Computers in Biology and Medicine 191 (2025) 110138

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A new lightweight deep learning model optimized with pruning and dynamic quantization to detect freezing gait on wearable devices

Myung-Kyu Yi [a], Seong Oun Hwang [b,*]

[a] Department of Biomedical Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul, 04763, Seoul, Republic of Korea
[b] Department of Computer Engineering, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam-si, 13120, Gyeonggi-do, Republic of Korea

ARTICLE INFO

Keywords:
Efficient channel attention
Freezing of gait
Inertial measurement unit
Gated recurrent unit
Wearable sensors

ABSTRACT

Freezing of gait (FoG) is a debilitating symptom of Parkinson's disease that severely impacts patients' mobility and quality of life. To minimize the risk of falls and injuries associated with FoG, it is crucial to develop highly accurate FoG detection systems for portable wearable devices that enable continuous and real-time monitoring. However, achieving high accuracy in deep learning (DL) models typically requires a large number of parameters and integrating multiple sensors, posing challenges for deployment on resource-constrained wearable devices. To address these challenges, we propose a novel lightweight DL model that combines a convolutional neural network and a gated recurrent unit with residual attention and efficient channel attention mechanisms. Additionally, we incorporate pruning and dynamic quantization techniques, along with an innovative feature selection method, to optimize the model for wearable applications. Experimental results demonstrate that our proposed DL model outperforms state-of-the-art supervised DL models, achieving an F1 score of 0.994 while utilizing 29.9 times fewer parameters than existing models. The model's maximum memory usage is only 420.91 KB, making it well-suited for wearable devices. Furthermore, optimizations through pruning and dynamic quantization further reduced the model size by an additional 7.84 times, resulting in a final size of just 44.04 KB without sacrificing accuracy. As a result, the proposed DL model achieves high accuracy in FoG detection with minimal memory usage, enabling real-time monitoring on wearable devices and providing a practical solution for managing FoG in Parkinson's patients.

1. Introduction

Parkinson's disease is a degenerative brain disorder caused by a deficiency of dopamine, one of the brain's essential neurotransmitters [1]. Dopamine plays a crucial role in ensuring the smooth functioning of the basal ganglia, which is vital for coordinated movement. The hallmark symptoms of Parkinson's disease, including tremors and rigidity, result from this dopamine deficit [2]. As one of the major neurodegenerative disorders affecting the elderly, Parkinson's disease has been increasing in prevalence as the global population ages [3]. Currently, more than 8.5 million individuals worldwide are living with Parkinson's, a number that has doubled in the last 25 years, making it the fastest-growing neurological disorder [4]. Notably, the disease occurs more frequently in men, with an incidence rate 1.5 times higher than in women [5]. These concerning trends highlight the urgent need for increased awareness and dedicated research to combat Parkinson's disease.

One significant symptom, Freezing of Gait (FoG), refers to a temporary inability to walk or an inability to take more than a few steps, particularly when turning or changing direction [6]. Patients experiencing FoG often describe a sensation as if their feet are glued to the ground, and in severe cases, they may come to a complete halt while attempting to walk, increasing the risk of falls [7]. Given the high frequency of this symptom, developing a continuous monitoring system capable of detecting FoG in real-time is essential for effective symptom management [2].

Detecting FoG presents several challenges due to individual variability, use of multiple physiological signals, and the dynamic nature of FoG episodes [8]. Symptoms can vary significantly between individuals, making it difficult to identify consistent signal patterns indicative of FoG onset [7]. Moreover, FoG detection involves processing various sensor signals, including electromyography (EMG), electroencephalography (EEG), and inertial measurement unit (IMU) data, adding complexity to signal interpretation [9,10]. Additionally, FoG episodes often occur suddenly, requiring sophisticated analytical methods for accurate detection [8].

