Daily Activity Recognition and Tremor Quantification from Accelerometer Data for Patients with Essential Tremor Using Stacked Denoising Autoencoders

Qin Ni[1] · Zhuo Fan[1] · Lei Zhang[2] · Bo Zhang[1] · Xiaochen Zheng[3] · Yuping Zhang[1]

Abstract

Human activity recognition (HAR) has received more and more attention, which is able to play an important role in many fields, such as healthcare and intelligent home. Thus, we have discussed an application of activity recognition in the healthcare field in this paper. Essential tremor (ET) is a common neurological disorder that can make people with this disease rise involuntary tremor. Nowadays, the disease is easy to be misdiagnosed as other diseases. We have combined the essential tremor and activity recognition to recognize ET patients' activities and evaluate the degree of ET for providing an auxiliary analysis toward disease diagnosis by utilizing stacked denoising autoencoder (SDAE) model. Meanwhile, it is difficult for model to learn enough useful features due to the small behavior dataset from ET patients. Thus, resampling techniques are proposed to alleviate small sample size and imbalanced samples problems. In our experiment, 20 patients with ET and 5 healthy people have been chosen to collect their acceleration data for activity recognition. The experimental results show the significant result on ET patients activity recognition and the SDAE model has achieved an overall accuracy of 93.33%. What's more, this model is also used to evaluate the degree of ET and has achieved the accuracy of 95.74%. According to a set of experiments, the model we used is able to acquire significant performance on ET patients activity recognition and degree of tremor assessment.

Keywords Essential tremor · Activity recognition · Stacked denoising autoencoder · Acceleration data · Deep learning

1 Introduction

Essential tremor (ET) is a common neurological disorder that usually causes the hands, head and other parts of body tremor. It is a mild single symptom disease whose exclusive symptom is tremor and it belongs to postural tremor that the disease attack when patients remain a posture. ET has affected about 4% of the population above 65 years of age in the world [1]. ET patients have usually faced serious difficulties when they perform activities (e.g., drinking, eating and writing) in their daily lives. The unconscious body tremor cause great inconvenience even danger for them. The disease imposes large restrictions on patients' living and working styles. It would be very meaningful and useful for society and people if essential tremor can be treated effectively.

Qin Ni and Zhuo Fan contributed equally to this work and should be considered co-first authors.

✉ Lei Zhang
lei.zhang@dhu.edu.cn

Qin Ni
niqin@shnu.edu.cn

Zhuo Fan
1000459457@smail.shnu.edu.cn

Bo Zhang
zhangbo@shnu.edu.cn

Xiaochen Zheng
xiaochen.zheng@alumnos.upm.es

Yuping Zhang
yp_zhang@shnu.edu.cn

[1] College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, No. 100, Haisi Road, Fengxian District, Shanghai 201418, China

[2] College of Information Science and Technology, Donghua University, No. 2999, North Renmin Road, Songjiang District, Shanghai 201620, China

[3] Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, Madrid, Spain

However, this disease has neither clear etiology nor unified diagnostic criteria, which is usually misdiagnosed as tremor caused by other diseases, such as Parkinson's disease (PD) [2]. Generally speaking, the diagnosis of ET depends on neurologist's professional knowledge and experience through face-to-face interaction. Therefore, it is difficult and inconvenient for doctors to monitor the patients' status precisely at all times. It would be useful to find an efficient method that provides an auxiliary analysis for monitoring and evaluating the current status of ET. Human activity recognition offers valuable considerations.

Human activity recognition (HAR), as an important application, is widely used in many social fields, especially healthcare [3]. For example, HAR has combined with fall detection technology to monitor older people's health [4]. HAR is able to identify people's activities according to their behavior information collected by utilizing different tools, such as sensor [5] and camera [6]. Generally speaking, the mobile devices based on a variety of sensors (e.g., accelerometer, gyroscope, barometer) are widely used. On the one hand, mobile devices have no space constraints, which is able to collect data in flexible manners. On the other hand, sensors embedded into mobile devices are enough to satisfy the demands of experiments in the real laboratory environment. In this paper, we utilized smartwatch which was embedded in a tri-axis accelerometer to collect behavior information for ET patients. According to the accelerometer, acceleration data from subjects are acquired easily, which carries human movement tendency [7]. Then these data can be used to recognize ET patients' daily activities and evaluate the degree of ET.

Patients with an essential tremor have usually suffered postural tremor or action tremor. Its frequency is 4–12 Hz in common tremor and become 8–12 Hz when ET patients suffered seizures [8]. Until now, ET has neither clear etiology nor unified diagnostic criteria, which causes many difficulties in diagnosing and analyzing the disease. The main diagnostic method is face-to-face interview by neurologists, which depends on largely their experience. In these fields, HAR can be used to provide an auxiliary analysis for evaluating the degree of ET. Utilizing activity recognition technique to ensure ET patients' state and then evaluate the tremor degree. When patients have tremors of different degrees, their daily activities would show more obvious fluctuation in a variety of styles. And meaningful features can be extracted from ET patients' individual behavior way to evaluate the disease level. Thus, the combination of ET and HAR has a bright development prospect. In this paper, we focus on the activity recognition of ET patients and the auxiliary analysis of estimating ET degree.

Human activity recognition can be conducted by conventional machine learning methods, such as SVM, DT, KNN and so on [9–11]. So far HAR has achieved significant performance by using these machine learning methods. However, there is a drawback about machine learning methods that they need to extract features manually, which will no doubt consume a large amount of time and computing resource. Thus, deep learning models which are able to extract features automatically and acquire better performance than conventional machine learning methods based on large-scale dataset have received more and more attention. In our experiment, deep learning methods are utilized. However, the scale of the dataset is small due to the limitation of collecting data form ET patients, which will cause lower accuracy because model cannot learn enough knowledge from small-scale dataset. Therefore, we consider using data augmentation methods to enlarge data. The resampling techniques are used to alleviate above problem.

In this paper, stacked denoising autoencoder (SDAE) is exploited to conduct activity recognition based on ET patients. The SDAE, as a deep learning model, is able to extract features automatically. Moreover, it utilizes encoder and decoder to compress and reconstruct input data, which can learn a higher level of representations and extract more useful features. What's more, the data collected from ET patients usually mix up much noise, SDAE model is able to reduce the impact of noise efficiently. The dataset is acceleration data collected from ET patients by using tri-axis accelerometer of smartwatch. It is noted that the angle data between smartwatch and ground is also adopted to recognize activities in order to extract effective features. Six standard activities are performed, which include extending arms, touching nose, writing sentences, drawing spirals, pouring water and simulating drinking.

The main contribution in this study are described as follows:

1. In this paper, the human activity recognition is combined with essential tremor to explore the concrete application of HAR on healthcare, which uses behavior data of ET patients to recognize their daily activities.
2. The SDAE model is utilized to extract potential features of motion data from ET patients and reduce the negative impact of noise. Meanwhile, resampling techniques are used to solve small sample size problem. Result has achieved a significant performance.
3. Besides the activity recognition of ET patients by utilizing acceleration data, auxiliary analysis is also discussed for the degree of tremor assessment.

