Universal Domain Adaptation Benchmark for Time Series Data Representation

EUSIPCO 2025

Romain Mussard, Fannia Pacheco, Maxime Berar, Gilles Gasso & Paul Honeine

Univ Rouen Normandie, INSA Rouen Normandie, Normandie Univ, LITIS UR 4108

Summary

Introduction

Time Series and Distribution Shift

Time Series

- **Univariate:** sequence of real-valued observations

$$\{x_t\}_{t=1}^{T}, \quad x_t \in \mathbb{R}$$

- **Multivariate:** multiple synchronized channels

$$\{\mathbf{x}_t\}_{t=1}^{T}, \quad \mathbf{x}_t \in \mathbb{R}^d$$

Time Series and Distribution Shift

Why Distribution Shift Matters

Deep models often assume training and test data follow the same distribution:

$$\mathcal{D}_{\text{train}} = \mathcal{D}_{\text{test}}$$

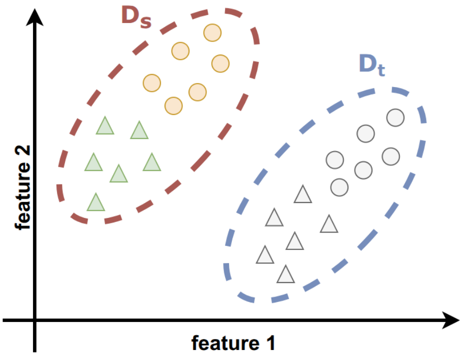
In practice, time series data frequently violates this assumption, leading to degraded performance.

Examples

- Sensor drift in biomedical signals (e.g., ECG).
- Seasonal or regime changes in financial time series.
- Different devices or acquisition protocols in speech signals.

The Challenges of Universal Domain Adaptation

Distribution Shift

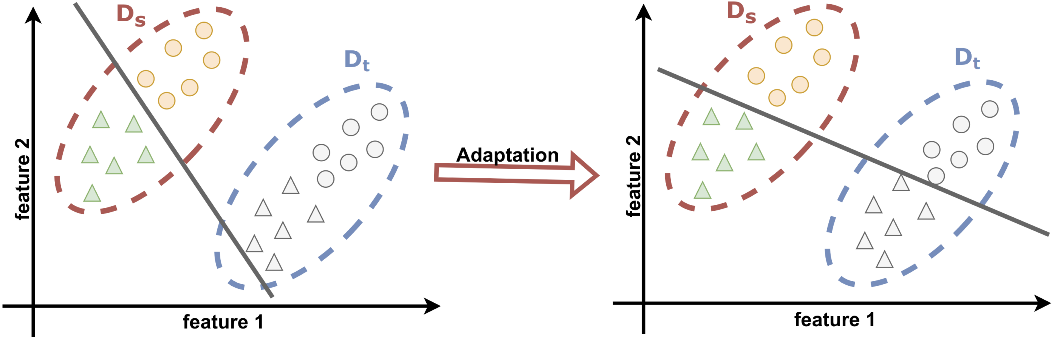


Covariate Shift

Datasets & Distributions

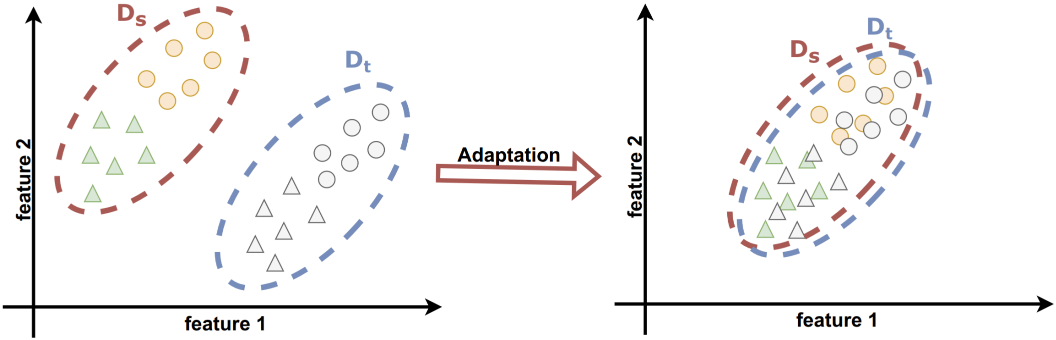
- Source Domain : $\mathcal{D}_s = \{x_i^s, y_i^s\}_{i=0}^{n_s} \sim P_s(x, y)$
- Target Domain : $\mathcal{D}_t = \{x_i^t\}_{i=0}^{n_t} \sim P_t(x, y)$
- $P_s(x) \neq P_t(x)$

