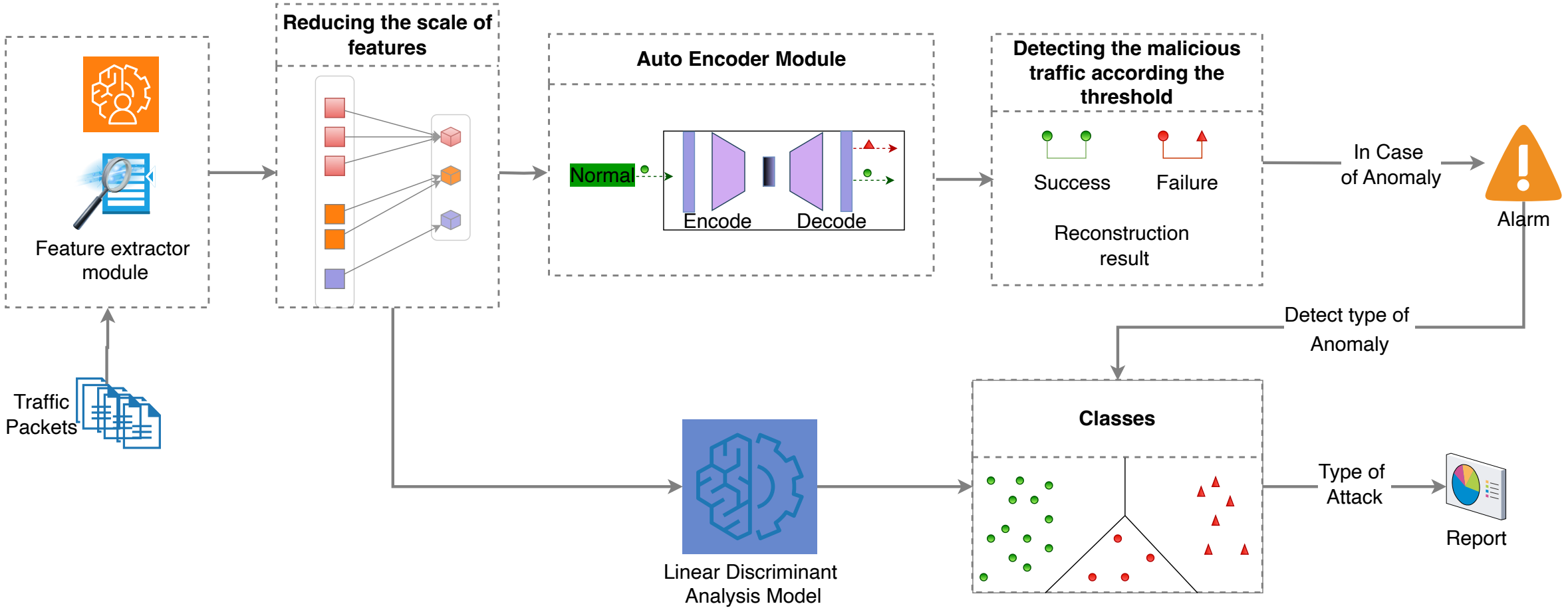
Limitations: Can generalize poorly without additional strategies like feature selection or classification.

Proposition solution:

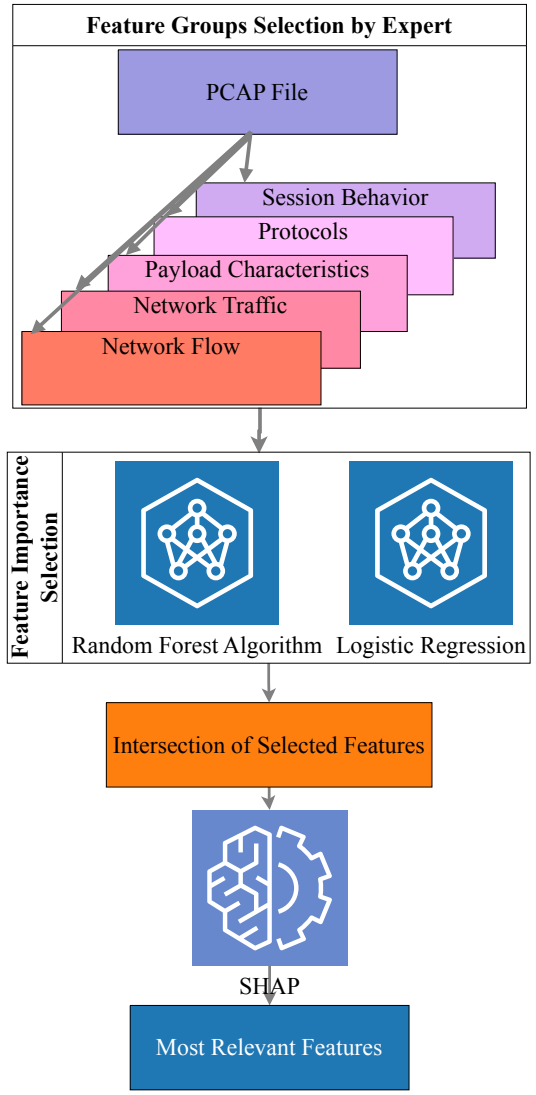
(AE-LDA)

