2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-8622-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/BIBM62325.2024.10821840

Adaptive Domain-Adversarial Multi-Instance Learning for Wearable-Sensor-Based Parkinson's Disease Severity Assessment

Zheyuan Xu
National Pilot School of Software
Yunnan University
Kunming, China
xuzy07122022@gmail.com

Xulong Wang
Department of Computer Science
University of Sheffield
Sheffield, United Kingdom
xl.wang@sheffield.ac.uk

Menghui Zhou
Department of Computer Science
University of Sheffield
Sheffield, United Kingdom
menghui.zhou@sheffield.ac.uk

Jun Qi
Department of Computing
Xi'an JiaoTong-Liverpool University
Suzhou, China
jun.qi@xjtlu.edu.cn

Yun Yang
National Pilot School of Software
Yunnan University
Kunming, China
yangyun@ynu.edu.cn

Po Yang*
Department of Computer Science
University of Sheffield
Sheffield, United Kingdom
po.yang@sheffield.ac.uk

***Abstract*—Wearable sensors combined with machine learning provide an effective solution for assessing Parkinson's Disease (PD) severity. However, time-series data from wearable sensors often lack window-level labels for PD severity, resulting in weak supervision, which introduces the challenge of label noise. Additionally, patient variability causes distributional discrepancies, further complicating the learning process. To address these issues, we propose Adaptive Domain-Adversarial Multi-Instance Learning (ADAMIL), which combines and refines Multiple-Instance Learning (MIL) with domain-adversarial techniques. We improve traditional MIL by incorporating self-attention mechanisms and learnable positional encoding, enabling ADAMIL to capture temporal dependencies more effectively, thus making it better suited for mitigating label noise in weakly supervised time-series data. Furthermore, ADAMIL refines domain-adversarial learning to autonomously align latent distributions, ensuring robust domain-invariant feature learning without relying on predefined labels. Experimental results show that ADAMIL achieves 85.29% accuracy and 80.57% F1-score in fine-grained PD severity classification, outperforming existing methods. Notably, this performance is achieved using only a single wrist-worn sensor, underscoring its potential for practical use in clinical and home settings. The code is available at https://github.com/xzxzy12345XZY/ADAMIL.**

***Index Terms*—Parkinson's Disease, Domain-Adversarial Learning, Multi-Instance Learning, Wearable Sensors.**

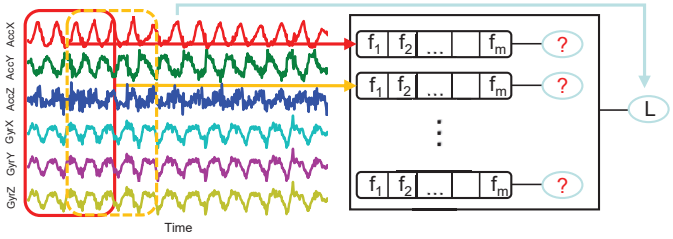
I. INTRODUCTION

Parkinson's Disease (PD) is a common neurodegenerative disorder characterized by a combination of motor and non-motor symptoms, such as tremors, bradykinesia, rigidity, and sleep disturbances [1]. It is the second most prevalent neurodegenerative disease, affecting 2-3% of individuals over the age of 65 [2]. Although there is currently no cure for PD, early diagnosis and timely intervention are crucial for alleviating symptoms and slowing disease progression [3].

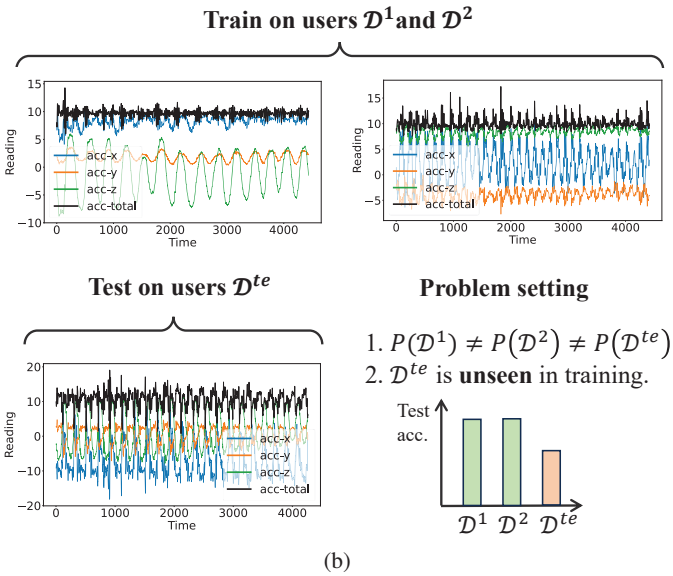
Traditionally, The diagnosis of Parkinson's Disease (PD) primarily relies on the Movement Disorder Society's revised Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [4], a widely recognized clinical assessment tool that is quite time-consuming. Additionally, PD diagnosis using specialized imaging techniques like PET and MRI involves high costs. Recently, methods combining wearable technology with machine learning have shown significant potential for the quantitative assessment of PD [5]–[10]. Machine learning models like Support Vector Machines (SVM) and LightGBM (LGBM), and deep learning models such as CNN-LSTM (CNNLSTM) and LSTM-CNN (LSTMCNN), have achieved some success in early PD detection and classification of motor symptom severity [11]–[13]. However, the practical application of models for Parkinson's Disease (PD) severity assessment based on wearable sensor data faces two major challenges: label noise and distributional differences [14], [15].

In evaluating Parkinson's Disease (PD) using sensor data, label noise remains a significant challenge. Most existing methods use a sliding window to segment time-series data, assigning uniform, coarse-grained labels to each segment (see Fig. 1a). This ignores the dynamic nature of PD symptoms, resulting in mismatched instances and labels, thereby introducing substantial noise [15]. As shown in Fig. 1a, not every segment consistently reflects the severity level, which exacerbates label inconsistency.