(Unsupervised) Domain Adaptation



Domain Adaptation

(Unsupervised) Domain Adaptation



Domain Adaptation

(Unsupervised) Domain Adaptation in Practice

Model :

$$h(\cdot) = f \circ g(\cdot),$$

with :

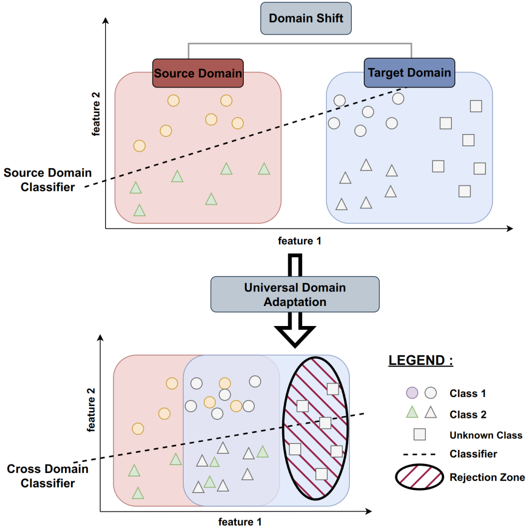
- $g(\cdot) : \mathcal{X} \rightarrow \mathcal{Z}$ a feature extractor
- $f(\cdot) : \mathcal{Z} \rightarrow \mathcal{Y}$ a classifier

Risk Minimization :

$$\min_{f,g} \frac{1}{n_s} \sum_{i=1}^{n_s} \mathcal{L}_{ce}(f(g(x_i^s)), y_i^s) + \lambda D(g(x_i^s), g(x_i^t))$$

- $\mathcal{L}_{ce}$ is the cross-entropy on source domain
- D align the distributions over the representation space $\mathcal{Z}$

Universal Domain Adaptation



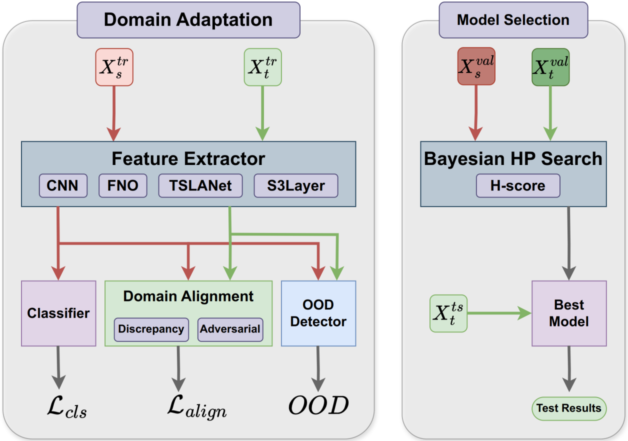
Objectives :

- **Alignment:** Align source and target common samples
- **OOD Discovery:** Detect target private samples (OOD samples)

No prior assumption is made regarding target set distribution !

