Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

May 2023, Volume 1, Number 1

Early Parkinson's Disease Diagnosis Based on Sequence Analysis

Armin Salimi-Badr [a*] Code Orcid: 0000-0001-6613-7921, **Mohammad Hashemi** [b] Code Orcid: 0009-0005-1608-7213

[a] Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran, a_salimibadr@sbu.ac.ir

[b] Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran, mohamm.hashemi@mail.sbu.ac.ir

ABSTRACT

In this paper, a neural approach based on Long Short-Term Memory (LSTM) neural networks is proposed to diagnose patients suffering from Parkinson's Disease (PD). Considering the movement disorders caused by PD, the proposed method investigates the gait cycle pattern of subjects based on vertical Ground Reaction Force (vGRF) measured by 16 wearable sensors placed in subjects' shoes. In this study, it is shown that the temporal patterns of the gait cycle are different for healthy persons and patients. Therefore, by using a recurrent structure like LSTM, able to analyze the dynamic nature of the gait cycle, the proposed method extracts the temporal patterns to diagnose patients from healthy persons. To reduce the number of data dimensions, the sequences of corresponding sensors measuring vGRF in different feet are combined by subtraction. This method analyzes the temporal pattern of time series collected from different sensors, without extracting special features representing statistics of different parts of the gait cycle. Indeed, the method can extract temporal features based on learning, without using expert knowledge. Finally, the Accuracy and F1 Score of the model trained with all data is \$99.87\%\$, and \$96.66\%\$ respectively.

KEYWORDS

Parkinson's disease, gait cycle, Long Short-Term Memory.

1. INTRODUCTION

The symptoms of some chronic diseases, including the Parkinson's Disease (PD), are emerged and diagnosed many years after infection. This late disease diagnosis causes to decrease the effectiveness of some medical or non-invasive treatment methods. PD, the second most common neurological disease, is a slow and progressive neurodegenerative disorder that affects the human central nervous system by destroying dopaminergic neurons in the Substantia Nigra pars compacta, one nucleus of a complex subcortical brain structure named Basal Ganglia [1, 2, 3, 4, 5, 6]. It is estimated that the number of patients suffering from PD in the world is close to 5 million [4]. The usual symptoms are related to movement disorders including slowness (Bradykinesia), lack of the power of starting voluntary movements (Akinesia), rigidity, and the resting tremor [1]. It is reported that the mortality rate is 3.44% in Iran [7].

Recently Machine Learning methods are effectively applied to detect patients suffering from PD. Considering the movements' symptoms of PD, some previous methods study the speech disorders (vocal impairment) [8, 9, 5]and some other ones consider the gait cycle changing [4, 10]. The gait cycle includes stance and swing phases. In a normal and healthy situation, the stance phase composes 60% of the gait cycle. (Fig. 1) [4, 10]. A normal gait cycle of a healthy person is composed of a pattern that seems to be time-invariant during walking. However, during patients' walking the pattern of the gait cycle changes significantly [10].

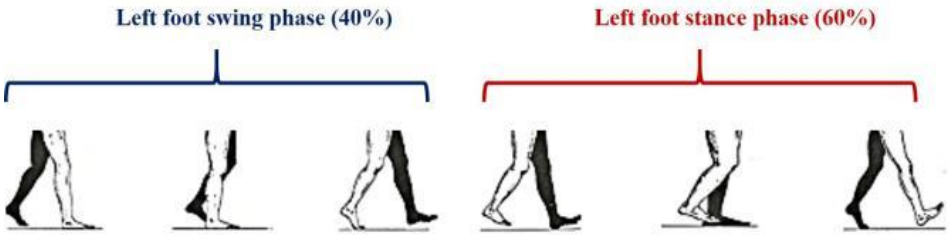


Figure 1: Gait cycle.

In this study, we have analyzed the gait cycle to for PD patient detection. The gait cycle pattern is recorded based on vertical Ground Reaction Force (vGRF) through time, by 16 wearable sensors situated in each shoe.

In [10], assuming the repetition of the gait cycle pattern during healthy persons' walking and changing during patients' walking, a method based on measuring the similarity of the time-series corresponding to stance phases is proposed. In this method, Continuous Dynamic Time Warping (CDTW) technique is used for measuring the similarity between gait cycles. Next, the obtained data is used as input of a binary classifier. Only data corresponding to the sum of 8 sensors is used in [10] for discriminating healthy persons from patients. Different supervised and unsupervised methods including K-Nearest Neighbors (KNN), Decision Tree, Support Vector Machine (SVM), Gaussian-Mixture Model (GMM) are utilized.

