University of Glasgow | School of Computing Science

THE AWARDS 2020 | UNIVERSITY OF THE YEAR

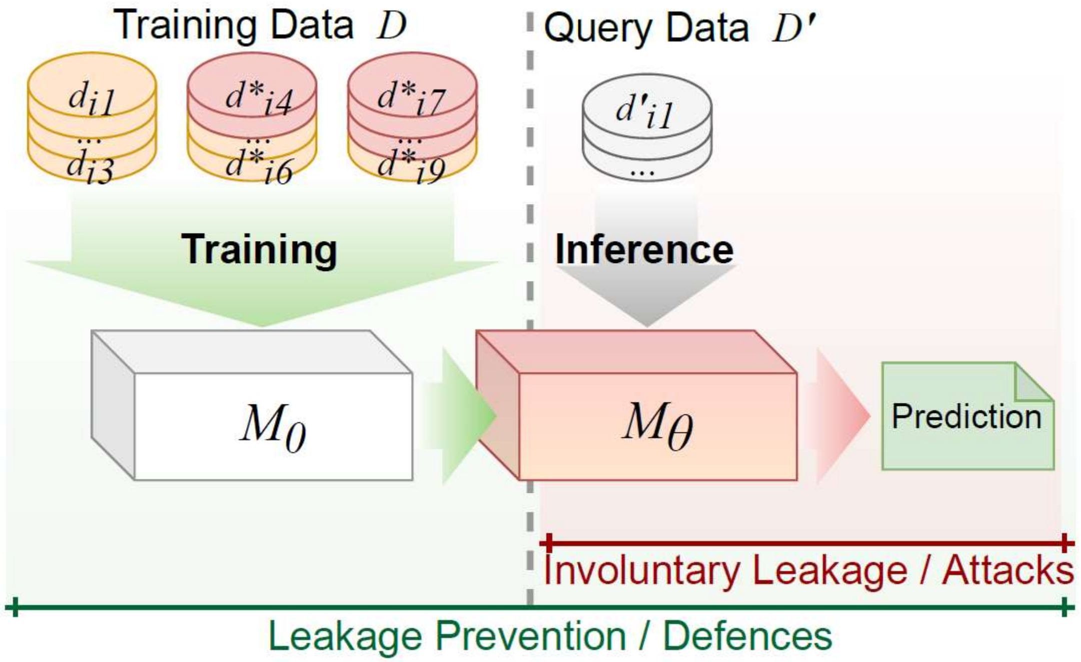
Privacy Concerns in CDSS

Dr. Fani Deligianni,
fani.deligianni@glasgow.ac.uk
Lecturer (Assistant Professor)
Lead of the Computing Technologies for Healthcare Theme
https://www.gla.ac.uk/schools/computing/staff/fanideligianni

WORLD CHANGING GLASGOW

Data Leakage in Neural Networks

- Neural network inherent ability to store information
- Information is represented in the neural network weights
- Original data can be recreated with high accuracy.



Jegorova et al. ‘Survey: Leakage and Privacy at Inference Time’, https://arxiv.org/abs/2107.01614, 2021.