To address these challenges, machine learning (ML) algorithms such as Support Vector Machines (SVM), Decision Tree (DT), Random Forest

* Corresponding author.
E-mail address: sohwang@gachon.ac.kr (S.O. Hwang).

https://doi.org/10.1016/j.compbiomed.2025.110138
Received 10 October 2024; Received in revised form 24 March 2025; Accepted 2 April 2025
Available online 16 April 2025

(RF), and K-Nearest Neighbors (KNN), along with deep learning (DL) models like Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (CNN-BiLSTM), have been leveraged to analyze sensor data, significantly improving the classification of FoG events [10–22]. For example, in the widely used Daphnet dataset [23] for FoG detection, Elbatanouny et al. (2024) [24] achieved an F1 score of 0.97 using the XGBoost (eXtreme Gradient Boosting) model. Similarly, Ronald et al. (2021) [12] reported an F1 score of 0.94 with the iSPLInception model, while Lisha et al. (2023) [25] obtained an F1 score of 0.958 on a multimodal dataset [14] using the CNN-BiLSTM model. Despite these notable advancements in FoG detection, several limitations persist within existing research.

1. Existing works detecting FoG have achieved a maximum F1 score of 0.97 [24]. However, further improvements in accuracy are necessary to reduce the risks associated with FoG. Even minor errors can have serious consequences for patients, particularly in critical situations, such as falls caused by FoG episodes.
2. Existing works do not consider the implementation of FoG systems into wearable devices. They primarily focused on improving accuracy by receiving diverse and extensive sensor data on computationally powerful servers and analyzing them using high-performance models. While these approaches achieve high performance, they overlook the practical implementation challenges on wearable devices with limited computational resources. For example, devices like the STM32 Nucleo-144 [26], commonly used in wearables, offer only 1 MB of flash memory [27].

To address these challenges, it is crucial to develop lightweight DL models that maintain high accuracy while operating efficiently within the constraints of resource-limited devices. Such advancements would enable real-time monitoring and personalized treatment for Parkinson's patients, significantly enhancing their quality of life. In this paper, therefore, we propose a new lightweight DL model to detect FoG episodes, which combines a convolutional neural network (CNN) and a gated recurrent unit (GRU) with residual attention (ATT) [28] and efficient channel attention (ECA) [29] mechanisms. The combination of these components is intended to complement each other, effectively capturing essential data characteristics for detecting FoG. The combination of CNN and GRU enables the simultaneous learning of local patterns and temporal variations in the signal, allowing the model to flexibly capture individual-specific patterns and characteristics [30]. The ECA mechanism dynamically adjusts the importance of each channel, minimizing the impact of noisy channels while emphasizing those carrying meaningful information [29]. This is crucial in extracting critical features from diverse physiological signals specific to each individual. The ATT mechanism focuses on significant features in the input data, highlighting critical patterns associated with the dynamic nature of FoG signals and improving prediction accuracy [28]. The ECA and ATT mechanisms enhance feature extraction while maintaining computational efficiency, making the model suitable for resource-constrained environments [28,29]. Finally, we employ the proposed feature selection method in combination with improved data preprocessing and pruning/quantization procedures to reduce the model's size and complexity without sacrificing accuracy. The major contributions of this paper are as follows.

1. In contrast to existing works that primarily depend on high-performance computing systems for real-time FoG detection, we propose a lightweight DL model specifically designed for resource-constrained wearable devices, utilizing minimal sensors for real-time detection. Furthermore, we significantly reduce the model size through effective pruning and dynamic quantization techniques. Our findings demonstrate that the proposed model can be successfully implemented in wearable devices, enabling real-time symptom monitoring and effectively reducing the risk of falls for patients.
2. To address the challenges of FoG detection arising from the variability and complexity of biosignals, we propose an enhanced data preprocessing method combined with an innovative feature selection algorithm, enabling the efficient extraction of key information essential for accurate detection. As a result, the performance of the FoG detection system is significantly improved, contributing to enhanced patient safety and better treatment outcomes through real-time monitoring.

The purpose of the proposed lightweight DL model is to enhance performance, particularly in terms of the F1 score, enabling accurate detection and prediction of FoG, even in imbalanced datasets. By improving robustness, the model becomes more practical for real-world applications with wearable devices, thereby maximizing the effectiveness of treatment and symptom management. Ultimately, this contributes to improving patient safety and quality of life.

