2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS) | 979-8-3503-3071-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICPADS60453.2023.00139

Modeling Parkinson's Disease Aided Diagnosis with Multi-Instance Learning: An Effective Approach to Mitigate Label Noise

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

Fengtao Nan
National Pilot School of Software
Yunnan University
Kunming, China
fengtaonan@gmail.com

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

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

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

***Abstract*—An effective auxiliary diagnostic model for the severity of Parkinson's disease (PD) could help hospitals reduce their workload, particularly in nations or regions where medical resources are limited. However, a critical challenge persists that hampers the progress of such endeavors. Previous studies have employed label propagation techniques that assign uniform labels to all activity signal segments of a patient, neglecting the complex expression of PD symptoms, thereby introducing label noise. To confront this challenge, we have collected an extensive set of PD activity signals from a clinical setting and have proposed an efficient and robust framework for assessing PD severity. Specifically, we gathered wearable device data on 14 daily activities from 70 PD patients, based on the Unified Parkinson's Disease Rating Scale Part III. Our data analysis indicates that many segments within the activities were incorrectly labeled, significantly impairing the classification performance of the model. We introduced a novel framework based on Multi-Instance Learning with a Re-weighted Discriminative Instance Mapping (RDIM) to model PD auxiliary diagnosis, aiming to eliminate the impact of label noise present in the data. The results demonstrate that our framework achieves an accuracy of 80.88% in classifying the severity of PD, effectively addressing the label noise caused by coarse-grained label propagation.**

***Index Terms*—Parkinson's Disease, Multi-Instance Learning, Wearable Devices, Label Noise, Inexact Supervision**

I. Introduction

Parkinson's Disease (PD) is currently the second most prevalent neurodegenerative disorder, primarily characterized by symptoms such as tremors, bradykinesia, and gait freezing. These manifestations severely impair the quality of life of patients. Numerous studies suggest that appropriate treatments can slow down the progression of the disease. physicians often assess patients using the Unified Parkinson's Disease Rating Scale (UPDRS) [1].

In medical practice, physicians primarily rely on a patient's clinical symptoms and patient's recollection of their medical history and propose treatment recommendations. This estimation of PD is time-consuming and complex, and relies on the subjective experience of the physician.

With the rapid advancement of the Internet of Things (IoT) [2]–[5] and Cyber-Physical Systems (CPS) [6], there has been a growing focus on the assessment of Parkinson's Disease (PD) in recent literature [7]–[11]. The deployment of wearable sensors to monitor and assess PD through analysis of patient movements has become a significant area of investigation. For instance, prior research has predominantly concentrated on detecting discrete symptoms, such as tremor, bradykinesia [12]–[16], and gait freezing [17]–[19]. While these studies have contributed valuable insights into the severity of individual PD symptoms, they have not provided a holistic evaluation of the patient's overall disease stage. This limitation stems from the fact that the comprehensive assessment of PD severity, as stipulated by the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS), necessitates the observation of a spectrum of symptoms manifested across various activities. Existing research has largely been restricted to the examination of isolated symptoms or singular activities, hence not capturing the full spectrum of activities and symptoms required for an all-encompassing assessment of the disease's progression.

Despite the promising potential of machine learning in diagnosing and treating PD, there exists a significant gap in the current research regarding the assessment of PD

2690-5965/23/$31.00 ©2023 IEEE
DOI 10.1109/ICPADS60453.2023.00139

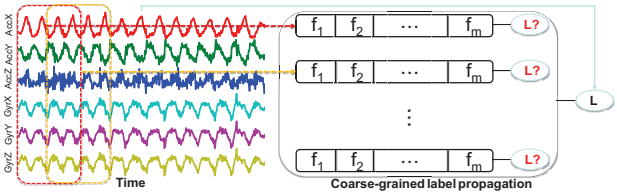

Fig. 1: Feature extraction and label propagation

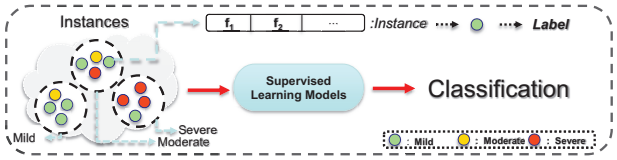

Fig. 2: Traditional supervised learning: the labels (i.e. Mild PD, Moderate PD or Severe PD) are available for each instance (i.e. ball). A dashed circle encloses all balls from one patient. The object for classification is an instance.

using a single wearable device. Specifically regarding data label processing methods, the label assignment exhibits a significant oversight. In contrast to activity recognition tasks that assign uniform coarse-grained labels to sliding windows (e.g., running, ascending stairs), such labeling is inappropriate for PD severity assessments due to the non-periodic nature of symptom manifestation, a significant difference often overlooked by current techniques using sliding windows.

As illustrated in Fig. 1, employing sliding window techniques for feature extraction from time-series data of accelerometers and gyroscopes in wearable devices allows the generation of multiple instances from a single activity session of a patient. The challenge of inexact supervision [20] arises due to the absence of specific labels for each instance, hindering the application of conventional supervised learning algorithms for instance-level model training. Previous work has addressed this by propagating coarse-grained labels of the overall condition to each instance.

However, the assignment of a single coarse-grained label, often based on the Hoehn and Yahr scale [21], to these instances overlooks the complex dynamics of PD, leading to mismatch between instance and label. Such an approach fails to account for the non-periodic nature of motor symptom expression in PD, potentially introducing label noise as not all segments captured by a fixed sliding window may reflect the label's characteristics. This issue poses a significant obstacle to the implementation of accurate supervised learning methodologies for the evaluation of PD severity, where the heterogeneity of motor symptoms cannot be adequately represented by a uniform label across all instances.The training process using the label propagation mode is shown in Fig. 2.