In [4], some human understandable statistical features measuring different aspects of the gait cycle are extracted from vGRF time-series. Indeed, based on a statistical analysis of each time-series a feature vector is extracted. Afterward, this feature vector is used as the input of a binary classifier to diagnose patients from healthy ones. Similar to [10], only two sequences corresponding to the sum of 8 sensors in each foot is used in [4]. Different machine learning algorithms are investigated for constructing the binary classifier.
In [11] a novel type-2 fuzzy neural network has been proposed to detect patients suffering from PD. This method has utilized the human understandable features to provide an interpretable solution. Authors have extracted some interpretable fuzzy rules for classifying patients. Although this method has a good interpretablity, its performance is lower than methods that apply deep learning approaches.
In [12] a temporal neural network model learned by a Q-Learning approach is proposed. This method detects anomalies in the gait cycle to detect patients based on analyzing the temporal information of vGRF time-series. The best accuracy of this model is 92.19%.
Recently deep learning methods are applied to detect patients suffering from PD. In contrary to the previous methods that use the human-designed features extracted from the vGRF time-series, these deep learning methods try to extract high-level features from these time-series. In [13], a convolutional neural network composed of one dimensional convolutional layers (1dCNN) is employed to detect patients. This 1dCNN reaches the final accuracy of 98.7%. In [14] and [15], a neural network consists of a combination of Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) are proposed. The best accuracy of these models is 98.61% for [14] and 99.22% for [15].
Generally, most previous methods utilize the sum of all sensors in each foot as the input data. Moreover, they tried to extract feature vectors based on some human knowledge, like gait cycle statistics or gait cycle repetition. In this study, we propose that the data from different sensors in either foot and their temporal patterns are useful for diagnosing patients. Furthermore, the spatial correlation among different sensors data during time is also important and is different between patients and healthy people. Therefore, in this study, we propose a neural approach based on analyzing the temporal and spatial patterns of different sensors data to classify patients and healthy persons. Recurrent Neural Networks (RNNs) are suitable tools to analyze sequences based on their recurrent structures [16]. Therefore, to consider the temporal dependency, the proposed structure has Long Short-Term Memory (LSTM) cell layers [17]. These LSTM cell layers encode the input sequence to a vector (sequence-to-vector architecture [18]). Next, the obtained vector is fed to some Dense layers to extract more compact and higher-level features based on applying different nonlinear maps. Finally, the output neuron is sigmoid as a binary classifier. This paper is the expanded version of our conference paper [19]. Comparing with the conference version [19], there are three main differences as follows:

1. More explanations about the method are provided;
2. More experiments are added;
3. Comparisons with the related studies are provided.

The rest of this paper is organized as follows: first the proposed method including data description, data preprocessing, and the proposed structure, is explained in section 2. Afterward, experimental results are presented in section 3. Finally, the paper is concluded in section 4.

2. PROPOSED METHOD

In this section, the proposed method and the applied preprocessing are presented.

2.1. Data and Proposed Structure

We utilized the gait cycle data provided by PhysioNet[1]. It is based on three similar data obtained in different experimental studies from different subjects [20, 21, 22]. The dataset includes vGRF obtained from 16 sensors placed under feet of control and patient subjects (8 per foot) as shown in Fig. 2. It contains data obtained from 93 patients with PD and 72 healthy control subjects. From each, there are more than one walking records in different situations. Thus, in total, there are 214 profiles of patients and 92 ones recorded from healthy control ones. Table 1 summarizes the details of these datasets.

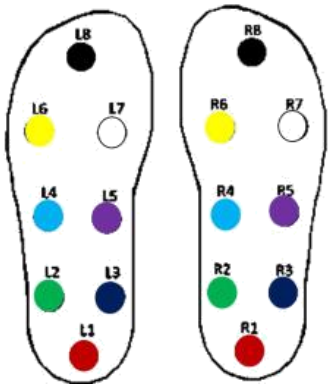


Figure 2: Placement of different sensors under feet.