Our implementation is available in the public domain: https://github.com/kainos14/FoG. The remainder of this paper is structured as follows: Section 2 reviews the prior research on FoG detection. Section 3 describes the materials and methods proposed for FoG detection. Section 4 presents and discusses the experimental results. Finally, Section 5 provides the conclusions of this study and suggests potential directions for future research.

2. Related works

As shown in Table 1, various studies have explored Freezing of Gait (FoG) detection using wearable technologies [11,40]. The most widely used method for FoG detection involves IMUs, which consist of sensors such as accelerometers, gyroscopes, and magnetometers [6,11–14,22,25,31–39]. These sensors can be placed on different areas of the lower limbs to capture both linear and angular movements. The IMU is favored for its small size, continuous signal collection, and reliability. Brain wave testing is another innovative and non-invasive approach, mainly using EEG [14,15]. This method utilizes an EEG cap to monitor real-time physiological changes in the brain before and during FoG episodes. The force sensing resistor (FSR) can also be used in gait analysis [16–18]. As wearable sensors, FSRs change their resistance in response to pressure or force. They are valued for their reliability, flexibility, and cost-effectiveness in gait analysis. However, FSRs face limitations in sensitivity and predictability, specifically in FoG detection. Vision-based approaches can accurately identify gait anomalies using a camera [9]. These methods may be preferable over IMU-based approaches thanks to fewer restrictions on user activities. However, challenges remain, such as protecting personal privacy, limitations in field-of-view cameras, and the need for high-performance computational resources. Lastly, the multimodal approach [19] integrates the above sensor types (IMUs, EEGs, and FSRs) to detect FoG events. This method reduces detection latency and improves analysis by integrating multiple data types, enhancing accuracy and robustness. By incorporating various modalities, the model can compensate for errors or data loss from a single sensor and conduct a more thorough investigation into the causes of gait freezing. However, this approach also presents challenges. One of the key issues is the need for data synchronization because different sensors may operate using varying time units or sampling rates. Furthermore, the model's complexity and computational cost increase, making real-time processing more difficult, as it must account for the unique characteristics of each modality while learning the overall patterns [27].

The core of FoG detection lies in the classification algorithm. Advancements in sophisticated DL models and various types of sensors, have enhanced the detection of FoG [12,14,15,20,21]. Ronald et al. [12] proposed a novel DL architecture for detecting FoG, leveraging the principles of Google's Inception ResNet model. They achieved the best F1 score of 0.94 using the Daphnet dataset. Shi et al. [22] proposed a robust model using time–frequency analysis. They used

Table 1
Comparative analysis of FoG detection and prediction methods using wearable sensors.

Study	# of subject	Sensor type (Location)	Model	Performance	Wearables	Pruning & Qtz.
Moore et al. (2013) [31]	25	acc (lumber, thight, shank, foot)	Threshold -based system	Sens. 0.843 Spec. 0.78	×	×
Mazilu et al. (2014) [32]	10	acc, gyro (waist, shank)	DT	Sens. 0.875 Spec. 0.970	×	×
Sijobert et al. (2014) [33]	7	acc, gyro (shank)	Threshold -based system	Sens. 0.839 F1 0.72	×	×
Pepa et al. (2014) [34]	18	acc (hip)	Fuzzy rule -based systems	Sens. 0.875 Spec. 0.970	×	×
Esfahani et al. (2021) [35]	10	acc (ankle)	LSTM	Sens. 0.926 Spec. 0.956	×	×
Ronald et al. (2021) [12]	10	acc (ankles, thighs, trunk)	iSPL Inception	F1 0.940	×	×
Li et al. (2022) [36]	10	acc (back, thigh, ankle)	CNN +LSTM	Sens. 0.875 Spec. 0.923 ACC. 0.920	×	×
Shi et al. (2022) [22]	63	acc, gyro (ankles)	CNN	Sens. 0.878 Spec. 0.864 ACC. 0.871	×	×
Naghavi et al. (2022) [37]	7	acc, gyro (ankles)	CNN	Sens. 0.630 Spec. 0.986	×	×
Ghuman et al. (2022) [38]	10	acc (thigh)	CNN	ACC. 0.837	O	×
Lin et al. (2022) [39]	10	acc (ankle, thigh, truck)	CNN	Sens. 0.888 F1 0.8534	O	×
Zhang et al. (2022) [14]	12	eeg, emg, acc, sc (leg, arm,waist)	Ensemble	F1 0.820	×	×
Lisha et al. (2023) [25]			CNN +LSTM	F1 0.958	×	×
Elbatanouny et al. (2024) [24]	10	acc (ankles, thighs, trunk)	XGBoost	F1 0.970	×	×

