Multimedia Tools and Applications (2022) 81:33851–33866
https://doi.org/10.1007/s11042-022-12659-9

Two-channel lstm for severity rating of parkinson's disease using 3d trajectory of hand motion

Aite Zhao[1] · **Jianbo Li**[2]

Received: 17 January 2021 / Revised: 14 April 2021 / Accepted: 9 February 2022 /
Published online: 21 April 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Hand movement is one of the important bases for the severity rating of Parkinson's disease. While observing hand motion of patients, medical specialists evaluate the degree of motor deterioration according to established rating scales. This diagnostic procedure is inefficient and can be easily affected by different doctors' subjectivity, even though several studies showed rating scales are reliable. In this paper, we propose an automatic method based on hand exercise data including finger-tapping and fist movements, which is recorded by ordinary camera. We estimate 3D hand pose from regular RGB images and proposed a two-channel long short-term memory model to learn the patterns of 3D position changing trajectory of hand joints. Experiments on our dataset, the proposed method outperforms literature including popular machine learning methods with 95.7% of the precision, 95.8% of the sensitivity and 92.8% of the specificity respectively on average. We believe the quantitative evaluation of hand movement will benefit the clinical PD diagnosis.

Keywords Parkinson's disease · Severity rating · 3D hand pose · Hand movements

1 Introduction

The severity assessment of Parkinson's disease (PD) could be improved significantly if objective hand motion information can be obtained in usual surroundings such as the medical institution and the private home. Generally, while observing patients' motions such as walking, upper-limb motion, fist and finger-tapping movement (repeatedly opening and closing the thumb and index finger), medical specialists diagnose the severity of patients with PD. The entire process requires a doctor to accompany and spend a lot of time, which is inefficient and subjective may lead to error diagnosis. Although there has been a unified

✉ Jianbo Li
lijianbo@qdu.edu.cn

Aite Zhao
zhaoaite@qdu.edu.cn

[1] School of Business, Qingdao University, Qingdao, China

[2] College of Computer Science and Technology, Qingdao University, Qingdao,China

evaluation criteria for severity rating of PD, the hand movement vary from person to person and can not be quantified, which is hard to visually and accurately report the patient's disease to their families increasing the difficulty of evaluation. Therefore, the development of computer-assisted diagnosis and computer-expert system is very important. Modern computer-assisted diagnosis, such as automatic medical image processing and large scale medical data analysis, has been widely used in corresponding medical fields, where feature exacting plays a key role in these systems.

For evaluating severity of PD from hand data, doctors pay attention to the patient's hand, especially the trajectory of the hand movement, and determine disease state of the patient according to the subtle change of the hand and the speed of movement, thus it is important to estimate the hand joint position of the patient. We utilize 3D hand pose estimation method to extract the 3D coordinate information of the hand joints from the hand motion video consists of multiple single color images, which records the moving trail of 21 hand joint points. With these ordered 3D coordinates, machine learning methods can model the hand moving process and quantify temporal hand data.

Machine learning methods used in literature, such as naive Bayesian (NB), random forest (RF) and support vector machine (SVM), have been employed and achieved promising results in gesture recognition [4, 7, 14, 30]. Other studies have established complex deep learning networks for hand gesture recognition, which achieved considerable classification results [22, 23]. However, these approaches are not specifically designed to deal with temporal sequential data, whereas the hand data recorded by devices (cameras, depth sensors, wearable devices) contains important temporal information that is critical for severity rating of PD.

As one of the most popular deep learning models [18, 20, 21, 31, 34], Long Short-Term Memory (LSTM) has recently reached various fields with a capability of handling time series with long intervals and multi-classification problems, which is considered suitable for modeling hand information.

Our contribution to the field is a proposal of a new combinatorial technique based on computer vision. We measure hand skeleton parameters and characterize hand movement by means of data captured by RGB cameras (without depth information). We fuse the 3D hand pose estimation method with the LSTM algorithm for modeling hand motion. Our proposal makes the design of intelligent, low-cost and noninvasive systems possible. In addition, the digital camera is energy efficient thanks to built-in image sensors relying entirely on visible light.

Hand gesture can certainly be visualized by digital cameras but gesture study entails a complexity not readily observable; in fact, it is riddled with difficulties arising from the variability of poses and actions. One way current research has devised to counteract this variability effect is to identify the main features of hand movement through the process of tracking and pose analysis using image processing and machine learning algorithms.

The main contributions of this paper are summarized as follows:

– A two-channel LSTM model is designed for severity rating of Parkinson's disease based on image data. can process the image data of two different hand movements at the same time, and analyze the performance of the two movements to obtain accurate severity score.
– A 3D hand pose estimation method is adopted for hand motion modeling and hand gesture recognition, which can extract joint points of hand bones from 2D images to avoid the influence of surrounding environment such as light and color on hand images.

- The fusion model of two-channel LSTM and 3D hand pose estimation method can be applied successfully to a variety of hand movement data and accomplish two hand movement recognition tasks, which outperforms the other methods in the literature.
- Our proposal makes the design of intelligent, low-cost and noninvasive systems possible.

This paper is organized based on the following sections. Section 2 describes existing research on computer assisted PD severity assessment, Section 3 introduces the proposed model, and Section 4 depicts the used dataset. Section 5 shows the experiment. We discuss the related technology and possible future research in Section 6 and conclude the whole paper in Section 7.

2 Related work

The development of modern computer-assisted medical diagnostic systems is driven by the presence of huge amounts of medical data coupled with the utilization of machine learning methods.

The hand of patients with Parkinson's disease is easily affected by muscle stiffness, and often presents a peculiar flexion posture, which makes it difficult to unfold the palm; while the extension of small joints between other fingers makes it difficult to clench the palm [24, 25, 36]. In view of these pathological manifestations, this project will collect the hand movement data of patients, analyze the influence of early Parkinson's disease on various characteristics of hand and finger movement of patients, and evaluate the severity level [8].