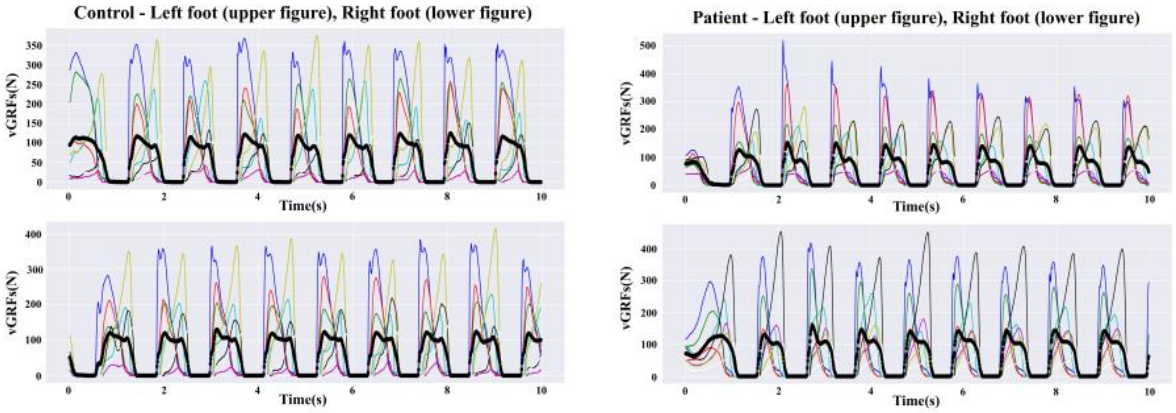
We show the outputs of each sensor for the left and right feet of one patient and one healthy control subject in Fig. 3 (different colors represent corresponding sensors in two feet, indicated in Fig. 2 and the solid black line in the middle of each figure is the

[1] https://physionet.org/content/gaitpdb/1.0.0/

mean value of all sensors outputs). It is shown that not only the temporal shape of different sensors outputs might change for a patient with PD, the correlations of these sequences from different sensors are also changed. For example, in a patient, the sequences of L1, L2, and L4 in the left foot and R1, R2, and R4 in the right foot are correlated and strong, while in a healthy person data, L3 and L8 from the left foot and R3 and R8 from the right foot are very strong. To properly analyze these sequences, we need a structure able to temporally (through time) and spatially (correlations among different sequences) analyze these sensors outputs. To provide spatial and temporal analysis we propose a neural structure based on recurrent layers (to temporal and spatial analysis through time) and forward layers (to extract higher-level features).

Table 1: Information of different datasets used for evaluation of the method (CO: Healthy Control Subject, PD: Patients with Parkinson's Disease).

Data	Subjects	Total subjects	Female	Male
Ga [21]	PD	29	9	20
	CO	18	8	10
Si [20]	PD	35	13	22
	CO	29	11	18
Ju [22]	PD	29	13	16
	CO	25	14	12



(a) (b)

Figure 3: Measured output sequences from different sensors of a healthy control subject and a patient. (a) Healthy, (b) Patient. The output sequences of the left foot is presented above the sequences belong to the right foot in each sub-figure. Different colors represent the output of corresponding sensors in each foot (in accordance with the indicated colors in Fig. 2). The black solid line in the middle of each figure is the calculated mean of outputs obtained from all sensors of each foot.

Our proposed structure is shown in Fig. 4. It is composed of four parts: 1- Input layer, presenting the input sequences from different sensors to the network; 2- LSTM layers, to analyze sequences temporally and spatially (correlation among different sequences through time); 3- Dense layers composed of neurons with ReLU nonlinear functions to extract features hierarchically; 4- Output layer consists of one neuron with sigmoid nonlinear function as a binary classifier. Therefore, the network receives the sequences as input, analyzes them through the time using LSTM cells and extracts a vector based on this analysis. Next, by applying hierarchical nonlinear maps, higher level features are extracted. The output binary classifier neuron classifies the input instance into two classes: 1- Healthy (output is equal to zero), and 2- Patient (output is equal to one). To learn the network's parameters, "Adam" algorithm [23] is used.
The loss function to train the network is the "binary cross-entropy" defined as follows [24]:

$$CE = -\frac{1}{N}\sum_{k=1}^{N} t_k \log y_k + (1 - t_k)\log(1 - y_k) \quad (1)$$

where N is the number of training instances, t_k is the desired label for k_{th} sample (0 or 1), and y_k is the network's output for k_{th} sample (between 0 and1).