The rest of this paper is organized as follows: Sect. 2 will introduce existing related works about essential tremor and activity recognition. Section 3 describes the overall framework about the architecture of the stacked denoising autoencoder model. Section 4 will design the experimental procedure and propose the methods of data preprocessing. And

the results of the experiment will be presented and analyzed. Finally, the conclusion will be drawn in Sect. 5.

2 Related Work

2.1 Essential Tremor

Essential tremor, as an one of the most common neurological disorder, has neither clear etiology nor unified diagnostic criteria. Thus, the ET may be misdiagnosed as other tremor disease (e.g. Parkinson's disease) in clinical diagnosis [12]. For diagnosing the ET precisely, the study [13] utilized the machine learning framework based on principal components analysis (PCA) and support vector machine (SVM) to identify PD and ET automatically. A decision tree (DT) based algorithm [14] was proposed to classify the ET patients' state and then predict when tremor would reappear. Moreover, to evaluate the ET severity reasonably, the rating scales were defined for the clinical assessment of ET [15]. The essential tremor rating scale was also used to test the performance of deep brain stimulation in the treatment of ET [16]. However, the rating of ET was usually made subjectively, which was difficult to achieve enough reliability [17]. Thus, besides the rating scale, researches have proposed other methods to extract features of ET for evaluating and diagnosing. The tremor signals were collected by different sensors, such as accelerometer and gyroscope, to analyze the severity of essential tremor [18, 19]. By utilizing these signals, features were extracted by specific methods which involved conventional machine learning algorithms or deep learning models. In this paper, a wearable smartwatch embedded with accelerometer was used to collect activity signals from ET patients. Moreover, the SDAE model was utilized to recognize ET patients' activities based on their signals collected by accelerometer.

2.2 Human Activity Recognition

In human activity recognition, there are variety of tools to collect human activity information. Recently, the smart devices are used widely. For example, smartphones and smartwatches which are embedded with sensors have enabled the human activity data to be gathered large scale. By utilizing these devices which are embedded with sensors, the study [20] presented a wearable interial sensor network to collect motion signals of human activities and its associated algorithm achieved high accuracy. After collecting data, there is an essential question that how to extract useful features from raw data. Feature extraction is the core of recognition algorithms, which have a significant impact on performance, computation time and complexity [21]. And conventional machine learning classification methods are often used to classify based on the extracted features. For instance, in [22], the paper exploited Relief-F and sequential forward floating search (SFFS) to select proper features from a range of previously extracted features. Then the activity recognition was performed by applying Naive Bayes and k-nearest neighbor. There are also other machine learning methods, such as decision tree (DT) [23], SVM [24], random forest (RF) [25] and K-means [26]. A paper exploited discrete cosine transform to extract effective features and used PCA to reduce the dimension of feature [27]. At last, they applied multi-class support vector machine (SVM) to recognize human activity. Decision tree (DT) classifier [28] has the preferable performance in recognizing daily activities. These methods have already achieved significant performance. However, a key point should be paid more attention to, which conventional machine-learning algorithm need to extract features manually [29], which increases the complexity of model, time consuming and energy. Thus, deep learning model which extracts features automatically has received more considerable attention. There are a variety of deep learning models applied in human activity recognition, such as convolutional neural network (CNN) [30], autoencoder [31], recurrent neural network (RNN) [32, 33], long short-term memory (LSTM) [34, 35], deep belief networks (DBN) [36, 37] and so on. These methods were able to be used to extract features from data automatically. In this study, stacked denoising autoencoder is utilized for ET patients' activity recognition. Compared to other models, SDAE model is able to improve effectively the problem of gradient disappearance and compress data by utilizing encoder, which extracts more representative features. Moreover, stacked denoising autoencoder is able to reach robust performance when processing noisy data.

3 Methodology and Framework

The study has followed the Declaration of Helsinki and was approved by the ethics committee of the Getafe University Hospital (Madrid, Spain). All subjects in the experiment have signed the informed consent. In this section, the overall framework of the experiment is described in detail. The principle of SDAE model is also be introduced. The framework includes two parts: data preprocessing and SDAE model, which is illustrated in Fig. 1.

3.1 Data Preprocessing

The acceleration data from ET patients include many unconscious tremors, which would cause large confusion for the activity recognition. From Fig. 1, it is obvious that the curves of raw data collected from ET patients contains abnormal

Fig. 1 The overall framework of SDAE-based activity recognition
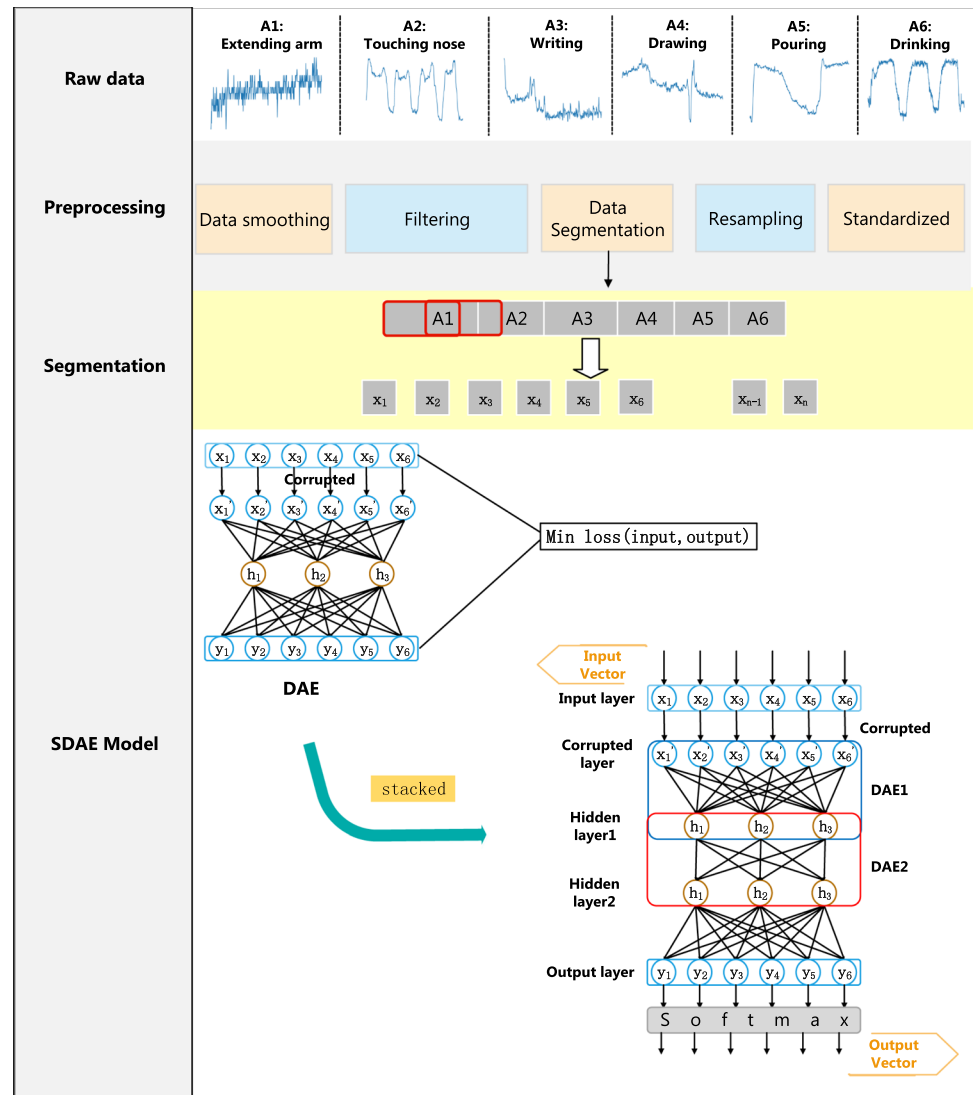
trembles. Thus, it is necessary to reduce the impact of insignificant tremor by preprocessing data.

