Imperial College London

April 19th 2018

Replacement AutoEncoder

A Privacy-Preserving Algorithm for Sensory Data Analysis

Hamed Haddadi

Joint work with:

Mohammad Malekzadeh and Richard G. Clegg

Queen Mary University of London

Imperial College London

Context

- Location (~50m)
- Microphone

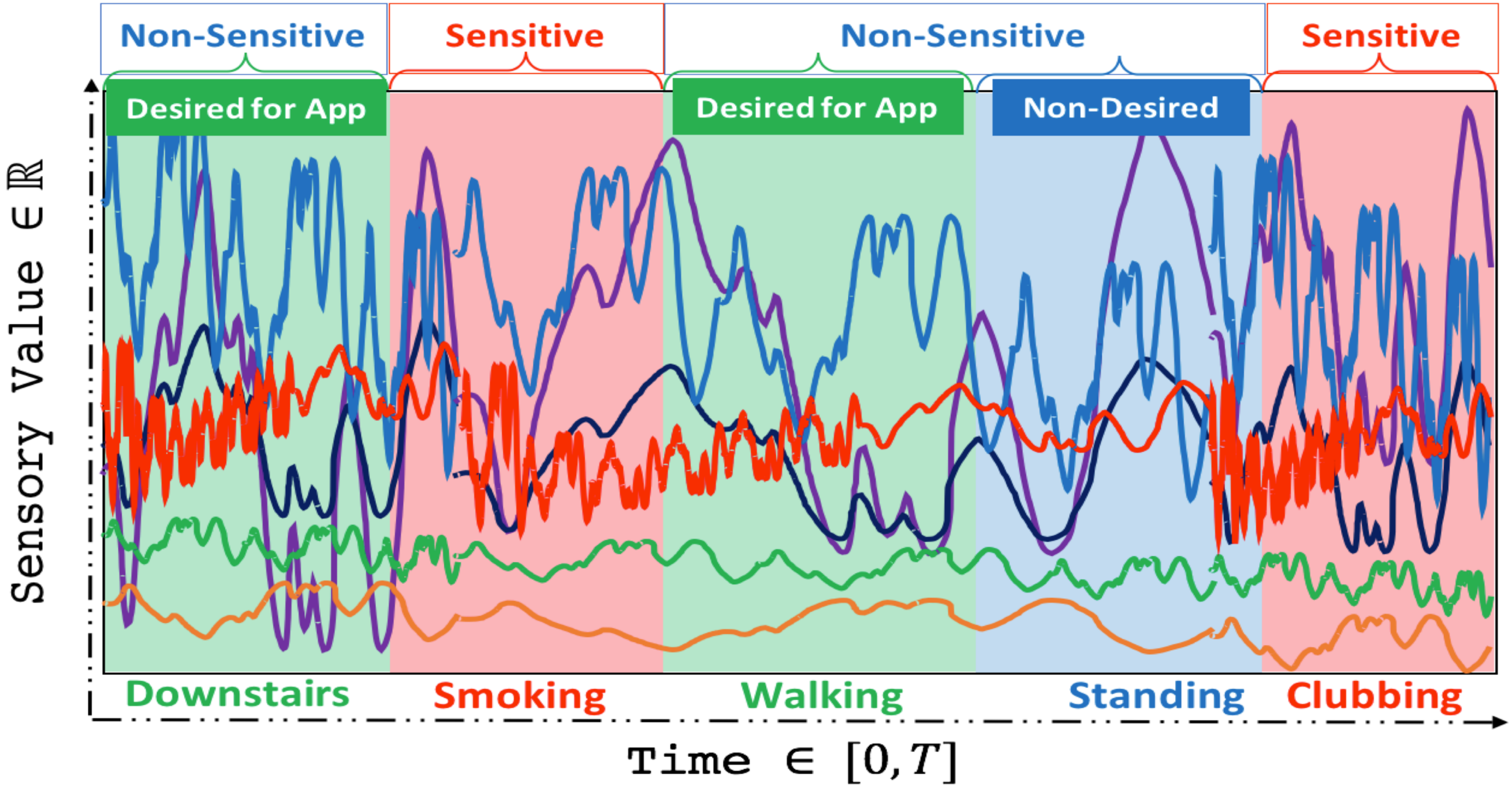
1973

- Location (~3m)
- Microphone
- **Gyroscope**
- **Accelerometer**
- Barometer
- Magnetometer
- Thermometer
- Proximity
- Ambient Light
- Humidity

