Feature Selection Based on Euclid Distance and Neuro-fuzzy System

Seok-Woo Jang
Department of Software, Anyang University, Anyang-si, Republic of Korea
Email: swjang7285@gmail.com

Sang-Hong Lee
Department of Computer Science & Engineering, Anyang University, Anyang-si, Republic of Korea
Email: shleedosa@gmail.com

***Abstract*—This article suggests the method to distinguish normal persons and a Parkinson's disease patients by their sole pressure sensor data using NEWFM (Neural Network with Weighted Fuzzy Membership Functions). To make the features to be used as initial input data of NEWFM, the left and right sole pressure sensor data were extracted at the 1st step. In the 2nd step, the frequency scales of the characteristics extracted in the 1st step were divided into individual scales by the FFT (Fast Fourier Transform) using the Hamming method. In the final step, 1 to 15 dimensions were extracted as the characteristics from the values of the individual frequency scales produced in the 2st step by the PCA (Principal Component Analysis). The 75.90% in accuracy performance was acquired from the 8 dimensions with the highest performance, using them as the characteristics.**

***Index Terms*—Parkinson's disease, gait, FFT, PCA, NEWFM**

I. Introduction

Parkinson's disease, one of the representative degenerative brain diseases, is caused by the lack of dopamine, a neurotransmitter secreted from substantia nigra of the brain, and its clinical characteristics include resting tremor, rigidity, bradykinesia, postural instability, etc. [1]-[8]. To diagnose Parkinson's disease based on bradykinesia, analysis of the figure motion [4] and gate pattern [5]-[10] have been used. In addition, computer keystroke [11], speech signal processing [12], [13], and Handwriting [14] have been used to diagnose Parkinson's disease. In preprocessing method, phase space reconstruction [9], wavelet transform [13], and recurrence plot [15] have been used to diagnose Parkinson's disease too.

In this article, normal persons and Parkinson's disease patients were distinguished by their sole pressure sensor data using NEWFM [16]-[19]. To make the characteristics to be used as the initial input data of NEWFM, the left and right sole pressure sensor data were extracted at the 1st step. In the 2nd step, the frequency scales of the characteristics extracted in the 1st step were divided into individual scales by the FFT using the Hamming method. In the final step, 1 to 15 dimensions were extracted as the characteristics from the values of the individual frequency scales produced in the 2nd step by the PCA.

In this article, 1 to 15 dimensions were used as the characteristics and the performance of each characteristic was acquired. The highest performance was acquired when 8 dimensions among them were used as the characteristics. 8 fuzzy membership functions were suggested for the 8 characteristics in this article so that the characteristics could be interpreted [16]-[19].

II. Experimental Data and Preprocessing

A. Experimental Data

In this article, normal people and Parkinson's disease patients were distinguished using the experimental data acquired from PhysioBank. The experimental data were acquired from the sensors on the sole, 8 for each side, of 93 Parkinson's disease patients (mean age: 66.3 years; 63% men) and 73 normal persons (mean age: 66.3 years; 55% men) shown in Fig. 1. 92 experimental group data were acquired from 73 persons and 214 experimental group data were acquired from the 93 Parkinson's disease patients. The experimental data of about 2 minutes was acquired in the individual experimental group data. The frequency of the acquired experimental group data was 100 Hz.

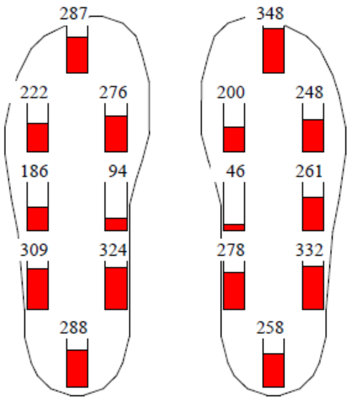

Figure 1. Pressure data collected through the sensors on the sole.

