Appendix A. Preliminary Experiments on 5 datasets

Preliminary experiments were conducted on five datasets (more information about the dataset can be found in Sec. C), with results reported in terms of accuracy and F1 score below:

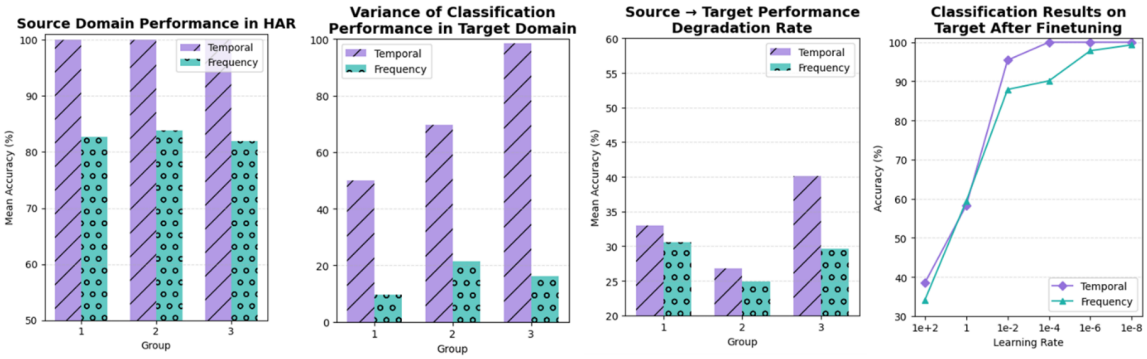


Figure 5: HAR Dataset (Accuracy)

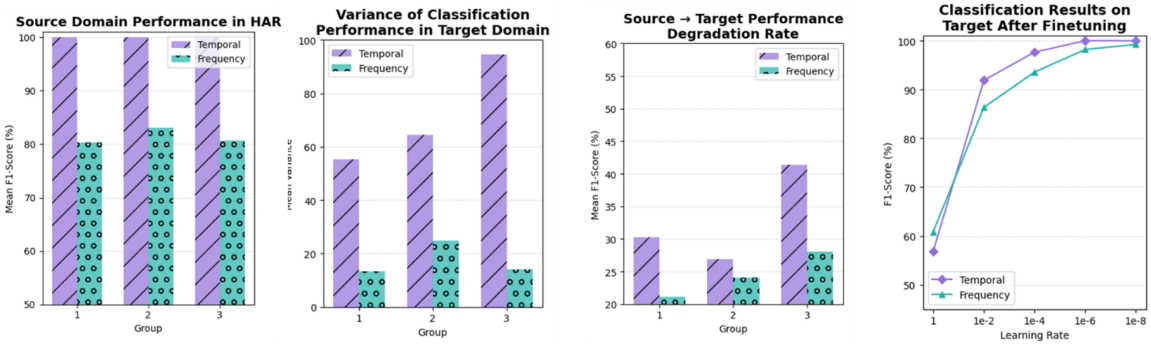


Figure 6: HAR Dataset (F1 Score)

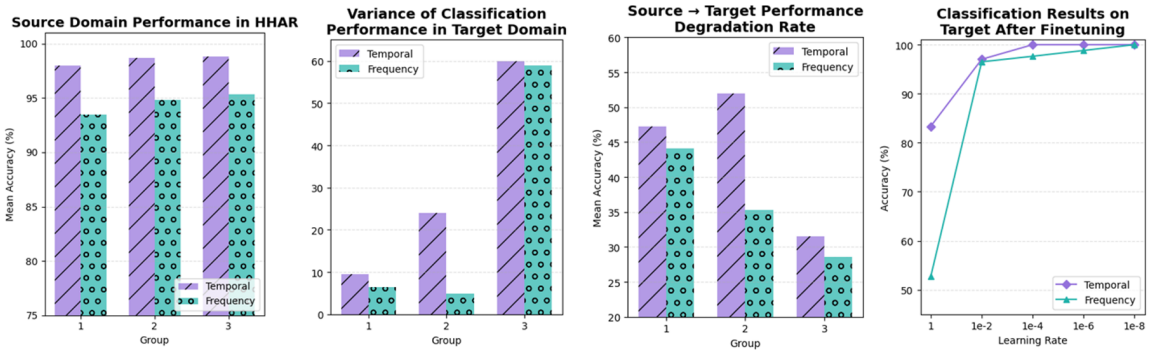


Figure 7: HHAR Dataset (Accuracy)

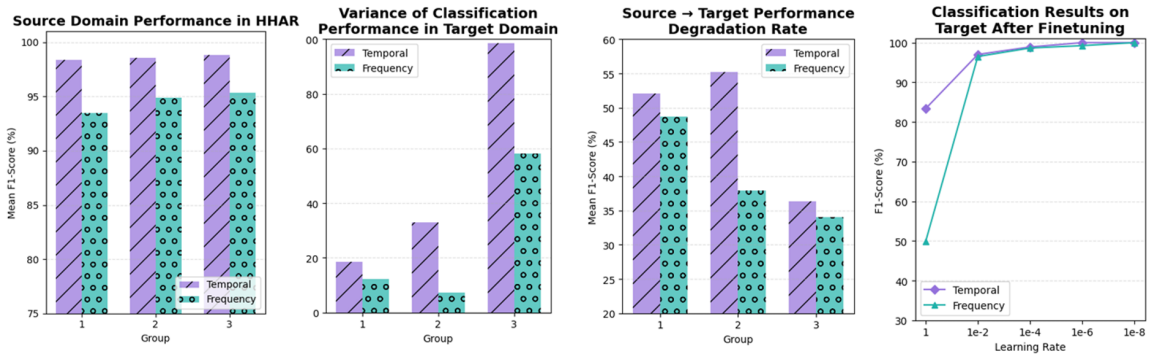


Figure 8: HHAR Dataset (F1 Score)

