

Detection of Brain Abnormalities in Parkinson's Rats by Combining Deep Learning and Motion Tracking

Houchi Li¹, Quan He, and Lei Wu²

Abstract—Parkinson's disease (PD) is a chronic neurodegenerative disease that affects the central nervous system. PD mainly affects the motor nervous system and may cause cognitive and behavioral problems. One of the best tools to investigate the pathogenesis of PD is animal models, among which the 6-OHDA-treated rat is a widely employed rodent model. In this research, three-dimensional motion capture technology was employed to obtain real-time three-dimensional coordinate information about sick and healthy rats freely moving in an open field. This research also proposes an end-to-end deep learning model of CNN-BGRU to extract spatiotemporal information from 3D coordinate information and perform classification. The experimental results show that the model proposed in this research can effectively distinguish sick rats from healthy rats with a classification accuracy of 98.73%, providing a new and effective method for the clinical detection of Parkinson's syndrome.

Index Terms—Parkinson's disease, 6-OHDA, CNN-BGRU, 3D coordinate.

I. INTRODUCTION

PARKINSON'S syndrome is a relatively common clinical neurodegenerative disease, that mostly occurs in middle-aged and elderly people, and causes a range of symptoms, including motor, sensory and speech impairments. These symptoms have a series of negative effects on patients with Parkinson's disease (PD), and greatly affect their quality of life. According to the National Institutes of Health (NIH), there are approximately 4 to 6 million Parkinson's patients worldwide, and 40% to 60% of Parkinson's patients suffer from mental syndromes such as depression, autism, dementia, delusions, and mania [1], [2]. There are many early symptoms of Parkinson's disease, such as handshaking, walking, and other motor abnormalities or sensory abnormalities e.g.

olfactory loss [3]. However, it is difficult for patients to identify the early symptoms of PD, thus they often cannot obtain timely treatment [4]. Since early treatment is more effective and can slow down disease progression, it is important to develop highly accurate and reliable health information systems that are required to detect PD [5]. The clinical diagnosis of Parkinson's syndrome is mainly based on magnetic resonance imaging (MRI) and olfactory tests combined with dopamine transporter scans (DaTscans) [6].

Ideally, animal models can accurately simulate pathological, histological, and biochemical changes and their resulting functional disorders, providing important auxiliary tools for analysing the pathogenesis and treatment principles of human diseases [7]. The most commonly employed Parkinson's disease model is the rat model of 6-hydroxydopamine (6-OHDA) injury. The rat model of medial forebrain tract (MFB) injury was established by unilateral injection of 6-OHDA. Then apomorphine was injected, and successfully modelled rats were selected as experimental subjects by rotation test [8]. The introduction of 6-OHDA caused damage to dopaminergic neurons of the nigrostriatal pathway in the rat brain and reduced dopamine secretion, resulting in motor and nonmotor deficits. Researchers have developed a variety of methods for the behavioural assessment of Parkinson's disease rats, such as the cylinder test, gait adjustment test, olfactory asymmetry test, and forced swim test [9], [10], [11].

In traditional animal experiments, animal behaviour data are mainly collected by visual observation and manual recording, which have a great influence on subjectivity, heavy workload, and tediousness, and are prone to errors during the observation process [12]. Later, with the development of modern electronic technology and computer science, collection of behavioural data of animals gradually joined the ranks of automation or semiautomation. In particular, machine vision technology research on animal behaviour acquisition has rapidly developed in recent years [13], [14]. However, with or without the use of automated collection of animal behaviour data, these methods pay minimal attention to atypical behaviour. Since animals exhibit behaviours with complex spatiotemporal patterns, researchers usually need to increase the sampling size to ensure robust results. However this step is not achievable when making diagnostics. In this research, an optical motion tracking system was used to collect rat behaviours in an open field. Several retroreflective markers

Manuscript received 22 September 2022; revised 24 November 2022 and 22 December 2022; accepted 13 January 2023. Date of publication 19 January 2023; date of current version 3 February 2023. (Corresponding author: Lei Wu.)

Houchi Li is with the School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411100, China (e-mail: 18175681159@163.com).

Quan He is with the Department of Biology and Chemistry, National University of Defense Technology, Changsha 410073, China (e-mail: hqatw213@126.com).

Lei Wu is with the Hunan Engineering Research Center for Intelligent Decision Making and Big Data on Industrial Development, Hunan University of Science and Technology, Xiangtan 411100, China (e-mail: wulei@hnust.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3237916

were mounted on the head and bodies of rats so that the high-resolution infrared cameras could obtain stable signals and accurately resolve their spatial coordinates [15], [16]. Compared with traditional 2D image analysis [17], 3D coordinates can effectively reduce the unavoidable information loss in the process of projecting a 3D real-world scene into a 2D image. Moreover, the motion tracking system provides coordinates of markers with a submillimeter spatial resolution at a frame rate of 240 per second. We expect this fine coordinate variation to reveal objective status by characterizing the quality of behaviours [18].