2018

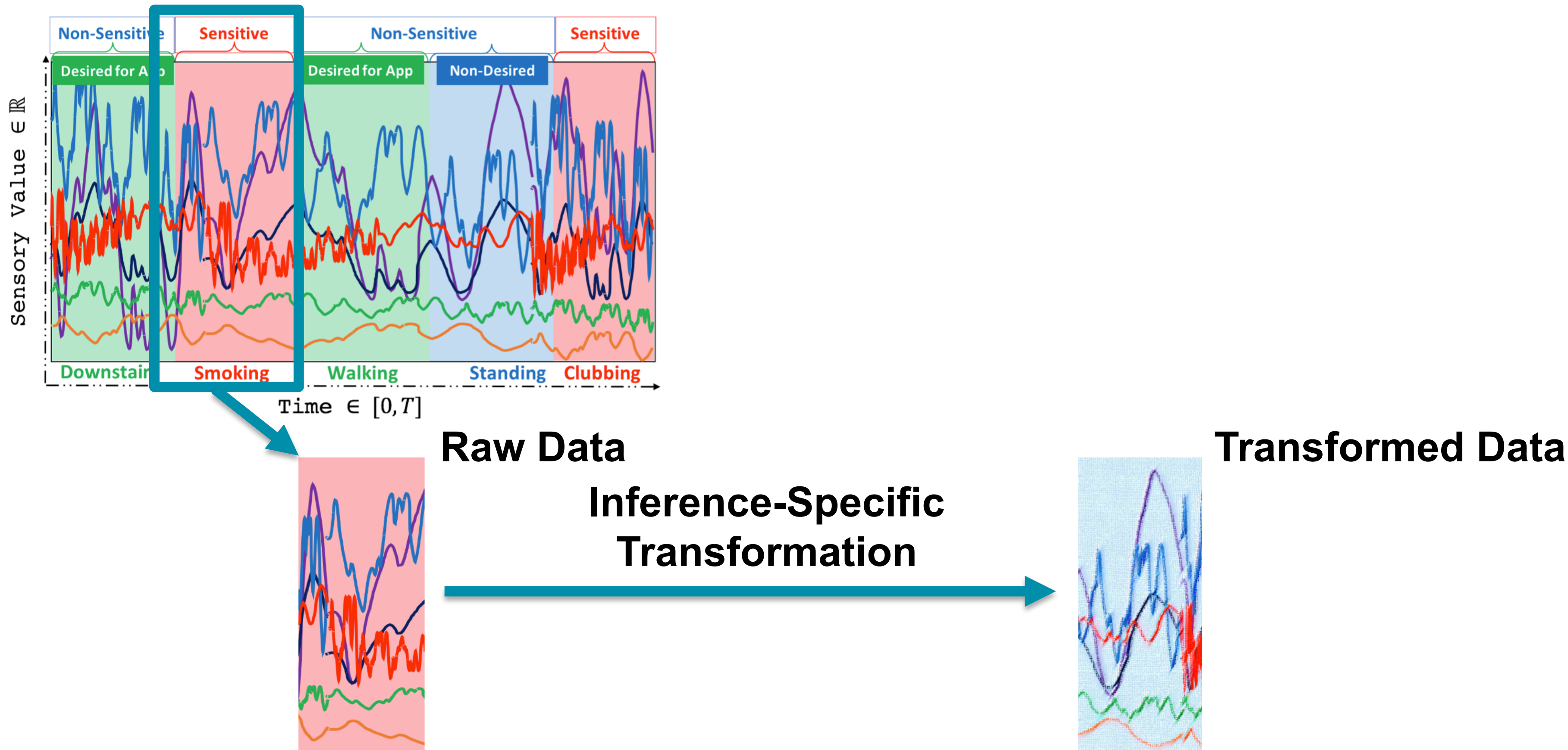
Imperial College London

Temporal Inferences on Sensory Data



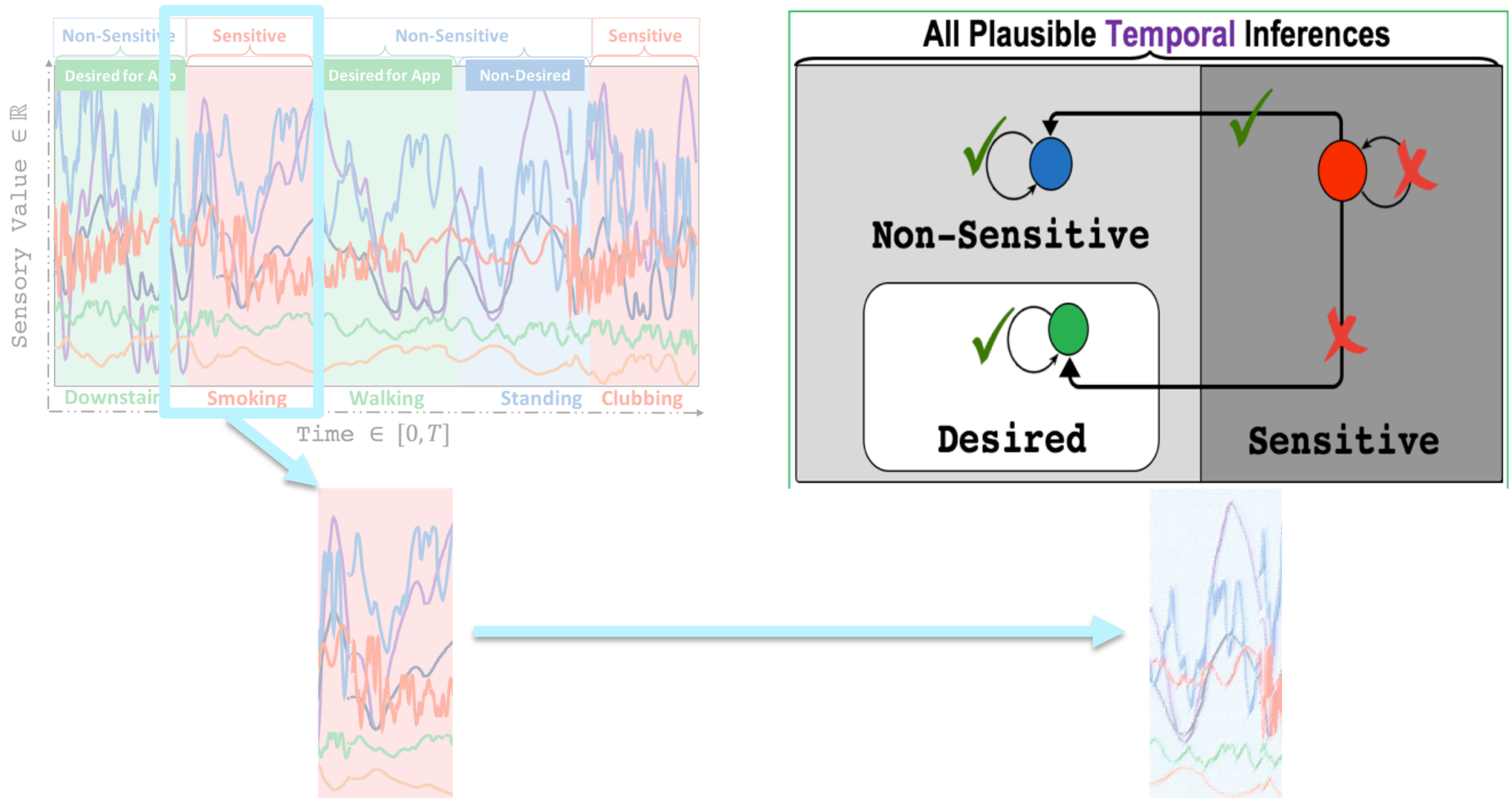
Imperial College London

Real-Time Transformation based on Corresponding State



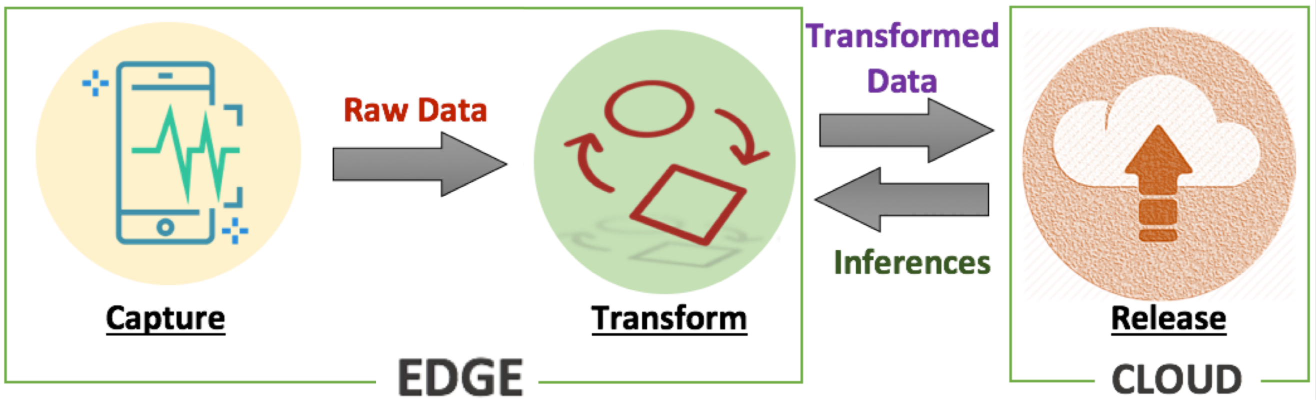
Imperial College London

Replacing Sensitive Sections with Non-Desired ones



Imperial College London

We propose a Hybrid Architecture



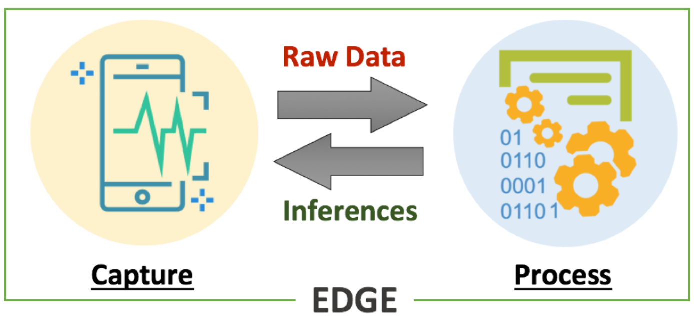
- We apply a privacy-preserving **transformation** on raw data at the **Edge**
- Then send transformed data to the **Cloud** for performing **deep analysis** and receive the **promised services.**