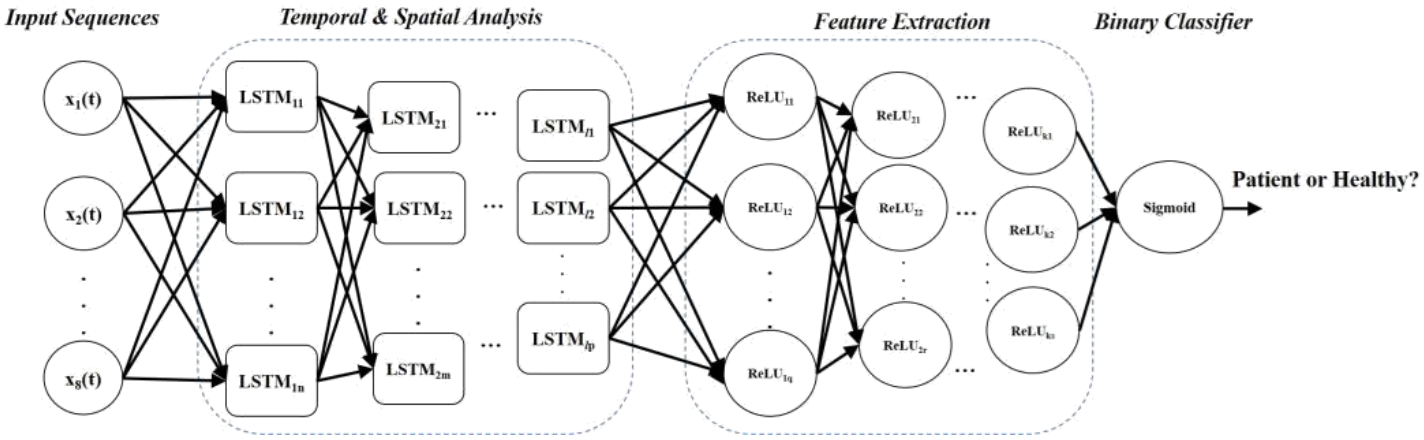


Figure 4: The proposed architecture for classifying patients and healthy persons by investigating the gait cycle based on temporal and spatial analysis of different sensors sequences.

2.2. Preprocessing

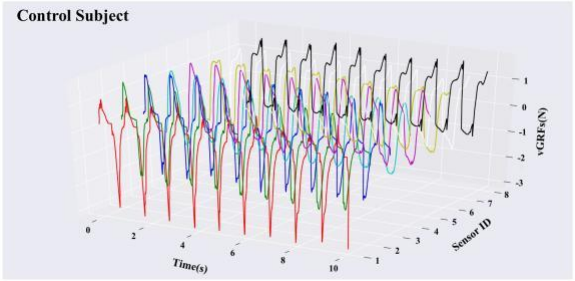
By looking at Fig. 3, it is obvious that each sequence includes a lot of zero values. These zero values are related to the swing phases of each foot. During this period, the other foot is in the stance phase and its sensors measure vGRF values greater than zeros. To reduce the number of input sequences and also remove these zero values, yet preserving the information of different sequences, we integrate the corresponding sensor's output sequences by subtracting the output of a sensor on the right foot from the output of the corresponding sensor on the left foot. Therefore, the number of sequences is reduced from 16 to 8.

In the provided dataset, each recorded instance corresponds to the walking of a subject in 2 minutes. We do not require all of these temporal samples and only a ten-second part of the sequence represents the gait cycle pattern. Therefore from each sensor output data, just a sequence related to 10 seconds of subject's walking is cropped and used as an input instance to the network. This ten-second split could start from every part of the walking and there would not exist any limitation on its start and finish time.

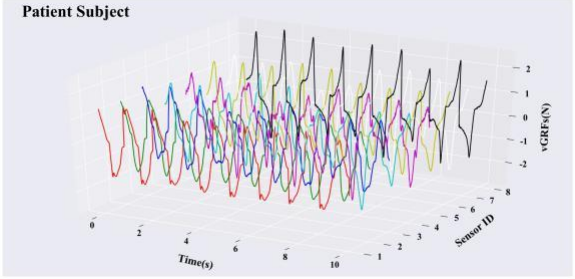
To normalize data, we normalize each sequence x as follows:

$$x_a = \frac{x - mean(x)}{std(x)} \tag{2}$$

This normalization approach does not memorize any parameter from the training set (like max and min values or mean and variance values in other normalization methods) to apply on new unseen instances. It can normalize each sequence based on its own mean and variance through time. Therefore, Fig. 5 compares two normalized and integrated instances belong to a patient and a healthy control subject.



(a)



(b)

Figure 5: The integrated sequences of different pairs of sensors by subtracting outputs of sensors placed on the right foot from the left one. (a) A healthy control subject, (b) A patient.