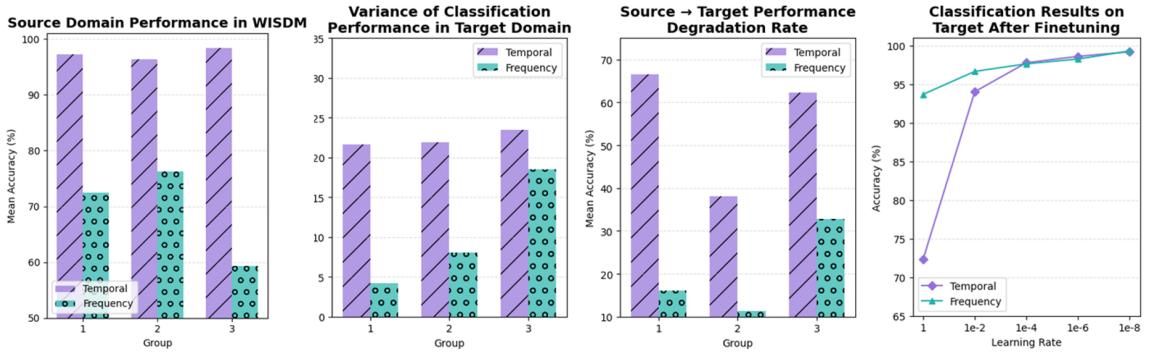


Figure 9: WISDM Dataset (Accuracy)

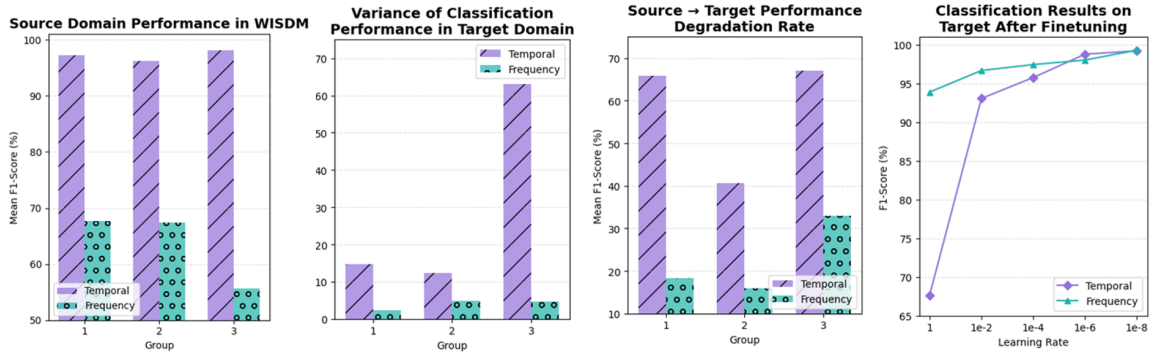


Figure 10: WISDM Dataset (F1 Score)

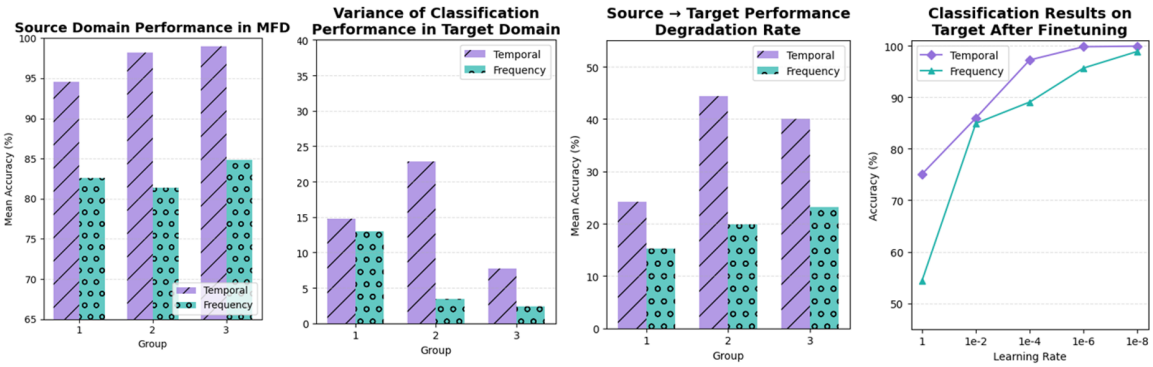


Figure 11: MFD Dataset (Accuracy)

