Up-Sampling Active Learning: An Activity Recognition Method for Parkinson's Disease Patients

Peng Yue[1], Xiang Wang[2], Yu Yang[2], Jun Qi[3], Po Yang[1]

[1] Department of Computer Science Faculty of Engineering,University of Sheffield, Sheffield, UK

[2] National Pilot School of Software,Yunnan University, Kunming, China

[3] Department of Computing, Xi'an JiaoTong-Liverpool University, Suzhou, China

Po.Yang@sheffield.ac.uk, Jun.Qi@xjtlu.edu.cn

Abstract: Parkinson's Disease (PD) is the second most common neurodegenerative disease. With the advancement of technologies of big data, wearable sensing and artificial intelligence, automatically recognizing PD patients' Physical Activities (PAs), health status and disease progress have become possible. Nevertheless, the PA measures are still facing challenges especially in uncontrolled environments. First, it is difficult for the model to recognize the PA of new PD patients. This is because different PD patients have different symptoms, diseased locations and severity that may cause significant differences in their activities. Second, collecting PA data of new PD patients is time-consuming and laborious, which will inevitably result in only a small amount of data of new patients being available. In this paper, we propose a novel up-sampling active learning (UAL) method, which can reduce the cost of annotation without reducing the accuracy of the model. We evaluated the performance of this method on the 18 PD patient activities data set collected from the local hospital. The experimental results demonstrate that this method can converges to better accuracy using a few labeled samples, and achieve the accuracy from 44.3% to 99.0% after annotating 25% of the samples. It provides the possibility to monitor the condition of PD patients in uncontrolled environments.

Keywords: Activity Recognition, Active Learning, Parkinson's Disease, Cross-Subject

1. Introduction

Parkinson's Disease (PD), one of the common chronic diseases in the elderly, is a progressive neurological disorder caused by the loss of dopamine-producing nerve cells in the brain[1]. The estimated number of PD patients will reach 8 million in 2030[2].One of the main challenges is how to monitor the condition of PD patients for a long time to improve efficiently the healthcare of these patients. In the course of treatment, the dosage of the medicine given by the doctor usually vary with the patient's condition. Especially the motor state of the PD patients in late stage may generate different symptoms from patient to patient [3-5]. Usually, patients need to go to the hospital to assess their condition under the guidance of neurologists. In view of the high correlation between the human Physical Activity (PA) index and the PD diagnosis score, the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) is utilised by physicians for the diagnosis of functional symptoms of PD in the early attempts, such as tremor and postural balance disorders [6]. Physical activity recognition (PAR) can provide clinicians with quantitative profiles of motor function behavior in natural environments in long-term period, which can further make treatment strategies objectively adapted.

However, due to the complicated and subjective scoring process, clinically automatic identification of the PAR for PD patients is urgently needed to reduce requirements for physicians as well as save time and effort. As such, researches on objective scoring via computer-aided diagnosis systems have been laid considerably theoretical foundations in previous studies [7-11]. Among them, most attentions focus on either designing standalone novel wearable sensors to achieve highly accurate PAR [12-14], or investigating advanced machine learning algorithms for training features from observed wearable sensory data of human body positions into specific several PA subjects [15-17]. Additionally, some researchers investigate how to attach wearable sensors into the most suitable positions for the best performance [18-21]. While these typical technologies have been capable of achieving satisfactory results, the majority concern on controlled environments (e.g., hospitals) but not uncontrolled environments (e.g., home). However, most patients measure PA after medication in controlled environments and thus the results may impact the overall diagnosis of the symptom [22]. Therefore, lifelong PA monitoring in an uncontrolled environment is extremely significant.

Nevertheless, measuring lifelogging PA currently may have the following obstacles that would affect their practical usefulness. First, in order to achieve high performance, some specific wearable sensors have to be distributed over the human body [23-25], but the accuracy would usually be greatly reduced by using consumer wearable devices in uncontrolled environments. Second, the types of PA are predefined and verified in manual-designed controlled environments, but for the PA data in the uncontrolled environment, PD patients may not be required to accurately record and label data samples for such a long time [26, 27]. Third, the individual conditions and life patterns of each patient are different, and thus the accuracy of PA measurement would be greatly affected by these uncertain factors, making traditional methods inapplicable. Thus, PAR for PD diagnosis in uncontrolled environments is still a very challenging issue [28].

Targeting on the challenges above, we propose a novel up-sampling active learning (UAL) framework by using only 5%-20% of the labeled samples to achieve the training effect of using all the labeled data. The ensemble classifier is exploited to train PA data of the PD patients collected from our local hospital. Under the circumstances that minimize the annotation cost, we explored the effect of different sampling strategies and up-sampling on active learning performance that is able to dynamically discover new patient PA. The proposed method is capable of adapting PD patients' PAR and compared the effects of different sampling strategies and up-sampling on the experiment results. The experimental results demonstrate that by using the active learning of Best-versus-Second Best (BvSB) sampling strategy, the model can converge to the better accuracy. For the 6 new PD patients in the test set, after annotating 5.63% to 25.00% of the data, the model accuracy can reach 99%. For the more severe PD patients, the proportion of labeled samples required to achieve 99.0% classification accuracy is 17.26% with 23.15% samples without annotation.

The remainder of this paper is organized as follows. Section 2 reviews related work on activity recognition. We describe the details of data acquisition in section 3, the section 4 introduces the details of our proposed method up-sampling active learning (UAL), the experimental results are in the section 5, and the section 6 states our conclusions and future work.

2. Related Work

PAR leveraging advanced sensors has been investigated extensively over the past decade. Among the sensor-based methods, devices such as accelerometers, gyroscopes, magnetometers, ambient sensors, and RFID tags are widely used [29-31]. These technologies have shown great potential and are beneficial to medical care services such as clinical interventions for the elderly and patients with chronic diseases [7, 9, 32, 33]. In recent years, increasing number of researches have focused on PAR for PD diagnosis in naturalistic settings with easy-to-easy devices such as mobile phones or wrist band. Cheng et al. [34] proposed a deep neural network to distinguish gait activities (walking, jogging, etc.) from stationary activities (sitting, standing, etc.) between PD patients and healthy subjects with 98% of accuracy. Albert et al. [35] used standard machine learning algorithms to recognize PAs of PD patients and healthy subjects respectively and analysed the different features of PA between PD patients and healthy subjects. MPower [36], a typical mobile app developed using Apple's ResearchKit library, is to remotely monitor PAs of PD patients and then analyze daily changes in their symptom severity and medicine conditions. The platform is a promise to conduct in naturalistic environments.