UniDABench : Universal Domain Adaptation Benchmark

Benchmark



Benchmark Overview

Model Selection in UniDA

Algorithm 1 Model Selection Procedure

Require: Number of runs N_r, number of validation scenarios N_{val}, number of test scenarios N_{eval}

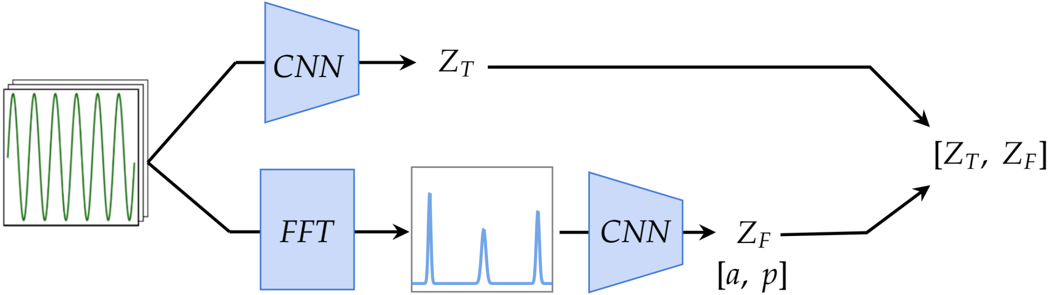
Require: Dataset $\mathbf{D} = \{\mathcal{D}^1, \mathcal{D}^2, \cdots \mathcal{D}^d\}$

1: $\mathcal{S}_{val}, \mathcal{S}_{eval} \leftarrow \text{select}(\mathbf{D}, N_{val}), \text{select}(\mathbf{D}, N_{eval})$
2: Initialize **scores** $\leftarrow [\,]$ *#Vector to store H-scores*
3: **for** $n \leftarrow 1$ to N_r **do**
4:     **for** each pair $\{\mathcal{D}^s, \mathcal{D}^t\} \in \mathcal{S}_{val}$ **do**
5:         Split $\mathcal{D}^t$ into $\{\mathcal{D}^t_{tr}, \mathcal{D}^t_{ts}\}$
6:         Train model using $(\mathcal{D}^s, \mathcal{D}^t_{tr})$
7:         $h \leftarrow \text{compute\_H\_score}(\mathcal{D}^t_{ts})$
8:         $\text{append}(\mathbf{scores}, h)$
9:     **end for**
10: **end for**
11: **return** $\arg\max_i \ \mathbf{scores}_i$ *#Select best models*

UniDA State-of-the-Art

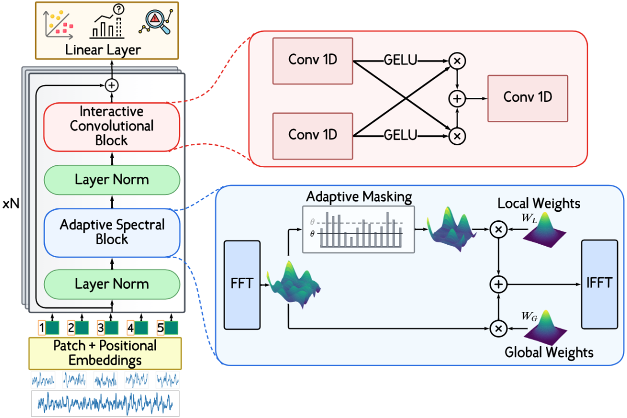
Method	Alignment	OOD Discovery	Threshold Based
UAN [1]	Adversarial Discriminator	Non Adversarial Discriminator	✓
DANCE [2]	Neighborhood Clustering	Entropy Separation	✓
OVANet [3]	Entropy	One-vs-All Rejection	✗
UniOT [4]	Neighborhood Clustering	Optimal Transport	✓
PPOT [5]	Prototypical Wasserstein	Softmax Thresholding	✓
UniJDOT [6]	Optimal Transport	Auto-thresholding	✗