Privacy attacks - Datasets

Attacks against the dataset

- Re-identification attack
- Dataset reconstruction attack
- Tracking attack

Privacy attacks - Algorithmic

Attacks against the dataset

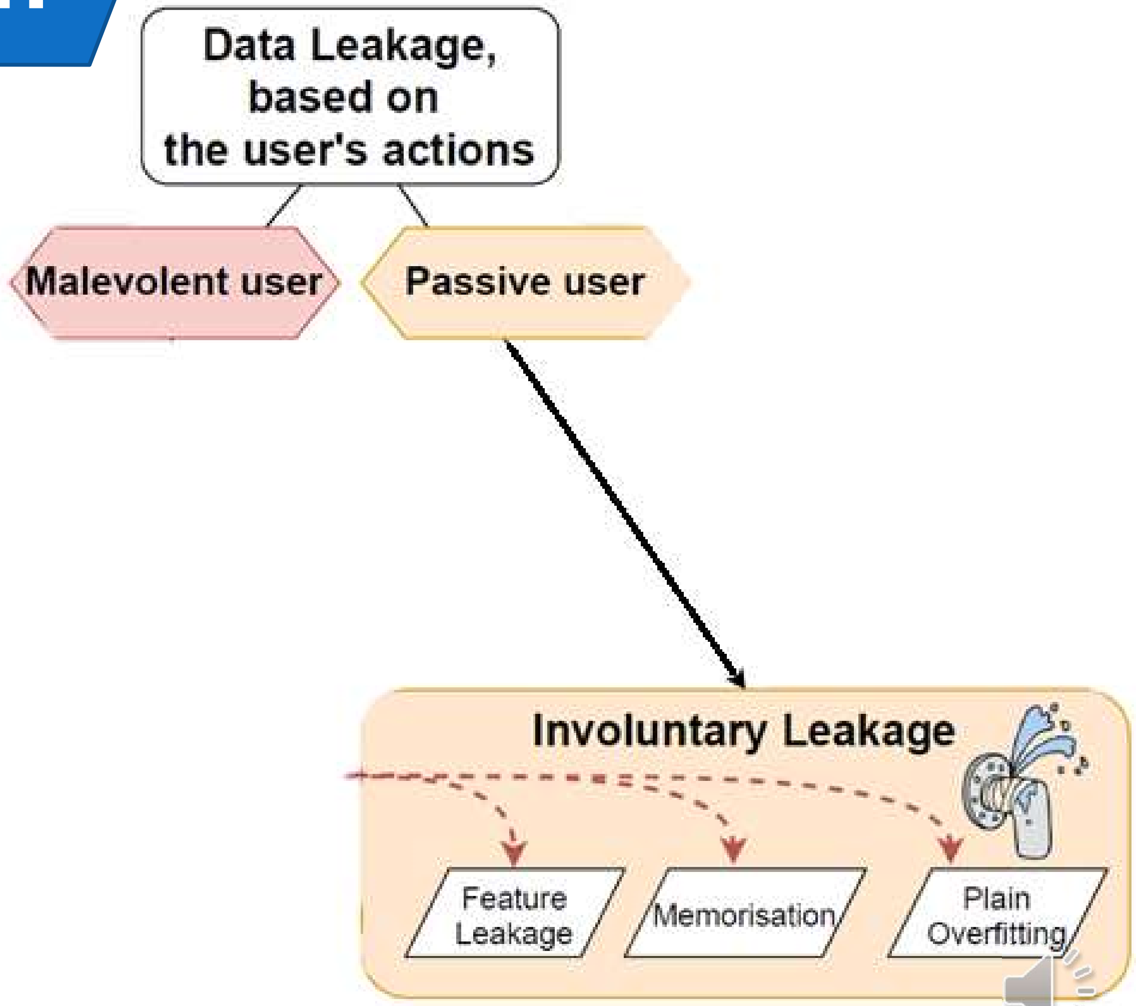
- Re-identification attack
- Dataset reconstruction attack
- Tracking attack