Deep learning is robust to noise and has high classification accuracy compared with traditional machine learning methods. The study [37] combined Linear Discriminant Analysis (LDA) and Long Short Term Memory (LSTM), and performed non-steady-state circuit tests including stairs, slopes and direction changes in mild PD patients and healthy subjects, achieved 80% F1 score when using LSTM only for the lower body data. However, deep learning requires a large amount of data, large amount of calculation, high cost and long time. Research [38] solves the problem that CNN is not suitable for small-scale and large-scale intra-class noise data sets through data enhancement. Using the enhanced data for model training, the recognition accuracy of CNN on 25 PD patient data reached 86.88%, and the accuracy of CNN was 7% higher than standard machine learning methods.

On the other hand, supervised learning methods require a large number of correct annotated data sets for training, which is often difficult to achieve in a naturalistic environment. Conventional inertial sensor data is difficult to annotate, so regardless of the actual experimental environment or the real environment, researchers generally use video to assist them in annotating [13]. In addition, self-recall and experience sampling are also commonly used to obtain labeled data, but such methods are prone to errors. In order to reduce the amount of data required by the classifier, semi-supervised learning [39-40] and transfer learning [41-43] have also been widely used in the field of PAR. The semi-supervised method predicts unlabeled samples, selects the samples that the classifier has the most confidence in the prediction results, and then adds them to the training set for training. One drawback to this is that the accuracy of the classifier will be reduced when incorrect predictions are added to the training set. Transfer learning leverages the knowledge learned from source domain to predict the target domain which has a small number of labels. This method usually requires that the source domain and the target domain are very similar, which is often difficult to achieve in the real setting.

In contrast to semi-supervised learning, active learning reduces the amount of data required by the model by detecting the most informative samples and asking users to query a label. This method improves the performance of the model and reduces the requirement of the model on the amount of data. A large body of work has used active learning to tackle the issue of label acquisition in activity recognition. For example, Stikic et al. explored and analyzed the effects of semi-supervised learning and active learning on reducing the label requirement of model [44]. Liu et al. analyzed the feasibility of active learning to search for the most informative samples to be labeled in activity recognition, by using the lowest confident level and high disagreement between two classifiers as the active sampling [45]. Study [46] derives a hierarchical Bayesian model, which combines active learning and transfer learning. Finally, they conclude that this method can use fewer tags on the target domain to achieve faster learning. AALO [47] label overlapped activities by active learning. The combination of deep learning and active learning has also been widely used to identify human activities [48-51]. However, there are few studies on the activity recognition of PD patients, especially for new PD patients. In this paper, for ensuring that the classifier converges to good accuracy with fewer annotated examples, we propose an up-sampling active learning method, which makes it possible to recognize the activity of PD patients in a real environment.

3. Data Collection

The participants were recruited from patients with PD symptoms who visited the First People's Hospital of Yunnan Province neurology clinic from August 2020 to January 2021. A total of 18 patients voluntarily participated in the experiment under the premise of obtaining the written informed consent of each patient. Neurologists label the PD patients with any of the (0–4)

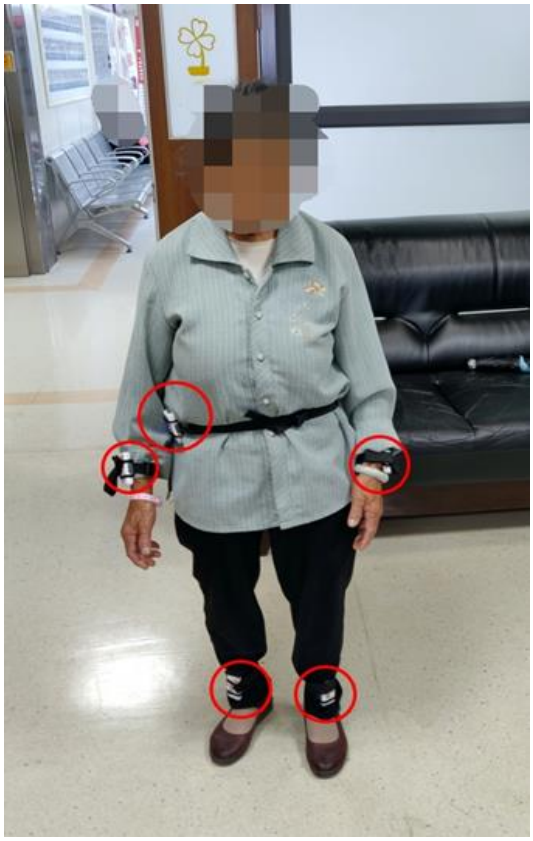


Fig. 1 Sensors wearing position of a PD patient in hospital

Table 1 The description of each activity

NO.	Activity name	Activity description
1	Fingers tapping	Quick pinch with thumb and index finger.
2	Hands closing–opening	Clench your hands and open them quickly.
3	Pronation–supination	Both wrists rotate quickly to the left and right.
4	Leg agility	Sitting in a chair and raising your heels repeatedly.
5	Right hand flip	Continuously pat the left hand with the palm and back of the right hand.
6	Left hand flip	Continuously pat the right hand with the palm and back of the left hand.
7	Finger_to_nose(l)	First touch the tip of your nose with your left index finger, then touch the doctor's index finger, and repeat.
8	Finger_to_nose(r)	First touch the tip of your nose with your right index finger, then touch the doctor's index finger, and repeat.
9	Stand and hand raised flat	Stand and hand raised flat for 30s.
10	Walking back and forth	Walk back and forth in a room.
11	Free walking	Walk freely in the crowd.
12	Sit-to-stand	Cross your hands in front of your chest and stand up from your seat.
13	Drink water	Drink water from a cup.
14	Pick up object	Pick up things from the ground.

MDS-UPDRS score based on the intensity and prevalence of these motor symptoms. After being rated by a professional doctor, among the 18 patients, 8 had mild symptoms, 6 had moderate symptoms, and 4 had severe symptoms with age from 31 to 82 years old, 52% male and 48% female.

