Addressing Malware Family Concept Drift with Triplet Autoencoder

Numan Halit Guldemir · Oluwafemi Olukoya · Jesús Martínez-del-Rincón

The Centre for Secure Information Technologies (CSIT) · Queen's University Belfast

About the Presenter

Numan Halit Guldemir

PhD student at Queen's University Belfast

numanhg

nguldemir01@qub.ac.uk

- *Numan Halit Guldemir is supported by the Republic of Türkiye Ministry of National Education (MoNE-1416/YLSY).*

INTRODUCTION

Machine Learning in Cybersecurity

- Machine learning has become a key tool in cybersecurity. These systems, when **well-trained**, are **highly effective** at identifying threats.

- **However,** ML models quickly become **outdated** as thousands of **new malware** are created daily, and existing ones **evolve**.

- This constant change, known as **concept drift**, significantly impacts model performance over time.

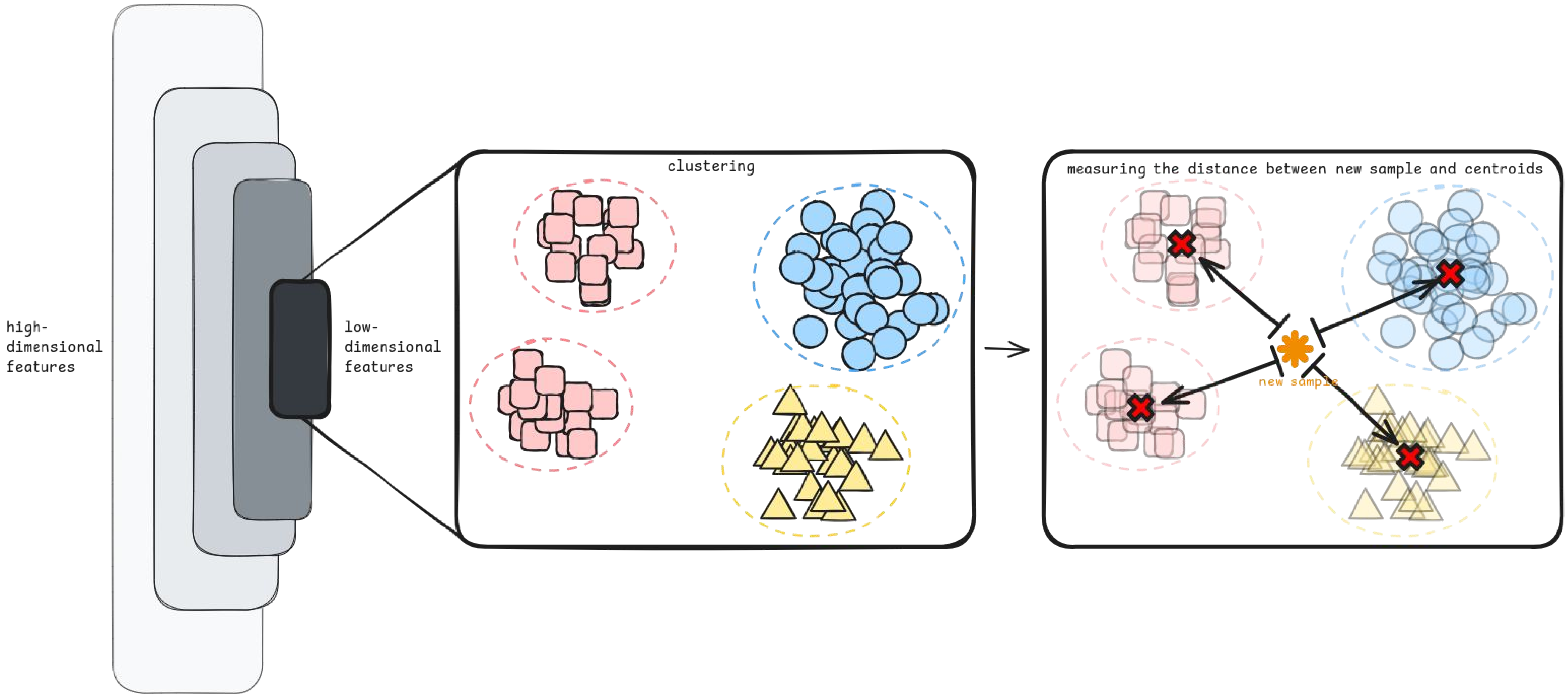
INTRODUCTION

Concept Drift in Malware Analysis

- There are 2 types of concept drift in the field of malware classification.
    - **Emerging malware families:** New, previously unseen malware families.
    - **Evolving malware families:** Variants within existing families.
- In this work, we focus on the first one, and develop an adaptive model that differentiates between known and new malware families.