However, the high-resolution kinetics data of markers require effective data mining algorithms or techniques. Machine learning techniques have long demonstrated strong analytical capabilities for behavioural data processing. For example, a support vector machine (SVM) is usually employed for animal behaviour classification, and a behaviour classification model is constructed by extracting various features during animal movement [19]. The hidden Markov chain model (HMM) can predict and judge the transition between animal behaviour modules and behaviour modules, and construct a reasonable behaviour sequence [20]. As an important branch of machine learning, deep learning greatly simplifies the overall algorithm analysis and learning process of traditional machine learning, and can automatically extract hidden features in datasets. Deep learning has been widely employed in animal behavior recognition. Stern et al. [21] utilized a convolutional neural network to detect whether *Drosophila* is in contact with the egg-laying substrate frame by frame, and the error rate of the result was only 0.072%. Arac et al. [22] used a CNN model composed of GoogLeNet followed by a long-short-term memory neural network to monitor the grasping of food particles by mouse paws. In addition to video images, deep learning can also be applied to other types of data, such as the introduction of three-dimensional marker coordinates [23], which can automatically learn the hidden laws of coordinate points and identify and classify them.

In this work, we propose a deep learning framework of CNN-BGRU to extract and classify the obtained 3D coordinate information spatiotemporal features. On this dataset, the classification accuracy of our proposed deep learning framework for Parkinson's disease is 98.73%, demonstrating the sufficiency of information required for diagnosing from high-resolution motion data. This method can be used to develop an effective diagnostic method without rotation tests on Parkinson's models and provides a new effective strategy for the transition from animal experimental research of Parkinson's disease to human clinical trials and treatments.

II. MATERIALS AND METHODS

A. Parkinson's Rat Model Construction and Data Collection

6-OHDA is a chemical agent similar to norepinephrine, which has strong neurotoxicity. Local injection into the rat brain can damage the dopaminergic neurons in the substantia

TABLE I
ROTARY TEST RESULTS

NO.	5min	10min	15min	20min	25min	30min	avg r/min
R1	43	84	128	176	229	277	9.2
R2	40	85	135	185	234	284	9.5
R3	44	91	139	194	232	285	9.5
R4	58	111	161	212	265	316	10.5
R5	50	113	180	242	308	365	12.2
R6	39	112	178	248	306	370	12.3

nigra striatum pathway in the rat brain and reduce the secretion of dopamine, resulting in motor and nonmotor defects. Therefore, in clinical research on Parkinson's disease, the animal model of Parkinson's disease that contributes the most is undoubtedly the unilateral MFB-injected 6-OHDA-injured rat [9]. Therefore, in this research, 12 adult male Sprague-Dawley rats (starting body weight of 250-300g) were divided into two groups: 6 rats with medial forebrain bundle (MFB) injury were selected as the experimental group and 6 normal rats were used as the control group. 6-OHDA was injected under the medial forebrain bundle (MFB) of rats in the experimental group to simulate Parkinson's disease.

After the injection of 6-OHDA into the lateral brain of rats, the dopamine secretion of both sides of the brain is inconsistent, and the stimulation of apomorphine, a dopamine receptor agonist, will cause hypersensitivity of the right brain receptor, resulting in asymmetric behaviour. All rats underwent a rotation test performed 25 days after surgery. After injection of 1.25 mg/kg apomorphine, each rat was placed in a freely movable behavioural chamber. At least 40 minutes after injection, the contralateral rotation was counted for 10 minutes. When the rotation speed exceeded 7 r/min, the modelling was considered successful. The results for the rotation of the selected rats are shown in Table IV.

The daily behaviour of these six rats was not significantly different from that of normal rats, but all of them rotated to the damaged side after injection of apomorphine. In particular, the rats numbered R6 showed the phenomenon of standing instability and dumping during rotation, and the actual number of rotations may be greater than the observed times. In this experiment, the rats in the control group were selected as a reference, and no rotation was observed in this group after injection of apomorphine.

