EarDA: Towards Accurate and Data-Efficient Earable Activity Sensing

Shengzhe Lyu, Yongliang Chen, Di Duan, Renqi Jia, Weitao Xu*
Department of Computer Science, City University of Hong Kong
Hong Kong SAR, China

***Abstract*—In the realm of smart sensing with the Internet of Things, earable devices are empowered with the capability of multi-modality sensing and intelligence of context-aware computing, leading to its wide usage in Human Activity Recognition (HAR). Nonetheless, unlike the movements captured by Inertial Measurement Unit (IMU) sensors placed on the upper or lower body, those motion signals obtained from earable devices show significant changes in amplitudes and patterns, especially in the presence of dynamic and unpredictable head movements, posing a significant challenge for activity classification. In this work, we present *EarDA*, an adversarial-based domain adaptation system to extract the domain-independent features across different sensor locations. Moreover, while most deep learning methods commonly rely on training with substantial amounts of labeled data to offer good accuracy, the proposed scheme can release the potential usage of publicly available smartphone-based IMU datasets. Furthermore, we explore the feasibility of applying a filter-based data processing method to mitigate the impact of head movement. *EarDA*, the proposed system, enables more data-efficient and accurate activity sensing. It achieves an accuracy of 88.8% under HAR task, demonstrating a significant 43% improvement over methods without domain adaptation. This clearly showcases its effectiveness in mitigating domain gaps.**

***Index Terms*—IMU, earable, domain adaptation**

I. Introduction

Earable devices, which significantly enhance people's everyday listening experience, have experienced a noticeable surge in recent years. With technological advancements, many newly launched devices have been integrated with various sensors (e.g., AirPods Pro 2, Bose QuietControl 30 and Sony WF-1000XM5), giving rise to an emerging research area known as earable sensing. Compared to other sensing modalities, earable sensing holds a significant advantage as it seamlessly integrates into users' daily lives. Consequently, it has attracted researchers' attention and has been shown to be effective in various tasks, including health care [1], [2], speech enhancement [3], [4] and activity recognition [5]–[7].

Existing works in this realm are either developed based on dedicated or commercial off-the-shelf (COTS) devices. The first type requires specially designed devices, such as eSense [8] and OpenEarable [9]. Despite their effectiveness, the requisite highly customized hardware significantly restricts their practicability. On the other hand, with the emergence of sensor-rich earable products like Apple Airpods, an increasing number of researchers have proposed plausible solutions based on these COTS devices [3], [10]. However, the obtained sensory data is highly heterogeneous across different manufacturers, and there are no unified datasets applicable to all devices, which suggests a dilemma where developing new sensing applications often necessitates collecting the whole dataset from scratch. As such, *how to develop an accurate earable sensing methodology in a data-efficient manner based on COTS devices* remains an unsolved ad-hoc question.

To fill this research gap, in this paper, we propose *EarDA*, an earable human activity recognition (HAR) system based on domain adaption. To the best of our knowledge, we are the first to propose such a system that can utilize knowledge learned from other sensing modalities (e.g., smartphones) with abundant training data to enable effective earable-based HAR with limited data. However, directly applying well-trained frameworks from other modalities to earable data is impracticable due to the significant domain gap that exists between these two modalities (e.g., smartphones are always carried in pockets while earable devices are attached to the user's head). Thereafter, we specially design a domain adaptation training framework to leverage HAR features learned from public datasets with substantial sensory data to boost the performance on new earable devices. It effectively reduces the required human labor intensity on collecting abundant earable training data due to the absence of unified datasets. In addition, we encounter another challenge that head movements introduce distinct signal variations, which overwhelm the informative signal for activity recognition. To mitigate this, we carefully investigate the obtained signals in different scenarios and devise a filter-based solution that effectively mitigates the resulting effects.

We have built a prototype of *EarDA* atop Apple AirPods Pro 2, and it is evaluated on four common daily activities (walking, upstairs, standing, and jogging) based on IMU measurements. The training is conducted with a small volume of earable data (i.e., 5 minutes) with the assistance of four public smartphone-based IMU datasets. It recognizes activities at an accuracy of 89%, with a significant increment of 43%, compared with cases only training with the public dataset and adopting the model on earables without domain adaption, demonstrating the proposed system's effectiveness. Incorporating a filter-based data processing method to mitigate head movement enhances the overall accuracy by 4%, highlighting the system's ability to cooperate with the head interference. In short, we make the

* Corresponding author.