METHOD

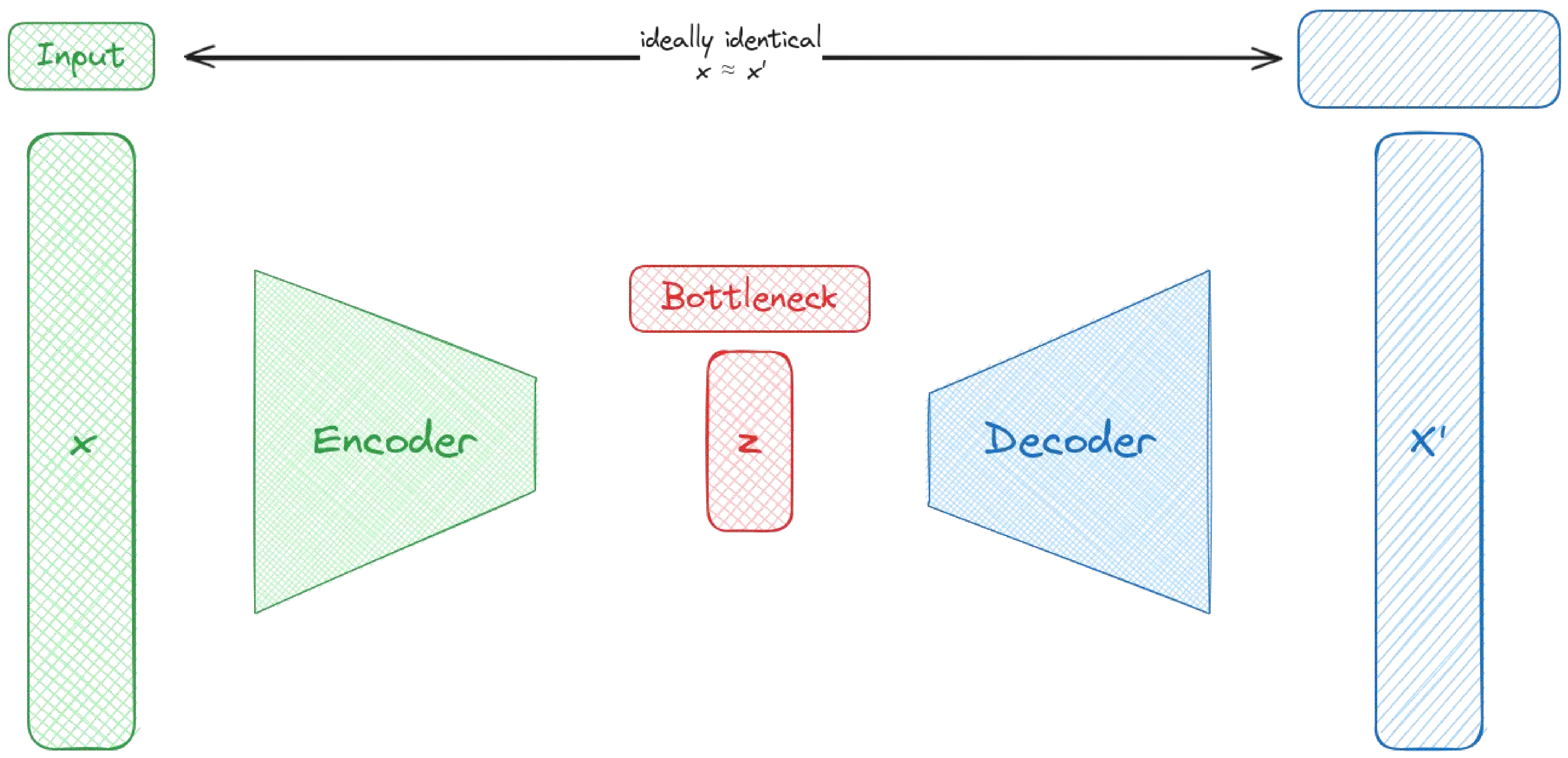
Overview



METHOD

Autoencoder

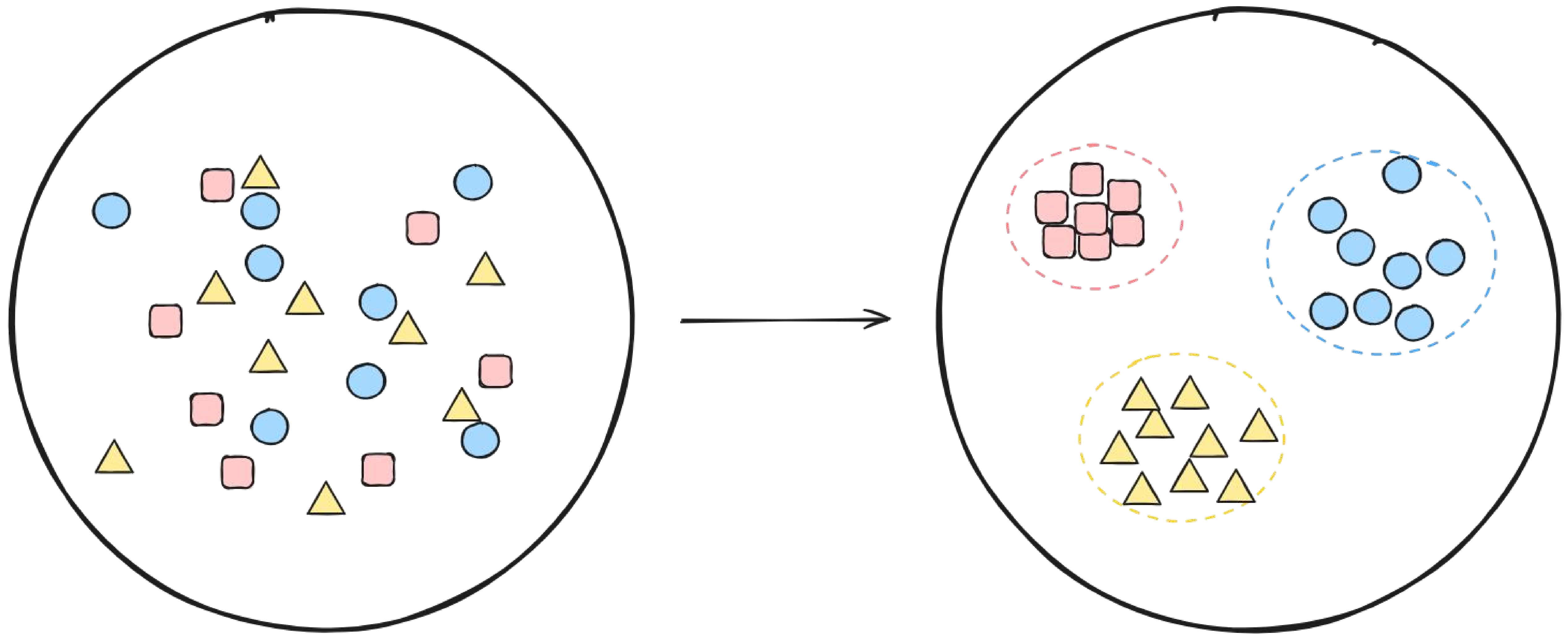
- High-dimensional data makes distances hard to define; known as the **curse of dimensionality.**
- As dimensionality increases, distances between data points become **less informative**, making it **difficult to distinguish clusters**.
- Autoencoders address this by reducing data to a **lower-dimensional space** while retaining essential features.



METHOD

Metric Learning

- Metric learning focuses on **defining a distance metric** to distinguish data points.
- It enables the **grouping of similar** samples while clearly **separating distinct** classes.



