UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

itee PhD information technology electrical engineering

DIETI UNINA

Alfredo Nascita

Explaining and Improving DL Models for Network Traffic Analysis:

Unveiling the Black Box via XAI

Tutor: Prof. Valerio Persico

Cycle: XXXVII Year: III

itee PhD information technology electrical engineering

My Background

- MSc degree in **Computer Engineering**, University of Napoli Federico II
- PhD start date – end date : 01/11/2021 - 31/10/2024
- DIETI Research group/laboratory: **Traffic** Group/**ARCLab**
- Scholarship type: **UNINA**
- Period abroad: **Huawei Technologies France**, Paris (15/01 - 14/07/2024)

Summary of Study Activities

- **9 courses**
  - 6 PhD Courses
  - 1 MSc Course
  - 2 External Courses

- **33 Seminars**
  - Network Security, Deep Learning, Artificial Intelligence

- **3 PhD Schools**
  - 2022 eXplainable AI Summer School (XAISS), Delft, Netherlands
  - 2022 PhD school of Network Traffic Measurement Analysis Conference (TMA), Enschede, Netherlands
  - 2023 PhD school of Network Traffic Measurement Analysis Conference (TMA), Napoli, Italy

Summary of Study Activities

- **8 Conferences**
  - Conference on emerging Networking EXperiments and Technologies
  - IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops
  - IEEE Conference on Computer Communications Workshops
  - IEEE Symposium on Computers and Communications
  - …

- **Tutorship Activities**
  - Bachelor's and Master's Degree courses in Computer Engineering

Research Area(s)

- **Main Research Area**

  Network Traffic Analysis with focus on explainability of Deep Learning (DL) models for analyzing Internet traffic

- **Other Projects**

  DL for attack classification and anomaly detection in Internet of Things networks

  analysis of of intrusion detectors in different network conditions

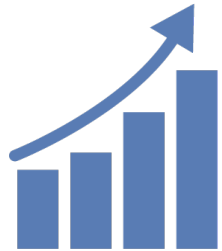
  

  design of DL solutions for class incremental traffic classification

Research Field of Interest

- Network Traffic Analysis (NTA)
  - Collecting and examining network data
  - Understanding and improving network performance

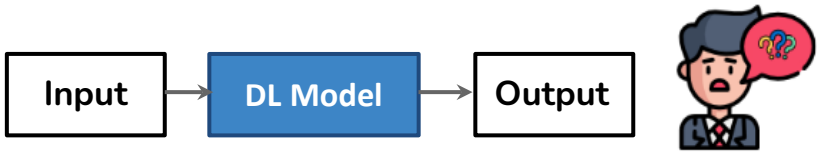
- Challenges
  - Rapid traffic growth
  - Networks' dynamicity
  - Encryption protocols



Research Activity: Overview

Deep Learning is a promising strategy to face these challenges but…

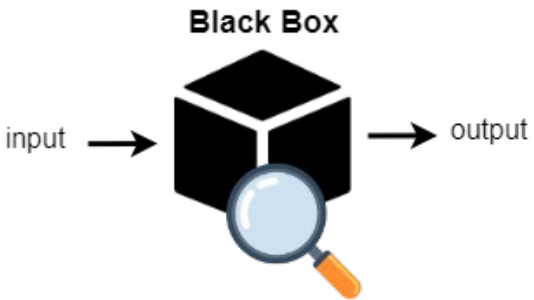
- Architectures' complexity
- Black-box nature
- Lack of Interpretability



Network operators do not trust using DL tools in real scenarios as long as they struggle to understand the logic behind their decisions

