Freezing of Gait Detection Using Discrete Wavelet Transform and Hybrid Deep Learning Architecture

1st Nguyen Thi Hoai Thu
School of Electronic and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
thunguyen@knu.ac.kr

2nd Dong Seog Han
School of Electronic and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
dshan@knu.ac.kr

***Abstract*—Freezing of gait (FoG) detection using wearable sensors plays an important role in both online and offline monitoring of Parkinson's disease patients. In a FoG detector, feature extraction is commonly considered as a critical part for distilling the sensor signals before the FoG classification. Manually extracted features with domain knowledge are widely used in conventional machine learning methods while recent deep learning algorithms introduce the automatic feature learning approach. In this paper, we propose a FoG detection framework, in which hand-crafted features are used as input to a hybrid deep learning model for further feature learning and classification task. The hand-crafted features with time-frequency representation are extracted from the raw sensor signal by using a multi-level discrete wavelet transform (DWT). A hybrid deep learning architecture constructed from two algorithms: convolutional neural network (CNN) and bidirectional long short-term memory network is then deployed to extract deep features and classify FoG events. For performance comparison purposes, experiments on different input data types and machine learning methods are carried out on the Daphnet public dataset.**

***Index Terms*—freezing of gait, deep learning, wearable sensors, discrete wavelet transform**

I. Introduction

Freezing of gait (FoG) is a gait disorder that can suddenly happen in short episodes at Parkinson's disease (PD) patients and causes the patients to be unable to move the feet forward in spite of their intention [1]. FoG can cause impaired balance control, thus give rise to the risk of falls and reduce the functional independence. In order to bring down the risk factors of FoG, various treatment methods have been studied, including both pharmacological and non-pharmacological. A common non-pharmacological strategy is using a rhythmic auditory stimulation system [2] or rhythmic auditory cues and visual cues [3] to help the patient return to a normal gait. In these systems, a context-aware wearable sensor-based FoG detector is deployed to online recognize the FoG events, and the stimulator will start providing a rhythmic ticking sound when a FoG event is detected [4]. In addition to online recognition of FoG, an off-line FoG detector also plays an important role in helping PD patients and clinicians quantitatively assess the PD associated signs [5].

With the significant contribution in the therapy and assistance to PD patients, there has been a growing interest in wearable sensor-based freezing of gait detection. With the observation that during FoG, due to the 'misfiring oscillators', the legs are forced to move too fast, Moore *et al*. [6] have used the frequency characteristics extracted from the vertical leg movement to help the detection of the FoG events. Authors in [4] use a hand-crafted method to extract a set of features from mobile sensor data in both time and frequency domains, then apply different machine learning algorithms (e.g., random forest, k-nearest neighbor) to detect FoG events from the extracted features. Similarly, El-Attar *et al*. [7] deploy discrete wavelet transform (DWT) for feature extraction and apply two machine learning techniques (*i.e.*, support vector machine, artificial neural network) for freezing detection. Bächlin *et al*. [8] use the fast Fourier transform (FFT) to calculate the power spectral density (PSD) from sensor signals. An energy threshold and a freeze index threshold are then applied to the extracted PSD for FoG detection. These methods have gain reasonable performance but are highly dependent on human domain knowledge such as the freeze index and freeze threshold.

In recent years with the rising of deep learning algorithms, great effort has been devoted to the study of applying deep learning (DL) to FoG detection. One of the first studies that applied the deep learning method for detecting FoG is proposed in [9]. The authors implement an 1-D convolutional neural network (CNN) and train it on the spectral data representation of two adjacent signal windows. Zeng and Gao *et al*. [10] propose a continuous attention-based long short-term memory network (LSTM) to adaptively focus on essential parts of the signal, thus, help to improve the detection accuracy. Bikias *et al*. [11] propose a model called DeepFoG structured from a CNN to automatically extract important features and detect FoG from a wrist-worn inertial measurement unit (IMU).