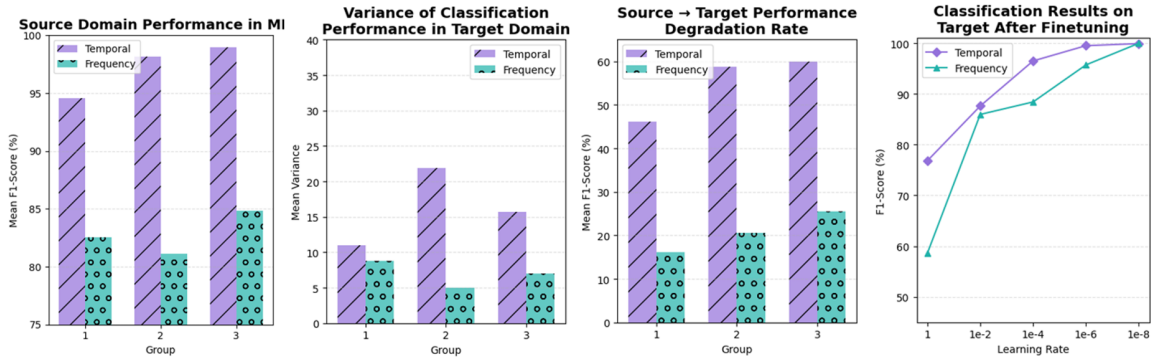


Figure 12: MFD Dataset (F1 Score)

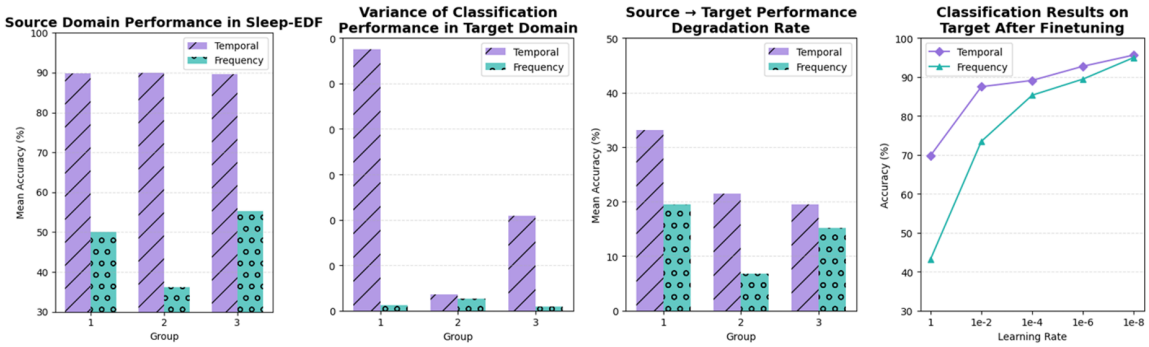


Figure 13: Sleep-EDF Dataset (Accuracy)

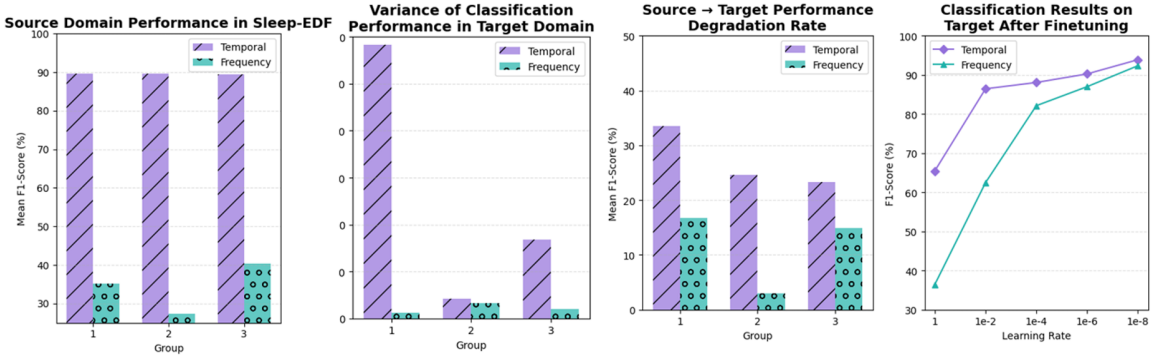


Figure 14: Sleep-EDF Dataset (F1 Score)

Appendix B. Algorithms

An overview of TF-DAN is in Alg. 1. Moreover, we enhance the nearest neighbor algorithm of VQ-VAE to make it suitable for our UDA task. We utilize nearest neighbor function ρ (Alg. 2) in both the training and adaptation phases, while voting function V (Alg. 3) is applied during the inference stage. Unlike Alg. 1, we want to illustrate a more comprehensive explanation of implementation details, so both of these algorithms are implemented following the PyTorch style.

Algorithm 1 Overview of TF-DAN

Input: data x_i; label y_i^s; Dual-stream encoder E; decoder U; classifier C; frequency block B_S; temporal block B_T; time step T; input channel M; nearest neighbor function ρ (Alg. 2); voting function V (Alg. 3)

Extract $z_e \leftarrow E(x_i)$

First: Training Phase

Get $e_{h,j}, e_{p,q} \leftarrow \rho(z_e, [B_S; B_T], y_i^s)$

$x_i' \leftarrow U(e_{h,j})$

Compute objective functions $\mathcal{L}_{CE}$, $\mathcal{L}_A$ and $\mathcal{L}_D$

Update E, B_S, B_T and C with

$$\nabla(\mathcal{L}_{CE} + \mathcal{L}_A + \mathcal{L}_D)$$

Second: Adaptation Phase

Get $e_{p,q} \leftarrow \rho(z_e, [B_S; B_T])$

$x_i' \leftarrow U(e_{p,q})$

Compute objective functions $\mathcal{L}_{MSE}$ and $\mathcal{L}_A$

Update E, B_S, B_T and U with $\nabla(\mathcal{L}_{MSE} + \mathcal{L}_A)$

Third: Inference Phase

Get $e_{p',q'} \leftarrow V(z_e)$

Output: p'

