Addressing Data Security in IoT: Minimum Sample Size and Denoising Diffusion Models for Improved Malware Detection

Chiara Camerota
Lorenzo Pappone
Flavio Esposito
Tommaso Pecorella

University of Florence
St. Louis University

INTRODUCTION

Back to Agenda Page

Internet of Things (IoT)

- IoT connects billions of devices through a network
- Devices are limited in resources and often lack robust security measures
- Machine Learning (ML) helps to address these limitations

Malware detection (MD)

- MD identifies dangerous software on devices or networks
- ML techniques improve the task in terms of accuracy
- Several ML methodologies can be customized for specific study cases

Back to Agenda Page

MOTIVATIONS

Back to Agenda Page

Motivations

IoT malware detection often **lacks large, diverse datasets**

Collecting data over the **shortest period** and then classifying them over a **longer timeframe**

Motivations

Data augmentation generates **synthetic images** to increase the train set size

Existing models (e.g., GAN) struggle with accuracy, **often misclassifying benign data as threats**

Back to Agenda Page

A

CONTRIBUTIONS

Back to Agenda Page

Introducing diffusion model as generative model for traffic based images

New method for sample train size definition based on the confusion matrix without assuming any distribution

Accurate results in terms of model accuracy and false/true positive rates

Back to Agenda Page

METHODOLOGY

Back to Agenda Page

Minimum Sample Size

Minimum number of training data points needed to achieve a **desired level** of accuracy

No distribution on the index and the data are made, i.e. the **performed test is non-parametric**

Based on the **confusion matrix** and **F1-score**

F1-score formulation

F1-score in terms of **true positive** (TP), **false negative** (FN) and **false positive** (FP) values is as follows:

$$F_1 \geq \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + 1 \cdot \mathrm{FN} + \mathrm{FP}}$$

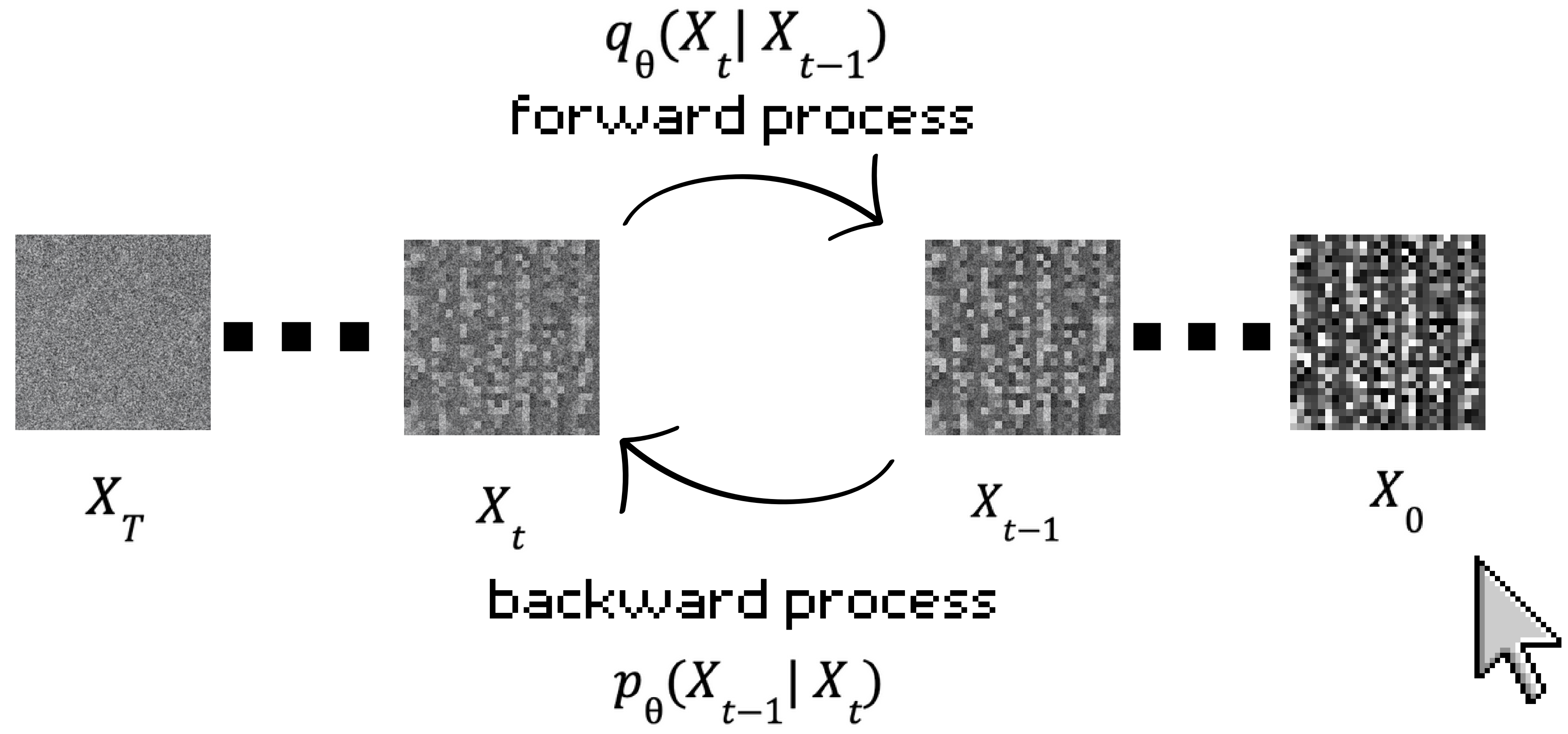
Performing the **McNemar-Bowker test**, the sample size outcome is defined for each class