acc: Accelerometer, gyro: Gyroscope, eeg: Electroencephalography, emg: Electromyography, sc: Skin Conductance, Sens.: Sensitivity, Spec.: Specificity, ACC. : Accuracy, Qtz.: Quantization

the continuous wavelet transform on signals collected from IMUs positioned on the lower extremities of 63 Parkinson's patients during FoG episodes. Experimental results showed that the proposed model obtained an F1 score of 91.5%. Bajpai et al. [15] conducted an in-depth study using an EEG and an IMU for early FoG prediction. They designed an ensemble model comprising two neural networks, EEGFoGNet and IMUFoGNet, and evaluated it across different prediction horizons (PHs) and ensemble weight configurations. Their proposed model obtained an F1 score of 0.85 at a 1-s PH, but its lowest F1 score was 0.73 at a 5-s PH. Zhang et al. [14] developed a FoG detection system that distinguishes between Parkinson's disease patients and ordinary individuals with an accuracy of 0.93. This approach involves recording gait acceleration (ACC), EEG, EMG, and skin conductance (SC).

Recent advancements in ML and DL have notably increased detection accuracy. Nevertheless, there remains a significant need for further improvements. Most studies concentrate on achieving high accuracy using advanced DL models and processing various sensor data on powerful servers. While these methods are effective, they demand considerable computational resources, making them less practical for resource-constrained wearable devices. Notable exceptions are the works of Ghuman et al. (2022) [38] and Lin et al. (2022) [39], who concentrated on implementing DL models specifically designed for wearables. Ghuman et al. employed a CNN with two convolutional layers on an Arduino Nano 33 Sense board [41], achieving an accuracy of 83.7% while using 384,952 bytes of memory. Lin et al. proposed a lightweight edge DL model consisting of three convolutional layers. This model achieved a sensitivity of 0.888 and a specificity of 0.807 while running on a 32-bit ARM Cortex-M4 microcontroller [42], utilizing 478 kB of memory. Despite these advancements, the accuracy of such models still requires improvement to be applicable in real-world scenarios. Therefore, it is essential to develop a lightweight and highly optimized DL model that balances accuracy and efficiency for wearable devices. Such advancements would facilitate real-time FoG detection with minimal sensors and reduced battery consumption, ultimately allowing for personalized interventions. This could significantly improve mobility and quality of life for Parkinson's patients.

3. Materials and methods

Fig. 1 presents the flowchart of the proposed FoG detection method, which consists of eight phases: data cleaning, segmentation, data balancing, feature extraction, feature selection, model training, pruning/dynamic quantization, and performance evaluation. First, raw signals from the accelerometer and gyroscope are collected and preprocessed to clean and segment the data. Next, time and frequency domain features are extracted from the segmented data. The proposed DL model is then applied to detect the presence of a FoG episode. Finally, the

