Hindawi
International Journal of Intelligent Systems
Volume 2023, Article ID 9921809, 18 pages
https://doi.org/10.1155/2023/9921809

Research Article

A Spatio-Temporal Siamese Neural Network for Multimodal Handwriting Abnormality Screening of Parkinson's Disease

Aite Zhao, Huimin Wu, Ming Chen, and Nana Wang

College of Computer Science and Technology, Qingdao University, Qingdao, China

Correspondence should be addressed to Aite Zhao; aitezhao@qq.com

Received 13 January 2023; Revised 21 March 2023; Accepted 23 March 2023; Published 14 April 2023

Academic Editor: Subrata Kumar Sarker

Currently, hand motion recognition of single-modality data has been extensively explored for the analysis of various contact and noncontact sensors, and it is recognized that all the existing technologies have both strengths and limitations. As a significant motor symptom, hand tremor is usually utilized for the diagnosis and evaluation of Parkinson's disease; furthermore, a multimodal analysis of the handwriting pattern of the patient has made up for the one-sided way of learning the hand movement in a single measurement dimension. Especially, considering a variety of measurement resources, it shows promising performance in recognizing handwriting patterns of Parkinson's disease. In this work, a novel Spatio-temporal Siamese neural network (ST-SiamNN) is proposed to learn the handwriting differences between healthy individuals and patients with Parkinson's disease, process data onto multiple sensors, and enhance the characteristics of handwriting in Parkinson's disease. Uniquely, it is a discriminative model of multilabel and multinetwork constructed by a Siamese network, which consists of four modules: a preprocessor for handwritten data enhancement, a Siamese bidirectional memory neural network (SiamBiMNN) for temporal and texture feature extraction and difference enhancement, a Siamese octave convolutional neural network (SiamOctCNN) for spatial feature extraction and difference enhancement, and a decision-making layer to rejudge the output features of the Siamese networks to obtain more accurate auxiliary diagnosis results. The framework proposed in this article is verified on two handwritten datasets of multiple modalities, i.e., images, smart pen signals, and graphics tablet signals, which are compared with several state-of-the-art studies.

1. Introduction

1.1. Background. Parkinson's disease (PD), also known as tremor paralysis, is a common neurodegenerative disease in the middle-aged and elderly. Tremor, myotonia, and decreased movement are the main clinical features of the disease. Because the flexion and extension of the upper limb are regulated by the central nervous system, the hand movements of patients with Parkinson's disease will be abnormal due to the degeneration of motor neurons. Therefore, the analysis of handwriting parameters is very valuable for better understanding the mechanism of hand dyskinesia and the development of Parkinson's disease. Patients with Parkinson's disease of different severity have their own unique writing modes. With the development of the disease, the writing data became more irregular [1, 2]. The hand tremor of PD patients is characterized by a static and obvious tremor, which is aggravated when nervous or excited, and the tremor is reduced or stopped during random exercise. Additionally, PD patients have small finger movements, usually showing the thumb and index finger as a "pill rubbing action," which is a potential factor leading to abnormal handwriting. Therefore, as one of the most effective detection methods of early Parkinson's disease, handwriting has been widely used in clinical diagnosis. Moreover, the hand movement away from PD patients with different severities has distinguished features, which can be evaluated by the unified Parkinson's disease score scale (UPDRS) and the Hoehn and Yahr score scale (H & Y) [3]. This article proposes an automatic learning framework that detects the potential for Parkinson's disease based on handwriting characteristics.

1.2. Prior Technologies and Limitations. Previously, there were a number of machine learning methods and technologies applied in the field of data classification [4–6]. With regard to the challenging PD diagnosis problem, these methods provide superior specifications. Examples of such classifiers are support vector machines (SVMs) [7], random forest [8], decision trees [9], naive Bayes [10], artificial neural networks (ANNs) [11–13], the sequential minimal optimization (SMO) method [14], and convolutional neural networks (CNNs) [15–19]. However, these methods ignore the temporal and spatial changes of handwriting, especially the signals. For example, the hand activity signals collected by a pressure pen that change with time are time series. The emerging models such as long short-term memory (LSTM) and gated recurrent unit (GRU) can successfully capture these microchanges and better describe and interpret the handwriting data of a single modality [20–23].

Several studies focus on the automated hand motion analysis using the handwriting data collected by multisensor; various dynamic attributes of handwriting, such as pen pressure, stroke speed, and in-air time, have been analyzed and evaluated for PD detection [15, 20, 24]. Previous studies have reported that using one of these data resources results in promising detection and diagnosis of PD by handwriting recognition. For instance, using handwriting data in a time series, the authors in [20] investigated the distinction between healthy controls and PD patients via one-dimensional convolutions and BiGRUs. However, handwriting recognition requires rich, fine-grained features because the differences between multiple handwriting patterns are usually much more subtle than those between common action categories. In order to capture subtle features without losing other key biomarkers, we chose multimodal handwritten data to describe multimodal hand motion features.

1.3. Research Motivation. A novel Spatio-temporal Siamese neural network (ST-SiamNN) is designed here to discover relevant and decisive attributes of signals or images of handwriting, which can integrate spatial and temporal descriptors, enhance the unique feature of the original data, and amplify the difference between healthy people and PD patients. The motivation for designing this structure is to match the input data in pairs and to increase the difference between the dissimilar data by comparing the distance between the output vectors of the Siamese network. In order to ensure the accuracy of classification, we add a similarity label to mark whether the paired data are similar or not, which is another attribute of the input data.

The ST-SiamNN is a discriminative model constructed by two Siamese networks, which consists of four modules: a preprocessor for handwritten data enhancement, a Siamese bidirectional memory neural network (SiamBiMNN) for temporal feature extraction and difference enhancement, a Siamese octave convolutional neural network (SiamOctCNN) for spatial feature extraction and difference enhancement, and a tree-type decision-making layer for feature classification. The curvature and signal-to-noise ratio (SNR) of the data are calculated to distinguish the handwriting of participants. Weight is added to the data with a large curvature and a small signal-to-noise ratio to improve the significance of the feature so that it can be trained by the Siamese network. The proposed recurrent cell in the SiamBiMNN can be able to capture the texture information of handwritings, which is a supplement to the original gated recurrent unit.

The three groups of data used in this article are shown in Figure 1, i.e., images, smart pen signals, and graphics tablet signals. It can be distinctly seen that the handwriting image of PD patients is disordered and unsmoothed, and the fluctuation range of handwritten signals in PD patients is also quite different from that of healthy people, e.g., the amplitude of finger grip strength in PD patients is significantly greater than that in healthy people. In order to achieve more precise recognition results, the handwriting data with diverse types need to be jointly exploited rather than individually handled. The motivation behind the proposed multimodal data analysis framework is that it plays a crucial role in processing multimodal data and generating joint predictions for handwriting recognition.

Supplementarily, there is a major challenge in the process of recognizing the handwriting of patients with PD. When the hand movement of patients with mild Parkinson's disease is only slight, which is very close to the handwriting of healthy subjects, the model will be misdiagnosed and affect the overall recognition rate. Therefore, it is very significant to enhance the difference between healthy subjects and PD patients and preprocess the data before training.

1.4. Main Contributions. The main contributions of our work are summarized as follows:

(i) Two Siamese neural networks, i.e., a Siamese bidirectional memory neural network (SiamBiMNN) and a Siamese octave convolutional neural network (SiamOctCNN), are proposed for Spatio-temporal and texture feature extraction, which can use the similarity label and the classification label to describe the input handwritings of healthy individuals and PD patients and enhance the feature saliency via two loss functions.

(ii) We design a data preprocessor for image binarization and weight calculation, including binarization, curvature, and SNR calculation of the original handwritten data. Binarization is to remove other interference factors from the handwritings, and the weight calculation of curvature and SNR is to enhance the feature saliency of handwritings.

(iii) We set up a tree-type classification classifier in the decision-making layer to evaluate the output features of the Siamese network and output accurate diagnosis results. The proposed hybrid model can enhance the significance of features as well as extract critical Spatio-temporal and texture information from handwritings for PD detection.

The rest of this article is organized as follows: Section 2 reviews the related work in recent years. Section 3 introduces

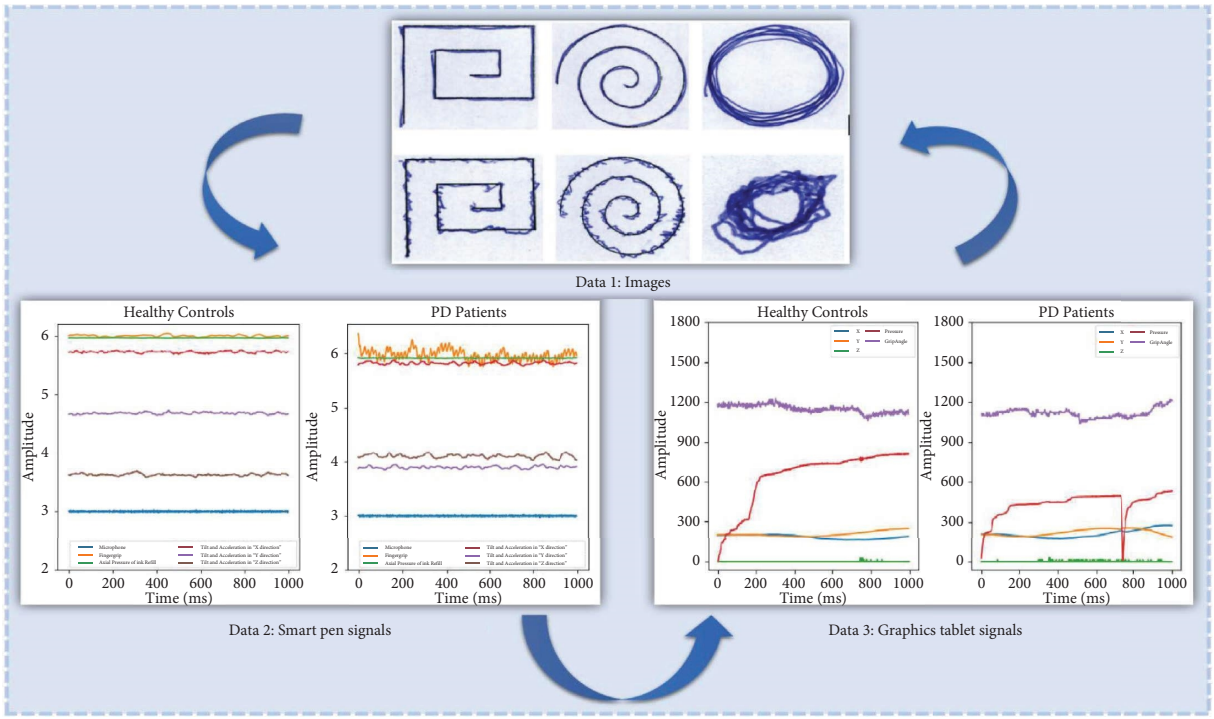


Figure 1: The multimodal handwriting data applied in this project: we evaluate the proposed network over three data types, i.e., images, smart pen signals, and graphics tablet signals, and we fuse the three modalities of the handwriting data for PD detection.

the internal structure and parameters of the proposed ST-SiamNN model. Section 4 provides the evaluation results of the system. The conclusion discusses the final results and the future work in Section 5.