Algorithm 2 Finding Nearest Neighbor Function ρ

Input: Query Q, Target T, Labels $label$
Initialization:
$index\_list \leftarrow []$
$k \leftarrow$ Total embeddings for each category
$h \leftarrow$ Total classification categories
$Q \leftarrow Q$.unsqueeze(1).repeat$(1, k, 1)$
for $i = 1$ **to** Q.size(0) **do**
  $T \leftarrow T[label[i] \times k : (label[i] + 1) \times k]$.unsqueeze(0)
  $tmp\_index \leftarrow (Q[i] - T)$.pow(2).sum(2).sqrt().min(1)[1][0]
  $index \leftarrow \text{int}(tmp\_index) + label[i] \times k$
  $index\_list$.append$(index)$
end for
$index\_tensor \leftarrow$ torch.tensor$(index_list)$
$e_{h,j} \leftarrow T[index\_tensor]$
if During Training Phase **then**
  {*Find the nearest neighbor from other categories.*}
  $index\_list \leftarrow []$
  $Q \leftarrow Q$.unsqueeze(1).repeat$(1, k \times (h-1), 1)$
  for $i = 1$ **to** Q.size(0) **do**
    $map\_original\_list \leftarrow$ list(range$(k \times h)$)
    del map_original_list$[label[i] \times k : (label[i] + 1) \times k]$
    $start\_index \leftarrow label[i] \times k$
    $end\_index \leftarrow (label[i] + 1) \times k$
    $target\_2 \leftarrow$ torch.cat$((target[: start_index], target[end_index :]), \text{dim} = 0)$
    $T \leftarrow target\_2$.unsqueeze(0)
    $tmp\_index \leftarrow (Q[i] - T)$.pow(2).sum(2).sqrt().min(1)[1][0]
    $index \leftarrow$ map_original_list$[tmp_index]$
    $index\_list$.append$(index)$
  end for
  $index\_tensor \leftarrow$ torch.tensor$(index_list)$
  $e_{p,q} \leftarrow T[index\_tensor]$
  Output: $e_{h,j}, e_{p,q}$
else
  Output: $e_{h,j}$
end if

Algorithm 3 Voting Mechanism from Hierarchical Embedding Table

Input: Query Q.
Initialization:
$index\_list \leftarrow []$
$k \leftarrow$ Total embeddings for each category
$h \leftarrow$ Total classification categories
$Q \leftarrow Q.\text{unsqueeze}(1).\text{repeat}(1, \text{HET.size}(0), 1)$
$T \leftarrow \text{HET.unsqueeze}(0).\text{repeat}(Q.\text{size}(0), 1, 1)$
$indexes \leftarrow (Q - T).\text{pow}(2).\text{sum}(2).\text{sqrt}().\text{argsort}(\text{dim} = 1)[:, : 5]$
for j **in** $indexes$ **do**
  $index \leftarrow j//hk$
  $counter \leftarrow \text{Counter}(index.\text{tolist}())$
  $most\_common\_index \leftarrow \text{counter.most\_common}(1)[0][0]$
  $index\_list.\text{append}(\text{int}(most\_common\_index))$
end for
$index\_tensor \leftarrow \text{torch.tensor}(index\_list)$
Output: $T[index\_tensor]$

Appendix C. Dataset Details for UDA Benchmark

We assess the performance of TF-DAN on five distinct UDA benchmark datasets, each characterized by its unique features. The datasets considered include:

1. HAR Anguita et al. (2013): This dataset incorporates measurements from a 3-axis accelerometer, 3-axis gyroscope, and 3-axis body acceleration. Data is collected from 30 participants at a sampling rate of 50 Hz and uses non-overlapping segments of 128-time steps to predict activity labels. The objective is to classify time series into six activities: walking, walking upstairs, walking downstairs, sitting, standing, and lying down.

2. HHAR Stisen et al. (2015): Comprising 3-axis accelerometer measurements from 9 participants at a frequency of 50 Hz, this dataset employs non-overlapping segments of 128-time steps for classification. Activity labels include biking, sitting, standing, walking, walking upstairs, and walking downstairs.

3. WISDM Kwapisz et al. (2011): Featuring 3-axis accelerometer measurements from 36 participants at a frequency of 20 Hz, similar to the HAR dataset, we use non-overlapping segments of 128-time steps for classification. The dataset includes six activity labels: walking, jogging, sitting, standing, walking upstairs, and walking downstairs.

4. Sleep-EDF Goldberger et al. (2000): This task involves classifying electroencephalography (EEG) signals into five stages (Wake, N1, N2, N3, REM). Comprising EEG readings from 20 healthy subjects, we select a single channel (Fpz-Cz) as Ragab et al. (2023).

5. MFD Lessmeier et al. (2016): Collected by Paderborn University to identify incipient faults using vibration signals, this dataset consists of data collected under four different

operating conditions. Each condition is treated as a separate domain, and we use five cross-condition scenarios to evaluate domain adaptation performance. Each sample in the dataset comprises a single univariate channel with 5120 data points.

The summary of the datasets is in Table 2. These datasets span diverse applications and challenges, enabling a comprehensive evaluation of TF-DAN's effectiveness and robustness across various domains.

Table 2: Summary of datasets. Ragab et al. (2023)

Dataset	#Subjects/Domains	#Class	#Channels	Length	#Train	#Test
HAR	30	6	9	128	2300	990
HHAR	9	6	3	128	12716	5218
WISDM	30	6	3	128	1350	720
Sleep-EDF	20	5	1	3000	14280	6310
MFD	4	3	1	5120	7312	3604

Appendix D. UDA on Benchmark Datasets

We engage in activity prediction through an Unsupervised Domain Adaptation approach, utilizing benchmark datasets such as HAR, HHAR, and WISDM. Additionally, we delve into specific tasks within the medical and mechanical engineering domains, focusing on the Sleep-EDF and MFD datasets, respectively.