In this paper, in order to leverage the advantages of the two approaches: hand-crafted feature extraction and deep learning-based feature extraction, we propose a hybrid deep learning model that is deployed on the time-frequency domain of the sensor signal. First, a multi-level discrete wavelet transform

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2020-0-01808) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation); and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144).

is applied to extract the time-frequency presentation from the raw signal. Then, a model constructed from two deep learning algorithms: convolutional neural network (CNN), bidirectional long short-term memory (BiLSTM) is used to further extract deep features automatically and perform the classification task.

The rest of the paper is organized as follows: Section II describes the multi-level discrete wavelet transform and structure of the hybrid deep learning network. Experiments on a public dataset and comparison results are shown in section III. Finally, section IV concludes the contribution of this paper and some notes for future work.

II. Proposed FoG Detection Framework

The proposed framework consists of two main parts: 1) multi-level discrete wavelet transform (DWT), 2) hybrid deep learning model. Firstly, a time-frequency representation is extracted from the raw acceleration signal by using a multi-level wavelet decomposition. Secondly, the extracted representation is fed into a hybrid deep learning network which contains a CNN and BiLSTM. During the training process, the network will automatically learn spatial and temporal features from the time-frequency representation. The detailed architecture of our proposed FoG detection framework is shown in Fig. 1.

A. Multi-level Discrete Wavelet Transform

By analyzing the power spectrum of acceleration signal, it was found out that frequencies in the 3-8Hz band which appear due to the shank movement during FoG event do not commonly happen during volitional standing [6]. Moreover, the power of this frequency band during the FoG event which happens during walking, turning is higher than during rest. Thus, a time-frequency representation of the sensor signal can provide useful information for detecting FoG events.

The wavelet transform has been widely used in conventional signal processing to analyze the signal and provide a time-frequency localization [12]. Instead of dividing the signal into shorter windows of equal size and compute the transform separately on each individual window like the short-time Fourier transform (STFT), the wavelet transform uses a wavelet called mother wavelet, scales and shifts it throughout the signal to calculate the convolution of the wavelet and the input signal. Therefore, the wavelet transform provides a sharper time resolution for high frequency components than low frequency components. This multi-resolution characteristic of the wavelet transform can help to detect some patterns that are not visible in the raw signal.

In this paper, a 6-level discrete wavelet transform (DWT) with the Haar mother wavelet is used to extract the time-frequency representation. The mother wavelet function of the Haar wavelet can be described as

$$\psi\left(t\right) = \begin{cases} 1 & 0 \leq t \leq \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

The scale parameter of the DWT is set to integer powers of 2 (*i.e.*, $2^i, i = 1, 2, 3...$). Thus, the mother wavelet is scaled and shifted by power of 2 as

$$\psi_{i,k}\left(n\right) = \frac{1}{\sqrt{2^i}}\psi\left(\frac{n - 2^i k}{2^i}\right) \tag{2}$$

where i and k are the scale parameter and shift parameter, respectively and both are integers. The representation extracted from the 6-level wavelet decomposition is considered as a pseudo 2-D image and is input to the hybrid deep learning model.

B. CNN-BiLSTM Deep Neural Network

CNN and BiLSTM are powerful deep learning algorithms and are widely used in recent research such as computer vision and natural language processing. Being first proposed by LeCun *et al.* [13] for handwritten digit recognition task, CNN has gained tremendous attention since its strength in processing data which has the form of multiple arrays. By calculating the convolution between the feature map and the kernel, it can extract the local connection while reducing the number of learnable parameters. BiLSTM consists of two LSTM layers in the backward and forward directions, on the other hand, is usually applied to sequential data such as time series for exhibiting temporal dynamic behaviour in both past and future states.

In this paper, the output from the multi-level wavelet decomposition is considered as a 2-D array and is split into 4 smaller frames in the time direction. A CNN subnet is first applied to each individual frame for extracting spatial and local temporal features. The subnet also plays an important role in distilling the input data, thus help to increase the parallelization and speed up the processing time of the BiLSTM layer and the overall deep learning model. The detailed structure of this subnet is described in Fig. 2. The CNN contains multiples convolutional layers followed by a parametric rectified linear unit (PReLU) activation layer. Max pooling layers are used to reduce the size of the features maps while dropout layers are used as a regularization mechanism to avoid the overfitting problem.