- Explainable AI-enhanced feature selection,
- autoencoder-based anomaly detection, and
- Linear Discriminant Analysis (LDA) for more comprehensive classification.

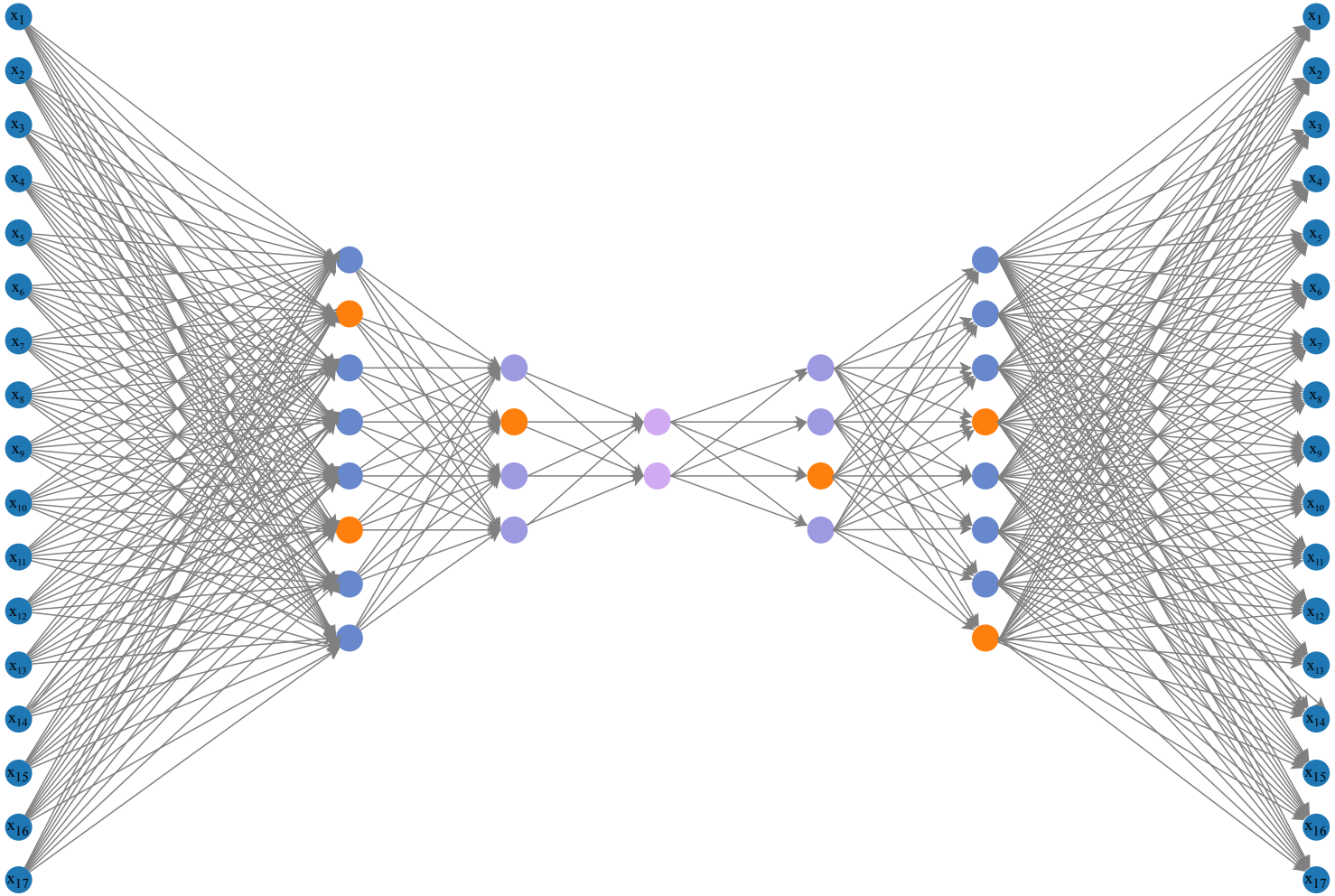
Workflow of Anomaly Detection



Feature Selection:



Autoencoder Structure:



Linear Discriminant Analysis (LDA):

Definition: LDA is a dimensionality reduction technique that projects data onto a lower-dimensional space, maximizing class separability.

Goal: Find a linear combination of features that best separates two or more classes.

How It Works:

- Feature Separation: Projects data to a new axis where inter-class differences are maximized, and intra-class variance is minimized.
- Decision Rule: Uses a probabilistic decision rule to classify new observations based on the linear combination of features.

Application in Our Paper:

- Anomaly Classification: In our AE-LDA approach, LDA is used to classify detected anomalies into known attack types (e.g., DoS Hulk, ARP MitM).
- Training: Trained on labeled network traffic to differentiate between benign and malicious patterns.
- Integration: Complements autoencoder-based anomaly detection by providing a detailed classification of recognized attacks.

Key Advantages:

- Interpretability: Provides clear and interpretable decision boundaries.
- Efficiency: Works well with high-dimensional data while reducing computational complexity.

Dataset and evaluation methodology

Datasets Overview:

- **CICIDS2017**: Provides a comprehensive set of network traffic data, including benign and multiple attack types (e.g., DoS Hulk, DoS Slowloris). Offers a mix of known and unknown threats for a well-rounded evaluation.
- **Kitsune Dataset:** Specializes in IoT network traffic and contains various attack scenarios like ARP MitM, SSDP Flood, and Active Wiretap. Used to benchmark zero-day attack detection capabilities.

Data Splitting Strategy:

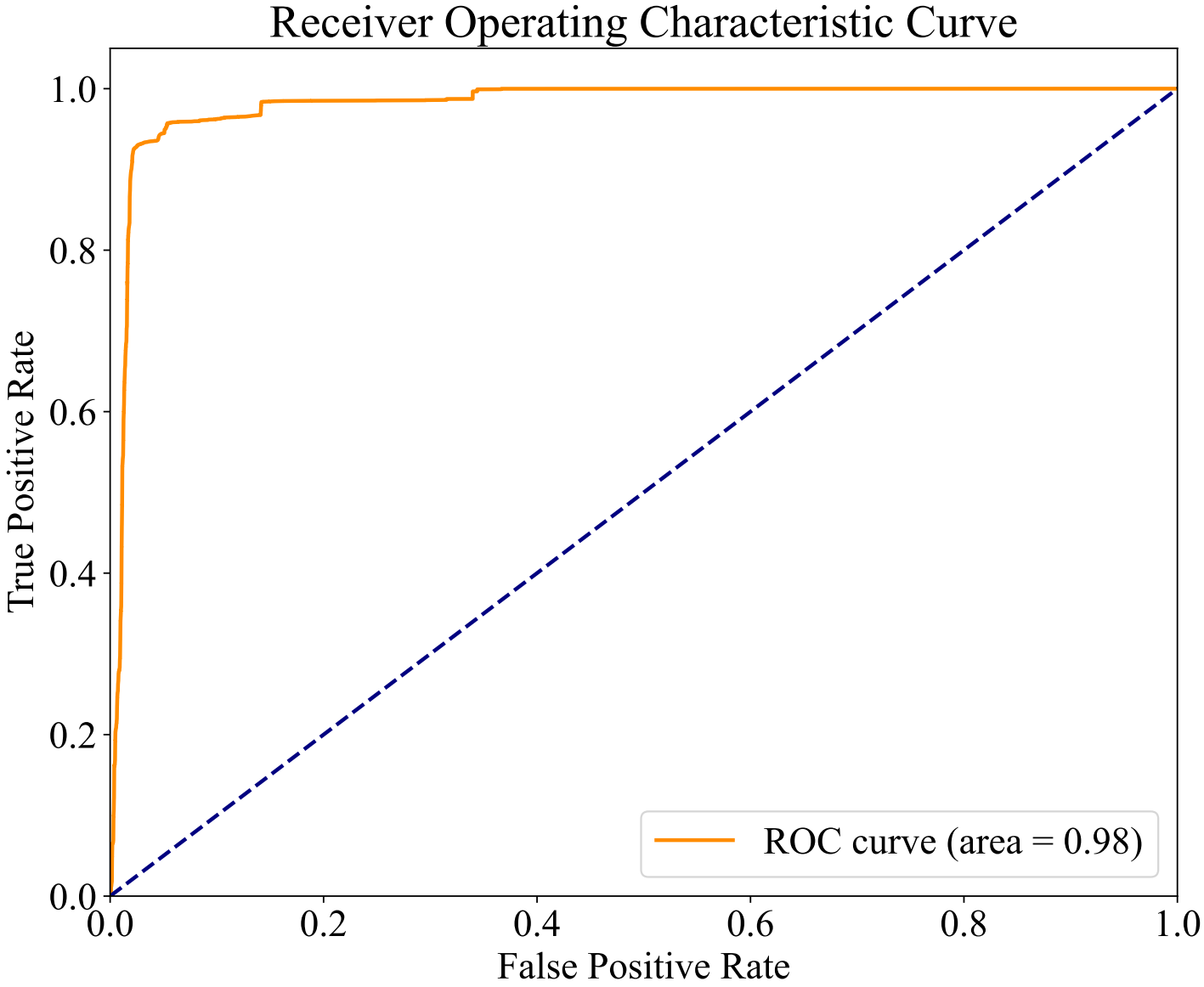
- **Training Set:** For training the autoencoder, only benign data was used to accurately characterize normal traffic patterns.
- **Test Set:** Contains a balanced mix of benign and malicious traffic, including attacks, allowing a thorough evaluation of both known and unknown threats.