2. Related Work

In this section, handwriting recognition methods in recent five years have been explored, including some of the latest classification methods, deep neural networks, and fusion algorithms.

To the best of our knowledge, biomarkers of Parkinson's disease can be analyzed through various forms of human-machine interaction, including precise grip strength, finger tapping test (FTT), and handwriting. Several traditional machine learning approaches [25–27] can be used to distinguish the individual writing patterns between different handwritings. For instance, support vector machines (SVMs), k-nearest neighbor (KNN), and neural networks have been evaluated to be effective on handwritten data from paper documents, photographs, touch screens, and other devices [25]. Apart from the independent method, the fusion of random forest (RF) and histogram of oriented gradient (HOG) was proposed to extract the impactful information of the image and reduce the input dimension [26]. Similarly, the CNN and SVM were also combined and utilized in the process of handwritten numeral recognition [27].

However, these studies only analyze the global handwriting differences at a macro level and ignore the local subtle changes in the assessment of hand movement disorders, such as the frequency of hand tremors, the force changes during hand movement, and the minor deviations in handwriting testing. In this article, the output of different forms of the handwriting of PD patients will be comprehensively considered to identify and evaluate the disease status. Due to the fine-grained feature learning and dual label contrast training of the fusion method, these subtle differences can be easily captured and quantified, which can be used as a reliable basis for the final distinction between PD patients and healthy controls.

On the other hand, with the rapid development of deep learning-oriented handwriting recognition algorithms, it has become the mainstream trend in the academia and industry [28, 29]. The handwriting test has multiple types and outputs, e.g., images and pressure signals. The convolutional neural network (CNN) model is often used to extract spatial information from images [30, 31]. Gazda et al. presented an end-to-end CNN-based method for PD diagnosis on handwriting images [30]. The ALexNet with the CNN architecture was utilized to refine the diagnosis of PD [31]. A new, improved CNN was proposed for handwritten character recognition, including a batch normalization layer and a residual network structure [28]. The handwriting signals collected by the pressure pen have continuity and contain temporal and spatial features. Using time series data, the long short-term memory (LSTM) can obtain the micro-motion information of PD patients' hands by using continuous and variable time-series data so as to improve the detection sensitivity and objectively analyze the complexity and diversity of motion behavior. Voigtlaender et al. proposed a multidimensional recurrent neural network (MDRNN) to process videos (3D) and images (2D) for handwriting recognition tasks [32].

Moreover, several hybrid models also consider the advantages of the CNN and LSTM, which contribute to handwriting recognition. The CNN and the bidirectional LSTM (BiLSTM) were cooperatively utilized to process handwriting data for PD detection [15]. The combination of one-dimensional convolutions and the bidirectional gated recurrent units (BiGRUs) was proposed to learn the potential information and pattern of handwriting in identifying Parkinsonian symptoms [20]. Learning invariant feature representations of handwritings, a CNN- and RNN-based method to incorporate pixel-level rectification was presented for handwriting recognition [29]. Even though these methods have made up for the shortcomings of traditional machine learning by integrating different deep models based on big data, they still lack the ability to track and analyze the texture of handwritten data and detect slight abnormalities in handwritten data of patients with mild Parkinson's disease. We compared two misclassified samples: one from healthy people and one from patients with mild Parkinson's disease. The results are shown in Figure 2. We can see that the handwriting of PD patients is similar to that of healthy subjects, and the degree of curvature of the handwriting curve is also similar. The handwriting of PD patients partially deviates from the standard test line. This situation of similar samples increases the difficulty of identification.

Inspired by the advantages of existing advanced methods, this article reports a Spatio-temporal Siamese neural network to learn the pattern of different handwritings. The experimental results show that the proposed method is superior to the existing state-of-the-art machine learning approaches.

3. Spatio-Temporal Siamese Neural Network

For achieving successful handwriting recognition of PD patients and healthy controls in the real environment, we here propose a hybrid multilevel network, a Spatio-temporal Siamese neural network (ST-SiamNN), for image enhancement, feature extraction, and classification. Firstly, we introduce the preprocessor for image enhancement. And then, we describe the architecture of the SiamBiMNN and the SiamOctCNN for interclass similarity minimization and Spatio-temporal feature extraction, followed by detailed discussion on the individual components. Finally, we introduce the tree-typedecision-making layer for distinguishing and scoring the fused Spatio-temporal features before supplying the final diagnostic result.

3.1. The Overall Structure. Our proposed hybrid model is shown in Figure 3. In this framework, the handwriting data are fed into the preprocessor first, which consists of binarization, signal-to-noise ratio (SNR) calculation, and curvature calculation to handle the input for enhanced data generation. Afterward, the binarized image or the signal is multiplied by the reciprocal of the SNR to highlight the noise. The calculated curvature map is added to form fusion data and input to the feature extractor for training. For the feature extraction module, a Siamese bidirectional memory neural network (SiamBiMNN) is utilized for analyzing the temporal dynamic changes and adding the texture information of the handwritten data, while a three-layer Siamese octave convolutional neural network (SiamOctCNN) is designed to consider the spatial features in the high- and low-frequency ranges of the input data. The twin network uses two loss functions and two types of labels (a similarity label and a classification label) to distinguish the handwriting of normal people and PD patients. After obtaining the features from the SiamBiMNN and SiamOctCNN, we create a tree-type decision-making layer to describe multiple classification outcomes. As illustrated in Figure 3, the Siamese-based network is a discriminant framework with detection outputs to train the handwriting recognition network.

First, we formulate the problem for handwriting recognition. The image sequences are defined as $X = \{x_i \in \mathbb{R}^{W*H}, i = 1, 2, 3, \ldots, N\}$ with corresponding 2-class label sequences. L and N are the sample numbers of the input data, and W and H are the width and height of each frame of the image. For the signal data, it will be divided into several training samples; each sample contains the signal value in a period of time, where W is the characteristic number of each frame signal, H is the length of time series, and $W * H$ is the dimension of a training sample.

3.2. Data Enhancement. As shown in Figure 4. All the handwritten images are binarized first, the noise is removed, and the image feature is enhanced. The curvature is a measure of how much a point on a curve bends. The more curved curve has the larger curvature. For two-dimensional discrete digital images, the mean curvature (Figure 5) is calculated using the following formula:

$$I' = \frac{\left(1 + I_x^2\right) I_{yy} - 2 I_x I_y I_{xy} + \left(1 + I_y^2\right) I_{xx}}{2\left(1 + I_x^2 + I_y^2\right)^{(3/2)}}, \tag{1}$$

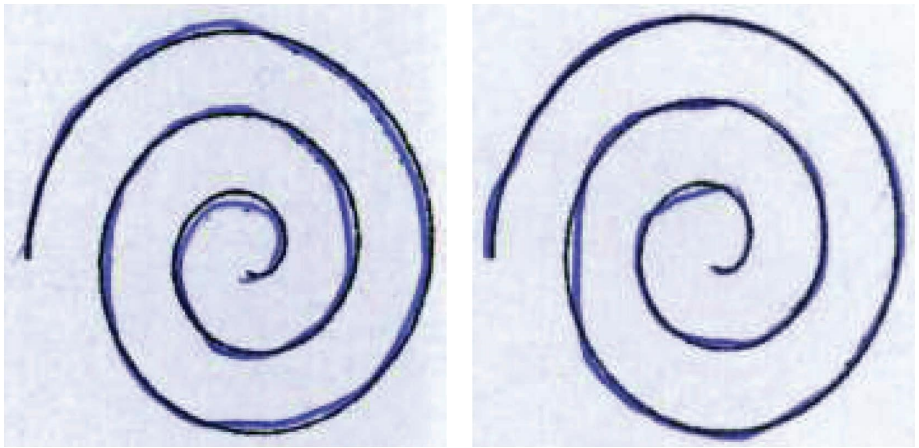
where I is the input image, x and y are the coordinates of the image, and I_x and I_y are the component sets of the input image in two dimensions. Generally, the average curvature is calculated by discretizing the formula. Another method is to avoid discretization by quadric surface fitting:

$$I'(x, y) \approx f(x, y) = C_5 x^2 + C_4 y^2 + C_3 xy + C_2 x + C_1 y + C_0. \tag{2}$$

Then, the coefficient C_i is determined by the least square method. After that, we substitute C_i into the above formula to get the final processed image I', as shown in the following equation:

$$I' \approx \frac{\left(1 + C_2^2\right) C_4 - C_2 C_1 C_3 + \left(1 + C_1^2\right) C_5}{\left(1 + C_1^2 + C_2^2\right)^{(3/2)}}. \tag{3}$$

For data enhancement, we also calculate the signal-to-noise ratio (SNR) of the image to highlight the more disordered signal in PD patients. The SNR refers to the ratio of the original test-question image and the answer result image



(a) (b)

Figure 2: Comparison of handwriting of PD patients and healthy subjects. (a) PD sample (mild). (b) Non-PD sample.

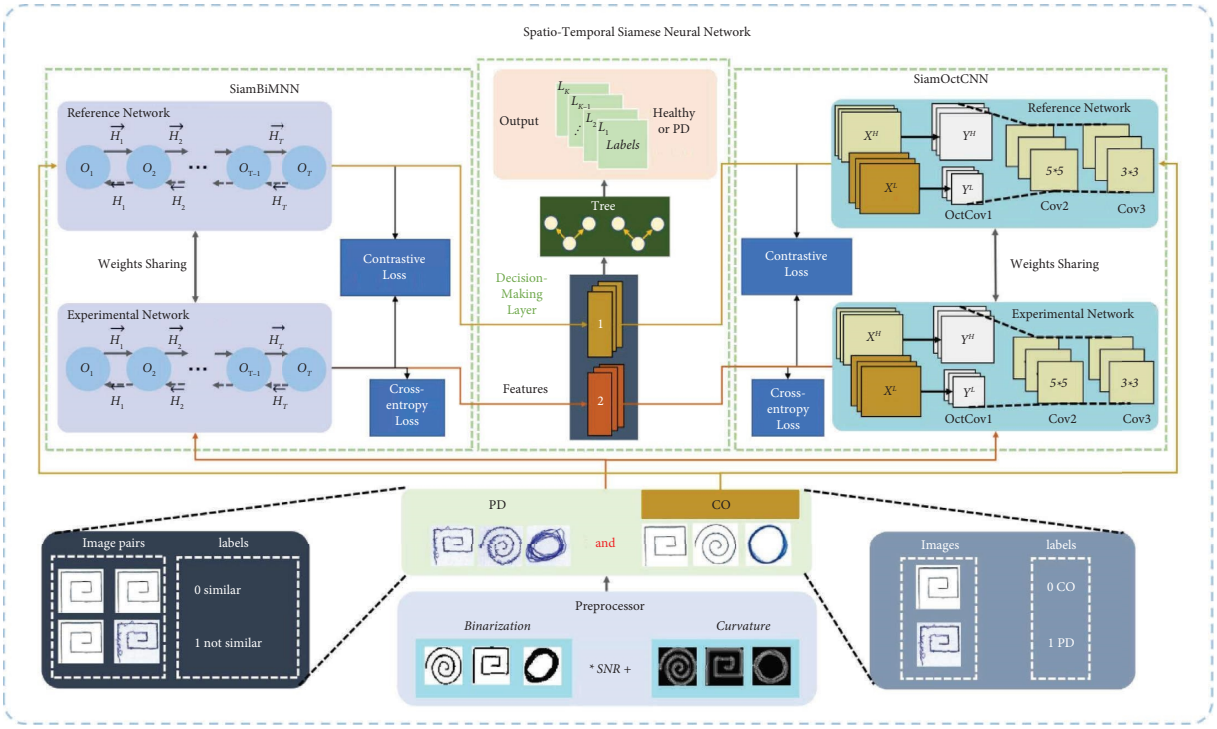


Figure 3: The structure of the ST-SiamNN. There are four modules in this model: preprocessor for image enhancement, SiamBiMNN for temporal feature extraction, SiamOctCNN for spatial feature extraction, and decision-making for classification.