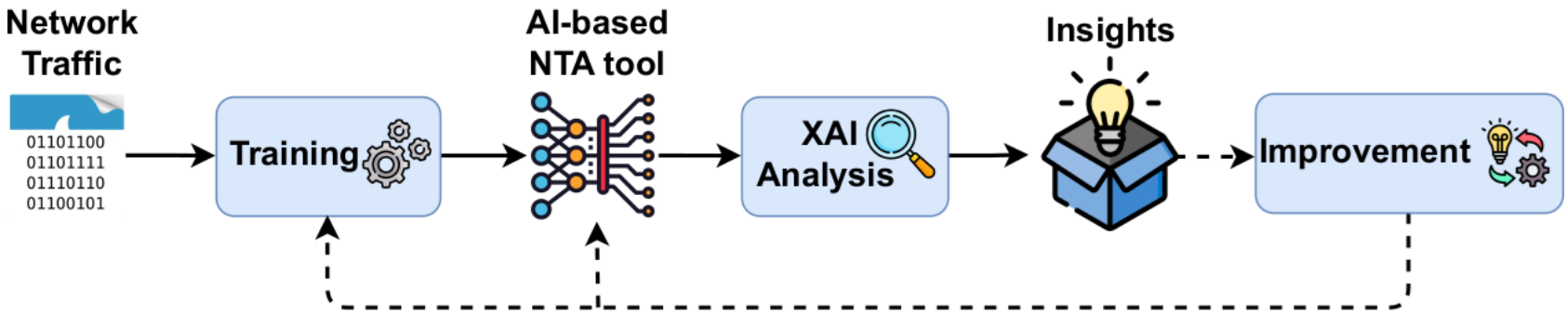
eXplainable Artificial Intelligence (XAI)

- Analyze data and models
- Justify model behaviors
- Enhance trust in decisions



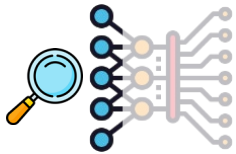
PhD Thesis Overview: Overview

Methodological Key Steps

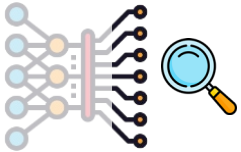


Explainability Aspects

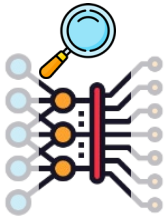
A Role of Traffic Input



B Reliability of Outputs



C Knowledge Localization in Inner Layers



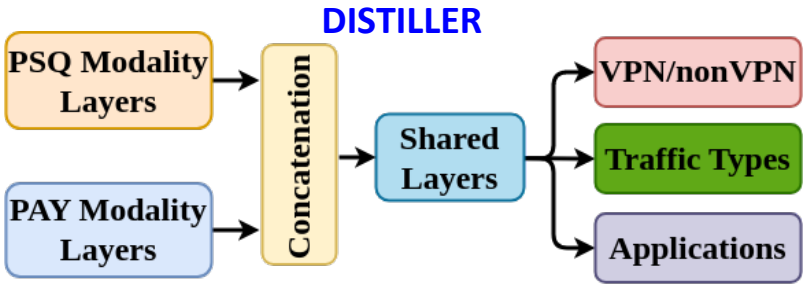
D Incremental Models

A

Understanding the Role of Traffic Input

- Multimodal: different traffic views
  - PSQ: Fields of the first 32 Packets
  - PAY: 784 Bytes of L4 payload
- Multitask: multiple TC tasks simultaneously
  - VPN/non VPN
  - Traffic Types
  - Applications