Time Series Oriented Representation Layer



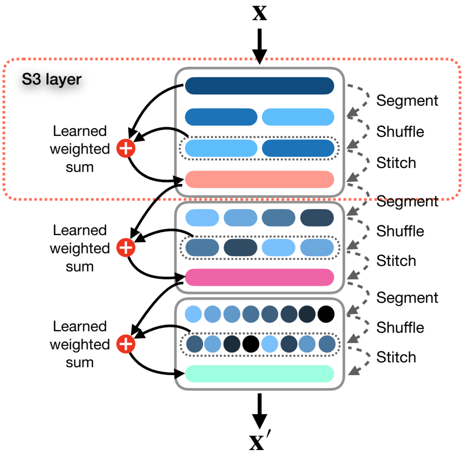
Time-Frequency Feature Extractor

Time Series Oriented Representation Layer



TSLANet: Rethinking Transformers for Time Series Representation Learning [7]

Time Series Oriented Representation Layer



Segment, Shuffle, and Stitch: A Simple Layer for Improving Time-Series Representations [8]

Time Series Datasets

Dataset	# Users/Domains	# Channels	# Classes	Sequence Length	Training set	Testing set
UCIHAR	30	9	6	128	2300	990
HHAR	9	3	6	128	12716	5218
EDF	20	1	5	3000	14280	6130

Experimental Results

Results

Methods	CNN	FNO	TSLANet	S3
UAN	60.1	**62.3**	32.3	52.2
OVANet	24.9	31.2	**32.0**	27.3
PPOT	37.6	**62.1**	6.9	35.9
DANCE	53.0	**54.1**	2.0	46.7
UniOT	47.7	49.1	**53.5**	37.6
UniJDOT	61.0	**64.6**	59.7	54.1

H-score (%) for HAR dataset

Methods	CNN	FNO	TSLANet	S3
UAN	54.2	**57.6**	30.2	52.1
OVANet	**55.4**	42.2	23.5	52.7
PPOT	**45.0**	36.9	3.8	36.8
DANCE	**51.9**	41.4	19.2	50.6
UniOT	**41.4**	41.2	32.3	37.7
UniJDOT	44.3	**55.6**	55.0	50.4

H-score (%) for EDF dataset

Methods	CNN	FNO	TSLANet	S3
UAN	**47.8**	41.1	34.0	41.9
OVANet	27.0	**43.7**	28.3	23.3
PPOT	40.6	**48.6**	2.1	39.4
DANCE	40.6	41.8	0.0	**42.4**
UniOT	51.0	**54.3**	39.9	44.6
UniJDOT	56.6	**61.2**	55.7	57.8

H-score (%) for HHAR dataset

$$\textbf{H-score} = \frac{2A_cA_u}{A_c + A_u}$$

A_c is the accuracy of known classes.

A_u is the accuracy of target unknown classes.

Results

H-scores (%) for HAR

Scenario	UAN*	OVANet†	PPOT*	DANCE*	UniOT†	UniJDOT*
$12 \rightarrow 16$	56.8	41.8	**61.0**	49.9	48.9	50.6
$13 \rightarrow 3$	67.3	16.4	**78.0**	71.4	64.4	76.3
$15 \rightarrow 21$	78.4	53.7	36.8	66.6	53.2	**81.5**
$17 \rightarrow 29$	63.5	32.0	72.8	65.3	52.9	**78.3**
$1 \rightarrow 14$	63.8	33.5	62.8	**64.8**	58.2	39.2
$22 \rightarrow 4$	63.0	36.7	68.4	41.2	54.0	**72.4**
$24 \rightarrow 8$	50.7	40.8	**58.2**	38.6	50.7	47.0
$30 \rightarrow 20$	47.2	9.2	**54.0**	49.4	51.9	53.4
$6 \rightarrow 23$	66.6	12.4	75.8	68.5	52.8	**78.7**
$9 \rightarrow 18$	65.4	43.8	53.2	25.4	48.5	**68.6**
Mean	62.3	32.0	62.1	54.1	53.5	**64.6**