Back to Agenda Page

Denoising Diffusion Probabilistic Models

Creates synthetic data by gradually **adding and removing noise** to achieve high-quality outputs

DDPM processes



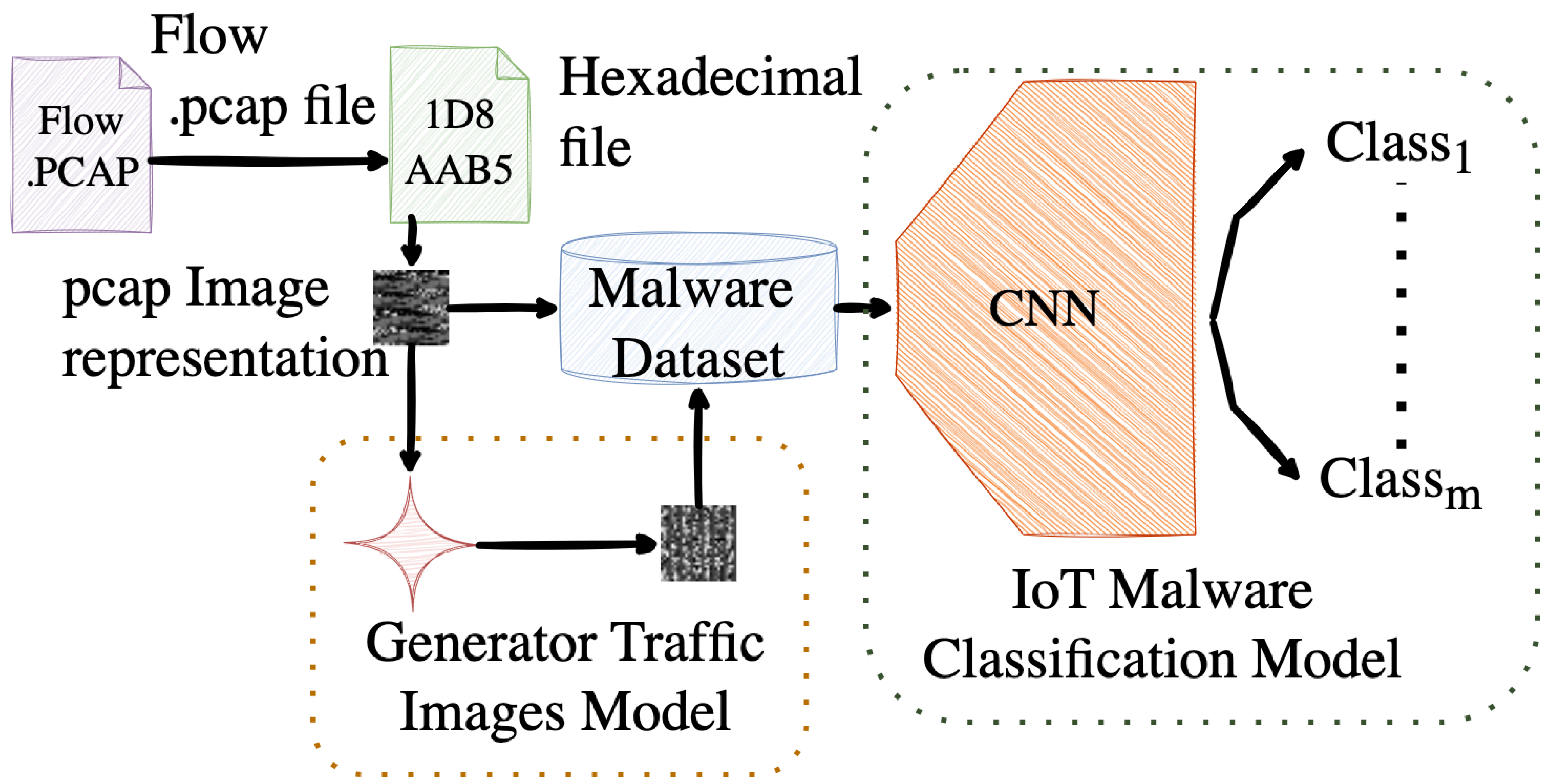
Why is DDPM better than a GAN?