Fig. 3 illustrates the issue of label noise from coarse-

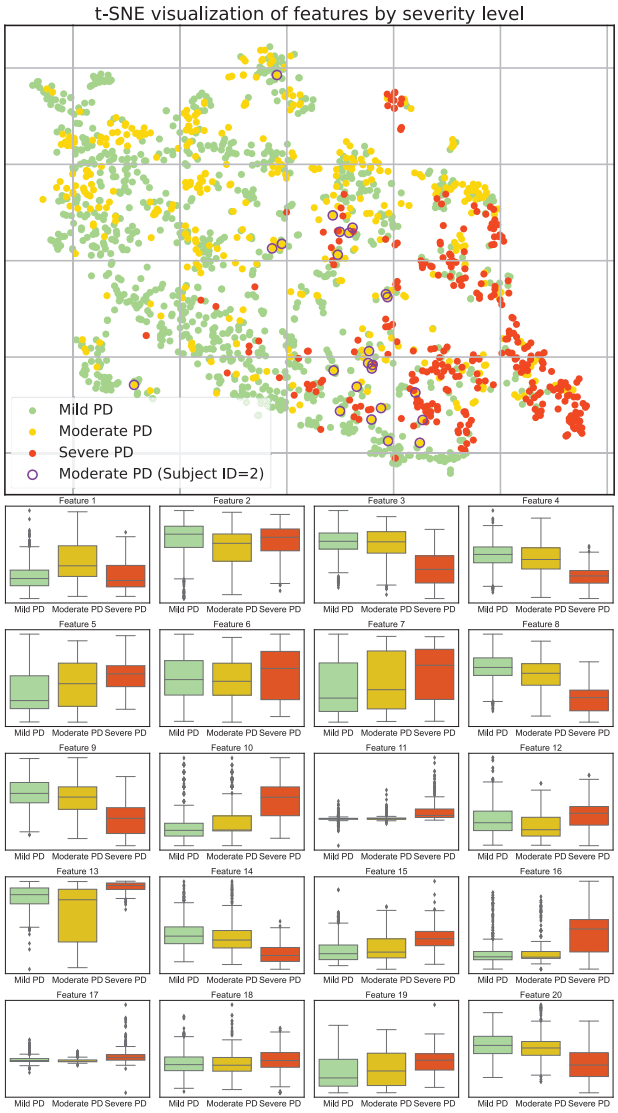

Fig. 3: t-SNE and Boxplot Visualization of PD Features by Severity Level.

grained label propagation, by presenting a reduced-dimension visualization of the activity data for a patient was labeled with moderate PD. Despite all instances being labeled 'moderate', they are dispersed and overlap with instances labeled as 'mild' and 'severe'. This dispersion within the moderate category, which is also evident in other categories but especially pronounced in moderate cases, suggests that not all instances should be uniformly labeled, underscoring the inaccuracy of coarse-grained labels.

This preliminary indication is reinforced by further examination via boxplot analysis, which reveals significant overlaps on critical clinical features between these categories. The convergence of evidence from these analyses implies that the feature distributions of certain data

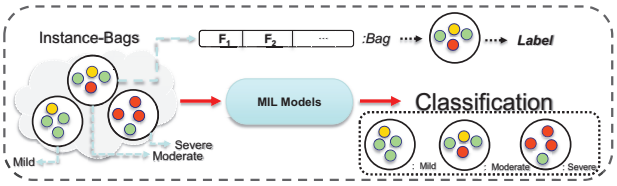

Fig. 4: Multi-instance learning: each bag (i.e. circle) consists of several instances (i.e. balls) and labels (i.e. Mild PD, Moderate PD or Severe PD) are only available for bags.The object for classification is an instance bag.

points currently labeled as 'Moderate PD' may align more closely with those characteristic of 'Mild PD' or 'Severe PD'. This implies that a patient with a moderate stage of PD may not consistently exhibit symptoms of that level within a fixed time window, reflecting the non-periodic nature of the disease's progression. Consequently, it suggests that not all instances from the patient should be uniformly assigned a coarse-grained label of 'moderate'.

To seek an effective solution, we extensively reviewed the relevant literature. Upon comprehensive research, it emerged that Multiple Instance Learning (MIL) stands as a principal method in addressing the inexact supervision challenge.

Compared to traditional supervised learning, MIL presents distinct characteristics. In conventional supervised learning, each instance is assigned a specific label. However, in MIL, instances are organized into "bags", where each "bag" encompasses multiple instances, but the label is allocated to the entire bag rather than individual instances within it. Formally, let's represent a set of samples and their respective labels as $D = \{(X_i, y_i)\}_{i=1}^{N}$. Here, each sample X_i is essentially a collection of instances, denoted as $X_i = (x_{i1}, x_{i2}, \ldots, x_{iK})$ where $x_{ij} \in \mathbb{R}^N$. The associated label y_i pertains to the entire bag X_i, not any individual instance $\mathbf{x}_{ij}$. Consequently, the aim in this context is to learn a bag classifier.

In this work, we categorize the Multiple-instance classification (MIC) methods according to how the information existent in the multiple instance (MI) data is exploited [22].