* Models trained with FNO † Models trained with TSLANet

Results CNN

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
HAR 12 → 16	**56.9**	20.5	26.2	39.4	22.2	45.8
HAR 13 → 3	72.3	39.7	57.8	**87.3**	61.4	77.6
HAR 15 → 21	76.3	20.6	39.6	**87.5**	61.8	79.0
HAR 17 → 29	68.2	16.5	47.6	**74.6**	48.9	64.8
HAR 1 → 14	**80.0**	17.2	24.5	19.2	55.8	46.0
HAR 22 → 4	73.4	40.2	46.7	**82.3**	52.2	70.4
HAR 24 → 8	42.7	21.3	27.0	**58.0**	44.1	54.2
HAR 30 → 20	37.7	15.4	27.2	32.9	35.1	**45.2**
HAR 6 → 23	24.5	11.6	34.2	00.0	50.5	**68.8**
HAR 9 → 18	**69.0**	46.3	45.2	49.0	45.3	58.3
HAR mean	60.1	24.9	37.6	53.0	47.7	**61.0**

H-score (%) for HAR dataset

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
HHAR 0 → 2	**46.8**	31.1	30.6	36.5	42.0	41.3
HHAR 0 → 6	41.4	34.0	31.8	49.8	61.6	**64.2**
HHAR 1 → 6	53.0	29.1	67.4	40.5	63.0	**78.3**
HHAR 2 → 7	36.2	13.6	12.3	**42.1**	15.8	14.7
HHAR 3 → 8	53.8	27.3	48.8	49.2	70.6	**72.3**
HHAR 4 → 5	62.2	37.4	35.6	25.6	52.9	**67.5**
HHAR 5 → 0	**13.1**	07.4	02.2	**12.7**	11.0	00.0
HHAR 6 → 1	73.4	20.4	70.0	57.4	81.3	**87.2**
HHAR 7 → 4	50.1	38.5	61.9	24.6	62.7	**74.3**
HHAR 8 → 3	48.1	31.5	44.9	**67.3**	49.5	**66.6**
HHAR mean	47.8	27.0	40.6	40.6	51.0	**56.6**

H-score (%) for HHAR dataset

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
EEG 0 → 11	39.9	18.6	17.3	**44.6**	33.1	25.2
EEG 12 → 5	58.2	**67.0**	45.5	57.2	42.9	52.6
EEG 13 → 17	32.3	41.1	**44.1**	42.5	41.4	33.8
EEG 16 → 1	56.4	**59.2**	53.7	48.3	43.6	54.6
EEG 18 → 12	**44.6**	43.2	32.9	39.2	43.1	35.4
EEG 3 → 19	55.6	**59.9**	41.8	43.0	40.8	40.2
EEG 5 → 15	61.6	**64.3**	59.2	55.7	33.0	43.4
EEG 6 → 2	57.5	**59.2**	36.1	57.8	53.3	43.8
EEG 7 → 18	**66.8**	61.2	61.7	63.2	45.7	58.9
EEG 9 → 14	69.6	**80.3**	57.5	67.2	37.3	55.2
EEG mean	54.2	**55.4**	45.0	51.9	41.4	44.3

H-score (%) for EDF dataset

$$\textbf{H-score} = \frac{2A_cA_u}{A_c + A_u}$$

A_c is the accuracy of known classes.

A_u is the accuracy of target unknown classes.

Results FNO

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
HAR 12 → 16	56.8	29.2	**61.0**	49.9	36.5	50.6
HAR 13 → 3	67.3	45.5	**78.0**	71.4	51.9	76.3
HAR 15 → 21	78.4	14.5	36.8	66.6	56.4	**81.5**
HAR 17 → 29	63.5	20.7	72.8	65.3	53.0	**78.3**
HAR 1 → 14	63.8	28.4	62.8	**64.8**	44.8	39.2
HAR 22 → 4	63.0	64.0	68.4	41.2	57.4	**72.4**
HAR 24 → 8	50.7	24.0	**58.2**	38.6	47.8	47.0
HAR 30 → 20	47.2	24.8	**54.0**	49.4	45.2	53.4
HAR 6 → 23	66.6	26.4	75.8	68.5	50.6	**78.7**
HAR 9 → 18	65.4	34.5	53.2	25.4	47.5	**68.6**
HAR mean	62.3	31.2	62.1	54.1	49.1	**64.6**