The behavioural data were collected using the OptiTrack optical capture device, which consists of 13 cameras. We set up reflective markers on the head and torso to provide sufficient information for identifying asymmetric movements. The sampling frequency of streaming data collection was 120HZ, to ensure the stability and accuracy of the data, there were enough samples, and each rat was recorded for at least one hour. In addition, the brain-injured rat group was recorded at different stages. To ensure that the rats were not disturbed as much as possible, all experiments were conducted in a constant temperature, quiet and dark environment [23]. Each state of the rat corresponds to 6 reflective markers, each light marker has three-dimensional spatial coordinates, and the dataset contains a total of 1.46 million coordinate points. Figure 1(a) and

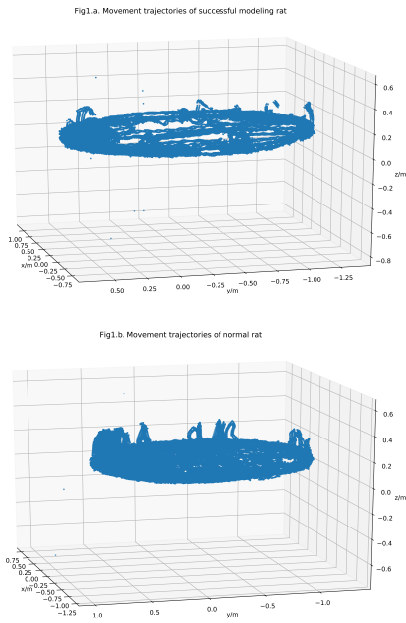


Fig. 1. Figure 1.a and Figure 1.b show the movement trajectories of PD rats and the movement trajectories of normal rats.

Figure 1(b) show the movement trajectories of PD rats and normal rats, respectively.

All animals were kept in a pathogen-free environment and fed ad libitum, the diameter of the site is 2m. The procedures for the care and use of animals were approved by the Ethics Committee of the Hunan Drug Safety Evaluation and Research Center and all applicable institutional and governmental regulations concerning the ethical use of animals were followed.

B. Data Preprocessing

Considering that there are broken frames in the original data record, to ensure continuity within the sample as much as possible, this research will exclude sample data with excessive discontinuity, which generally does not exceed 1.5 times the total length of the sample. Before training our CNN-BGRU model, we needed to refine the input data as the neural network is sensitive to different data. Therefore, in this work, we remove outliers from the input data and then standardize this data. Standardization can ensure that the data of each dimension conforms to the normal distribution, that is, the standard deviation is 1 and the mean is 0. The standardization equation is expressed as follows:

$$X = (X - \mu)/\sigma \quad (1)$$

where μ is the mean value of all sample data, and σ is the standard deviation of all sample data.

C. Deep Learning Network Structure

This research developed a two-step diagnostic framework for Parkinson's disease. The first step is to preprocess the input data to remove outliers and normalize the input dataset with standard deviation to fit a standard normal distribution and then

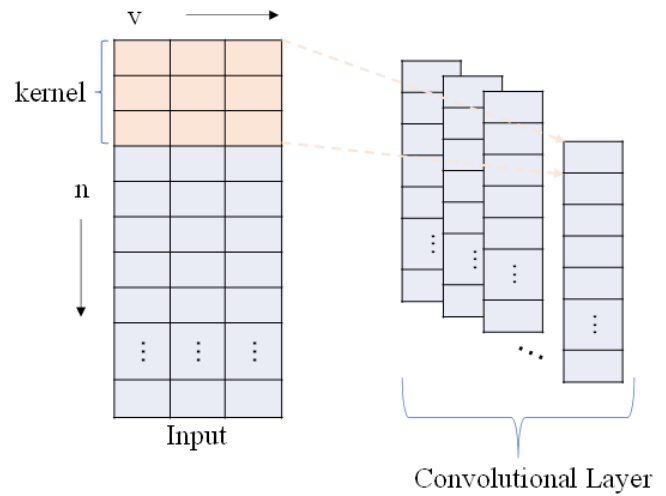
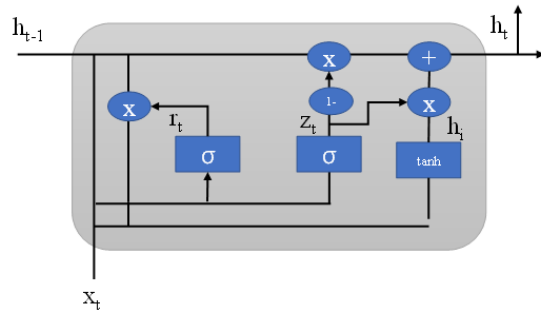


Fig. 2. Typical convolution operation on time series data.

feed the processed input data into the training phase. Inspired by the network structure of CNN-LSTM [24], this research developed a hybrid model combining the CNN-BGRU model and achieved better results.

1) *Convolutional Neural Networks*: A time series has a strong one-dimensional structure: variables that are close to each other are highly correlated in space or time. The advantage of local correlation is that feature extraction and local combination are performed before identifying spatial and temporal objects. Therefore, it is particularly important to extract local features from the constructed local regions [25]. As one of the most famous deep learning models, the convolutional neural network extracts rich local features by using various filters in convolutional layers, pooling layers, normalization layers, and fully-connected layers, thereby improving various performances of the task. A one-dimensional convolutional neural network extracts effective and representative features from time series data by using multiple filters to perform one-dimensional convolution operations. The details of a typical convolution operation on time series data are shown in Figure 2. The convolution layer convolves with the input signal through a convolution kernel to generate the feature map of the next layer. where n and v represent the length of the time series and the number of features, respectively.

