Temporal pyramid attention-based spatiotemporal fusion model for Parkinson's disease diagnosis from gait data

Xiaomin Pei\(^1\,2\,3\), Huijie Fan\(^2\,3\), Yandong Tang\(^2\,3\)

\(^1\)School of Information and Control Engineering, Liaoning Shihua University, Fushun, China
\(^2\)State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China
\(^3\)Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, China

Abstract

Parkinson's disease (PD) is currently an ongoing challenge in daily clinical medicine. To reduce diagnosis time and arduousness and even assess PD levels, a temporal pyramid attention-based spatiotemporal (PAST) fusion model for diagnosis of PD is produced by using gait data from ground reaction forces. This model is innovative in two aspects. First, by using the temporal pyramid attention module, multiscale temporal attention is obtained from raw sequences. Second, 1D convolutional neural network and bidirectional long short-term memory layers are used together to learn spatial fusion features from multiple channels in the spatial domain to obtain multichannel, multiscale fusion features. Experiments are performed on the PhysioBank data set, and the results show that the proposed PAST model outperforms other state-of-the-art methods on classification results. This model can assist in the diagnosis and treatment of PD by using gait data.

1 | INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative brain disorder. This disorder predominately affects dopaminergic neurons in the substantia nigra, which control balance and movement [1]. It is the second most common neurodegenerative disease seen in people over 60 years of age, just after Alzheimer's. In 2015, PD affected 6.2 million people and resulted in approximately 117,400 deaths globally [2]. With increasing public awareness, computer-aided diagnosis (CAD) of PD has been developed to assist clinicians in diagnosing it in the early stage.

The primary symptoms of PD are tremor, postural instability, muscle rigidity, and slowness of movements. These symptoms alter gait in the early stage of PD [3]. Gait parameter analysis as an automatic, non-invasive method has been widely used for the diagnosis of PD by observing different kinds of gait features, such as stride fluctuations in PD, time-stamped gait data, wrist sensors, gait rhythm signals, and signal turn counts.

Ground reaction forces (GRFs) are a kind of gait variable that can be easily measured by force-sensitive resistors on feet non-invasively. GRFs are different among persons and even different for the same person at different times [4]. Therefore, handcrafted feature extraction methods have limits for gait analysis using GRFs. Although many researchers are focused on gait data analysis, they still lack objective tools to assist physicians in gait evaluation. There are still challenges for detecting PD and predicting the PD severity rate from gait data.

With the current deluge of the success of deep learning and the urgent need for PD diagnosis, we propose a spatiotemporal fusion model to classify and estimate PD severity levels from gait reaction force data. The flow chart of our temporal pyramid attention-based spatiotemporal (PAST) fusion model is shown in Figure 1.

The main contributions of this work are summarized as follows:

1. A data-level fusion method is proposed by incorporating multiple sources of raw sensor data to acquire more definite, informational, and comprehensive fused data than that of the primitive sources.
2. The combination of 1D convolutional neural network (CNN) and bidirectional long short-term memory
(BiLSTM) layers is studied to extract spatial features for multichannel sensor data, to obtain more comprehensive multisensor fusion features.

(3) With the proposed temporal pyramid attention module, global and local temporal attention can be obtained from input sequences to enhance the multiscale temporal characteristics of the input signals. This module is a self-attention module, with no need for additional training parameters.

Extensive experiments are performed to validate the effectiveness of our model, which realizes state-of-the-art results on the PhysioBank data set [5].

2 | RELATED WORKS

In PD diagnosis, large quantities of human activity data have been collected, and numerous feature extraction and machine learning methods have been explored in previous works. Here, we divide these methods into traditional handcrafted feature extraction methods and deep learning methods.

In traditional handcrafted feature extraction methods, Song-Hong Lee et al. [6] used both gait characteristics of VGRF for PD patients’ feet and wavelet-based feature extraction to classify PD patients and healthy controls. Daliri et al. [7] applied a short-time Fourier transform (STFT) on each input signal using the same vertical ground reaction force (VGRF) input data. Ertegrul et al. [8] proposed an algorithm based on shifted 1D local binary patterns with multilayer perceptron to classify the feature vectors. Wei Zeng et al. [9] used phase space reconstruction, empirical mode decomposition, and neural networks to classify gait patterns between patients with PD and healthy controls. Although the aforementioned methods achieve satisfactory results on specific data sets, they are limited in their capacity for different data sets because features are manually defined.