DL-based multimodal/multitask traffic classification engine

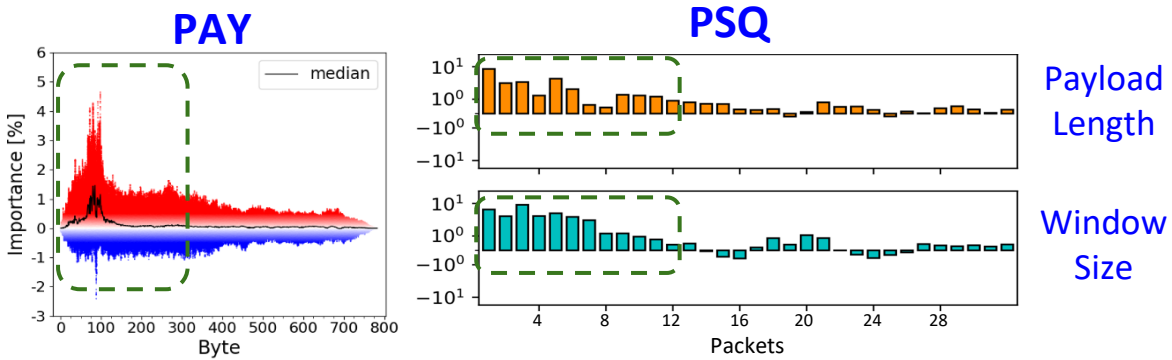


Analysis

- 2 interpretability techniques: SHAP and Integrated Gradients
- from local (single sample) to global explanations (group of samples)
  - Relative importance of each modality
  - Packets importance (PSQ)
  - Bytes importance (PAY)

A

Refining Model Complexity



Improvement

- No degradation classification capabilities (f1 score)
- -35% spatial complexity (number of parameters)
- -58% training times (run time per epoch)
- Improved earliness: 12 packets (vs. 32)

B

Investigating and Improving the Reliability of Traffic Classifiers' Outputs

Calibration of Probabilistic Outputs

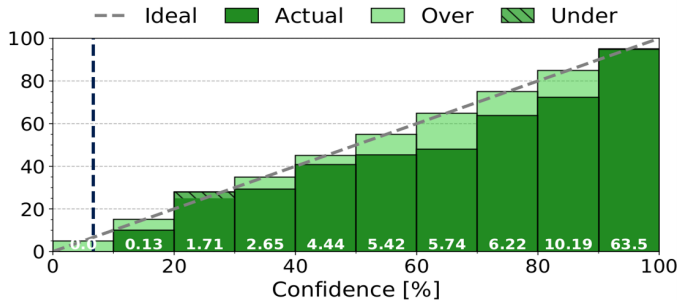
Analysis

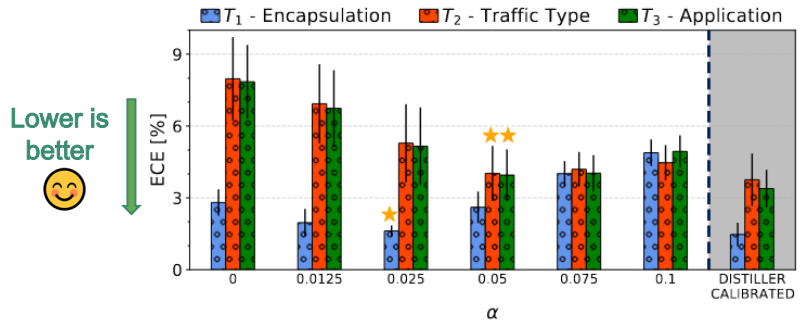
- Expected Calibration Error *(ECE)*
- Reliability diagrams

Improvement

- Label smoothing
- No degradation of classification capabilities (f1 score)
- -50% ECE for all the 3 tasks