979-8-3503-6338-8/24/$31.00 ©2024 IEEE
DOI 10.1109/CSCAIoT62585.2024.00005

following contributions:

- We take the first step toward understanding the domain gap existing in the public wearable datasets, especially smartphone-based datasets and IMU data from COTS earables.
- We propose *EarDA* with novel techniques that mitigate the domain gap between COTS earable devices and public datasets and influences introduced by irregular head movements.
- We conduct extensive evaluations and showcase the superiority of the proposed system. The overall accuracy achieves 88.8%, and the ablation study demonstrates the effectiveness of the proposed components.

II. Feasibility Study

A. Impact of Location Diversity

The disparity among sensory data collected by different modalities mainly stems from the different sensing locations. To better illustrate such discrepancies, we collect IMU data from a smartphone placed in the pocket of the trousers and a naturally-worn earable device when performing different actions. We visualize their sensory data in Figure 1. It can be observed from Figure 1(a) and Figure 1(b) that there are notable differences in both of the acceleration and gyroscope measurements. The smartphone's sensor records the movements and vibrations of the lower body, reflecting the motion patterns and gestures associated with leg and hip movements. On the other hand, the earable device captures a combination of head and body movements, providing a different perspective on overall body motion. For example, the maximum range of acceleration during jogging is observed to be twice that of earables, as shown in Figure 1(d, e). When measuring acceleration signals in the lower body, the amplitude of the signals tends to be larger. Nevertheless, apart from all the challenges, the presence of similar patterns, such as periodic peaks in both domains, underscores the potential to transfer the knowledge across domains.

B. Impact of Head Movement

Head movement is unavoidable when performing daily activities. However, the introduced signal fluctuations pose extra challenges in sensing tasks. As shown in Figure 1(c), even when the lower body and upper body are experiencing periodical movements during walking, unpredictable interference appears in IMU signals collected by the ear-worn device due to the head movement. In addition, the impact of head movements is still obvious while the users tried their best to control their heads during the data collection, as shown in Figure 1(b). This can significantly affect the HAR tasks because the noise and variations distort the IMU data patterns, making it more difficult to accurately classify the user's activity. Fortunately, due to the filtering function of the upper body, the frequency components of earable IMU data always fall in a certain range, suggesting the usage of filter-based data processing methods might help to mitigate the interference from head movements.

III. System Design

A. System Overview

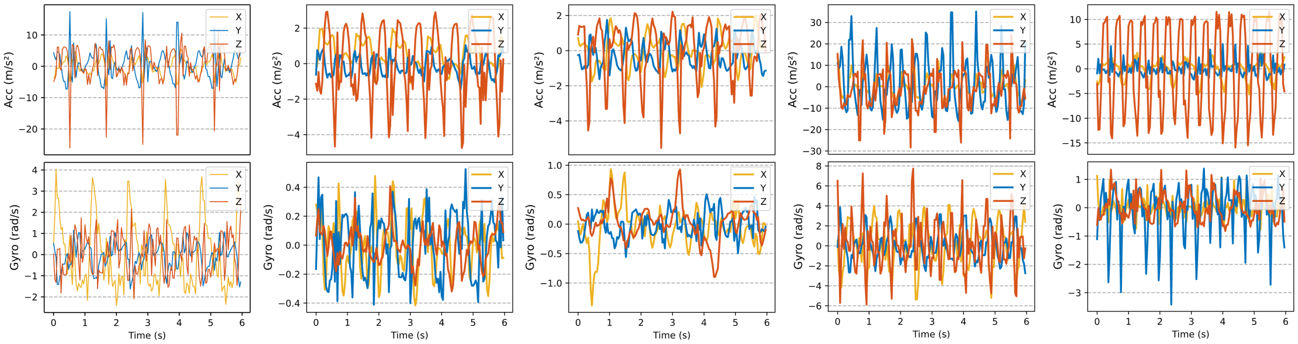
The overview of *EarDA* is shown in Figure 2. We design an effective adversarial-based domain adaptation module that can extract domain-invariant features across the source domain with substantial data (i.e., public datasets) and the target domain with limited data (i.e., earable data). Moreover, we specially design a filter-based data pre-processing pipeline to mitigate the influence caused by head movements.