Manuscript received January 27, 2020; revised June 18, 2020.

doi: 10.12720/jait.11.3.155-160

TABLE I. EXPLANATIONS ABOUT THE MEASURED DATA

Input Order	Explanation
The first input	Time
The second to the ninth inputs	8 data collected from the sensors on the left sole shown in Fig. 1
The tenth to the seventeenth inputs	8 data collected from the sensors on the right sole shown in Fig. 1
The eighteenth input	The entire input data collected from the left sole
The nineteenth input	The entire input data collected from the right sole

Table I shows that the data acquired at one time include total 19 inputs. This article used the sum of the 18th input and the 19th input for the experiment. The experimental group data shown in Table II was acquired from the sum of the 18th input and the 19th input and they consisted of 2048 sums of the 18th input and the 19th input. For the experiment in this article, 420 data of the health persons and 974 data of the Parkinson's disease patients were allotted to the training set and the test set in the ratio of 5:5, as shown in Table II. Fig. 2 shows the difference of sole pressure sensor data of normal people and Parkinson's disease patients.

TABLE II. THE EXPERIMENTAL GROUPS USED FOR THE IDENTIFICATION OF THE PARKINSON'S DISEASE (IN THE RATIO OF 5:5)

Class	Training Set	Test Set	Total Number
Parkinson's disease	487	487	974
Normal	210	210	420
Total Number	697	697	1394

Figure 2. Graph of pressure data collected through the sensors on the sole.

Fig. 3 shows the big difference of sole pressure sensor data of normal people and Parkinson's disease patients by phase space reconstruction (PSR). The sole pressure sensor data of normal people is closer than that of Parkinson's disease patients from square point in Fig. 3.

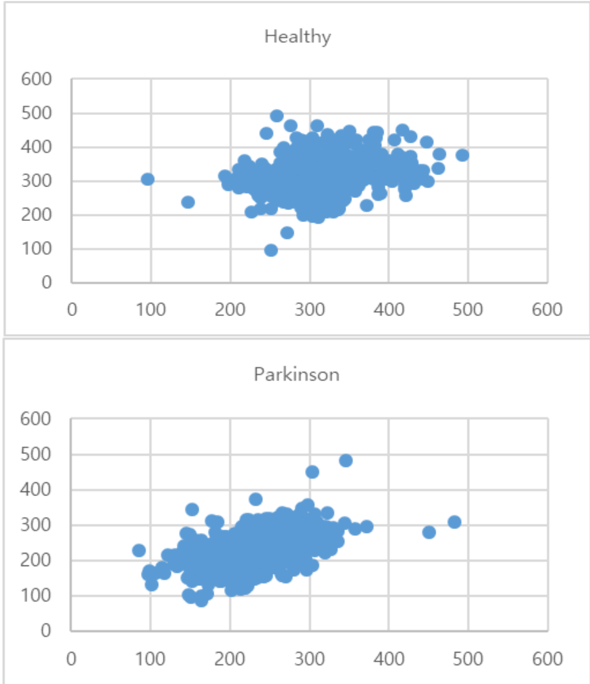

Figure 3. Graph of pressure data collected through the sensors on the sole by PSR.

B. *Feature Selection Based on Euclid Distance*

FFT (fast Fourier transform) was made as a method for rapid generation DFT (discrete FT). FFT was made to lessen the number of calculations of DFT and it can be considered as a method to make DFT rapidly by removing the repeated calculation in the DFT equations. FFT can be divided into the methods that separate the time domain and the frequency domain where N is a number of instances. $f(i,c)$ is ith feature with class 1 or 2. The Hamming method-based FFT was carried out to disassemble the 2048 data that constitute the individual groups shown in Table II into 1024 frequency domains.

$$ED(i) = \left| \frac{\sum f(i,1)}{N} - \frac{\sum f(i,2)}{N} \right| \quad (1)$$

where N is a number of instances. $f(i,c)$ is ith feature with class 1 or 2

C. *Principal Component Analysis*

The purpose of PCA (principal component analysis) to explain the whole change with *m* principal components by consisting of the principal components by means of the first-order association of the *p* given (measured) parameters and ranking the component as the 1st component, the 2nd component, …, the *m*th component in the order of how much each component contributes to the explanation of the entire change. In this article, 1 to 15 components were extracted by PCA from the 1024 characteristics that were generated by FFT.