To address this issue, the Multiple-Instance Learning (MIL) has been introduced. In the MIL, a patient's activity session is treated as a 'bag' that contains multiple sliding window instances, with the label for the bag representing the overall severity of the session. The advantage of this approach is that it does not require precise labeling for each individual instance but instead uses bag-level labels to guide the learning process, thereby effectively handling weak supervision and label uncertainty.



(a)



(b)

Fig. 1: sensor signals from patients with moderate Parkinson's disease. (a) Feature extraction and label propagation. (b) Accelerometer signals from three patients with moderate Parkinson's disease during right-hand flips show significant variation.

However, the standard Multiple-Instance Learning (MIL) has several limitations [15]–[20]. It assumes that instances are independently and identically distributed (i.i.d.) and possess permutation-invariant properties, which prevents it from capturing temporal correlations and sequential dependencies between instances. In multivariate time-series classification (MTSC), instances often exhibit temporal correlations and sequential dependencies, posing a significant challenge to the direct application of the standard MIL. Therefore, we propose a general MIL mechanism for time-series data that integrates a self-attention mechanism and learnable positional encoding. This approach effectively captures the temporal correlations and sequential dependencies among instances, accurately represents contextual information to obtain bag embeddings, and avoids the need for label propagation, thereby mitigating label noise.

Aside from its flexibility, MIL, as an extension of traditional supervised learning, also inherits some of its limitations. Notably, the performance of MIL tends to degrade significantly when there is a distributional mismatch between the training bags and the testing bags [14]. In practical applications of Parkinson's Disease (PD) assessment, sensor signals are influenced by individual patient differences. PD symptoms vary widely, and even patients with the same severity level can exhibit significant variability in sensor signals (as shown in Fig. 1b), indicating the presence of multiple underlying data distributions. Traditional models often assume independent and identically distributed (i.i.d.) data, which limits their ability to generalize effectively in the presence of out-of-distribution (OOD) scenarios [14].

To address the issue of data distribution differences, Domain-Adversarial Learning (DAL) methods have been proposed in recent years. A typical example is the Domain-Adversarial Neural Network (DANN) [21]–[29], which employs adversarial training to enable the model to learn domain-invariant features between the source domain (training set) and the target domain (test set), thereby reducing distributional discrepancies and enhancing the model's generalization capability. However, methods like DANN typically require known domain labels, which are often difficult to obtain in practice.

To overcome these challenges, we propose a domain-adversarial learning approach capable of automatically identifying latent distributions. By leveraging iterative adversarial training and unsupervised clustering, the model can learn domain-invariant features without predefined domain labels. This approach not only effectively reduces the distributional discrepancy between the source and target domains but also improves adaptability across different patients, thereby enhancing generalization capability.

To simultaneously address the two major challenges of label noise and distributional differences, this paper proposes an innovative approach that integrates a multi-instance learning mechanism with a domain-adversarial learning method capable of automatically identifying distributions. This approach is named ADAMIL (Adaptive Domain-Adversarial Multi-Instance Learning). ADAMIL offers the following advantages:

1) **Label Noise and Weak Supervision:** Traditional methods suffer from significant label noise due to assigning uniform coarse-grained labels to sliding windows, often misrepresenting segment conditions. ADAMIL addresses this by directly learning from bag-level data under weak supervision. Leveraging self-attention and learnable positional encoding, ADAMIL captures temporal order and correlations, enhancing bag-level feature representation and reducing label noise.
2) **Distributional Differences:** ADAMIL directly addresses distributional differences by automatically identifying latent distributions and learning domain-invariant features without relying on predefined domain labels, thereby enhancing model robustness and generalization.

Our experiments on a private wearable sensor dataset and a public wearable dataset demonstrate that the proposed ADAMIL method outperforms existing models in fine-grained PD severity classification and PD vs. differential diagnosis (DD) classification tasks, in terms of both accuracy and stability. To summarize, our contributions are as follows:

- **Novel Perspective:** We identify two major challenges in wearable sensor-based Parkinson's Disease (PD) assessment: label noise and distributional differences. Traditional methods struggle with these issues, which hinder accurate severity evaluation and underscore the need for more advanced approaches.
- **Innovative Method:** We propose ADAMIL (Adaptive Domain-Adversarial Multi-Instance Learning), an end-to-end approach that combines multi-instance learning with domain-adversarial techniques. ADAMIL utilizes self-attention and learnable positional encoding to effectively capture temporal correlations, while automatically identifying latent distributions to learn domain-invariant features, enhancing robustness without relying on predefined domain labels.
- **Empirical Validation:** We validate ADAMIL through extensive experiments on private and public datasets for fine-grained severity classification and PD vs. differential diagnosis. Results show that ADAMIL achieves superior accuracy and stability, highlighting its potential for real-world clinical applications.

II. PROPOSED METHOD

In this section, we describe how the proposed ADAMIL method is applied to the task of assessing Parkinson's disease severity. First, in Section A, we define the two main challenges faced by this task: label noise and distributional differences. Next, in Sections B and C, we introduce an innovative Multiple-Instance Learning (MIL) method to address label noise, detailing the data preprocessing and feature extraction strategies. Then, in Sections D, E, F, and G, we present a novel Domain-Adversarial Learning (DAL) approach, which aims to both model and automatically identify latent distributions, aligning them to address distributional differences across domains. The proposed ADAMIL method is illustrated in Figure 3, and the complete training and inference procedure is summarized in Algorithm 1.

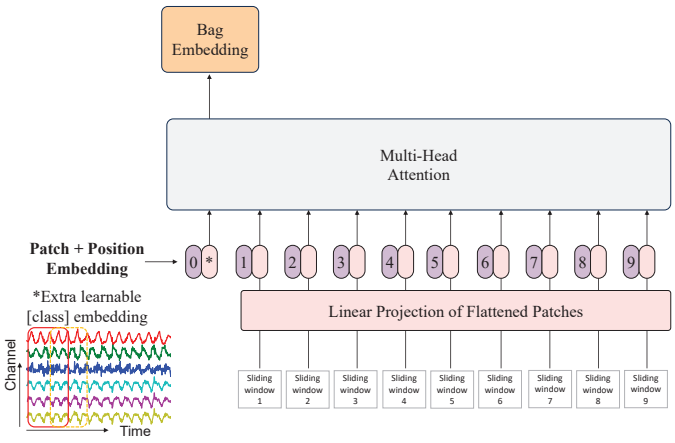


Fig. 2: Multi-instance learning feature extractor incorporating learnable positional encoding and self-attention mechanisms.