B. Domain-invariant Feature Extraction

a) Feature Extractor: A bi-directional LSTM model is designed to construct a generalized feature extractor due to its architecture, which is adapted to capturing the complex temporal dynamics and dependencies inherent in sequential data. To boost the model's robustness against device orientation and data heterogeneity, the acceleration and gyroscope magnitudes are calculated as inputs. It serves as a pre-processing step to abstract away the device's orientation, providing a more orientation-invariant representation of the motion. This approach enhances the model's generalization ability across different devices and domains. It can be seen in Figure 1 that the acceleration readings are always steadier than gyroscope readings during the same activity. The values derived from the accelerometer and gyroscope always fall on different ranges as the gyroscope is more sensitive to movement. To mitigate the impact, accelerations are normalized through the division by gravity, after which the range differences between acceleration readings and gyroscope readings can be narrowed.

b) Domain Adaptation: As mentioned previously, the normal LSTM classifier trained on the source domain cannot be directly applied to the target domain. As a result, the model should retain the capability to extract domain-invariant features regardless of which domain the input data falls on. In this work, we present a domain adaptation module that takes substantial data from the source domain and limited data from the target domain as input. Then, two distinct pathways are derived from the feature extractor, one label predictor, and one domain classifier. The label predictor is tasked with the objective of predicting the correct activity label for the input IMU sequence. In contrast, the domain classifier, aided by the gradient reversal layer, attempts to determine the domain of the input data.

Inspired by [11], we integrate a Gradient Reversal Layer (GRL) to assist in learning domain-invariant features of the two domains. Specifically, during the forward pass, the GRL acts as an identity function, allowing data to pass through unchanged. However, during the backward pass, it reverses the gradient sign before passing it back to the preceding layers. This effectively encourages the LSTM layers to learn representations that are indistinguishable from the domain classifier. Consequently, this process leads to the minimization of domain discrepancy, resulting in the capability of domain-



(a) walk, phone, slight head movement

(b) walk, earable, slight head movement

(c) walk, earable, random head movement

(d) jog, phone, slight head movement

(e) jog, earable, slight head movement

Fig. 1: IMU measurements by smartphone (left pocket) and earable (head) for different activities and head movements.

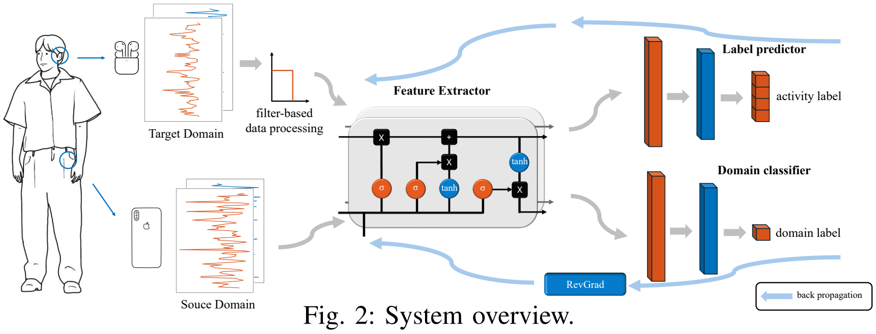


Fig. 2: System overview.

invariant feature extraction. More formally, the loss function for the network utilizing GRL can be written as:

$$
\begin{aligned}
E(\theta_f, \theta_y, \theta_d) = & \sum_{i=1..N} L_y\left(G_y(G_f(x_i;\theta_f);\theta_y), y_i\right) \\
& - \lambda \sum_{i=1..N} L_d\left(G_d(G_f(x_i;\theta_f);\theta_d), y_i\right) \\
= & \sum_{i=1..N} L_y^i(\theta_f, \theta_y) - \lambda \sum_{i=1..N} L_d^i(\theta_f, \theta_d)
\end{aligned}
$$

Here, G_f, G_y, G_d are the functions representing the feature extractor, label predictor, and domain classifier over the parameters of $\theta_f, \theta_y, \theta_d$, respectively. λ is a trade-off parameter that controls the importance of the domain adaptation loss relative to the label prediction loss, which is selected to be 0.3 in our work. L_y is the label predictor loss and L_d is the domain classifier loss. This loss function aims to minimize the label prediction error while also encouraging the feature extractor to learn domain-invariant features. The presence of the GRL is implied in the negative sign before the domain classifier loss term, which reflects the gradient reversal during the backpropagation process.

C. Mitigation of Head-movement

The head movement interference has been a challenge in performing HAR tasks with earables because the signals recorded by ear-worn devices are usually unpredictable with the interference added. One approach to mitigate the impact of head movements on HAR tasks is to employ signal processing like applying filters. It must be admitted that eliminating the motion caused by head movements from raw IMU data is not easy, as the frequencies of body or head movements spread over a big range and overlap each other in a specific range. As studied in [12], human activity frequencies are always between 0 and 20 Hz and 98% of the frequency spectrum is below 10 Hz. Nonetheless, due to the filtering function of the body structure, the signals recorded on the head are usually smaller, always below 5.8 Hz, presented in [13]. This leads to the potential of mitigating head movement by low-pass filtering. We observed similar frequency component distribution by comparing the frequency spectrum under different kinds of head movements. As illustrated in Figure 3, the dominant frequency components of gyroscope data are concentrated in low-frequency bins (e.g., below 5 Hz). However, when the same activity is accompanied by more rigorous head movements, noticeable lobes appear at higher frequency ranges. Considering previous literature on regular human activities [12], [13] and the observed sensory data, we empirically devise a low-pass filter with a cut-off frequency of 5 Hz. It can effectively filter out unwanted disturbances occupying high-frequency ranges, including those introduced by erratic head movements, while simultaneously retaining informative signals related to body movements.