It is well known that Parkinson's disease biomarkers can be analyzed through various forms of human-computer interaction, including precise grip strength, finger tapping test (FTT), hand and finger movement, and writing [6, 26, 33]. Yokoe et al.[33] measured 14 parameters of FTT motion using touch sensor data, showing significant differences between PD patients and non PD patients.

The reaction time of patients with Parkinson's disease is often slower than that of unaffected people of the same age. When complex reactions are needed, it increases the difficulty of execution sequence and hand movement. Through the time series analysis of key holding time and support vector machine (SVM), Giardo et al. found that there were significant differences between PD patients and the control group, but the subsequent research on the classification characteristics of larger groups only obtained 78% accuracy [9].

However, these studies did not evaluate other hand movement symptoms, such as writing images, finger opening and closing images, and tremor force sensitive data. In this project, various hand movement data of patients with Parkinson's disease will be considered to comprehensively identify and evaluate their disease status.

In order to evaluate hand movement, the researchers collected different types of hand movement data as the basis for the severity evaluation of PD according to evaluation criteria. Arora et al. organized a finger tapping test to tap the screen alternately keeping a regular rhythm using the speed and frequency features [3]. Sano et al. utilized magnetic sensors to quantify the severity of symptoms related to the finger tapping of PD imitators with high accuracy, which calculated features from the obtained waveforms [28]. Alty et al. assessed whether Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) graded one PD bradykinesia can be accurately detected by analyzing a standard finger tapping clinical assessment employing electromagnetic tracking sensors [1]. Another low-cost sensor accelerometer was adopted to quantify finger movement, which aimed to predict score of Unified Parkinson's Disease Rating Scale (UPDRS) and evaluate motor disability in PD patients [29].

Thus, hand motion is quantified into different dimensions of data according to multifarious sensors, for which a variety of machine learning methods are adopted to model its inherent information, and finally evaluated the state of the diseased object. Keisuke et al. measured and evaluated finger tapping movements for the assessment of motor function using log-linearized Gaussian mixture networks (LLGMNs), which verified that the impairment status of patients could be classified correctly with a high accuracy rate 93.1% employing 12 LLGMNs [12]. Ariyanto et al. classified finger movement pattern using designed artificial neural networks based on EMG signals [2]. Kupryjanow et al. came up with an alternative measurement technique for determining UPDRS sub-scores related to motor tests that a SVM recognized hand gestures [17]. Printy et al. trained SVM classifier to grade motor impairment severity in PD based on UPDRS utilizing the finger tapping test data captured by smartphone [27].

Compared to wearable equipment, the camera has low-cost and non-invasive favorable factors, the captured image is also very intuitive to reflect the hand movement. The most related work with us has been applying machine learning methods in processing hand moving videos. Khan et al. proposed a computer-vision method for assessment of tapping symptoms with severity levels ("0: normal" to "3: severe") through motion analysis of index-fingers, which tracked the index-finger motion between the video-frames to produce a tapping time-series [13]. SVM classifier was employed to classify the tapping time-series with a 88% accuracy. A 3-dimensional motion camera system was designed for measuring finger tapping in PD, which can provide hand information marker in real time [16].

The video data collected by the camera, composed of multiple ordered images including temporal and spatial information. With these temporal data, LSTM is suitable for changing series to obtain movement information for patients with PD, which can improve the detection sensitivity and analyze the complexity and diversity of the motor behavior objectively. Inspired by the existing advanced methods, this paper reports a model to learn the pattern of hand movement, and to our best knowledge, it is the first to propose the utilization of hand videos to diagnose the severity of PD. The experimental results show that the proposed method is superior to the aforementioned machine learning approaches.

3 The proposed method

The hand movement evaluation model of PD includes the fusion of two algorithms, 3D hand skeleton joint points extraction and hand motion learning and classification algorithm.

3.1 3D hand pose estimation

We utilize a new approach to learn full 3D hand pose estimation from single color images in videos without the need for any special equipment and capitalize on the capability of deep networks to learn sensible priors from data in order to resolve ambiguities [37]. By experimenting on a variety of test sets, the feasibility of 3D hand pose estimation on single color images has been demonstrated.

The 3D hand pose estimation method consists of three deep networks that cover important sub-tasks on the way to the 3D pose; The first network provides a hand segmentation to localize the hand in the image. Based on its output, the second network localizes hand key

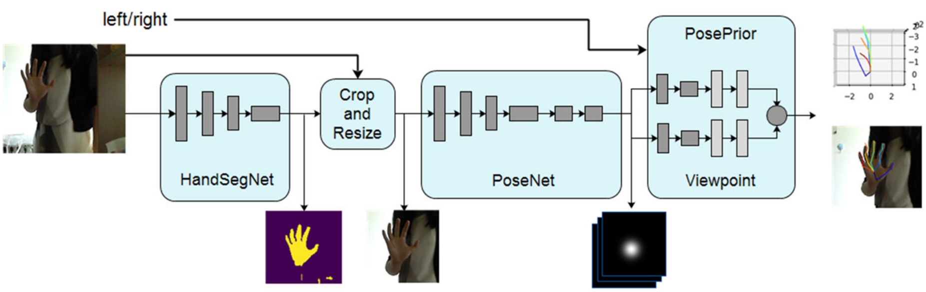

Fig. 1 The 3D hand pose estimation model

points in the 2D images. The third network finally derives the 3D hand pose from the 2D key points. The specific details of the structure was shown in Fig. 1.