3.1.1 Filtering

In experimental design progress, six standard activities were performed by twenty ET patients, namely, extending arms, touching nose, writing sentences, drawing spirals, pouring water and simulating drinking. All experimental data are collected by leveraging tri-axis accelerometer of smartwatch, $x_i = \{ac_i^x, ac_i^y, ac_i^z, an\}$, an represents the angle data between smartwatch with ground. Before using input data to train model, the preprocessing procedure has to be carried out. For subjects with ET, their activity information contains much unconscious tremor which has a negative impact on performance of recognition. From Fig. 2, it can been seen that the curves of raw data collected from ET patients contains more trembles than healthy subjects.

Thus, data preprocessing is an essential step for reducing noise influence and enhancing data equality.

In this progress, data smoothing was first executed due to the many fluctuations in activity signals to decrease signal jitters by moving average filter which is described as follows:

$$y(n) = \frac{x(n) + x(n-1) + \cdots + x(n-N+1)}{N}, \quad (1)$$

where N represents the size of the filter, which is set to 5. $x(n)$ is input data and $y(n)$ is output data.

The result of data smoothing is illustrated in Fig. 3. We choose a section of *x-axis* signals to demonstrate smooth result. It is obvious that the raw data signals have more intense tremble which represents much noise than smooth signals. Then a third-order low-pass Butterworth filter with cutoff frequency at 12 Hz was used to remove

Fig. 2 The data signals of patients and healthy subjects

(a) Data signals of healthy people

(b) Data signals of ET patients
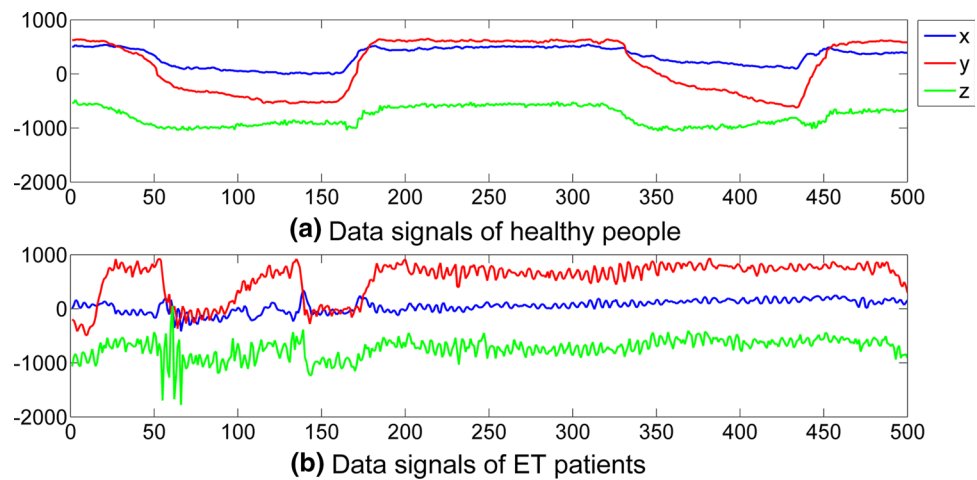
Fig. 3 The smooth result of raw data

(a) Raw Data Signals

(b) The Smoothing Signals
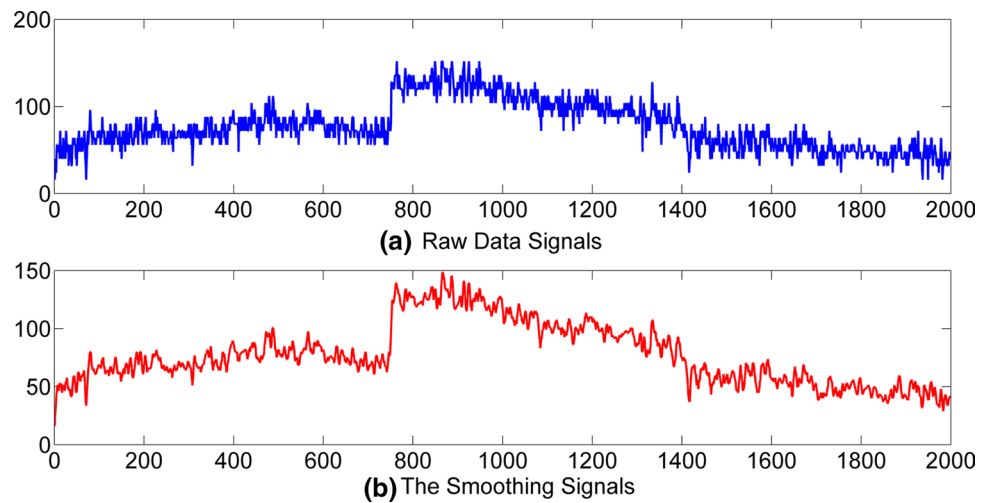
high-frequency noise. And a third-order median filter was used to remove abnormal noise. We also used a third-order high-pass Butterworth filter with cutoff frequency at 0.005 Hz to alleviate the influence of gravity. Moreover, to promote the performance of the model and alleviate the impact of imbalanced samples and small sample size problem in the experiment, the data segmentation and resampling techniques were used.