IV. Evaluation

A. Implementation

To better understand the impact of sensor locations and head movements on the performance of HAR tasks, we collected a dataset with Apple Airpods Pro 2. The Airpods are chosen

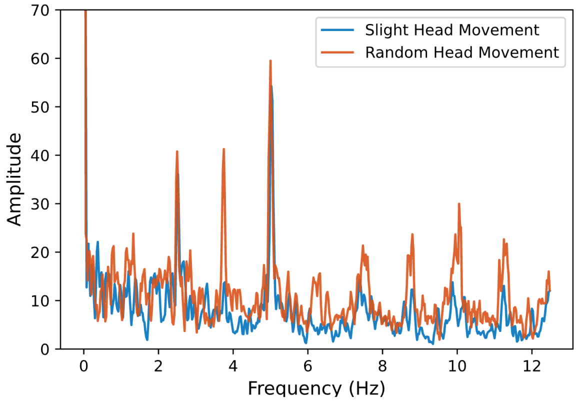
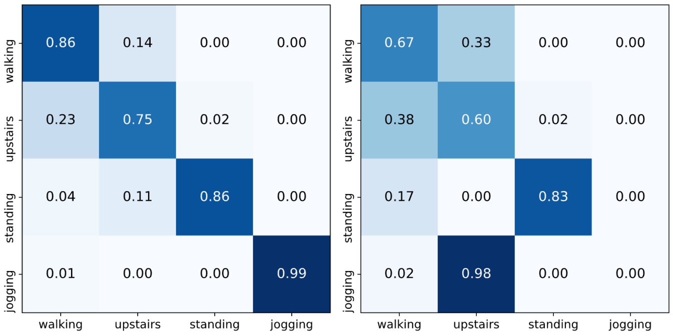


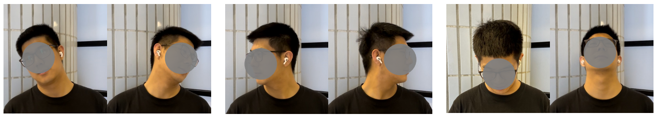
Fig. 3: Frequency spectrum of gyroscope measurements collected by earables during jogging.



(a) confusion matrix with domain adaptation.

(b) confusion matrix without domain adaptation.

Fig. 4: Confusion matrix on the test set with and without applying domain adaptation.



(a) roll

(b) yaw

(c) pitch

Fig. 5: Three evaluated head movements in the experiment.

because: 1) the APIs are provided to access accurate and real-time 6-axis motion data captured from the accelerometer and gyroscope embedded in the earbuds, 2) IMU signals from each earbud in a pair are fused together automatically in API to represent the motion signals of the head, and 3) it is widely utilized in HAR tasks on earable devices as it is one of the most popular COTS earable products. The sampling rate of Airpods was set to the highest value allowed by API, which is 25 Hz.

A two-layer bidirectional LSTM model with a hidden size of 16 is implemented as the feature extractor. The model takes two-dimensional inputs with a sequence length of 100 from each domain. Fully connected layers with ReLU activation are employed for both the label predictor and the domain classifier. The gradient reversal layer is only implemented as part of the domain classifier. The model was trained with a batch size of 32 for 200 epochs on the computer with an NVIDIA GeForce RTX 2080Ti GPU.

B. Experimental Setup

1) Earable Data Collection: During the data collection with Airpods, the participant was asked to perform four different activities *(walking, walking upstairs, standing, jogging)* while wearing the Airpods as usual. The total time for each activity is 14 minutes, constituting 4 minutes of slight head movement, 4 minutes of random head movements, and 2 minutes for each of rolling, yawing, and pitching. During the collection period, the participant was first asked to control their head in a steady state for 4 minutes while doing all four activities, leading to slight head movement interference on captured data. After that, the participant performed all activities casually for another 4 minutes. Head movements are unavoidable in daily life, so random interferences are added to IMU data collected. Rolling (tilting), yawing (waving), and pitching (nodding) are chosen as specific head movements, as depicted in Figure 5. We want to evaluate the impact of each kind of head movement on the performance of cross-domain HAR tasks separately.