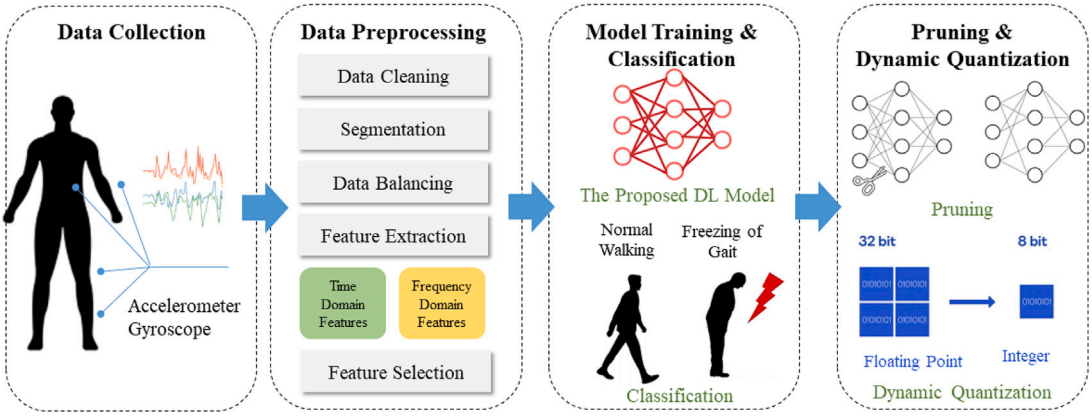

Fig. 1. The flowchart of the proposed FoG detection method.

Table 2
The list of selected features for the proposed method.

Feature	Description	Equation
Mean	Average value of the signal	$\frac{1}{N}\sum_{i=1}^{N} x_i$
Variance	Variance value of the signal	$\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$
Standard deviation	Standard deviation of the signal	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}$
Root mean square	Square root of the average of the squared values of the signal	$\sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2}$
Mean absolute value	Average of the absolute values of the signal	$\frac{1}{N}\sum_{i=1}^{N} \lvert x_i\rvert$
Kurtosis	Tailedness of a distribution of the signal	$\frac{E[(X-\mu)^4]}{\sigma^4}$
Skewness	Asymmetry of a distribution's shape about to its mean	$\frac{E[(X-\mu)^3]}{\sigma^3}$
Sum magnitude area	The sum of the magnitudes of the signal	$\sum_{i=1}^{N} \lvert x_i\rvert$
Slope	Rate of signal change over time	$\frac{y_2-y_1}{x_2-x_1}$
Entropy	A measure of randomness in the frequency distribution of the signal	$-\sum_n P(n)\log_2 P(n)$
Energy	Total energy of the signal	$\sum_{n=1}^{N} \lvert x[n]\rvert^2$
Peak frequency	The frequency at the highest amplitude in the signal's power spectrum	$max(X(n))$
Freeze index	FoG to non-FoG power ratio	$\frac{\sum_{f=3\ Hz}^{f=8\ Hz} P(f)}{\sum_{f=0.5\ Hz}^{f=3\ Hz} P(f)}$
Power	Combined power within the FoG and non-FoG bands of the signal	$\sum_{f=3\ Hz}^{f=8\ Hz} P(f)+\sum_{f=0.5\ Hz}^{f=3\ Hz} P(f)$

model undergoes pruning and dynamic quantization to reduce its size while maintaining accuracy, followed by a performance evaluation. The following sections provide a detailed explanation of each phase in our method.

3.1. Data preprocessing

Data processing comprises several stages, including data cleaning, segmentation, data balancing, feature extraction, and feature selection, as seen in Fig. 1. Streamlined data preprocessing is vital to ensuring the calculations are fast and understandable when deploying wearable devices. We designed an optimized data preprocessing method that enhances interpretability with minimal raw data processing. Several critical steps during the data cleaning stage are performed. They involve removing missing values and outliers and normalization by transforming the data to ensure they are in a suitable format for analysis, including scaling them. Then, raw data are segmented using sliding windows of a fixed length based on the sampling rate, as shown in Fig. 2.

We employ a sliding window with a fixed length to minimize memory usage. The window size is chosen according to the sampling rate and the minimum duration of freezing events. Overlapping is a widely used method for analyzing FoG signals [43]. Overlapping windows enhance FoG signal analysis by preserving boundary information and capturing temporal patterns, improving prediction accuracy and model performance for biosignals. However, larger overlaps can improve accuracy but also lead to increased data redundancy, which results in higher memory usage, greater computational demands, and longer processing times. On the other hand, smaller overlaps enhance computational efficiency and processing speed, but they may compromise signal continuity and the accuracy of feature extraction [43]. To minimize power consumption, we adopted the 50% overlap method, commonly used for deploying DL models on wearable devices in [38]. FoG windows are labeled as such if all the data, or more than 50% of that window's data, are FoG data. Each window is analyzed to identify sections corresponding to each activity defined in the dataset.