In deep learning, recurrent neural networks (RNNs) are widely used to process time series and sequences. Many RNN
variants, such as LSTM (long short-term memory) [10] and GRUs (gated recurrent units) [11], have been proposed to improve RNN. BiLSTM [12] is a further development of LSTM that can access both the preceding and following contexts by combining the forward hidden layer and the backward hidden layer. It can also access both past and future contexts. The attention mechanism (AM) [13] can highlight important information from contextual information by setting different weights. The combination of LSTM and AM can further improve the classification accuracy. Aite Zhao et al. [14] proposed a two-channel model that combines LSTM and CNN to learn the spatiotemporal patterns of the gait data, which was defined as a decision-level fusion method. Maachi et al. [15] proposed a decision attention network for action classification, which constructed a spatial pyramid and utilized multiscale information to obtain more accurate attention scores. Hanchao Li et al. [17] introduced a feature pyramid attention module and a global attention upsampling module, which exploited the impact of global contextual information as a guidance for low-level features to select category localization details.

Inspired by the aforementioned methods, we present a temporal PAST fusion model for diagnosis of PD from gait data. It involves a temporal pyramid attention module for multiscale attention extraction in the time domain and a spatiotemporal domain fusion method for multichannels, which results in higher classification accuracy than other methods. It is an end-to-end deep learning method for which the input is gait data and the output is the classification results for PD levels.

3 | OUR APPROACH

In this section, we describe the PAST model in detail. First, this method uses a temporal pyramid attention model to extract multiscale temporal attention. Second, considering spatial relations between channels, it uses a spatiotemporal fusion model to learn the spatial and temporal features of the sequences. Then, features are classified into PD levels by the classification layer.

3.1 | Temporal pyramid attention model

The spatial pyramid attention module has been widely used to obtain better feature representation by combining multiscale features [16,17]. Motivated by this model, we present a temporal pyramid attention model to extract multiscale temporal attention from raw inputs.

The temporal pyramid attention model is composed of multiple streams for extracting multiscale attention, as shown in Figure 1. Considering that attention covers range, we define attention as global attention $G_{\text{attention}}$ and local attention $L_{\text{attention}}$. The top stream calculates $G_{\text{attention}}$ and the other streams calculate $L_{\text{attention}}$ with different time intervals. Global attention $G_{\text{attention}}$ is defined as comparisons between values of every time and the global maximum. $X_{\text{Global}}$ is obtained by the global max-pooling layer and then upsampled to the same dimension as input $X$. $G_{\text{attention}}$ is calculated using the 1-norm by Equation (1). Local attention $L_{\text{attention}}$ is defined as comparisons between values for every time and the local maximum, as Equation (2). $X_{\text{Local}}$ is obtained by a max-pooling layer with certain time intervals and then upsampled to the same dimension as input $X$. Different time intervals are used to calculate $X_{\text{Local}}$ to obtain multiscale temporal attention by using multistreams.

$$G_{\text{attention}} = \|X - X_{\text{Global}}\|_1 \quad (1)$$

$$L_{\text{attention}} = \|X - X_{\text{Local}}\|_1 \quad (2)$$

3.2 | Spatial feature learning

VGRF sensors are at specific positions on human feet. There is great relevance among multiple sensors, so we present a spatial feature learning model to learn about the spatial features among sensors at time $t$. To maintain the temporal order of the sequences, 1D CNN is used to learn about the relations among sensors, while BiLSTM is used to fuse multiple features to obtain a new representation for all channels.

As shown in Figure 2, we use 1D convolutional filters to slide over channels of sensors and learn representations for spatial features. Then, we obtain a group of feature maps $C = \{X_1, X_2, \ldots, X_n\}$ at time $t$, $t \in (1, 2, \ldots, T)$. We treat values on every feature map for one point $i, i \in (1, 2, \ldots, N)$ as a vector $f_i$. BiLSTM is used to learn the relations between every channel at time $t$, and the relations between every point from $I$ to $N$ and $N$ to $I$ are learnt. Then, we obtain a whole representation for all the channels at $t$.

3.3 | Temporal feature learning

LSTM is used to learn the relation between every time from 1 to $T$. The model we propose consists of two LSTM layers, and the inputs to the first LSTM are features for each time $t$. The outputs of the first LSTM layer are \( T \) length sequences. For the second LSTM layer, the output is a one-dimensional vector. Fully connected layers are used to map the vectors to the dimension space, while the softmax layer is used to classify vectors into labels of different PD levels.