Attacks against the algorithm

- Adversarial attack
- Model-inversion/reconstruction attack

Leakage - Users Action

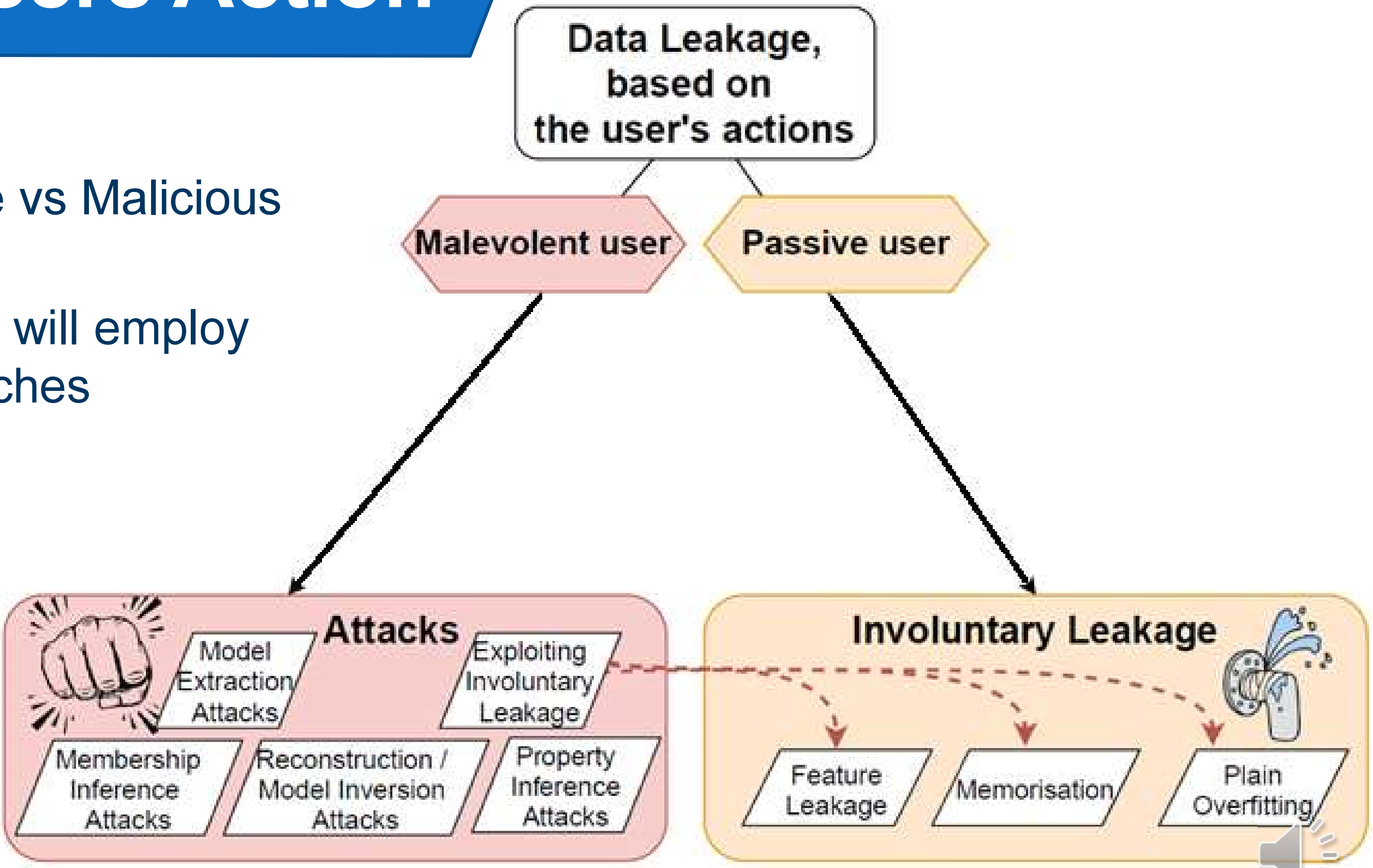
- Involuntary Leakage vs Malicious Attacks
- Data leakage can be exploited even if it is involuntary
- Feature leakage, memorization and plain overfitting are all related to the ability of deep learning models to ‘memorise’



Jegorova et al. ‘Survey: Leakage and Privacy at Inference Time’, https://arxiv.org/abs/2107.01614, 2021.