Then, we perform oversampling such as SMOTE [43] to balance the data and prevent bias towards a majority class. SMOTE improves data diversity and reduces the risk of overfitting by generating new synthetic data through interpolating existing minority class samples rather than simply duplicating data. This approach expands the feature space of the minority class and enhances the model's generalization performance by addressing biased decision boundaries in imbalanced datasets.

After balancing the dataset, we extract statistical domain features from each sliding window. A Butterworth bandpass filter is then applied to isolate and analyze frequency bands for each window, followed by the extraction of frequency domain features. FoG episodes are characterized by greater variability and abnormal signal patterns compared to normal gait. Statistical features such as mean and standard deviation help identify this variability, while skewness and kurtosis capture asymmetric distributions or specific concentrated patterns. Energy is also an essential feature, as it reflects the reduction in movement

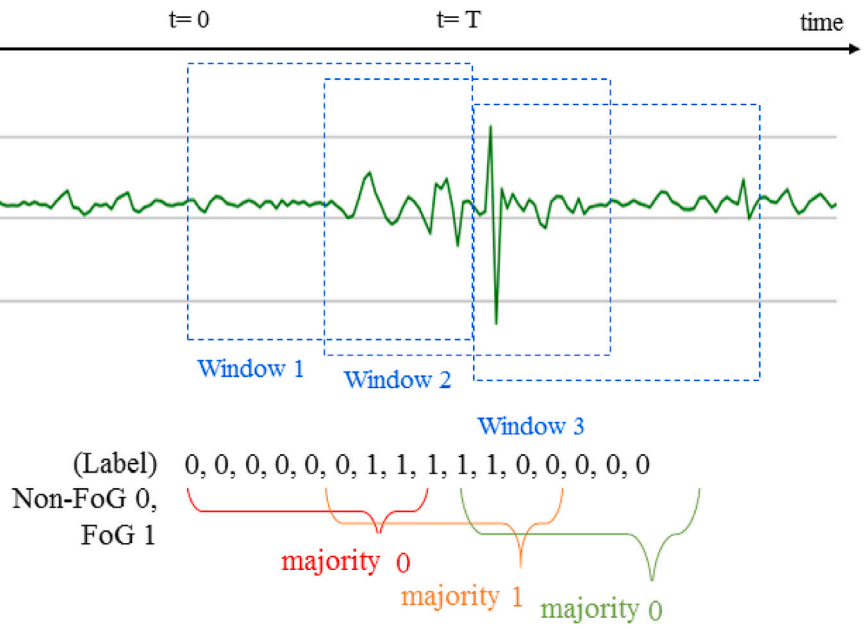

Fig. 2. The proposed overlapping window technique.

intensity during FoG episodes. In the frequency domain, FoG typically exhibits increased energy in low-frequency ranges (0.5–3 Hz) and reduced activity in high-frequency ranges (3–8 Hz) [22]. Features such as the center of frequency and dominant frequency can effectively detect unique changes occurring during FoG events. The complete list of features extracted is in Table 2. Finally, the proposed feature selection is performed as the last preprocessing stage. Feature selection improves computational efficiency and model performance by reducing input variables, lowering complexity, and optimizing hardware usage. It also enhances generalization and predictive accuracy by focusing on significant features, minimizing noise, and preventing overfitting.