(handwriting image). The smaller SNR means the greater noise, illustrating that the testing signal has a greater probability of coming from PD patients. The calculation process of the SNR is as follows:

$$\mathrm{SNR} = 10 * \log_{10}\left[\frac{\sum_{x=1}^{W}\sum_{y=1}^{W}(f(x,y))^2}{\sum_{x=1}^{W}\sum_{y=1}^{W}(g(x,y))^2}\right], \tag{4}$$

where H and W are the number of pixels on the height and width of the image, respectively, and $f(x, y)$ and $g(x, y)$ are the pixel values of the original image and the noise image at the points (x, y), respectively. Afterward, we multiply the binary image by the SNR and add the curvature image to get the final fusion image as the input of the Siamese network.

3.3. SiamBiMNN. The Siamese neural network is a coupled framework, which consists of two neural networks with the same structure and weights. It takes two samples as input and outputs a representation embedded in high-dimensional space to compare the similarity of the two samples. Each neural network usually has a deep structure, which can be composed of convolutional neural networks, recurrent

Figure 4: The handwriting data after binarization.

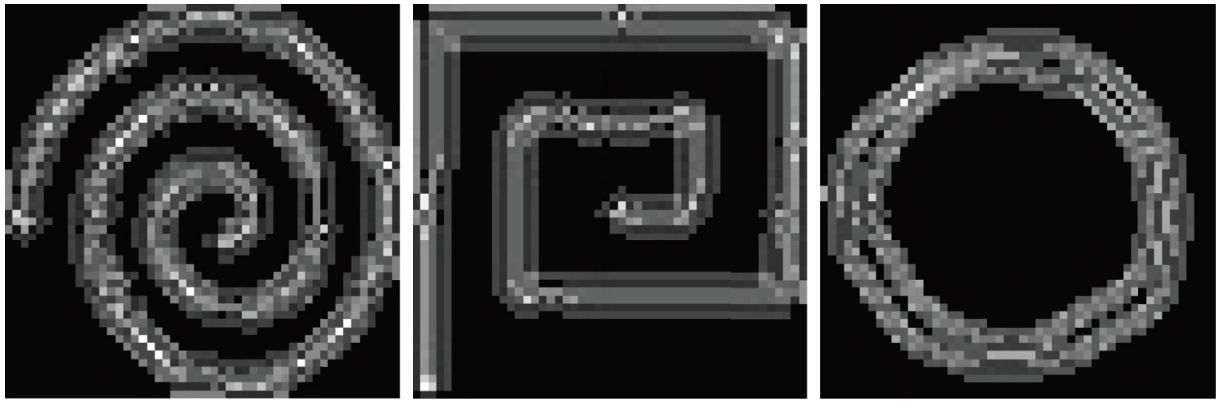


Figure 5: The handwriting data after curvature calculation.

neural networks, and so on. Therefore, we design two different Siamese networks and add similarity labels on the basis of retaining the original labels to compare the similarity of data features to complete the task of handwriting recognition.

We start from the SiamBiMNN for framework details description.

To model nonlinear dynamic processes, long-short term memory (LSTM) is widely used to describe the temporal dynamic behavior of time series with scalable memory cells and three nonlinear gates (input, forget, and output) [33]. Although LSTM networks are efficient at learning long-term temporal dependencies, they cannot be sensitive to subtle texture differences when processing data in a single sample spatial domain. For $50 * 50$ size of the handwritten images, we chose the simplest GRU-like gating mechanism, which is suitable for handling short sequences.

The overall framework of the SiamBiMNN is illustrated in Figure 6. Two groups of data (one group is only healthy controls, and the other group includes all the data) are input into two networks. The first network is the reference network, and only the contrastive loss is used to train the similarity of CO vs. CO and CO vs. PD. The second network is the experimental network. In addition to training the similarity of CO vs. CO and CO vs. PD with contrastive loss, the cross-entropy loss is also used to complete the task of feature extraction. Finally, we choose the output features of the experimental network as the output of the decision-making layer.

The two groups of the data reach the two networks with T expanded nodes in the SiamBiMNN. As the forward MNN and the backward MNN are combined to form the BiMNN, the BiMNN is composed of the forward MNN and the backward MNN, which can consider the feature inputs of $t + 1$ and $t - 1$ at the same time. Because the output of the forward and backward MNNs contains the context information, we take the final output O_{2T} as a part of the fusion feature. The outstanding point is that the MNN expanded unit can extract temporal features, as well as analyze texture information, which makes the expanded unit have more abundant and enhanced output.

3.3.1. The SiamBiMNN Cell. Before introducing the novel memory cell, we will briefly review the internal structure of the GRU [34]. To model nonlinear dynamic processes, GRUs are widely used to describe the dynamic behavior of time series. The scalable memory cell of the GRU consists of two nonlinear gates, i.e., the update gate and the output gate. It can effectively learn long-term dependence because memory cells in GRU can maintain their state for a long time and regulate the incoming and outgoing information flow.

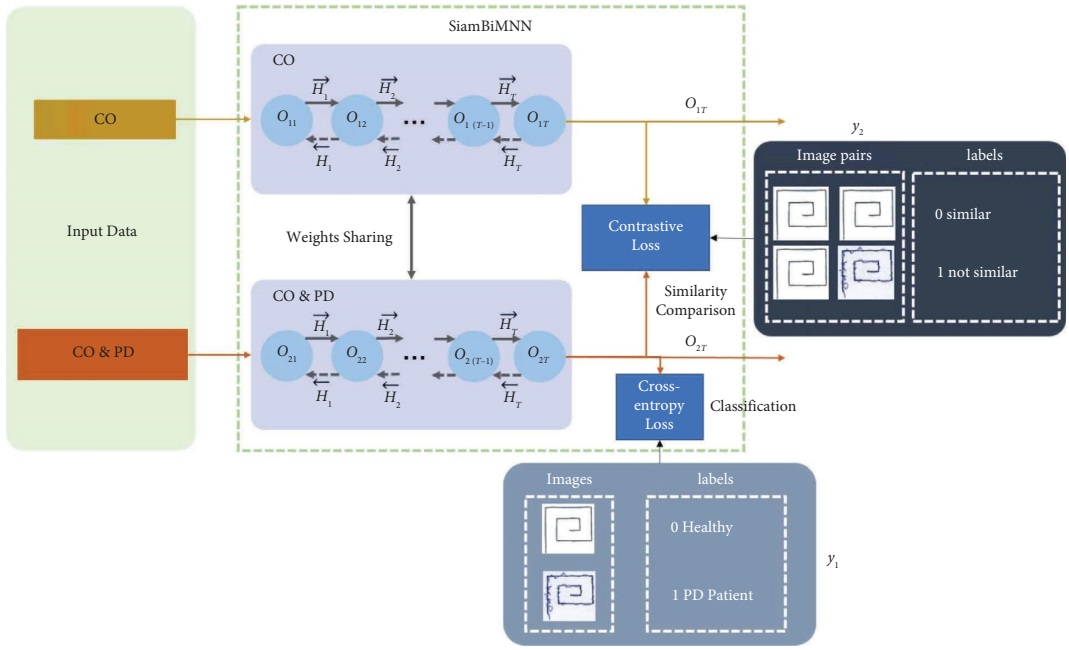


Figure 6: The structure of the SiamBiMNN. The input layer of the SiamBiMNN is divided into two entrances, corresponding to two bidirectional memory neural networks in the feature extraction layer. By inputting the wrapped handwriting data (CO, PD, and CO), the SiamBiMNN improves the significance of the similarity of the output features of the two networks by reducing the contrastive loss and improves the difference of the nonsimilar samples by reducing the cross-entropy loss.

Unlike the original GRU cell, we build a path (ctemp) before the output O_t so as to extract more discriminative features. We witness that the new path in the multigated memory cell can capture the state of the previous node and the input of the current node at the same time. ctemp restricts these two states to $[-1, 1]$, which highlights the difference in feature changes.

Figure 7 illustrates the internal structure of the proposed memory unit and the operations of all the gates. The outputs and inputs in each unit are demonstrated in the following equation:

$$\text{Equation Group}\begin{cases} z_t = \sigma(W_z \cdot [O_{t-1}, x_t]), \\ r_t = \sigma(W_r \cdot [O_{t-1}, x_t]), \\ \tilde{h}_t = \tanh(W \cdot [r_t \odot O_{t-1}, x_t]), \\ \text{ctemp} = \tanh\left(W_{\text{ctemp}} \cdot [O_{t-1}, x_t]\right), \\ c_t = (1 - z_t) \otimes \tilde{h}_t + z_t \otimes O_{t-1}, \\ O_t = c_t \otimes \sigma(\text{ctemp}), \\ O'_t = g\left(W\left[\overrightarrow{O_t}, \overleftarrow{O_t}\right] + b\right), \end{cases} \tag{5}$$

where W_z, W_r, W, and W_{ctemp}, are the weight parameters that need to be learned in the training process. x_t and O_t are the current input and output of the hidden unit at time $t \in \{1, 2, 3 \ldots, T\}$. σ and tanh are the activation functions. z_t and r_t indicate the output of the update gate and the reset gate at time t, which determine whether or not the previous hidden state should be ignored and updated. ctemp represents the temporary state in order to determine x_t and

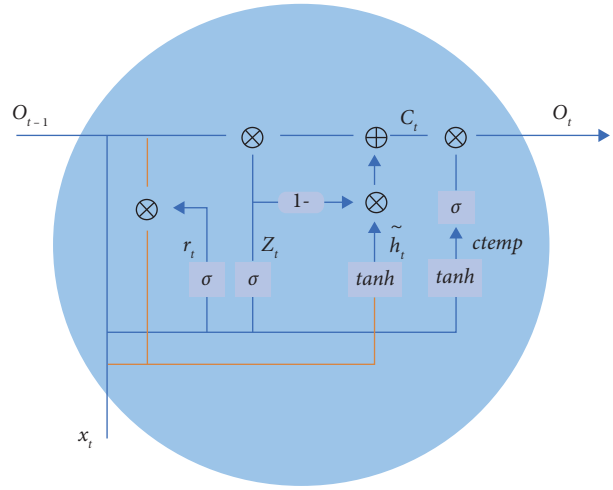


Figure 7: The memory cell in the SiamBiMNN.