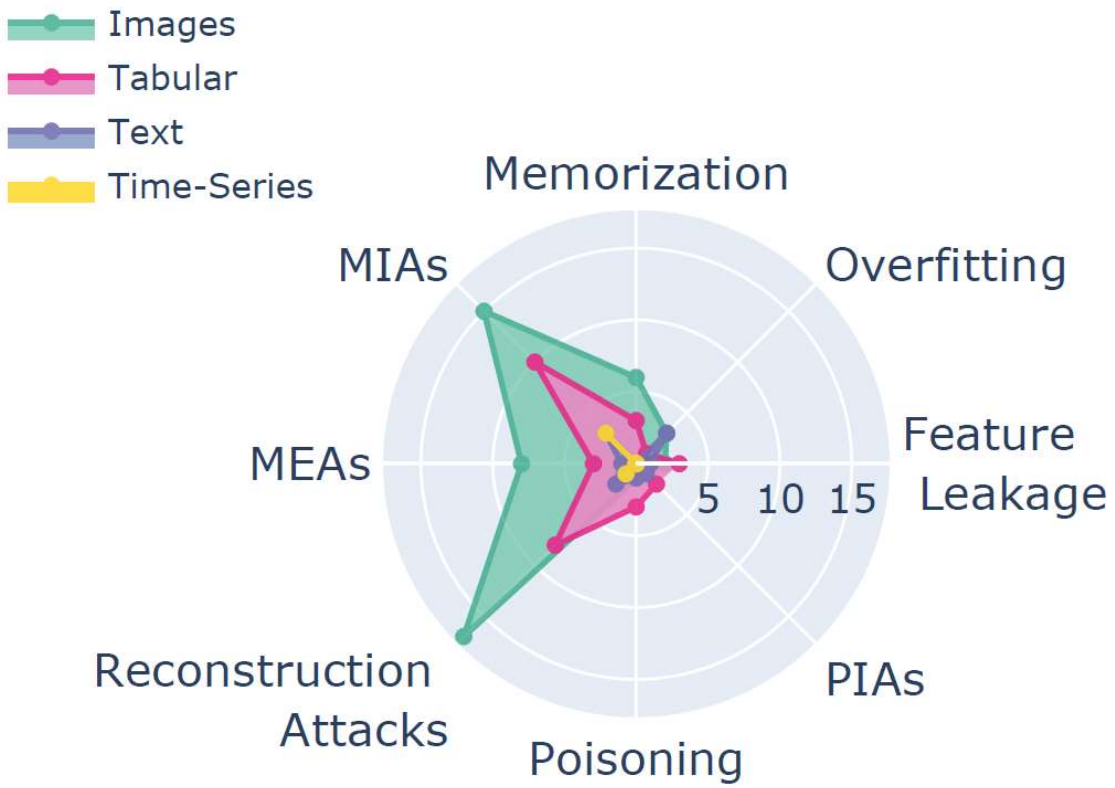
Leakage - Users Action

- Involuntary Leakage vs Malicious Attacks
- A malicious attacker will employ sophisticate approaches


Data Leakage, based on the user's actions
Malevolent user
Passive user
Attacks
Model Extraction Attacks
Exploiting Involuntary Leakage
Membership Inference Attacks
Reconstruction / Model Inversion Attacks
Property Inference Attacks
Involuntary Leakage
Feature Leakage
Memorisation
Plain Overfitting