First, PD patients wear a shimmer sensor on the right wrist. The sensor is placed in this position to imitate the wristband that people wear in daily life (In fact, we put 5 sensors on the patient's limbs and waist as shown in Fig. 1, but in order to simulate the daily environment, we only used one sensor in this study). Then, we use the 200hz frequency to collect the accelerometer, gyroscope and magnetometer data of the three axes. The full scale range of the sensor is ±2.0g, and sensitivity is 600 mV/g. A Lenovo ThinkPad E440 laptop connects to sensors via Bluetooth and stores data through a software called ConsensysPro.
We collected a total of 14 activities including some activities in the third part of MDS-UPDRS and some daily activities, and each action was collected from 20 seconds to 90 seconds. For daily activities, in order to reflect the differences in the habits of different patients in the real environment, we briefly tell the patients the actions that need to be performed instead of specifying the way to perform them. The activities in the third part of MDS-UPDRS can reflect the difference in the degree of patient disease, so we told the patient in detail how to perform. We described the details of all the activities in Table 1.

4. Methodology

4.1 Problem Formulation

We define activity recognition as a classification problem. Unlike the conventional classification, we will divide PD patients into two parts according to the severity of the patient. The first part simulates the previous patients, including mild, moderate and severe patients, and the other simulates the new PD patients. And then we will perform the same preprocessing and feature extraction on the data of all PD patients. Next, the classification model first learns from the previous PD patient activities data, and then apply the learned model to identify the activities of the new PD patient, which is in line with the actual life situation. We aim to evaluate the performance of the activity recognition model when it faces PD patients in real setting and analyze the reasons behind them. Finally, we propose how to solve this problem through active learning methods.

4.2 Data Preprocessing and Feature Extraction

In order to remove the noise in the raw data, for all signal data, we use the band-pass filtering and standardization by zero-mean normalization (z-score), as Eq. (1) where μ and σ represent the mean and standard deviation of data, respectively.

Table 2 Features extracted from time domain and frequency domain

Category	Feature sets
Time domain	Mean, Standard deviation, Variance, Skewness, Kurtosis, Root mean square, Energy, Median, Range, Correlation
Frequency domain	FFT energy, Mean amplitude, Max amplitude, Spectral entropy

$$x^* = \frac{x-\mu}{\sigma} \tag{1}$$

$$SMV = \sqrt{x_i^2 + y_i^2 + z_i^2} \tag{2}$$

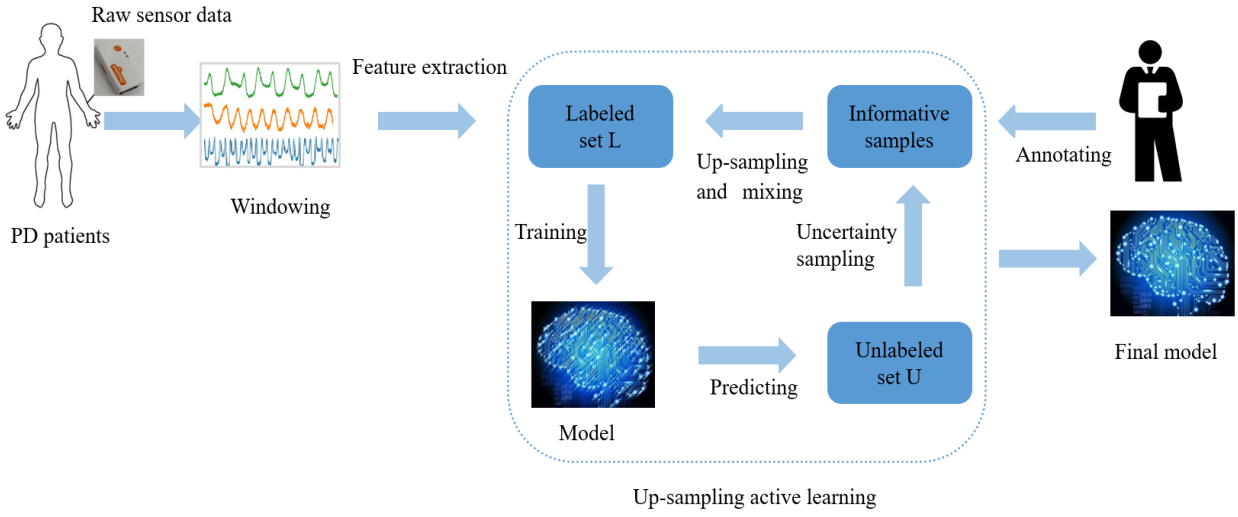
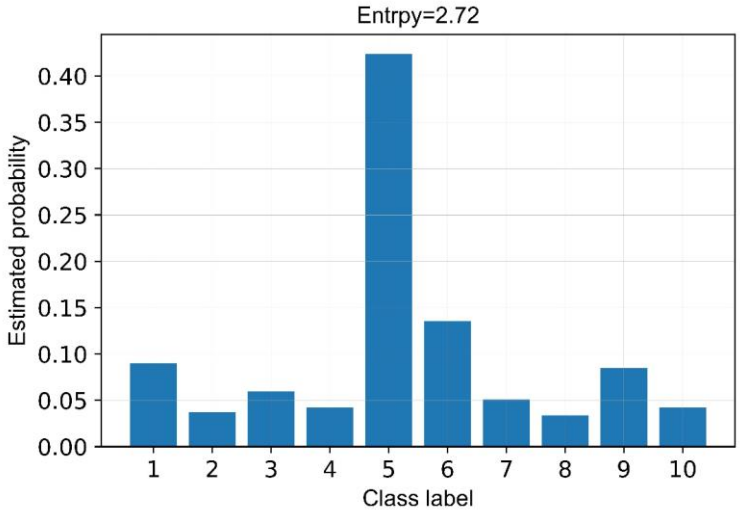


Fig. 2 Overall framework for up-sampling active learning activity recognition model

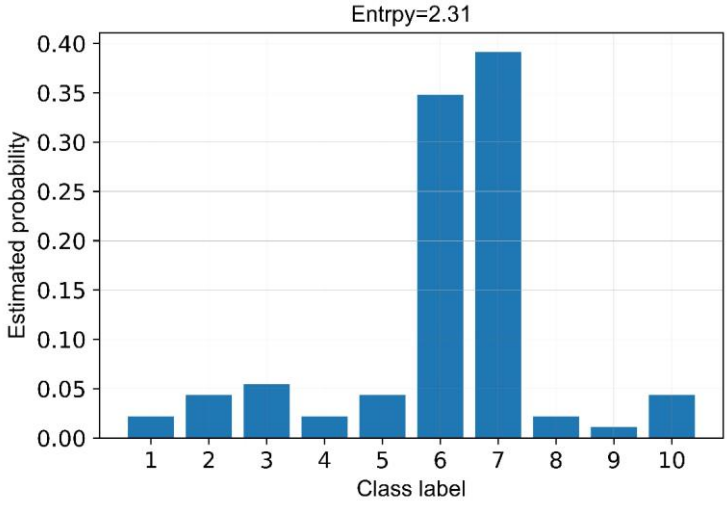
There are 9 dimensions of raw data (a Shimmer sensor consists of three sensors and each of which has three axes of X, Y and Z). Then we got 3 extra axes after calculating signal magnitude vector (SMV) by Eq. (2), which helps to measure the intensity of activities. On each axe, as shown in Table 2, we extract independently the time domain and frequency domain features based on the sliding windows. Based on our experience, we chose a sliding window size of 3 seconds and the data was segmented according to the window overlap ratios of 50%. Of course, we have tried other sliding window sizes and overlap rates, but they have no significant impact on the results, and they are not the focus of our research.