Imperial College London

Privacy-Utility Tradeoff

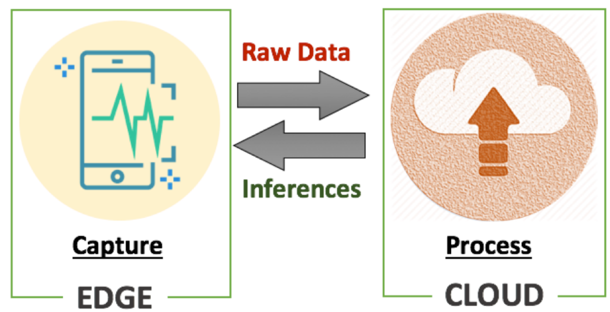
I. Edge-Based

Perfect Privacy vs. Low Utility



II. Cloud-Based

Low Privacy vs. Perfect Utility

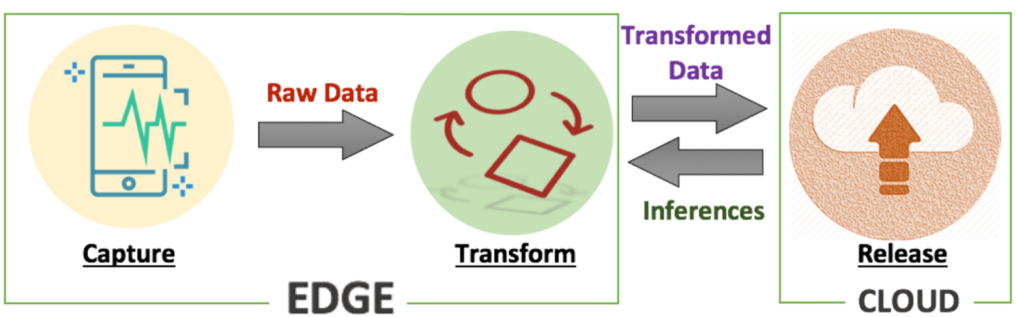


III. Hybrid

Good Privacy

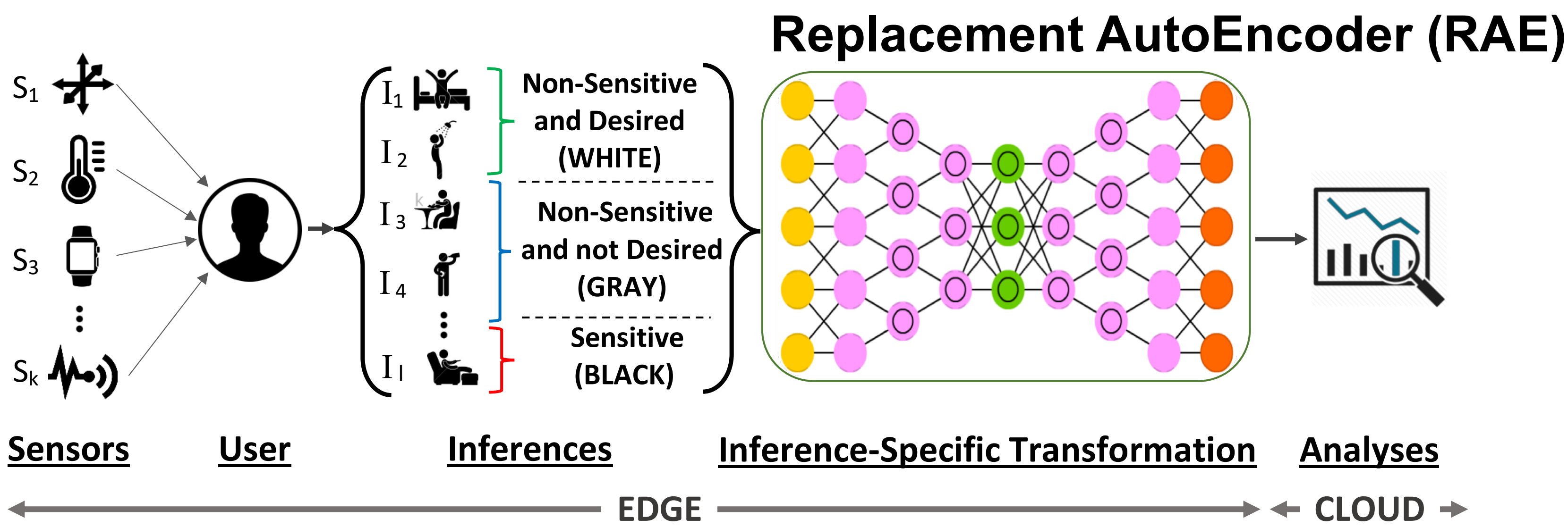
and

Good Utility



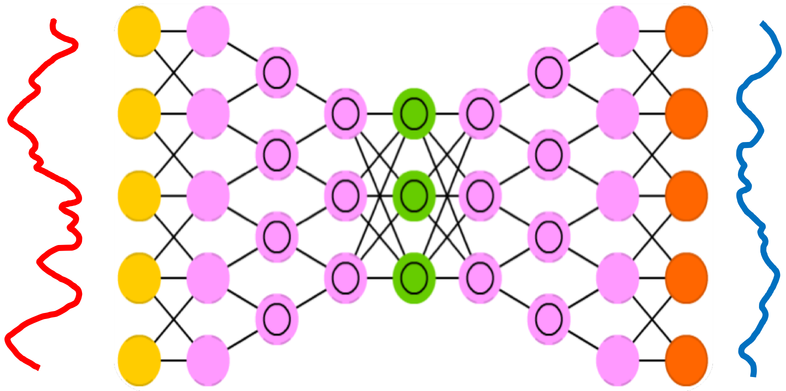
Imperial College London

The High-Level Architecture of the Solution



Imperial College