A. Problem Formulation

The sensor time series is represented as $X_1, \ldots, X_n$, where $X_i = x_i^1, \ldots, x_i^T$ consists of T sliding windows. Each $x_i^t \in \mathbb{R}^{6\times d}$ represents tri-axial acceleration and angular velocity over d time points. The label for each sequence X_i is denoted as $y_i \in \mathcal{Y} = 1, \ldots, C$. Our objective is to learn a mapping from the feature space $\mathcal{X}$ to the label space $\mathcal{Y}$ using the training dataset $D^{\text{tr}} = (X_1, y_1), \ldots, (X_n, y_n)$, effectively approximating $\mathbb{P}^*(y \mid X)$ for accurate PD severity prediction.

This task presents two major challenges. First, traditional methods often use label propagation to map sliding window segments to coarse-grained labels, assuming uniform conditions across all instances in a bag. However, the dynamic nature of Parkinson's Disease (PD) symptoms makes this assumption unrealistic, leading to significant label noise and reducing the reliability of learned features. Second, distributional shift between training and testing data ($\mathbb{P}^{\text{tr}}(X) \neq \mathbb{P}^{\text{te}}(X)$) further complicates model generalization. Such shifts can result from differences in patient behaviors, sensor placements, data acquisition conditions, or the complex manifestations of Parkinson's Disease symptoms, thereby impacting the model's ability to generalize to unseen data.

Algorithm 1 ADAMIL

Training stage:
Input: Training dataset $D^{\text{tr}} = \{X_i, y_i\}_{i=1}^{n_{\text{tr}}}$.
Model: ADAMIL h.
Pre-process: Segment signals using sliding window, followed by filtering and normalization.

for $epoch = 1$ to $epochs$ **do**
  for $local\_epoch = 1$ to $local\_epochs$ **do**
    Forward propagation to calculate Eq. (3).
    Backward propagation to update $\theta_h^{(2)}$.
  end for
  for $local\_epoch = 1$ to $local\_epochs$ **do**
    Forward propagation to calculate Eq. (4).
    Backward propagation to update $\theta_h^{(3)}$.
  end for
  Calculate the initial centroids $\tilde{\mu}_k$ and pseudo-domain labels $\tilde{d}_i'$ by Eq. (5).
  for $iteration = 1$ to $iterations$ **do**
    Update cluster centers μ_k and pseudo-domain labels d_i' by Eq. (6).
  end for
  for $local\_epoch = 1$ to $local\_epochs$ **do**
    Forward propagation to calculate Eq. (7).
    Backward propagation to update $\theta_h^{(4)}$.
  end for
end for

Return: The optimal ADAMIL model h.

Inference stage:
Input: Test dataset $D^{\text{te}} = \{X_i, y_i\}_{i=1}^{n_{\text{te}}}$.
Model: The optimal ADAMIL model h.
Output: Final diagnosis decisions by Eq. (8).