- **Instance-Space (IS) paradigm**:This paradigm posits that discriminative information predominantly exists at the instance level. Accordingly, the learning mechanism is specifically designed to function at this finer granularity. A discriminative instance-level classifier, denoted as $f(x)$, is developed to distinguish instances within bags. A bag classifier then aggregates these instance-level decisions to ascertain the bag's label by evaluating its constituent instances. A crucial element of this approach is the assignment of coarse-grained bag labels to each instance [23]. However, this process is prone to introducing label noise, which can compromise the precision of instance label predictions and, consequently, the overall accuracy of bag label predictions. The IS paradigm concentrates on local, instance-level attributes, thereby neglecting the collective characteristics of the entire bag.
- **Bag-Space (BS) paradigm**: This paradigm asserts that discriminative information is inherent at the bag level. Owing to the non-vectorial composition of bags, BS approaches employ non-standard learning algorithms tailored for such data structures. Detailed exposition of these methods is beyond the scope of this paper.
- **Embedded-Space (ES) paradigm**: Each bag X is represented by a single feature vector that encapsulates the aggregate information of the bag. This transformation maps the original bag space to a vectorial embedded space where traditional discriminative classifiers can be trained. Such a transformation redefines the original MIC challenge into a conventional supervised learning task, with each feature vector associated with a specific label, permitting the application of standard classifiers like Knn, Neural Networks, or Support Vector Machines (SVMs). This concept is visualized in Fig. 4. Within this paradigm, commonly implemented techniques include Simple MI [24] and Vocabulary-based methods [25], [26], However, these methods often lose key instance information within bags.

Our research addresses the challenges of label noise in the Instance-Space (IS) paradigm and detail loss in the Embedded-Space (ES) paradigm. Drawing on insights from key studies [27], [28], we have developed an integrated approach that effectively tackles these issues. We propose the Re-weighted Discriminative Mapping (RDIM) framework, an ES-based approach that refines bag representation by mapping to a feature space optimized for discriminative clarity. This is achieved by correlating original features with a selected subset from the Re-weighted Discriminative Instance Pool (RDIP), enhancing accuracy of bag classification as demonstrated in Fig. 5.

In summary, the primary contributions of this paper encompass three aspects:

1) We have transformed the problem of PD severity classification under inexact supervision into a multi-instance classification problem, thus enabling the application of multi-instance learning for the diagnosis of PD.
2) We have introduced the Re-weighted Discriminative Bag Mapping (RDIM) framework for multi-instance learning, specifically targeting the reduction of label noise and improving the discriminative power for PD severity classification.
3) We have validated the RDIM approach using a dataset from 70 PD patients engaging in 14 diverse daily activities, which has proven the method's robustness and accuracy in evaluating PD severity.

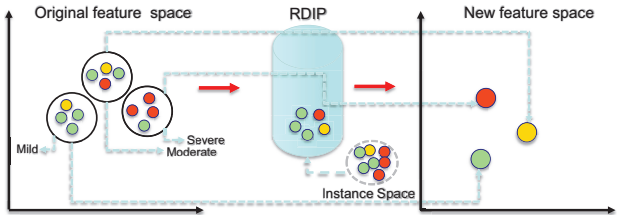

Fig. 5: Re-weighted discriminative bag mapping: the informative instances are selected as a re-weighted discriminative instance pool to ensure that the bags in the new mapping space can be easily distinguished.

The structure of this paper is as follows: Section II provides a comprehensive overview of the PD dataset, detailing data collection, preprocessing, and labeling methods. Section III delineates our proposed Re-weighted Discriminative Bag Mapping framework for assessing PD severity. Section IV explicates the experimental setup and documents the results to substantiate our hypothesis. Concluding discussions are encapsulated in Section V.

II. DATA COLLECTION AND PREPROCESSING

A. Data Collection

Between 2020 and June 20, 2022, a total of 70 PD activity datasets were collected within the Neurology Department at the First People's Hospital of Yunnan Province, China. All subjects provided voluntary written informed consent for their participation. Based on the Unified Parkinson's Disease Rating Scale (UPDRS) and its measures for PD motor symptoms, 14 activity paradigms were designed to evaluate PD status. Disease staging followed the Hoehn and Yahr (H&Y) classification, with 41 patients in the mild stages (H&Y stages 1 and 2), 17 in the moderate stage (H&Y stage 3), and 12 in the severe stages (H&Y stages 4 and 5).

To capture motion data, SHIMMER3 sensors were placed on patients in five locations: right hand, left hand, waist, right ankle, and left ankle, as shown in Fig. 6(a). This study primarily focuses on the data from the right-hand sensor. Data communication was facilitated via Bluetooth with a laptop computer. The Shimmer sensors were equipped with the ConsensysPRO software, recording 3D accelerometer, gyroscope, and magnetometer data at a frequency of 200Hz and offering real-time data visualization. Throughout the data collection process, participants were clearly instructed on the 14 activity paradigms by the research team, with data collection durations ranging from 20 to 50 seconds. A comprehensive breakdown of these activities is provided in Fig. 6(b).

B. Data Preprocessing and Feature Extraction

In order to mitigate the noise in the collection process of wearable sensors, a band-pass filter was applied to the 3D

(a)

No.	Activity Name	Activity Description
1	Finger tap	Tap your thumb with the index finger of both hands in rapid succession.
2	Hands opening and closing	Open and close the hand in rapid succession.
3	Wrist rotation	Extend hands forward, palms facing down, then quickly alternate between turning palms up and down.
4	Right-hand flip	The palm and back of the right hand alternately slap the left palm.
5	left-hand flip	The palm and back of the left hand alternately slap the right palm.
6	Left finger to nose	Touch the nose with the index finger of the left hand, then touch the doctor's index finger 30 cm from the patient.
7	Right finger to nose	Touch the nose with the index finger of the right hand, then touch the doctor's index finger 30 cm from the patient.
8	Upright and hand raised flat	Raise your hands and stand shoulder height for 20 seconds.
9	Walk back and forth	Walk straight for 10 m, turn around, then walk in the opposite direction.
10	From Sit to Stand	Stand up from the straight-backed chair with arms on the chest.
11	Drink water	Pick up a paper cup from the table and drink from it.
12	Pick up things	Pick up items on the ground.
13	Sit	Sit in a chair in your natural state with your hands on your knees.
14	Standing	Stand with your arms in a natural vertical position.