D. Overview of Classifying Normal Persons and a Parkinson's Disease Patients

Fig. 4 shows a process for classifying normal persons and a Parkinson's disease patients. As shown in the diagram in Fig. 4, NEWFM is used to classify normal persons and a Parkinson's disease patients as the classifier.

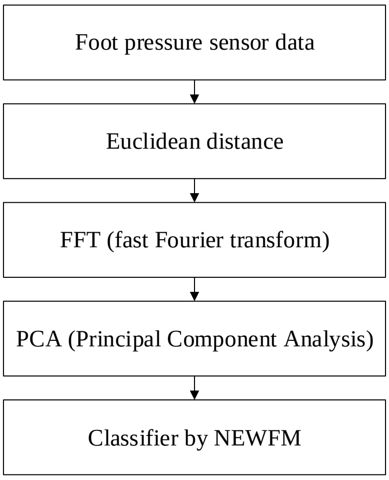

Figure 4. Diagram of model proposed for classifying normal persons and a Parkinson's disease patients by sole pressure sensor data.

III. NEURAL NETWORK WITH WEIGHTED FUZZY MEMBERSHIP FUNCTION (NEWFM)

NEWFM is a kind of fuzzy neural networks using the Bounded Sum of Weighted Fuzzy Membership functions (BSWFMs) [16]-[19]. Fig. 5 explains the structure of the NEWFM that is composed of three layers (input, hyperbox, and the class layer). An hth input can be used as I_h={A_h=(a_1, a_2, … , a_n), *class*}, where *class* is classification node and A_h is n features of an input. 1 to 15 dimensions extracted by the PCA were used as inputs in Fig. 5.

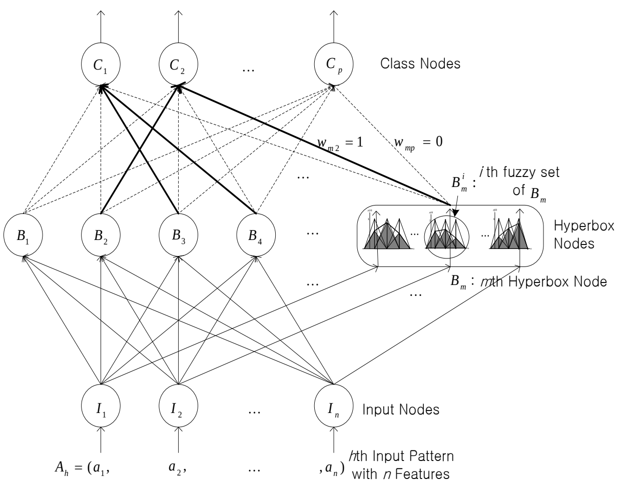

Figure 5. Structure of NEWFM.

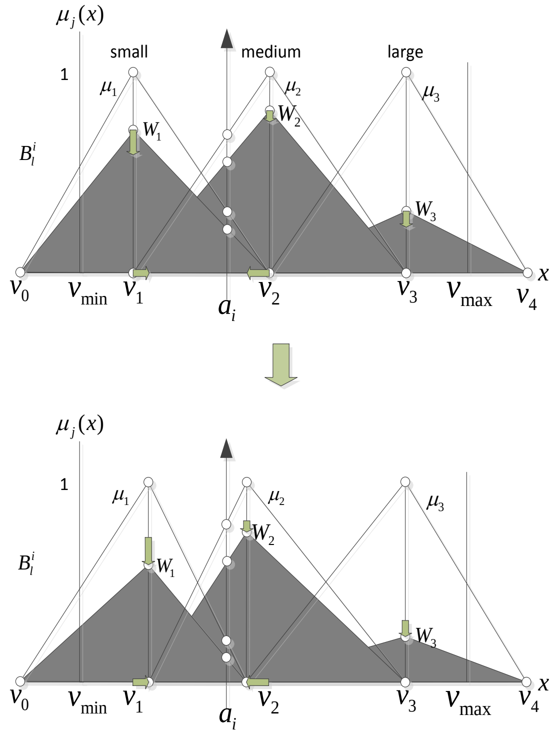

Figure 6. Example of before and after *Adjust*(B_l) operation.