2) *BGRU*: However, the features extracted by a CNN are short-term and local. Although a CNN can sufficiently extract spatial features, it does not consider the temporal correlation between data, and the experimental data that we analyse are long-term time series data. Therefore, to accurately identify the symptoms of Parkinson's disease, it is necessary to consider the time background of the data. A recurrent neural network (RNN) [26] is a traditional deep learning method specialized in processing time series data. However, RNNs can handle certain short-term dependencies, but cannot handle long-term dependencies. Therefore, when the time series is long, the gradient at the back of the sequence has difficult propagating back to the previous sequence, which leads to the problem of gradient disappearance. To solve the problem of RNN gradient



(a) GRU structure diagram.

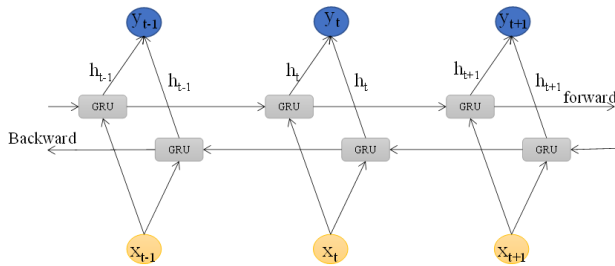


Fig. 3.b BGRU network structure diagram.

(b) BGRU network structure diagram.

Fig. 3. Figure 3 (a) and (b) show the network structures of GRU and BGRU respectively.

disappearance, Hochreiter and Schmidhuber [27] proposed a long short-term memory (LSTM) recurrent neural network, which combines short-term memory and long-term memory by gating, overcoming the shortcomings of traditional RNNs. Cho, Kyunghyun, et al. [28] proposed a variant GRU that works well based on LSTM networks. The internal unit of GRU is similar to the internal unit of LSTM, with the exception that a GRU combines the input gate and forget gate in LSTM into a single update gate. The structure of a GRU is shown in Figure 3.a.

The detailed calculation formula is as follows:

$$z_t = \sigma(W_z \cdot [h(t-1), x_t]) \quad (2)$$

$$r_t = \sigma(W_r \cdot [h(t-1), x_t]) \quad (3)$$

$$h_i = \tanh(W \cdot [r_t * h(t-1), x_t]) \quad (4)$$

$$h_t = (1 - z_t) * h(t-1) + z_t * h_i \quad (5)$$

where x , h , z , and r are the input vector, output vector, update gate state, and reset gate state, respectively. The update gate is responsible for controlling the influence of the state information h_{t-1} of the previous moment on the state of the current moment, and the reset gate is responsible for controlling the degree of disregarding the state information h_{t-1} of previous moment. W_z , W_r , and W represent the weight matrix corresponding to the target state, h_i is the hidden layer, the output of the hidden unit includes the update part and previous part, \tanh represents the tangent hyperbolic function, and σ represents the sigmoid activation function.

GRU can not only overcome the vanishing gradient problem existing in a traditional RNN, but also has a relatively simple structure, lower complexity, and faster convergence speed compared with the LSTM network. The model structure of the bidirectional GRU [29] is similar to that of the GRU model. A recurrent neural network composed of two independent GRU networks, the model utilizes both past information and future information. In this paper, we use the bidirectional GRU model. The structural principle is shown in Figure 3.b, and its network consists of two subnetworks: forwards state and backwards state, which represent forwards transmission and backwards transmission, respectively.

D. CNN+BGRU

The CNN-BGRU-based hybrid model proposed in this paper includes the CNN and BGRU. The CNN is employed for feature extraction, and the BGRU is utilized to process data with a time series. This paper divides the hybrid model into five parts: input layer, convolutional layer, pooling layer, BGRU layer, fully connected layer, and sigmoid output layer. To prevent overfitting, we add a dropout layer between each fully connected layer to improve the generalization ability of the model. The flow chart of the model procedure proposed in this study is shown in Figure 4. First, the data normalized by the standard deviation are input into the first convolution layer to propose abstract features. The designed 1DCNN consists of three convolution layers, and the number of convolution kernels is 64, 48 and 32. The kernel_size of each convolution kernel is 3, and the stride of the convolution kernel is 1. The activation function is a rectified linear unit (ReLU), which adds nonlinear factors to the network, strengthens the representation ability of the network, and solves problems that cannot be solved by linear models. The convolution layer operation of the i th filter is defined as follows:

$$x_j^l = \emptyset \left(\sum_{i=0}^{K_j} W_{i,j}^l x_j^{l-1} + b_j^l \right) \quad (6)$$

where x_j^l represents the characteristic vector corresponding to the j th convolution kernel of the l th layer, $W_{i,j}^l$ is the weight of the filter, b_j^l is the bias of the filter, K_j is the receptive field of the current neuron, and \emptyset is the ReLU activation function.