The collected dataset is segmented into non-overlapping windows, each encompassing sequences of 4 seconds. Each sequence corresponds to a single activity label. In consideration of the variability introduced by device orientation, we compute the magnitudes of acceleration and gyroscope data to serve as a more stable input feature set for the deep learning model. Subsequently, the processed 840 data samples are randomly partitioned into distinct subsets for model training (10%), validation (10%), and testing (80%). As a result, the training set contains only 84 samples while the testing set contains 672 samples.

2) Public Datasets: The source domain refers to smartphone-based publicly available IMU datasets. Thus, four datasets are chosen as they are widely utilized in previous works [14], [15]. The four publicly available datasets comprise IMU signals gathered from individuals engaging in various activities while wearing smartphones. These datasets vary not only in the model of the smartphones used but also in the range of activities performed by the participants. Additionally, some of these datasets are further diversified by sensor locations. By incorporating these datasets into the experiment, the adaptability and effectiveness of the algorithm across diverse domains can be facilitated.

a) MotionSense: The MotionSense dataset [16] was collected by accelerometer and gyroscope sensors of Apple iPhone6s in 50 Hz. There are six different activities designed for the dataset *(walking, walking upstairs, walking downstairs, standing, jogging and sitting)*, and the smartphones were kept in the participants' pockets of trousers.

b) HHAR: The Heterogeneity Human Activity Recognition (HHAR) dataset [17] contains IMU signals captured between 50-200 Hz for six different activities *(walking, walking upstairs, walking downstairs, standing, biking and sitting)*.

TABLE I: Performance of the model with different head movements after mitigation of head movements.

Head movement	**Slight**		**Random**		**Roll**		**Yaw**		**Pitch**		**Average**	
Metric	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**
Walking	0.90	0.82	0.76	0.83	1.00	0.63	0.92	0.92	0.88	0.92	0.82	0.82
Upstairs	0.77	0.81	0.89	0.79	0.62	0.76	0.80	0.89	0.86	0.92	0.78	0.81
Standing	0.92	0.96	1.00	1.00	1.00	0.96	1.00	0.86	1.00	0.86	0.97	0.94
Jogging	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
Overall	**0.90**	**0.89**	**0.91**	**0.91**	**0.90**	**0.84**	**0.93**	**0.92**	**0.94**	**0.93**	**0.89**	**0.89**

TABLE II: Performance of the model with different head movements.

Head movement	**Slight**		**Random**		**Roll**		**Yaw**		**Pitch**		**Average**	
Metric	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**
Walking	0.78	0.86	0.76	0.80	0.86	0.39	0.87	0.67	1.00	0.59	0.80	0.73
Upstairs	0.95	0.82	0.83	0.78	0.52	0.69	0.52	0.68	0.50	0.67	0.64	0.72
Standing	0.98	0.99	1.00	1.00	1.00	0.88	1.00	0.70	1.00	0.74	0.99	0.91
Jogging	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
Overall	**0.93**	**0.92**	**0.90**	**0.90**	**0.84**	**0.74**	**0.85**	**0.76**	**0.88**	**0.75**	**0.86**	**0.84**

There are six models of smartphones in the dataset (three models of Samsung Galaxy and one model of LG), carried by the users around their waist.

c) UCI HAR: The UCI HAR dataset [18] was obtained from a waist-mounted Samsung Galaxy S2 smartphone. 6-axis IMU signals were collected under six different activities *(walking, walking upstairs, walking downstairs, standing, lying down and sitting)* at 50 Hz.

d) Shoaib: The Shoaib dataset [19] consists of the motion data of seven daily activities *(walking, walking upstairs, walking downstairs, standing, jogging, biking and sitting)*. During data collection, five Samsung Galaxy SII models were placed in five body positions. The sampling rate is 50 Hz.

We randomly extract samples corresponding to four activities of interest: walking, walking upstairs, standing, and jogging. A balanced dataset comprising 10,000 instances is constructed, ensuring an equal representation of 2,500 samples for each activity to prevent class imbalance during the training process. In line with the sampling frequency of the IMU signals from the Airpods, the data is down-sampled to 25 Hz. The pre-processing steps are consistent with those applied to the Airpods IMU signals, involving the segmentation of the data into non-overlapping sequences of 100 samples each. The prepared dataset is then divided, allocating 80% for training, 10% for validation, and the remaining 10% for testing. Finally, the training set comprises 8,000 samples, which is 100 times larger than the training samples collected from earables.

C. Overall Performance

Table I summarizes the system's overall performance on the earable data test set, including both accuracy and F1-score. Figure 4(a) further visualizes the confusion matrix. The proposed framework demonstrates an exceptional ability to classify all four activities, regardless of the original sensor domain. The model achieves an average accuracy of 88.8% and an F1-score of 0.89. While excelling in classifying jogging activities with a 99% accuracy, the model exhibits lower performance for upstairs activities, sometimes misclassified as less-strong activities like walking.