The Adjust(B_l) operation adjusted the weights and the center of membership functions in Fig. 6. W_1, W_2, and W_3 are moved up or down, v_1 and v_2 are moved up to a_i, and v_3 stays in the same position [17]. After finishing Adjust(B_l), each of all fuzzy sets in hyperbox node B_l in Fig. 5 contains three *Weighted Fuzzy Membership functions* (*WFMs*). The WFM means grey membership functions in Fig. 7. The *bounded sum* of WFMs (BSWFM) in the ith fuzzy set of $B_l^i(x)$ denoted as $\mu_b^i(x)$ defined by:

$$\mu_b^i(x) = \sum_{j=1}^{3} B_l^i(\mu_j(x)). \tag{2}$$

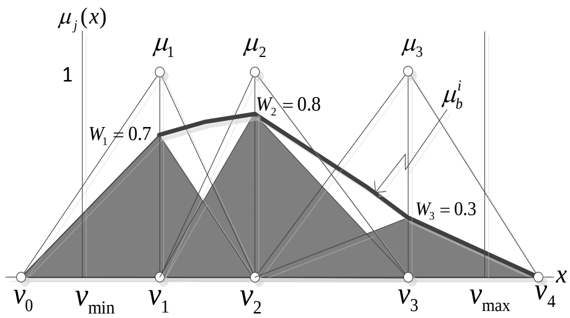

Figure 7. Example of the 3 BSWFMs.

The BSWFM means bold line in Fig. 7. The two BSWFMs graphically show the difference between a normal person and Parkinson's disease patient for each input feature.

IV. Experimental Results

This article carried out the Hamming method-based FFT to extract 1024 characteristics for each frequency domain from the 2048 initial sole pressure sensor data. From the 1024 extracted characteristics, 1 to 15 characteristics were extracted by PCA. The 1 to 15 characteristics were used as the input data to distinguish Parkinson's disease patients from normal persons.

Fig. 8 shows the examples of fuzzy membership functions on the 15 dimensions among the 1 to 15 dimensions produced by PCA. These describes the BSWFM in [16]. The difference in the sole pressure sensor data of the normal persons and the Parkinson's disease patients could be visualized and analyzed accordingly.

This article compared performance of NEWFM with that of Backpropagation (BP), Support Vector Machine (SVM), and Multilayer Perception (MP). In Equation (1), TP (True Positive) means the cases where the sole pressure of the Parkinson's disease patient was identified as that of the Parkinson's disease patient and TN means the cases where the sole pressure of the normal person is identified as that of the normal person. On the contrary, FP (False Positive) means the cases where the sole pressure of the Parkinson's disease patient was identified as that of the normal person and FN (False Negative) means the cases where the sole pressure of the normal person as that of the Parkinson's disease patient. Table III, Table IV, Table V, and Table VI show the performance results, and Table VII, Table VIII, Table IX, and Table X show the accuracy, specificity and sensitivity defined in Equation (3).

$$Sensitivity = \frac{TP}{TP + FN} \times 100$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100$$