3. EXPERIMENTAL RESULTS

3.1. Evaluation Metrics

To evaluate the performance of the proposed structure on the data, the usual metrics utilized in the classification task are used. One of the most popular metrics in classification tasks is ”Accuracy” that shows the average performance of the method to properly classify instances and defined as follows [25, 26]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP (True Positive) is the number of instances belongs to the positive class (here PD patient) classified correctly as members of the positive class, TN (True Negative) is the number of instances belongs to the negative class (here healthy persons) classified correctly as members of the negative class, FP (False Positive) is the number of instances belongs to the negative class but classified wrongly as members of the positive class, and finally, FN (False Negative) is the number of instances belonging to the positive class but classified wrongly as members of the negative class.
Moreover, considering the different weight of effects of ”FP” rate versus ”FN” rate in our medical application the following metrics are also utilized [25]:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F_1 Score = \frac{2.Precision.Recall}{Precision + Recall} \tag{6}$$

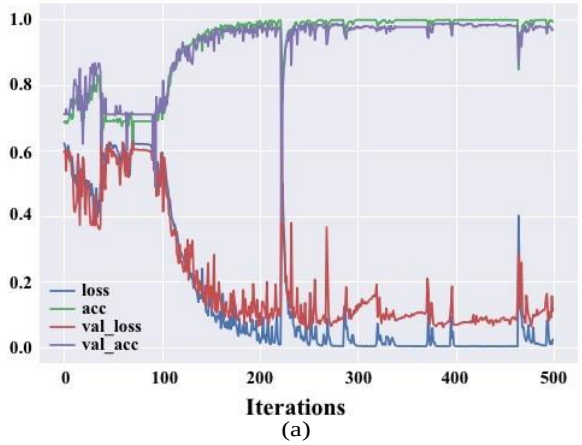
3.2. Experiments

The recorded data from each subject is divided into 10-second splits. These 10-second splits are used as the dataset to train and validate the proposed network. We choose randomly 70% of samples for training the network (2065 samples), 20% to test its generalization (585 samples), and finally 10% as validation data to avoid overfitting during training (300 samples).
To investigate the effect of depth level, the number of neurons in each layer, activation functions, proposed normalization approach, and both recurrent and feed-forward parts of the proposed architecture, we used 6 structures introduced in Table 2. All structures have one sigmoid neuron as the output layer for binary classification. Considering the number of sequences which is equal to 8, the number of neurons in different layers are set as multiplies of 8. To extract higher and more compact features by the network, the number of neurons decreases from the first hidden layer to the output layer. We investigate the effect of two non-saturating nonlinear functions: 1- Rectified Linear Unit (ReLU), and 2- Exponential Linear Unit (ELU). Since ReLU neurons may die, it is expected that the ELU function would be more accurate [27].
Based on the presented results of Table 2, we can summarize the following consequences:

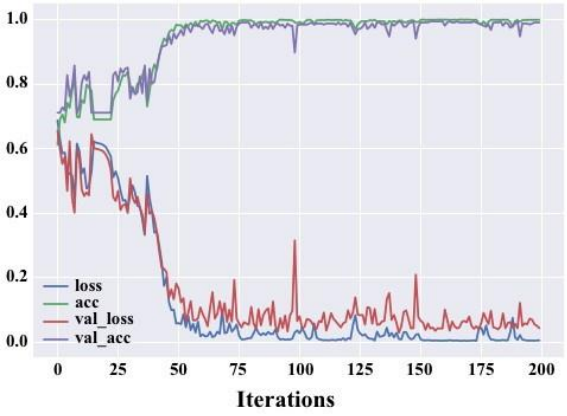
1. It is necessary to have dense layers for feature extraction comparing results obtained from 5th structure with the other ones’. Using just LSTM layers decrease the accuracy and precision significantly;
2. According to the results obtained from 6th architecture and comparing them with the other structures’, it is seen that the problem requires more than one LSTM layers to work properly;
3. By comparing results obtained from structures indexed as 1,2, and 4, we can conclude that using the proposed normalization leads to improving the performance of the method. Moreover, if we do not apply the normalization approach, to compensate for its effect we should add more dense layers (compare 2 with 4);
4. Finally, as we have expected, the structure with ELU performs better than the structure with ReLU (compare 2 with 3).
5.