4.3 Up-sampling Active Learning

In the real setting, there are great diversities in performing a same activity between different PD patients. In this work, one of our aims is to train a robust and adaptive activities recognition model for each PD patient, which usually requires abundant labeled data with possible variations. Because annotating the activities of PD patients is time-consuming and laborious, we introduce active leaning to select the most informative samples, so as to ease the burden of activities data annotations. Besides, the proportion of activity data of a new patient is very small, which leads to model to ignore them, so up-sampling is also integrated into active learning to improve the model convergence speed.



(a)



(b)

Fig. 3 The estimated probability of prediction results to two unlabeled sample and its entropy

Based on the above considerations, we propose an UAL pipeline (as illustrated in Fig. 2). We first extract features by sliding window technique from preprocessed data. Starting from an initial labeled PD patients' activity data, we iteratively update the training set by adding new patient activity samples. The steps are as follows: firstly, on previous patients dataset, we train an initial classification model using LightGBM [51]. Then the model will predict the new PD patients' activities and several informative sample will be chosen according to uncertainty sampling strategy BvSB. Next, selected samples will be passed to an oracle to be annotated. Finally, we add the newly annotated sample to previous labeled dataset after up-sampling. This

iteration (the part in the dotted box in Fig. 2) will stop when a certain condition is satisfied to get the final model. The classifier LightGBM and the uncertainty sampling strategy BvSB are demonstrated as below:

LightGBM is an efficient gradient boosting decision tree algorithm (GBDT) proposed by the Microsoft team. It is an improved algorithm for GBDT and an integrated learning algorithm based on Boosting. The traditional Boosting algorithm has some drawbacks, especially in scalability and operating speed, and the emergence of LightGBM has solved these problems. Compared with other Boosting algorithms (such as XG Boost[52]), it can shorten the training time by more than 20 times without reducing the prediction accuracy, and the memory overhead is also greatly reduced. After considering the training time, memory usage and model accuracy, we chose LightGBM as the classification algorithm instead of using SVM, XGBoost, CNN, LSTM, etc., which can make our method more adaptable to the real environment.

Uncertainty sampling strategy: If we predict new patient activities using trained model by other patient activities data, we will able to select the most informative sample to annotate according the uncertainty sampling strategy. By mixing these samples with previous training set and adding them to the model to train, the decision boundary of the model will be changed, which allows the model to have better prediction results for new patient activities. Uncertainty of a prediction result can be quantified by entropy measure, least confident, and BvSB[53]. Among them, BvSB is proved to be excellent in activity recognition based on active learning[10]. Uncertainty sampling strategy using entropy measure is subject to small probability values of unimportant classes. The histogram Fig. 3 shows the estimated probability distribution of two unlabeled samples in a 10-calss classification problem, which explains why entropy cannot calculate the uncertainty well. In Fig. 3 (b), the classifier hesitates between class 6 and 7, but in Fig. 3 (a), the classifier is relatively confident in its prediction results. A prediction result like Fig. 3(a), will be considered to have higher entropy, even if the model is much more confident about the prediction of the example. In the case of multiple classifications, this situation will be even worse. To avoid above problems, we adopted the BvSB as a metrics. BvSB uses the difference between the probability values of the two classes that have the highest estimated probability value as an indicator to measure uncertainty, the sampling criterion can be described as Eq. (3), where $P(y_B|x_i, F_\theta)$and $P(y_{SB}|x_i, F_\theta)$denote the two highest estimated class probabilities output from classifier and u denote all unlabeled data sets. So, BvSB sampling strategy tends to select this prediction situation in Fig. 3(b).

$$x_i^{BvSB} = \arg\min_{x_i, i \in u}(P(y_B|x_i, F_\theta) - P(y_{SB}|x_i, F_\theta)) \quad (3)$$

5. Experiments Results

In this section, we evaluate and analyze the unreliability and instability of directly predicting the activity of a new patient using classifier trained by other subjects' activities dataset. Then, also on the PD patient activity data set collected in hospital, we demonstrated and analyzed the performance of active learning with up-sampling.

5.1 General Experimental Settings

All our experimental data comes from the data set mentioned in section 3. After performing data preprocessing and feature extraction described in section 4, we get 11115 samples from 18 PD patients. We selected 12 patients as the previous patients, and the remaining 6 patients were simulated as new patients in turn. For convenience, the classification accuracy (CA), the proportion of correctly classified samples to the number of test samples, is used as an indicator to evaluate model performance.

5.2 The unreliability and instability of Directly Predicting

In order to evaluate the performance of the PD patient activity recognition model trained by other subjects' activities dataset, we conducted the following experiments.

(1) As differences between patients with different symptoms, different ages etc. will cause unstable results due to the different division of the training set and test set. At the same time, in actual situations, patients are usually tested by wearing the device alone. The Leave-One-Out (LOO) validation method is explored in our experiment, which may derive more stable results and avoid over-fitting in real environment. The LOO method is a special case of cross-validation. Obviously, since there is only one way to divide m samples into m subsets (each subset contains one sample), the method is not affected by the way that how random samples are divided.

(2) Then we use Lightgbm package in the Scikit-learn library to train a PAR model eliminate randomness and select the best hyperparameters of the classification model. The learning rate is set to 0.05, the maximum number of decision trees is 300, and the model converges at 100 training epochs.

(3) We subjectively believe that different patients will have different ways of performing activities, which will make it difficult to apply an activity recognition model to all patients. But this view is ambiguous. In order to intuitively see the difference between different patients, we designed the following scheme: Firstly, we reduce the dimensions of the extracted 161-dimensional features to two dimensions through the t-SNE algorithm. Then we draw the same activity of different patients into a scatter plot based on the reduced dimensionality data. Here, for convenience of display, we only randomly selected 6 patients and 4 activities.