Approach Highlights:

- Autoencoder Ensemble: Trained on benign data to recognize standard traffic patterns.
- LDA Module: Separately trained on labeled traffic to distinguish different types of anomalies.

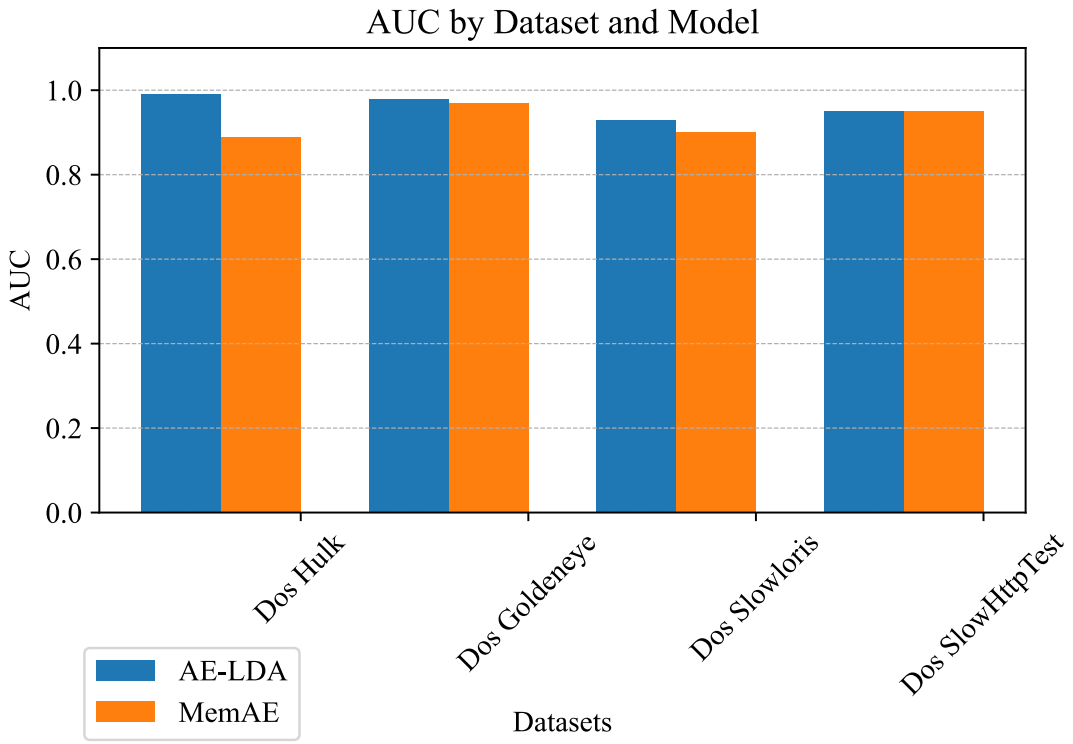
Experimental Evaluation

- Datasets: CICIDS2017 and Kitsune
- Metrics: Accuracy, AUROC, Detection Time
- Setup: Trained the autoencoder on benign traffic and used LDA for anomaly classification.