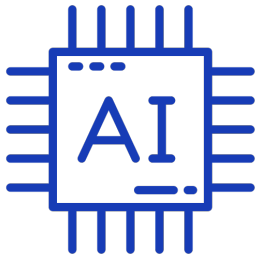
Reliability Diagram





B Deployment on Resource-constrained Devices

Compression techniques:
- Knowledge Distillation, Pruning, Quantization

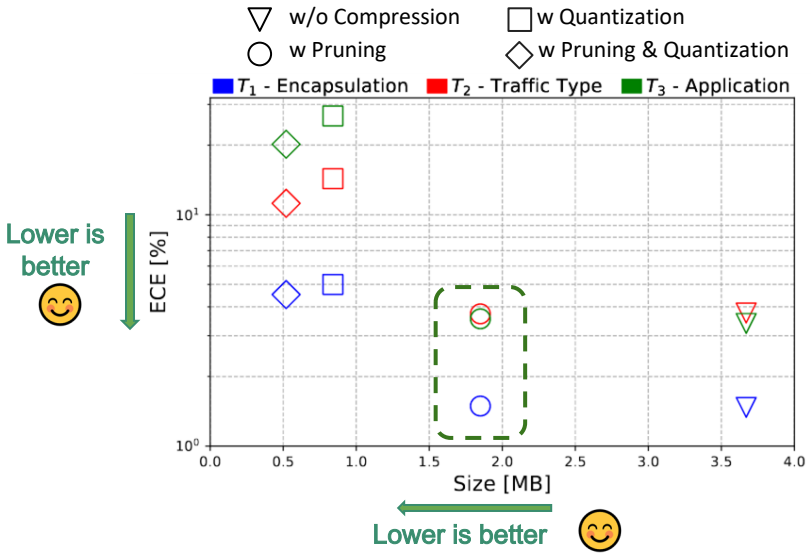


Analysis

Impact on outputs
- Performance
- Calibration

Improvement

Pruning: trade off complexity and calibration
- -50% memory occupation
- calibration unchanged



c

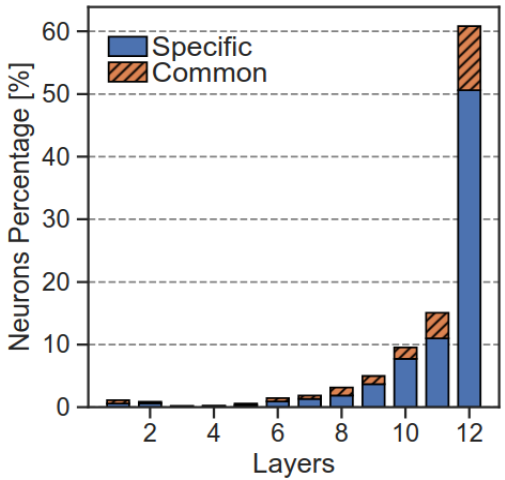
Localizing Knowledge in the Core of LLMs

Analysis

- BERT-based LLM (12 inner layers)
- Identify crucial neurons for classification task

Steps:

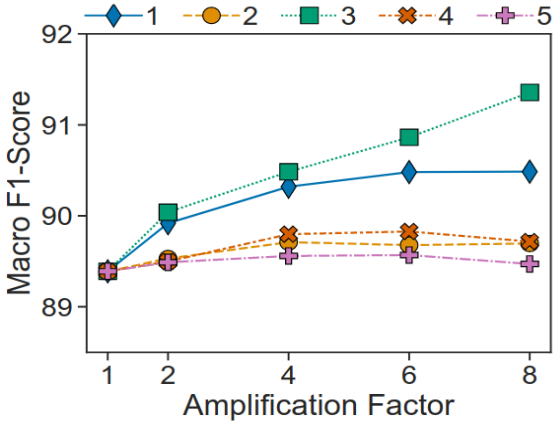
- Collect most frequent neurons
- Discard common neurons
- Identify per-class (specific) knowledge