More **stable and predictable** image generation

Generate loyal synthetic data and **reduce false positives**

Back to Agenda Page

Work flow



Back to Agenda Page

RESULTS

Back to Agenda Page

DDPM-generated data have:

7% higher F1-score

5% less variance

higher average AUROC

than the GAN-generated data

Back to Agenda Page

Public Available

Datasets

EDGE-IIOTSET
8 classes

Malicious Network Traffic PCAPs and binary visualization images Dataset (MNT)
4 classes

Back to Agenda Page

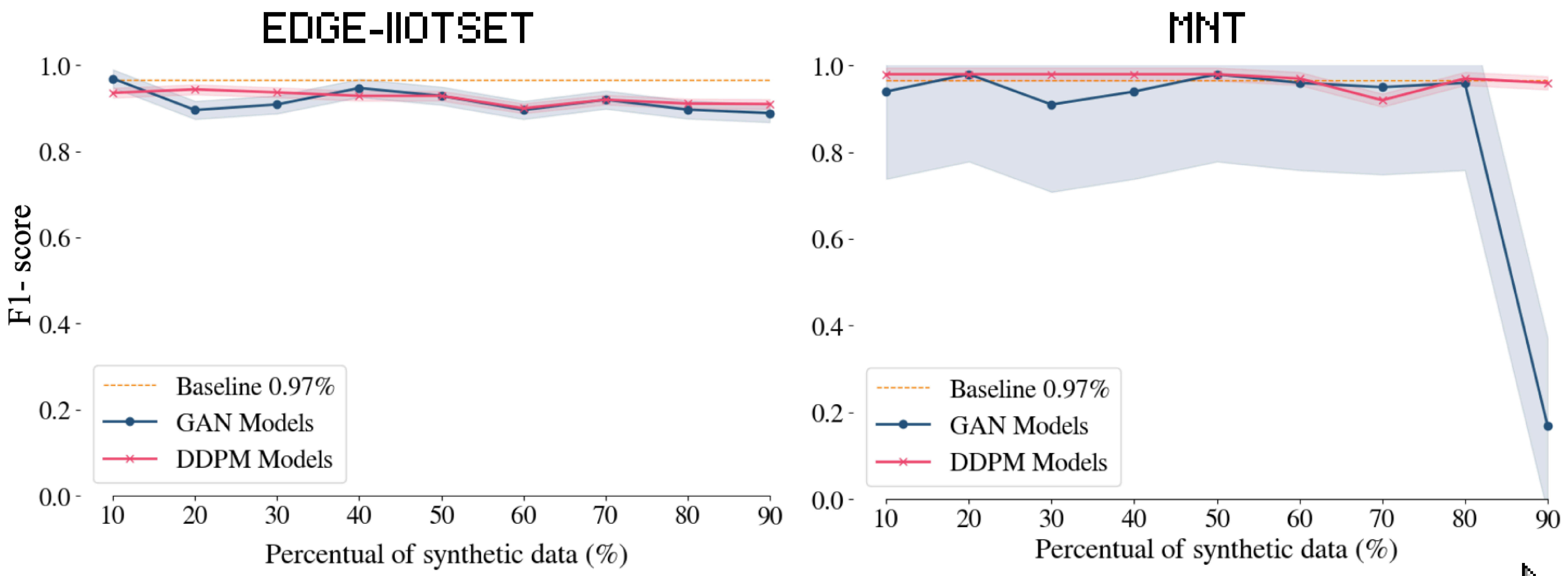
F1-score of the test sets

Class sample size	EDGE-IIOTSET (threshold 735)	MNT (threshold 580)
≤ threshold	0.6	0.7
real unbalanced train set	0.73	0.58
our train set	0.93	0.97