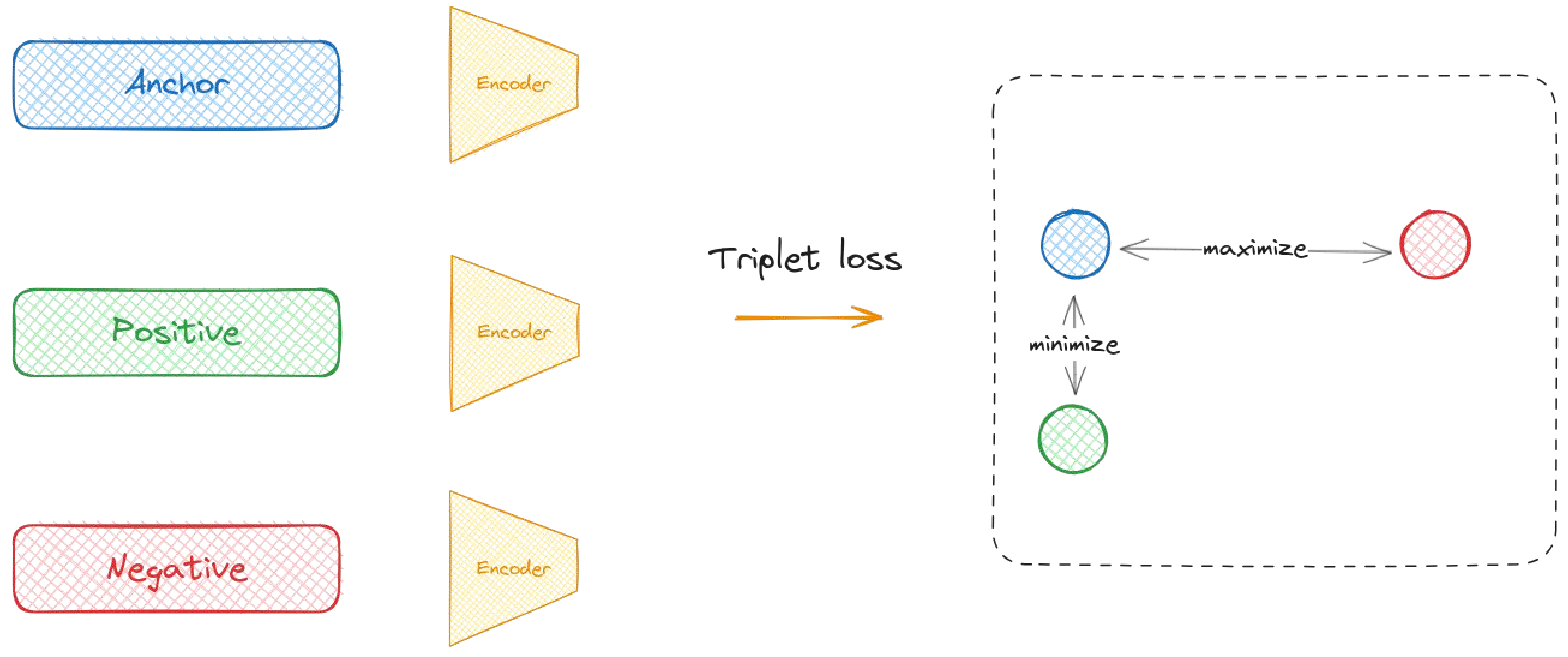
METHOD

Triplet Loss

- Optimizes the distance between data points to create distinct clusters.

- Uses three points (triplets) for training:
  - **Anchor**: A sample from a known malware family.
  - **Positive**: Another sample from the same family.
  - **Negative**: A sample from a different family.

- Ensures that the **anchor-positive** distance is **smaller** than the **anchor-negative** distance by a specified margin.