For each dataset, we present prediction results for 10 randomly selected source $\mapsto$ target pairs. To ensure robustness, we conduct the experiments with 5 random initializations and report the mean and standard deviation values. The results are organized into tables:

- Table 3: Mean accuracy and average Macro-F1 on the target domains for the HAR dataset.
- Table 4: Mean accuracy and average Macro-F1 on the target domains for the HHAR dataset.
- Table 5: Mean accuracy and average Macro-F1 on the target domains for the WISDM dataset.
- Table 6: Mean accuracy and average Macro-F1 on the target domains for the Sleep-EDF dataset.
- Table 7: Mean accuracy and average Macro-F1 on the target domains for the MFD dataset.

Table 3: Prediction accuracy for HAR Dataset between various subjects. Shown: mean accuracy and macro F1 over 5 random initializations.

	Mean Accuracy (%)									
Method	$2 \mapsto 9$	$1 \mapsto 14$	$1 \mapsto 10$	$4 \mapsto 9$	$21 \mapsto 29$	$25 \mapsto 28$	$30 \mapsto 2$	$4 \mapsto 3$	$2 \mapsto 11$	$9 \mapsto 18$
w/o UDA	48.28	81.44	52.81	68.97	50.96	84.35	54.95	66.02	77.89	30.91
DeepCoral	50.63	75.00	57.50	58.44	76.25	82.91	46.87	93.12	90.63	46.88
CDAN	66.88	88.95	56.87	63.13	**89.58**	85.21	54.37	97.29	85.42	58.86
DIRT-T	69.68	60.62	**62.81**	52.81	85.62	74.37	55.00	84.58	80.21	59.03
HoMM	35.00	58.96	23.75	37.81	39.37	73.75	41.88	72.71	65.47	41.27
CoDATS	59.06	79.58	54.69	67.50	81.87	88.75	71.56	88.12	68.23	63.89
AdvSKM	51.25	78.54	57.19	59.06	76.67	84.37	47.18	91.04	98.96	74.65
CLUDA	65.91	57.14	42.22	50.00	61.54	74.14	52.17	98.08	81.77	67.71
RAINCOAT	70.31	63.54	62.50	73.13	84.16	88.75	**87.50**	96.46	**100.0**	75.69
Ours	**73.12**	**90.01**	61.87	**80.08**	87.23	**88.79**	86.94	**100.0**	**100.0**	**76.17**
	Mean Macro F1									
w/o UDA	0.374	0.802	0.524	0.685	0.351	0.840	0.500	0.569	0.714	0.190
DeepCoral	0.440	0.733	0.590	0.554	0.714	0.832	0.492	0.927	0.910	0.440
CDAN	0.621	0.879	0.591	0.642	**0.900**	0.846	0.523	0.969	0.850	0.610
DIRT-T	0.675	0.501	0.645	0.458	0.861	0.706	0.491	0.811	0.810	0.580
HoMM	0.313	0.550	0.224	0.318	0.296	0.730	0.453	0.677	0.573	0.366
CoDATS	0.538	0.789	0.538	0.685	0.797	**0.899**	0.721	0.866	0.660	0.600
AdvSKM	0.452	0.767	0.583	0.549	0.737	0.846	0.519	0.893	0.990	0.730
CLUDA	0.664	0.557	0.389	0.511	0.570	0.756	0.481	0.980	0.810	0.670
RAINCOAT	0.645	0.614	0.626	0.724	0.831	0.899	**0.864**	0.963	**1.000**	**0.760**
Ours	**0.727**	**0.888**	**0.649**	**0.778**	0.894	**0.905**	0.848	**1.000**	**1.000**	0.728

Table 4: Prediction accuracy for HHAR Dataset between various subjects. Shown: mean accuracy and macro F1 over 5 random initializations.

	Mean Accuracy (%)									
Method	$7 \mapsto 6$	$1 \mapsto 3$	$0 \mapsto 2$	$2 \mapsto 3$	$2 \mapsto 6$	$7 \mapsto 2$	$4 \mapsto 0$	$5 \mapsto 0$	$7 \mapsto 0$	$4 \mapsto 2$
w/o UDA	78.04	**98.51**	64.51	50.32	45.11	32.37	32.81	30.42	33.92	19.16
DeepCoral	79.08	88.24	84.23	54.32	45.28	34.45	28.13	42.04	38.62	23.74
CDAN	96.04	93.01	76.19	60.27	31.88	37.05	29.09	22.84	25.09	27.16
DIRT-T	93.79	95.09	77.83	**66.22**	50.69	38.10	32.22	24.70	27.81	26.41
HoMM	84.63	88.91	68.38	45.83	44.03	35.94	32.37	34.60	29.60	23.21
CoDATS	88.95	95.16	79.61	61.09	35.90	38.54	21.80	33.85	32.41	**36.31**
AdvSKM	83.71	82.07	78.94	43.45	36.67	39.95	33.49	34.60	24.91	19.05
CLUDA	92.43	96.51	79.84	59.83	56.18	37.80	38.84	34.93	44.59	35.29
RAINCOAT	89.90	95.65	**87.82**	60.04	40.21	43.32	**46.46**	30.36	27.90	24.33
Ours	**97.04**	96.91	87.54	65.78	**57.01**	**51.46**	46.28	**42.38**	**44.97**	35.31
	Mean Macro F1									
w/o UDA	0.783	**0.985**	0.600	0.410	0.359	0.310	0.290	0.220	0.337	0.135
DeepCoral	0.761	0.874	0.860	0.498	0.419	0.320	0.260	0.380	0.409	0.230
CDAN	0.961	0.930	0.700	0.563	0.325	0.320	0.270	0.202	0.265	0.257
DIRT-T	0.936	0.950	0.760	**0.628**	0.441	0.340	0.300	0.207	0.303	0.283
HoMM	0.836	0.881	0.625	0.408	0.398	0.377	0.318	0.306	0.315	0.192
CoDATS	0.883	0.951	0.730	0.580	0.366	0.360	0.200	0.328	0.315	**0.356**
AdvSKM	0.821	0.791	0.720	0.388	0.333	0.410	0.330	0.279	0.270	0.157
CLUDA	0.928	0.965	0.820	0.544	0.506	0.360	0.400	0.305	0.426	0.345
RAINCOAT	0.903	0.955	**0.870**	0.553	0.397	0.440	0.450	0.288	0.331	0.235
Ours	**0.987**	0.967	0.814	0.611	**0.544**	**0.518**	**0.453**	**0.409**	**0.444**	0.381