The convolved and activated output is dimensionally reduced by a maximum pooling layer with a pooling window size of 2. The pooling layer models the obtained feature map and converts it to a more abstract feature form. Then two convolution layers are used to further extract higher-level features. Next, the features extracted by the convolutional neural network are mapped to a fully connected layer with 256 neurons, and the dropout layer can be set to prevent overfitting. The fully connected layer maps the extracted distributed feature representation to the sample label space as the input of the BGRU network. Two BGRU layers are employed to model temporal features, and the number of neurons is 64 and 32. After the features are extracted by the BGRU network, the output signal is sent to two fully connected layers, the number of units is 256 and 64, and the

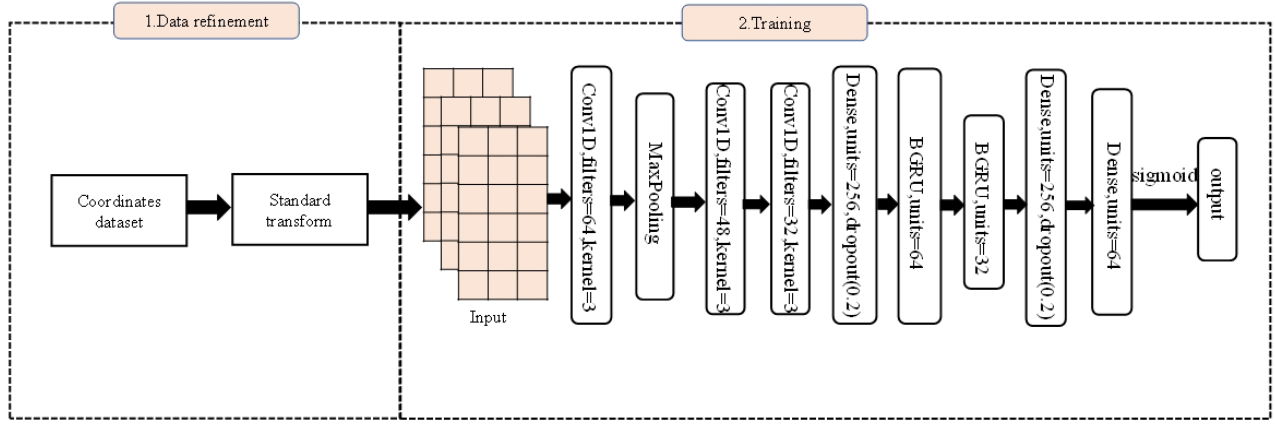


Fig. 4. Flow chart of the experimental procedure.

dropout layer is added after the first fully connected layer. Next, the signal is classified by the sigmoid activation function. The gradient descent algorithm uses Adam, the learning rate is 0.01, the batch_size is 200, and each training is 50 epochs.

E. Experimental Environment

This experiment uses an Intel i5-9400 6-core 6-thread processor, 16 GB memory, win10 operating system, anaconda3 as the experimental platform, and Python language programming. The deep learning framework used Keras in TensorFlow 2.0 and completed four hours of training using an NVIDIA 1080 GPU.

III. EXPERIMENTS AND RESULTS

A. Model Evaluation Metrics

The performance measures applied in this paper are the F1-score, recall, precision, accuracy, and confusion matrix. Accuracy is the ratio of correctly classified samples to the total number of samples. When activities are correctly classified, they can be classified as true positives (TPs) and true negatives (TNs), and when they are misclassified, they can be classified as false negatives (FNs) and false positives (FPs). Performance measures can be defined in terms of TP, TN, FP, and FN.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

Precision is the ratio of correctly predicted positives to the total number of samples classified as positives.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Recall is the ratio of correctly predicted positive samples to the actual number of positive samples.