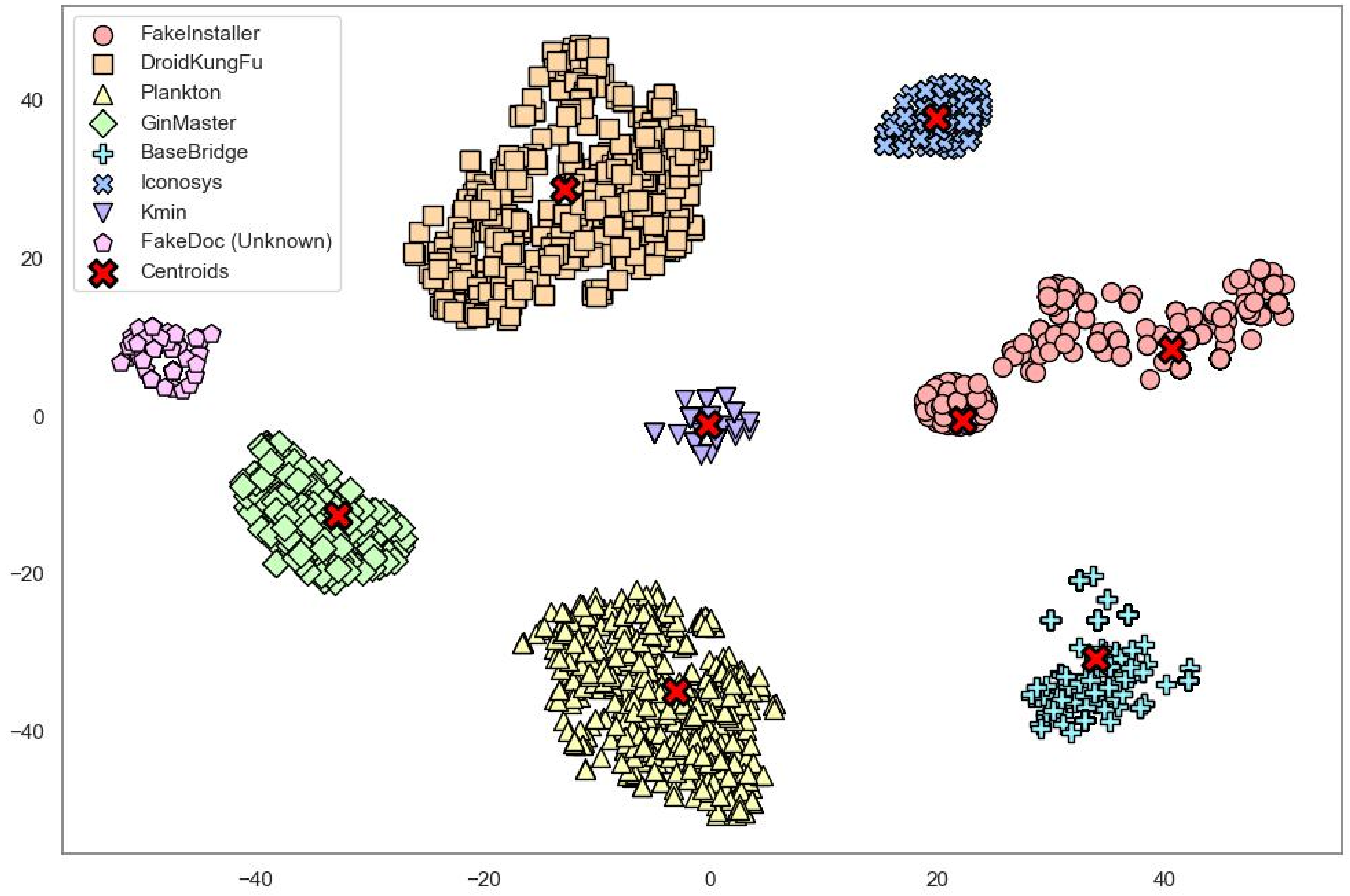


METHOD

DBSCAN

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a clustering algorithm that groups data points based on their density.

- Why use DBSCAN?
  - **To detect sub-clusters:** Identifies distinct clusters within a single malware family, allowing differentiation between multiple variants.
  - **To detect outliers**: Flags low-density points as outliers, capturing mislabelled samples or anomalies.



Datasets

Drebin
3,317 malware – 8 families

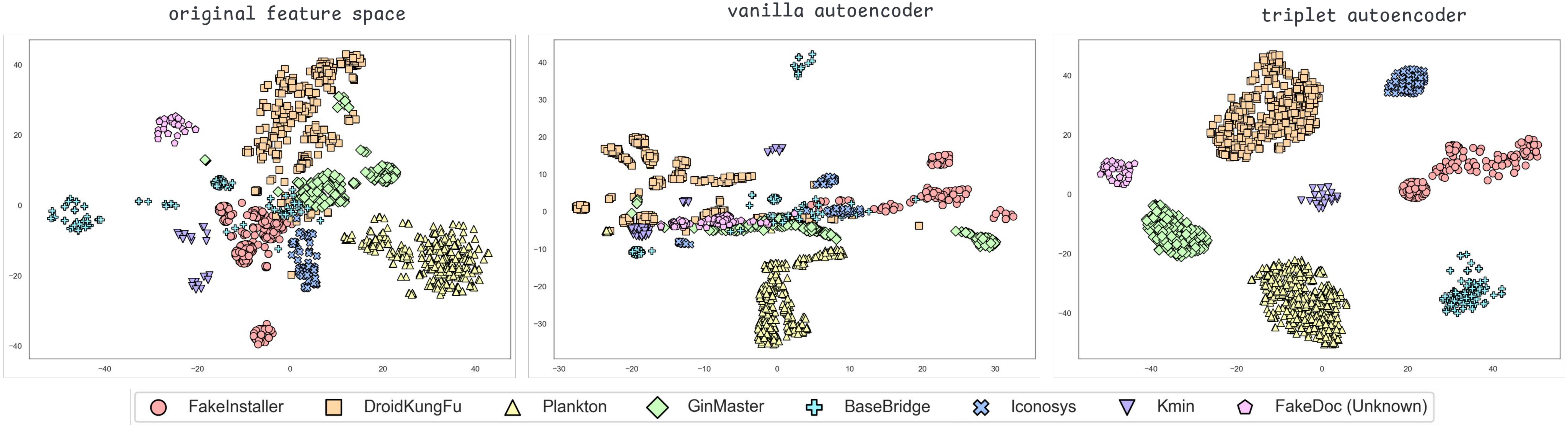
family	sample size
FakeInstaller	925
DroidKungFu	667
Plankton	625
GingerMaster	339
BaseBridge	330
Iconosys	152
Kmin	147
FakeDoc	132