London

Why Autoencoders?



- To Discover the most salient features of the data.
  - A bottleneck: **Middle Layer Size << data dimension. e.g. : 50 << 1000**

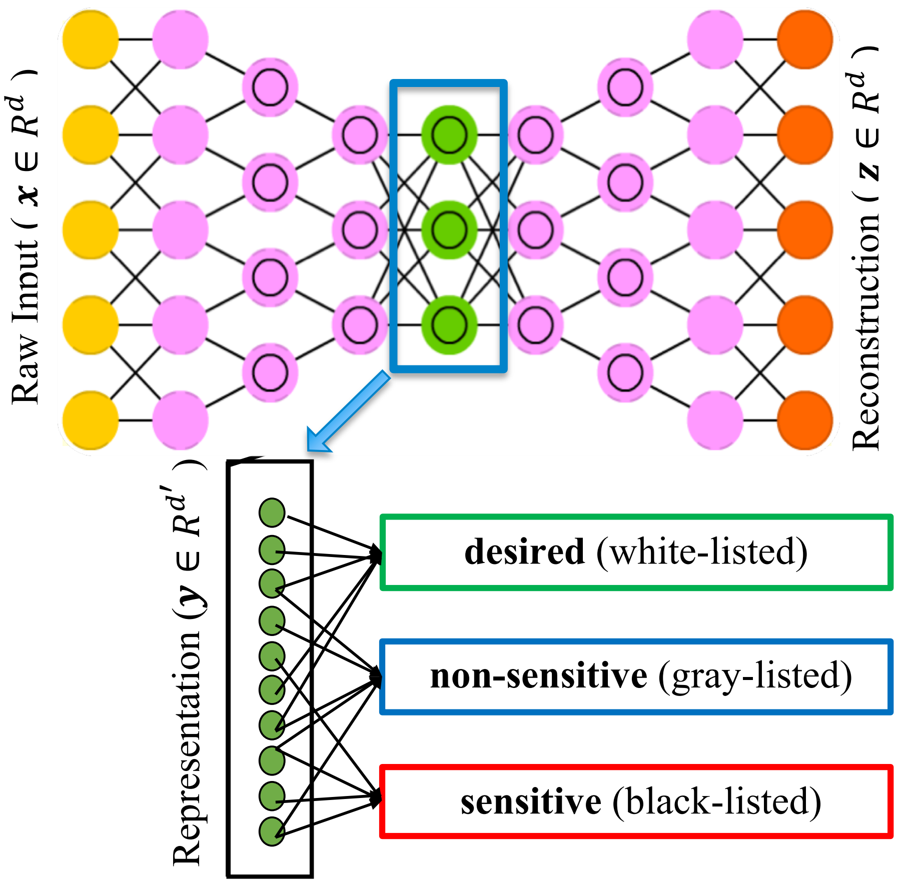
- Force to discover the **essence of the data** in order to **reconstruct** it.

Imperial College London

Key Idea: Remove Black Feature While Decoding Data

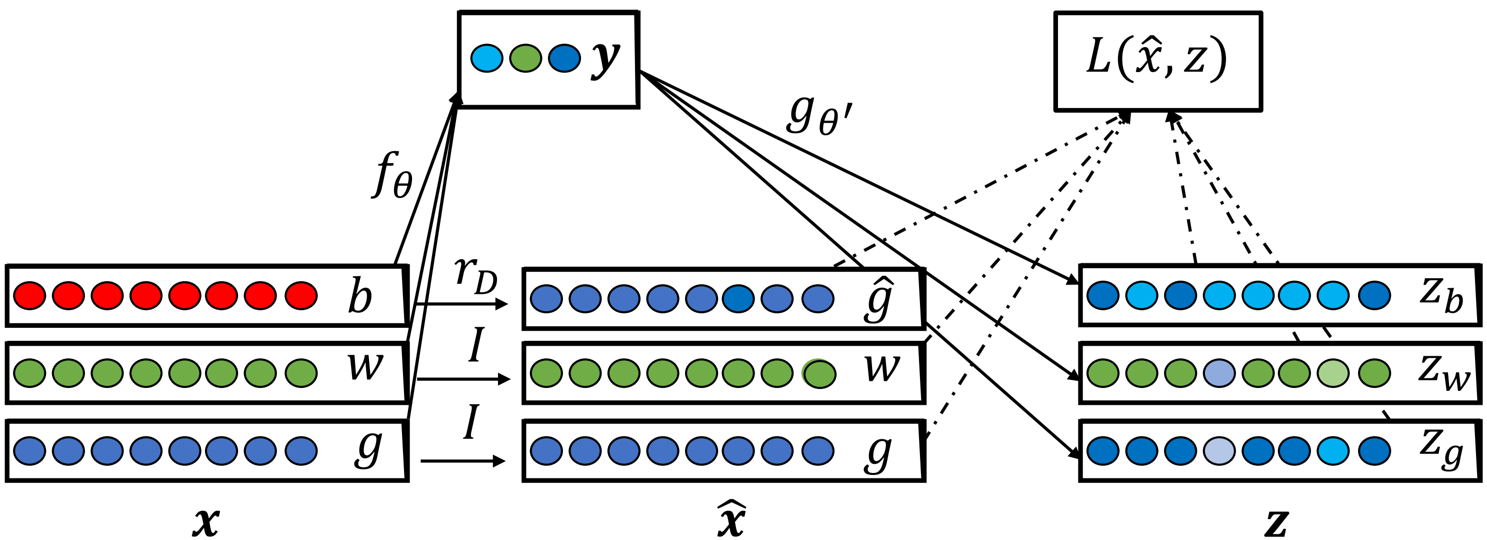
1. Encode raw data, x, into a new representation, y.