Table 2: Comparison of different structures

Index	LSTM layers	Dense layers	Activation function	Normalization	Accuracy (%)			Precision (%)			Recall (%)		
					Training	Validation	Test	Training	Validation	Test	Training	Validation	Test
1	64, 32	16, 8	ReLU	No	70.21	71.66	70.59	69.88	71.76	70.75	99.78	99.06	98.77
2	64, 32	16, 8	ReLU	Yes	97.67	93.00	94.87	99.28	98.00	97.97	97.32	92.01	94.62
3	64, 32	16, 8	ELU	Yes	99.61	97.00	97.77	99.71	97.66	98.29	99.71	98.12	98.53
4	64, 32	16, 8, 4	ReLU	No	99.83	99.25	98.97	99.46	97.29	97.94	99.99	99.99	98.75
5	64,32	None	None	No	69.00	71.00	69.91	68.99	71.00	69.98	99.92	99.99	99.75
6	32	16, 8	ReLU	No	80.29	80.00	78.80	83.86	80.00	78.26	45.32	41.38	40.91



(a)



(b)

Figure 6: The changing of error and accuracy values during training for two structures: (a) using ELU nonlinear functions and normalization (indexed 3 in Table 2), (b) using ReLU nonlinear functions and does not use the normalization layer but has one more dense layer (indexed 4 in Table 2).

Fig. 6 shows the changing loss function and accuracy for training and validation datasets during the training of our two best structures indexed 3 and 4 in Table 2. The average F1 score is equal to 99.27 and 99.34 for index 3 and 4 respectively.

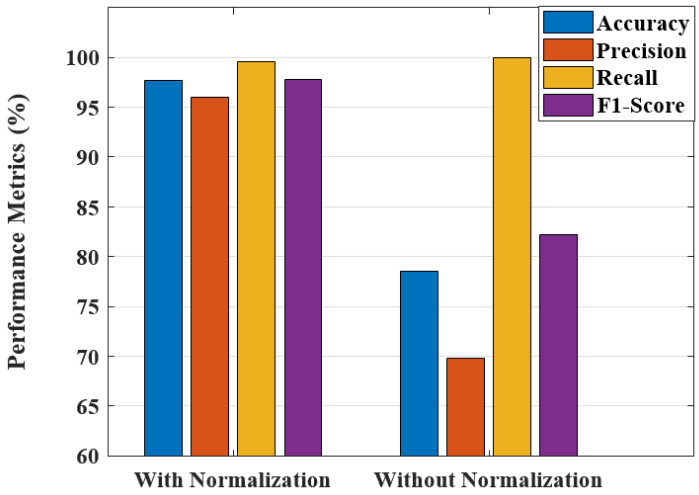


Figure 7: Comparison of the average performance by applying normalization and withoutusing it.

The results reported above are based on separating all samples to compare different structures. Indeed, it is possible that two different sequences from one person selected as the samples in two different datasets (one in training and the other in test). Therefore, up to now, the generalization ability of the model is investigated in classifying sequences. To investigate the generalization ability of the model in classifying persons instead of just sequences, we study the performance of the best model on data considering "persons". Indeed, we have randomly selected 80 patient profiles and 79 healthy control profiles for training, 6 patient profiles and 5 healthy control profiles for validation, and finally 9 patient profiles and 8 healthy control profiles for the test (to preserve all data balanced, we do not use some profiles related to the patients). Next, we have extracted ten-second sequences from each person's file to form different datasets. In this way, 1533 ten-second sequences are extracted for training, 92 ten-second sequences are extracted for validation, and 174 sequences are extracted for the test. In this experiment, different sequences belong to one person are considered only in one dataset. Therefore, it would be possible to investigate the

generalization ability of the model on persons. Table 3 compares the performance of the best structures with and without applying the proposed sample-wise normalization. It is shown that the normalization method is necessary to improve the generalization. The average "Accuracy" is 97.66% and the average F1 score is 97.78% in the case of applying the sample-wise normalization. Fig. 7 compares the average performance of the method in these two cases: using normalization and without it.

Table 3: Comparison of the best structure in classifying persons with and without applying the normalization

LSTM layers	Dense layers	Activation function	Normalization	Accuracy (%)			Precision (%)			Recall (%)		
				Training	Validation	Test	Training	Validation	Test	Training	Validation	Test
64, 32	16, 8, 4	ReLU	No	78.73	77.17	77.58	69.78	70.83	69.76	99.99	99.99	99.99
64, 32	16, 8, 4	ReLU	Yes	98.56	93.48	91.95	97.28	89.47	88.75	99.87	99.99	96.66

The performance of the final model is compared with some previous methods in Tables 4 and 5. Following the previous studies, the final model is trained with all data and its performance on all data is reported. Based on the reported results in these Tables, the proposed model outperforms the other deep learning and non-deep learning methods in terms of Accuracy, But its F1 Score is lower than [13] and [15]. The lower F1 Score and higher Accuracy shows that the model has higher FP or FN rate in comparison with [13] and [15]. Since, both previous studies have utilized CNNs in their structure, it could be concluded that the CNN that investigates patterns in a neighborhood has a better generalization than the dense structure in our model. Moreover, the proposed method has fewer parameters comparing to the other methods based on deep structures.