Table 5: Prediction accuracy for WISDM Dataset between various subjects. Shown: mean accuracy and macro F1 over 5 random initializations.

	Mean Accuracy (%)									
Method	$4 \mapsto 5$	$11 \mapsto 16$	$12 \mapsto 23$	$18 \mapsto 23$	$26 \mapsto 29$	$28 \mapsto 27$	$4 \mapsto 11$	$28 \mapsto 21$	$12 \mapsto 26$	$17 \mapsto 26$
w/o UDA	42.03	13.73	45.00	58.33	**50.00**	8.00	32.89	59.62	54.88	43.90
DeepCoral	76.81	15.69	39.17	61.67	21.67	68.00	27.63	28.85	48.17	65.24
CDAN	60.87	17.65	61.67	23.33	15.00	76.00	44.74	61.54	48.78	65.85
DIRT-T	73.91	6.86	63.33	56.67	39.17	46.00	42.11	41.35	53.66	63.41
HoMM	57.97	3.92	32.50	45.83	39.17	52.00	32.24	31.73	40.85	43.90
CoDATS	56.52	30.39	52.50	60.83	27.50	66.00	54.61	31.73	64.02	**70.12**
AdvSKM	61.59	23.53	29.17	25.00	36.67	78.00	24.34	17.31	35.98	56.71
CLUDA	62.86	15.38	54.84	48.39	6.67	36.00	47.37	34.62	48.78	51.22
RAINCOAT	65.22	19.61	63.33	63.33	21.67	84.00	43.42	84.62	57.32	64.63
Ours	**87.96**	**42.32**	**66.77**	**69.69**	49.75	**85.21**	**72.58**	**84.64**	**64.04**	65.77
	Mean Macro F1									
w/o UDA	0.099	0.083	0.176	0.226	0.133	0.033	0.329	0.388	0.223	0.160
DeepCoral	0.704	0.166	0.176	0.308	0.136	0.519	0.300	0.225	0.234	0.456
CDAN	0.366	0.277	0.340	0.156	0.218	0.337	0.383	0.541	0.257	0.422
DIRT-T	0.492	0.096	0.382	0.274	0.255	0.496	0.276	0.346	0.255	0.417
HoMM	0.501	0.020	0.201	0.268	0.268	0.421	0.229	0.245	0.237	0.281
CoDATS	0.496	**0.283**	0.384	0.508	0.151	0.291	0.414	0.266	0.310	0.502
AdvSKM	0.548	0.271	0.191	0.160	0.269	0.458	0.204	0.154	0.221	0.438
CLUDA	0.611	0.126	0.359	0.275	0.111	0.370	0.262	0.321	0.236	0.233
RAINCOAT	0.461	0.265	**0.519**	0.283	0.162	**0.713**	0.333	0.691	0.267	0.398
Ours	**0.819**	0.254	0.517	**0.548**	**0.311**	0.588	**0.705**	**0.730**	**0.369**	**0.731**

Table 6: Prediction accuracy for Sleep-EDF Dataset between various subjects. Shown: mean accuracy and macro F1 over 5 random initializations.