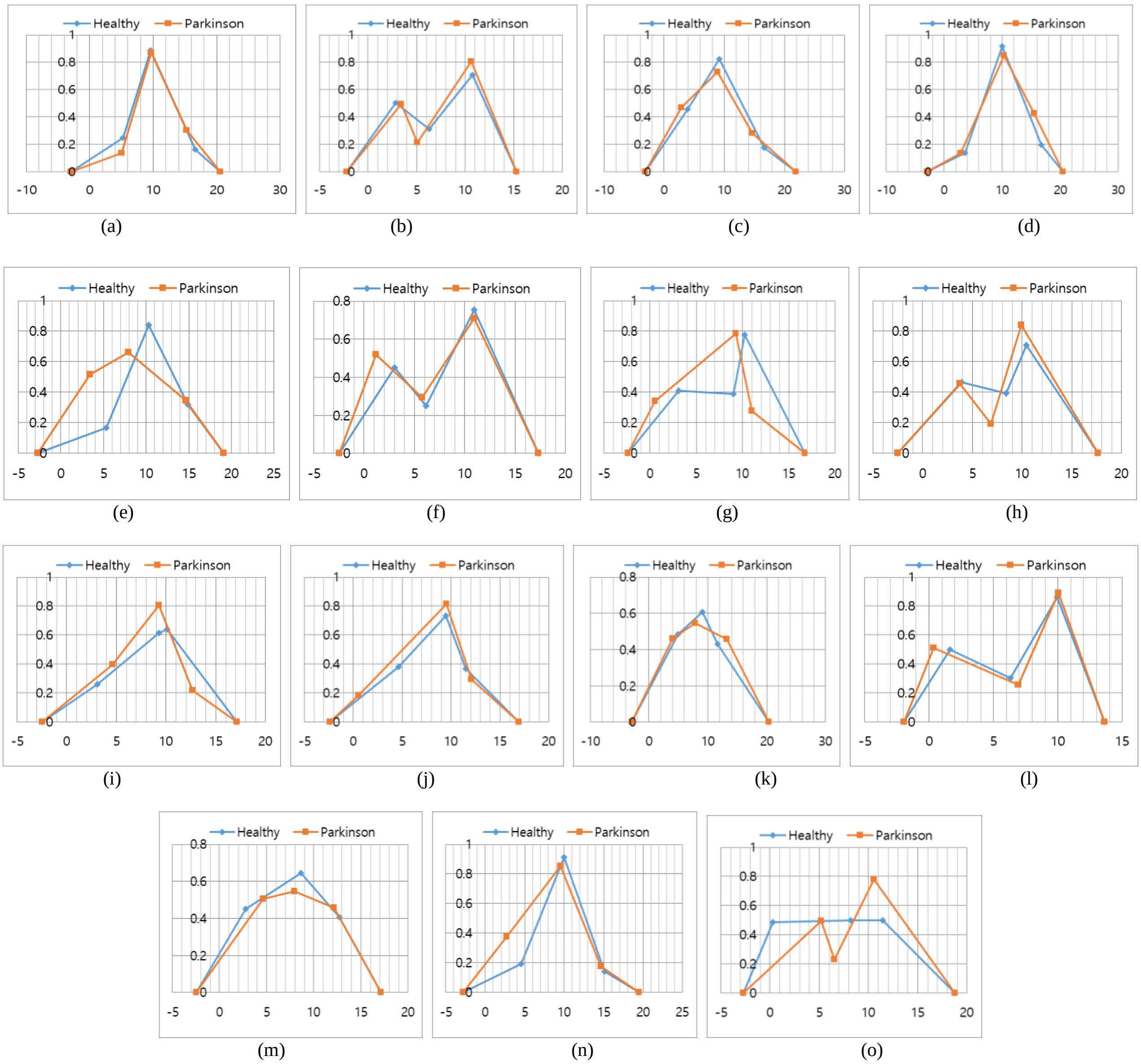

Figure 8. Examples of the BSWFM of the 15 features.