Algorithm 1 Proposed Feature Selection

Require: $S = \{s_1, s_2, \ldots, s_n\}$: The set of all features.
Ensure: $E \leftarrow$ The set of selected features
T_h: The predefined threshold
$S' \leftarrow$ EmptyList
$E \leftarrow$ EmptyList
$S' \leftarrow \{x \in S \mid x \geq T_h\}$
$S_{sorted} \leftarrow$ Sort S' in descending order
$A_{total} \leftarrow$ Evaluated performance of the model with S_{sorted}
for $i \leftarrow 1$ **to** $|S_{sorted}|$ **do**
  $S_{temp} \leftarrow Remove\ s_i\ from\ S_{sorted}$
  $A_i \leftarrow$ Evaluated performance of the model with S_{temp}
  if $A_{total} \leq A_i$ **then**
    $E \leftarrow E$ append s_i
  end if
end for
return E

Algorithm 1 shows the pseudocode of the proposed feature selection method. The primary objective is eliminating redundant or irrelevant features and selecting the minimum number required to achieve high performance. The algorithm initially filters out low-impact features based on a threshold value, T_h. The features are then sorted, and each feature is removed (one after the other) to evaluate the model's performance without it. If performance without a corresponding feature is better than the total, the feature is not included in the final set; otherwise, it is added. So, the most valuable features for model performance can be selected by applying the feature selection technique, and wearable implementation can be made easy by achieving high performance with a few features. It is well known that there is a correlation when the value is more significant than 0.3 [44]. However, we adopt a threshold T_h of 0.2, as even lower correlation values can reflect meaningful relationships to avoid overlooking subtle patterns crucial for FoG detection. Preliminary experiments across various thresholds showed that $T_h = 0.2$ provided the most optimal balance [45]. The proposed method simplifies the feature space through an initial filtering process focusing on correlation-based linear relationships. Then it involves indirectly capturing the nonlinear relationships through performance evaluation using the proposed DL model. This approach offers an advantage over Least Absolute Shrinkage and Selection Operator (LASSO) [46], which only considers linear relationships, and Recursive Feature Elimination (RFE) [46], which can be computationally expensive when dealing with many features.

3.2. Model description

This section describes the proposed DL model. First, the proposed DL model architecture is introduced, followed by the GRU, ECA, ATT, pruning, and dynamic quantization.

3.2.1. The architecture of the proposed DL model

The architecture of the proposed DL model is shown in Fig. 3. We build a baseline DL model called CGNET by combining commonly used 1D CNN and GRU layers with an optimized data preprocessing method. Then, we further enhanced CGNET by integrating ATT and ECA modules. As a result, the proposed DL model consists of an input layer, a 1D convolution layer, the ATT module, the GRU layer, the ECA module, Flatten, Dense, and output layers. The input signals pass through the convolution layer of the channel. Then, the data passes through the residual attention module consisting of Permute, Dense, Permute, and Multiply layers. The purpose of the residual attention module is to focus attention on important parts of the input sequence. It can enhance the ability to grasp nuances and dependencies in sequence data. These extracted features are combined in a concatenation layer and fed into the GRU layer through a Flatten layer. Then, data go through the ECA module, consisting of Global Average Pooling (GAP), Reshape, 1D convolution, Sigmoid activation, Reshape, and Multiply layers. The ECA module learns the importance of specific channels, enabling the network to emphasize more relevant features. Then, they go through Flatten and Dense layers using Sigmoid activation as the classification function to obtain the normalized output.