In the hand segmentation section, the hand localization is considered as a segmentation problem, and the network HandSegNet is trained on a hand pose dataset, which allows us to crop and normalize the inputs in size simplifying the learning task for the PoseNet. For hand segmentation, the two-dimensional human detection problem is regarded as a fractional graph to estimate the center position of human body. The most likely position is used as the center of a fixed size cut image. Because the size of the hand changes rapidly in the whole image, and largely depends on the clarity, hand location is regarded as a segmentation problem. HandSegNet provides a hand model that allows clipping and normalizing the size of the input image, which simplifies the learning task of PoseNet. This network uses convolution layer to express texture information and spatial information. Firstly, the spatial constraints between components are expressed by the response graph of each component. Response graph and feature graph are transmitted together as data in the network. The network is divided into several stages, and each stage has supervision training to avoid the problem that the network is too deep to optimize. At the same time, the input characteristics and responses are processed at multiple scales. It can not only ensure the accuracy, but also consider the long-distance relationship between the various components. Finally, the segmented hand image is obtained.

The purpose of PoseNet is predicting an initial score map from image feature representation produced by an encoder-decoder architecture. We regard the location of 2D key points as the estimation of 2D score map, and train a network to predict the score, in which each

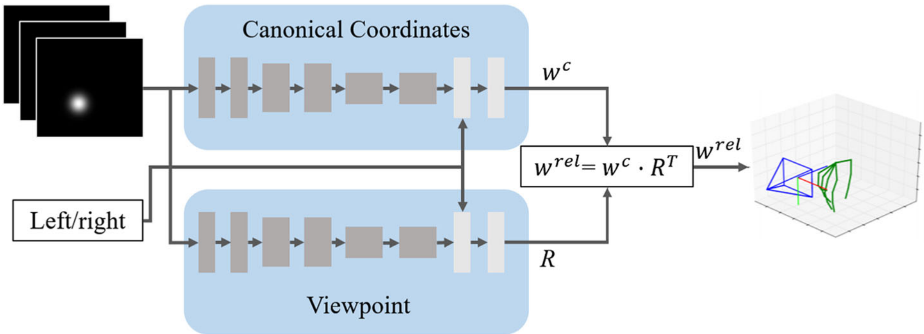

Fig. 2 The PosePrior Network

score map contains information about the possibility of a joint point appearing in a certain spatial position. The PoseNet uses an encoder-decoder architecture. Given the representation of image features generated by the encoder, the initial fractional image is predicted and continuously refined in resolution. We use HandSegNet as initialization weight, and get the model of hand joint detection through training.

The last network, PosePrior network learns to predict relative, normalized 3D coordinates conditioned on potentially incomplete or noisy score maps, which has two parallel processing streams that use 6 convolutional layers followed by two fully-connected layers. After these three networks have been well trained, we can get three-dimensional coordinates of the 21 key joints of hand from each frame of videos. We encapsulate these hand joint coordinates for the next step. The structure of the PosePrior Network is shown in Fig. 2. The network structure for the pose prior has two parallel processing channels, which process the stack of J score maps in 6 convolutions with ReLU function. Information on whether the image shows a left or right hand is concatenated with the feature representation and processed further by two fully-connected layers which yields estimations for viewpoint R and canonical coordinates w^c. Both estimations combined lead to an estimation of w^{rel}. The formula is as follows:

$$w^{rel} = w^c \cdot R^T \tag{1}$$

3.2 The two-channel LSTM model

LSTM (Long Short-Term Memory) is a deep learning model, which has repetitive modules that learn long-term dependency information from the input data [10]. The internal structure of repetitive module in LSTM has four interactive operations (3 $sigmoid$ and 1 $tanh$), which enables it by extracting valid information from dynamic data.

The basic cell of LSTM has three gates (input, forget and output). Every gate has a sigmoid activation function and a pointwise multiplication operation. We choose a variant of LSTM called the Gated Recurrent Unit (GRU) as learning model, which combines the input gate and forgot gate into an update gate and mixes cell state and hidden state. Figure 3

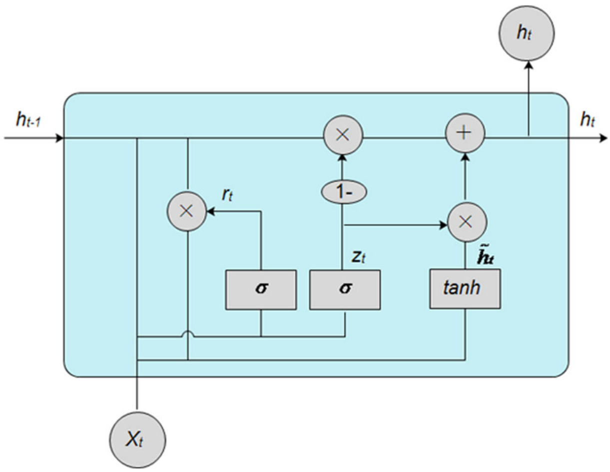

Fig. 3 The GRU model

shows the structure of a GRU cell and illustrates the operations of the gates. The basic cell of the GRU is defined as the following equations:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{2}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{3}$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t]) \tag{4}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{5}$$

where z_t denotes the output of update gate to the network at time step t, where σ is the logistic sigmoid function. Moreover, x_t and h_{t-1} are the input and the previous hidden state, respectively. In addition, W_r and W_z are weight matrices which are learned and r_t denotes the reset gate and the actual activation of the proposed unit h_t is then computed by (5).The update gate z selects whether the hidden state is to be updated with a new hidden state $\tilde{h}$. The reset gate r decides whether the previous hidden state is ignored.