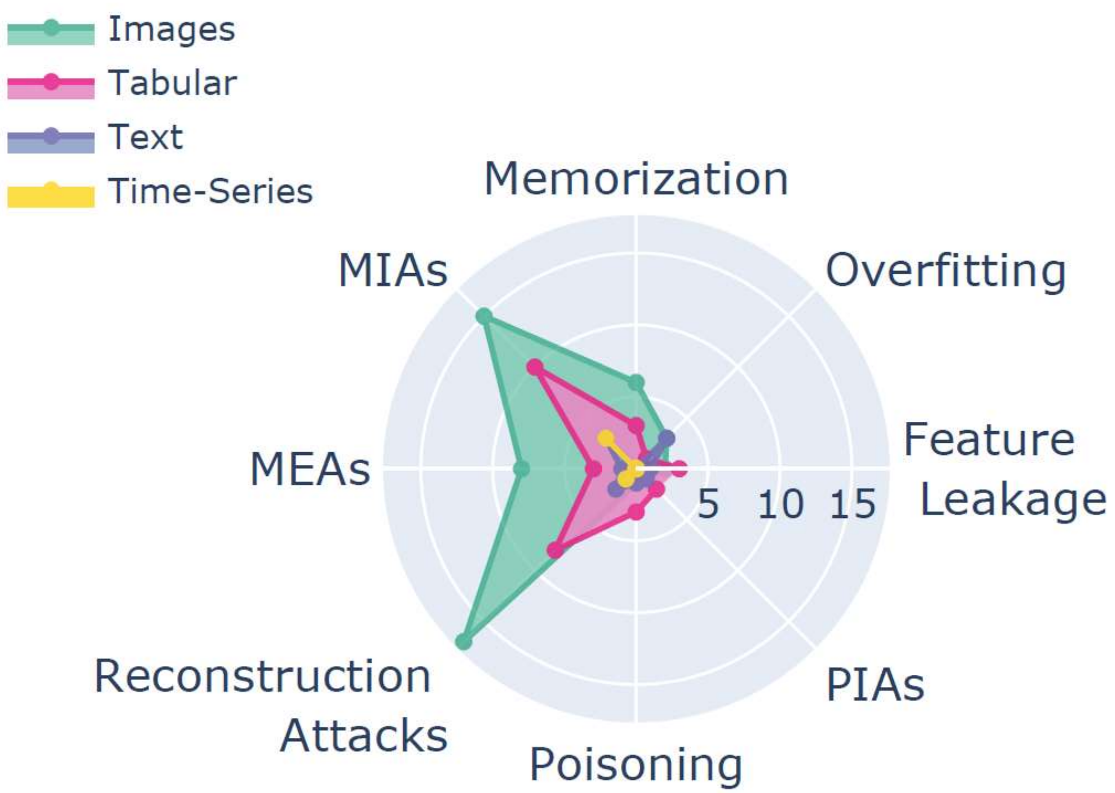
Jegorova et al. ‘Survey: Leakage and Privacy at Inference Time’, https://arxiv.org/abs/2107.01614, 2021.