3.1.2 Data Segmentation

Data segmentation is an essential procedure for human activity recognition due to the continuity of human activities. In reality, human activities are usually continuous, just a single data sample at a time point cannot reflect concrete tendency and feature of an activity. Thus, data segmentation should be carried out before training neural network. In our experiment, the sliding window is utilized to segment dataset, which has a significant influence on recognition performance. To be specific, the dataset is segmented by

integrating n samples as a sequence according to sampling rate, namely

$$x_{ac} = \{ac_k^x, \ldots, ac_{k+n-1}^x, ac_k^y, \ldots, ac_{k+n-1}^y, ac_z^x, \ldots, ac_{k+n-1}^z\}. \tag{2}$$

$$x_{an} = \{an_k, an_{k+1}, \ldots, an_{k+n-1}\}. \tag{3}$$

$$x_i = \{x_{ac}, x_{an}\}. \tag{4}$$

where n is set to 128 in this paper, which is decided by sample rate of 25 Hz and time interval of 5.12 s. In this progress, a sliding window whose size is 128 with a 25% overlap is utilized to segment data.

3.1.3 Resampling

After segmentation, we found that the samples existed problem of imbalance, in which the pouring activity had 1544 samples while the drawing activity had 945 instances. Generally speaking, the problem of unbalanced samples would

drop the accuracy of model. If dataset is unbalanced, the model might tend to output the classification with more data and has a worse fitting ability for the category with less data. Thus, to improve the unbalanced samples problem, the common method is resampling technique, which included random undersampling, random oversampling and synthetic minority oversampling technique (SMOTE) [38–40].

Random undersampling drops directly majority class samples to balance the training dataset, but it may discard some meaningful information. Random oversampling can obtain desired sample size by duplicating some instances from minority class examples. However, it would cause overfitting problem. The cluster-based oversampling utilizes K-means algorithm to cluster minority class and majority class, respectively. Then the oversampling technique is used in every clustering to ensure they have same number of samples. But it may cause overfitting problem too. The SMOTE is an improved method based on random oversampling, which analyses and synthesizes samples from minority class, then adds these new samples to dataset. More specifically, this method chooses a representative data subset which has been taken as instance to create similar and new examples. Then these examples would be added to the original dataset. On one hand, it is able to alleviate the overfitting problem, on the other hand, it don't discard meaningful information.

What's more, the dataset includes acceleration data and angle data, which have different units. Thus, data standardization is used to unify dimensions. In the end, the whole dataset was divided into training set, validation set and testing set with the radio of 7:1:2.

3.2 Stacked Denoising Autoencoder

The stacked denoising autoencoder is a kind of deep learning model, which consists of multiple denoising autoencor (DAE). Its architecture is illustrated in Fig. 1. At first, an autoencoder(AE) includes input layer, hidden layer and output layer. The input data will be transformed into output by encoding and decoding. The goal of AE is to minimize the error between input and output. But it's easy to cause the over-fitting problem. The DAE proposed to corrupt raw input by adding a denoising factor before encoding to solve the problem. Generally speaking, it is not necessarily good that reconstructing the original data from features of raw data encoding. But the ability of reconstructing the raw data from corrupted data would be more useful.

The whole training progress of SDAE contains two procedures: pretraining and fine-tuning. In pretraining progress, all DAEs in SDAE model will be trained respectively, which utilizes encoder and decoder to compress and reconstruct raw input. And the output of a DAE will become the input of next DAE. The encoder is able to map the input data to hidden layer by applying a sigmoid function, which translates input into useful features. The decoder is able to reconstruct the raw input by applying a sigmoid function too. Then parameters are updated by minimizing the loss between raw input and output. Moreover, DAE have an unique advantage that reduces the impact of noise. To avoid over-fitting or other problems, a denoising factor a is usually used to corrupt raw input. In this procedure, the noisy data are also possible to be corrupted, which reduce the noise of data.

Let x_i represent the input data which have already been preprocessed. In this paper, x_i is a 1×512 vector, the whole dataset is represents as follows:

$$S_n = \{x_0, x_1, \ldots, x_n\}. \tag{5}$$

The encoder and decoder progresses are described as follows:

$$y = f(W_1 x_i' + b_1). \tag{6}$$

$$z = g(W_2 y + b_2). \tag{7}$$

At first, each x_i is corrupted by a denoising factor to obtain x_i'. Then in encoder progress, the x_i' is mapped into hidden layer by applying a sigmoid function f to acquire data representation y. And in encoder progress, the y is mapped into output layer by applying a sigmoid function g too. Among Eqs. (2) and (3), W_1 and W_2 are the weight matrix and b_1 and b_2 are bias.

o compute the error between output z and raw input x_i, the loss function is described as follows:

$$L(x_i, z) = -\sum_{i=0}^{n} (x_i log(z) + (1 - x_i) log(1 - z)), \tag{8}$$

where the loss function we utilize is cross entropy. The error is computed by comparing original data x_i and output data z of decoder progress. Then parameters of each layer are updated by applying gradient descent. And the pretraining process is done.

In fine-tuning progress, the whole network is trained as a multilayer perceptron (MLP) in a supervised manner. It is noted that the highlight is that the parameters updated in pretraining progress are shared with fine-tuning progress. Backpropagation and gradient decent are applied to minimize loss and update parameters. Moreover, at the top of the network, a softmax layer is added to identify activity types of ET patients. The every node of the softmax layer represents a class and the sum of output values of each node is 1, which is written as:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_c}}. \tag{9}$$

where z_i presents the output of ith instance. C represents the number of activity categories. The output value of node is the probability of belonging to a class current node represents.

4 Experiments and Results

In this section, we will describe our experimental design and analysis of the results, which include the protocol of collecting data, performance of activity recognition model, comparative experiments and so on.

4.1 Experimental Design

To acquire the activity information of ET patients, twenty patients (11 men and 9 women) with essential tremor were chosen to perform six activities, and their age ranges from 38 to 81. It is noted that 75% of them had a family history of ET and they fight the tremor for $10.35 + 7.09$ years [41, 42]. Meanwhile, in these patients, eleven were under tremor therapy, two with primidone, six with propranolol, one both drugs, and two with clonazepam. Furthermore, to show the difference between patients with healthy adults, five healthy subjects were included. What's more, these activities all subjects performed contained extending arms, touching nose, writing sentences, drawing spirals, pouring water and simulating drinking. Their concrete descriptions are given in Table 1.

The experimental data are ET patients' acceleration data collected by using a smartwatch which contains a tri-axis accelerometer and Bluetooth 4.0. Thus, the activity information we collected includes three-axis arm movement acceleration data. Meanwhile, we also considered the influence of angle between smartwatch with ground. All acceleration data were collected with a sampling rate of 25 HZ which was conducted by accelerometer and 228,548 reliable instances were supported as input data.

4.2 Performance Evaluation

After data preprocessing and segmentation, the dataset is imported into the model for activity recognition. Meanwhile, cross-validation is used to evaluate the performance of results. In this step, we will discuss the result in terms of different perspectives.