5.3 Results and Discussions on Cross-subject Prediction

Table 3 exhibits the classification accuracy of the model trained by previous dataset for each new patient's activity. The table suggests: 1) If the model is trained by previous data but apply to never seen new patient data, the prediction results will be relatively poor and unstable. It is worth mentioning that the model has a classification accuracy of 99% on the test set of the previous data. 2) The model has different classification accuracy on different patients, the highest is 90.1%, but the lowest is only 44.3%.

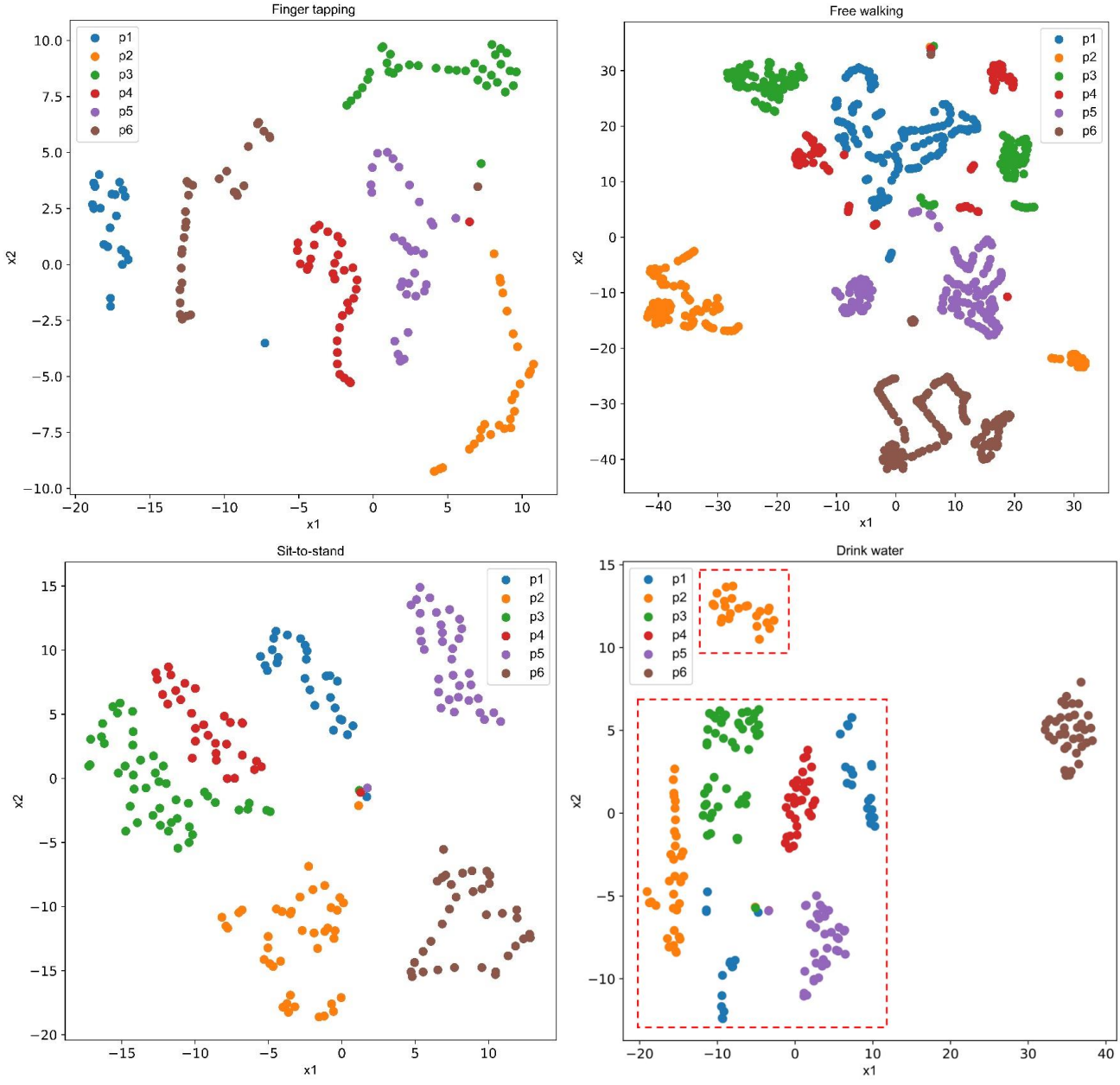


Fig. 4 Feature distribution of 4 kinds of activities in two-dimensional space

recognize the PAs of patients with severe PD.

The main reason for the large fluctuations in the performance of the model among different people is that different patients have different symptoms, severity, and activity habits. These factors firstly lead to differences in the data collected by the sensors, leading to different feature spaces of the data. If a new sample is outside the decision boundary learned by the classifier, it will be difficult for the classifier to make a correct judgment. The classifier has a 90.1% accuracy for patient 4, which is probably because the patient's activity execution method is very similar to some patients in the training set. However, patient 6 was just the opposite of the above, so the classification accuracy is only 44.3%.

Through the visualization results shown in Fig. 4, we can better illustrate this problem, where the horizontal and vertical coordinates are the data obtained after dimensionality reduction of 161 features and the different colors indicate the data of different patients. It can be seen from Fig. 4 that different patients gather in different areas on a two-dimensional plane,

especially the three activities of “finger tapping”, “sit-to-stand”, and “drink water”. Take activity “free walking” as an example. If the model training set only includes the data of patients p1 to p5 (The part in the dotted box in the Fig. 4, then the data of patient p6 will be outside the decision boundary, and the classifier will have difficulty making accurate predictions.

Carefully observe the data distribution of "free walking", we can find that its spatial distribution of activity features is relatively chaotic. The data of patients p1, p3, p4, and p5 are clustered together, which shows that they perform this activity very similarly. However, patients p2 and p6 were gathered in other areas, which shows that they are very different from others in performing this activity. This is mainly because patients p1, p3, p4, and p5 are patients with mild or moderate PD, but patients p2 and p6 are patients with severe PD. And the mobility of severe patients is severely restricted, so they cannot complete this activity well. It is worth mentioning that even if patients 2 and 6 are both severe PD patients, their data distribution is still quite different, which explains why the training set contains data from severe PD patients, but the classifier still unable to accurately.