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

The F1-score is the harmonic mean of recall and precision, which is generally used when the dataset is unbalanced.

$$F1 - score = \frac{2(Precision * Recall)}{(Precision + Recall)} \quad (10)$$

The root mean square error (RMSE) measures the deviation between the observed value and the true value.

$$RMSE(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x_i) - y_i)^2} \quad (11)$$

B. Analysis of Experimental Results

Experiments were conducted using a locally obtained 3D coordinate dataset, and throughout the experiments, we selected 60% of the data as the training set, 20% of the data as the validation set and 20% of the data as the test set with a constant random seed for reproducing the results. We use the Adam optimizer to alleviate the gradient oscillation problem. We select the hyperparameters of the model on the validation set, and evaluate the effect of our model through the test set. Training and validation accuracy and loss curves for our model method, considering an increase in epochs, are demonstrated in Figure 5. The figure shows that the accuracy and loss curves quickly converge, and the training accuracy and validation accuracy achieve a good fit, which proves that the model proposed in this paper has good robustness in this dataset. To verify the advantages of our proposed model, we compare the proposed model with the deep learning models in [27], [28], [29], [30], [31], [32], and [33] according to the model evaluation metrics, and both use our dataset for training. The comparison results are shown in Table IV. It can be seen that the accuracy, precision, recall, and F1 scores of our proposed model are 98.73%, 98.45%, 98.33%, and 98.40%, respectively. The performance of the model proposed in this paper is similar to that of CNN-Bi-LSTM and significantly outperforms other models. GRU and LSTM have similar accuracies. GRU simplifies the model structure while maintaining the computational effect of LSTM, thereby reducing the amount of computation. Thus GRU has a shorter running time and higher efficiency. A comparison of the experimental results based on the GRU model and BGRU model, reveal that the accuracy rate of the BGRU is higher than that of the GRU, mainly as the BGRU obtains the state information before and after the input sequence, which

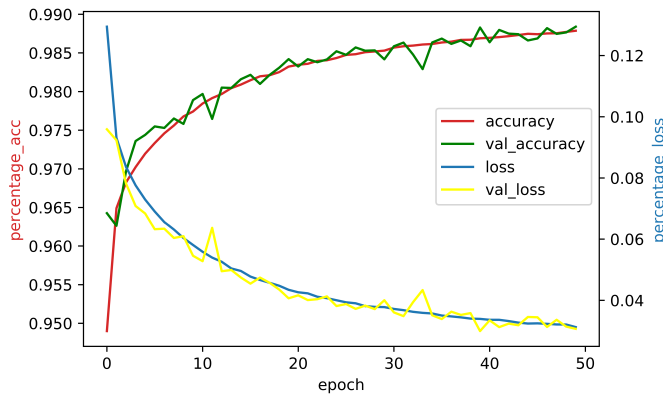


Fig. 5. ACC and LOSS curves of training.

TABLE II
PERFORMANCE OF DEEP LEARNING MODELS IN
BINARY CLASSIFICATION TASKS

Methods	Accuracy	Precision	Recall	F1-score
LSTM [27]	93.51%	93.01%	90.23%	82.91%
GRU [28]	94.78%	94.18%	92.39%	88.25%
BGRU [29]	95.85%	95.37%	94.17%	94.74%
BP [30]	94.52%	93.68%	92.24%	89.32%
CNN [31]	95.77%	95.15%	94.00%	94.42%
CNN+BLSTM [32]	98.93%	98.72%	98.55%	98.65%
CNN+GRU [33]	96.61%	95.45%	95.43%	95.14%
Our model	98.73%	98.45%	98.33%	98.40%

TABLE III
PERFORMANCE COMPARISON BETWEEN THE MACHINE LEARNING
MODEL AND THE MODEL IN THIS STUDY FOR
BINARY CLASSIFICATION TASKS

Methods	Accuracy	RMSE
Catboost [34]	61.23%	0.622
GaussianNB [35]	96.70%	0.181
Xgboost [36]	89.64%	0.322
RF [37]	73.02%	0.519
KNN [38]	95.18%	0.211
AdaBoostClassifier [39]	74.84%	0.500
Our model	98.73%	0.109

TABLE IV
10-FOLD CROSS-VALIDATION RESULT

NO.	Accuracy	Loss
1	98.70%	0.0336
2	98.58%	0.0363
3	98.60%	0.0358
4	98.81%	0.0310
5	98.55%	0.0371
6	98.46%	0.0394
7	98.66%	0.0340
8	98.89%	0.0285
9	98.65%	0.0346
10	98.74%	0.0323
AVG	98.66%	0.0343

can more accurately extract coordinate information. Thus the calculation results of the bidirectional neural unit model have higher accuracy.

To obtain the optimal model, we also tested the machine learning model in [34], [35], [36], [37], [38], and [39] on this dataset. The classification accuracies of Catboost, GaussianNB, Xgboost, random forest(RF),

K-nearest neighbour(KNN), and AdaBoostClassifier were 96.70%, 61.23%, 89.64%, 73.02%, 95.18% and 75.84%, respectively. We also compared the root mean square error (RMSE). The comparison results are shown in Table IV. Compared with the traditional machine learning method, the model proposed in this paper has higher accuracy, smaller deviation and has higher robustness on this dataset. Moreover, deep learning technology can automatically perform feature extraction on the original dataset, so it does not rely on artificial design features, which improves efficiency while saving manpower.