4.2.1 Experimental Result

In experiment, the estimation metrics this paper adopted are accuracy, precision, recall and F-measure. The accuracy is the proportion between the number of right labels model predicted and the number of true labels. The Precision of an activity is the proportion between the number of right label model predicted and the number of this activity's label model predicted. The recall of an activity is the proportion between the number of right label model predicted and the number of this activity's realistic label. The F-measure is the combination of the precision and the recall. They can be defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

In our study, there are six standard activities from ET patients, which contains extending arms, touching nose, writing sentences, drawing spirals, pouring water and

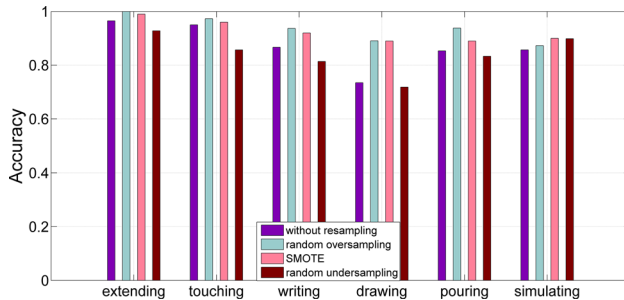**Table 1** The execution protocol of six activities when collecting data

Activities	Description
Extending arms	Keep both arms horizontally extended for 20 s and then relaxed on the lap for 10 s. This task is repeated three times
Touching nose	Repeat the finger-nose test (touching the nose with the index finger with the arm raised to 90°) 10 times, and then keep the hands relaxed on the lap for 10 s. This task is repeated twice
Writing sentences	Write "THIS IS A SAMPLE OF MY WRITING", and then keep both hands relaxed on the lap for 10 s. This task is repeated three times
Drawing spirals	Draw a large and a small spirals, and three lines sequentially with a 10-s rest between every two drawings. This task is repeated twice
Pouring water	Pour a full glass of water into another empty cup and then pour back for three times, and then rest for 10 s. This task is repeated three times
Simulating drinking	Simulate the action of drinking a full glass of water five times and then resting for 10 s. This task is repeated twice

Table 2 Performance of each activity class recognition

	Acc (%)	Pre (%)	Rec (%)	F1 (%)
Extending arms	**99.11**	96.54	99.11	97.81
Touching nose	85.57	95.03	85.57	90.05
Writing sentences	82.28	86.67	82.28	84.42
Drawing spirals	**80.20**	73.49	80.20	76.70
Pouring water	85.89	85.36	85.89	85.77
Simulating drinking	87.80	85.71	87.80	86.74

Bold values indicate the performance gaps between two activities

Acc accuracy, *Pre* precision, *Rec* recall)

Table 3 Performance of each activity class recognition after applying random oversampling

	Acc (%)	Pre (%)	Rec (%)	F1 (%)
Extending arms	99.03	97.77	99.03	98.40
Touching nose	95.32	95.32	95.32	95.32
Writing sentences	91.04	93.27	91.04	92.15
Drawing spirals	90.52	91.42	90.52	90.97
Pouring water	91.23	89.78	91.23	90.50
Simulating drinking	93.15	92.57	93.15	92.86

Fig. 4 Accuracy of each activity class by applying resampling and without resampling

Fig. 5 The confusion matrix of activity recognition for ET patients

simulating drinking. To analyze the performance of human with essential tremor activity recognition intuitively, the recognition accuracy of each activity is shown in Table 2.
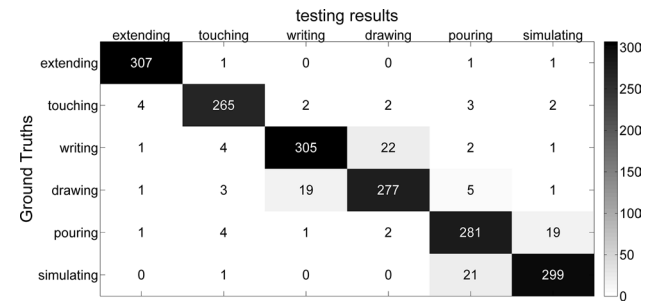
According to Table 2, it shows that there are obvious gaps in different activities. For example, the extending arms activity recognition achieved a highest accuracy of 99.11%, but the drawing activity recognition just reached 80.20%. We assume that there is a problem of unbalanced samples, which drops the accuracy of recognition. According to statistics result, the number of each class is 37,780, 32,551, 38,241, 30,352, 49,519 and 40,104. The maximum gap is 19,167 which is almost one-tenth of the total. Thus, in order to improve this problem, we try to apply resampling techniques to increase data quantity.

4.2.2 Performance Enhancement with Resampling

The problem of unbalanced samples resulted in the big gaps between accuracy of different activities recognition. Thus, we applied the random oversampling, synthetic minority oversampling technique and random undersampling technique to enhance performance in our experiment. After oversampling, the number of each class is 49,519. After undersampling, the number of each class is 30,352. The Fig. 4 shows the effect of performance enhancement with resampling.

According to Fig. 4, the performance of recognition has received significant enhancement by applying random oversampling. However, it is noted that the accuracy of pouring activity recognition drops.

After random oversampling, the performance of activity recognition experiment has achieved 93.33% as measured by accuracy, 93.32% as measured by precision, 93.33% as measured by recall, 93.32% as measured by F-measure. To understand the experimental result clearly, the concrete performance of each activity recognition is shown in Table 3.

From Table 3, it is obvious that recognition accuracy of every activity has reached high level. Moreover, the extending activity acquire highest accuracy of 99.03%, and the pouring activity reach a lowest accuracy of 90.52%. Thus, utilizing random oversampling is able to alleviate the imbalanced sample problem and improve recognition accuracy.

As is shown in Fig. 5, the confusion matrix exhibits the recognition results intuitively. The most obvious errors are that 22 instances of writing sentences are incorrectly recognized as drawing spirals, 21 instances of simulating drinking are classified as pouring water. 19 instances of drawing spirals are misrecognized as writing sentences and 19 instances of pouring water are classified as simulating drinking. Therefore, it can be concluded that similar activities are easier to be confused and more difficult to recognize.

4.2.3 Performance Analysis of Reducing Noise

In this paper, we have studied the ET people's behavior data for activity recognition and tremor degree evaluation. It's worth noting that subjects' activity information includes much unconscious tremor which generates unnecessary noise in behavior data. To alleviate the negative impact of noise, three methods were applied, which were data smoothing, filtering and denoising function of SDAE model. Data smoothing is able to decrease signal jitters by moving average filter. Filter can remove high-frequency noise or abnormal noise. SDAE model uses denoising factors to corrupt raw data, which gives a chance to eliminate noise implement function of denoising. Table 4 shows different performance when applying the three methods.

According to the experimental result, we can see the performance and effectiveness of the three methods. When denoising methods are all not used, the accuracy of activity recognition has only achieved 84.36%. It can be seen that noisy data have a negative impact on performance of recognition. Three denoising methods were applied to reduce noise respectively. After data smoothing, the recognition accuracy has improved by 5% and applying filters is able to improve by more than 6% by removing high-frequency noise or abnormal noise. In experiments, denoising factors were also used to eliminate noise and avoid overfitting. Though the accuracy improvement is not high after setting denoising factor, it is still indispensable to improve recognition performance and avoid overfitting. Thus, three denoising methods in this study are useful and played an important role in experiments.