F1-score 0.8
alpha 0.05
beta 0.128

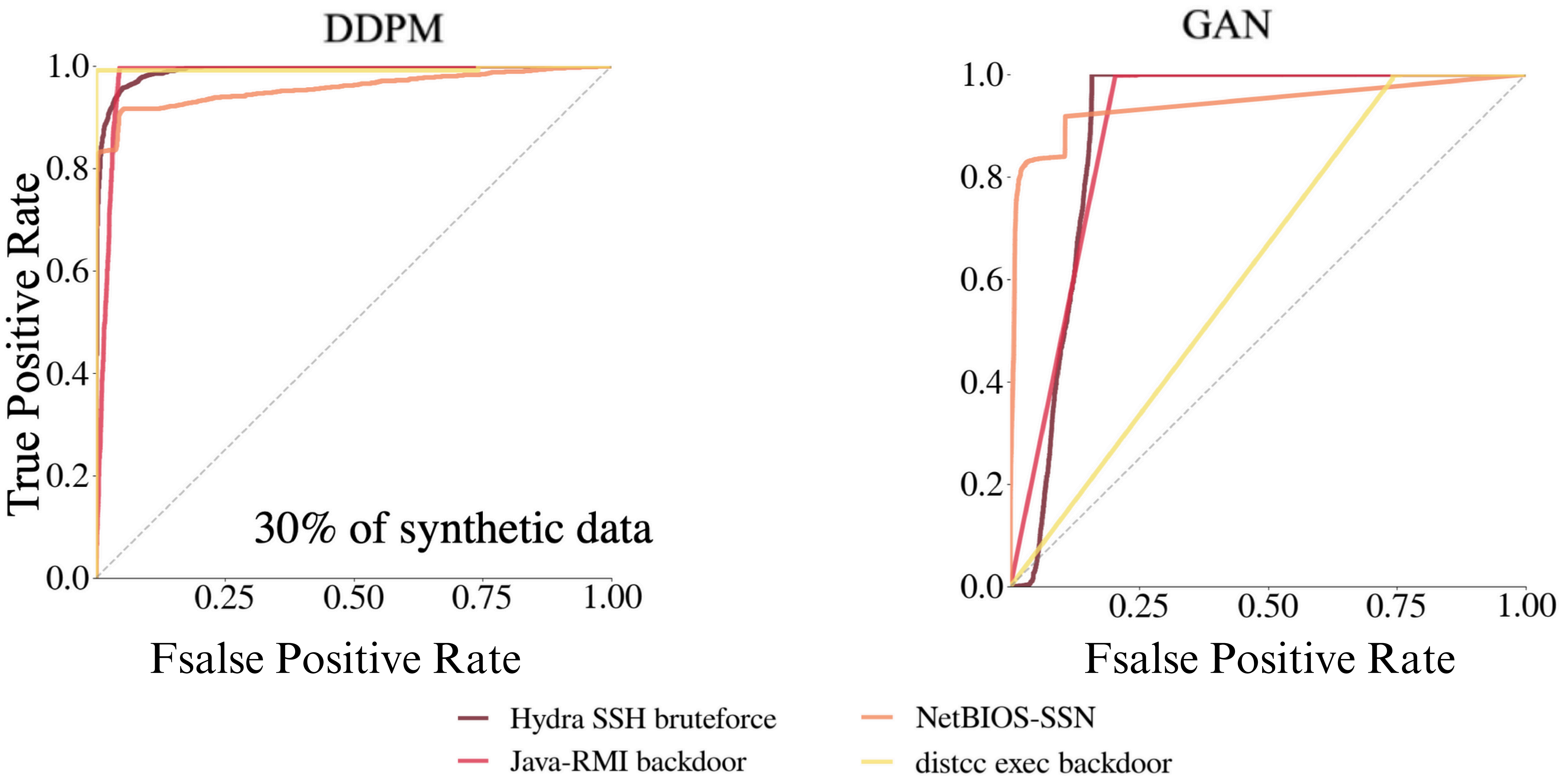
Back to Agenda Page

Resilience of F1-score accross synthetic data levels



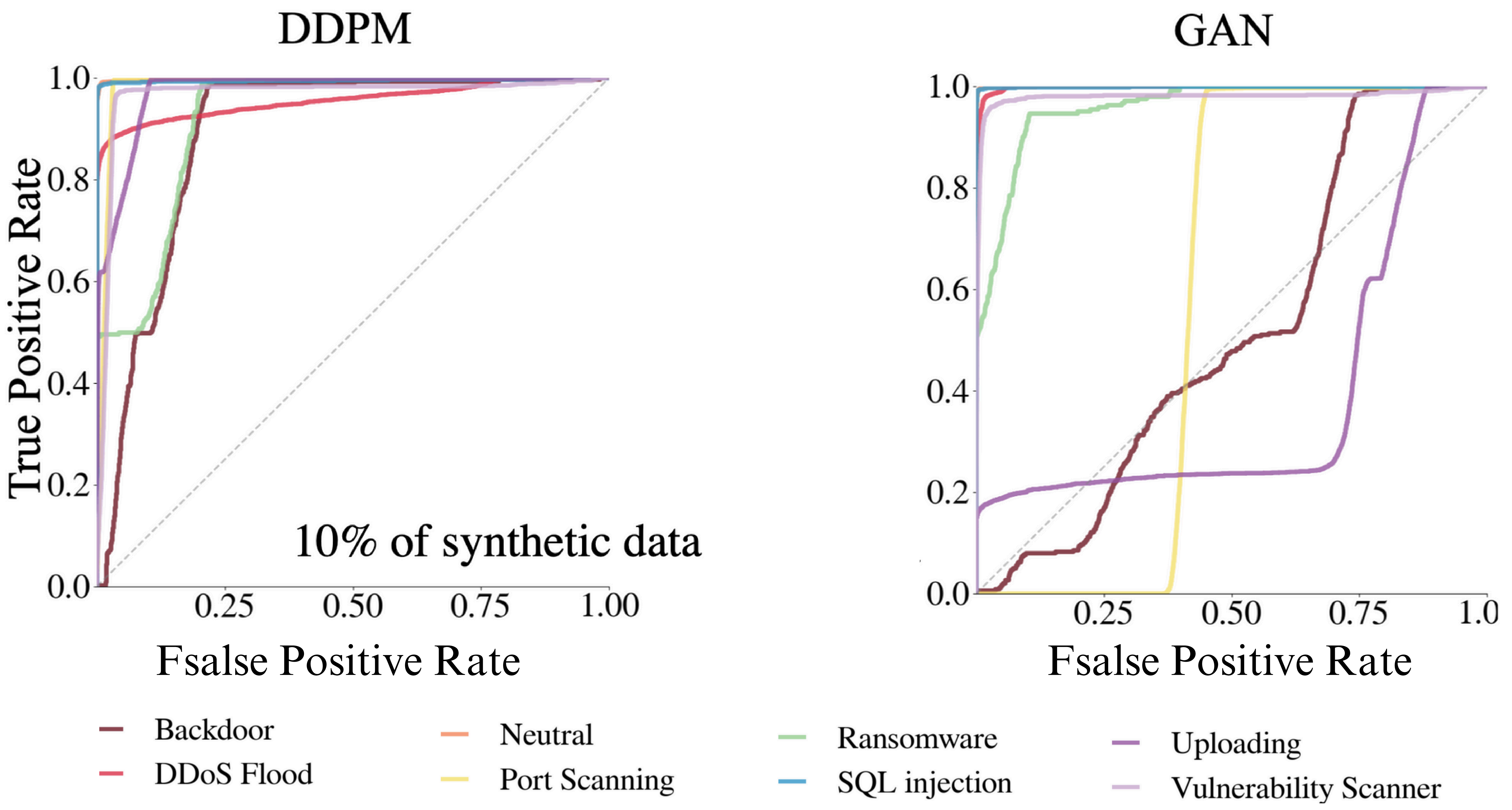