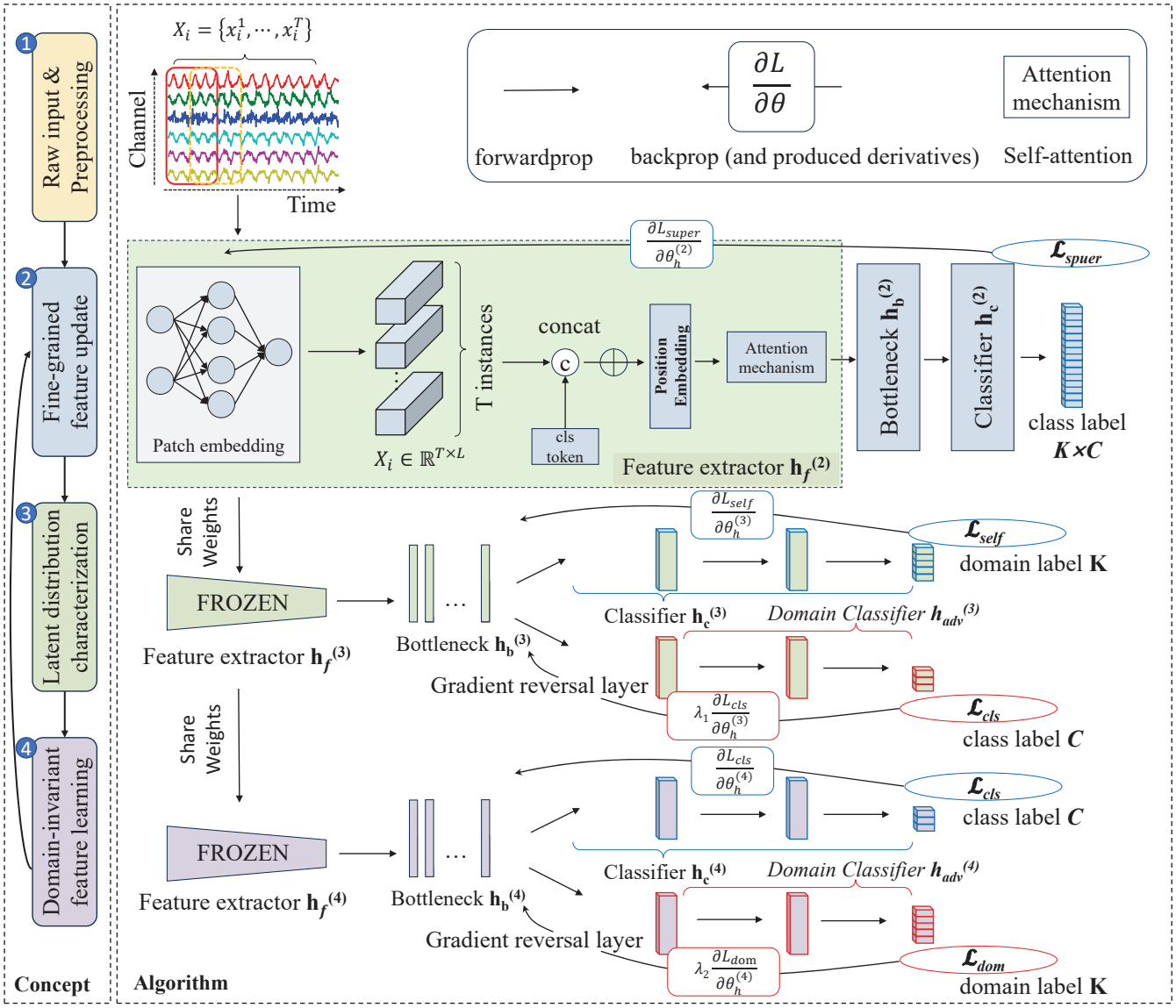


Fig. 3: Workflow of the ADAMIL method.

B. Data Preprocessing

We apply a band-pass filter (0.3 to 20 Hz) and Z-score standardization to reduce noise in sensor data. The time-series data is segmented using a sliding window (1-second length, 50% overlap) [30], resulting in the training set D^{tr}. This segmentation ensures overlapping segments, providing sufficient temporal context for feature extraction.

C. Multi-Instance Learning for Feature Extraction

Our method, illustrated in Figure 2, begins by segmenting sensor data into T sliding windows. Each window is linearly projected into a $1 \times L$ patch embedding to extract local features.

To aggregate global information, we introduce a learnable `cls token`, inspired by the [CLS] token in ViT [31], which adaptively integrates features from all patches. A learnable positional encoding is added to both the `cls token` and patch embeddings, enabling the model to capture temporal order, addressing traditional MIL's lack of sequential awareness.

All embeddings are then processed by a multi-head self-attention mechanism, allowing the `cls token` to interact with each patch and aggregate global sequence information. The final output of the `cls token` forms the bag embedding, representing the entire time series by integrating both local and temporal relationships.

This design enhances conventional MIL by effectively capturing temporal dependencies under weak supervision. By leveraging self-attention and learnable positional encoding, our innovative approach allows multi-instance learning to extract meaningful temporal patterns without ignoring sequence in-

formation, even with coarse-grained labels.

$$\mathbf{p}_i^t = f_{\text{patch}}(x_i^t), \quad t = 1, 2, \ldots, T \tag{1}$$

$$X_i^{\text{bag embedding}} = h_f^{(2)}\Big([\text{CLS}] + \text{PE}_0, \sum_{t=1}^{T}(\mathbf{p}_i^t + \text{PE}_t)\Big) \tag{2}$$

D. Fine-grained Feature Update

Before describing the latent distributions, we perform a fine-grained feature update to capture information that mixes both domain and class characteristics. As illustrated Fig. 3 (blue), we introduce the concept of "pseudo-domain-class labels" to expand the label space, increasing the original number of labels to $K \times C$, where K represents the number of latent distributions and C denotes the number of classes. This allows us to fully exploit the knowledge embedded within both domains and classes. During the first iteration, we assign a domain label $d' = 0$ to each sample and consider each class within each domain as a new class $s \in \{1, 2, \ldots, s\}$. We then perform pseudo-domain-class label assignment to obtain discrete values for supervision: $s = d' \times C + y$.

Let $h_f^{(2)}, h_b^{(2)}, h_c^{(2)}$ represent the feature extractor, bottleneck, and classifier, respectively, where the superscript 2 indicates the current designated stage. The supervised loss is then computed using cross-entropy:

$$\mathcal{L}_{\text{super}} = \mathbb{E}_{(X,y)\sim\mathbb{P}^{\text{tr}}}\ell\left(h_c^{(2)}(h_b^{(2)}(h_f^{(2)}(X))), s\right). \tag{3}$$

E. Latent Distribution Representation

This step involves characterizing latent distributions in the dataset and assigning domain labels d' to each sample. As shown in Fig. 3 (green), we utilize the feature extractor from Step 2 with fixed parameters. To obtain class-independent features, we employ a modified domain-adversarial training strategy that uses class labels to guide the extraction of domain-specific features while minimizing class dependency. Subsequently, a self-supervised pseudo-labeling approach is applied to derive domain labels for each sample based on the learned class-independent representation.

First, we train the model using the feature extractor $h_f^{(3)}$, the bottleneck $h_b^{(3)}$, and the classifier $h_c^{(3)}$, where the superscript 3 indicates the current stage. In the first iteration, we set $d' = 0$.

$$\begin{aligned}\mathcal{L}_{\text{self}} + \mathcal{L}_{\text{cls}} = \mathbb{E}_{(X,y)\sim\mathbb{P}^{\text{tr}}}\Big[&\ell(h_c^{(3)}(h_b^{(3)}(h_f^{(3)}(X))), d') \\ &+ \ell(h_{\text{adv}}^{(3)}(R_{\lambda_1}(h_b^{(3)}(h_f^{(3)}(X)))), y)\Big]\end{aligned} \tag{4}$$

To learn class-invariant features, we use the Gradient Reversal Layer (GRL) from the Domain-Adversarial Neural Network (DANN) [21]. The GRL leaves the input unchanged during forward propagation but multiplies the gradient by a negative scalar $-\lambda_1$ during backpropagation. This encourages the bottleneck $h_b^{(3)}$ to capture features that minimize class differences, ensuring class invariance.