The output of the CNN extracted from each frame is then flattened into an 1-D array and input to the BiLSTM layer for extracting global temporal features in both future and past time directions. The outputs from the forward and backward layers are then concatenated and input to a fully connected layer (dense layer). Finally, another fully connected layer with 2 hidden nodes and a *softmax* activation function is used for classifying FoG and non FoG events.

III. Experimental Results

A. Daphnet - Freezing of Gait Dataset

The public Daphnet dataset [8] is a benchmark dataset for automatic detection of FoG events from patients with PD. The dataset was collected from ten PD patients with an age range from 59 to 75. Three accelerometer sensors were

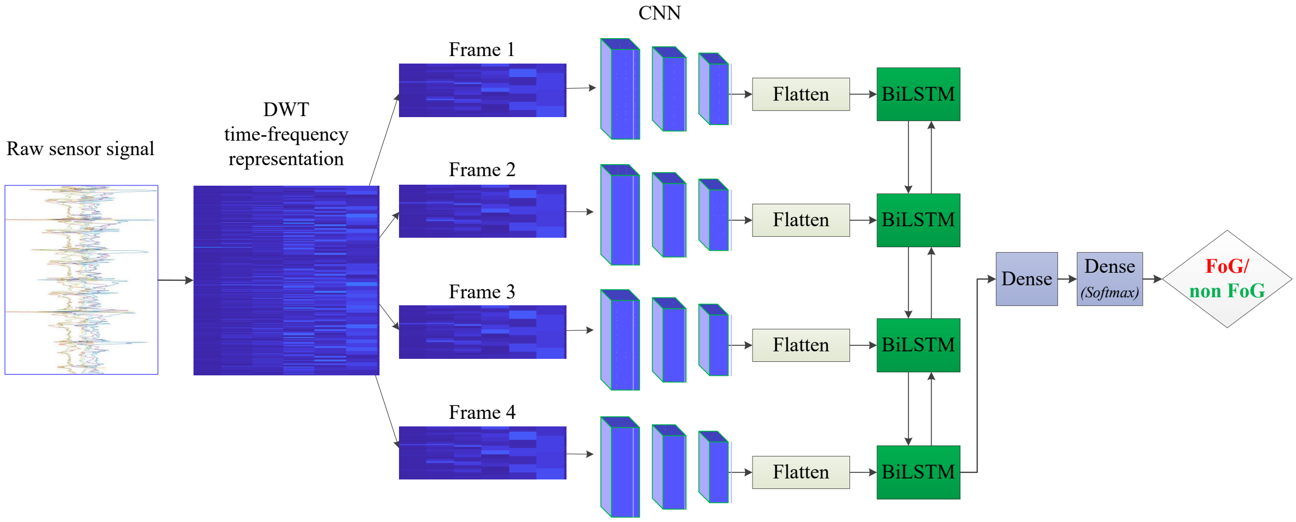

Fig. 1. Detailed architecture of the proposed hybrid deep learning-based framework for freezing of gait detection.

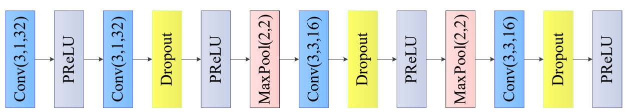

Fig. 2. Detailed structure of the convolutional neural network block in the proposed FoG detection framework.

attached to the patients' shank, thigh and waist to measure 3-D acceleration. The data were acquired at a sampling rate of 64 Hz. Three labels were used to annotate the collected data: idle, normal activities with no freeze (e.g., standing, walking, turning) and freeze.

From the raw data, we remove the samples with idle notation and insufficient sensor data such as data samples at the end of subject 10. The clean data is split into fix-length windows with a size of 2s (128 samples), an overlap of 50% is applied on adjacent windows. Finally, we obtain 1,724 FoG windows (10.19%) and 15,199 non-FoG windows (89.81%).

B. Experiment settings

Baseline models: In order to make a comparison with the proposed model, we implement a set of models including conventional machine learning methods: k-Nearest Neighbors (k-NN), support vector machine (SVM), and deep learning networks: CNN, stacked LSTM and stacked BiLSTM. All the deep learning models that are considered in this study are implemented and trained from scratch using the TensorFlow framework. Other conventional machine learning methods are implemented using the Scikit-learn module.