4 | EXPERIMENTS

4.1 | VGRF signal and data set

The VGRF signal is obtained from the feet sensors. In our experiments, data from the public data set PhysioNet are used.
This data set consists of three PD gait sub-data sets—the Ga data set [18], the Si data set [19], and the Ju data set [20] — which were contributed by three groups of researchers, Frenkel-Toledo [18] et al., Yogev [19] et al., and Hausdorff [20] et al. The database includes measures of VGRFs of 93 patients with idiopathic PD (mean age: 66.3 years, 63% men) with moderate disease severity (H&Y stage 2–3) and 73 healthy controls (CO) (mean age: 66.3 years, 55% men). Every participant wears a pair of shoes with eight force sensors under each foot. The sensors measure forces in Newton as functions of time (seconds). The outputs of all sensors are digitized and recorded at 100 samples per second. The records also include two signals (input columns 18 and 19) that each represents the sum of eight sensor outputs for each foot. VGRF signals are divided into several segments and labelled by subject category according to the participants PD level. VGRF signal segments are the input samples of our model.

4.2 | Rating scales for PD

The Hoehn and Yahr scale (H&Y) [21, 22] is used worldwide for scaling functional disability and objective signs in PD. It was originally designed on a five-point scale (1–5). A modified H&Y scale that includes 0.5 increments was introduced for clinical trials—Stage 1.0: unilateral involvement only; Stage 1.5: unilateral and axial involvement; Stage 2.0: bilateral involvement without impairment of balance; Stage 2.5: mild bilateral disease, some postural instability, physically independent; Stage 4.0: severe disability, still able to walk or stand unassisted; Stage 5.0: wheelchair bound or bedridden unless aided. In our experiments, we only classified PD with levels 1.0, 1.5, 2.5 and CO.

4.3 | Implementation details

Experiments were implemented using the Tensorflow and Keras deep learning framework and tested on an Intel Core i7 computer with 16 GB RAM. During training, the root mean square optimizer was used to optimize the network. The initial learning rate was set to 0.01 and reduced by multiplying it by 0.1 every 50 epochs. The batch size was 32. We used cross-entropy loss to compute the cross-entropy loss between network predictions and target values for classification, which is defined by Equation (3). $X_{ij}$ was the network response for a given category, $T_{ij}$ was the target value of that category, and $C$ was the total number of categories; here, $C$ was 4. The output was four numbers with 0,1,2, and 3, for which 0 indicated a healthy person, and PD levels 1.0,1.5, and 2.5 were labelled with 1,2, and 3, respectively. The cross-entropy loss was calculated as the probability of a given observation being assigned to a given category, summed over all categories and observations, and normalized by the number of all observations.

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} \left( T_{ij} \log(X_{ij}) + (1 - T_{ij}) \log(1 - X_{ij}) \right) \quad (3)$$
For preprocessing, we reshaped VGRF signal segments as $100 \times 19 \times N$, and each sequence had $100 \times 19$ dimensions. One hundred samples were collected in two seconds, and the interval was 0.02 seconds. We found that the characteristics were not obvious at the beginning of the signal, so we discarded 50 seconds at the start. To enlarge the data set for training, we separated the time series with 50 shifting movements each time, and then we obtained a data set of $100 \times 19 \times 60,000$, where 60,000 was the number of sequences. We used fivefold cross validation to validate our model, for which we also ensured that the training data and testing data were from different persons.

4.4 | Model design

4.4.1 | The design of the temporal attention model

In this section, we discuss the design of the temporal attention model. Local attention was calculated at several local time intervals. We first selected $T = 2, 4, 8, 16, \text{and} 32$ samples as timescales. The enhanced classification results by models with different time scales are shown in Figure 3; we found that $T = 2$ had no effect on improving classification results, while $T = 8, 16, \text{and} 32$ had little effect compared with $T = 4$; we also increased time intervals to 64 samples, but it had little influence on the classification results. We used two intervals with different combinations, which had the same effect as $T = 4$. Therefore, we chose $T = 4$ samples as local time intervals to learn the local attention of the series.

Global attention and local attention were used separately to show the effectiveness of the temporal pyramid attention model. Experiments were performed for PD level classification with different attention models. The results are shown in Figure 4.