4.2.4 Hyperparameter Analysis

For deep neural networks, different hyperparameters would result in different performance of classifier. Thus, to acquire the best accuracy of recognition, we discussed different parameters which included pretraining learning rate, the number of hidden layer and the number of hidden layer's neuron and gave the optimal choices of hyperparameters for model.

Table 4 Performance change when applying different denoising method

	Acc (%)	Pre (%)	Rec (%)	F1 (%)
no-denoising	84.36	84.53	84.36	84.43
DS	89.29	89.31	89.29	89.24
Filter	90.53	90.50	90.53	90.51
DF	85.13	85.28	85.13	85.18
DS+Filter +DF	93.33	93.32	93.33	93.32

DS data smoothing, *DF* denoising factor

(a) Pretraining learning rate

The influence of pretraining learning rate on accuracy is shown in Fig. 6. With the changes in the pretraining learning rate, the performance has significant fluctuations. We test multiple values which are set to 1×10^{-4}, 1×10^{-5}, 1×10^{-6}, 1×10^{-7} respectively. We choose the accuracy and precision to observe the change of performance.

According to Figs. 6 and 7, the best performance is reached at 1×10^{-6} level by comparing with other parameters. Each activity has a significant accuracy. It is noted that the performance at 1×10^{-7} level has just small differences with the 1×10^{-6} level. Meanwhile, considering the accuracy and precision of activity recognition, the learning rate is set to 1×10^{-6} though the pretraining learning rate is set very low. Because the goal of pretraining is to acquire proper initial values of parameters and ensure that the fine-tuning process has better astringency and performance, the
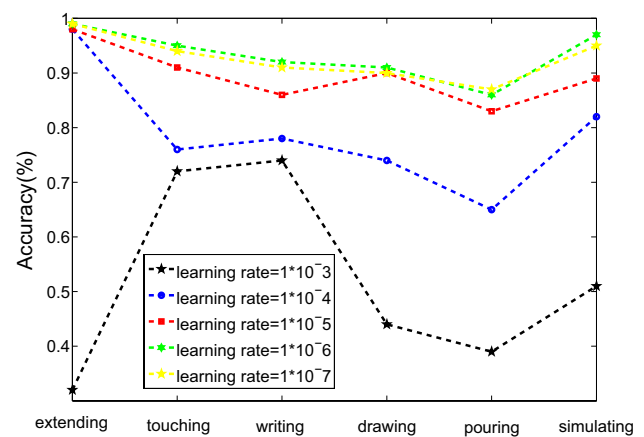

Fig. 6 The recognition accuracy of each activity when adopting different pretraining learning rates
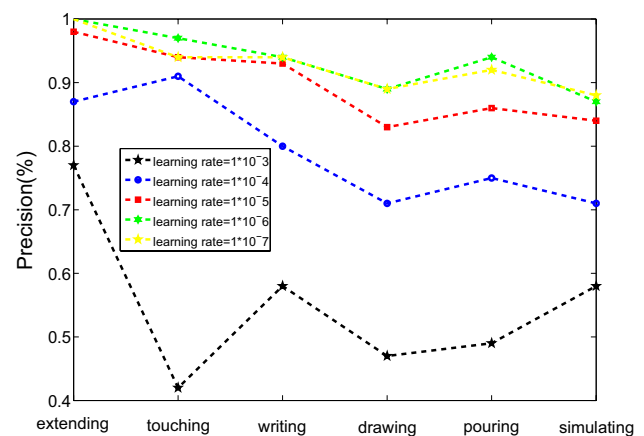

Fig. 7 The recognition precision of each activity when adopting different learning rates

pretraining learning late does not need to be too high. With the increasing of learning rate, the experiment results have acquired more and more unsatisfactory performance. Just as at 1×10^{-3} level, the result has achieved a worst accuracy. In addition, it is obvious that the pouring activity recognition has achieved the lowest accuracy. A reasonable assumption is that the pouring activity has similar features to other activities, which is be misrecognized easily.

(b) The number of hidden layer

In Fig. 8, it shows that the different performance of ET patients' activity recognition when selecting different numbers of the hidden layer. The main gap of accuracy also lies in the pouring activity. And with the changes in the number of the hidden layer, the recognition accuracy has shown obvious fluctuation. Thus, in order to achieve the best performance and acquire a trade-off between accuracy with time-cost, we utilize 2 hidden layers to construct network.

(c) The number of hidden layer's neuron

Besides learning rate and the number of the hidden layer, we also take into account the number of hidden layer's neuron. With the increasing of neurons, the dimension of variables in hidden layers is growing. Then features extracted from data would contain more useful information and the time-cost would increase too. The Fig. 9 shows that the concrete performance when choosing different numbers of neuron.

According to Fig. 9, there are obvious fluctuation about the performance of recognition when selecting different number of neuron. Through compare, the best performance is achieved when the number of hidden layer's neuron is set to 1000. It is noted that the recognition performance of pouring activity outperforms other choice when the layers' neuron is set to 2500, but the performance of writing activity has achieved the worst accuracy. Moreover, the run time is nearly 10 times of other choice when choosing the 2500
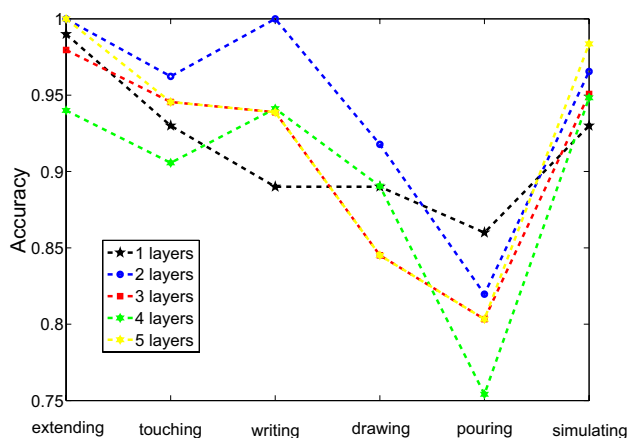

Fig. 8 The performance of ET patients' activity recognition when selecting different numbers of hidden layer
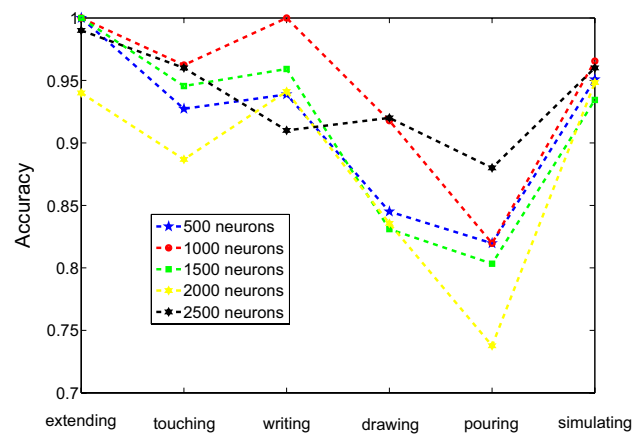

Fig. 9 The performance of ET patients' activity recognition when selecting different numbers of hidden layer's neuron

neurons. Thus, we take into account accuracy, run time and computational complexity, the hidden layer's neuron is set to 1000 to create the network.