2. Decode y into a privatized version, z, using only ***white*** and ***grey*** features which have been learnt by the model.



Imperial College London

Training the RAE



$$\theta^*, {\theta'}^* = \arg_{\theta,\theta'} \min \frac{1}{n} \sum_{i=1}^{n} L(\hat{\boldsymbol{x}}^{(i)}, \boldsymbol{z}^{(i)})$$

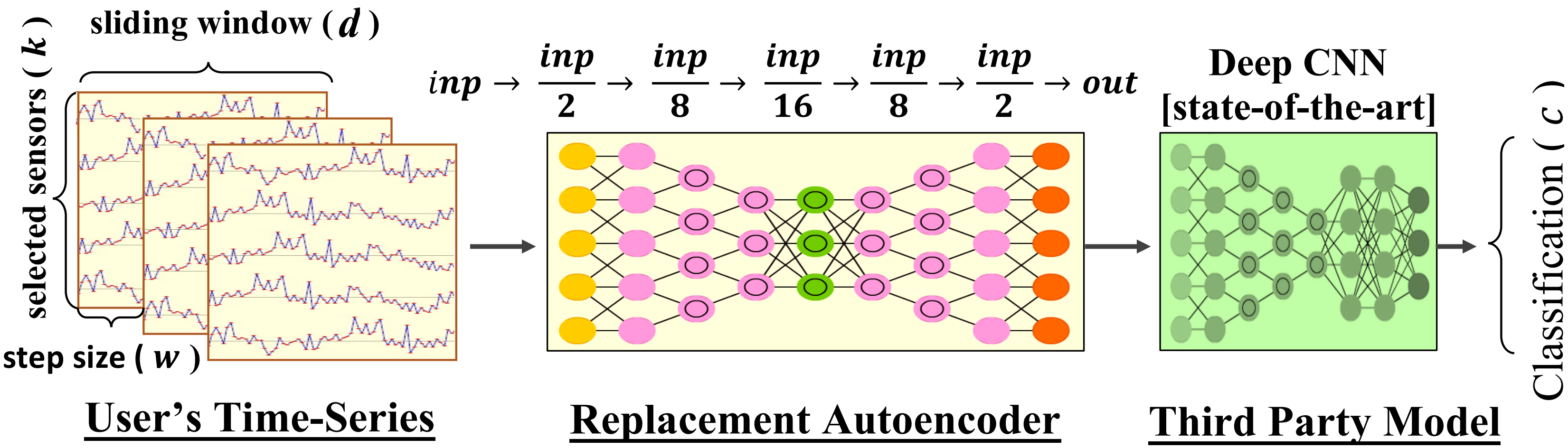
Imperial College London

A Real-World Case Study

- Activity Recognition
- Three datasets of Sensor Data generated by wearable devices.

	#	*Name of Dataset*		
		Opportunity	**Skoda**	**Hand-Gesture**
Activities	**0**	null	null	null
	1	open door1	write notes	open window
	2	open door2	open hood	close window
	3	close door1	close hood	water a plant
	4	close door2	check front door	turn book
	5	open fridge	open left f door	drink a bottle
	6	close fridge	close left f door	cut w/ knife
	7	open washer	close left doors	chop w/ knife
	8	close washer	check trunk	stir in a bowl
	9	open drawer1	open/close trunk	forehand
	10	close drawer1	check wheels	backhand
	11	open drawer2		smash
	12	close drawer2		
	13	open drawer3		
	14	close drawer3		
	15	clean table		
	16	drink cup		
	17	toggle switch		
Subjects		**4 people**	**1 person**	**2 people**
Sampling Rate		**30 Hz**	**98 Hz**	**32 Hz**
Dimension (d)		**113**	**60**	**15**

Experimental Setup



- RAE : A 7-layers Deep Autoencoder

- Activity Recognizer: A Deep Convolutional Autoencoder
  - We implemented the state-of-the-art for activity recognition using sensory data[1]

[1] J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition." in *IJCAI*, 2015, pp. 3995–4001.