We propose a two-channel LSTM model for severity rating in finger-tapping and fist movement, each channel consists of two layers of LSTM, each of which includes the same number of expanded nodes (called time steps or cells) according to the input data. The LSTM networks of two channels are trained separately to predict classification probability values, which are fed into the $softmax$ classifier for calculating classification accuracy rates. Finally, the two accuracy rates are weighted to arrive at the final classification results, according to each training accuracy, the dual channel LSTM model minimizes the loss to obtain the optimum weights and biases. The proposed two-channel LSTM model is illustrated in Fig. 4.

The two-channel LSTM model inputs hand movement (finger-tapping, fist) data in batches, each channel has same number of LSTM cells based on the number and time variation of the input. By independent training, the output feature of the last cell from the second layer of LSTM (O_N) is chosen to predict the probability of each class. Taking the maximum probability value and fusing accuracy from the two channels, the model will output predicted severity of PD. We only explain the detailed design of the finger-tapping channel, because the structure of the two channels is similar.

In the finger-tapping channel, we employ a two-layer LSTM to obtain deeper feature, each LSTM layer has $N(N = 10)$ cells with the last output O_N. The hand skeleton data will be input into the LSTM after normalized, which makes the data fixed in length according the time steps. Experiments show that 10 frames of video data include sufficient hand movement information of the patient (patients with different severity differ in their motor

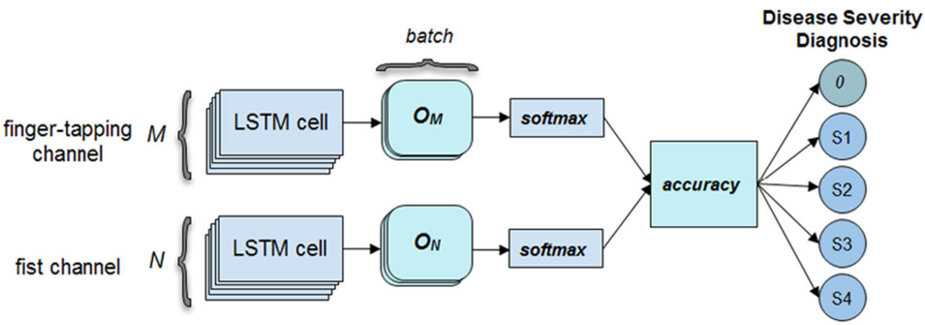

Fig. 4 The two-channel LSTM model

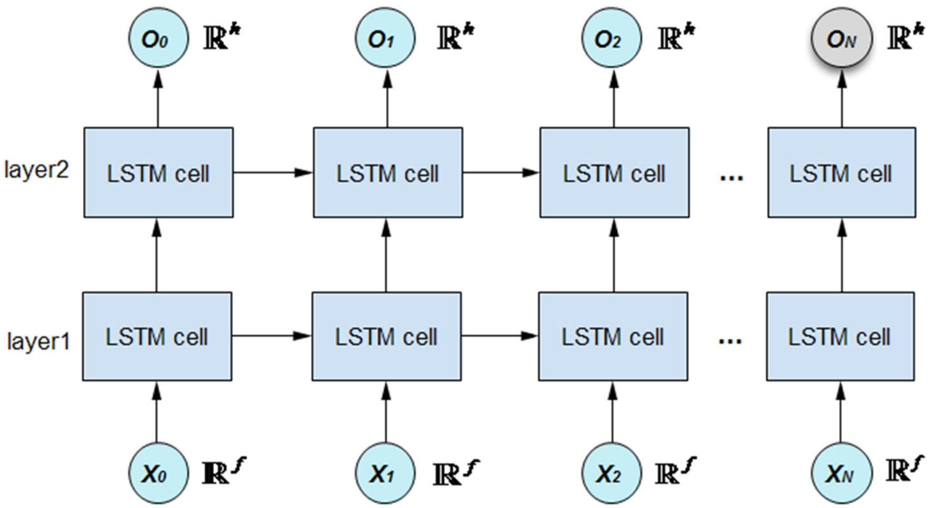

Fig. 5 The structure of the two-layer LSTM. The input of the LSTM is $x_N, x_N \in \mathbb{R}^f$ $(f = 63)$, which from different time periods are entered into N $(N = 10)$ LSTM cells. The output of Nth cell can denote the hand moving sequence

cycles). After repeatedly training, a feature vector is produced to represent the hand motion. Figure 5 depicts the structure of the used two-layer LSTM.

A training sample is a sequence of $Nf - dimensional$ coordinate vectors, K training samples are input into the first LSTM layer at a time, the dimension of each input is $N \times K \times f$. LSTM will be expanded according to time steps and every sample in one training sample can be input to one LSTM cell respectively. Every LSTM cell takes $K \times f$ data as the input, the output of which is an $h - dimensional$ feature vector (hidden layer output) that can be adjusted to an appropriate value. The second layer of LSTM uses the same setting. At time N, the output feature $(O_{10} \in \mathbb{R}^h)$ contains information of previous cells, which is taken as the basis for classification. By utilizing a multi-class classifier $softmax$, a $g - dimensional(g$ classes) vector is generated based on the weights and biases to map the output of the LSTM to a probability distribution and diagnose finger-tapping severity of PD.

3.3 The fusion model

We combine the above methods into an end-to-end model, and evaluate the patient's severity by training. The process of fusion model is shown in Fig. 6. After testing on our dataset, the proposed model is verified to be feasible.