Next, we calculate the centroid for each domain with class-invariant features. The symbol $\tilde{\mu}_k$ denotes the initial centroid of the k-th latent domain, while δ_k is the k-th element of the logit soft-max output. We then obtain the pseudo domain labels using a nearest centroid classifier with a distance function D:

$$\begin{aligned}\tilde{\mu}_k &= \frac{\sum_{X_i\in\mathcal{X}^{\text{tr}}}\delta_k(h_c^{(3)}(h_b^{(3)}(h_f^{(3)}(X_i))))\, h_b^{(3)}(h_f^{(3)}(X_i))}{\sum_{X_i\in\mathcal{X}^{\text{tr}}}\delta_k(h_c^{(3)}(h_b^{(3)}(h_f^{(3)}(X_i))))}, \\ \tilde{d}_i' &= \arg\min_k D(h_b^{(3)}(h_f^{(3)}(X_i)), \tilde{\mu}_k)\end{aligned} \tag{5}$$

Finally, after obtaining $\tilde{d}_i'$, we iteratively update the cluster centers μ_k and pseudo-domain labels $\tilde{d}_i'$ until convergence or the maximum number of iterations is reached, where $\mathbb{I}(a) = 1$ when a is true, otherwise 0:

$$\begin{aligned}\mu_k &= \frac{\sum_{X_i\in\mathcal{X}^{\text{tr}}}\mathbb{I}(\tilde{d}_i' = k)\, h_b^{(3)}(h_f^{(3)}(X_i))}{\sum_{X_i\in\mathcal{X}^{\text{tr}}}\mathbb{I}(\tilde{d}_i' = k)}, \\ d_i' &= \arg\min_k D(h_b^{(3)}(h_f^{(3)}(X_i)), \mu_k)\end{aligned} \tag{6}$$

F. Domain-Invariant Representation Learning

After obtaining the latent distributions, we learn domain-invariant representations for generalization.

$$\begin{aligned}\mathcal{L}_{cls} + \mathcal{L}_{dom} = \mathbb{E}_{(X,y)\sim\mathbb{P}^{\text{tr}}}\Big[&\ell(h_c^{(4)}(h_b^{(4)}(h_f^{(4)}(X))), y) \\ &+ \ell(h_{adv}^{(4)}(R_{\lambda_2}(h_b^{(4)}(h_f^{(4)}(X)))), d')\Big],\end{aligned} \tag{7}$$

This step (purple in Fig. 3) differs from Step 3 by employing a GRL with negative scalar $lambda_2$ to learn domain-invariant features. The GRL directs the bottleneck $h_b^{(4)}$ to minimize domain discrepancies, thereby enhancing generalization across domains. The domain classifier is optimized with domain labels to enforce domain invariance, while the class classifier uses class labels to ensure precise disease severity classification. Both domain loss L_{dom} and classification loss L_{cls} are updated to improve domain alignment.

G. Training and Inference

During the training phase, Steps 2-4 are repeated until convergence or the maximum number of iterations is reached. Note that $h_f^{(3)}$, $h_f^{(4)}$, and $h_c^{(2)}$ represent the same feature extractor, with its parameters being updated only in Step 2. The parameters of the feature extractor module are shared across the various stages to ensure consistency.

For inference, the final patient label is predicted using the bag embedding derived from the entire sequence. Specifically, the trained feature extractor $h_f^{(4)}$, followed by the bottleneck $h_b^{(4)}$ and the classifier $h_c^{(4)}$, is used to directly produce the patient-level prediction:

$$\hat{y}_{\text{patient}} = h_c^{(4)}(h_b^{(4)}(h_f^{(4)}(X_i))) \tag{8}$$

where X_i represents the entire temporal sequence of activities for the i-th patient. This approach directly utilizes the full temporal data of the patient's activities to predict Parkinson's disease severity, avoiding the need for individual window-level aggregation.

III. EXPERIMENTS

A. Datasets

This study uses two datasets to assess Parkinson's Disease (PD) severity: a private dataset collected by our team and the publicly available Parkinson's Disease Smartwatch (PADS) dataset [32].

Private Dataset: The private dataset was collected using Shimmer devices (200 Hz) worn on the right wrist, capturing three-axis acceleration and angular velocity. It includes 35 healthy controls (HC), 63 mild, 21 moderate, and 16 severe PD cases. Data was collected across 16 distinct activities, as detailed in Table I.

PADS [32]: The PADS dataset was collected using Apple Watch Series 4 (100 Hz) worn on the right wrist, capturing similar sensor data. It includes 276 PD and 114 differential diagnosis (DD) cases, involving 14 activities, as detailed in Table II.

TABLE I: ACTIVITY SETUP IN PRIVATE DATASET

Num	Activity Name	Activity Description
1	FingerTap	Tap thumb with index finger in rapid succession.
2	HandOpen	Open and close the hand repeatedly.
3	WristRotation	Rotate wrist by turning palms up and down.
4	RHandFlip	The palm and back of the right hand alternately slap left palm.
5	LHandFlip	The palm and back of the left hand alternately slap left palm.
6	FingerNoseL	Touch the nose with the index finger of the left hand, then touch the doctor's index finger 30 cm from the patient.
7	FingerNoseR	Touch the nose with the index finger of the right hand, then touch the doctor's index finger 30 cm from the patient.
8	HandRaise	Raise hands to shoulder height for 20 seconds.
9	WalkBack	Walk straight for 10 m, turn around, and walk back.
10	SitStand	Stand up from a chair while keeping arms crossed.
11	DrinkWater	Pick up a paper cup from a table and drink from it.
12	PickUpThings	Pick up items from the ground.
13	Sit	Sit in a chair in a natural posture.
14	Stand	Stand naturally with arms positioned at sides.
15	ArmSwing	Swing arms back and forth using the shoulder joint.
16	DrawingSpirals	Draw four turns of a 12 cm spiral on paper.