TABLE III. PERFORMANCE RESULTS WITH BP

The sole pressure of the Parkinson's disease patient	TP	FN
	364	123
The sole pressure of the normal person	FP	TN
	95	115

TABLE IV. PERFORMANCE RESULTS WITH SVM

The sole pressure of the Parkinson's disease patient	TP	FN
	332	155
The sole pressure of the normal person	FP	TN
	100	110

TABLE V. PERFORMANCE RESULTS WITH MP

The sole pressure of the Parkinson's disease patient	TP	FN
	307	180
The sole pressure of the normal person	FP	TN
	76	134

TABLE VI. PERFORMANCE RESULTS WITH NEWFM

The sole pressure of the Parkinson's disease patient	TP	FN
	416	71
The sole pressure of the normal person	FP	TN
	97	113

TABLE VII. PERFORMANCE RESULTS WITH BP

Epochs	Number of Hidden Nodes	Accuracy	Specificity	Sensitivity
5000	4	63.99	62.86	64.48
	6	64.85	61.90	66.12
	8	66.66	57.62	70.43
	10	66.28	58.57	69.61
10000	4	65.14	55.24	69.40
	6	68.58	58.57	72.90
	8	**68.72**	**54.76**	**74.74**
	10	66.71	50.48	73.72
15000	4	66.28	53.33	71.87
	6	68.58	59.05	72.69
	8	66.14	51.90	72.28
	10	66.57	48.57	74.33
20000	4	66.68	54.33	71.89
	6	68.88	59.55	72.79
	8	66.54	52.10	72.38
	10	66.67	48.67	74.23

(Learning rate: 0.01, momentum: 0.7)

TABLE VIII. PERFORMANCE RESULTS WITH SVM

	Accuracy	Specificity	Sensitivity
Performance (%)	63.41	41.5	76.85

(Step size: 0.01, output layer: 2, threshold: 0.1)

TABLE IX. PERFORMANCE RESULTS WITH MP

	Accuracy	Specificity	Sensitivity
Performance (%)	63.27	42.68	80.16

(Hidden layers: 3, momentum: 0.7)

TABLEX. PERFORMANCE RESULTS WITH NEWFM

	Accuracy	Specificity	Sensitivity
Performance (%)	75.90	61.41	81.09

V. CONCLUDING REMARKS

In this article, the Hamming method-based FFT was carried out with the sole pressure sensor data of normal persons and Parkinson's disease patients to extract 1024 characteristics for each frequency domain. The generated 1024 characteristics were reduced to 8 dimensions by means of PCA and they were used as the input data for NEWFM. The method proposed in this article can distinguish normal persons and Parkinson's disease patients using their sole pressure sensor data in real-time. NEWFM shows the characteristics of 8 features with the fuzzy membership functions by deriving their BSWFMs. The accuracy of 75.90% was acquired from the 8 dimensions with the highest performance, using their BSWFMs as the characteristics.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

The first author, Seok-Woo Jang, wrote the program and the manuscript for the paper.

The corresponding author, Sang-Hong Lee, conducted the main research design of feature selection for the paper.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No. NRF-2019R1F1A1055423).

REFERENCES

[1] W. C. Koller, *et al.*, "Falls and Parkinson's disease," *Clin Neuropharmacol*, vol. 12, pp. 98-105, 1989.

[2] D. W. Kim, K. Lee, D. Lee, and K. H. Lee, "A k-populations algorithm for clustering categorical data," *Pattern Recognition*, vol. 38, no. 7, pp. 1131-1134, 2005.

[3] M. E. Morris, "Movement disorders in people with Parkinson disease: A model for physical therapy," *Phys. Ther.*, vol. 80, pp. 578-597, 2000.

[4] Á. Jobbágy, P. Harcos, R. Karoly, and G. Fazekas, "Analysis of finger-tapping movement," *Journal of Neuroscience Methods*, vol. 141, pp. 29-39, 2005.

[5] C. Lee, *et al.*, "Dynamic foot pressure measurement in Parkinson's disease with foot scan system," *Journal of Korean Neurological Association*, vol. 25, no. 2, 2007.

[6] Y. Yang, Y. Lee, S. Cheng, P. Lin, and R. Wang, "Relationships between gait and dynamic balance in early Parkinson's disease," *Gait & Posture*, vol. 27, pp. 611-615, 2008.