O_{t-1}, while c_t denotes the final state of the SiamBiMNN. We then extract features from c_t. O'_t is the final output of the SiamBiMNN, including the output $(\overrightarrow{O_t}, \overleftarrow{O_t})$ of two unidirectional models.

After obtaining the temporal features, we combine them with the output of the preprocessor to get the fusion vectors of three features and then classify them by the softmax classifier to obtain the detection results.

3.4. SiamOctCNN. The structure of the SiamOctCNN is illustrated in Figure 8, which includes one octave convolution layer and two ordinary convolution layers.

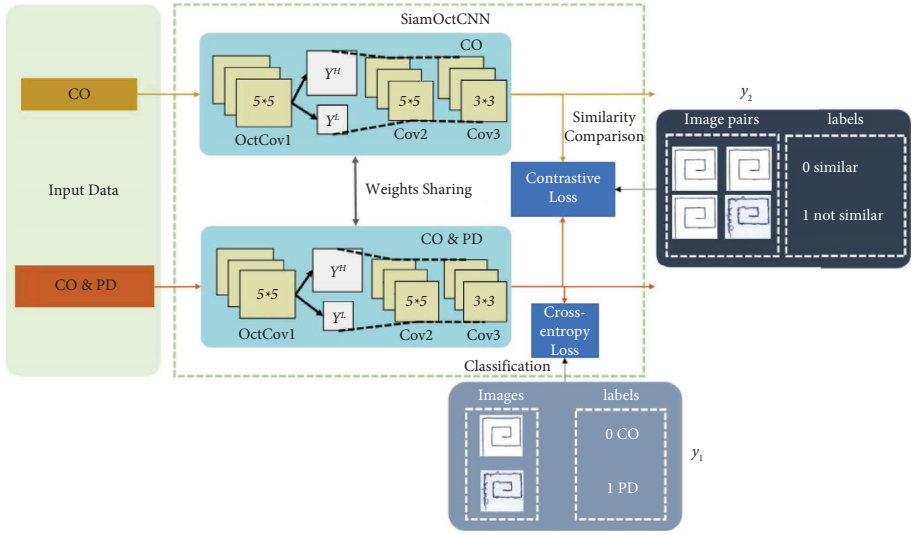


FIGURE 8: The structure of the SiamOctCNN. The input layer of the SiamOctCNN is divided into two entrances, corresponding to two octave convolutional neural networks in the feature extraction layer. Each convolution network in the SiamOctCNN consists of three layers: one octave convolution layer and two ordinary convolution layers. The output features include high- and low-frequency spatial features of handwriting data.

3.4.1. Octave Convolutional Layer. Unlike the convolution operation, octave convolution (Octconv) [35] is used to store and process feature maps with low spatial resolution and slow spatial change so as to reduce the cost of memory and computation. Different from the existing multiscale methods, Octconv is a single, general-purpose, plug-and--play convolution unit that can directly replace ordinary convolution without any adjustment to the network structure. It also proposes orthogonal and complementary methods for better topology or reducing redundant image groups or deep convolution channels. For processing multisensor data, the convolution feature map is decomposed into two sets of different spatial frequencies: fine detail coding is usually used at a higher frequency, and global structure coding is usually used at a lower frequency.

Equation (6) shows the calculation process of the Octconv unit. The input of Octconv consists of high-frequency feature X^H and low-frequency feature X^L, with the two outputs $Y = \{Y^H, Y^L\}$. The high-frequency output Y^H is the sum of the feature $Y^{H\longrightarrow H}$ obtained by convolution of the high-frequency input feature itself and a mutual feature $Y^{L\longrightarrow H}$ of the low frequency. The high- and low- frequency output features are derived as follows:

$$\begin{cases} Y^H = f\left(X^H; W^{H\longrightarrow H}\right) + \text{upsample}\left(f\left(X^L; W^{L\longrightarrow H}\right), 2\right), \\ Y^L = f\left(X^L; W^{L\longrightarrow L}\right) + f\left(\text{pool}\left(X^H, 2\right); W^{H\longrightarrow L}\right), \end{cases} \tag{6}$$

where $f(X^H, W)$ is a convolution operation with parameters W, $\text{pool}(X^H, k)$ is an average pooling operation with kernel size $k \times k$ and stride k, and $\text{upsample}(X^H, k)$ is an upsampling operation by a factor of k via nearest interpolation.

Although octave convolution has two inputs, we cannot manually distinguish the frequency of the input data, so the first input data is all high-frequency information by default, the middle convolution layer outputs the feature map, which contains low- and high-frequency information, and the last layer convolution restores the normal feature map. The output of the Octconv is demonstrated in Figure 9, and we use the Fast Fourier Transform to represent the spectral low- and high-frequency features. High-frequency features represent places where gray-scale changes quickly, and low-frequency features represent places where gray-scale changes slowly.

3.5. Loss Function. In the Siamese network, the loss functions of cooperative training are used to reduce two losses: similarity loss (contrastive loss) and classification loss (cross-entropy loss), which will be described in detail below.

3.5.1. Contrastive Loss. In the Siamese network, the loss function used is a contrastive loss, which can effectively handle the relationship between paired data in the Siamese neural network. In our designed model, images are sent to the Siamese network in pairs, with similar pairs marked as negative pairs and dissimilar pairs marked as positive pairs. By minimizing the contrast loss between the reference network and the experimental network, the difference between dissimilar pairs is widened, thereby achieving the effect of expanding the distance between classes.

The loss function is mainly used for dimensionality reduction, that is, after dimensionality reduction (feature extraction), the two samples are still similar in the feature space; however, the original samples after dimensionality reduction are still different in the feature space. Likewise, the loss function can be a good representation of how well a pair of samples match.

This loss function can effectively handle the relationship between paired data in a dual neural network. The comparative loss is expressed as follows:

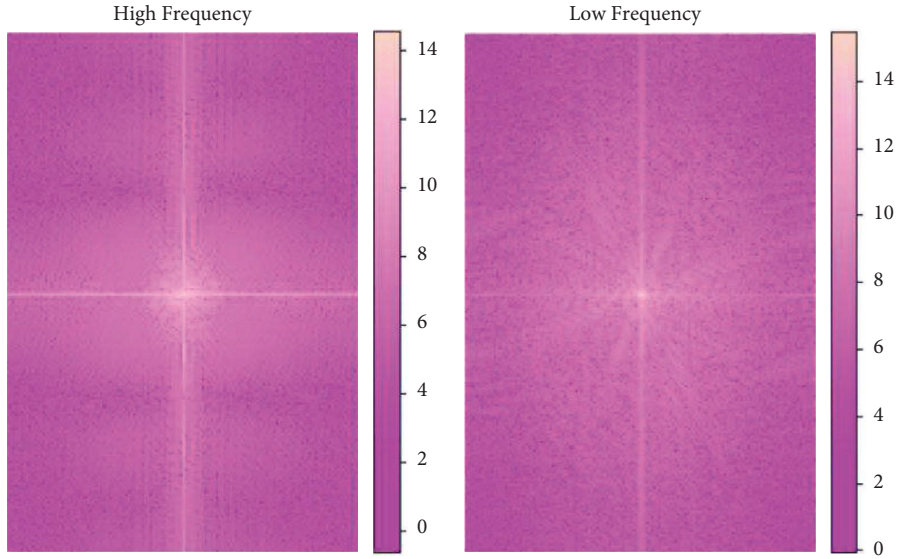


Figure 9: The (a) high- and (b) low-frequency features of the handwriting data.

$$L_1(W,(Y,X_1,X_2)) = \frac{1}{2N}\sum_{n=1}^{N} YD_W^2 + (1-Y)\max(m-D_W,0)^2, \tag{7}$$

where $D_W = \|X_1 - X_2\|_2 = (\sum_i^N (X_1^i - X_2^i)^2)^{(1/2)}$. It represents the Euclidean distance of two sample features X_1, X_1 and X_2, X_2; Y is the label of whether the two samples match; $Y = 1$ represents that the two samples are similar or match; $Y = 0$ represents that they do not match; and m is the set threshold.

By observing the expression of comparative loss, we can find that this loss function can well express the matching degree of paired samples and can also be used to train a feature extraction model.

When $Y = 1$ (i.e., the samples are similar), the loss function is only $L_S = 1/2N\sum_{n=1}^{N} YD_W^2$. If the Euclidean distance of the original similar samples in the feature space is large, the current model is not good, so the loss is increased.

When $Y = 0$ (i.e., the samples are not similar), the loss function is $L_D = (1-Y)\max(m-D_W,0)^2$; that is, when the samples are not similar, the Euclidean distance of the feature space is smaller and the loss value will become larger.

Setting a threshold margin means that we only consider the dissimilar features whose Euclidean distance is between 0 and margin. When the distance exceeds margin, the loss is regarded as 0; however, for similar features that are close to each other, we need to increase their loss so as to constantly update the matching degree of paired samples.

3.5.2. Cross-Entropy Loss. Cross-entropy describes the distance between two probability distributions. The closer the cross-entropy is, the closer they are. While cross-entropy describes the distance between two probability distributions, the output of a neural network is not necessarily a probability distribution. Therefore, softmax regression converts the result of forwarding propagation of the neural network into a probability distribution. Softmax is often used in multiclassification processes. It normalizes the outputs of multiple neurons to the (0, 1) interval, enabling multiclassification. The calculation process of cross-entropy loss is as follows:

$$L_2 = -\frac{1}{N}\sum_{i=1}^{N}\left[y^{(i)}\log\hat{y}^{(i)} + \left(1-y^{(i)}\right)\log\left(1-\hat{y}^{(i)}\right)\right], \tag{8}$$

where L_2 represents the classification loss of the handwriting data. During optimization, we use the Adam optimizer to process the first and second moments of the gradient to quickly reduce the loss. $y^{(i)}$ denotes the i^{th} true label of a training batch in the temporal data, while $\hat{y}^{(i)}$ represents the i^{th} predicted label.

3.6. Decision-Making Layer. The gradient boosting decision tree (GBDT) generates weak classifiers through multiple iterations and trains each classifier based on the residuals of the previous classifier. The training process of the GBDT is shown in Figure 10. The requirements for weak classifiers are usually simple enough, low variance, and high bias because the training process is intended to continuously improve the accuracy of the final classifier by reducing the bias.