BODMAS
16,458 malware – 7 families

family	sample size
berbew	1741
dinwod	1942
ganelp	1413
mira	1526
sfone	3218
sillyp2p	3012
small	3606

EVALUATION

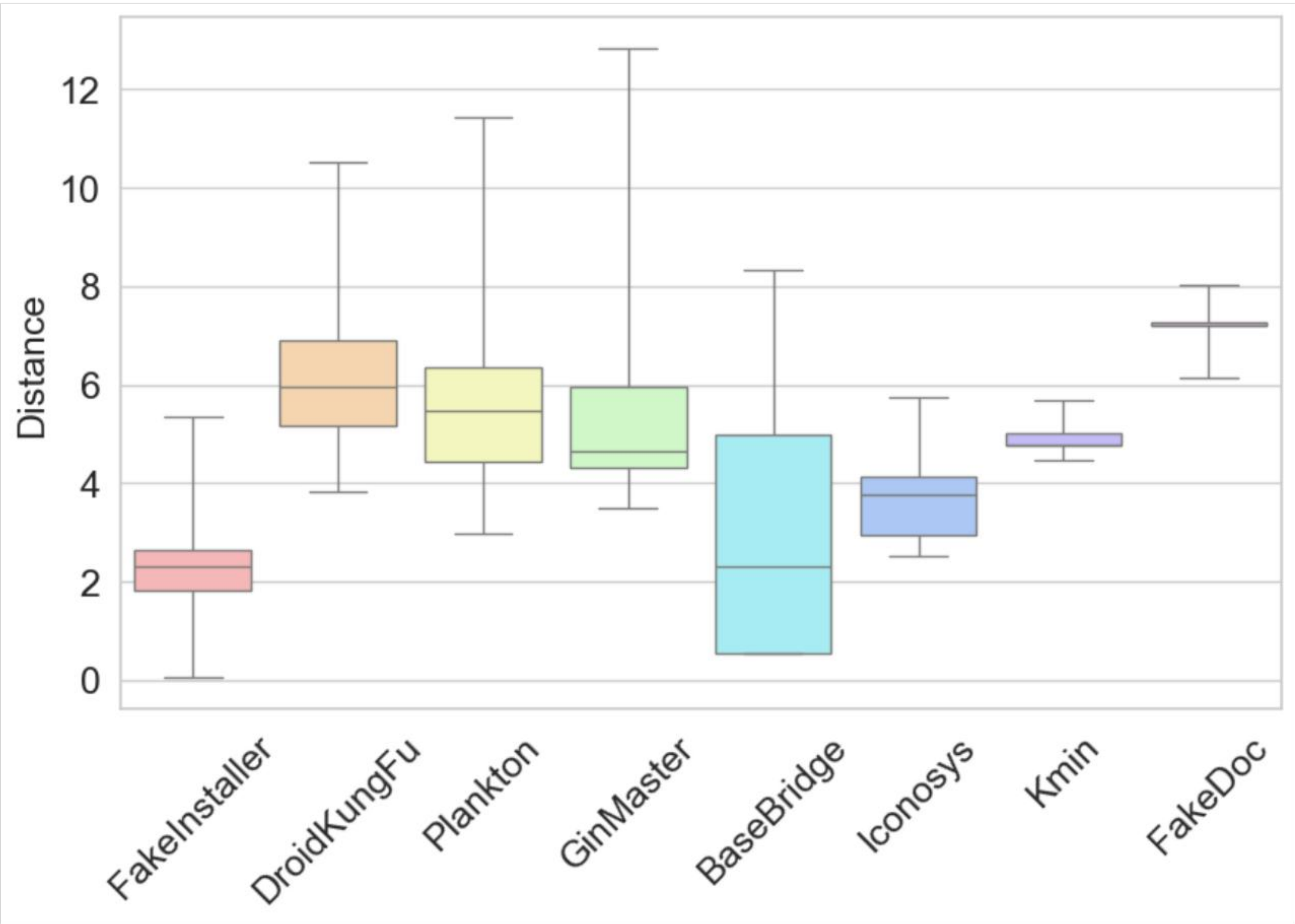
t-SNE Plot of Feature Spaces



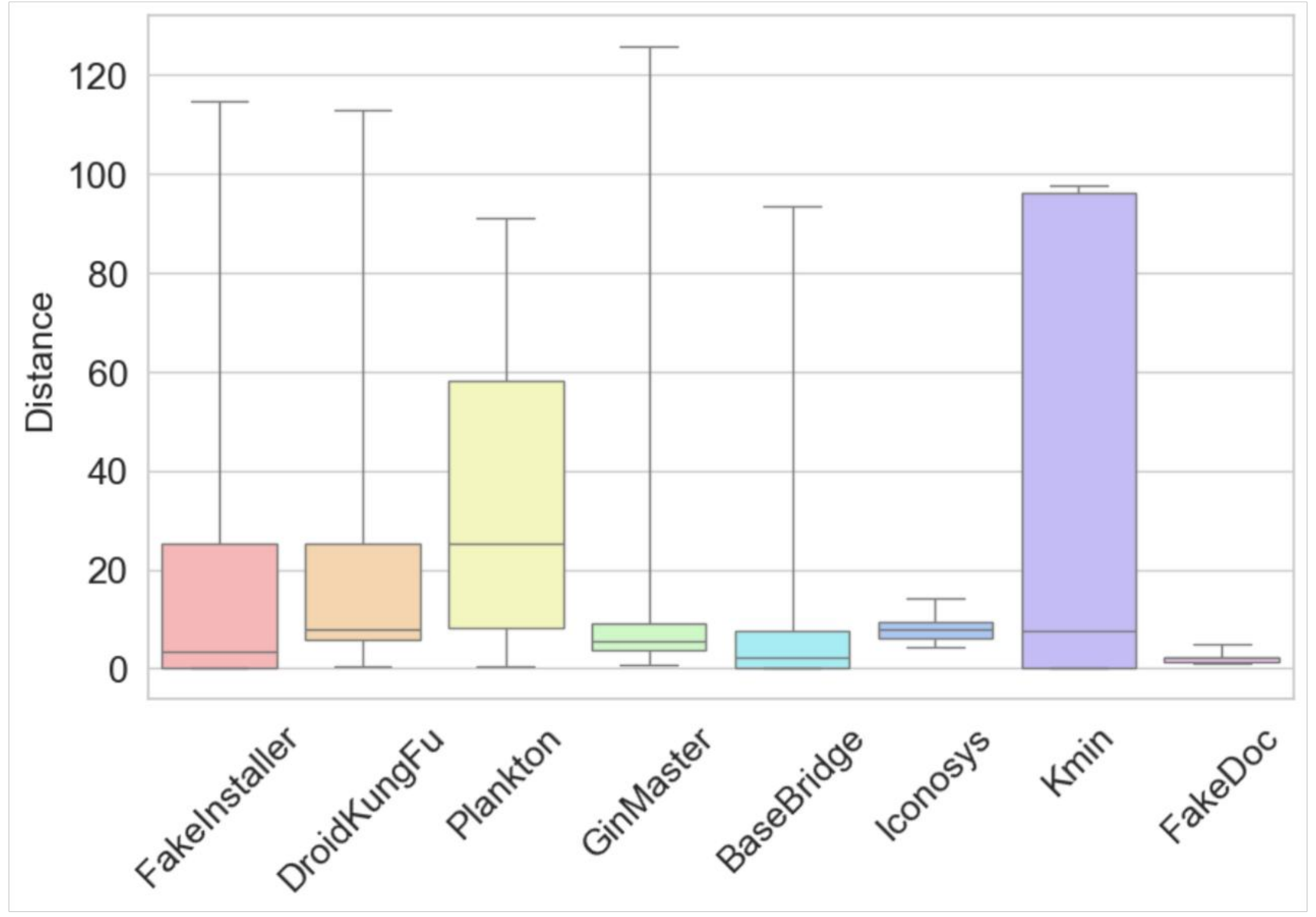
EVALUATION

Distances From Centroids

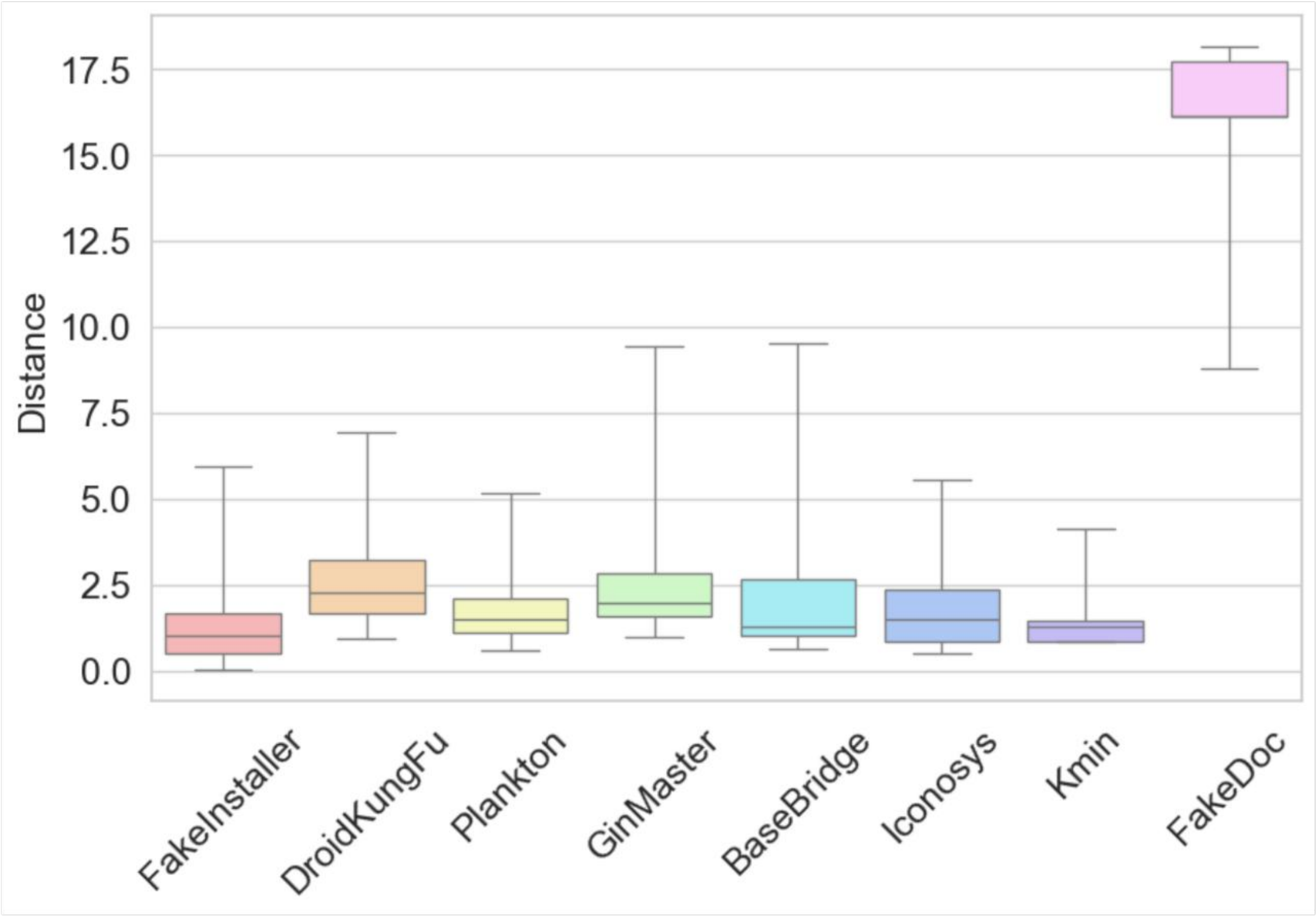
original feature space



vanilla autoencoder



triplet autoencoder



EVALUATION

Results

Drebin

family selected as unknown	f1 score
FakeInstaller	0.95
DroidKungFu	0.90
Plankton	0.87
GingerMaster	0.85
BaseBridge	0.98
Iconosys	0.65
Kmin	0.62
FakeDoc	0.66

BODMAS

family selected as unknown	f1 score
berbew	0.99
dinwod	0.96
ganelp	0.97
mira	0.58
sfone	0.51
sillyp2p	0.83
small	0.96

Takeaways

- The **triplet autoencoder** combined with **DBSCAN** clustering significantly improves accuracy in identifying **previously unseen malware families.**
- DBSCAN improves clustering quality by effectively handling **outliers** and ensuring clear separation of **different malware variants**.
- The method demonstrates strong generalizability across diverse datasets (Android and Windows PE), confirming its effectiveness in **various environments.**

Thank you!