According to above a series of experiments, Table 5 has given the optimal hyperparameters of the neural network. Because some hyperparameters have little influence on recognition performance in experiments. Thus we have discussed above three hyperparameters.

4.2.5 Comparison with Other Methods

To demonstrate the superiority of SDAE model on ET patients' activity recognition, we compared it with some conventional machine learning algorithms, such as support vector machine (SVM), decision tree (DT), K-nearest neighbors (KNN), other deep neural network, such as CNN, and some effective algorithms propose by previous researchers [43, 44]. These experiments all adapted unified data processing procedure. By comparing with other methods, the results are shown in Table 6.

From Table 6, it is obvious that the SDAE model outperforms other conventional machine learning methods (SVM, DT, KNN). Moreover, SDAE model, as a deep learning

Table 5 Optimal hyperparameters for SDAE model

Hyperparameter	Value
Number of hidden layer	2
Number of units per layer	1000
Pretraining learning rate	1×10^{-6}
Fine-tuning learning rate	0.01
Iteration	200
Denoising factor	0.4
Data segment size (in s)	5.12

Table 6 Performance comparison between SDAE and other methods

Methods	Acc (%)	Pre (%)	Rec (%)	F1 (%)
SVM	82.00	83.15	82.00	82.13
DT	82.29	83.39	82.29	82.48
KNN	79.71	81.34	79.71	79.85
CNN	81.33	79.85	81.33	80.27
LSTM	76.68	77.09	76.68	76.70
Zhao et al. [43]	83.68	84.90	83.68	83.23
Panwar et al. [44]	86.57	86.92	86.64	86.62
SDAE	**93.33**	**93.32**	**93.33**	**93.32**

Bold values indicate that method this paper proposed achieves the best performance in comparative experiments

Table 7 Performance of ET rating evaluation based on SDAE model

Level	Acc (%)	Pre (%)	Rec (%)	F1 (%)
0	80.23	85.84	80.23	82.94
1	82.73	82.14	82.73	82.43
2	94.21	90.62	94.21	92.38
3	100.00	98.18	100.00	99.08

There are four levels: 0, 1, 2 and 3. The 0 level represents health subjects

model, is able to extract features automatically compared with machine learning methods. Though CNN and LSTM is also able to extract features automatically, its recognition accuracy is weaker than SDAE. Generally speaking, deep learning models can acquire better performance than conventional machine learning methods. But in this experiment, CNN and LSTM model achieve lower accuracy. We assume that the scale of sample is still so small that they cannot learning enough knowledge though applying resampling techniques. A deep network architecture using residual bidirectional long short-term memory is proposed to enhance the recognition rate [43]. A novel methodology based on deep learning framework is used to extract useful features automatically and reduce the computational cost [44]. They all have achieved lower performance than our method. Therefore, in this situation, utilizing SDAE model can achieve more robust performance.

4.2.6 Essential Tremor Evaluation

In data gathering process, besides human activity labels, two neurologists were invited to estimate ET severity of each subject. They rated ET by Fahn–Tolosa Marin Tremor Rating Scale (FTMTRS). The degree of ET contained 0, 1, 2 and 3. Therefore, we have also utilized these hierarchical tags of all samples to evaluate essential tremor extent based on the SDAE model. According to the experimental verification, the performance of ET rating evaluation has achieved an accuracy of 95.74% and performances of each level are shown in Table 7. And the confusion matrix is shown in Fig. 10.

From Table 7 and Fig. 10, it is obvious that the 2 and 3 levels have achieved significant recognition performance, but the recognition performance of 0 and 1 levels are weak. It can be concluded that the tremor of patients also can be regarded as data feature which play a significant role in activity recognition when the degree of tremor is large enough. Therefore, the degree of 0 and 1 levels has been

mistaken easily. According to Fig. 10, 24 instances of 0 level are misrecognized as 1 level and 62 instances of 1 level are misrecognized as 0 level. Thus, The higher the degree of tremor, the higher is the performance evaluation. And the patients with essential tremor of 1 level have not obvious symptoms.

5 Conclusion

In this paper, we discussed the combination of essential tremor and human daily activity recognition, which has a bright development prospect on providing an auxiliary analysis for evaluating the degree of ET. Utilizing HAR system to recognize ET patients' activities based-on stacked denoising autoencoder. The SDAE model is able to not only extract features automatically but also reduce the impact of noise of behavior information from ET patients. What's more, the resampling techniques have been used in order to alleviate the imbalanced sample and small-sample problems. The result we obtain demonstrate that the SDAE model is able to reach significant performance on human with essential tremor activity recognition and outperform other conventional machine learning methods. Moreover, we also perform ET degree evaluation based SDAE model, which rates ET according to four levels.

In the future, we consider studying a novel fusion model to improve performance of activity recognition and we also
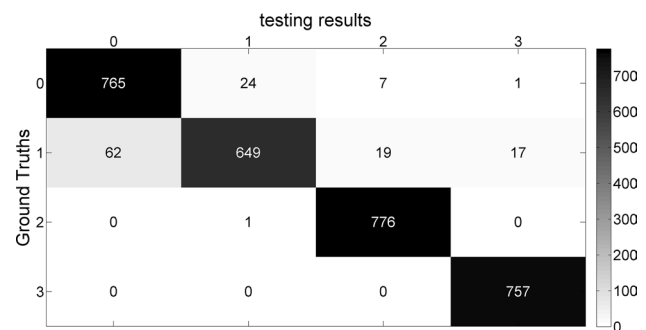
Fig. 10 The confusion matrix of ET degree evaluation

consider utilizing more kinds of sensors to collect data in a future experiment. Furthermore, we will focus on the application of human activity recognition in video.