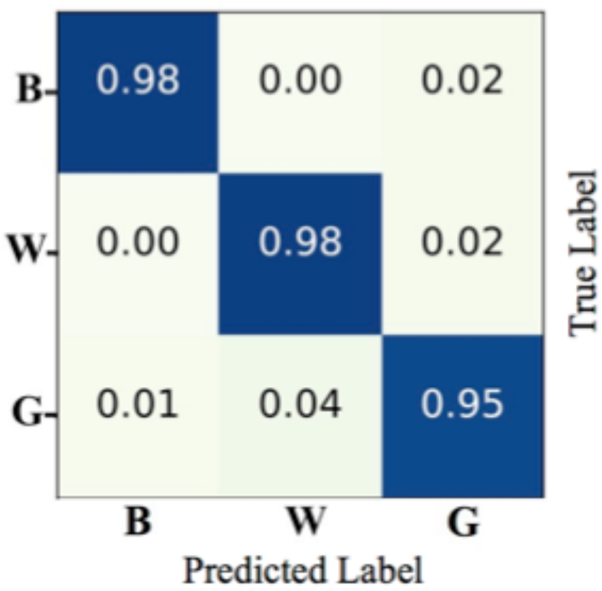
Imperial College London

Skoda Dataset : F1-Score for Activity Recognition

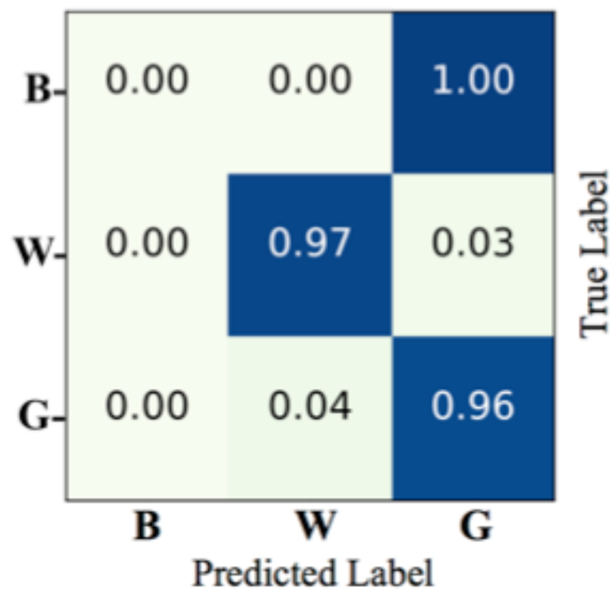
Hand	List of Inferences (Table I)	Original	Transformed	
Left	$I_w = \{4, 8, 9, 10\}$	97.92	96.32	White-Listed
	$I_b = \{1, 5, 6, 7\}$	**96.24**	**0.00**	Black-Listed
	$I_g = \{0, 2, 3\}$	94.34	93.42	Gray-Listed
Left	$I_w = \{2, 3, 5, 6, 7, 9\}$	96.52	93.23	White-Listed
	$I_b = \{4, 8, 10\}$	**97.88**	**0.00**	Black-Listed
	$I_g = \{0, 1\}$	93.86	94.85	Gray-Listed
Right	$I_w = \{1, 4, 10\}$	97.56	94.9	White-Listed
	$I_b = \{2, 3, 8, 9\}$	**97.97**	**0.00**	Black-Listed
	$I_g = \{0, 5, 6, 7\}$	92.33	88.23	Gray-Listed
Right	$I_w = \{2, 3, 5, 6, 7, 9\}$	95.76	91.06	White-Listed
	$I_b = \{4, 8, 10\}$	**97.39**	**0.00**	Black-Listed
	$I_g = \{0, 1\}$	94.31	92.39	Gray-Listed

Imperial College
London

Skoda Dataset : Confusion Matrix



Original Data



Transformed Data

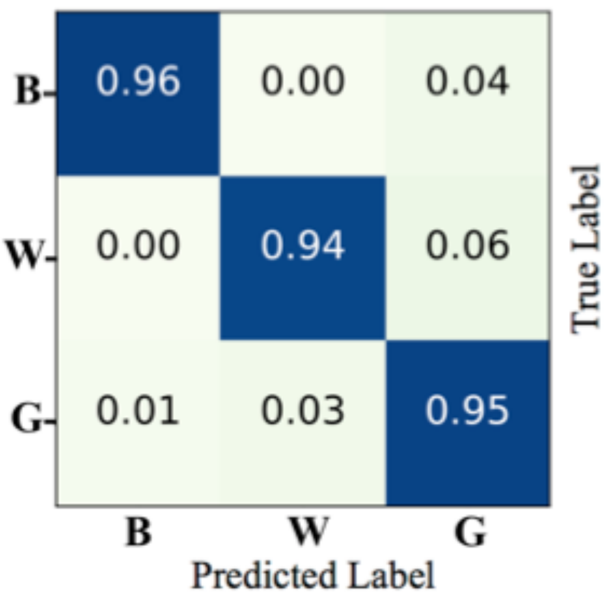
Imperial College London

Hand-Gesture Dataset : F1-Score

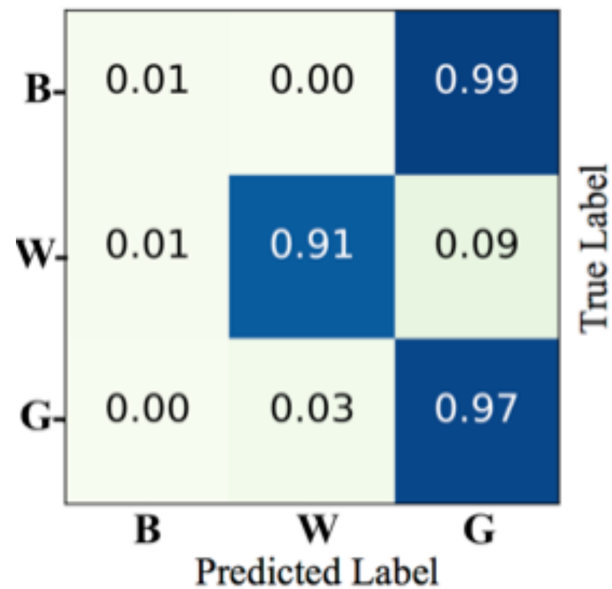
Subject	List of Inferences (Table I)	Original	Transformed	
#1	$I_w = \{1,2,3,4,9,10,11\}$	94.11	90.15	White-Listed
	$I_b = \{5,6,7,8\}$	**95.75**	**0.26**	Black-Listed
	$I_g = \{0\}$	95.04	96.54	Gray-Listed
#1	$I_w = \{1,3,4,5,6,7\}$	95.23	90.45	White-Listed
	$I_b = \{2,8,9,10,11\}$	**94.53**	**0.62**	Black-Listed
	$I_g = \{0\}$	95.04	97.46	Gray-Listed
#2	$I_w = \{1,3,4,5,6,7,8\}$	97.21	93.30	White-Listed
	$I_b = \{2,9,10,11\}$	**92.54**	**0.71**	Black-Listed
	$I_g = \{0\}$	95.89	97.53	Gray-Listed
#2	$I_w = \{2,3,5,6,7,9\}$	96.10	92.13	White-Listed
	$I_b = \{4,8,10\}$	**96.96**	**0.52**	Black-Listed
	$I_g = \{0,1\}$	95.70	97.56	Gray-Listed