5.4 Active Learning with Up-Sampling

In order to verify and evaluate the efficiency of our proposed method and the effectiveness of the up-sampling, we conducted the following experiments. Firstly, 70% of the data of each new patient is used as a candidate set, and the remaining 30% is used to test the accuracy on the new data set. Then, we train a classifier by previous patients training set. When the model converges, we make predictions on the test set and candidate set at the same time, and record the accuracy of the model on the test set. By calculating class probabilities of samples in the candidate set, we select samples for annotation according to the sampling strategy. Next, we copy the newly annotated data into the same five copies and mix them all into the previous training set for the classifier retraining in next iteration. It is worth noting that, after comprehensively considering the calculation amount and performance of the model, we select 5 samples for annotation in each iteration. We compare the performance of three different sampling strategies, BvSB, entropy and random, and compare the effect of up-sampling on the results.

Table 3 The classification accuracy to 6 new patients with the model trained by the previous 12 patients

Patient ID	1	2	3	4	5	6
Severity of illness	moderate	severe	mild	mild	mild	severe
Accuracy (%)	66.6±1.6	59.1±0.7	74.5±1.0	90.1±1.9	64.3±1.1	44.3±0.5

Table 4 The number of labeled samples and the proportion of the total samples required to achieve 99% accuracy with different methods

Method / Patient	UAL(Ours)	BvSB	entropy	random
Patient 1	**70 (19.18%)**	105 (28.77%)	90 (24.66%)	340 (93.15%)
Patient 2	**75 (15.96%)**	100 (21.28%)	105 (22.34%)	365 (77.66%)
Patient 3	**70 (15.91%)**	80 (18.18%)	105 (23.86%)	195 (44.32%)
Patient 4	**20 (5.63%)**	25 (7.04%)	35 (9.86%)	165 (47.14%)
Patient 5	**85 (20.48%)**	125 (30.12%)	175 (42.17%)	260 (63.41%)
Patient 6	**105 (25.0%)**	135 (32.14%)	170 (40.48%)	305 (72.62%)

Table 5 The highest classification accuracy value (the first number) that each method can achieve and the proportion of the total samples required to achieve corresponding highest accuracy (the second number)

Method / Patient	UAL(Ours)	BvSB	entropy	random
Patient 1	**100% (20.5%)**	100**%** (41.1%)	100**%** (34.2%)	100**%** (100.0%)
Patient 2	**99.51% (22.3%)**	**99.61% (23.4%)**	99.51**%** (26.6%)	99.51**%** (78.7%)
Patient 3	**100% (40.9%)**	100**%** (55.7%)	100**%** (68.2%)	100**%** (62.5%)
Patient 4	**100% (8.5%)**	**100% (8.5%)**	100**%** (25.4%)	100**%** (85.7%)
Patient 5	**100% (28.9%)**	100**%** (69.9%)	100**%** (75.9%)	99.78**%** (85.4%)
Patient 6	**99.89% (51.2%)**	99.78**%** (66.7%)	99.78**%** (57.1%)	99.57**%** (95.2%)

5.5 Results and Discussions on Active Learning with Up-Sampling

Fig. 5 presents the curves of the classification accuracy values of 6 new patients as a function of the number of labelled samples using three different sampling policies, where the x-axis is the number of annotated samples and the y-axis is the classification accuracy of the classifier on the test set.

The number of labeled samples and the proportion of the total samples required to achieve corresponding highest accuracy are listed in Table 5. The best results in table 4 and 5 are shown in bold. From Fig. 5, table 4 and 5, we can see that:

1) For all patients, our proposed UAL method can achieve 99% accuracy with the smallest number of labeled samples, followed by active learning methods based on BvSB and entropy, and the worst random sampling. We tend to randomly select some samples for labeling when active learning is not applied, which is often very inefficient. We attribute this result as follows: Firstly, the BvSB sampling strategy allows the model to quickly and accurately refine the decision boundary by selecting the

most informative samples, which allows the model to quickly adapt to samples of new patients and make high-precision predictions. Secondly, through the up-sampling, the weight of the newly labeled data in the model can be increased, which allows the model to converge more quickly.

2) The sampling strategy based on BvSB tends to improve the accuracy faster than the sampling strategy based on entropy. It's because BvSB strategy is inclined to choose samples meeting Eq. (3) in section 4, such samples are just on the decision boundary of the two classes. But the entropy sampling scheme tends to select samples whose estimated probabilities are scattered over all classes with similar probability values, such samples may not have high uncertainty as described in section 4.

3) For patients 2 and 6, when we didn't annotate their activities data, the classification accuracy was only 59.1% and 44.3% respectively. Because they are patients with severe PD and exhibits symptoms rigidity and bradykinesia, they are very slow to perform an action, which is very different from patients with mild PD. For example, patient 4, he is a mild PD patient, he performs these activities smoothly and standardly, which is the same as most people in the training set, so the classifier's classification accuracy for him is 90% at the beginning. But with our method UAL, the classifier only needs 20 additional labeled samples to achieve a classification accuracy of 99%, which fully demonstrates the efficiency of UAL.

4) It can be seen from Table 5 that for patients 1, 3, 4, and 5, all methods except random sampling can achieve 100% classification accuracy, but our proposed method requires the least number of labeled samples. For patients 2 and 6, our method UAL and BvSB method are almost the same in accuracy, but our method requires fewer labeled samples. The difference between UAL and BvSB is that we have added up-sampling to increase the weight of the data in the training process, which allows the model to improve accuracy faster.