(b)

Fig. 6: (a) depicts a schematic of the patient wearing the device, (b) provides a detailed description of the 14 daily activities used for PD severity assessment.

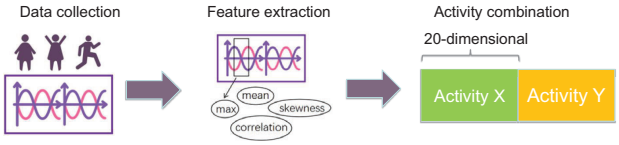

Fig. 7: Activity combination

accelerometer and gyroscope readings, resulting in a six-dimensional dataset. Data segmentation used a 1.5-second sliding window with a 50% overlap. From these windows, both time-domain (mean, max, min, standard deviation, Root Mean Square (RMS), variance, etc.) and frequency-domain (spectral entropy, Fast Fourier Transform (FFT) energy, etc.) features were extracted. Following feature selection, 20 features were identified for in-depth analysis.

C. Activity Combination

Our methodology enriches PD severity assessment by incorporating an activity concatenation phase after feature extraction, aggregating data across various activities to improve accuracy. Instead of relying on single-activity data, our approach combines 20-dimensional feature sets from multiple activities into a comprehensive 40-dimensional feature vector, enhancing the dataset's feature richness and diagnostic granularity (see Fig. 7).

III. METHODOLOGY

A. Problem Formulation

To address the challenge of label noise in PD severity classification from patient activity data, we frame the task as a Multiple Instance Classification (MIC) problem. We

TABLE I: NOTATIONS

Notation	Meaning
$\boldsymbol{\beta} = \{\boldsymbol{\beta}_i\}_{i=1}^{n}$	The dataset
$\mathbf{n}$	Number of bags
$\boldsymbol{\beta}_i$	The i-th bag
$\boldsymbol{y_i} \in \mathcal{Y} = \{0, 1, 2\}$	Bag label
$\boldsymbol{\mathcal{X}}$	Instance space
$\boldsymbol{x}_{ij}$	The j-th instance in the i-th bag
$\boldsymbol{\mathcal{P}}$	Re-weighted Discriminate Instance Pool
$\mathbf{m}$	$\mathcal{P}$ Size
$\boldsymbol{\beta}_i^{\phi} = \left[s\left(\boldsymbol{\beta}_i, x_1^{\phi}\right), \cdots, s\left(\boldsymbol{\beta}_i, x_m^{\phi}\right)\right]$, where $x_k^{\phi} \in \mathcal{P}$	Bag mapping
$s(\boldsymbol{\beta}_i, x_k^{\phi})$	The i-th bag similarity to x_k^{ϕ}

introduce the RDIM framework to mitigate label inconsistency and enhance predictive accuracy in PD severity classification.

B. Preliminaries And Overall Framework

Given a dataset β of activity data from 70 PD patients, each activity paradigm corresponds to a distinct β. With n representing the number of patients and equivalently the number of bags, each β_i symbolizes a bag. Within each bag β_i, there exist instances x_{ij}, each reflecting features derived from a sliding window of signal data.

Formally, a bag β_i is defined as $\beta_i = \{x_{ij}\}_{j=1}^{n_i}$, where n_i represents the number of instances within β_i and x_{ij} signifies the j-th instance in the i-th bag, with each instance $x_{ij} \in \mathbb{R}^d$ characterizing a segment of activity through d extracted features. The label $y_i \in \mathcal{Y} = \{0, 1, 2\}$ designates the PD severity level for bag β_i, with 0 indicating mild PD, 1 moderate PD, and 2 severe PD.

The instance space $\mathcal{X}$ encompasses all instances x_{ij} from β. During training, a re-weighted discriminative instance pool $\mathcal{P} \subseteq \mathcal{X}$, a selected subset of size m, is utilized to transform each bag β_i into a new feature space, represented as singular instances β_i^{ϕ}. The new instance representation is $\beta_i^{\phi} = [s(\beta_i, x_1^{\phi}), \ldots, s(\beta_i, x_m^{\phi})]$, where $s(\beta_i, x_k^{\phi})$ is the similarity measure between bag β_i and the k-th instance x_k^{ϕ} in $\mathcal{P}$. A KNN-based supervised classifier is then trained in this new feature space, with symbol representations detailed in TABLE I.

For testing, test bags are mapped into new instances using the Re-weighted Discriminative Instance Pool (RDIP) established during the training phase. The trained classifier subsequently assigns the final class labels. Identifying the optimal instance set $\mathcal{P}$ is crucial in this process.