It can be seen that temporal pyramid attention obtained better classification results than single global attention or local attention, for all data sets. Although the enhancement of classification accuracy was less than 1%, it was important in the PD level classification, because it indicated a PD misdiagnosis rate [23,24].

4.4.2 | The selection of model parameters

We performed experiments to set parameters and optimize layers for the model. We selected different parameters and compared the classification accuracy. Finally, we chose parameters and layers with the best classification accuracy for our model, which are shown in Table 1.

4.5 | Algorithm evaluation

The first experiment was PD and healthy control classification. We used Python to realize the methods in papers [6–9, 14, 15, 25].

We compared our methods with these methods, and all the methods used the same data set for validation. Classification results are shown in Table 2. Accuracy (Acc.), Precision (Pr.), F1-core (F1), Recall (Re.) are recorded in Table 2. Compared with these methods, the PAST method achieved the highest performance in classifying PD patients and healthy controls. The classification accuracy for Ga dataset was 99.5%, for Ju dataset was 98.6%, for Si dataset was 99.5%, and for all the whole dataset was 99.2%. The PAST method also had the highest precision, F1-score, and recall for all data sets.

The second experiment was the PD level classification. The results are shown in Table 3. The PAST method also achieved...
the highest classification accuracy on classifying PD patients with different severity levels, with a classification accuracy of 99.7% for the Ga data set, 98.5% for the Ju data set, 99.5% for the Si data set, and 98.9% for the whole data set. The PAST method had the highest precision, F1-score, and recall for all data sets. According to the results in the second experiment, we created confusion matrices for different data sets to show the accuracy of predicting PD rating levels by the PAST model in Figure 5.

4.6 Ablation study

In this section, we conducted ablation experiments to investigate the effect of different components in our model. First, we evaluated the necessity of the temporal pyramid attention model; we only used the spatiotemporal fusion model, which we called the ST model. Second, we evaluated the necessity of the CNN–BiLSTM layer for spatial feature learning, where only the stacked LSTM model was used, which we called the S-LSTM model. Third, we evaluated the necessity of the BiLSTM layer for feature learning, where CNN was used for spatial feature learning and LSTM was used for temporal feature learning, which was called the C-LSTM model. The results of these models for PD level classification on the whole data set are shown in Table 4. We also increased the layers and parameters in every model for our experiments. The results in Table 4 were the best for our model, so we considered that the classification results only had relevance with the structure of the model.

The PD classification accuracy for the C-LSTM model was 95.72%, which was the lowest; the classification result for the ST model was higher than S-LSTM and C-LSTM; the PAST model evidently achieved the highest classification accuracy on classifying PD patients with different severity levels, with a classification accuracy of 99.7% for the Ga data set, 98.5% for the Ju data set, 99.5% for the Si data set, and 98.9% for the whole data set.
result of 99.18%, which verified PAST had the best performance for PD classification; all of the components in the PAST model were necessary for PD level classification.

**TABLE 4** Classification results for different models

| Method | Acc. (For the whole data set) |
|--------|-------------------------------|
| ST     | 98.02%                        |
| S-LSTM | 97.12%                        |
| G-LSTM | 95.72%                        |
| PAST   | 99.18%                        |

**5 | CONCLUSION**

We presented a temporal PAST fusion model for diagnosis of PD from gait data. A temporal pyramid attention module was proposed to enhance the multiscale temporal characteristics. The 1D CNN and BiLSTM layer was used to learn the spatial domain features of the sequences. With this structure, we obtained a data-level fusion method for multiple channels. We performed experiments on the PhysioBank data set. From the experimental results, we can see that classification accuracy was enhanced by our methods for both PD CO classification and PD level classification. Moreover, the PAST model can also be oriented towards CAD of other diseases [26] in future works by using spatiotemporal signals.

**ACKNOWLEDGEMENT**

This work was supported in part by the National Natural Science Foundation of China under grants 61873259 and U1613214 and the Liaoning Natural Science Foundation 2019ZD0066.