TABLE II: ACTIVITY SETUP IN PADS [32]

Num	Activity Name	Activity Description
1	Relaxed1	Resting with eyes closed while sitting (first half).
2	Relaxed2	Resting with eyes closed while sitting (second half).
3	RelaxedTask1	Resting while patient is calculating serial sevens (first half).
4	RelaxedTask2	Resting while patient is calculating serial sevens (second half).
5	StretchHold	Lift and extend arms, then hold.
6	LiftHold	Remain arms lifted.
7	HoldWeight	Hold a weight, alternating arms.
8	PointFinger	Point index finger at examiner's hand, repeat for both sides.
9	DrinkGlas	Simulate drinking from an empty glass, alternating hands.
10	CrossArms	Cross and extend both arms.
11	TouchIndex	Bring both index fingers to each other.
12	TouchNose	Tap your nose alternately, then extend your arms.
13	Entrainment1	Foot stomping at the pace set by the examiner (first half).
14	Entrainment2	Foot stomping at the pace set by the examiner (second half).

B. Experimental setup

Experiments are conducted using a 4-fold cross-validation strategy, where participants from each class are evenly divided, with three groups used for training and one for testing. The Adam optimizer is used with a learning rate selected from 5×10^{-3} or 5×10^{-4}, and the predefined number of domain distributions K is chosen from the range (2, 23). All models are implemented in PyTorch and trained on an NVIDIA RTX 3090 GPU, with comparative methods adhering to their original configurations.

C. Experimental results

Table III and IV compare the performance of our proposed ADAMIL method with state-of-the-art models, including the handcrafted feature-based LGBM [11] and deep learning models such as CNN-LSTM [12] and LSTM-CNN [13]. The evaluation metrics—accuracy, macro-precision, macro-recall, and macro-F1—are applied across four classification tasks: (i) HC/PD/Mild/Moderate/Severe and (ii) PD vs. differential diagnosis (DD) tasks, representing comprehensive scenarios in Parkinson's disease assessment.

Unlike traditional models that use sliding windows paired with coarse-grained labels—often leading to label noise—and assume independent and identically distributed (i.i.d.) data, ADAMIL is specifically designed to learn under weak supervision from coarse-grained labels, while automatically identifying latent distributions and learning domain-invariant features. This allows ADAMIL to handle non-stationary and complex sensor signals effectively, mitigating label noise and addressing distributional shifts.

The experimental results show that ADAMIL achieves superior performance across all tasks, with an accuracy of 85.29% and an F1 score of 80.57% in fine-grained PD severity classification, and an accuracy of 81.44% with an F1 score of 75.75% in PD vs. DD classification. These results demonstrate ADAMIL's effectiveness in capturing multiple latent distributions, handling significant data variability, and improving generalization and predictive accuracy compared to other models.

D. Feature visualization

Figures 4a and 4b show the feature visualizations for fine-grained severity classification. Figure 4a depicts features from traditional methods, with scattered and misaligned training and test samples, indicating poor generalization. In contrast, Figure 4b shows features from our model, where training and test samples are well-aligned and form distinct clusters, demonstrating enhanced consistency and generalization.

IV. CONCLUSION

In this paper, we presented ADAMIL, a novel method for assessing Parkinson's Disease (PD) severity and performing differential diagnosis (DD). ADAMIL overcomes the limitations of traditional methods by addressing label noise under weak supervision and mitigating distributional discrepancies due to patient variability. This is achieved through the integration of Multiple-Instance Learning (MIL) for noisy labels and domain-adversarial techniques for distributional differences.

ADAMIL employs self-attention and learnable positional encoding to model temporal dependencies in sensor data.

TABLE III: Model Performance Comparison for Fine-Grained PD Severity Assessment

HC vs. Mild vs. Moderate vs. Severe in Private Dataset																	
Num	Activity Name	LGBM[11]				CNNLSTM[12]				LSTMCNN[13]				Ours			
		Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
1	FingerTap	54.55	54.67	39.03	39.58	48.48	36.72	27.78	20.96	48.48	28.74	33.33	28.41	**66.67**	**66.29**	**68.19**	**66.36**
2	HandOpen	66.67	61.01	50.00	50.98	48.48	36.72	31.25	25.96	54.55	56.71	37.92	38.21	**66.67**	**69.52**	**61.39**	**63.67**
3	WristRotation	63.64	58.33	46.25	45.60	69.70	38.53	47.22	41.96	72.73	61.04	52.22	49.17	**69.70**	**67.08**	**66.53**	**66.62**
4	RHandFlip	54.55	27.84	37.78	31.77	48.48	24.54	36.67	29.11	51.52	24.18	38.33	29.57	**69.70**	**73.27**	**67.64**	**68.87**
5	LHandFlip	66.67	64.42	46.67	46.63	63.64	52.50	46.11	45.25	66.67	86.01	51.25	55.03	**72.73**	**74.05**	**62.36**	**65.65**
6	FingerNoseL	65.52	39.58	40.62	37.65	51.72	37.50	33.33	29.17	51.72	28.78	33.48	30.78	**65.52**	**69.09**	**72.32**	**70.00**
7	FingerNoseR	63.64	76.50	48.47	52.50	60.61	42.45	43.33	41.87	69.70	51.67	52.78	51.32	**75.76**	**88.47**	**65.42**	**68.78**
8	HandRaise	46.88	11.72	25.00	15.96	50.00	27.50	36.88	29.17	56.25	63.39	40.62	39.66	**68.75**	**67.95**	**65.42**	**64.58**
9	WalkBack	45.45	11.36	25.00	15.62	54.55	62.50	41.25	40.95	54.55	62.50	41.25	40.95	**78.79**	**89.38**	**71.53**	**76.30**
10	SitStand	56.67	29.00	33.33	28.66	63.33	80.61	51.88	56.13	53.33	26.04	48.21	33.77	**76.67**	**81.61**	**76.98**	**77.85**
11	DrinkWater	61.76	46.27	49.48	47.14	73.53	66.00	53.47	53.04	76.47	66.67	59.72	60.20	**85.29**	**94.05**	**74.72**	**80.57**
12	PickUpThings	57.14	43.71	45.83	41.22	60.71	29.76	35.71	31.90	64.29	50.89	49.26	47.74	**82.14**	**90.55**	**75.30**	**80.97**
13	Sit	65.52	36.16	44.44	38.77	62.07	47.29	54.17	50.31	44.83	35.71	27.78	20.00	**65.52**	**69.55**	**68.33**	**68.78**
14	Stand	55.56	28.78	34.72	30.83	48.15	36.54	37.50	32.46	44.44	57.81	48.61	46.21	**74.07**	**79.55**	**65.97**	**69.59**
15	ArmSwing	61.54	55.73	49.62	50.00	57.69	37.50	36.11	32.05	57.69	29.97	37.63	33.33	**73.08**	**81.25**	**67.87**	**71.31**
16	DrawingSpirals	62.50	31.29	40.06	35.09	54.17	28.12	33.81	30.44	62.50	68.75	65.91	66.20	**70.83**	**86.46**	**69.60**	**71.38**