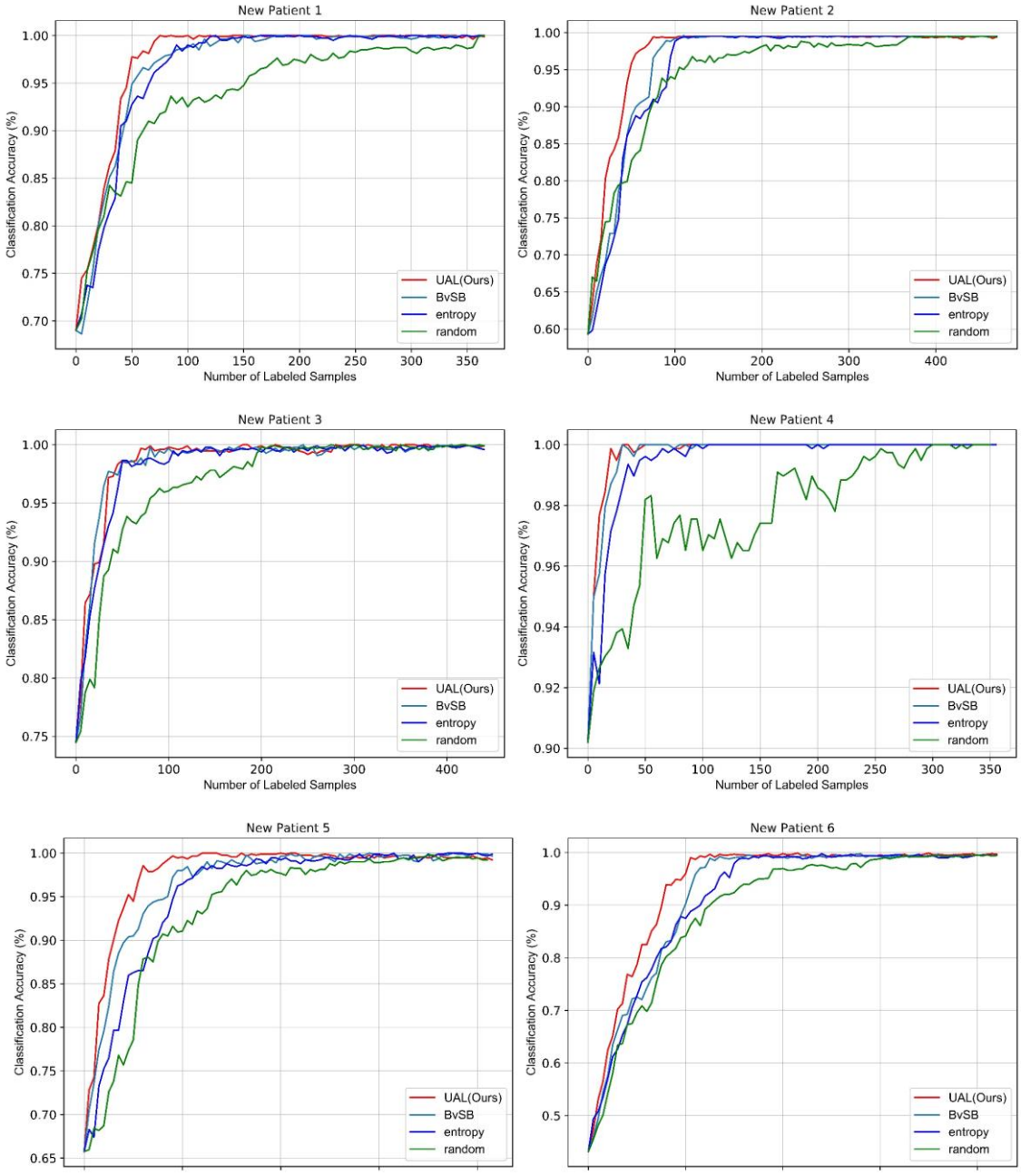


Fig. 5 The curves of the classification accuracy values of 6 new patients as a function of the number of labelled samples

6. Conclusions

In this paper, we analyzed the difficulties in identifying the activities of PD patients. An up-sampling active learning method was proposed for effective long-term monitoring of PD patients. This method iteratively selects the most informative samples for labeling, and then adds them to the training set after up-sampling. We conducted experiments on the data set collected from the local hospital to evaluate the performance of UAL. The experimental results show that the effect of UAL is different for patients with different symptoms. For the 6 new PD patients after annotating 5.63% to 25.00% of the data, the model accuracy can reach up to 99%. Among them, for severe PD patients, this method has a more obvious performance improvement and need more labeled data. Compared with other active learning sampling strategies, the UAL method that combines BvSB sampling strategy and up-sampling trick can converge to the better accuracy, which represents a lower annotation cost. In future work, we plan to combine active learning and transfer learning to further reduce annotation costs and improve the robustness of the model, and analyze the progress of the patient's condition through the patient's activity data.

References

1. Jankovic J: Parkinson's disease: clinical features and diagnosis. 79 (4):368-376 (2008)
2. Dorsey ER, Constantinescu R, Thompson J, Biglan K, Holloway R, Kieburtz K, Marshall F, Ravina B, Schifitto G, Siderowf AJN: Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. 68 (5):384-386 (2007)
3. Nutt JG, Gancher ST, Woodward WRJN: Does an inhibitory action of levodopa contribute to motor fluctuations? 38 (10):1553-1553 (1988)
4. Merello M, Lees AJJoN, Neurosurgery, Psychiatry: Beginning-of-dose motor deterioration following the acute administration of levodopa and apomorphine in Parkinson's disease. 55 (11):1024-1026 (1992)
5. Maetzler W, Klucken J, Horne MJMD: A clinical view on the development of technology-based tools in managing Parkinson's disease. 31 (9):1263-1271(2016)
6. Albani G, Ferraris C, Nerino R, Chimienti A, Pettiti G, Parisi F, Ferrari G, Cau N, Cimolin V, Azzaro CJS: An integrated multi-sensor approach for the remote monitoring of Parkinson's disease. 19 (21):4764 (2019)
7. De Pessemier T, Martens LJMT: Heart rate monitoring, activity recognition, and recommendation for e-coaching. Multimedia Tools and Applications, 77(18), 23317-23334 (2018)
8. Ryder J, Longstaff B, Reddy S, Estrin D Ambulation: A tool for monitoring mobility patterns over time using mobile phones. In: 2009 International Conference on Computational Science and Engineering, 2009. IEEE, pp. 927-931 (2009)
9. Emmanouil G Spanakis, Dimitris Kafetzopoulos, Po Yang, Kostas Marias, Zhikun Deng, Manolis Tsiknakis, Vangelis Sakkalis, Feng Dong: MyHealthAvatar: Personalised and Empovermnet Health Services through Internet of Things technologies. In: 2014 4th International Conference on Wireless Mobile Communication and Healthcare Transforming Healthcare through Innovatins in Mobile and Wireless Technologies (MOBIHEALTH), 2014. IEEE, pp.331-334, (2014)
10. Bi H, Perello-Nieto M, Santos-Rodriguez R, Flach PJIJoB, Informatics H: Human Activity Recognition Based on Dynamic Active Learning. IEEE Journal of Biomedical and Health Informatics, 25(4), 922-934 (2020)
11. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y: Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: A systematic review. 87, 138-153 (2018)
12. Zhang M, Sawchuk AA USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp 1036-1043 (2012)
13. Twomey N, Diethe T, Kull M, Song H, Camplani M, Hannuna S, Fafoutis X, Zhu N, Woznowski P, Flach P: The SPHERE challenge: Activity recognition with multimodal sensor data. arXiv preprint arXiv:1603.00797. (2016)
14. Martín H, Bernardos AM, Iglesias J, Casar J: Activity logging using lightweight classification techniques in mobile devices. 17 (4):675-695 (2013)
15. Kwapisz JR, Weiss GM, Moore S: Activity recognition using cell phone accelerometers. 12 (2):74-82 (2011)
16. Xu H, Pan Y, Li J, Nie L, Xu XI. Activity recognition method for home-based elderly care service based on random forest and activity similarity. IEEE Access, 7, 16217-16225 (2019)
17. Cook DJ, Krishnan NC, Rashidi PJ: Activity discovery and activity recognition: A new partnership. 43 (3):820-828 (2013)
18. Khan AM, Tufail A, Khattak AM, Laine T: Activity recognition on smartphones via sensor-fusion and kda-based svms. 10 (5):503291 (2014)
19. Ouchi K, Doi M: Indoor-outdoor activity recognition by a smartphone. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing. pp. 600-601 (2012)
20. Ha S, Choi S: Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In: 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, pp 381-388 (2016).
21. Bianchi V, Bassoli M, Lombardo G, Fornacciari P, Mordonini M, De Munari I: IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment. 6 (5):8553-8562 (2019)
22. Mutegeki R, Han DS: A CNN-LSTM approach to human activity recognition. In: 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), IEEE, pp 362-366 (2020)
23. Alawneh L, Mohsen B, Al-Zinati M, Shatnawi A, Al-Ayyoub M A Comparison of Unidirectional and Bidirectional LSTM Networks for Human Activity Recognition. In: 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE, pp 1-6 (2020)