Results & Discussion

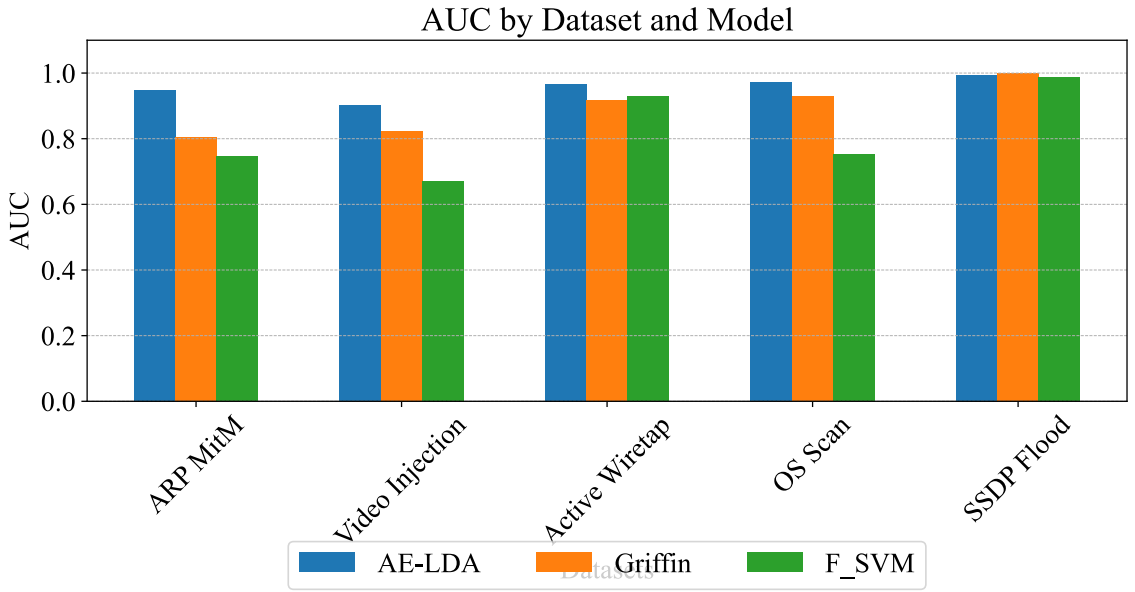
CICIDS2017 Results: AE-LDA achieved higher AUROC compared to other models like OCSVM and MemAE.



Model	AUROC
OCSVM [22]	0.7684
AE [23]	0.8758
MemAE [17]	0.9101
AE-LDA	0.98

Results & Discussion

Kitsune Results: AE-LDA excels at detecting ARP MitM, Video Injection, and Active Wiretap attacks.



Method	AE-LDA	Griffin	pcStream2	F_SVM	F_RF
ARP MitM	0.9487	0.8048	0.7219	0.7452	0.6512
Video Injection	0.9007	0.8237	0.5816	0.6718	0.6139
Active Wiretap	0.9669	0.9188	0.7413	0.9281	0.7634
OS Scan	0.9713	0.9281	0.7513	0.7517	0.7212
SSDP Flood	0.9945	0.9999	0.9971	0.9876	0.8674

Conclusion

Summary: AE-LDA significantly improves network anomaly detection by combining explainable AI with robust autoencoders and LDA.

Future Work: Refining the model to differentiate between faults and malicious anomalies, and predicting attacks based on network behavior patterns.

Questions and Discussion

We would like to express our gratitude to the funding organization German Federal Ministry of Education and Research with their funding program ”Forschung an Fachhochschulen” contract number COSMIC-X 02J21D144 for their support in accordance with the regulations associated with this project.

Thank you!

For additional information please contact us:

Fatemeh Stodt
fatemeh.stodt@etu.unistra.fr