TABLE IV: Model Performance Comparison for PD vs. DD Classification

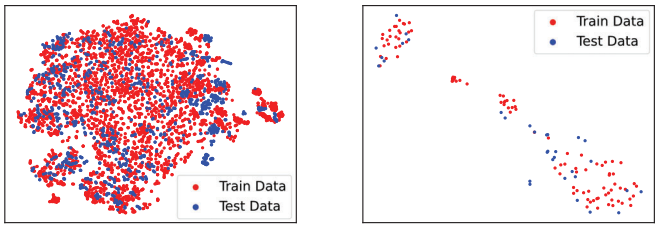
PD vs. DD in PADS [32]																	
Num	Activity Name	LGBM[11]				CNNLSTM[12]				LSTMCNN[13]				Ours			
		Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
1	Relaxed1	70.41	35.20	50.00	41.32	72.45	69.75	55.45	53.04	71.43	85.57	51.72	44.90	**80.61**	**82.74**	**69.24**	**71.70**
2	Relaxed2	72.45	73.67	54.45	50.81	74.49	71.12	60.89	61.45	77.55	73.28	70.06	71.21	**78.57**	**74.29**	**73.79**	**74.02**
3	RelaxedTask1	76.53	78.07	62.34	63.19	71.43	69.12	52.72	47.71	76.53	81.53	61.34	61.69	**76.53**	**71.79**	**69.34**	**70.26**
4	RelaxedTask2	76.53	87.50	60.34	60.00	70.41	35.20	50.00	41.32	72.45	73.67	54.45	50.81	**77.55**	**73.59**	**69.07**	**70.47**
5	StretchHold	71.13	61.02	52.12	47.61	77.32	87.91	60.71	60.77	74.23	86.70	55.36	52.01	**81.44**	**78.36**	**74.22**	**75.75**
6	LiftHold	71.13	35.57	50.00	41.57	71.13	35.57	50.00	41.57	75.26	75.14	59.27	59.07	**80.41**	**77.16**	**72.44**	**74.06**
7	HoldWeight	70.41	35.20	50.00	41.32	75.51	74.29	61.62	62.31	73.47	68.42	60.17	60.61	**78.57**	**74.43**	**75.79**	**75.01**
8	PointFinger	73.47	70.41	58.17	57.56	70.41	35.20	50.00	41.32	74.49	71.12	60.89	61.45	**78.57**	**76.11**	**68.79**	**70.61**
9	DrinkGlas	70.41	35.20	50.00	41.32	77.55	74.80	67.07	68.75	77.55	75.82	66.07	67.76	**80.61**	**77.81**	**73.24**	**74.82**
10	CrossArms	71.43	64.84	54.72	52.36	75.51	71.24	64.62	65.91	74.49	74.72	58.9	58.35	**80.61**	**79.51**	**71.24**	**73.41**
11	TouchIndex	71.13	35.57	50.00	41.57	72.16	69.50	52.85	48.17	73.20	86.32	53.57	48.74	**77.32**	**73.21**	**67.08**	**68.65**
12	TouchNose	73.20	68.33	56.75	55.66	74.23	68.86	60.66	61.34	77.32	73.21	67.08	68.65	**81.44**	**79.06**	**73.16**	**75.09**
13	Entrainment1	73.20	86.32	53.57	48.74	71.13	35.57	50.00	41.57	76.29	70.97	69.54	70.15	**80.41**	**82.19**	**68.19**	**70.62**
14	Entrainment2	74.23	76.96	56.42	54.36	71.13	35.57	50.00	41.57	74.23	72.94	57.48	56.42	**77.32**	**72.35**	**70.26**	**71.10**

Its domain-adversarial component autonomously aligns latent distributions, ensuring robust, domain-invariant feature learning without predefined labels. Experimental results show that ADAMIL outperforms existing methods in both fine-grained PD severity assessment and PD vs. DD classification, confirming its effectiveness.

Notably, ADAMIL achieves these results using only a single wrist-worn sensor, underscoring its efficiency, practicality, and suitability for applications in both clinical and home use. This highlights the practical viability and scalability of such methods in real-world scenarios, particularly in contexts where ease of deployment, cost-effectiveness, and minimal hardware requirements are critical. Future work will enhance the MIL component by incorporating domain-specific prior knowledge and automate the selection of latent distributions (K), improving adaptability and minimizing manual tuning for broader real-world applications.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (No. 62061050), the National Natural Science Foundation of China (62301452) and the Scientific Research Fund of the Yunnan Provincial Department of Education (No. 2024Y034).



(a) (b)

Fig. 4: t-SNE visualization for fine-grained Parkinson's disease severity assessment during the drinking water activity: (a) handcrafted features [11], (b) ADAMIL method.

REFERENCES

[1] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of neurology, neurosurgery & psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[2] W. Poewe, K. Seppi, C. M. Tanner, G. M. Halliday, P. Brundin, J. Volkmann, A.-E. Schrag, and A. E. Lang, "Parkinson disease," *Nature reviews Disease primers*, vol. 3, no. 1, pp. 1–21, 2017.

[3] Y. Zhou, M. Pang, W. Huang, and B. Wang, "Early diagnosing parkinson's disease via a deep learning model based on augmented facial expression data," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 1621–1625.

[4] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.