Table 4: Comparison of the final performance of the proposed method with some previous approaches (non-deep structures) on all data.

METHOD	**PROPOSED METHOD**	**[28]**	**[29]**	**[30]**	**[12]**	**[11]**
ACCURACY	99.87%	88.89%	84.48%	89.92%	91.53%	97.61%

Table 5: Comparison of the final performance of the proposed method and recent methodsbased on deep structures on all data.

Method	Reported Accuracy	Reported F1	Number of parameters
1D-CNN [13]	98.70%	98.90%	857120
CNN with LSTM (in parallel) [14]	98.61%	N/A	66505078
CNN with LSTM (sequentially)[15]	99.22%	98.04%	89201292

4. CONCLUSIONS

In this paper, we proposed a neural approach to classify patients with Parkinson's Disease (PD) and healthy persons based on analyzing the gait cycle pattern. Gait cycle data is obtained by 8 wearable sensors recording vertical Ground Reaction Force (vGRF), placed under each foot of a subject. The recorded data corresponding to 10 seconds of the subject's walking is presented as the input to the proposed method. Since the recorded data from different sensors behave differently for different healthy person and persons, gait cycle pattern recognition task requires analyzing all sequences temporally (considering correlations through time) and spatially (considering correlation among measured data of different sensors). To investigate sequences spatially and temporally, without windowing the input sequence, we proposed a neural structure based on recurrent cells layers (we used LSTM) for sequential analyzing and some dense layers to extract higher-level features.

To integrate sequences from corresponding sensors of different feet and remove extra zero values related to the swing phase of each foot, we subtract the right foot sensor data from the left one. This preprocessing leads to a sequence including the information of both feet simultaneously and finally reduces the input space dimensions. The performance of some different structures are investigated. It is shown that using the proposed normalization approach can lead to a decreasing False Positive rate and without this normalization, we need more dense layers to extract more abstract and higher-level features. The best accuracy and F1 score are equal to 97.66% and 97.78%. The best-reported accuracy corresponding to the state-of-the-art is about 97% [10, 4].

Finally, as the future trends of this study, we propose 1- to investigate the precision of this method to diagnose patients with similar diseases (like Huntington's Disease), 2- to study vocal impairment and gait cycle simultaneously, 3- to check the performance of the approach for data obtained from other sensors like IMU.