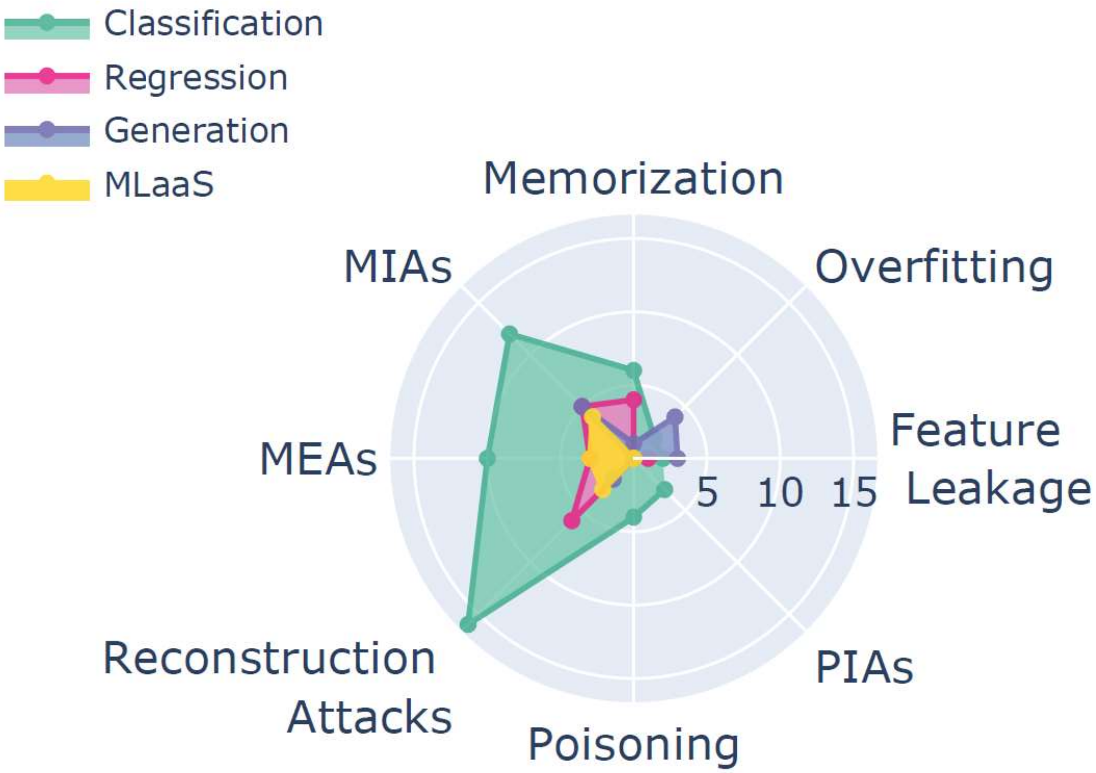
Privacy attacks - Statistics



Jegorova et al. ‘Survey: Leakage and Privacy at Inference Time’, https://arxiv.org/abs/2107.01614, 2021.

Privacy attacks - Statistics

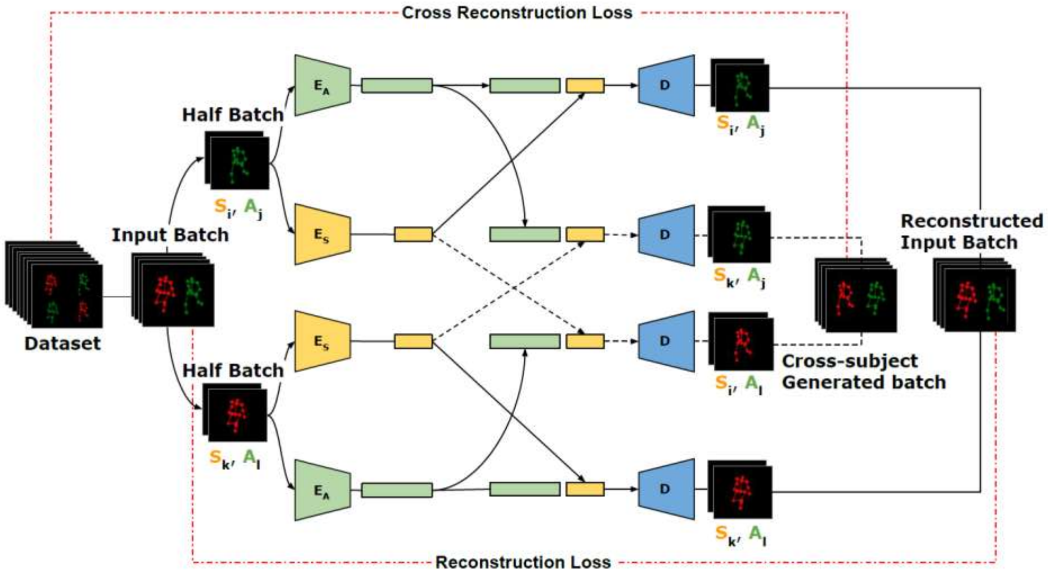




Jegorova et al. ‘Survey: Leakage and Privacy at Inference Time’, https://arxiv.org/abs/2107.01614, 2021.