[5] P. Yang, G. Bi, J. Qi, X. Wang, Y. Yang, and L. Xu, "Multimodal wearable intelligence for dementia care in healthcare 4.0: A survey," *Information Systems Frontiers*, pp. 1–18, 2021.

[6] J. Qi, R. Cai, Q. Liu, W. Wang, J. Ma, and J. Chen, "A dynamic bayesian multi-channel fusion scheme for heart rate monitoring with ballistocardiograph signals in free-living environments," *IEEE Journal of Selected Areas in Sensors*, 2024.

[7] H. Wu, J. Qi, E. Purwanto, X. Zhu, P. Yang, and J. Chen, "Multi-scale feature and multi-channel selection toward parkinson's disease diagnosis with eeg," *Sensors*, vol. 24, no. 14, p. 4634, 2024.

[8] X. Wang, Y. Zhang, M. Zhou, T. Liu, J. Qi, and P. Yang, "Spatio-temporal similarity measure based multi-task learning for predicting alzheimer's disease progression using mri data," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 940–943.

[9] Y. Zhao, X. Wang, X. Peng, Z. Li, F. Nan, M. Zhou, J. Qi, Y. Yang, Z. Zhao, L. Xu *et al.*, "Selecting and evaluating key mds-updrs activities using wearable devices for parkinson's disease self-assessment," *IEEE Journal of Selected Areas in Sensors*, 2024.

[10] M. Zhou, X. Wang, T. Liu, Y. Yang, and P. Yang, "Integrating visualised automatic temporal relation graph into multi-task learning for alzheimer's disease progression prediction," *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[11] X. Peng, Y. Zhao, Z. Li, X. Wang, F. Nan, Z. Zhao, Y. Yang, and P. Yang, "Multi-scale and multi-level feature assessment framework for classification of parkinson's disease state from short-term motor tasks," *IEEE Transactions on Biomedical Engineering*, 2024.

[12] B. Vidya and P. Sasikumar, "Parkinson's disease diagnosis and stage prediction based on gait signal analysis using emd and cnn–lstm network," *Engineering Applications of Artificial Intelligence*, vol. 114, p. 105099, 2022.

[13] X. Wang, J. Huang, M. Chatzakou, K. Medijainen, A. Toomela, S. Nõmm, and M. Ruzhansky, "Lstm-cnn: An efficient diagnostic network for parkinson's disease utilizing dynamic handwriting analysis," *Computer Methods and Programs in Biomedicine*, vol. 247, p. 108066, 2024.

[14] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and S. Y. Philip, "Generalizing to unseen domains: A survey on domain generalization," *IEEE transactions on knowledge and data engineering*, vol. 35, no. 8, pp. 8052–8072, 2022.

[15] Z. Xu, F. Nan, J. Qi, Y. Yang, X. Wang, and P. Yang, "Modeling parkinson's disease aided diagnosis with multi-instance learning: An effective approach to mitigate label noise," in *2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2023, pp. 936–943.

[16] M. Stikic, D. Larlus, S. Ebert, and B. Schiele, "Weakly supervised recognition of daily life activities with wearable sensors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2521–2537, 2011.

[17] X. Guan, R. Raich, and W.-K. Wong, "Efficient multi-instance learning for activity recognition from time series data using an auto-regressive hidden markov model," in *International Conference on Machine Learning*. PMLR, 2016, pp. 2330–2339.

[18] X. Wang, Y. Yan, P. Tang, X. Bai, and W. Liu, "Revisiting multiple instance neural networks," *Pattern recognition*, vol. 74, pp. 15–24, 2018.

[19] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *International conference on machine learning*. PMLR, 2018, pp. 2127–2136.

[20] J. Early, G. K. Cheung, K. Cutajar, H. Xie, J. Kandola, and N. Twomey, "Inherently interpretable time series classification via multiple instance learning," *arXiv preprint arXiv:2311.10049*, 2023.

[21] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of machine learning research*, vol. 17, no. 59, pp. 1–35, 2016.

[22] Y. Chang, A. Mathur, A. Isopoussu, J. Song, and F. Kawsar, "A systematic study of unsupervised domain adaptation for robust human-activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–30, 2020.

[23] D. Cook, K. D. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: A survey," *Knowledge and information systems*, vol. 36, pp. 537–556, 2013.

[24] M. A. A. H. Khan, N. Roy, and A. Misra, "Scaling human activity recognition via deep learning-based domain adaptation," in *2018 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 2018, pp. 1–9.

[25] X. Qin, Y. Chen, J. Wang, and C. Yu, "Cross-dataset activity recognition via adaptive spatial-temporal transfer learning," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 4, pp. 1–25, 2019.

[26] Y. Du, J. Wang, W. Feng, S. Pan, T. Qin, R. Xu, and C. Wang, "Adarnn: Adaptive learning and forecasting of time series," in *Proceedings of the 30th ACM international conference on information & knowledge management*, 2021, pp. 402–411.

[27] W. Lu, J. Wang, X. Sun, Y. Chen, and X. Xie, "Out-of-distribution representation learning for time series classification," *arXiv preprint arXiv:2209.07027*, 2022.

[28] A. R. Sanabria and J. Ye, "Unsupervised domain adaptation for activity recognition across heterogeneous datasets," *Pervasive and Mobile Computing*, vol. 64, p. 101147, 2020.

[29] J. Wang, Y. Chen, L. Hu, X. Peng, and S. Y. Philip, "Stratified transfer learning for cross-domain activity recognition," in *2018 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 2018, pp. 1–10.

[30] G. Das, K.-I. Lin, H. Mannila, G. Renganathan, and P. Smyth, "Rule discovery from time series." in *KDD*, vol. 98, no. 1, 1998, pp. 16–22.

[31] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[32] J. Varghese, A. Brenner, M. Fujarski, C. M. van Alen, L. Plagwitz, and T. Warnecke, "Machine learning in the parkinson's disease smartwatch (pads) dataset," *npj Parkinson's Disease*, vol. 10, no. 1, p. 9, 2024.