Back to Agenda Page

MNT Dataset - ROC curve



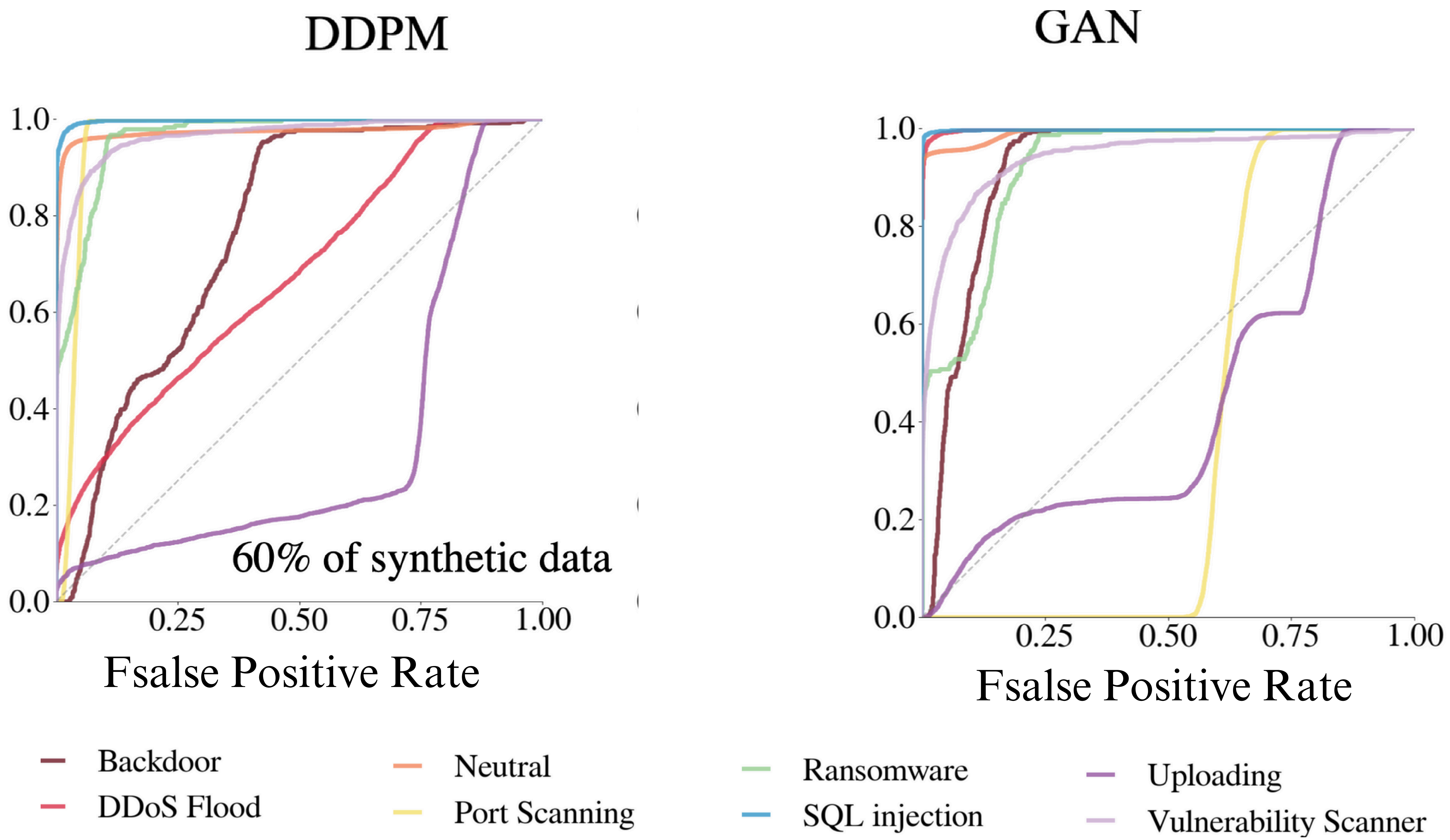
Back to Agenda Page

EDGE-IIOTSET - ROC curve



Back to Agenda Page

EDGE-IIOTSET - ROC curve



Back to Agenda Page

A

CONCLUSIONS AND FUTURE DIRECTIONS

Back to Agenda Page

sion • Conclusion • Conclusion • Conclusion • Conclusion • Conclusion • Conclusion • Conclusion Conclusion • Conclusion • Concl