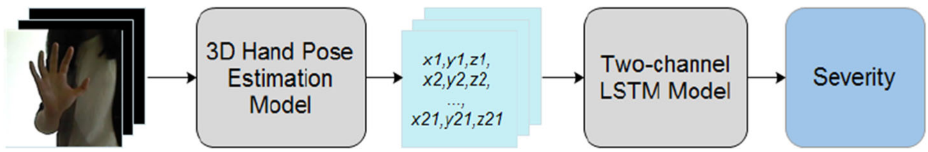

Fig. 6 The Fusion Model

Table 1 The severity illustration in UPDRS

Hand Movements	Severity 0	Severity 1	Severity 2	Severity 3	Severity 4
Finger-tapping	normal	11∼14times/5s	7∼10times/5s	3∼6times/5s	0∼2times/5s
Fist movement	normal	amplitude reduction	moderate disability	serious disability	almost no action

4 Data collection

Our dataset included hand motion videos from 12 subjects, who simulated finger-tapping and fist movements of PD patients at different levels of illness according to UPDRS standard. There were 5 rating levels describing different PD severity, each level contained 12 subjects involved in video data collection, because subjects can simulate different PD hand behaviors to obtain more data. The severity illustration in UPDRS is introduced in Table 1.

The raw data were obtained using RGB camera with 30 fps, subjects were placed 0.5 meters away from the camera performing 500 frames of every action. Every frame was an RGB image with a resolution of 640×480, which were in chronological order and excluded incomplete hand movements.

Opencv and Python libraries were adopted to record hand motion and crop the video into frames so that the hand skeleton can be extracted easily.

The dataset had also explained the specific situation of each participant including gender, age, height, weight, walking speed, and a measure of disease severity or duration. The simulated severity assignment of disease is shown in Table 2.

5 Experiment

In our experiment, We evaluate the performance of our learning method on PD severity diagnosis task on our dataset. The evaluation is done by training and testing all sequences belonging to the dataset repeatedly and calculating its classification accuracy with respect to the true labels. In addition, we tune some key model parameters to obtain optimal results verifying the availability of our method in this field.

The experimental results of the experiment were evaluated against several state-of-the-art methods. The following is a list of all the comparison methods:

Classifiers:

– NB (naive Bayesian) is a series of simple probability classifiers based on Bayesian theorem under the assumption of strong independence between features.
– SVM (support vector machine) is a kind of generalized linear classifier that classifies data by supervised learning.
– DT (decision tree) refers to a decision support tool using a tree-like model of decisions and their possible consequences.

Table 2 The number of people divided by severity

Severity 0(Healthy)	Severity 1	Severity 2	Severity 3	Severity 4
12	12	12	12	12

- GBDT (gradient boosting decision tree) is an iterative decision tree algorithm, which is composed of multiple decision trees. The results of all the trees are accumulated to determine the final label.
- LR (logistic regression) is a kind of nonlinear regression models and a machine learning method for probability estimation.
- RF (random forest) is a classifier with multiple decision trees.
- KNN (k-nearest neighbor) is a classifier that finds k nearest instance and votes to determine the class name of the new instances.

Deep Models:

- GRU (gated recurrent unit) is a gating mechanism in recurrent neural networks, which has fewer parameters than LSTM.
- LSTM (long-short term memory) is a model with expandable nodes that are suitable for temporal data.
- BiLSTM (bidirectional long-short term memory) is composed of a forward LSTM and a backward LSTM.
- RBF (radial basis function) is a neural network with RBF as activation function.
- CNN (convolutional neural network) is a network for learning spatial features.

We conducted three experiments, severity classification of PD in finger-tapping, fist movement and the above two hand movements. Details will be depicted as follows.

5.1 Experimental Settings

The two models presented in this paper had some adjustable critical parameters, in the 3D hand pose estimation model, the parameters from HandSegNet and PoseNet were defined in advance, HandSegNet allowed us to crop and normalize the inputs in size, which was trained for hand segmentation on R-train with a batch size of 8 and using ADAM solver [15]. The network was initialized using weights of Wei et al. [32] for layers 1 to 16 and then trained data using a standard softmax cross-entropy loss. The learning rate were different in 3 training epochs. From the training images, a 256×256 crop was taken randomly.

The PoseNet was trained using a L2 loss. The learning rate were also different in 3 training epochs. The model employed normal distributions with a variance of 25 pixels for generating ground truth of the score maps, the mean was equal to the given location of key point. By normalizing the resulting maps, each map contained values from 0 to 1 with the key point visible.

We modified the best crop size of hand segment, detect key points in 2D image and estimate most likely 3D pose and preprocessed the data into LSTM input formats. The specific data tensor of the network and the operations were detailed explained in [37].

In the two-channel LSTM model, the time step was adjusted to 10 containing enough hand moving information, after the L2-normalization, 3D positions of 21 hand skeleton joints (63 dimensions) were input into the 1st LSTM Layer, the LSTM network made 0.5 dropout avoid over fitting and 128 hidden layer number of features when training the data with appropriate training iterations, and outputs of the two-channel LSTM network were weighted and summed to gain the combination forecasting result. In addition, we chose 0.001 learning rate and 128 batch size for model training, which was verified to be the optimal scheme. Our experimental scene is depicted in Fig. 7

We implemented our architecture on an Intel Core i5 computer with 31.3 GB RAM and NVIDIA Tesla K40c GPU, in the testing stage, one training sample took 2.1 ms on average.

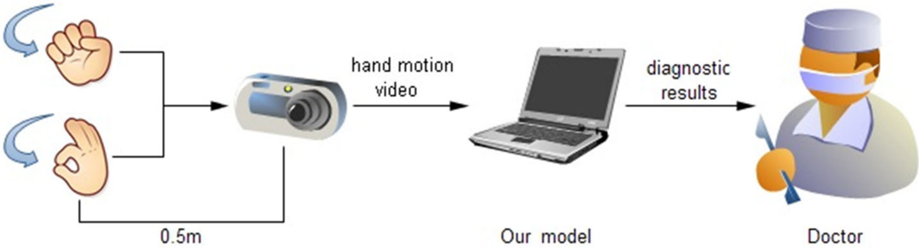

Fig. 7 The experimental scene of PD severity diagnosis

For the evaluation, three measurements are used including the Sensitivity, the Specificity and the Accuracy. These measurements are defined as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{8}$$

where TP is the number of true positives, FN is the number of false negatives, TN is the number of true negatives and FP is the number of false positives.