24. Chen L, Hoey J, Nugent CD, Cook DJ, Yu ZJ: Sensor-based activity recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6), 790-808 (2012)
25. Thomaz E, Essa I, Abowd GD: A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp 1029-1040 (2015)
26. Merck CA, Maher C, Mirtchouk M, Zheng M, Huang Y, Kleinberg S Multimodality sensing for eating recognition. In: PervasiveHealth, 2016. pp 130-137
27. Qi J, Yang P, Min G, Amft O, Dong F, Xu L: Advanced internet of things for personalised healthcare systems: A survey. Pervasive and Mobile Computing, 41, 132-149 (2017)
28. Peng X, Yang Y, Wang X, Li J, Qi J, Yang P: Experimental analysis of artificial neural networks performance for accessing physical activity recognition in daily life. In 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom) (pp. 1348-1353). IEEE. (2020).
29. Vrigkas M, Nikou C, Kakadiaris IAJFiR, AI A review of human activity recognition methods. Frontiers in Robotics and AI, 2, 28 (2015)
30. Krishnan NC, Cook DJ: Activity recognition on streaming sensor data. Pervasive and mobile computing, 10, 138-154. (2014)
31. Zhang M, Sawchuk AAJIjoB, Informatics H: Human daily activity recognition with sparse representation using wearable sensors. 17 (3):553-560 (2013)
32. Basilakis J, Lovell NH, Redmond SJ, Celler BG.: Design of a decision-support architecture for management of remotely monitored patients. ieee transactions on information technology in biomedicine, 14(5), 1216-1226. (2010)
33. Bao L, Intille SS: Activity recognition from user-annotated acceleration data. In International conference on pervasive computing. Springer, pp 1-17 (2004)
34. Cheng WY, Scotland A, Lipsmeier F, Kilchenmann T, Jin L, Schjodt-Eriksen, J, et al: Human activity recognition from sensor-based large-scale continuous monitoring of Parkinson's disease patients. In 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE) pp 249-250 (2017)
35. Albert, MV, Toledo S, Shapiro M, Kording K: Using mobile phones for activity recognition in Parkinson's patients. Frontiers in neurology, 3, 158 (2012)
36. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, Trister, AD: The mPower study, Parkinson disease mobile data collected using ResearchKit. Scientific data, 3(1), 1-9 (2016)
37. Kazemimoghadam M, Fey NP An Activity Recognition Framework for Continuous Monitoring of Non-Steady-State Locomotion of Individuals with Parkinson's Disease. Applied Sciences, 12(9), 4682 (2022)
38. Kaur S, Aggarwal H, Rani R: Diagnosis of Parkinson's disease using deep CNN with transfer learning and data augmentation. Multimedia Tools and Applications, 80(7), 10113-10139 (2021)
39. Balabka D: Semi-supervised learning for human activity recognition using adversarial autoencoders. 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. pp 685-688 (2019)
40. Ma Y, Ghasemzadeh H Labelforest: non-parametric semi-supervised learning for activity recognition. Proceedings of the AAAI Conference on Artificial Intelligence. vol 01. pp 4520-4527 (2019)
41. Qin X, Chen Y, Wang J, Yu C: Cross-dataset activity recognition via adaptive spatial-temporal transfer learning. 3 (4):1-25 (2019)
42. Wang J, Zheng VW, Chen Y, Huang M: Deep transfer learning for cross-domain activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(4), 1-25 (2019)
43. Khan MAAH, Roy N, Misra A: Scaling human activity recognition via deep learning-based domain adaptation. In: 2018 IEEE international conference on pervasive computing and communications (PerCom),. IEEE, pp 1-9 (2018)
44. Stikic M, Van Laerhoven K, Schiele B: Exploring semi-supervised and active learning for activity recognition. In: 2008 12th IEEE International Symposium on Wearable Computers. IEEE, pp 81-88 (2008)
45. Liu R, Chen T, Huang L: Research on human activity recognition based on active learning. In: 2010 International Conference on Machine Learning and Cybernetics. IEEE, pp 285-290 (2010)
46. Diethe T, Twomey N, Flach PA: Active transfer learning for activity recognition. In: ESANN (2016).
47. Hoque E, Stankovic J AALO: Activity recognition in smart homes using Active Learning in the presence of Overlapped activities. In: 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops, 2012. IEEE, pp 139-146
48. Hossain HS, Al Haiz Khan MA, Roy N:DeActive: scaling activity recognition with active deep learning. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(2), 1-23. (2018)
49. Wang D, Shang Y: A new active labeling method for deep learning. 2014 International joint conference on neural networks (IJCNN). IEEE, pp 112-119 (2014)
50. Zhou S, Chen Q, Wang X: Active deep networks for semi-supervised sentiment classification. In: Coling 2010: Posters. pp 1515-1523 (2010)
51. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-YJ: Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30, 3146-3154 (2017)
52. Chen T, Guestrin C: Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp 785-794 (2016)
53. Joshi AJ, Porikli F, Papanikolopoulos N Multi-class active learning for image classification. 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp 2372-2379 (2009).