EDGE-IIOTSET and MNT

DDPM models achieve higher F1 scores and lower false-positive rates across both datasets compared to GAN

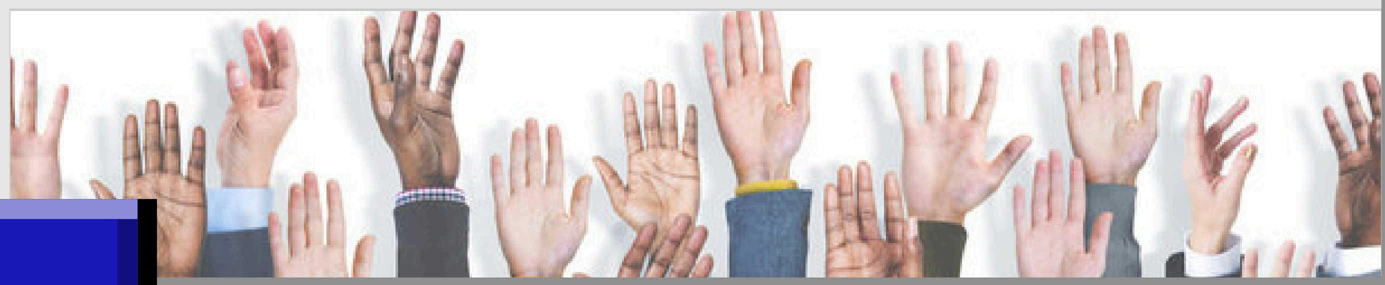
Explainable AI (XAI)

Confirm that DDPM images closely match real-world data

Future Directions

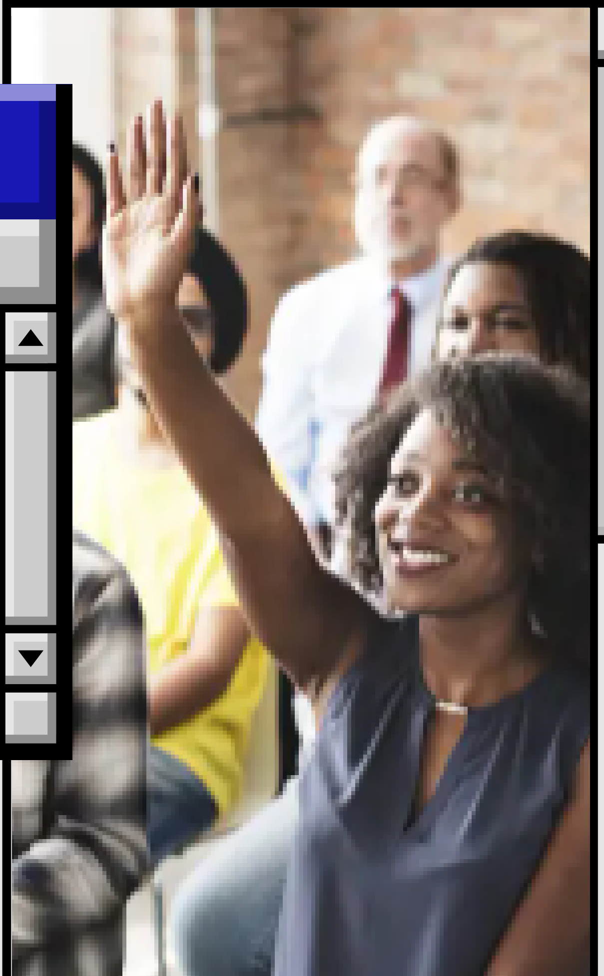
Explore further optimization of DDPM for specific IoT applications and validate across more diverse datasets

Back to Agenda Page



Back to Agenda Page

Q & A session



Thank you!

Please, for any clarification write to

chiara.camerota@unifi.it