5.2 Finger-tapping severity classification

In this section, we evaluated 5 severity rating levels of finger-tapping movement including normal state based on UPDRS scale. The 12 subjects were placed at a distance of 0.5 meters from the camera to do the finger-tapping movement with the motion frequency of PD patients (we chose the right hand), everyone completed the action after 500 frames of video were collected. By testing the 3D hand pose estimation model, 3D hand coordinates were obtained to input into the two-layer LSTM by batch with the batch size 128. After being trained with 200000 training iterations, the two-layer LSTM generated 128 dimensional feature vector from every 10 frames, which provided finger-tapping severity classification with reliable diagnostic basis.

We compared our experimental results with other modern machine learning methods using the open-source Python-sklearn library to implement these methods including naive Bayesian (NB), k-nearest neighbor (KNN), logistic regression (LR), random forest (RF), decision tree (DT), support vector machine (SVM, we used one-vs-all strategy), and gradient boosting decision tree (GBDT). By comparison, the performance of our fusion model outperformed these methods. After verification, 10 samples (time steps 10) were considered as a training sample as well as the testing sample. 80% of the data can be regarded as the training data; the rest used for testing. It can be seen that KNN and RF achieved a higher accuracy. The classification results of PD severity of finger-tapping movement are shown

Table 3 The classification results of the traditional machine learning methods

NB	KNN	LR	DT	RF	SVM	GBDT	Our Method
78.9%	90.5%	87.4%	83.1%	92.5%	88.2%	85.5%	93.2%

Table 4 The classification results of the deep learning methods

LSTM	GRU	BiLSTM	CNN	RBF	Our Method
90.98%	91.26%	91.69%	90.42%	89.36%	93.2%

in Table 3. We obtained the specificity of 89.2% and sensitivity of 95.6% with the proposed method.

In addition, we also compared several deep learning methods for finger-tapping severity classification, the results are illustrated in Table 4. The comparison method is explained according to the temporal feature extraction method and the spatial feature extraction method. We can see that the performance of temporal feature extraction methods, i.e., LSTM, GRU and BiLSTM, are superior than that of spatial feature extraction methods, i.e., CNN and RBF, which shows that the regular change of hand movement has strong influence and significant feature. The result of BiLSTM shows that obtaining the information of the previous frame and the next frame increases the amount of hand motion information and achieves the best effect. The proposed method extracts skeleton information from hand image to eliminate other noise and interference information for achieving more accurate classification.

5.3 Fist movement severity classification

In this section, we evaluated 5 severity rating levels of fist movement including normal state based on UPDRS scale. The 3D hand pose estimation model generated 3D hand coordinates to enter into the two-layer LSTM by batch with the batch size 128. After training with 150000 iterations, a 128 dimensional feature vector was output from the two-layer LSTM in every 10 frames, which provided fist movement severity classification with reliable diagnostic basis. We also compared our experimental results with other modern machine learning methods. The classification results of PD severity of fist movement are shown in Table 5 with 80% training data. Moreover, we obtained the specificity of 90.4% and sensitivity of 98.2% with the proposed method.

For the deep learning models shown in Table 6, we find that the performance of BiLSTM ranks the first among temporal models. The performance of CNN and LSTM is similar, indicating that the temporal and spatial characteristics are also significant. The learning ability of RBF is slightly weaker than other deep models.

5.4 Severity classification of finger-tapping and fist movement

In this section, we evaluated 5 severity rating levels of the two hand movements including normal state based on UPDRS scale. The two kinds of hand motion videos were input into the trained fusion model with 300000 training iterations, the classification accuracy of the two movements summed up according to the 50% weights in every training iteration, and

Table 5 The classification results of the traditional machine learning methods

NB	KNN	LR	DT	RF	SVM	GBDT	Our Method
80.9%	91.3%	88.5%	85.3%	94.4%	89.0%	87.5%	96.8%

Table 6 The classification results of the deep learning methods

LSTM	GRU	BiLSTM	CNN	RBF	Our Method
94.33%	95.28%	95.97%	94.58%	90.39%	96.8%

the optimal solution was obtained using the other settings of the above two experiments. In addition, we combined the original hand skeleton joints from the two movements using the above machine learning methods to analyze and extract important features. The classification results of the two hand movements are shown in Table 7. It is verified that our model is superior to these fusion feature classification methods (the popular classifiers). Through the joint training of these two movements, we obtained the specificity of 98.9% and sensitivity of 93.5% with the proposed method.

For the deep learning models shown in Table 8, we can see that the performance of all the methods has an increase after data fusion. For training the fist and the finger-tapping movements data, all the models have information gain. Our proposed method still shows its strong learning ability.

Additionally, we also divided the proposed model into two parts for ablation experiment, and the results are shown in Table 9. The model was split into fist-channel LSTM and finger-tapping channel LSTM (with hand pose estimation (HPE)), fist-channel LSTM and finger-tapping channel LSTM (without hand pose estimation). We can see that the best result is to consider the two hand movements at the same time, and the second is to analyze the fist movement after the hand pose estimation. It can be concluded that hand pose estimation and two channels structure can promote accurate hand feature classification.

6 Discussion

The life quality of people suffered from PD depends on the management of their condition in the form of tailored treatment plans. Devising such plans is a challenge, as objective information about fluctuations in disease state is not accessible in clinical practice beyond recall by the individual. Current best practice in automated assessment of PD is to obtain data in medical institution conditions, where big amounts of clinician-validated behavior can be observed. While such applications show good performance in this setting, it is unlikely that they make objective, efficient and detailed analysis in daily lives. In order to address this issue, assessment applications are dedicated to developing more superior models for handling different situations ([11], [19], [5], [35]). On the basis of these methods, we can present new viewpoints to solve the problem of PD severity rating.