Imperial College London

Hand-Gesture : Confusion Matrix



Original Data



Transformed Data

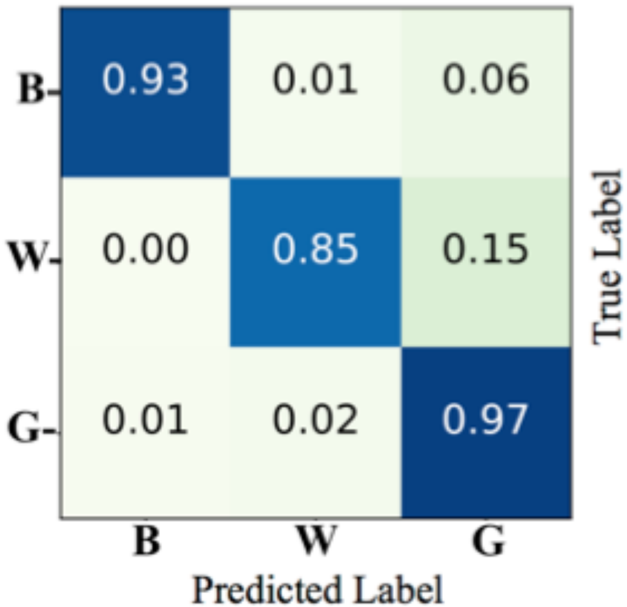
Imperial College London

Opportunity Dataset: F1-Score

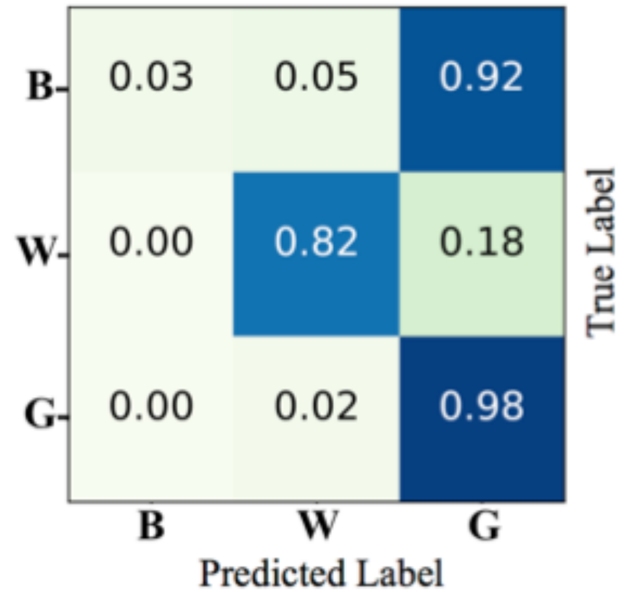
Subject	List of Inferences (Table I)	Original	Transformed	
#1	I_w={9,10,11,12,13,14,15,16,17}	71.75	64.32	White-Listed
	I_b={1,2,3,4,5,6,7,8}	**79.15**	**0.21**	Black-Listed
	I_g={0}	88.93	89.70	Gray-Listed
#1	I_w={1,2,3,4,5,6,7,8,15,17}	76.87	75.93	White-Listed
	I_b={9,10,11,12,13,14}	**71.49**	**1.32**	Black-Listed
	I_g={0,16}	84.44	82.08	Gray-Listed
#3	I_w={9,10,11,12,13,14,16}	74.92	77.07	White-Listed
	I_b={1,2,3,4,15,17}	**76.16**	**0.92**	Black-Listed
	I_g ={0,5,6,7,8}	84.98	81.58	Gray-Listed
#3	I_w={1,2,3,4,5,6,7,8,15,17}	70.32	65.05	White-Listed
	I_b={9,10,11,12,13,14,16}	**74.92**	**6.31**	Black-Listed
	I_g={0,1}	93.72	92.95	Gray-Listed