To prevent the model from overfitting, we adopt 10-fold cross-validation. The cross-validation results are shown in Table IV. We randomly divide the dataset into 10 mutually exclusive subsets of the same size, and each time, we randomly select 9 copies as the training set and the remaining copies as the test set. All data will participate in training and prediction to effectively avoid overfitting. Our model has high accuracy for each subset, which fully proves the stability of our model.

IV. CONCLUSION AND OUTLOOK

This paper proposes a deep learning framework that integrates CNN and BGRU networks for the automatic extraction and classification of behavioural coordinate information for Parkinson's disease. We exploit the robustness of the CNN for feature extraction and the strength of the BGRU for the classification of time series data. The network model of the CNN combined with the BGRU has a deep depth in both time and space, and the model can automatically perform feature extraction on the original dataset, so it does not rely on manual design features and saves manpower. The main contribution of this research is to use high-precision quantitative methods to characterize the rat model, via accurate data mining behaviour of the underlying information, bypassing the interference of macro behaviour, combined with deep learning technology to identify rat abnormalities. At the same time, in order to prevent overfitting phenomena, we also added 10-fold cross-validation. Different from traditional video capture of animal behaviour, our experiment captures the three-dimensional coordinate information of rats based on optical marker capture devices, avoiding the loss of information in the process of converting video to 2D images, and the coordinate information captured by optical markers is more precise. Our research provides a new and effective method for the transition from animal experimental research on Parkinson's disease to human clinical trials and treatments. Our future goal is to apply this method to clinical trials and provide a new paradigm for the clinical diagnosis of Parkinson's disease.

REFERENCES

- [1] W. Dauer and S. Przedborski, "Parkinson's disease: Mechanisms and models," *Neuron*, vol. 39, no. 6, pp. 889–909, 2003.
- [2] C. A. Davie, "A review of Parkinson's disease," *Brit. Med. Bull.*, vol. 86, no. 1, pp. 109–127, 2008.
- [3] J. Wingate, I. Kollia, L. Bidaut, and S. Kollias, "A unified deep learning approach for prediction of Parkinson's disease," 2019, *arXiv:1911.10653*.
- [4] W. Wang, J. Lee, F. Harrou, and Y. Sun, "Early detection of Parkinson's disease using deep learning and machine learning," *IEEE Access*, vol. 8, pp. 147635–147646, 2020.