In this work we suggest a novel methodology to assess different severity of PD, which can not only simplify complicated diagnosis process, but also assist physicians to rate disease level effectively. It can also quantify hand data characteristics and provide a detailed objective description of diseases.

Table 7 The classification results of the traditional machine learning methods

NB	KNN	LR	DT	RF	SVM	GBDT	Our Method
79.9%	92.1%	82.3%	80.0%	90.3%	89.3%	89.5%	97.2%

Table 8 The classification results of the deep learning methods

LSTM	GRU	BiLSTM	CNN	RBF	Our Method
94.81%	95.19%	96.21%	96.3%	93.63%	97.2%

In our experiments we showed that deep learning seems particularly suitable to discover disease features and has been applied in similar fields, such as speech or gait, where labeled data is easily accessible. According to the survey, there are most researches focus on the diagnosis of PD according to data processing and feature recognition before disease diagnosis, only a few pay attention to treatment or rehabilitation period after diagnosis. In addition, patient data is not public and is difficult to acquisition limiting the application of excellent algorithms in medical diagnosis. Equipment for monitoring and collecting patient's action clinical signs and symptoms should be developed to obtain more disease information. We will collect data from PD patients' daily lives to describe and analyze PD deeply for evidence-based medicine.

7 Conclusion

In this paper, we reported our research of predicting the severity of Parkinson's disease using the serial neural network consists of two deep learning models using hand movement data. We proposed a two-channel LSTM to model the hand data (single image) recorded by RGB camera. The proposed model can capture the motion trajectory of hand skeleton over a period of time and evaluate the severity of Parkinson's disease. To overcome the limits of single feature analysis, we chose a pattern recognition strategy to identify two groups of features that provide specific information on visual sensor based recorded movement impairments. The model was evaluated and the best classification results were calculated. In general, PD motor symptoms appear first in upper extremities. With disease progression symptoms also affect lower extremities. We have been learning abnormal gait performance of patients with Parkinson, which will help to understand and diagnose PD more deeply.

In the future, the classification experiment will aim to differentiate between different diseases or disease stages. A large set of patient and control data will be used to do an individual assessment for early diagnosis and disease monitoring. The goal is to build up a system for an immediate rating of the movement impairment. Therefore, a recording of more patient and control data is required for multi-feature analysis. With these data, we will use a variety of motion features of PD, including upper limb movements and lower limb movements (which will involve non-motor features, such as brain images, speech features), to establish a comprehensive auxiliary diagnosis and rehabilitation system of PD based on the combination of multi-feature and multi-model.

Table 9 The classification results of the individual models

fist-channel LSTM	finger-tapping channel	fist-channel LSTM	finger-tapping channel	Our Method
(with HPE)	LSTM (with HPE)	(without HPE)	LSTM (without HPE)	
96.8%	93.2%	94.33%	90.98%	97.2%

Acknowledgements This research was supported in part by National Key Research and Development Plan Key Special Projects under Grant No. 2018YFB2100303, Shandong Province colleges and universities youth innovation technology plan innovation team project under Grant No. 2020KJN011, Shandong Provincial Natural Science Foundation under Grant No. ZR2020MF060, Program for Innovative Postdoctoral Talents in Shandong Province under Grant No. 40618030001, National Natural Science Foundation of China under Grant No. 61802216, and Postdoctoral Science Foundation of China under Grant No.2018M642613. National Natural Science Foundation of China under Grant No.62106117, and Shandong Provincial Natural Science Foundation under Grant No.ZR2021QF084.

References

1. Alty JE, Cosgrove J, Lones MA, Smith SL, Possin K, Schuff N, Jamieson S (2016) Clinically 'slight' bradykinesia in parkinson's disease is accurately detected using evolutionary computation analysis of finger tapping. In: International Parkinsons and Movement Disorders Society Congress
2. Ariyanto M, Caesarendra W, Mustaqim KA, Irfan M, Pakpahan JA, Setiawan JD, Winoto AR (2015) Finger movement pattern recognition method using artificial neural network based on electromyography (emg) sensor. In: Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT), 2015 International Conference on. IEEE, pp 12–17
3. Arora S, Venkataraman V, Zhan A, Donohue S, Biglan KM, Dorsey ER, Little MA (2015) Detecting and monitoring the symptoms of parkinson's disease using smartphones: a pilot study. Parkinsonism Relat Dis 21(6):650–653
4. Camgöz NC, Kindiroglu AA, Akarun L (2014) Gesture recognition using template based random forest classifiers. In: ECCV Workshops (1), pp 579–594
5. Chen HL, Wang G, Ma C, Cai ZN, Liu WB, Wang SJ (2016) An efficient hybrid kernel extreme learning machine approach for early diagnosis of parkinson's disease. Neurocomputing 184(C):131–144
6. Diaz M, Ferrer MA, Impedovo D, Pirlo G, Vessio G (2019) Dynamically enhanced static handwriting representation for parkinson's disease detection. Pattern Recognition Letters
7. Escalante HJ, Morales EF, Sucar LE (2016) A naive bayes baseline for early gesture recognition. Pattern Recogn Lett 73:91–99
8. Gage H, Hendricks A, Zhang S, Kazis L (2003) The relative health related quality of life of veterans with parkinson's disease. J Neurol Neurosurg Psychiatry 74(2):163
9. Giancardo L, Sánchez-Ferro A, Arroyo-Gallego T, Butterworth I, Mendoza CS, Montero P, Matarazzo M, Obeso JA, Gray ML, Estépar RSJ (2016) Computer keyboard interaction as an indicator of early parkinson's disease. Sci Rep 6(10):1–10
10. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
11. Ireland D, Wang Z, Lamont R, Liddle J (2016) Classification of movement of people with parkinsons disease using wearable inertial movement units and machine learning. Stud Health Technol Inform 227:61
12. Keisuke S, Toshio T, Akihiko K, Masaru Y, Saburo S (2009) Measurement and evaluation of finger tapping movements using log-linearized gaussian mixture networks. Sensors 9(3):2187–201
13. Khan T, Nyholm D, Westin J, Dougherty M (2014) A computer vision framework for finger-tapping evaluation in parkinson's disease. Artif Intell Med Artif Intell Med 60(1):27–40
14. Kim H, Lee S, Lee D, Choi S, Ju J, Myung H (2015) Real-time human pose estimation and gesture recognition from depth images using superpixels and svm classifier. Sensors 15(6):12410–12427
15. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. Computer Science
16. Krupicka R, Szabo Z, Jirina M (2011) Motion camera system for measuring finger tapping in parkinson's disease. Springer, Berlin Heidelberg
17. Kupryjanow A, Kunka B, Kostek B (2010) Updrs tests for diagnosis of parkinson's disease employing virtual-touchpad. In: Database and Expert Systems Applications, pp 132–136
18. Li F, Ge R, Zhou H, Wang Y, Liu Z, Yu X (2020) Tesia: A trusted efficient service evaluation model in internet of things based on improved aggregation signature. Concurrency and Computation: Practice and Experience
19. Li Y, Yang L, Wang P, Zhang C, Xiao J, Zhang Y, Qiu M (2017) Classification of parkinson's disease by decision tree based instance selection and ensemble learning algorithms. Journal of Medical Imaging & Health Informatics 7(2)