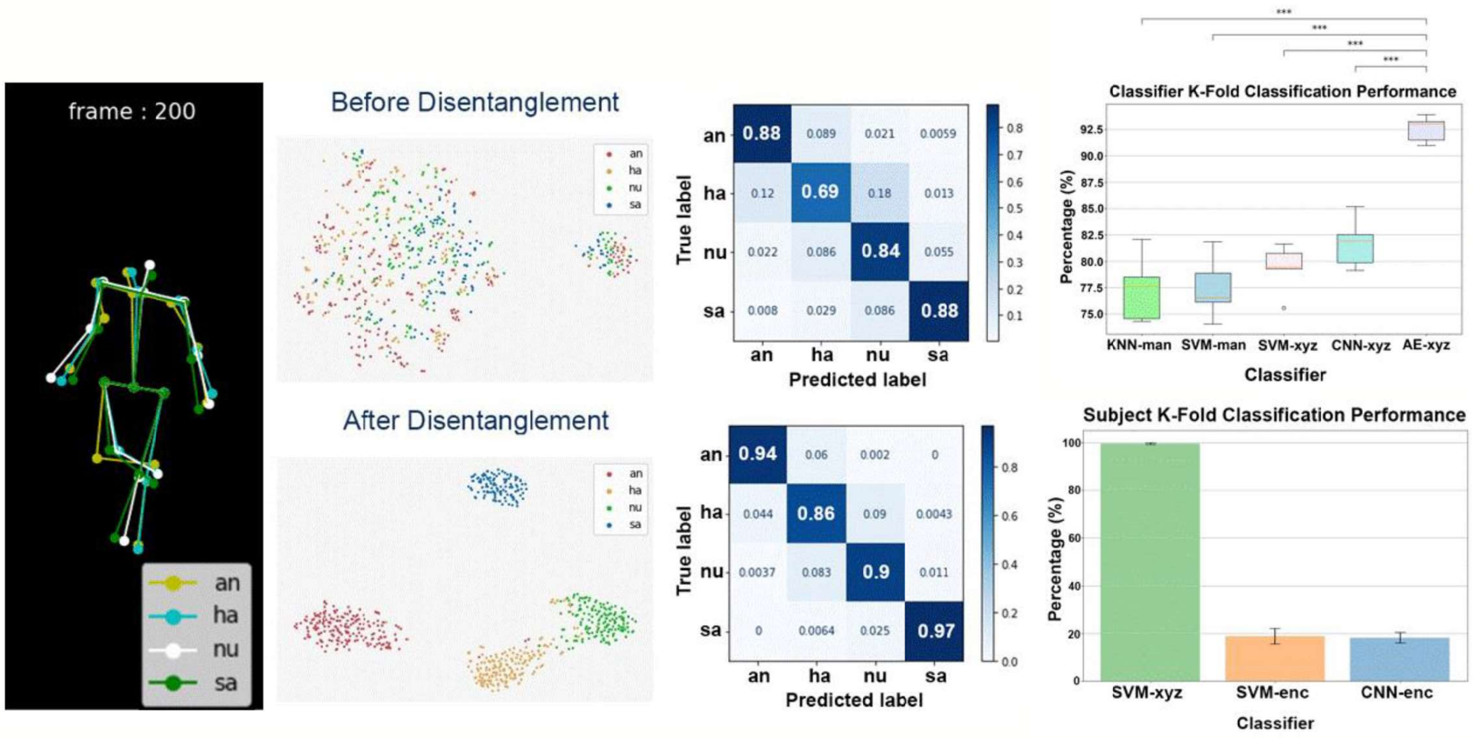
Privacy and Disentangled Representations

- Separate latent representations of identity and the characteristic of interest
- Biometrics information are filtered out in a measurable way
- Inherent designs of explainable, privacy-preserved classification



Malek-Podjaski et al. ‘Towards Explainable, Privacy-Preserved Human-Motion Affect Recognition’, IEEE Symposium Series on Computational Intelligence. https://arxiv.org/abs/2105.03958, 2021

Privacy and Disentangled Representations



Malek-Podjaski et al. ‘Towards Explainable, Privacy-Preserved Human-Motion Affect Recognition’, IEEE Symposium Series on Computational Intelligence, https://arxiv.org/abs/2105.03958, 2021

Summary

- Deep Neural Networks can memorise information with relation to the training data
- This property results in inherent vulnerabilities that can be exploited by a malicious attacker
- Considering privacy early in the development of clinical decision support systems is important
- Disentanglement can separate biometrics from the features of interest and allow to filter this information early in the processing pipeline

References

- Kaissis et al. ‘Secure, privacy-preserving and federated machine learning in medical imaging’, Nature Machine Intelligence, 2020.
- Hitaj et al. ‘Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning’, ACM CCS’17, 2017.
- Jegorova et al. ‘Survey: Leakage and Privacy at Inference Time’, https://arxiv.org/abs/2107.01614, 2021.
- Malek-Podjaski et al. ‘Towards Explainable, Privacy-Preserved Human-Motion Affect Recognition’, IEEE Symposium Series on Computational Intelligence, https://arxiv.org/abs/2105.03958, 2021.