C. How To Find The Best Distinguishing Instance

$$\max_{\mathcal{P} \subseteq \mathcal{X}} \sum_{y_i \neq y_j} d(\boldsymbol{\beta}_i^{\phi}, \boldsymbol{\beta}_j^{\phi}) \tag{1}$$

$$\min_{\mathcal{P} \subseteq \mathcal{X}} \sum_{y_i = y_j} d(\boldsymbol{\beta}_i^{\phi}, \boldsymbol{\beta}_j^{\phi}) \tag{2}$$

The optimal RDIP (Re-weighted Discriminative Insantacne Pool) would result in mapped instances with the maximal discriminative power. Therefore, the optimality of an RDIP can be assessed by measuring whether the mapped instances possess the highest degree of discrimination. The function d(·,·) represents the distance measure between two mapped vectors. Given two samples in the new mapped space, denoted by $\boldsymbol{\beta}_i^{\phi}$ and $\boldsymbol{\beta}_j^{\phi}$, in the new feature space, distinct class separation and same-class proximity are crucial.

In an endeavor to transform the multi-objective optimization task into a single-objective one, we introduce the label relationship matrix, $\boldsymbol{\Delta}$, defined as $\Delta = [Q_{ij}]_{n \times \mathbf{n}}$ wherein:

$$\mathcal{Q}_{ij} = \begin{cases} 1, & y_i \neq y_j \\ -1, & y_i = y_j \end{cases} \tag{3}$$

Thus, the composite optimization objective is given by:

$$J(\mathcal{P}) = \frac{1}{2} \sum_{i,j} d_{ij} Q_{ij} \tag{4}$$

$$\mathcal{P}_* = \arg\max_{\mathcal{P} \subseteq \mathcal{X}} J(\mathcal{P}) \quad \text{s. } t. \mid \mathcal{P} \mid = m \tag{5}$$

In equation (4), d_{ij} denotes the distance between $\boldsymbol{\beta}_i^{\phi}$ and $\boldsymbol{\beta}_j^{\phi}$. The term Q_{ij} represents the label relationship between $\boldsymbol{\beta}_i^{\phi}$ and $\boldsymbol{\beta}_j^{\phi}$. If the labels are the same, Q_{ij} is -1. If they are different, Q_{ij} is 1. A greater value of $J(\mathcal{P})$ suggests better optimization of equations (1) and (2), implying that instances of the same class are closer in the new feature space, while those from different classes are further apart. The ultimate goal now is to find the optimal instance set $\mathcal{P}$ that maximizes the value of $J(\mathcal{P})$. In order to optimize equation (5) and find this set $\mathcal{P}$, we can compute each instance (i.e., x_{ij}) within the instance space $\mathcal{X}$. We then collect the top m instances with the highest scores to constitute the optimal instance set $\mathcal{P}$.

$$J(\mathcal{P}) = \frac{1}{2} \Sigma_{i,j} \parallel \boldsymbol{\beta}_i^{\boldsymbol{\phi}_x} - \boldsymbol{\beta}_j^{\boldsymbol{\phi}_x} \parallel^2 Q_{ij} \tag{6}$$

Thus, we can obtain equation (6) as our current optimization function. However, the label relationship Q_{ij} treats any pair of β_i and β_j indiscriminately. This can cause the instance selection process to be influenced by data imbalance, leading to a bias towards classes with larger quantities. Such a bias means that these classes would be more distinguishable in the new feature space, making it easy to overlook the distinguishing capabilities of minority classes in the new space. Therefore, we reweight the label relationship matrix Q to obtain L.

$$Ł_{ij} = \begin{cases} \frac{1}{|G| \times \sqrt{\frac{|l_{\beta_i}| + |l_{\beta_j}|}{2 \times |\beta|}}}, & y_i \neq y_j \\ \frac{1}{|S| \times \sqrt{\frac{|l_{\beta_i}| + |l_{\beta_j}|}{2 \times |\beta|}}}, & y_i = y_j \end{cases} \tag{7}$$

Algorithm 1 RDIP: Re-weighted Discriminate Instance Pool

Input: Training bag dataset β
The instance set $\mathcal{X}$ collected from β
The number of selected instance m
Output: $\mathcal{P} = \{p_1, \ldots, p_m\}$: A set of instances

1 $\mathcal{P} = \emptyset$, $\tau = 0$
2 $Q \leftarrow$ Apply all bag labels in β to obtain the label relationship matrix via Eq. (3)
3 $L \leftarrow$ Assign weights to the values in the corresponding Q label relationship matrix to obtain re-weighted label relationship matrix via Eq. (7)
4 **for** *each instance* x_k *in* $\mathcal{X}$ **do**
5 $\quad f(x_k, L) \leftarrow$ Apply Eq. (8) to measure the score
6 $\quad$ **if** $|P| < m$ *or* $f(x_k, L) > \tau$ **then**
7 $\quad\quad \mathcal{P} \leftarrow \mathcal{P} \cup \{x_k\}$
8 $\quad$ **end**
9 $\quad$ **if** $|P| \geq m$ **then**
10 $\quad\quad \mathcal{P} \leftarrow \mathcal{P}/\{\arg\min_{x_k \in \mathcal{P}} f(x_k, L)\}$
11 $\quad$ **end**
12 $\quad \tau \leftarrow \min_{x_k \in \mathcal{P}} f(x_k, L)$
13 **end**
14 **return** $\mathcal{P}$

Given that $|l_{\beta_i}|$ and $|l_{\beta_j}|$ represent the number of instances for labels β_i and β_j in the current β dataset, respectively, and $|\beta|$ denotes the total number of data instances in the dataset. While $|G|$ stands for the count of elements in the matrix Q that are greater than 0, $|S|$ also signifies the count of elements in the matrix Q that are smaller than 0. From these insights, we can infer that if both β_i and β_j correspond to minority class data, a larger weight will be allocated when evaluating $J(\mathcal{P})$. This methodology accentuates the significance of the minority class data's perspectives. In contrast, for the majority class data, a diminished weight is applied in the calculation of $J(\mathcal{P})$, attenuating the dominance of the majority class. This approach takes into account the impact of data imbalance on the final instance set $\mathcal{P}$.