20. Liu X, Xia Y, Yu H, Dong J, Jian M, D. Pham T (2021) Region based parallel hierarchy convolutional neural network for automatic facial nerve paralysis evaluation. IEEE Transactions on Neural Systems and Rehabilitation Engineering, pp 2325–2332
21. Liu X, Xia Y, Yu H, Dong J, Jian M, Pham T (2020) Region based parallel hierarchy convolutional neural network for automatic facial nerve paralysis evaluation. IEEE Trans Neural Syst Rehabilitation Eng 10:2325–2332
22. Molchanov P, Gupta S, Kim K, Kautz J (2015) Hand gesture recognition with 3d convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 1–7
23. Molchanov P, Yang X, Gupta S, Kim K, Tyree S, Kautz J (2016) Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 4207–4215
24. Om MS, Duarte M, Diana R, Divyanshu D, Bahman J (2018) Botulinum toxin in essential hand tremor - a randomized double-blind placebo-controlled study with customized injection approach. Parkinsonism Relat Dis 56:65–69
25. Papadopoulos A, Kyritsis K, Klingelhoefer L, Bostanjopoulou S, Delopoulos A (2019) Detecting parkinsonian tremor from imu data collected in-the-wild using deep multiple-instance learning. IEEE J Biomed Health Inform 24(9):2559–2569
26. Parziale A, Senatore R, Cioppa AD, Marcelli A (2021) Cartesian genetic programming for diagnosis of parkinson disease through handwriting analysis: Performance vs. interpretability issues. Artif Intell Med 111:1–13
27. Printy BP, Renken LM, Herrmann JP, Lee I, Johnson B, Knight E, Varga G, Whitmer D (2014) Smartphone application for classification of motor impairment severity in parkinson's disease. Conf Proc IEEE Eng Med Biol Soc 2014:2686–2689
28. Sano Y, Kandori A, Shima K, Yamaguchi Y, Tsuji T, Noda M, Higashikawa F, Yokoe M, Sakoda S (2016) Quantifying parkinson's disease finger-tapping severity by extracting and synthesizing finger motion properties. Med Biol Eng Comput 54(6):953–965
29. Stamatakis J, Ambroise J, Crémers J, Sharei H, Delvaux V, Macq B, Garraux G (2013) Finger tapping clinimetric score prediction in parkinson's disease using low-cost accelerometers. Computational Intelligence and Neuroscience,2013,(2013-4-16) 2013(2):1
30. Tsironi E, Barros P, Weber C, Wermter S (2017) An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition. Neurocomputing
31. Wang Y, Dong X, Li G, Dong J, Yu H (2021) Cascade regression-based face frontalization for dynamic facial expression analysis. Cognitive Computation 99(3)
32. Wei SE (2016) Convolutional pose machines: A deep architecture for estimating articulated poses. PhD thesis
33. Yokoe M, Okuno R, Hamasaki T, Kurachi Y, Akazawa K, Sakoda S (2009) Opening velocity, a novel parameter, for finger tapping test in patients with parkinson's disease. Parkinsonism Relat Dis 15(6):440–444
34. Yu X, Li F, Li T, Wu N, Zhou H (2020) Trust-based secure directed diffusion routing protocol in wsn. J Ambient Intell Humaniz Comput J Amb Intel Hum Comp 99(5):1–13
35. Zeng W, Liu F, Wang Q, Wang Y, Ma L, Zhang Y (2016) Parkinson's disease classification using gait analysis via deterministic learning. Neurosci Lett 633:268–278
36. Zhou Y, Jenkins ME, Naish MD, Trejos AL (2018) Characterization of parkinsonian hand tremor and validation of a high-order tremor estimator. IEEE Trans Neural Syst Rehabilitation Eng 26(9):1823–1834
37. Zimmermann C, Brox T (2017) Learning to estimate 3d hand pose from single rgb images. ICCV 2017

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.