Author Contributions Methodology, QN, ZF; software, ZF; validation, QN, LZ; formal analysis, BZ; investigation, LZ; resources, QN, XZ; data curation, XZ; writing—original draft preparation, QN, ZF; writing–review and editing, ZF, YZ; visualization, ZF; supervision, BZ, YZ; project administration, Q.N.; funding acquisition, BZ, QN.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. Thanvi, B., et al.: Essential tremor—the most common movement disorder in older people. Age Ageing **35**(4), 344–349 (2006)
2. Poon, C., et al.: Combined measures of movement and forcé variability distinguish Parnkinson's disease from essential tremor. Clin. Neurophysiol. **122**(11), 2268–2275 (2011)
3. Alhami, M.F., et al.: Hamon: an activity recognition framework for health monitoring support at home. In: 2011 IEEE International Instrumentation and Measurement Technology Conference (2011)
4. Chelli, A., Pätzold, M.: A machine learning approach for fall detection and daily living activity recognition. IEEE Access **7**, 38670–38687 (2019)
5. Wang, A., et al.: A comparative study on human activity recognition using inertial sensors in a smartphone. IEEE Sens. J. **16**(11), 4566–4578 (2016)
6. Yan, Y., et al.: Egocentric daily activity recognition via multitask clustering. IEEE Trans. Image Process. **24**(10), 2984–2995 (2015)
7. Cheng, J., et al.: A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals. IEEE J. Biomed. Health Inform. **17**(1), 38–45 (2013)
8. Yang, X., et al.: Detection of essential tremor at the S-band. IEEE J. Transl. Eng. Health Med. **6**, 2168–2372 (2018)
9. Gaglio, S., et al.: Human activity recognition process using 3-D posture data. IEEE Trans. Hum. Mach. Syst. **45**(5), 586–597 (2015)
10. Wannenburg, J., Malekian, R.: Physical activity recognition from smartphone accelerometer data for user context awareness sensing. IEEE Trans. Syst. Man Cybern. Syst. **47**(12), 3142–3149 (2017)
11. Jain, A., Kanhangad, V.: Human activity classification in smartphones using accelerometer and gyroscope sensors. IEEE Sens. J. **18**(3), 1169–1177 (2018)
12. Jain, S., et al.: Common misdiagnosis of a common neurological disorder. JAMA Neurol. **63**(8), 1100–1104 (2006)
13. Zhang, L., et al.: Classification of Parkinson's disease and essential tremor based on structural MRI. In: 2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA) (2017)
14. Shukla, P., et al.: Towards closed-loop deep brain stimulation: decision tree-based Essential Tremor patient's state classifier and tremor reappearance predictor. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2014)
15. Elble, R., et al.: Reliability of a new scale for essential tremor. Mov. Disord. **27**(12), 1567–1569 (2017)
16. Blomstedt, P., et al.: Deep brain stimulation in the posterior subthalamic area in the treatment of essential tremor. Mov. Disord. **25**(10), 1350–1356 (2010)
17. Pg, B., et al.: Assessing tremor severoty. J. Neurol. Neurosurg. Psychiatry **56**, 868–873 (1993)
18. Thanawattano, C., et al.: Temporal fluctuation analysis of tremor signal in Parkinson's disease and Essential tremor subjects. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2015)
19. Matsumoto, Y., et al.: Analysis of EMG signals of patients with essential tremor focusing on the change of tremor frequency. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2012)
20. Hsu, Y., et al.: Human daily and sport activity recognition using a wearable inertial sensor network. IEEE Access **6**, 31715–31728 (2018)
21. Nweke, H.F., et al.: Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. Expert Syst. Appl. **105**, 233–261 (2018)
22. Gupta, P., Dallas, T.: Feature selection and activity recognition system using a single triaxial accelerometer. IEEE Trans. Biomed. Eng. **61**(6), 1780–1786 (2014)
23. Pärkkä, J., et al.: Personalization algorithm for real-time activity recognition using PDA, wireless motion bands, and binary decision tree. IEEE Trans. Inf. Technol. Biomed. **14**(5), 1211–1215 (2010)
24. Chen, Z., et al.: Robust human activity recognition using smartphone sensors via CT-PCA and online SVM. IEEE Trans. Ind. Inf. **13**(6), 3070–3080 (2017)
25. Feng, Z., et al.: A random forest-based ensemble method for activity recognition. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2015)
26. Elshourbagy, M., et al.: Enhanced bag of words using multilevel k-means for human activity recognition. Egypt. Inform. J. **17**(2), 227–237 (2016)
27. He, Z., Jin, L.: Activity recognition from acceleration data based on discrete consine transform and SVM. In: Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics, San Antonio, 11–14 October (2009)
28. McCarthy, M.W., James, D.A., Lee, J.B., Rowlands, D.D.: Decision-tree-based human activity classification algorithm using single-channel foot-mounted gyroscope. Electron. Lett. **51**, 675–676 (2015)

29. Wang, J., et al.: Device-free wireless localization and activity recognition: a deep learning approach. IEEE Trans. Veh. Technol. **66**(7), 6258–6267 (2017)

30. Xi, R., et al.: Deep dilation on multimodality time series for human activity recognition. IEEE Access **6**, 53381–53396 (2018)

31. Wang, L.: Recognition of human activities using continuous autoencoders with wearable sensors. Sensors **16**(2), 189 (2016)

32. Inoue, M., et al.: Deep recurrent neural network for mobile human activity recognition with high throughput. Artif. Life Robot. **23**(2), 173–185 (2018)

33. Yao, S., Hu, S., Zhao, Y., Zhang, A., Abdelzaher, T.: DeepSense: a unified deep learning framework for time-series mobile sensing data processing. In: Proceedings of the 26th International Conference on World Wide Web (2017)

34. Milenkoski, M., et al.: Real time human activity recognition on smartphones using LSTM networks. In: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (2018)

35. Yu, S., Qin, L.: Human activity recognition with smartphone inertial sensors using Bidir-LSTM networks. In: Proceedings of the 2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE) (2018)

36. Yalçın, H.: Human activity recognition using deep belief networks. In: 2016 24th Signal Processing and Communication Application Conference (SIU) (2016)

37. Zhang, L., Wu, X., Luo, D.: Real-time activity recognition on smartphones using deep neural networks. In: Proceedings of the 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom) (2015)

38. Chawla, N.V., et al.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**(1), 321–357 (2002)

39. Galar, M., et al.: A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans. Syst. Man Cybern. Part C (Applications and Reviews) **42**(4), 463–484 (2012)

40. Abdi, L., Hashemi, S.: To Combat multi-class imbalanced problems by means of over-sampling techniques. IEEE Trans. Knowl. Data Eng. **28**(1), 238–251 (2016)

41. Zheng, X., Vieira, A., Marcos, S.L., Aladro, Y., Ordieres-Meré, J.: Activity-aware essential tremor evaluation using deep learning method based on acceleration data. Parkinsonism Relat. Disord. **58**, 17–22 (2019)

42. Zheng, X., Vieira Campos, A., Ordieres-Meré, J., Balseiro, J., Labrador Marcos, S., Aladro, Y.: Continuous monitoring of essential tremor using a portable system based on smartwatch. Front. Neurol. **8**, 96 (2017). https://doi.org/10.3389/fneur.2017.00096

43. Zhao, Y., Yang, R., Chevalier, G., Xu, X., Zhang, Z.: Deep residual bidir-LSTM for human activity recognition using wearable sensors. Math. Probl. Eng. **2018**(ArticleID 7316954), 13 (2018). https://doi.org/10.1155/2018/7316954

44. Panwar, M., Dyuthi, S.R., et al.: CNN based approach for activity recognition using a wrist-worn accelerometer. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2017)