Now, we employ the equation(8) to optimize the aforementioned expression and use L to assign weights to the values in the corresponding Q label relationship matrix, where Equation (8) represents the functional form of Equation (6).

$$\max_{\mathcal{P} \subseteq \mathcal{X}} \sum_{x_k^{\phi} \in \mathcal{P}} f(x_k^{\phi}, L) \quad s.t. |\mathcal{P}| = m \tag{8}$$

In the end, from the instance space $\mathcal{X}$, we procure the top m instances with the highest scores, constructing the optimal instance set $\mathcal{P}$, referred to as RDIP.

Algorithm 1 details the proposed approach for obtaining the Re-weighted Discriminative Instance Pool (RDIP). Initiated with an empty instance set $\mathcal{P} = \varnothing$ and a minimal score threshold $\mathcal{T} = 0$ (line 1), it first computes the label relationship matrix Q and its corresponding L (lines 2-3), which encapsulates the re-weighted relationship information between the labels of the bags i and j. For every instance $x_{\mathrm{k}} \in \mathcal{X}$, its discriminative score $f(x_k, L)$ is computed based on L via equation (8) . If $f(x_k, L)$ surpasses the minimal discriminative score when $\mathcal{P}$ is at $\mathcal{T}$, or if the size of $\mathcal{P}$ is smaller than m (indicating $\mathcal{P}$ isn't full), x_k is selected as a member of $\mathcal{P}$ (lines 6-7).Conversely, if $\mathcal{P}$ overflows, the instance with the lowest discriminative score, $\arg\min_{\chi_k \in \mathcal{P}} f(x_k, L)$, is removed to retain its size (lines 9-10). Subsequently, the minimum discriminative score $\mathcal{T}$ in $\mathcal{P}$ is updated (line 12). This loop continues until the optimal RDIP $\mathcal{P}$ is obtained.

D. Bag Mapping Via RDIP

$$\boldsymbol{\beta}_i^{\phi} = [\mathrm{s}(\boldsymbol{\beta}_i, \mathrm{x}_1^{\phi}), \cdots, \mathrm{s}(\boldsymbol{\beta}_i, \mathrm{x}_{\mathrm{m}}^{\phi})] \tag{9}$$

$$s(\beta_i, x_k^{\phi}) = \max_{x_{ij} \in \boldsymbol{\beta}_i} exp(-\|x_{ij} - x_k^{\phi}\|^2 / \sigma^2) \tag{10}$$

Once the RDIP is formulated using the chosen instances, utilizing the instances within RDIP and through equation (9), we map all bags from the original space to the new feature space. The function $\mathrm{s}(\boldsymbol{\beta}_i, \mathrm{x}_{\mathrm{k}}^{\phi})$ within the equation is derived from equation (10). Here, σ is a predefined scaling factor. Once every $\boldsymbol{\beta}_i \in \boldsymbol{\beta}$ is mapped using the optimal RDIP, it can be subjected to any single-instance learner, such as KNN, without constraints.

IV. EXPERIMENTS

A. Experimental Setup

In this study, three comparative methods are employed: the approach of coarse-grained label propagation, which serves as a baseline; α (Simple MI);and β (Vocabulary-based methods). For α, each bag X is mapped to the average of the instances inside, mathematically represented as $\mathcal{M}(X) = \frac{1}{|X|} \sum_{\vec{x} \in X} \vec{x}$. In contrast, for β, we implement k-means clustering on instances within each bag, With the clustering statistics of the instances within the bag, the bag representation is formed using the information from the clustering of the instances. For instance, if a bag contains 24 instances clustered into five groups, the bag is represented as $(15, 1, 3, 5, 0)$, where each number corresponds to the count of instances in each cluster, we set the n-clusters range to $(0, 11)$.

Following the establishment of the Re-weighted Discriminative Instance Pool (RDIP), data represented by bags are transposed from the original to a novel feature space, enabling the transfer of bag labels to corresponding instances in this space. This transposition facilitates the application of various supervised classifiers within the multi-instance learning (MIL) paradigm. For our empirical analysis, we selected the instance-based K-nearest neighbor (KNN) classifier , which has demonstrated substantial classification efficacy in activity recognition scenarios, particularly with limited user data [29].

TABLE II: ACCURACY FOR CLASSIFICATION

Activity	KNN	KNN+α	KNN+β	KNN+RDIM	patients
1	62.85	62.85	64.28	**68.57**	70
2	58.57	65.71	61.43	**64.28**	70
3	65.71	64.28	64.28	**74.28**	70
4	60.0	**68.57**	58.57	67.14	70
5	64.28	58.57	71.42	**74.28**	70
6	66.17	61.76	52.94	**73.52**	68
7	52.85	62.85	51.42	**65.71**	70
8	59.42	65.21	49.27	**65.21**	69
9	67.64	**75.00**	64.70	73.52	68
10	66.10	71.18	69.49	**71.18**	59
11	64.61	61.53	62.71	**67.69**	65
12	71.42	73.21	64.28	**76.78**	56
13	54.54	59.09	47.27	**65.90**	44
14	50.0	50.00	54.76	**71.42**	43
3+9	63.23	69.11	70.58	**80.88**	68
3+5	68.57	62.85	71.42	**74.28**	70
3+6	63.23	61.76	60.29	**69.11**	68
5+6	60.29	63.23	57.35	**64.70**	68
5+9	64.70	**70.58**	67.64	69.11	68
6+9	65.15	65.15	65.15	**72.72**	66
Average	62.47	64.62	61.46	**70.51**	-