- [5] H. W. Loh et al., "Application of deep learning models for automated identification of Parkinson's disease: A review (2011–2021)," *Sensors*, vol. 21, no. 21, p. 7034, Oct. 2021.
- [6] M. M. Ponsen, D. Stoffers, E. C. Wolters, J. Booij, and H. W. Berendse, "Olfactory testing combined with dopamine transporter imaging as a method to detect prodromal Parkinson's disease," *J. Neurol., Neurosurg. Psychiatry*, vol. 81, no. 4, pp. 396–399, Apr. 2010.
- [7] M. Gerlach and P. Riederer, "Animal models of Parkinson's disease: An empirical comparison with the phenomenology of the disease in man," *J. Neural Transmiss.*, vol. 103, nos. 8–9, pp. 987–1041, Aug. 1996.
- [8] R. Deumens, A. Blokland, and J. Prickaerts, "Modeling Parkinson's disease in rats: An evaluation of 6-OHDA lesions of the nigrostriatal pathway," *Experim. Neurol.*, vol. 175, no. 2, pp. 303–317, Jun. 2002.
- [9] L. O. Soto-Rojas et al., "A sequential methodology for integral evaluation of motor and non-motor behaviors in parkinsonian rodents," *MethodsX*, vol. 7, Jan. 2020, Art. no. 100821.
- [10] M. E. Johnson and L. Bobrovskaya, "An update on the rotenone models of Parkinson's disease: Their ability to reproduce the features of clinical disease and model gene–environment interactions," *NeuroToxicology*, vol. 46, pp. 101–116, Jan. 2015.
- [11] J. Boix, D. von Hieber, and B. Connor, "Gait analysis for early detection of motor symptoms in the 6-OHDA rat model of Parkinson's disease," *Frontiers Behav. Neurosci.*, vol. 12, p. 39, Mar. 2018.
- [12] P. N. Lehner, "Design and execution of animal behavior research: An overview," *J. Animal Sci.*, vol. 65, no. 5, pp. 1213–1219, Nov. 1987.
- [13] J. D. Crall, N. Gravish, A. M. Mountcastle, and S. A. Combes, "BEETag: A low-cost, image-based tracking system for the study of animal behavior and locomotion," *PLoS ONE*, vol. 10, no. 9, Sep. 2015, Art. no. e0136487.
- [14] Z. T. Pennington et al., "EzTrack: An open-source video analysis pipeline for the investigation of animal behavior," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, Dec. 2019.
- [15] S. M. Leonard, H. Xin, T. M. Brown-Brandl, and B. C. Ramirez, "Development and application of an image acquisition system for characterizing sow behaviors in farrowing stalls," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104866.
- [16] D. D. Brown, R. Kays, M. Wikelski, R. Wilson, and A. P. Klimley, "Observing the unwatchable through acceleration logging of animal behavior," *Animal Biotelemetry*, vol. 1, no. 1, pp. 1–16, 2013.
- [17] M. Lorbach, E. I. Kyriakou, R. Poppe, E. A. van Dam, L. P. J. J. Noldus, and R. C. Veltkamp, "Learning to recognize rat social behavior: Novel dataset and cross-dataset application," *J. Neurosci. Methods*, vol. 300, pp. 166–172, May 2018.
- [18] Y. Bi and L. Wu, "Progress of animal behavior studies with optical motion capture system," in *Proc. 2nd Int. Conf. Biomed. Eng. Bioinf.*, Sep. 2018, pp. 102–105.
- [19] Z. Wang, S. A. Mirbozorgi, and M. Ghovanloo, "An automated behavior analysis system for freely moving rodents using depth image," *Med. Biol. Eng. Comput.*, vol. 56, no. 10, pp. 1807–1821, Oct. 2018.
- [20] A. B. Wiltschko et al., "Mapping sub-second structure in mouse behavior," *Neuron*, vol. 88, no. 6, pp. 1121–1135, Dec. 2015.
- [21] U. Stern, R. He, and C.-H. Yang, "Analyzing animal behavior via classifying each video frame using convolutional neural networks," *Sci. Rep.*, vol. 5, no. 1, pp. 1–13, Sep. 2015.
- [22] A. Arac, P. Zhao, B. H. Dobkin, S. T. Carmichael, and P. Golshani, "DeepBehavior: A deep learning toolbox for automated analysis of animal and human behavior imaging data," *Frontiers Syst. Neurosci.*, vol. 13, p. 20, May 2019.
- [23] L. Wu, Q. He, H. Cao, and L. Liu, "Deep learning on Parkinson's rat behavior point sets in open field," in *Proc. Int. Symp. Artif. Intell. Med. Sci.*, Sep. 2020, pp. 117–121.
- [24] J. Mallela et al., "Voice based classification of patients with amyotrophic lateral sclerosis, Parkinson's disease and healthy controls with CNN-LSTM using transfer learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 6784–6788.
- [25] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10. New York, NY, USA: ACM Digital Library, 1995, p. 1995.
- [26] G. Li, H. Wang, S. Zhang, J. Xin, and H. Liu, "Recurrent neural networks based photovoltaic power forecasting approach," *Energies*, vol. 12, no. 13, p. 2538, 2019.
- [27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [28] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.
- [29] Y. Zhang et al., "Epileptic seizure detection based on bidirectional gated recurrent unit network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 135–145, 2022.
- [30] S. Ding, C. Su, and J. Yu, "An optimizing BP neural network algorithm based on genetic algorithm," *Artif. Intell. Rev.*, vol. 36, no. 2, pp. 153–162, 2011.
- [31] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Netw. Appl.*, vol. 25, pp. 743–755, Dec. 2019.
- [32] H. Li and L. Wu, "EEG classification of normal and alcoholic by deep learning," *Brain Sci.*, vol. 12, no. 6, p. 778, Jun. 2022.
- [33] M. Sajjad et al., "A novel CNN-GRU-based hybrid approach for short-term residential load forecasting," *IEEE Access*, vol. 8, pp. 143759–143768, 2020.
- [34] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: Gradient boosting with categorical features support," 2018, *arXiv:1810.11363*.
- [35] W. Lou, X. Wang, F. Chen, Y. Chen, B. Jiang, and H. Zhang, "Sequence based prediction of DNA-binding proteins based on hybrid feature selection using random forest and Gaussian Naïve Bayes," *PLoS ONE*, vol. 9, no. 1, Jan. 2014, Art. no. e86703.
- [36] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, and H. Cho, "Xgboost: Extreme gradient boosting," *R Package Version* vol. 1, no. 4, pp. 1–4, Aug. 2015.
- [37] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.
- [38] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *Proc. OTM Confederated Int. Conf. Move Meaningful Internet Syst.* Cham, Switzerland: Springer, 2003, pp. 986–996.
- [39] T. Gaber, A. Tharwat, A. E. Hassanien, and V. Snasel, "Biometric cattle identification approach based on Weber's local descriptor and AdaBoost classifier," *Comput. Electron. Agricult.*, vol. 122, pp. 55–66, Mar. 2016.