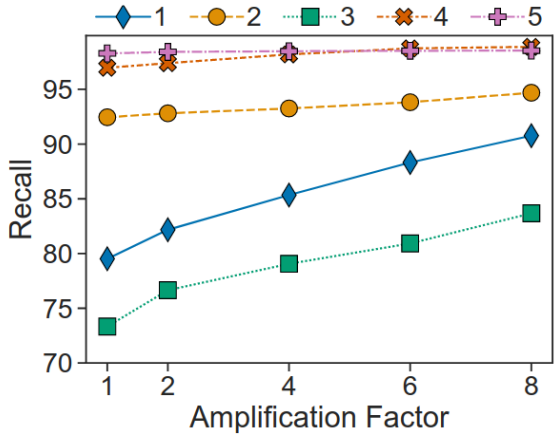


C Influencing Model Performance with Targeted Manipulations

Improvement

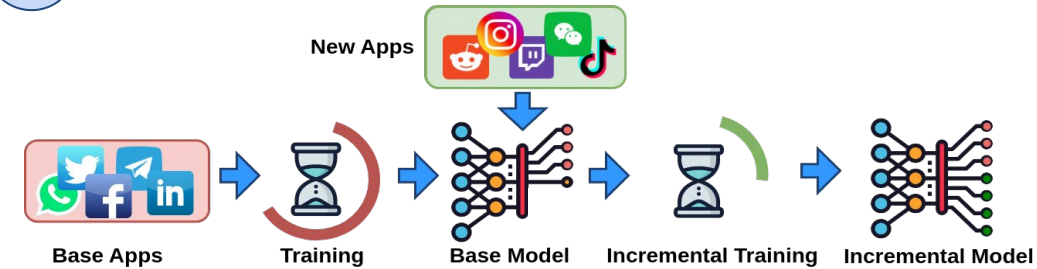
- Amplification of class-specific neurons' activations
- Slight improvement on overall performance
  - +≈2% macro f1 score
- Recall enhancement:
  - Class 1: +10%
  - Class 3: +8%
- No additional fine-tuning





D

XAI for Incremental Models



Class Incremental Learning (CIL)

Training from scratch

Performance of CIL approaches is not satisfactory (gap w.r.t. Scratch)

Analysis

- Base Models
- Incremental Models
- Comparison with Scratch Models

D Deriving Guidelines to Improve CIL Training

Base Models (starting point of CIL procedure)

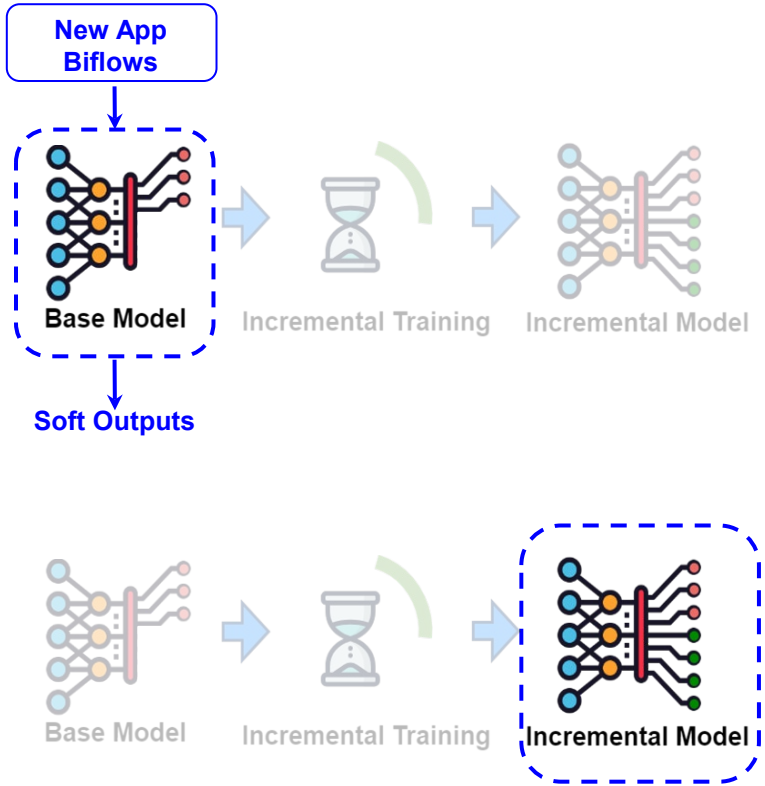
- Responses of base model for new app biflows
- Identified bias towards old apps

Scaling responses to transfer smoother knowledge to the incremental model

Incremental Models

- Backbone, Classification and Correction Layers
- Identified the reasons for the bias towards the new app

- Inclusion of old samples in validation set
- Penalize alignment between model components



Research Products

1. Machine and Deep Learning Approaches for IoT Attack Classification, A. Nascita, F. Cerasuolo, D. Di Monda, J. T. A. Garcia, A. Montieri, and A. Pescapè. INFOCOM 10th International Workshop on Security and Privacy in Big Data