D. Effectiveness of domain adaptation

To evaluate the effectiveness of the proposed domain adaptation component, keeping all the settings the same as *EarDA*, we discard the domain classifier and only use the 8,000 data samples from public datasets to train a classifier. The trained model is then applied to test the AirPods samples. As shown in Figure 4(b), the model's performance shows a significant decline, achieving an accuracy of only 46%. This decline in performance clearly indicates that the model trained on public IMU datasets cannot be directly applied to earable IMU data due to a substantial domain gap across different sensor positions. In Figure 4(b), we present the confusion matrix derived from the model's predictions. It can be observed that all jogging samples were wrongly classified as less intense activities such as walking or going upstairs.

In comparison, after adding the Gradient Reversal Layer in the back-propagation path of the domain classifier, the model exhibits strong potential in recognizing various activities, according to the confusion matrix in Figure 4(a). The model's performance on standing and jogging activities underlines its capability to discern static and highly dynamic activities with high precision. As mentioned, the amplitudes and patterns of IMU sequences while jogging show significant differences between earbuds and smartphones, which is addressed by domain adaptation. This significant improvement in performance highlights the effectiveness of bridging the domain gap and transferring knowledge even with limited earable domain data.

E. Effectiveness of head movement mitigation

To investigate the impact of pre-processing on system performance, we compare the results before and after applying low-pass filtering. The accuracy and F1-score without head movement mitigation are presented in Table II, while Table I displays the results with head movement mitigation. The test set is further divided into 5 groups based on the magnitude of head movements: slight head movement, random head movement, roll, yaw, and pitch.

From Table II, it is evident that the model's accuracy decreases when specific types of head movements are present, particularly under the investigation of rolling, yawing, or pitching. When participants make an effort to keep their heads steady during data collection (as seen in the slight head movement column), the accuracy reaches up to 93%. However, when larger magnitude head movements are present, the accuracy degrades to approximately 85%.

As expected, the filter effectively eliminates high-frequency components of the Airpods IMU data while preserving frequency components below 5 Hz, which are relevant to body activities. As shown in Table I, the model's performance in classifying all four activities under various head movement scenarios improves, with an overall accuracy increment of around 4%. Notably, after applying the 5 Hz low-pass filter, there is no degradation in accuracy, regardless of the magnitudes or kinds of head movement interference. It is crucial to note that the utilization of more advanced techniques or components could potentially isolate and mitigate even low-frequency head movements from the captured motion signals. However, those approaches would inevitably introduce a higher computational overhead and might necessitate a larger amount of data for training, which is not desired here.

V. Related Work

Until now, deep learning models deployed in wearable devices facilitate many ubiquitous applications. Various sensors like IMUs, microphones, and wireless modules are widely embedded in mobile devices, and there are existing works focusing on human activity recognition or human interaction detection based on sensor signals captured by mobile devices, especially smartphones [20]–[26]. In industry or academia, ear-worn devices have been a promising wearable technology for areas like speech enhancement [3], [4], eating episodes detection [1], respiration rate measurement [2]. Placed on the ear, earbuds equipped with the accelerometer and gyroscope enable the motion tracking of the body while they are less susceptible to motion disturbance as the upper body acts as a natural filter. They are widely utilized in the areas like human body movement classification or detection [5]–[7], [27], [28]. Furthermore, compared to traditional IMU sensors placed in other positions, earables can sense subtle head or facial movements [29]–[31].

Domain adaptation, a specialized branch of transfer learning, aims to train a model using the data of the source domain and achieve high accuracy on a target dataset, which differs significantly from the source. Domain adaptation has been studied in the realm of sensing systems. It has been proven to be effective in addressing the domain shift issue in wireless sensing [32]–[41] and wearable sensing [42]–[44]. Few attempts have been made to utilize domain adaptation in earable sensing. Nguyen et al. [45] addressed the lack of publicly available datasets in Respiratory Symptom Detection using earables. A domain adaptation layer was implemented in [46] to guide the model to learn common feature representations by aligning the distribution of both the source and target domains. Zhang et al. [47] sought to use domain adaptation to minimize the training overhead while maximizing the model's adaptability to individual user characteristics.

VI. Conclusion

In this work, we introduce *EarDA*, a novel system for earable-based human activity recognition (HAR) that effectively addresses the challenges of domain adaptation and mitigates the impact of head movements. Our system leverages the rich datasets available from smartphone-based IMU sensors and adapts them for use with COTS earable devices, which are becoming increasingly prevalent. Through extensive experiments, we have demonstrated that the system can improve the classification accuracy to 88.8% in activity recognition tasks, even with limited earable training data. The system's performance highlights the potential of domain adaptation in the realm of earable sensing and sets the stage for future work in this area.