Imperial College
London

Opportunity : Confusion Matrix



Original Data

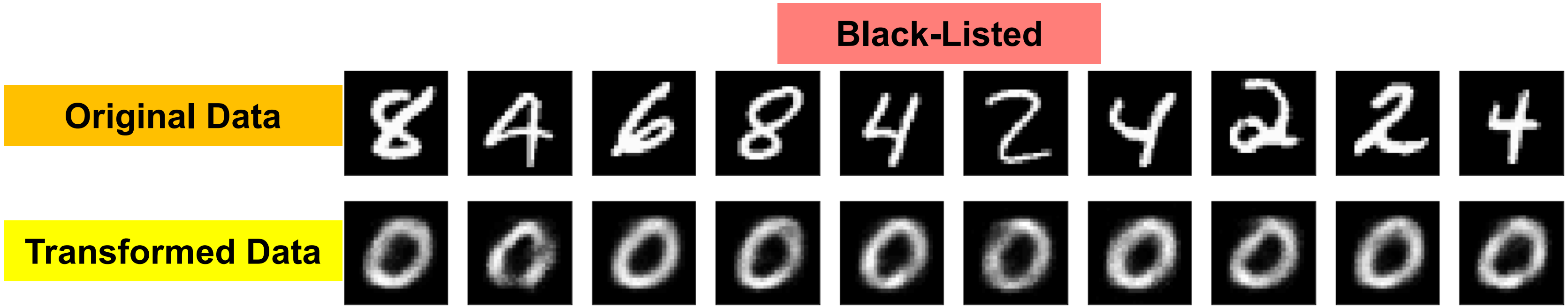


Transformed Data

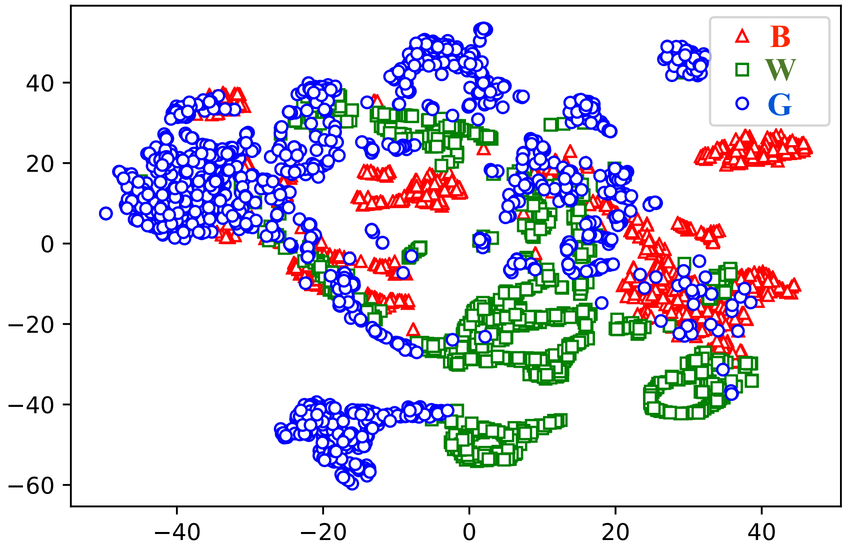
Imperial College London

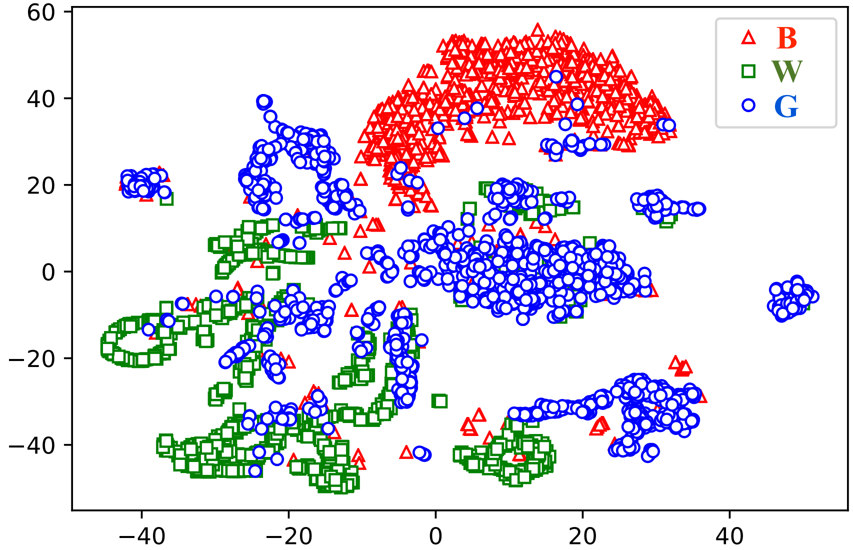
Visualization: An MNIST example

Digit 0 is a Gray-Listed data



t-Distributed Stochastic Neighbor Embedding t-SNE



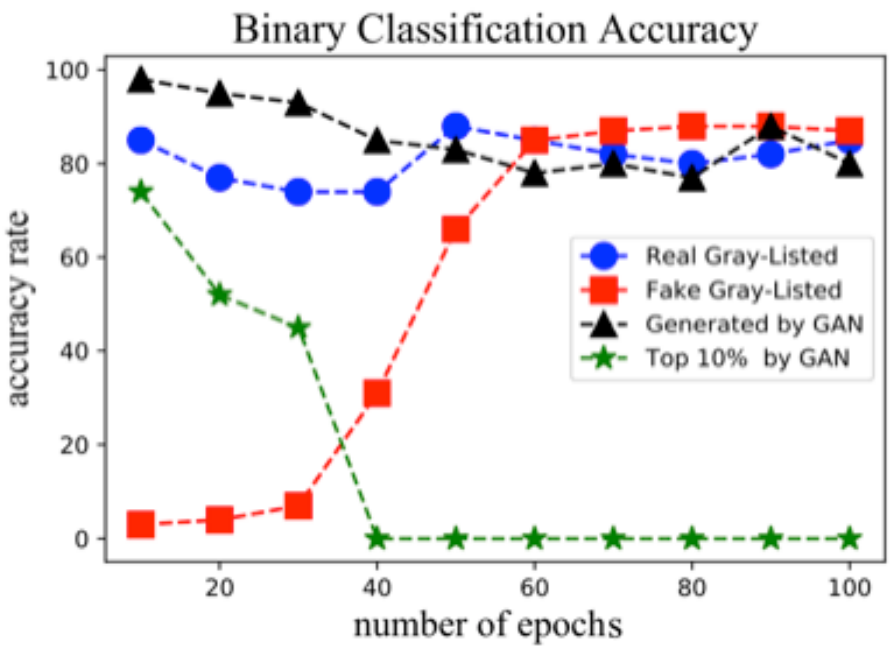


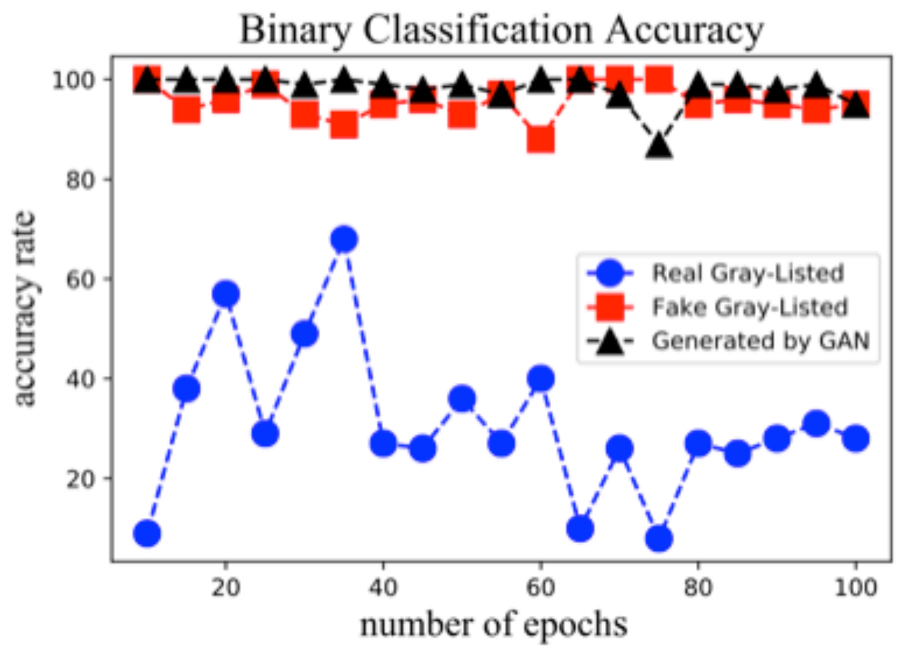
Original Data

Transformed Data

Imperial College London

Threat Model : Using GANs to detect Replaced intervals





When Adversaries have access To user Original Data

When Adversaries DON'T have access To user Original Data

Imperial College London

Conclusion

- RAE:
  - A hybrid architecture for locally transforming sensor data on edge devices.
- Inference-Specific Transformation:
  - Privacy-preserving data reconstruction based on learned features correspond to different inferences.

Take Home

- Encode data into a feature set, then replace sensitive information with non-sensitive not-desired features in the reconstruction (decoding) phase.

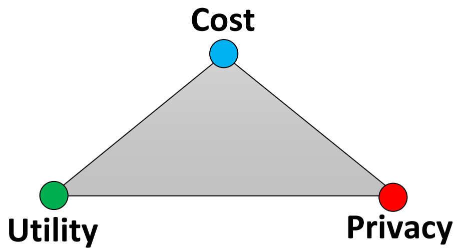
Imperial College London

Future Directions

- How we can provide a **statistical guarantee** (probabilistic bound) for sensitive information which can still be inferred from the transformed data?
  - Differential Privacy : **Composition Theorem?**
  - Mutual Information : **Joint Distributions?**

- Correlation among repeated measurement:
  - little by little information leakage

- What is the Complexity / Cost of the solution for running on Edge devices?



Imperial College London

Thank You!

Replacement AutoEncoder: A Privacy-Preserving Algorithm for Sensory Data Analysis

We are looking for postdocs and PhD students!

Hamed Haddadi

Imperial College London

https://haddadi.github.io/

h.haddadi@imperial.ac.uk

@realhamed