Generally, the cart tree is chosen as the weak classifier. Due to the high bias and requirements mentioned above, the depth of each classification regression tree will not be very deep. The final overall classifier is a weighted sum of weak classifiers from each round of training. The model can be described as follows:

$$F_m(x) = \sum_{m=1}^{M} T(x;\theta_m). \tag{9}$$

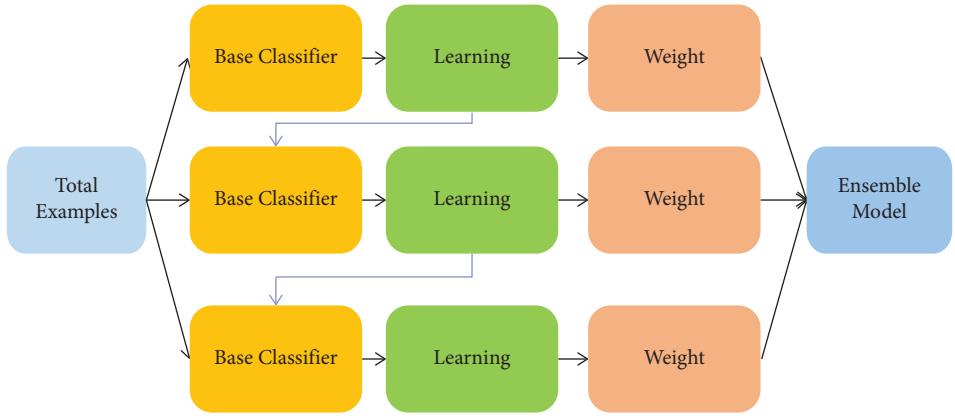


Figure 10: The training process of the GBDT.

The model trains m rounds in total, and each round produces a weak classifier $T(x;\theta_m)$. The loss function of weak classifier is as follows:

$$\begin{aligned}\widehat{\theta} &= \underset{\theta_m}{\operatorname{argmin}} \\ &= \sum_{i=1}^{N} L\left(y_i, F_{m-1}\left(x_i\right) + T\left(x_i;\theta_m\right)\right).\end{aligned} \tag{10}$$

4. Experiments

This section presents our experimental settings and the performance of the proposed ST-SiamNN, compared against several state-of-the-art methods on two challenging handwriting datasets.

The state-of-the-art technologies compared in this article are as follows:

(i) LSTM (long short-term memory) is a scalable model suitable for temporal data

(ii) BiLSTM (bidirectional long short-term memory) is a combination of forward and backward LSTMs

(iii) GRU (gate recurrent unit) is a lightweight variant of LSTM with fewer variables

(iv) CNN (convolutional neural network) is a feed-forward neural network with a deep structure and convolutional computation

(v) CRNN (convolutional recurrent neural network) [36] combines CNN and RNN networks, jointly trained to achieve end-to-end handwriting detection and recognition

(vi) TS-LSTM (temporal sliding LSTM) [37] contains short-term, medium-term, and long-term TS-LSTM networks

(vii) HiGRU (Hierarchical Gated Recurrent Unit) [38] has a lower-level GRU to model word-level input and a higher-level GRU to capture the context of utterance-level embeddings

(viii) EfficientNet [39] proposes a network scaling method that creatively uses composite coefficients to change network dimensions in network reconstruction, including network width, network depth, and image resolution

(ix) Vision Transformer (ViT) [40] and MobileViT [41] are lightweight universal visual transformers

(x) ConvNeXt [42] builds a pure convolutional network that outperforms the advanced transformer-based models

(xi) MobileNetV2 [43] has a deep separable convolution and adds linear bottleneck and inverted residual

4.1. Dataset Specifications. In this section, we give a brief description of the NewHandPD dataset [44] and the PARKINSON_HW dataset [45, 46] used in our experiment, which is shown in Figure 1.

4.1.1. NewHandPD Dataset. The NewHandPD dataset includes images acquired from two groups of individuals (i.e., the healthy group and the PD patient group) during handwritten exams, which aim at describing the individual skill when filling a form. The handwriting data were collected at Botucatu Medical School, São Paulo State University, from Brazil, and were intended to ask a person to perform some specific tasks that were supposed to be nontrivial to PD patients, i.e., drawing "spirals," "meanders," and "circles" (Figure 1).

This dataset was composed of 66 individuals (35 healthy controls and 31 PD patients). Each individual was asked to complete 12 exams, including 4 spirals, 4 meanders, and 2 circled movements (one circle in the air and another on the paper), and left and right-handed diadochokinesis. Totally, there were 9 images for each individual after the exam.

In addition to images, the dataset also included part of the signal data. The subjects also performed the so-called diadochokinese test, which was basically a test where the subjects held the pen with straight arms and performed hand-wrist movements. Since there were no drawings involved, only the signal generated though these movements was recorded by the pen. The signals were extracted from the

BiSP® smart pen (Figure 11), concerning four sensors and six features, microphone, fingergrip, axial pressure of ink refill, tilt and acceleration in "X direction," tilt and acceleration in "Y direction," and tilt and acceleration in "Z direction."

4.1.2. PARKINSON_HW Dataset. The PARKINSON_HW dataset consists of 62 PD patients and 15 healthy controls who attended the Department of Neurology in Cerrahpasa Faculty of Medicine at Istanbul University.

For all participants, three handwriting tests, i.e., the static spiral test, the dynamic spiral test, and the stability test on certain point, were implemented by using a Wacom Cintiq 12WX graphics table (Figure 12) (a graphics tablet integrated with an LCD monitor that displayed the PC's screen on the monitor and can only interact with a digitizing pen). This is a specially designed software for recording handwriting patterns and testing coordination of PD patients.

The static spiral test (SST) was frequently used for clinical research. In this test, three Archimedes spirals were displayed on a drawing board using the software, and patients were asked to use a digital pen to trace as many of the same spirals as possible. During testing, the aforementioned characteristics and other data used to designate patients were recorded in the dataset.

Unlike the SST, in the dynamic spiral test (DST), the Archimedes spiral simply appeared and disappeared (blinked) at certain intervals, forcing the patient to memorize the pattern and keep drawing. The purpose of this test was to determine changes in the patient's drawing performance and pause time, as it was difficult to track the Archimedes spiral in this condition. As a result of this test, it was observed that most of the Parkinson's patients continued to draw, but nearly all lost their patterns.

The purpose of the stability test on certain point (STCP) was to determine the stability of the patient's hand or the degree of hand tremor. There was a red dot in the middle of the screen, and the subject was asked to hold the digital pen on the dot without touching the screen.

4.2. Experimental Settings. The experiment was implemented on two handwriting datasets, and appropriate settings were arranged according to the features of each dataset. The device had a graphics card of GeForce RTX 2080, a memory of 31.1 GiB, and a CPU of Intel Xeon(R) W-2133. The settings were described in accordance with the dataset.

For the two sleep datasets, we shuffled and randomly selected 80% of the data for training and 20% for testing, with a data capacity of 1329 to 146976.80% of the data were for training in the experimental network and compared with all the CO data in the reference network. The remaining 20% of the data were used to compare with all CO data and were tested. The final testing time on each dataset was approximately 105 ms (NewHandPD-Image), 452 ms (NewHandPD-Signal), and 129 ms (PARKINSON_HW).

The computational cost (time complexity) of the model directly determines the forward time of the model and the training/prediction time of the model. If the complexity is too high, it can lead to a large amount of time spent on model training and prediction, and it is neither possible to quickly validate ideas and improve models nor to achieve rapid prediction. Time complexity refers to the number of operations of the model. The proposed model is divided into four modules to analyze the computational cost. For the preprocessor, according to equations (1)–(4), we can calculate the time complexity as $O(n^3 + n^2 + 2\log n + 1)$. For the SiamBiMNN, in the light of the two recurrent networks contained within the model, according to equation (5), the time complexity is $O(2 * n * d^2)$, n is the sequence length, and d is the representation dimension. For the SiamOctCNN, the time complexity is $O(2 * k * n * d^2)$, and k is the kernel size of the convolutions. For the decision-making layer, the time complexity is $O(n \log n * d * m)$, where n is the number of samples, d is the number of features, and m is the depth of the tree.

The NewHandPD (image) dataset contained 1329 samples. We first converted the image to a binary image and limited each value to (0, 1) for training. The input dimension and time step size in the SiamBiMNN were 50 and 50, which were the reshaped sizes of the original images, with a hidden output of 128 and a learning rate of 0.001. In the

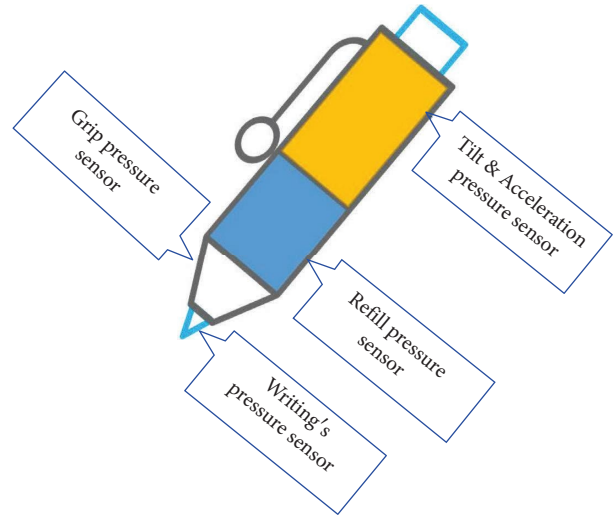


Figure 11: The biometric pen. It includes four sensors: the tilt and acceleration sensor, the refill pressure sensor, the grip pressure sensor, and the writing's pressure sensor.

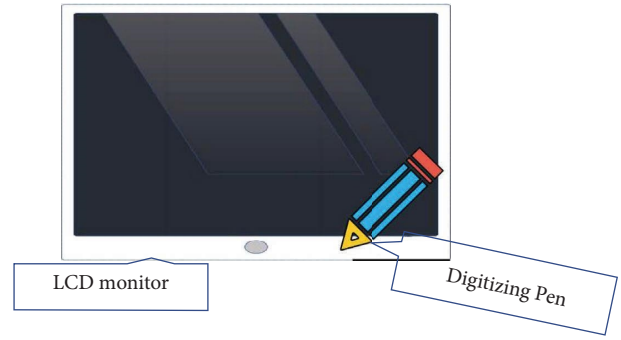


Figure 12: The digitized graphics tablet.

SiamOctCNN, the convolution kernels of the three layers were $5 * 5$ (octave convolution kernel), $5 * 5$, and $3 * 3$. In the decision-making layer, the maximum number of iterations for a weak learning machine was 100.

The NewHandPD (signal) contained 146976 samples. We reshaped the signal matrix to $6 * 100$. In the SiamBiMNN, the signal data were processed by the SiamBiMNN cell with the time step of 10, the input dimension of 60, and the hidden output of 128. In the SiamOctCNN, the size of the octave convolution kernel was $5 * 5$ and the regular convolution kernel was $5 * 5$ and $3 * 3$. The maximum number of iterations for a weak learning machine was 50 in the decision-making layer.

The PARKINSON_HW dataset contained 3208 samples, and the signal matrix was reshaped to $4 * 100$. In the SiamBiMNN, the signal data were processed by the SiamBiMNN cell with the time step of 10, the input dimension of 40, and the hidden output of 128. In the SiamOctCNN, the size of the octave and the regular convolution kernels was the same as that of the above two datasets. The maximum number of iterations for a weak learning machine was 100 in the decision-making layer.