Acknowledgment

The work was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 21201420 and CityU 11201422), the Innovation and Technology Commission of Hong Kong (Project No. PRP/037/23FX and MHP/072/23), NSF of Shandong Province (Project No. ZR2021LZH010).

References

[1] A. Bedri, R. Li, M. Haynes, R. P. Kosaraju, I. Grover, T. Prioleau, M. Y. Beh, M. Goel, T. Starner, and G. Abowd, "Earbit: using wearable sensors to detect eating episodes in unconstrained environments," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 1, no. 3, pp. 1–20, 2017.

[2] T. Röddiger, D. Wolffram, D. Laubenstein, M. Budde, and M. Beigl, "Towards respiration rate monitoring using an in-ear headphone inertial measurement unit," in *Proceedings of the 1st International Workshop on Earable Computing*, 2019, pp. 48–53.

[3] D. Duan, Y. Chen, W. Xu, and T. Li, "Earse: Bringing robust speech enhancement to cots headphones," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 7, no. 4, pp. 1–33, 2024.

[4] L. He, H. Hou, S. Shi, X. Shuai, and Z. Yan, "Towards bone-conducted vibration speech enhancement on head-mounted wearables," in *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*, 2023, pp. 14–27.

[5] T. Hossain, M. S. Islam, M. A. R. Ahad, and S. Inoue, "Human activity recognition using earable device," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 81–84.

[6] F. Kawsar, C. Min, A. Mathur, and A. Montanari, "Earables for personal-scale behavior analytics," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 83–89, 2018.

[7] C. P. Burgos, L. Gärtner, M. A. G. Ballester, J. Noailly, F. Stöcker, M. Schönfelder, T. Adams, and S. Tassani, "In-ear accelerometer-based sensor for gait classification," *IEEE Sensors Journal*, vol. 20, no. 21, pp. 12 895–12 902, 2020.

[8] F. Kawsar, C. Min, A. Mathur, A. Montanari, U. G. Acer, and M. Van den Broeck, "esense: Open earable platform for human sensing," in *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, 2018, pp. 371–372.

[9] T. Röddiger, T. King, D. R. Roodt, C. Clarke, and M. Beigl, "Openearable: Open hardware earable sensing platform," in *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, 2022, pp. 246–251.

[10] G. Li, Z. Cao, and T. Li, "Echoattack: Practical inaudible attacks to smart earbuds," in *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*, 2023, pp. 383–396.

[11] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International conference on machine learning*. PMLR, 2015, pp. 1180–1189.

[12] E. K. Antonsson and R. W. Mann, "The frequency content of gait," *Journal of biomechanics*, vol. 18, no. 1, pp. 39–47, 1985.

[13] C. N. Rinaudo, M. C. Schubert, W. V. Figtree, C. J. Todd, and A. A. Migliaccio, "Human vestibulo-ocular reflex adaptation is frequency selective," *Journal of neurophysiology*, 2019.

[14] A. Saeed, T. Ozcelebi, and J. Lukkien, "Multi-task self-supervised learning for human activity detection," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019.

[15] H. Xu, P. Zhou, R. Tan, M. Li, and G. Shen, "Limu-bert: Unleashing the potential of unlabeled data for imu sensing applications," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 220–233.

[16] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Mobile sensor data anonymization," in *Proceedings of the international conference on internet of things design and implementation*, 2019, pp. 49–58.

[17] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen, "Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition," in *Proceedings of the 13th ACM conference on embedded networked sensor systems*, 2015, pp. 127–140.

[18] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, 2016.

[19] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," *Sensors*, vol. 14, no. 6, pp. 10 146–10 176, 2014.

[20] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 1307–1310.

[21] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "Deepsense: A unified deep learning framework for time-series mobile sensing data processing," in *Proceedings of the 26th international conference on world wide web*, 2017, pp. 351–360.

[22] S. Liu, S. Yao, J. Li, D. Liu, T. Wang, H. Shao, and T. Abdelzaher, "Giobalfusion: A global attentional deep learning framework for multisensor information fusion," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020.

[23] T. Ni, X. Zhang, C. Zuo, J. Li, Z. Yan, W. Wang, W. Xu, X. Luo, and Q. Zhao, "Uncovering user interactions on smartphones via contactless wireless charging side channels," in *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2023, pp. 3399–3415.

[24] Y. Chen, T. Ni, W. Xu, and T. Gu, "Swipepass: Acoustic-based second-factor user authentication for smartphones," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, pp. 1–25, 2022.