H-score (%) for HAR dataset

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
HHAR 0 → 2	41.8	**55.4**	28.0	37.8	42.4	45.4
HHAR 0 → 6	41.6	**59.8**	36.6	32.8	52.5	**59.7**
HHAR 1 → 6	41.8	51.2	71.2	44.0	71.5	**78.7**
HHAR 2 → 7	23.1	23.4	08.5	**27.6**	16.2	25.7
HHAR 3 → 8	44.1	69.8	65.9	62.9	71.1	**77.9**
HHAR 4 → 5	45.3	27.2	54.2	50.4	58.3	**77.0**
HHAR 5 → 0	**24.5**	16.2	07.9	**23.8**	20.9	12.9
HHAR 6 → 1	55.0	27.1	63.1	56.4	81.5	**86.0**
HHAR 7 → 4	43.0	39.0	**77.9**	38.8	71.5	76.7
HHAR 8 → 3	50.6	67.4	**73.2**	43.3	57.2	71.8
HHAR mean	41.1	43.7	48.6	41.8	54.3	**61.2**

H-score (%) for HHAR dataset

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
EEG 0 → 11	**39.9**	25.5	21.3	33.9	34.8	38.1
EEG 12 → 5	58.5	46.1	60.0	39.5	39.1	**65.2**
EEG 13 → 17	42.3	**47.7**	15.5	36.6	43.4	42.9
EEG 16 → 1	**66.7**	45.6	54.3	45.5	44.6	60.5
EEG 18 → 12	**49.3**	39.7	16.9	35.7	41.0	42.8
EEG 3 → 19	**65.2**	40.4	36.6	42.6	40.3	54.8
EEG 5 → 15	45.4	38.7	47.6	32.8	32.0	**67.7**
EEG 6 → 2	**62.3**	50.9	9.9	49.0	50.9	49.9
EEG 7 → 18	**66.6**	41.4	58.8	50.9	45.3	64.3
EEG 9 → 14	**80.4**	46.3	47.7	47.6	40.9	69.2
EEG mean	**57.6**	42.2	36.9	41.4	41.2	55.6

H-score (%) for EDF dataset

$$\textbf{H-score} = \frac{2A_cA_u}{A_c + A_u}$$

A_c is the accuracy of known classes.
A_u is the accuracy of target unknown classes.

Results TSLANet

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
HAR 12 $\to$ 16	30.5	41.8	01.3	00.0	**48.9**	32.6
HAR 13 $\to$ 3	43.9	16.4	12.3	02.2	64.4	**73.5**
HAR 15 $\to$ 21	28.9	53.7	06.7	02.7	53.2	**83.1**
HAR 17 $\to$ 29	43.2	32.0	04.8	06.1	52.9	**82.7**
HAR 1 $\to$ 14	48.1	33.5	01.3	00.0	**58.2**	55.7
HAR 22 $\to$ 4	34.0	36.7	02.7	00.0	54.0	**83.1**
HAR 24 $\to$ 8	21.8	40.8	09.0	00.0	50.7	**66.2**
HAR 30 $\to$ 20	30.3	09.2	06.0	00.0	**51.9**	00.0
HAR 6 $\to$ 23	23.7	12.4	01.2	02.3	52.8	**61.1**
HAR 9 $\to$ 18	19.0	43.8	23.5	06.3	48.5	**58.9**
HAR mean	32.3	32.0	06.9	02.0	53.5	**59.7**