4.3. Handwriting Recognition. The experiment is implemented in two parts: single-modal handwriting recognition and multimodal handwriting recognition. Firstly, the signals and images in datasets NewHandPD and PARKINSON_HW are processed by individual components in the proposed algorithm, and four evaluation methods, i.e., precision, recall, $F1$ score, and accuracy, are used to compare the performance of each module and the whole module. Then, we compared the handwriting recognition results of PD patients and healthy subjects with various deep models and verified the high accuracy of the ST-SiamNN.

In the process of multimodal identification, we input six kinds of bimodal data into the SiamBiMNN and the SiamOctCNN and show the experimental results with four evaluation methods. In addition, these six collocations are spliced and mixed and input into different independent deep frameworks for training, and accurate recognition results are obtained.

4.3.1. Singlemodal Handwriting Recognition. We conducted ablation experiments on single-modal data. The ST-SiamNN is split into four components, i.e., the preprocessor, the SiamBiMNN, the SiamOctCNN, and the decision-making layer, and we selected seven different combinations among them to evaluate the experimental results on the two datasets.

The ablation Study on the NewHandPD dataset: this dataset includes bimodal handwriting data, i.e., image and signal. First of all, in experiment 1, we were faced with 1392 handwritten handwriting images. Each image was reshaped to a size of $50 * 50$ for training in the ST-SiamNN model. We enable different components in the ST-SiamNN to evaluate the performance of each module, and results of combined components on the NewHandPD dataset (image) are shown in Table 1. We use the "✓" to represent the components that are enabled and highlight the greatest performance of the combination.

Among the four evaluation indexes, i.e., precision, recall, $F1$ score, and accuracy, the best result is that all the components are activated. Among the seven combinations, the combination "preprocessor + SiamBiMNN + decision-making layer" ranks the second, and the recognition rate is 91.73%, followed by combination "SiamBiMNN + decision-making layer" and combination "preprocessor + SiamOctCNN + decision-making layer," which shows that the SiamBiMNN has a greater impact on the whole model, followed by the SiamOctCNN. Compared with the results of "preprocessor + decision-making layer," the accuracy of including the SiamNN has been greatly improved, which verifies the effectiveness of the model.

In experiment 2, the signal data corresponding to experiment 1 are input into the model for training, and the effect is slightly lower than the image data. Results of combined components on the NewHandPD dataset (Signal) are shown in Table 2.

For the evaluation methods of recall, $F1$ score, and accuracy, the whole model ST-SiamNN achieves the optimum results, and the components SiamBiMNN and SiamOctCNN almost attain a preferable classification result. For the precision, the combination "preprocessor + SiamBiMNN + decision-making layer" reaches the top, indicating that this combination is more sensitive to the detection of healthy subjects. Compared with the recall value, the whole model shows the highest results, which indicates that the ST-SiamNN is the most accurate for Parkinson's disease detection.

The ablation study on the PARKINSON_HW dataset: in this experiment, we firstly reshape the four-dimensional data of frames into training samples with 100 frames as a group, and each training sample is 400 dimensions. Results of combined components on the PARKINSON_HW dataset are shown in Table 3.

We can see that the results of the combination "preprocessor" and combination "preprocessor + decision-making layer" are still the lowest. After data preprocessing and twin network training, the accuracy is improved by 6%-7%. The combination "preprocessor + SiamOctCNN + decision-making layer" achieves the highest recall value of 85.86%. The whole model and the combination "preprocessor + SiamBiMNN + decision-making layer" obtain approximately 83% of the result. Overall, Siamese networks contribute the most to the proposed model.

The performance of seven different deep models on single-modal datasets is illustrated in Table 4. By comparing the experimental results of the datasets of the deep model, the results of datasets NewHandPD (Image) and NewHandPD (Signal) are much higher than that of the PARKINSON_HW dataset due to data inconsistency. The coordinate data included in the signal data of the PARKINSON_HW dataset does not change significantly, or the handwriting difference between healthy subjects and PD patients is small, which affects the experimental results.

In the dataset NewHandPD (image), although the performance of the CNN is superior to the three time-series

Table 1: Results on the NewHandPD dataset (image).

NewHandPD dataset							
Components				Evaluation			
Preprocessor	SiamBiMNN	SiamOctCNN	Decision-making layer	Precision (%)	Recall (%)	*F*1 score (%)	Acc (%)
			✓	85.19	89.84	87.45	87.59
✓			✓	90.14	87.67	88.89	87.97
✓	✓		✓	89.84	92.74	91.27	91.73
✓		✓	✓	86.96	92.31	89.55	89.47
	✓		✓	90.65	91.30	90.97	90.60
		✓	✓	86.23	91.54	88.81	88.72
✓	✓	✓	✓	**92.81**	**93.48**	**93.14**	**92.86**

The bold values in Table 1 show the highest scores of the compared individual components in the proposed model on the NewHandPD dataset (image).

Table 2: Results on the NewHandPD dataset (signal).

NewHandPD dataset							
Components				Evaluation			
Preprocessor	SiamBiMNN	SiamOctCNN	Decision-making layer	Precision (%)	Recall (%)	*F*1 score (%)	Acc (%)
			✓	86.57	87.07	86.82	83.41
✓			✓	88.31	86.03	87.16	84.13
✓	✓		✓	**91.19**	90.20	90.69	88.48
✓		✓	✓	89.81	88.67	89.24	86.56
	✓		✓	90.62	89.67	90.14	87.76
		✓	✓	87.18	90.94	89.02	85.90
✓	✓	✓	✓	91.17	**94.37**	**92.74**	**90.76**

The bold values in Table 2 show the highest scores of the compared individual components in the proposed model on the NewHandPD dataset (signal).

Table 3: Results on the PARKINSON_HW dataset.

PARKINSON_HW dataset							
Components				Evaluation			
Preprocessor	SiamBiMNN	SiamOctCNN	Decision-making layer	Precision (%)	Recall (%)	*F*1 score (%)	Acc (%)
			✓	79.37	73.71	76.43	71.18
✓			✓	77.84	77.44	77.63	72.90
✓	✓		✓	80.60	83.29	81.92	77.73
✓		✓	✓	77.10	**85.86**	81.24	75.55
	✓		✓	79.70	81.33	80.51	76.01
		✓	✓	79.90	78.50	79.19	74.30
✓	✓	✓	✓	**81.09**	83.38	**82.22**	**78.04**

The bold values in Table 3 show the highest scores of the compared individual components in the proposed model on the PARKINSON_HW dataset.

Table 4: Performance of deep models on the two datasets.

Dataset	LSTM (%)	BiLSTM (%)	CNN (%)	CRNN (%)	TS-LSTM (%)	HiGRU (%)	ST-SiamNN (%)
NewHandPD (image)	85.71	86.84	87.16	86.47	88.35	89.10	**92.86**
NewHandPD (signal)	86.46	87.74	86.39	87.60	87.48	88.31	**90.76**
PARKINSON_HW	73.99	74.45	71.18	72.74	74.61	76.32	**78.04**

It shows the ST-SiamNN obtains the best performance among those mentioned in deep models.

models LSTM, BiLSTM and CRNN, the accuracy of TS-LSTM and HiGRU has been increased due to the consideration of the improvement of an extensible unit of the time series model and the hierarchical structure. In the dataset NewHandPD (Signal), there is little difference between the results of each model. HiGRU has achieved 88.31% accuracy, ranking second, followed by BiLSTM. Each time series model focuses on different characteristics, which are inferior to the ST-SiamNN. For the dataset PARKINSON_HW, TS-LSTM, HiGRU, and BiLSTM still have a steady effect. Due to the lack of spatial features of handwriting, the CNN is less suitable for handwritten data than the temporal model.

We also use ROC curves and AUC values to compare the performance of these deep models (Figure 13). For the NewHandPD (image) dataset (Figure 13(a)), we can see that the ST-SiamNN ranks at the top and LSTM is at the bottom. The AUC value of TS-LSTM and HiGRU is about 4% lower than the ST-SiamNN. The CRNN and CNN achieve a similar AUC due to the role of the convolution layer. For the NewHandPD (signal) dataset (Figure 13(b)), the performance of deep models is divided into three levels: Level 1:

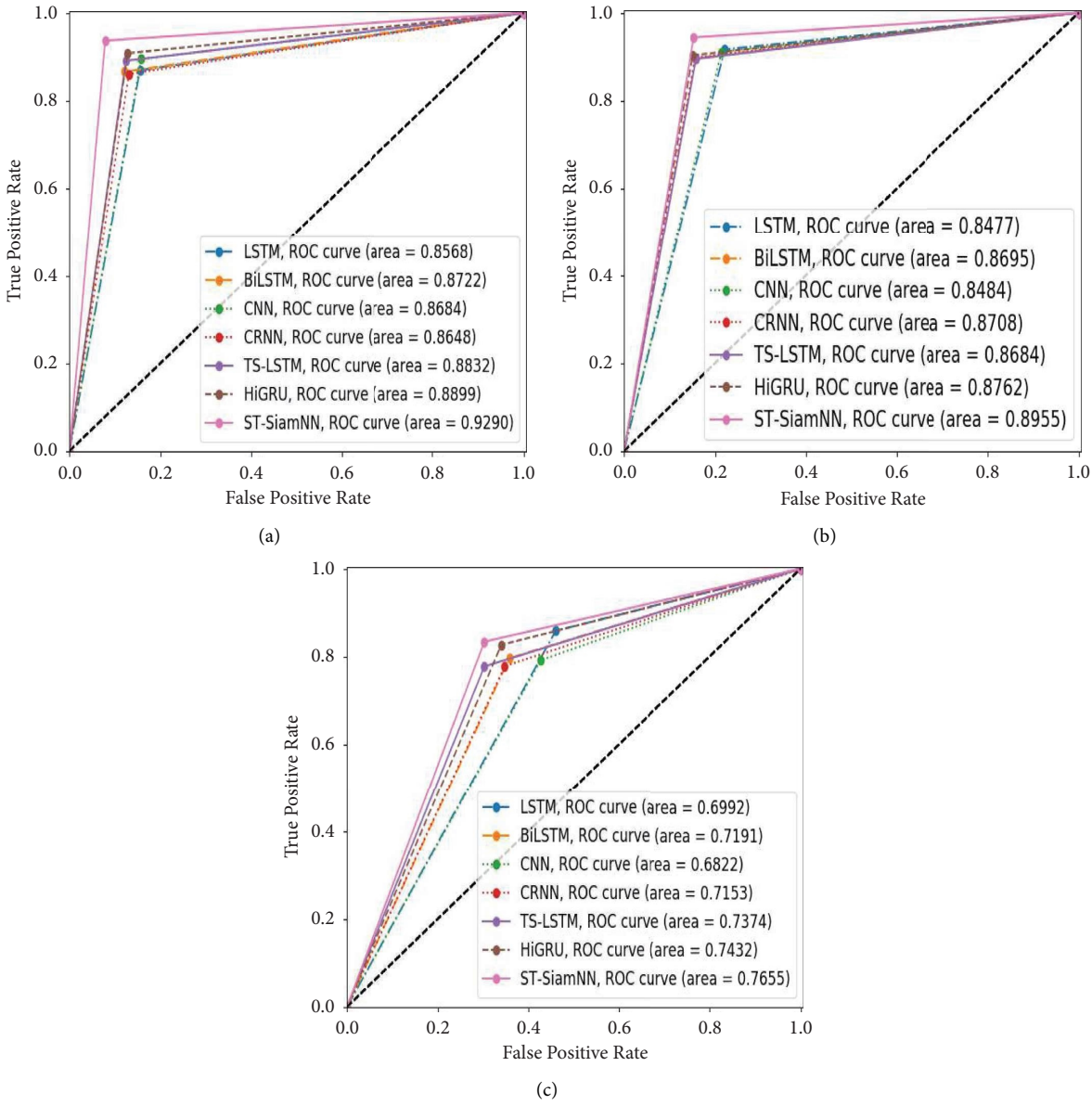


Figure 13: The ROC curve of the ST-SiamNN on the datasets. The AUC value is used to evaluate the classification effect and stability of each class. (a) The NewHandPD (image) dataset. (b) The NewHandPD (signal) dataset. (c) The PARKINSON_HW dataset.