[7] M. D. Lewek, R. Poole, J. Johnson, O. Halawa, and X. Huang, "Arm swing magnitude and asymmetry during gait in the early stages of Parkinson's disease," *Gait & Posture*, vol. 31, pp. 256-260, 2010.

[8] K. K. Patterson, W. H. Gage, D. Brooks, S. E. Black, and W. E. McIlroy, "Evaluation of gait symmetry after stroke: A comparison of current methods and recommendations for standardization," *Gait & Posture*, vol. 31, pp. 241-246, 2010.

[9] W. Zeng, C. Yuan, Q. Wang, F. Liu, and Y. Wang, "Classification of gait patterns between patients with Parkinson's disease and healthy controls using Phase Space Reconstruction (PSR), Empirical Mode Decomposition (EMD) and neural networks," *Neural Networks*, vol. 111, pp. 64-76, 2019.

[10] F. Liu, Q. Wang, Y. Wang, and Y. Zhang, "Parkinson's disease classification using gait analysis via deterministic learning," *Neuroscience Letters*, vol. 633, pp. 268-278, 2016.

[11] T. D. Pham, "Pattern analysis of computer keystroke time series in healthy control and early-stage Parkinson's disease subjects using fuzzy recurrence and scalable recurrence network features," *Journal of Neuroscience Methods*, vol. 307, pp. 194-202, 2018.

[12] D. Braga, A. M. Madureira, L. Coelho, and R. Ajith, "Automatic detection of Parkinson's disease based on acoustic analysis of speech," *Engineering Applications of Artificial Intelligence*, vol. 77, pp. 148-158, 2019.

[13] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, and H. Apaydin, "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255-263, 2019.

[14] A. Ammour, I. Aouraghe, G. Khaissidi, M. Mrabti, and F. Belhacen, "A new semi-supervised approach for characterizing the Arabic on-line handwriting of Parkinson's disease patients," *Computer Methods and Programs in Biomedicine*, 2019.

[15] L. C. S. Afonso, G. H. Rosa, C. R. Pereira, S. A. T. Weber, and J. P. Papa, "A recurrence plot-based approach for Parkinson's disease identification," *Future Generation Computer Systems*, vol. 94, pp. 282-292, 2019.

[16] J. S. Lim, "Finding features for real-time premature ventricular contraction detection using a fuzzy neural network system," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 522-527, 2009.

[17] J. S. Lim, D. Wang, Y. S. Kim, and S. Gupta, "A neuro-fuzzy approach for diagnosis of antibody deficiency syndrome," *Neurocomputing*, vol. 69, pp. 969-974, 2006.

[18] S. Lee, "Feature selection based on the center of gravity of BSWFMs using NEWFM," *Engineering Applications of Artificial Intelligence*, vol. 45, pp. 482-487, 2015.

[19] S. Lee, J. S. Lim, J. Kim, J. Yang, and Y. Lee, "Classification of normal and epileptic seizure EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance," *Computer Methods and Programs in Biomedicine*, vol. 116, pp. 10-15, 2014.

Seok-Woo Jang received the B.S., M.S., Ph.D. degrees in Computer Science from Soongsil University, Seoul, Korea, in 1995, 1997, and 2000, respectively. From October 2003 to January 2009, he was a Senior Researcher with the Construction Information Research Department at Korea Institute of Construction Technology (KICT), Ilsan, Korea. Since March 2009, he has been a Professor in the Department of Software, Anyang University, Korea. His primary research interests include robot vision, augmented reality, video indexing and retrieval, cluster computing, biometrics and pattern recognition.

Sang-Hong Lee received the B.S., M.S., and Ph.D. degrees in computer science from Gachon University, Korea in 1999, 2001, and 2012, respectively. He is currently an assistant professor in the department of computer engineering at Anyang University, Korea. His research focuses on neuro-fuzzy systems, stocks prediction systems, and biomedical prediction systems.