H-score (%) for HAR dataset

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
HHAR 0 $\to$ 2	39.8	18.8	01.2	00.0	42.4	**67.5**
HHAR 0 $\to$ 6	23.6	**43.7**	03.4	00.0	34.3	37.0
HHAR 1 $\to$ 6	48.2	28.9	01.2	00.0	51.5	**75.2**
HHAR 2 $\to$ 7	20.4	10.4	00.5	00.0	08.5	**22.7**
HHAR 3 $\to$ 8	48.8	70.0	02.2	00.0	70.0	**77.1**
HHAR 4 $\to$ 5	30.2	24.9	01.2	00.0	37.1	**67.4**
HHAR 5 $\to$ 0	**04.3**	02.5	00.2	00.0	**03.7**	02.0
HHAR 6 $\to$ 1	43.3	17.1	02.6	00.2	60.4	**67.8**
HHAR 7 $\to$ 4	32.3	33.5	05.3	00.0	39.0	**61.4**
HHAR 8 $\to$ 3	49.2	32.8	02.8	00.0	52.0	**79.1**
HHAR mean	34.0	28.3	02.1	00.0	39.9	**55.7**

H-score (%) for HHAR dataset

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
EEG 0 $\to$ 11	36.0	25.3	00.7	02.9	37.8	**53.2**
EEG 12 $\to$ 5	36.1	18.4	01.0	32.0	41.5	**64.3**
EEG 13 $\to$ 17	15.2	01.5	04.6	05.5	**17.9**	15.8
EEG 16 $\to$ 1	27.8	05.9	01.4	20.3	04.7	**49.1**
EEG 18 $\to$ 12	27.6	29.1	10.2	18.5	9.8	**39.0**
EEG 3 $\to$ 19	34.8	28.4	01.7	12.8	39.5	**73.4**
EEG 5 $\to$ 15	28.2	13.9	03.0	08.8	66.0	**76.4**
EEG 6 $\to$ 2	29.0	**55.7**	08.5	36.8	32.2	**55.8**
EEG 7 $\to$ 18	31.7	08.8	00.8	12.0	36.0	**61.9**
EEG 9 $\to$ 14	35.9	48.2	05.7	42.6	37.9	**61.4**
EEG mean	30.2	23.5	03.8	19.2	32.3	**55.0**

H-score (%) for EDF dataset

$$\textbf{H-score} = \frac{2A_cA_u}{A_c + A_u}$$

A_c is the accuracy of known classes.

A_u is the accuracy of target unknown classes.

Results S3Layer

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
HAR 12 → 16	**43.6**	26.0	30.2	<u>40.3</u>	21.8	<u>39.8</u>
HAR 13 → 3	<u>63.0</u>	30.0	42.1	**68.4**	34.0	60.6
HAR 15 → 21	37.2	35.3	36.8	34.9	**46.1**	<u>38.1</u>
HAR 17 → 29	**63.9**	30.5	56.4	57.3	36.4	<u>61.0</u>
HAR 1 → 14	**56.9**	28.2	9.8	46.2	39.0	<u>52.6</u>
HAR 22 → 4	51.8	18.5	32.4	**53.5**	41.8	**53.8**
HAR 24 → 8	<u>41.5</u>	21.1	33.5	28.8	34.5	**51.4**
HAR 30 → 20	<u>53.7</u>	25.1	42.3	50.9	33.6	**58.1**
HAR 6 → 23	51.0	22.6	43.2	43.4	<u>54.1</u>	**69.0**
HAR 9 → 18	**59.0**	35.8	32.1	43.1	35.0	<u>56.7</u>
HAR mean	<u>52.2</u>	27.3	35.9	46.7	37.6	**54.1**