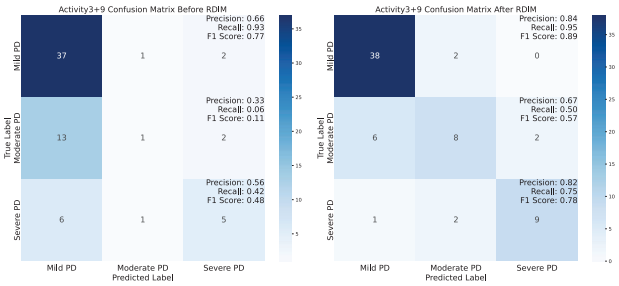

Fig. 8: Confusion matrix results before and after RDIM

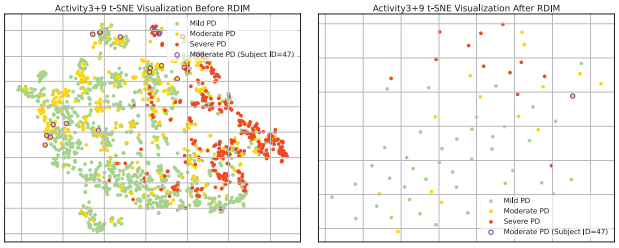

Fig. 9: t-SNE visualization before and after RDIM

In configuring the KNN model, we elected the "weighted" scheme for the weight parameter and confined the n-neighbors to the interval $(0, 11)$. The RDIP's critical parameter σ was tuned within $(10^0, 10^{20})$, the value of parameter m is associated with the total number of instances in the instance space, spanning a range of $(0, |\mathcal{X}| + 1)$, where $|\mathcal{X}|$ signifies the total instance count.

Our assessment metrics encompassed Accuracy, Recall, Precision, and F1 score. We utilized leave-one-out cross-validation for our evaluation methodology, designating one patient as the test set and the remainder for training in each iteration.

B. The Most Discriminating Activity And Visualization

In our experiment, we evaluated 14 individual activity patterns and 6 activity combination paradigms to determine which ones have the highest discriminatory efficacy for PD. We discerned that Activity Paradigms 3 (wrist flip), 5 (left hand flip), 6 (left fingertip to nose), and 9 (walk back and forth) correlate strongly with PD symptoms, as evidenced by their heightened discriminative capacity (see TABLE II). To enhance the richness of data features, we combined four activity paradigms to create activity combinations, with the 3 and 9 fusion manifesting the most potent discriminative ability.

The selection of highly discriminative activity paradigms in PD can be explained by the symptoms exhibited by PD patients. For example, in Activity Paradigm 9 (back-and-forth walking), PD patients demonstrate symptoms such as gait freezing, altered walking posture, irregular rhythm and stride, diminished or absent arm swing, as well as difficulties in turning and performing hand activities. These symptoms are also observed in Activity Paradigms 3, 5, and 6, which involve rapid and broad motions. Specifically, Activity Paradigm 6 requires participants to touch their nose with a fingertip and then touch a doctor's fingertip located 30 cm away. PD patients exhibit sluggishness and unresponsiveness during execution due to compromised connections between the basal ganglia and the cortical motor area, leading to difficulties in performing smooth and coordinated activities. Numerous clinical studies [30] have shown a strong correlation between disease stages and motor symptoms in PD patients.

A confusion matrix (see Fig. 8) will be presented, elucidating the recall, precision, and F1 score, thereby appraising the efficacy of our Re-weighted Discriminative Instance Mapping (RDIM) framework in PD severity classification. Further, t-SNE visualizations (see Fig. 9) underscore the improved discriminability within the transformed feature space, confirming the RDIM framework's capacity to mitigate data overlap and label noise issues.

V. CONCLUSION

In our research, we propose a novel instance selection methodology emanating from a multi-instance learning paradigm. This approach, through the utilization of specific instances, maps bag-level data from the original feature space to an innovative feature space, thereby effectively addressing the label noise issue originating from coarse-grained label propagation. Specifically, relying on the concept of Re-weighted Discriminative Instance Mapping (RDIM), we enhance the discernibility of data within this new feature space. The method's validity was affirmed through stringent testing on a PD dataset, which included 70 patients participating in 14 distinct activity paradigms,

particularly demonstrating its capability in overcoming the challenges associated with label noise from coarse-grained label propagation.

Looking ahead, we plan to develop a PD severity assessment application for use in a free-living environment. This tool is anticipated to not only assist physicians in diagnosing PD but also empower patients to conduct self-examinations.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (No. 62061050) and Young Scientists Fund of the National Natural Science Foundation of China (No.62301452).

REFERENCES

[1] 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.

[2] E. G. Spanakis, D. Kafetzopoulos, P. Yang, K. Marias, Z. Deng, M. Tsiknakis, V. Sakkalis, and F. Dong, "Myhealthavatar: Personalized and empowerment health services through internet of things technologies," in *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*. IEEE, 2014, pp. 331–334.

[3] J. Qi, P. Yang, M. Hanneghan, D. Fan, Z. Deng, and F. Dong, "Ellipse fitting model for improving the effectiveness of life-logging physical activity measures in an internet of things environment," *Iet Networks*, vol. 5, no. 5, pp. 107–113, 2016.

[4] J. Qi, P. Yang, A. Waraich, Z. Deng, Y. Zhao, and Y. Yang, "Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: A systematic review," *Journal of biomedical informatics*, vol. 87, pp. 138–153, 2018.