Since sensor data in the FoG detection task highly differ from person to person, the dataset suffers from high intra-class variance. Therefore, to measure the flexibility of the model in coping with new users, we use a validation method called leave-one-subject-out (LOSO) in which for each running time, data of 9 subjects are used for training and data of the remained subject is used for validating the model. Finally, predicted values from all 10 running times are combined together to assess the model.

Three performance metrics are used in this study to evaluate the models which are classification accuracy, binary F1-score and weighted F1-score. Binary F1-score is used to report results for the class of FoG events, and can be defined as

$$F_{1-\mathrm{b}} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{3}$$

where $\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ and $\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$, $(\mathrm{TP}, \mathrm{FP}, \mathrm{FN}$ are number of true positive, false positive and false negative prediction of the FoG class, respectively). Weighted F1-score is calculated as a weighted average of binary F1-score of all the classes

$$F_{1-\mathrm{w}} = \frac{\sum_i^C N_i \cdot \frac{2 \cdot \mathrm{Precision}_i \cdot \mathrm{Recall}_i}{\mathrm{Precision}_i + \mathrm{Recall}_i}}{\sum_i^C N_i} \tag{4}$$

where for each class i, N_i is the number of true instances and C is the number of classes (in this paper, $C = 2$).

C. Results and Discussion

To analyze the efficiency of using discrete wavelet transform-based features in comparison with the raw sensor signal, an experiment is carried out by applying the proposed CNN-BiLSTM model on the two types of input data. The results of this experiment are shown in Table I. By focusing on the frequency characteristic of freezing of gait events, the time-frequency representation extracted from the multi-level wavelet decomposition outperforms the raw input signal in

TABLE I
COMPARISON OF THE PROPOSED DEEP LEARNING MODEL PERFORMANCE ON TWO TYPES OF INPUT DATA

Input Data	**Accuracy**	F_{1-b}	F_{1-w}
Raw Data	87.9%	33.9%	87.29%
DWT features	90.01%	44.92%	89.47%

TABLE II
COMPARISON OF DIFFERENT MACHINE LEARNING MODELS ON THE TIME-FREQUENCY REPRESENTATION DATA

Model	**Accuracy**	F_{1-b}	F_{1-w}
k-NN ($k = 7$)	89.77%	32.50%	88.15%
SVM	**90.13**%	37.59%	88.83%
Stacked LSTM	88.87%	36.14%	87.97%
Stacked BiLSTM	89.41%	40.54%	88.72%
CNN	88.79%	39.72%	88.31%
(Proposed) CNN-BiLSTM	90.01%	**44.92**%	**89.47**%

terms of all three performance metrics. The result has strengthened our hypothesis of using time-frequency representation for FoG detection.

Results from testing the baseline models and proposed model on the DWT time-frequency representation data are shown in Table II and Fig. 3. A surprising finding is the two conventional machine learning models have better accuracy than most of the deep learning models in which the SVM gets the highest accuracy of 90.13% and closely followed by the proposed CNN-BiLSTM with 90.01% of classification accuracy. However, KNN and SVM seriously suffer from the class imbalance problem as their binary F1-score are only 32.50% and 37.59%, respectively. With the ability of learning spatial and temporal features in both past and future states, the proposed CNN-BiLSTM model achieves the best results in both binary and weighted F1-score.

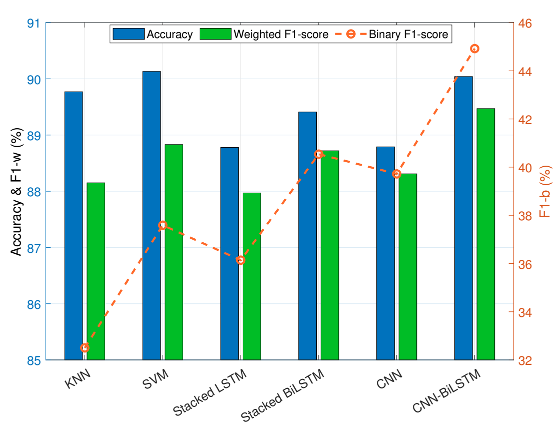

Fig. 3. A visualization of the comparison results on different machine learning algorithms.