**ORCID**

Xiaomin Pei https://orcid.org/0000-0003-3847-6162

**REFERENCES**

1. Paul, A., Yadav, K.S: Parkinson's disease: Current drug therapy and unraveling the prospects of nanoparticles. J. Drug Deliv. Sci. Technol. 58(101790), 1773–2247 (2020)
2. Benito-León, J., et al.: Under-reporting of Parkinson's disease on death certificates: a population-based study (NEDICES). J. Neurol. Sci. 347(1–2), 188–192 (2014)
3. Bhat, S., et al.: Parkinson's disease: cause factors, measurable indicators, and early diagnosis. Comput. Biol. Med. (102), 234–241 (2018)
4. Minamisawa, T., Elan D: Characteristics of temporal fluctuation of the vertical ground reaction force during quiet stance in Parkinson's disease. Gait Posture. 35(2), 308–311 (2012)
5. NIGMS, Physionet Database. http://physionet.org/pn3/gaitpdb/
6. Lee, S-H., et al.: Parkinsons disease classification using gait characteristics and wavelet-based feature extraction. Expert Syst. Appl. 39(8), 7338–7344 (2012)
7. Daliri, M.R.: Chi-square distance kernel of the gait for the diagnosis of Parkinson’s disease. Biomed. Signal Proc. Control. 8(1), 66–70 (2013)
8. Ertegrul, O.F., et al.: Detection of Parkinson’s disease by shifted one dimensional local binary patterns from gait. Expert Syst. Appl. 56, 156–163 (2016)
9. Zeng, W., et al.: Classification of gait patterns between patients with parkinson's disease and healthy controls using phase spaces reconstruction(PSR), empirical mode decomposition(EMD) and neural networks. Neural Netw. 111, 64–76 (2019)
10. Hochreiter, S., Schmidhuber, J: Long short-term memory. Neural Comput. 9, 1735–1780 (1997)
11. Liu, J., Wu, C., Wang, J.: Gated recurrent units based neural network for time heterogeneous feedback recommendation. Infor. Sci. 435, 50–65 (2018)
12. Liu, G., Guo, J.: Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing. 337(4), 325–338 (2019)
13. Qi, F., et al.: Spatio-temporal fall event detection in complex scenes using attention guided LSTM. Pattern Recog. Lett. 130, 242–249 (2020)
14. Zhao, A., et al.: A hybrid spatio-temporal model for detection and severity rating of parkinson's disease from gait data. Neurocomputing. 315, 1–8 (2018)
15. El Maachi, I., Bilodeau, G.-A., Bouachir, W.: Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait. Expert Syst. Appl. 143, 113075 (2020)
16. Du, Y., et al.: Interaction-aware spatio-temporal pyramid attention networks for action classification. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 373–389 (2018)
17. Wang, Y., et al.: Spatio temporal Pyramid Network for Video Action Recognition. Computer Vision & Pattern Recognition(CVPR) IEEE, pp. 529–1538 (2017)
18. Frenkel-Toledo, S., et al.: Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease. Mov. Disord. 20(9), 1109–1114 (2005)
19. Yohe, G., et al.: Dual tasking, gait rhythmicity, and parkinson's disease: which aspects of gait are attention demanding?. Eur. J. Neurosci. 22(5), 1248–1256 (2005)
20. Hausdorff, J.M., et al.: Rhythmic auditory stimulation modulates gait variability in parkinson's disease. Eur. J. Neurosci. 26(8), 2369–2375 (2007)
21. Tanaas, A., et al.: Statistical analysis and mapping of the unified Parkinson's disease rating scale to Hoehn and Yahr staging[J]. Parkinsonism Relat. Disord. 18(5), 697–699 (2012)
22. Rabey, J.M., Korczyn, A.D.: The Hoehn and Yahr Rating Scale for Parkinson's disease. In: Przuntek, H., Kraus, P.H., Klotz, P., Korczyn, A.D. (eds.), Instrumental Methods and Scoring in Extrapyramidal Disorders. Springer, Berlin, Heidelberg (1995)
23. Zhang, Y.: Can a smartphone diagnose Parkinson disease? A deep neural network method and telediagnosis system implementation. Parkinsons Disease, 1–11 (2017)
24. Okan Sakar, C., et al.: A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. Appl. Soft Comput. (74), 255–263 (2019)
25. Perumal, S.V., Sankar, R.: Gait and tremor assessment for patients with Parkinson's disease using wearable sensor. ICT Express. 2, 168–174 (2016)
26. Laroche, D., Tolambiya, A., et al.: A classification study of kinematic gait trajectories in hip osteoarthritis. Comput. Biol. Med. (55), 42–48 (2014)

How to cite this article: Pei X, Fan H, Tang Y. Temporal pyramid attention-based spatiotemporal fusion model for Parkinson's disease diagnosis from gait data. IET Signal Process. 2021;15:80–87. https://doi.org/10.1049/spl2.12018