2. A Comparison of Machine and Deep Learning Models for Detection and Classification of Android Malware Traffic, Giampaolo Bovenzi, Francesco Cerasuolo, Antonio Montieri, Alfredo Nascita, Valerio Persico, Antonio Pescapé, ISCC 2nd IEEE International Workshop on Distributed Intelligent Systems (DistInSys)

3. Improving Performance, Reliability, and Feasibility in Multimodal Multitask Traffic Classification with XAI, A. Nascita, A. Montieri, G. Aceto, D. Ciuonzo, V. Persico, A. Pescapé. IEEE Transactions on Network and Service Management (TNSM) 2023

4. On the Integration of Blockchain and SDN: Overview, Applications, and Future Perspectives - A. Rahman, A. Montieri, D. Kundu, Md R. Karim, Md J. Islam, S. Umme, A. Nascita, A. Pescapé, *Springer's Journal of Network and Systems Management*

5. Benchmarking Class Incremental Learning in Deep Learning Traffic Classification, G. Bovenzi, A. Nascita, L. Yang, A. Finamore, G. Aceto, D. Ciuonzo, A. Pescapé, D Rossi. Accepted for publication in IEEE Transactions on Network and Service Management (TNSM) 2023

Research Products

6. MCOTM: Mobility-Aware Computation Offloading and Task Migration for Edge Computing in Industrial Iot, W. Qin, H. Chen, L. Wang, Y. Xia, A. Nascita, A. Pescapé. Elsevier Future Generation Computer Systems (FGCS) journal

7. MEMENTO: A Novel Approach for Class Incremental Learning of Encrypted Traffic, F. Cerasuolo , A. Nascita , G. Bovenzi, G. Aceto , D. Ciuonzo , A. Pescapé , D. Rossi. Elsevier Computer Networks

8. Cross-Evaluation of Deep Learning-based Network Intrusion Detection Systems, C. Guida, A. Nascita, A. Montieri, A. Pescapé, 10th International Conference on Future Internet of Things and Cloud (FiCloud 2023)

9. Explainable Mobile Traffic Classification: the case of Incremental Learning, A. Nascita, F. Cerasuolo, G. Aceto, D. Ciuonzo, V. Persico, A. Pescapé, 19th International Conference on emerging Networking EXperiments and Technologies

10. Interpretability and Complexity Reduction in Iot Network Anomaly Detection Via XAI, A. Nascita, R. Carillo, F. Giampetraglia, A. Iacono, V. Persico and A. Pescapé, 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)

11. *Can XAI Tools Interpret Traffic Classifiers based on Deep Learning?,* A. Nascita, A. Montieri, G. Aceto, D. Ciuonzo, V. Persico, A. Pescapé, Secondo Convegno Nazionale CINI sull'Intelligenza Artificiale, Torino, Italy, February 2022.

Research Products

12. A Survey on Explainable Artificial Intelligence for Internet Traffic Classification and Prediction, and Intrusion Detection, A. Nascita, G. Aceto, D. Ciuonzo, A. Montieri, V. Persico, and A. Pescapé, submitted to IEEE Communications Surveys and Tutorials (under second review round)

13. *[hidden title],* A. Nascita, J. Krolikowski, V. Persico, A. Pescapé, D. Rossi, submitted to the IEEE International Conference on Computer Communications (INFOCOM) 2025 (under double-blind review process at the date of submission of this document)

Conclusions

Understanding and **improving** DL Models for Network Traffic Analysis

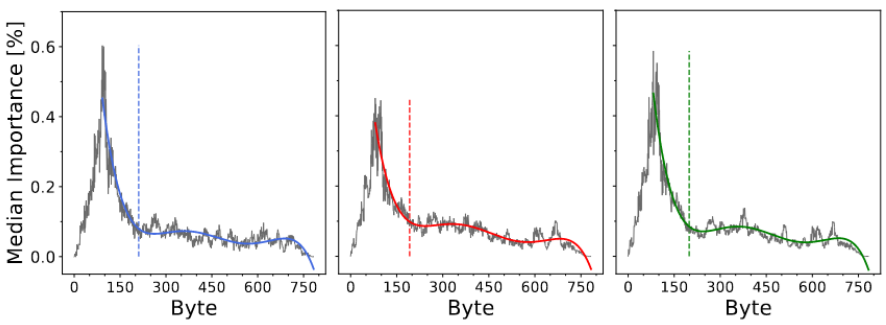
Different **explainability** aspects:

- Role of **Traffic Input**
  - Input Importance → -58% training times, -35% parameters
- Reliability of Classifier **Outputs**
  - Calibration techniques→ -50% ECE
- Knowledge Localization in **Inner Layers**
  - Class-specific neurons→ +8-10% Recall improvement
- **Incremental** Models
  - Differences w.r.t scratch→ Guidelines for improving incremental training

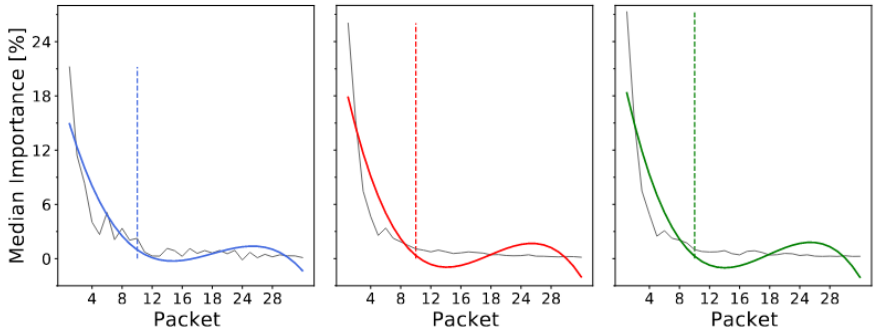
Thank you for the attention!

Backup Slides

Median Importance



(a) T_1 PAY. (b) T_2 PAY. (c) T_3 PAY.



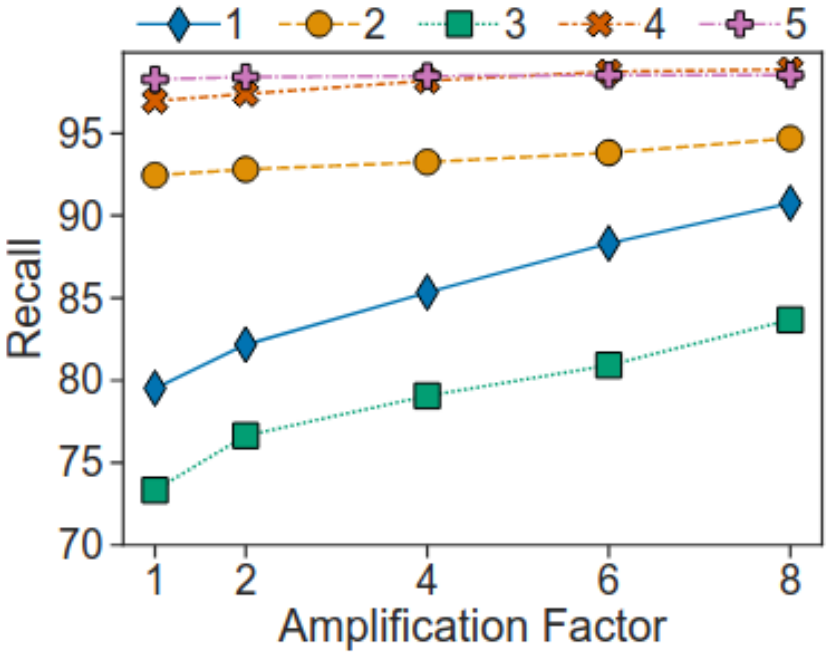
(d) T_1 PSQ. (e) T_2 PSQ. (f) T_3 PSQ.