5. REFERENCES

[1] J. Hall, Guyton and Hall Textbook of Medical Physiology. Guyton Physiology, Elsevier Health Sciences, 2010.
[2] A. Salimi-Badr, M. M. Ebadzadeh, and C. Darlot, "A possible correlation between the basal ganglia motor function and the inverse kinematics calculation," Journal of Computational Neuroscience, vol. 43, no. 3, pp. 295–318, 2017.
[3] A. Salimi-Badr, M. M. Ebadzadeh, and C. Darlot, "A system-level mathematical model of basal ganglia motorcircuit for kinematic planning of arm movements," Computers in Biology and Medicine, vol. 92, pp. 78–89, 2018.
[4] N. Khoury, F. Attal, Y. Amirat, L. Oukhellou, and S. Mohammed, "Data-driven based approach to aid parkinson's disease diagnosis," Sensors, vol. 19, no. 2, p. 242, 2019.
[5] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of parkinson's disease," Computer methods and programs in biomedicine, vol. 113, no. 3, pp. 904–913, 2014.
[6] S. Pan, S. Iplikci, K. Warwick, and T. Z. Aziz, "Parkinson's disease tremor classification–a comparison between support vector machines and neural networks," Expert Systems with Applications, vol. 39, no. 12, pp. 10764–10771, 2012.
[7] S.-M. Fereshtehnejad, A. Shafieesabet, M. Shafieesabet, G. A. Shahidi, A. Delbari, and J. L¨okk, "Mortality in iranian patients with parkinson's disease: Cumulative impact of cardiovascular comorbidities as one major risk factor," Parkinson's Disease, vol. 2015, 2015.
[8] R. Das, "A comparison of multiple classification methods for diagnosis of parkinson disease," Expert Systems with Applications, vol. 37, no. 2, pp. 1568–1572, 2010.
[9] F. Astr¨om and R. Koker, "A parallel neural network approach to prediction of parkinson's disease," Expert systems with applications, vol. 38, no. 10, pp. 12470–12474, 2011.
[10] N. Khoury, F. Attal, Y. Amirat, A. Chibani, and S. Mohammed, "Cdtw-based classification for parkinson's disease diagnosis.," in ESANN, 2018.
[11] A. Salimi-Badr, M. Hashemi, and H. Saffari, "A type-2 neuro-fuzzy system with a novel learning method for parkinson's disease diagnosis," Applied Intelligence, vol. 53, 2023.
[12] Y. Nancy Jane, H. Khanna Nehemiah, and K. Arputharaj, "A q-backpropagated time delay neural network for diagnosing severity of gait disturbances in parkinson's disease," Journal of Biomedical Informatics, vol. 60, pp. 169–176, 2016.
[13] I. El Maachi, G.-A. Bilodeau, and W. Bouachir, "Deep 1d-convnet for accurate parkinson disease detection and severity prediction from gait," Expert Systems with Applications, vol. 143, p. 113075, 2020.
[14] A. Zhao, L. Qi, J. Li, J. Dong, and H. Yu, "A hybrid spatio-temporal model for detection and severity rating of parkinson's disease from gait data," Neurocomputing, vol. 315, pp. 1–8, 2018.
[15] X. Liu, W. Li, Z. Liu, 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. 7221–7232, Mar.
318 2021.
[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436–444, 2015.
[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[18] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning, MIT press Cambridge, 2016.
[19] A. Salimi-Badr and M. Hashemi, "A neural-based approach to aid early parkinson's disease diagnosis," in 2020 11th International Conference on Information and Knowledge Technology (IKT), pp. 23–25, 2020.
[20] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff, "Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease," Movement disorders: official journal of the Movement Disorder Society, vol. 20, no. 9, pp. 1109–1114, 2005.
[21] G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon, and J. M. Hausdorff, "Dual tasking, gait rhythmicity, and parkinson's disease: which aspects of gait are attention demanding?," European journal of neuroscience, vol. 22, no. 5, pp. 1248–1256, 2005.
[22] J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi, "Rhythmic auditory stimulation modulates gait variability in parkinson's disease," European Journal of Neuroscience, vol. 26, no. 8, pp. 2369–2375, 2007.
[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
[24] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in Advances in neural information processing systems, pp. 8778–8788, 2018.
[25] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.
[26] S. S. Sadeghi and M. R. Keyvanpour, "Rusdr: Class imbalance-aware ensemble learning for drug repurposing," in IKT, 2019.
[27] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," arXiv preprint arXiv:1511.07289, 2015.
[28] O. F. Ertugrul, Y. Kaya, R. Tekin, and M. N. Almali, "Detection of parkinson's disease by shifted one dimensional local binary patterns from gait," Expert Systems with Applications, vol. 56, pp. 156–163, 2016.
[29] Y. Wu, P. Chen, X. Luo, M. Wu, L. Liao, S. Yang, and R. M. Rangayyan, "Measuring signal fluctuations in gait rhythm time series of patients with parkinson's disease using entropy parameters," Biomedical Signal Processing and Control, vol. 31, pp. 265–271, 2017.
[30] M. R. Daliri, "Automatic diagnosis of neuro-degenerative diseases using gait dynamics," Measurement, vol. 45, no. 7, pp. 1729–1734, 2012.

Armin Salimi-Badr received the B.Sc., M.Sc. and PhD degrees in Computer Engineering, all from Amirkabir University of Technology, Tehran, Iran in 2010, 2012, and 2018 respectively. He also obtained a PhD degree in Neuroscience from University of Burgundy, Dijon, France in 2019, where he was researching on presenting a computational model of brain motor control in the Laboratory 1093 CAPS (Cognition, Action, et Plasticité Sensorimotrice) of the Institut National de la Santé et de la Recherche Médicale (INSERM). He was a Postdoctoral Research Fellow at Biocomputing lab of Amirkabir University of Technology from October 2019 to September 2020. Currently, he is an Assistant Professor at Faculty of Computer Science and Engineering of Shahid Beheshti University, Tehran, Iran. He is also the founder and Chair of *Robotics Intelligent Autonomous Agents (RoIAA)* Lab in Shahid Beheshti University. He is also a *Senior Member* of IEEE and currently he is the Chair of *Professional Activities Committee* and a Board Member of *Computer Society* of IEEE Iran Section. His research interests include fuzzy neural networks, evolutionary algorithms, computational intelligence, machine learning, and robotics.

Mohammad Hashemi is currently a senior undergraduate computer engineering student at Shahid Beheshti University (National University), Tehran, Iran. His main research interests are Computer Vision, Deep Learning, and Machine Learning.