LSTM and CNN; Level 2: HiGRU, TS-LSTM, CRNN, and BiLSTM; Level 3: ST-SiamNN. Level 1 is the basic model, Level 2 is the upgraded model, and Level 3 is the mixed upgraded model. We can see that their performance is increasing. In Figure 13(c), it can be seen that the area covered by the ROC curve is obviously smaller than that of the first two datasets and so is the corresponding AUC value. CNN's AUC ranks last because the signal data provided by this dataset has fewer characteristics in space than in time. The hybrid models, i.e., TS-LSTM, HiGRU, and the proposed ST-SiamNN, still outperform other deep models.

Additionally, we compare five popular methods in recent years (EfficientNet, MobileViT, ConvNeXt, Vision Transformer (ViT), and MobileNetV2) and find that the classification results on the NewHandPD dataset are the best. As shown in Table 5, we use four criteria to evaluate the testing results. The overall effect of MobileNetV2 is higher than that of other methods, but compared to ViT based effects, it shows significant disadvantages, indicating that the visual transformer is not sensitive to handwriting data.

4.3.2. Multimodal Handwriting Recognition. In the multimodal handwriting recognition process, we input handwriting data of different modalities into two Siamese networks (the SiamBiMNN and the SiamOctCNN) for training. There are six collocations as follows:

(1) SiamOctCNN: NewHandPD (image)
SiamBiMNN: PARKINSON_HW

(2) SiamBiMNN: NewHandPD (image)
SiamOctCNN: PARKINSON_HW

Table 5: Performance comparison with state-of-the-art approaches on the NewHandPD dataset.

Method	Dataset	Precision (%)	Recall (%)	$F1$ score (%)	Accuracy (%)
EfficientNet	NewHandPD	86.93	87.65	84.02	86.53
MobileViT	NewHandPD	87.82	89.21	87.49	89.62
ConvNeXt	NewHandPD	90.30	91.47	90.35	90.38
Vision transformer	NewHandPD	82.18	84.99	81.63	82.97
MobileNetV2	NewHandPD	90.26	91.48	90.14	**90.89**

It shows the MobileNetV2 obtains the highest accuracy among state-of-the-art approaches on the NewHandPD dataset.

Table 6: Results of the ST-SiamNN on multimodal data.

Multimodal data		Evaluation			
SiamBiMNN	SiamOctCNN	Precision (%)	Recall (%)	$F1$ score (%)	Acc (%)
NewHandPD (image)	NewHandPD (signal)	94.21	93.44	93.83	94.36
NewHandPD (signal)	NewHandPD (image)	92.00	95.04	93.50	93.98
NewHandPD (image)	PARKINSON_HW	89.51	93.43	91.43	90.98
PARKINSON_HW	NewHandPD (image)	88.10	90.24	89.16	89.85
NewHandPD (signal)	PARKINSON_HW	84.83	87.75	86.27	82.24
PARKINSON_HW	NewHandPD (signal)	85.44	84.82	85.13	80.84

Table 7: Performance comparison with state-of-the-art approaches on fusion datasets.

Multimodal data	LSTM (%)	BiLSTM (%)	BiMNN (%)	CNN (%)	CRNN (%)	TS-LSTM (%)	HiGRU (%)	ST-SiamNN (%)
NewHandPD (image) + NewHandPD (signal)	86.95	88.84	90.01	87.59	88.10	89.38	89.59	**94.36**
NewHandPD (image) + PARKINSON_HW	84.59	85.12	86.17	86.69	85.11	84.32	86.19	**90.98**
NewHandPD (signal) + PARKINSON_HW	72.36	73.28	73.80	73.59	71.49	72.11	75.20	**82.24**

It shows that ST-SiamNN achieves the highest accuracy on fusion datasets.

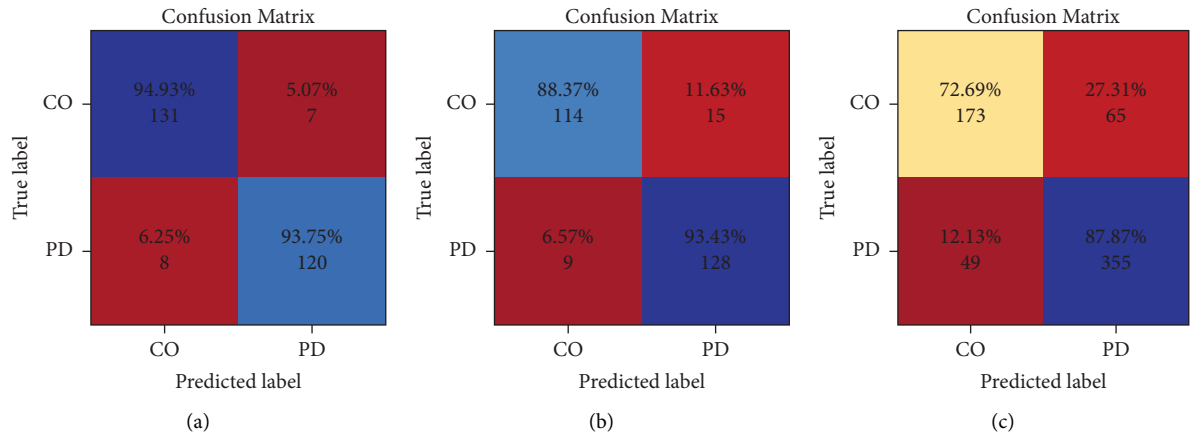


Figure 14: The confusion matrix of the ST-SiamNN on the datasets. (a) NewHandPD (image + signal). (b) NewHandPD (image) + HW. (c) HW + NewHandPD (signal).

(3) SiamOctCNN: NewHandPD (signal)
SiamBiMNN: PARKINSON_HW
(4) SiamBiMNN: NewHandPD (signal)
SiamOctCNN: PARKINSON_HW
(5) SiamOctCNN: NewHandPD (signal)
SiamBiMNN: NewHandPD (image)

Table 8: Performance comparison with single-modal approaches on NewHandPD.

Tests	Pereira et al. [47] (%)	Pereira et al. [48] (%)	Ribeiro et al. [49] (%)	Diaz et al. [50] (%)	Xu and Zhu [51] (%)	Zhu et al. [52] (%)	ST-SiamNN (%)
Spiral	77.53	78.26	89.48	94.44	81.17	77.45	**95.86**
Meander	87.14	80.75	92.24	91.11	78.18	70.86	**93.39**
Circle	—	68.04	—	88.89	—	—	**90.21**

ST-SiamNN achieves the highest accuracy on three tests.

(6) SiamBiMNN: NewHandPD (signal)
SiamOctCNN: NewHandPD (image)

The evaluation of the performance of the six bimodal data combinations is demonstrated in Table 6. The results in Tables 6 and 7 show that different signals (NewHandPD (image) + NewHandPD (signal)) from the same data set are pieced together into a sample through manual screening. Signals from different datasets (PARKINSON_HW + NewHandPD (image)) are directly processed into a matrix of appropriate size and input into the network. PD and non-PD detection are to analyze the input data. Because the data of NewHandPD (signal) and NewHandPD (image) are corresponding, we carry out data pairing preprocessing and send them to the SiamBiMNN and the SiamOctCNN for training. It can be seen that the accuracy of using fusion data is higher than that of using single data. On the contrary, due to the influence of similar data between classes in the PARKINSON_HW dataset, the performance of the model decreases slightly, which makes the result higher than that using a single PARKINSON_HW and lower than that using a single NewHandPD. Here, we are concerned that the accuracy rate of the BiMNN is about 1% higher than that without adding a path, and the convergence speed is faster, which is sufficient to prove the efficiency of its cell.

The comparison of the recognition accuracy of deep models is shown in Table 7. The deep model is evaluated by training the direct connection data. After fusing the two sets of data in NewHandPD, the performance of most models is improved slightly. After adding the data of PARKINSON_HW, the average accuracy of deep models is reduced, which shows the limitation of these single models and the disadvantage of the hard fusion method. The designed data fusion algorithm ST-SiamNN processes the input data separately and makes the training results reach the best. The extracted features can better show the differences in handwriting between PD patients and healthy subjects, which are superior to other deep models. The confusion matrix of classification is demonstrated in Figure 14. It can be clearly seen that the ST-SiamNN is successful in the detection of PD patients. In the combination "PARKINSON_HW + NewHandPD (signal)," the recognition rate of healthy subjects is only 72.69%, which indicates that the model is more sensitive to the handwriting data containing images, and the processing of multidimensional signal data will be the next research goal.

Since there is little work on multimodal data fusion in the field of handwriting recognition for Parkinson's disease, we only compare the classification effect of the single-modal methods on the NewHandPD dataset. As shown in Table 8, this experiment compares three different tests in the dataset separately. It can be seen that the performance of our method still exceeds that of other single-modal studies, including CNN + LSTM, SVM, random forest, statistical methods, and multi-BiGRU. Since there are few methods on the PARKINSON_HW dataset, we will not discuss them in this article. In summary, among the three handwriting tests, the accuracy of recognizing spirit is the highest, which is very recognizable and can be the best choice for PD detection. The accuracy of circle recognition is the lowest because it is difficult to distinguish PD from healthy people.

5. Conclusion

In this article, we propose a novel methodology for identifying multimodal handwritings of PD patients and healthy people using a Spatio-temporal Siamese neural network (ST-SiamNN) from images and signals obtained after the examination of different devices. The ST-SiamNN is based on a Siamese structure where we have embedded multiple networks to generate the most discriminant features. In this difficult problem, it is necessary to take advantage of both deep learning and traditional classifiers to capture the changes in multimodal PD handwriting data and produce the most discriminative feature vector. The experiment emphasizes that the proposed framework and the multimodel fusion method make it easier to screen and detect PD patients or individuals with potential PD from the subjects, and the high accuracy of the classification result also promotes handwritten training and the nonmanual data processing process, which provides a reference for auxiliary medical diagnosis and treatment.