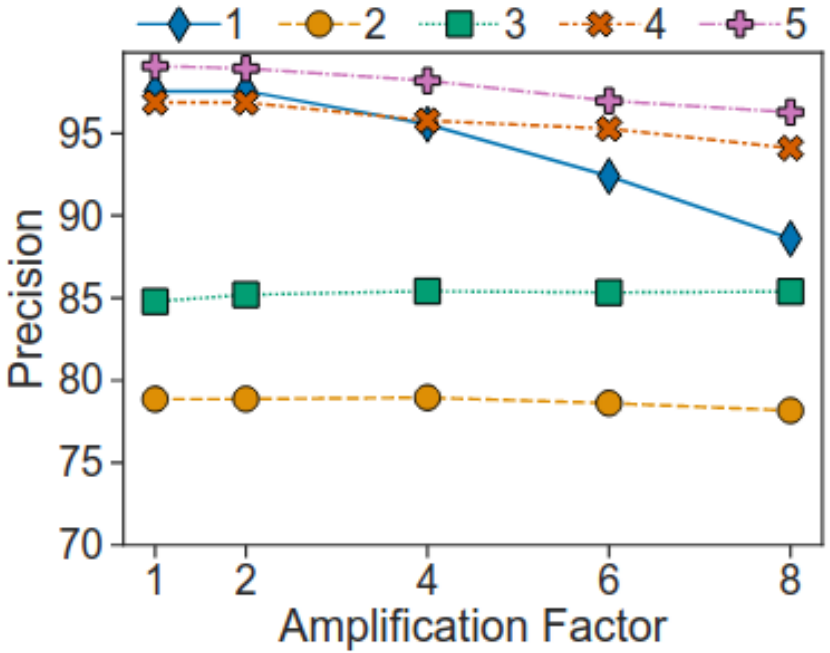
Label Smoothing / ECE

```
y_ls = (1 - α) * y_hot + α / K
```

$$\mathrm{ECE} \approx \sum_{m=1}^{M} (|B_m| / N) \, |\mathrm{acc}(B_m) - \mathrm{conf}(B_m)|$$

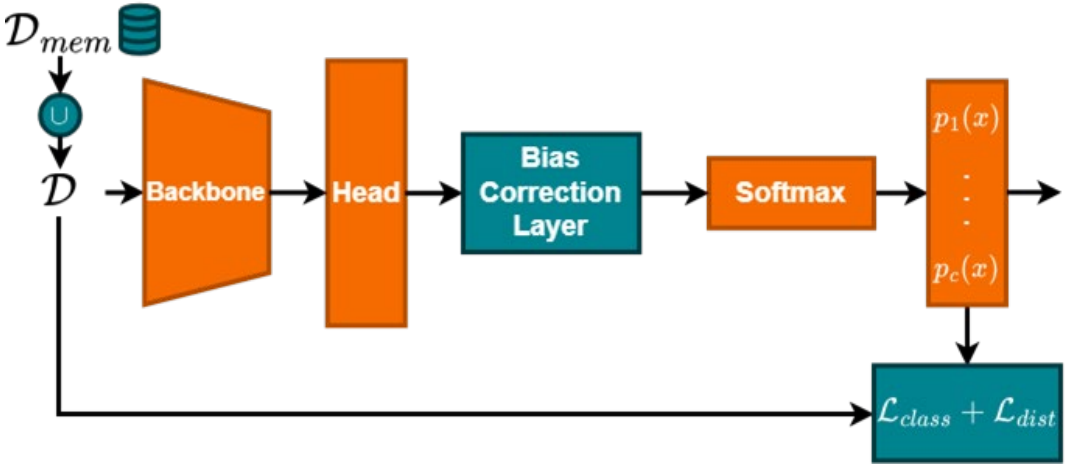
Precision/ Recall LLMs





CIL Approach: Bias Correction (BiC)

- **Fine-Tuning Family:** All weights are updated in the new training phase
- Strategies:
  - **Rehearsal:** Storage of old samples (memory)
  - **Regularization:** Knowledge Distillation
  - **Bias Correction:** Linear trainable layer to correct bias towards new apps



* Wu Yue et al., "Large scale incremental learning." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019

CIL Analyses