[25] T. Ni, X. Zhang, and Q. Zhao, "Recovering fingerprints from in-display fingerprint sensors via electromagnetic side channel," in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2023, pp. 253–267.

[26] T. Ni, Y. Chen, W. Xu, L. Xue, and Q. Zhao, "Xporter: A study of the multi-port charger security on privacy leakage and voice injection," in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, 2023, pp. 1–15.

[27] C. Min, A. Mathur, and F. Kawsar, "Exploring audio and kinetic sensing on earable devices," in *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*, 2018, pp. 5–10.

[28] M. Laporte, P. Baglat, S. Gashi, M. Gjoreski, S. Santini, and M. Langheinrich, "Detecting verbal and non-verbal gestures using earables," in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021.

[29] P. Zhu, Y. Zou, W. Li, and K. Wu, "Char: Composite head-body activities recognition with a single earable device," in *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2023, pp. 212–221.

[30] K. Li, R. Zhang, B. Liang, F. Guimbretière, and C. Zhang, "Eario: A low-power acoustic sensing earable for continuously tracking detailed facial movements," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1–24, 2022.

[31] S. Gashi, A. Saeed, A. Vicini, E. Di Lascio, and S. Santini, "Hierarchical classification and transfer learning to recognize head gestures and facial expressions using earbuds," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, 2021, pp. 168–176.

[32] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *Proceedings of the 17th annual international conference on mobile systems, applications, and services*, 2019, pp. 313–325.

[33] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas *et al.*, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th annual international conference on mobile computing and networking*, 2018, pp. 289–304.

[34] T. Gong, Y. Kim, J. Shin, and S.-J. Lee, "Metasense: few-shot adaptation to untrained conditions in deep mobile sensing," in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, 2019.

[35] T. Ni, J. Li, X. Zhang, C. Zuo, W. Wang, W. Xu, X. Luo, and Q. Zhao, "Exploiting contactless side channels in wireless charging power banks for user privacy inference via few-shot learning," in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, 2023.

[36] T. Ni, Y. Chen, K. Song, and W. Xu, "A simple and fast human activity recognition system using radio frequency energy harvesting," in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021, pp. 666–671.

[37] Z. Sun, T. Ni, H. Yang, K. Liu, Y. Zhang, T. Gu, and W. Xu, "Flora: Energy-efficient, reliable, and beamforming-assisted over-the-air firmware update in lora networks," in *Proceedings of the 22nd International Conference on Information Processing in Sensor Networks*, 2023, pp. 14–26.

[38] ——, "Flora+: Energy-efficient, reliable, beamforming-assisted, and secure over-the-air firmware update in lora networks," *ACM Transactions on Sensor Networks*, 2024.

[39] M. Han, H. Yang, T. Ni, D. Duan, M. Ruan, Y. Chen, J. Zhang, and W. Xu, "mmsign: mmwave-based few-shot online handwritten signature verification," *ACM Transactions on Sensor Networks*, 2023.

[40] T. Ni, G. Lan, J. Wang, Q. Zhao, and W. Xu, "Eavesdropping mobile app activity via {Radio-Frequency} energy harvesting," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 3511–3528.

[41] Z. Sun, T. Ni, Y. Chen, D. Duan, K. Liu, and W. Xu, "Rf-egg: An rf solution for fine-grained multi-target and multi-task egg incubation sensing," in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, 2024.

[42] A. Natarajan, G. Angarita, E. Gaiser, R. Malison, D. Ganesan, and B. M. Marlin, "Domain adaptation methods for improving lab-to-field generalization of cocaine detection using wearable ecg," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 875–885.

[43] A. Akbari and R. Jafari, "Transferring activity recognition models for new wearable sensors with deep generative domain adaptation," in *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*, 2019, pp. 85–96.

[44] D. Duan, H. Yang, G. Lan, T. Li, X. Jia, and W. Xu, "Emgsense: A low-effort self-supervised domain adaptation framework for emg sensing," in *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2023, pp. 160–170.

[45] N. Nguyen, A. Chakma, and N. Roy, "A scalable and domain adaptive respiratory symptoms detection framework using earables," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021.

[46] H. Li, Y. Zhang, J. Han, Y. Yan, Y. Liu, and H. Yang, "Adapsqa: Adaptive ecg signal quality assessment model for inter-patient paradigm using unsupervised domain adaptation," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2022.

[47] S. Zhang, T. Lu, H. Zhou, Y. Liu, R. Liu, and M. Gowda, "I am an earphone and i can hear my user's face: Facial landmark tracking using smart earphones," *ACM Transactions on Internet of Things*, vol. 5, no. 1, pp. 1–29, 2023.