[5] J. Qi, P. Yang, L. Newcombe, X. Peng, Y. Yang, and Z. Zhao, "An overview of data fusion techniques for internet of things enabled physical activity recognition and measure," *Information Fusion*, vol. 55, pp. 269–280, 2020.

[6] P. Yang, G. Yang, J. Liu, J. Qi, Y. Yang, X. Wang, and T. Wang, "Duapm: An effective dynamic micro-blogging user activity prediction model towards cyber-physical-social systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5317–5326, 2019.

[7] H. Wu, J. Qi, and Y. Yue, "Machine learning-based eeg signal classification of parkinson's disease," in *2023 IEEE 10th International Conference on Cyber Security and Cloud Computing (CSCloud)/2023 IEEE 9th International Conference on Edge Computing and Scalable Cloud (EdgeCom)*. IEEE, 2023, pp. 423–428.

[8] L. Tao, X. Wang, X. Peng, P. Yang, J. Qi, and Y. Yang, "Activity selection to distinguish healthy people from parkinson's disease patients using i-da," in *2021 17th International Conference on Mobility, Sensing and Networking (MSN)*. IEEE, 2021, pp. 66–73.

[9] L. Tao, X. Wang, F. Nan, J. Qi, Y. Yang, and P. Yang, "Effective severity assessment of parkinson's disease with wearable intelligence using free-living environment data," in *2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE)*. IEEE, 2023, pp. 1–10.

[10] M. Zhou and P. Yang, "Automatic temporal relation in multi-task learning," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 3570–3580.

[11] P. Yue, X. Wang, Y. Yang, J. Qi, and P. Yang, "Up-sampling active learning: An activity recognition method for parkinson's disease patients," in *International Conference on Pervasive Computing Technologies for Healthcare*. Springer, 2022, pp. 229–246.

[12] H. Dai, G. Cai, Z. Lin, Z. Wang, and Q. Ye, "Validation of inertial sensing-based wearable device for tremor and bradykinesia quantification," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 997–1005, 2020.

[13] J.-F. Daneault, B. Carignan, C. É. Codère, A. F. Sadikot, and C. Duval, "Using a smart phone as a standalone platform for detection and monitoring of pathological tremors," *Frontiers in human neuroscience*, vol. 6, p. 357, 2013.

[14] D. J. Wile, R. Ranawaya, and Z. H. Kiss, "Smart watch accelerometry for analysis and diagnosis of tremor," *Journal of neuroscience methods*, vol. 230, pp. 1–4, 2014.

[15] G. Rigas, A. T. Tzallas, M. G. Tsipouras, P. Bougia, E. E. Tripoliti, D. Baga, D. I. Fotiadis, S. G. Tsouli, and S. Konitsiotis, "Assessment of tremor activity in the parkinson's disease using a set of wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 478–487, 2012.

[16] J. Synnott, L. Chen, C. D. Nugent, and G. Moore, "Wi-ipd—objective home assessment of parkinson's disease using the nintendo wii remote," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1304–1312, 2012.

[17] J. C. Perez-Ibarra, A. A. Siqueira, and H. I. Krebs, "Identification of gait events in healthy subjects and with parkinson's disease using inertial sensors: An adaptive unsupervised learning approach," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 12, pp. 2933–2943, 2020.

[18] S. T. Moore, V. Dilda, B. Hakim, and H. G. MacDougall, "Validation of 24-hour ambulatory gait assessment in parkinson's disease with simultaneous video observation," *Biomedical engineering online*, vol. 10, no. 1, pp. 1–9, 2011.

[19] A. Weiss, S. Sharifi, M. Plotnik, J. P. van Vugt, N. Giladi, and J. M. Hausdorff, "Toward automated, at-home assessment of mobility among patients with parkinson disease, using a body-worn accelerometer," *Neurorehabilitation and neural repair*, vol. 25, no. 9, pp. 810–818, 2011.

[20] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National science review*, vol. 5, no. 1, pp. 44–53, 2018.

[21] M. M. Hoehn, M. D. Yahr *et al.*, "Parkinsonism: onset, progression, and mortality," *Neurology*, vol. 50, no. 2, pp. 318–318, 1998.

[22] J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artificial intelligence*, vol. 201, pp. 81–105, 2013.

[23] E. Frank and X. Xu, "Applying propositional learning algorithms to multi-instance data," 2003.

[24] L. Dong, "A comparison of multi-instance learning algorithms," Ph.D. dissertation, The University of Waikato, 2006.

[25] Sivic and Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proceedings ninth IEEE international conference on computer vision*. IEEE, 2003, pp. 1470–1477.

[26] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part IV 9*. Springer, 2006, pp. 490–503.

[27] J. Wu, S. Pan, X. Zhu, C. Zhang, and X. Wu, "Multi-instance learning with discriminative bag mapping," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 6, pp. 1065–1080, 2018.

[28] M. Yang, Y.-X. Zhang, X. Wang, and F. Min, "Multi-instance ensemble learning with discriminative bags," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 9, pp. 5456–5467, 2021.

[29] P. Paul and T. George, "An effective approach for human activity recognition on smartphone," in *2015 IEEE International conference on engineering and technology (ICETECH)*. IEEE, 2015, pp. 1–3.

[30] T. Exley, S. Moudy, R. M. Patterson, J. Kim, and M. V. Albert, "Predicting updrs motor symptoms in individuals with parkinson's disease from force plates using machine learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 3486–3494, 2022.