Considering the limitations of the current study, the size of the dataset can still be increased to include more categories and subjects, which will help better evaluate the generalization ability of the proposals. The conventional program will bring the bias of developers. Every answer or decision output in a deep learning system cannot be explained accurately, and the labeled data set of training is processed in advance, so fairness requires to be tested. When it is used in the medical scene of human life, the system should be transparent and interpretable enough; otherwise, people's trust in it will be greatly reduced.

Data Availability

The data supporting the findings of the current study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Aite Zhao (the corresponding author) supervised the study, wrote the original draft, and wrote and edited the review. Huimin Wu was involved in the methodology, visualized the software, and wrote the original draft. Ming Chen performed the formal analysis, was involved in the methodology, and visualized the software. Nana Wang was involved in the methodology, visualized the software, and wrote the original draft.

Acknowledgments

This research was supported in part by National Natural Science Foundation of China under grant no. 62106117, the China Postdoctoral Science Foundation under grant no. 2022M711741, and the Natural Science Foundation of Shandong Province under grant no. ZR2021QF084.

References

[1] N. L. Neuroscience, "My life with Parkinson's," *Nature*, vol. 503, no. 7474, pp. 29-30, 2013.

[2] S. Meoni, A. Macerollo, and E. Moro, "Sex differences in movement disorders," *Nature Reviews Neurology*, vol. 16, 2020.

[3] A. Zhao, J. Li, J. Dong et al., "Multimodal gait recognition for neurodegenerative diseases," *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 9439–9453, 2022.

[4] A. Zhao, J. Dong, J. Li, L. Qi, and H. Zhou, "Associated spatio-temporal capsule network for gait recognition," *IEEE Transactions on Multimedia*, vol. 24, pp. 846–860, 2022.

[5] A. Zhao and J. Li, "Two-channel lstm for severity rating of Parkinson's disease using 3d trajectory of hand motion," *Multimedia Tools and Applications*, vol. 81, no. 23, pp. 33851–33866, 2022.

[6] M. Yu, T. Liu, Z. Guan et al., "Salstm: an improved lstm algorithm for predicting the competitiveness of export products," *International Journal of Intelligent Systems*, vol. 37, no. 9, pp. 6185–6200, 2022.

[7] B. Sahu and S. N. Mohanty, "Cmba-svm: a clinical approach for Parkinson disease diagnosis," *International Journal on Information Technology*, vol. 1, no. 73, 2021.

[8] S. Xu and Z. Pan, "A novel ensemble of random forest for assisting diagnosis of Parkinson's disease on small handwritten dynamics dataset," *International Journal of Medical Informatics*, vol. 144, Article ID 104283, 2020.

[9] Y. Li, L. Yang, P. Wang et al., "Classification of Parkinson's disease by decision tree based instance selection and ensemble learning algorithms," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 2, pp. 444–452, 2017.

[10] L. Ali, S. U. Khan, M. Arshad, S. Ali, and M. Anwar, "A multi-model framework for evaluating type of speech samples having complementary information about Parkinson's disease," in *Proceedings of the 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, Swat, Pakistan, July 2019.

[11] V. Bevilacqua, C. Loconsole, A. Brunetti, G. D. Cascarano, and E. D. Sciascio, "A model-free computer-assisted handwriting analysis exploiting optimal topology ANNs on biometric signals in Parkinson's disease research," in *Proceedings of the Intelligent Computing Theories and Application*, Wuhan, China, August 2018.

[12] Y. Tang, L. Zhang, F. Min, and J. He, "Multiscale deep feature learning for human activity recognition using wearable sensors," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 2, pp. 2106–2116, 2023.

[13] W. Huang, L. Zhang, H. Wu, F. Min, and A. Song, "Channel-equalization-har: a light-weight convolutional neural network for wearable sensor based human activity recognition," *IEEE Transactions on Mobile Computing*, vol. 1, p. 1, 2022.

[14] P. Durga, S. Jebakumari, and D. Shanthi, "Diagnosis and classification of Parkinsons disease using data mining techniques," *International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)*, vol. 3, pp. 86–90, 2016.

[15] C. Taleb, L. Likforman-Sulem, C. Mokbel, and M. Khachab, "Detection of Parkinson's disease from handwriting using deep learning: a comparative study," *Evolutionary Intelligence*, vol. 1, no. 1, 2020.

[16] A. Zhao, Y. Wang, and J. Li, "Transferable self-supervised instance learning for sleep recognition," *IEEE Transactions on Multimedia*, vol. 1, p. 1, 2022.

[17] Y. Wang, Z. Lv, Z. Sheng, H. Sun, and A. Zhao, "A deep spatio-temporal meta-learning model for urban traffic revitalization index prediction in the covid-19 pandemic," *Advanced Engineering Informatics*, vol. 53, Article ID 101678, 2022.

[18] K. Wang, J. He, and L. Zhang, "Sequential weakly labeled multiactivity localization and recognition on wearable sensors using recurrent attention networks," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 4, pp. 355–364, 2021.

[19] C. Han, L. Zhang, Y. Tang, W. Huang, F. Min, and J. He, "Human activity recognition using wearable sensors by heterogeneous convolutional neural networks," *Expert Systems with Applications*, vol. 198, Article ID 116764, 2022.

[20] M. Diaz, M. Moetesum, I. Siddiqi, and G. Vessio, "Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and bigrus," *Expert Systems with Applications*, vol. 168, 2020.

[21] A. Al-Wahishi, N. Belal, and N. Ghanem, "Diagnosis of Parkinson's disease by deep learning techniques using handwriting dataset," *Advances in Signal Processing and Intelligent Recognition Systems*, vol. 1365, 2021.

[22] M. Xu, J. Du, Z. Xue, Z. Guan, F. Kou, and L. Shi, "A scientific research topic trend prediction model based on multi-lstm and graph convolutional network," *International Journal of Intelligent Systems*, vol. 37, no. 9, pp. 6331–6353, 2022.

[23] Y. Xiao, H. Yin, Y. Zhang, H. Qi, Y. Zhang, and Z. Liu, "A dual-stage attention-based conv-lstm network for spatio-temporal correlation and multivariate time series prediction," *International Journal of Intelligent Systems*, vol. 36, no. 5, pp. 2036–2057, 2021.

[24] X. Liu, L. Wang, L. Zheng, F. Du, and Q. Zou, "A dual-branch model for diagnosis of Parkinson's disease based on the independent and joint features of the left and right gait," *Applied Intelligence*, vol. 51, pp. 1–12, 2021.

[25] N. Abdul Hamid, "Handwritten recognition using svm, knn and neural network," 2017, https://arxiv.org/abs/1702.00723.

[26] J. Ouyang, "Combining extreme learning machine, rf and hog for feature extraction," in *Proceedings of the 2017 IEEE Third International Conference on Multimedia Big Data (BigMM)*, IEEE, Laguna Hills, CA, USA, April 2017.

[27] S. Ahlawat and A. Choudhary, "Hybrid cnn-svm classifier for handwritten digit recognition," *Procedia Computer Science*, vol. 167, pp. 2554–2560, 2020.

[28] Y. Jia, Y. Zhao, Y. Zhang, and S. Fan, "A lightweight handwriting recognition system based on an improved convolutional neural network," in *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition*, Xiamen China, October 2020.

[29] S. Xiao, L. Peng, R. Yan, and S. Wang, "Deep network with pixel-level rectification and robust training for handwriting recognition," *SN Computer Science*, vol. 1, no. 3, pp. 145–213, 2020.

[30] M. Gazda, M. Hire, and P. Drotar, "Multiple-fine-tuned convolutional neural networks for Parkinson's disease diagnosis from offline handwriting," *IEEE Transactions on Systems Man and Cybernetics*, vol. 52, 2021.

[31] D. C. Shubhangi and P. Gundagurti, "Deep learning based diagnosis of Parkinson's disease using cnn," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 6, pp. 351–355, 2020.

[32] P. Voigtlaender, P. Doetsch, and H. Ney, "Handwriting recognition with large multidimensional long short-term memory recurrent neural networks," in *Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Shenzhen, China, October 2017.

[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[34] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific Reports*, vol. 8, no. 1, pp. 6085–85, 2018.

[35] A. Zhao, J. Li, and M. A. Spidernet, "A spiderweb graph neural network for multi-view gait recognition," *Knowledge-Based Systems*, vol. 206, 2020.

[36] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2298–2304, 2017.

[37] I. Lee, D. Kim, S. Kang, and S. Lee, "Ensemble deep learning for skeleton-based action recognition using temporal sliding lstm networks," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, October 2017.

[38] W. Jiao, H. Yang, I. King, and M. R. Lyu, "HiGRU: hierarchical gated recurrent units for utterance-level emotion recognition," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 397–406, Association for Computational Linguistics, Minneapolis, Minnesota, June 2019.

[39] M. Tan, V. Quoc, and L. Efficientnet, "Rethinking model scaling for convolutional neural networks," 2019, https://arxiv.org/abs/1905.11946.

[40] A. Dosovitskiy, L. Beyer, K. Alexander et al., "An image is worth 16x16 words: transformers for image recognition at scale," *CoRR*, 2020, https://arxiv.org/abs/2010.11929.

[41] S. Mehta and M. Rastegari, "Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer," *CoRR*, 2021, https://arxiv.org/abs/2110.02178.

[42] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," *CoRR*, 2022, https://arxiv.org/abs/201.03545.

[43] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, Salt Lake City, UT, USA, June 2018.

[44] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics," in *Proceedings of the 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 340–346, Sao Paulo, Brazil, October 2016.

[45] M. E. Isenkul, B. E. Sakar, and O. Kursun, "Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease," in *Proceedings of the 2nd International Conference on E-Health and TeleMedicine-ICEHTM 2014*, Istanbul, Turkey, May 2014.

[46] B. E. Sakar, M. E. Isenkul, C. O. Sakar et al., "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.

[47] C. R. Pereira, D. R. Pereira, F. A. Silva et al., "A new computer vision-based approach to aid the diagnosis of Parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 136, pp. 79–88, 2016.

[48] C. R. Pereira, D. R. Pereira, G. H. Rosa et al., "Handwritten dynamics assessment through convolutional neural networks: an application to Parkinson's disease identification," *Artificial Intelligence in Medicine*, vol. 87, pp. 67–77, 2018.

[49] L. C. Ribeiro, L. C. Afonso, and J. P. Papa, "Bag of samplings for computer-assisted Parkinson's disease diagnosis based on recurrent neural networks," *Computers in Biology and Medicine*, vol. 115, Article ID 103477, 2019.

[50] M. Diaz, M. Moetesum, I. Siddiqi, and G. Vessio, "Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and bigrus," *Expert Systems with Applications*, vol. 168, Article ID 114405, 2021.

[51] Z. Xu and Z. Zhu, "Handwritten dynamics classification of Parkinson's disease through support vector machine and principal component analysis," *Journal of Physics: Conference Series*, vol. 1848, no. 1, Article ID 012098, 2021.

[52] Z. Zhu, S. Xu, and Z. Pan, "A cascade ensemble learning model for Parkinson's disease diagnosis using handwritten sensor signals," *Journal of Physics: Conference Series*, vol. 1631, no. 1, Article ID 012168, 2020.