H-score (%) for HAR dataset

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
HHAR 0 → 2	<u>42.0</u>	11.1	28.0	40.5	**43.2**	39.0
HHAR 0 → 6	39.3	<u>65.0</u>	37.7	51.0	47.8	**65.9**
HHAR 1 → 6	50.6	16.2	<u>62.9</u>	36.9	51.3	**74.7**
HHAR 2 → 7	27.7	12.5	<u>33.3</u>	**42.6**	15.7	25.1
HHAR 3 → 8	58.9	44.3	41.1	54.0	<u>63.5</u>	**77.7**
HHAR 4 → 5	43.8	06.4	33.3	25.1	<u>53.7</u>	**61.0**
HHAR 5 → 0	<u>15.6</u>	02.6	00.7	**16.9**	13.1	10.3
HHAR 6 → 1	49.0	04.7	<u>63.2</u>	55.9	56.7	**82.5**
HHAR 7 → 4	43.9	11.3	<u>62.0</u>	31.1	52.8	**74.9**
HHAR 8 → 3	48.6	59.1	31.4	**69.5**	47.9	<u>67.1</u>
HHAR mean	41.9	23.3	39.4	42.4	<u>44.6</u>	**57.8**

H-score (%) for HHAR dataset

Scenario	UDA	OVANet	PPOT	DANCE	UniOT	UniJDOT
EEG 0 → 11	36.4	20.1	22.6	<u>40.0</u>	27.7	**46.0**
EEG 12 → 5	**64.9**	57.5	35.4	57.0	40.9	<u>60.6</u>
EEG 13 → 17	34.8	**49.6**	17.1	<u>41.8</u>	36.2	36.1
EEG 16 → 1	**57.8**	51.4	50.9	45.3	40.2	<u>55.4</u>
EEG 18 → 12	36.1	**42.0**	25.7	<u>39.5</u>	14.5	37.1
EEG 3 → 19	**54.1**	<u>50.8</u>	31.5	42.1	37.3	48.9
EEG 5 → 15	50.0	**63.1**	54.8	<u>57.2</u>	50.1	50.5
EEG 6 → 2	56.8	**59.9**	22.9	<u>58.5</u>	24.9	53.1
EEG 7 → 18	**62.0**	54.9	50.3	**61.6**	50.6	60.8
EEG 9 → 14	<u>68.3</u>	**77.5**	56.7	62.7	54.2	55.9
EEG mean	<u>52.1</u>	**52.7**	36.8	50.6	37.7	50.4

H-score (%) for EDF dataset

$$\textbf{H-score} = \frac{2A_cA_u}{A_c + A_u}$$

A_c is the accuracy of known classes.

A_u is the accuracy of target unknown classes.

Conclusion

Conclusion

Wrap-Up:

- We proposed a framework for UniDA model training and testing
- We studied time-series–oriented architectures in UniDA
- Our study showed that recent architectures may not be well-suited for UniDA

Future Work:

- Focus on time series: How can we learn good representations of such data?
- Explore more methods and datasets
- Investigate whether foundation models improve performance

Thank You!

romain.mussard@univ-rouen.fr

Questions?

References i

[1] K. You, M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Universal domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[2] K. Saito, D. Kim, S. Sclaroff, and K. Saenko, "Universal domain adaptation through self-supervision," 2020.

[3] K. Saito and K. Saenko, "Ovanet: One-vs-all network for universal domain adaptation," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8980–8989, 2021.

[4] W. Chang, Y. Shi, H. Tuan, and J. Wang, "Unified optimal transport framework for universal domain adaptation," in *Advances in Neural Information Processing Systems* (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds.), vol. 35, pp. 29512–29524, Curran Associates, Inc., 2022.

[5] Y. Yang, X. Gu, and J. Sun, "Prototypical partial optimal transport for universal domain adaptation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 10852–10860, Jun. 2023.

[6] R. Mussard, F. Pacheco, M. Berar, G. Gasso, and P. Honeine, "Deep joint distribution optimal transport for universal domain adaptation on time series," *IEEE International Joint Conference on Neural Networks*, 2025.

[7] E. Eldele, M. Ragab, Z. Chen, M. Wu, and X. Li, "Tslanet: Rethinking transformers for time series representation learning," in *International Conference on Machine Learning*, 2024.

[8] S. Grover, A. Jalali, and A. Etemad, "Segment, shuffle, and stitch: A simple layer for improving time-series representations," *Advances in Neural Information Processing Systems*, vol. 37, pp. 4878–4905, 2025.