	Mean Accuracy (%)									
Method	$1 \mapsto 8$	$6 \mapsto 10$	$8 \mapsto 0$	$2 \mapsto 1$	$15 \mapsto 4$	$8 \mapsto 1$	$4 \mapsto 19$	$8 \mapsto 5$	$18 \mapsto 6$	$13 \mapsto 7$
w/o UDA	52.05	75.11	68.53	78.75	68.54	61.43	77.58	51.39	76.14	68.44
DeepCoral	61.82	71.09	66.41	78.07	69.90	62.66	72.74	43.62	76.17	68.85
CDAN	45.62	75.31	75.13	73.23	70.78	60.16	68.97	65.89	75.78	65.62
DIRT-T	49.06	77.97	84.83	77.92	68.75	69.84	80.56	70.25	72.72	61.77
HoMM	62.29	71.61	64.58	65.05	73.70	58.70	67.62	36.91	76.43	66.46
CoDATS	62.55	67.29	62.63	79.74	72.71	60.57	82.34	55.01	68.82	75.00
AdvSKM	67.34	71.20	59.31	79.53	69.32	60.26	70.62	38.35	74.09	66.04
CLUDA	46.81	53.64	51.01	60.47	57.65	45.64	48.58	43.40	47.31	47.93
RAINCOAT	59.74	77.08	72.98	78.33	69.90	66.30	71.83	64.78	76.17	65.78
Ours	**75.81**	**78.66**	**78.97**	**80.13**	**73.65**	**76.92**	**82.51**	**73.84**	**80.52**	**77.98**
	Mean Macro F1									
w/o UDA	0.409	0.694	0.632	0.677	0.564	0.560	0.619	0.559	0.651	0.576
DeepCoral	0.556	0.574	0.582	0.728	0.640	0.565	0.618	0.464	0.670	0.611
CDAN	0.400	0.590	0.636	0.687	0.596	0.495	0.529	0.573	0.664	0.572
DIRT-T	0.445	0.596	0.714	0.710	0.583	0.563	0.671	0.590	0.618	0.523
HoMM	0.548	0.582	0.572	0.662	**0.691**	0.540	0.551	0.402	0.643	0.591
CoDATS	0.555	0.534	0.522	0.696	0.668	0.497	0.719	0.489	0.627	0.630
AdvSKM	0.599	0.545	0.519	**0.740**	0.656	0.562	0.587	0.401	0.650	0.607
CLUDA	0.310	0.179	0.364	0.338	0.409	0.305	0.233	0.305	0.284	0.365
RAINCOAT	0.528	0.641	0.601	0.724	0.578	0.572	0.536	0.540	0.675	0.527
Ours	**0.715**	**0.749**	**0.702**	0.702	0.645	**0.790**	**0.751**	**0.757**	**0.720**	**0.665**

Table 7: Prediction accuracy for MFD Dataset between various subjects. Shown: mean accuracy and macro F1 over 5 random initializations.

	Mean Accuracy (%)									
Method	$0 \mapsto 1$	$0 \mapsto 3$	$1 \mapsto 2$	$1 \mapsto 0$	$3 \mapsto 0$	$2 \mapsto 0$	$3 \mapsto 2$	$0 \mapsto 2$	$2 \mapsto 1$	$1 \mapsto 3$
w/o UDA	41.73	51.39	67.04	42.06	39.84	28.97	79.69	61.71	88.46	98.45
DeepCoral	66.15	69.79	64.21	41.67	48.33	41.67	61.53	65.89	89.14	81.32
CDAN	47.36	68.79	76.00	46.61	50.04	49.33	70.24	62.69	90.62	99.44
DIRT-T	58.37	65.62	72.19	**81.10**	73.40	**70.65**	74.63	64.84	70.83	98.85
HoMM	65.59	68.34	65.29	42.56	47.84	36.64	62.35	59.90	82.66	81.81
CoDATS	60.66	62.72	**86.16**	41.74	45.59	42.58	79.97	54.91	81.03	**100.0**
AdvSKM	64.73	71.80	65.10	40.85	48.25	45.05	61.87	64.14	86.24	82.63
CLUDA	48.34	48.56	48.12	41.69	42.57	47.67	49.45	54.77	46.56	44.79
RAINCOAT	63.02	67.49	76.45	61.53	68.45	65.40	81.55	58.82	**92.30**	97.14
Ours	**73.96**	**84.28**	83.51	78.77	**84.98**	67.24	**87.22**	**67.33**	91.71	99.84
	Mean Macro F1									
w/o UDA	0.400	0.520	0.758	0.575	0.558	0.479	0.851	0.674	0.915	0.989
DeepCoral	0.496	0.551	0.688	0.477	0.503	0.473	0.667	0.607	0.919	0.856
CDAN	0.318	0.523	0.800	0.343	0.428	0.452	0.743	0.525	0.925	0.996
DIRT-T	0.492	0.634	0.788	**0.830**	0.756	0.742	0.789	**0.733**	0.777	0.992
HoMM	0.460	0.490	0.700	0.480	0.501	0.424	0.665	0.442	0.866	0.859
CoDATS	0.557	**0.689**	**0.871**	0.451	0.532	0.499	0.826	0.393	0.843	**1.000**
AdvSKM	0.450	0.633	0.685	0.473	0.504	0.501	0.674	0.560	0.896	0.866
CLUDA	0.408	0.339	0.333	0.252	0.295	0.323	0.345	0.383	0.325	0.311
RAINCOAT	**0.610**	0.655	0.806	0.692	0.737	0.719	0.850	0.581	**0.941**	0.979
Ours	0.580	0.623	**0.871**	0.826	**0.885**	**0.751**	**0.902**	0.577	0.925	0.993

Appendix E. Implementation Details for Hyperparameters

E.1. Learning rate

Table 8: Leaning rates of different components in TF-DAN.