IV. CONCLUSION

This paper has presented a wearable sensor-based freezing of gait detection framework that consists of multi-level discrete wavelet transform and a hybrid CNN-BiLSTM deep learning model. In order to utilize the frequency characteristics of FoG events, the discrete wavelet transform is used to generate a time-frequency representation from the raw sensor signal. The CNN-BiLSTM model is, then, used to further learn abstract features automatically and classify FoG events. The experiment results suggest that leveraging both knowledge-based hand-crafted features and deep learning-based automatic learning features could be useful for recognizing the freezing of gait events. For further studies, testing in out-of-clinic environments and class imbalance problem will be need to take into account.

REFERENCES

[1] R. Gilbert, “Freezing of gait in Parkinson’s disease,” American Parkinson Disease Association (APDA) (accessed on June 12 2021). [Online]. Available: https://www.apdaparkinson.org/article/freezing-gait-and-parkinsons-disease/
[2] T. Hashimoto, “Speculation on the responsible sites and pathophysiology of freezing of gait,” *Parkinsonism & Related Disorders*, vol. 12, pp. S55–S62, 2006.
[3] J. Nonnekes, A. H. Snijders, J. G. Nutt, G. Deuschl, N. Giladi, and B. R. Bloem, “Freezing of gait: a practical approach to management,” *The Lancet Neurology*, vol. 14, no. 7, pp. 768–778, 2015.
[4] S. Mazilu, M. Hardegger, Z. Zhu, D. Roggen, G. Tröster, M. Plotnik, and J. M. Hausdorff, “Online detection of freezing of gait with smartphones and machine learning techniques,” in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE, 2012, pp. 123–130.
[5] W. Maetzler, J. Domingos, K. Srulijes, J. J. Ferreira, and B. R. Bloem, “Quantitative wearable sensors for objective assessment of parkinson’s disease,” *Movement Disorders*, vol. 28, no. 12, pp. 1628–1637, 2013.
[6] S. T. Moore, H. G. MacDougall, and W. G. Ondo, “Ambulatory monitoring of freezing of gait in parkinson’s disease,” *Journal of neuroscience methods*, vol. 167, no. 2, pp. 340–348, 2008.
[7] A. El-Attar, A. S. Ashour, N. Dey, H. Abdelkader, M. M. Abd El-Naby, and R. S. Sherratt, “Discrete wavelet transform-based freezing of gait detection in parkinson’s disease,” *Journal of Experimental & Theoretical Artificial Intelligence*, pp. 1–17, 2018.
[8] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster, “Wearable assistant for parkinson’s disease patients with the freezing of gait symptom,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 436–446, 2010.
[9] J. Camps, A. Sama, M. Martin, D. Rodriguez-Martin, C. Perez-Lopez, J. M. M. Arostegui, J. Cabestany, A. Catala, S. Alcaine, B. Mestre *et al.*, “Deep learning for freezing of gait detection in Parkinson’s disease patients in their homes using a waist-worn inertial measurement unit,” *Knowledge-Based Systems*, vol. 139, pp. 119–131, 2018.
[10] M. Zeng, H. Gao, T. Yu, O. J. Mengshoel, H. Langseth, I. Lane, and X. Liu, “Understanding and improving recurrent networks for human activity recognition by continuous attention,” in *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. ACM, 2018, pp. 56–63.
[11] T. Bikias, D. Iakovakis, S. Hadjidimitriou, V. Charisis, and L. J. Hadjileontiadis, “DeepFoG: An imu-based detection of freezing of gait episodes in parkinson’s disease patients via deep learning,” *Frontiers in Robotics and AI*, vol. 8, 2021.
[12] I. Daubechies, “The wavelet transform, time-frequency localization and signal analysis,” *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.
[13] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, “Handwritten digit recognition with a back-propagation network,” in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., vol. 2. Morgan-Kaufmann, 1990, pp. 396–404.