Component	Training Phase	Adaptation Phase
Encoder	1e-4	2e-6
HET - temporal block	2e-4	2e-6
HET - frequency block	2e-4	1e-8
Classifier	1e-2	-
Decoder	-	2e-4

E.2. Parameter K

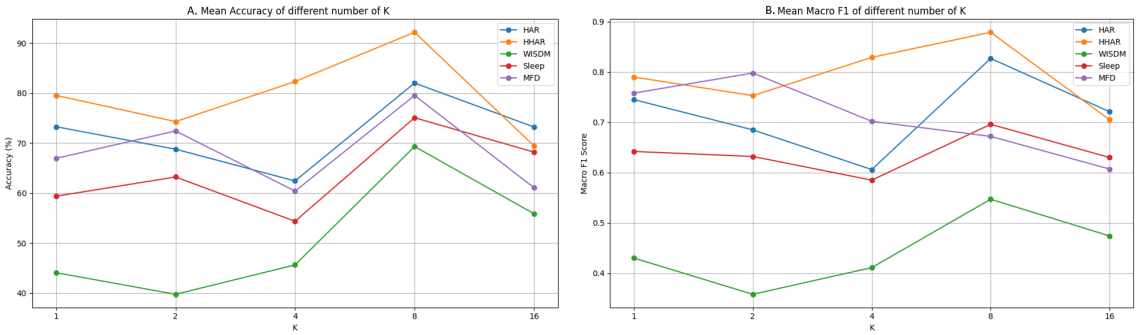


Figure 15: TF-DAN's performance in different K. We observed that for the majority of datasets, setting K to 8 yielded better performance, excluding MFD dataset in mean macro F1 score.

E.3. Parameter γ

Table 9: γ in different datasets.

HAR	HHAR	WISDM	Sleep-EDF	MFD
1.2	1.2	1	1	1.5

E.4. Training Batch size

Table 10: Batch sizes of different datasets in TF-DAN.

Dataset	Training Phase	Adaptation Phase
HAR	32	32
HHAR	32	32
WISDM	32	32
Sleep-EDF	32	32
MFD	32	32

Appendix F.

To further understand why TF-DAN succeeds in UDA tasks, we utilize principle component analysis (PCA) to visualize the embeddings in 2D and observe the distribution of embeddings from the temporal block and the frequency block. Fig. 16 shows that even though we initialize the embeddings of both blocks uniformly in the HET, the trained embeddings of the temporal block do not cluster as effectively as those of the frequency block.

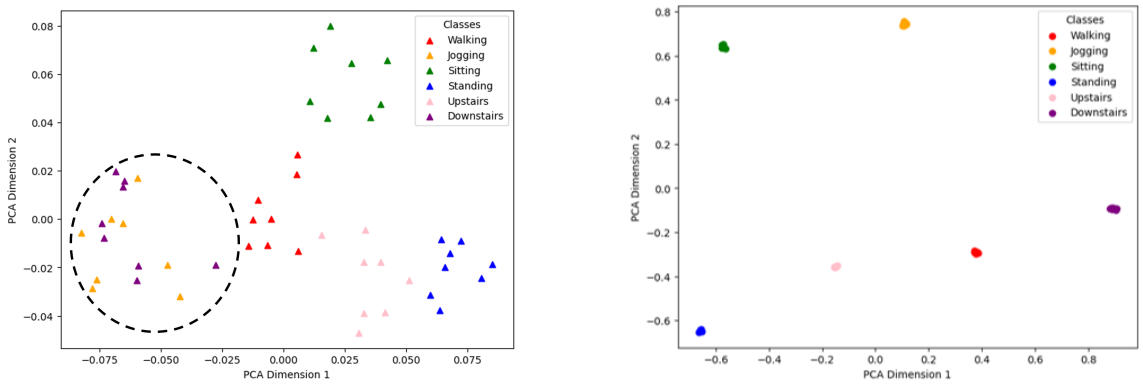
This may be due to the higher diversity and complexity of features in the time domain. These features include not only class-specific characteristics but also information such as confounders. In contrast, the frequency block contains more uniform and less diverse information, which allows it to learn the key features of the category more effectively during training. As a result, it demonstrates better clustering performance in the PCA visualization (Fig. 16(b)), aligning with the findings from earlier experiments in Section 3.1.

Appendix G. Computation Analysis

Two main factors affect TF-DAN's performance: (1) the size of the hierarchical embedding table and (2) the number of classification categories. The following will elaborate on these two aspects:

G.1. Size of the hierarchical embedding table

During the training phase, as the source domain has labels, we only need to calculate K nearest neighbors for each category, where K represents the number of embeddings per



(*a*) Temporal block (*b*) Frequency block

Figure 16: PCA visualization of (a) temporal features and (b) frequency features in HET from WISDM dataset.

category (Fig. 15). We determine the appropriate value of K through experimentation, considering both Mean accuracy and macro F1 score. We found that for the majority of datasets, setting K to 8 yielded better performance, excluding MFD dataset in mean macro F1 score. Accordingly, we speculate that other parameters of the model contribute to its superior performance at K=8.

G.2. Number of classification categories

During the adaptation phase, as the target domain lacks labels, we must compute all embeddings in the embedding table to obtain the closest embeddings. At this point, the time required by the model is directly influenced by the number of categories, leading to a significant impact.

Our study utilized an A100 GPU 40GB, with an average total training time of 0.5 GPU hours across the five datasets. Table 11 is the relevant parameter table for the 5 datasets:

Table 11: Epochs of training and adaptation phases in different datasets.

Dataset	Training Epoch	Adaptation Epoch
HAR	70	50
HHAR	80	70
WISDM	150	50
Sleep-EDF	200	100
MFD	150	100

Appendix H. Broader Impacts

Potential positive societal impacts. We may apply TF-DAN in smart elderly care facilities. Given the significant differences in behavior between the elderly population and

middle-aged adults, such as frequent nocturnal bathroom visits, slower mobility, and increased susceptibility to falls, leveraging the human activity recognition datasets (i.e, HAR, HHAR, WISDM, DSADS) as the source domain and adapting it to the elderly population for downstream tasks could be a crucial research direction and technological advancement in the future.