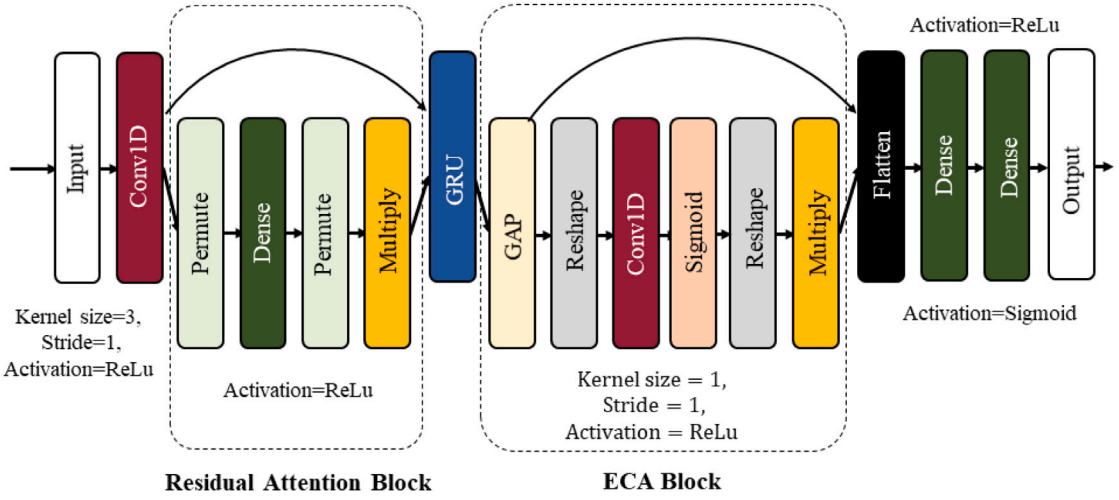

Fig. 3. The proposed attention-based DL model architecture.

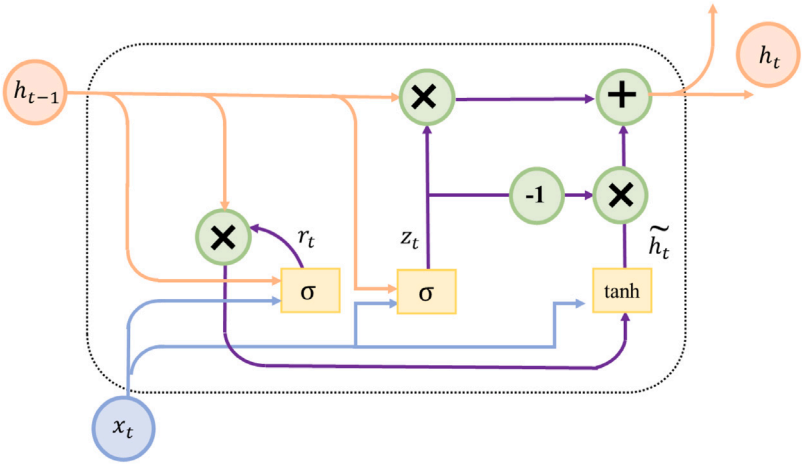

Fig. 4. The architecture of the GRU.

3.2.2. Gated recurrent unit

We employ a combination of CNNs and GRUs to extract spatial and temporal features as baseline models. GRU has a simpler structure than long short-term memory (LSTM) [30]. A GRU updates the hidden state and minimizes computations while keeping the LSTM's long-term dependency issue solved. The GRU simplifies the LSTM structure of three gates (input, output, and delete), which functions similarly to an LSTM, as illustrated in Fig. 4. There are just two gates in a GRU: reset and update. The GRU learns faster than LSTM, but in several evaluations, the GRU showed performance similar to LSTM. The equations for a GRU with different cell states are as follows:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{1}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{2}$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \tag{3}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_r \tag{4}$$

where z_t is the update gate, r_t denotes the reset gate, and in Eq. (4), $1 - z_t$ represents the input gate. Each gate plays a distinctive role. The input gate regulates how much of the prior state's data is kept. On the other hand, the reset gate controls how much information from the prior instant is received; σ is the sigmoid function, h_t is the output vector, and x_t is the current input values.

3.2.3. The efficient channel attention mechanism

ECA mechanism is a lightweight channel attention mechanism designed to improve upon the limitations of Squeeze-and-Excitation (SE) [47], which involves high computational costs and complexity due to learning correlations among all channels. ECA dynamically adjusts the kernel size based on the number of channels to learn effective channel interactions within an appropriate range. The core idea of the ECA mechanism is to determine channel-wise weights based on the global information of each channel. Global cross-channel interactions are beneficial because they allow the model to capture complex dependencies and relationships between different parts of the data, potentially leading to improved performance. However, this approach is computationally expensive, with a parametric complexity of $O(n^2)$. To address this, a more efficient method called local cross-channel interaction calculates the attention weight for a query channel by considering a dynamically selected set of nearby channels. This is achieved through a 1D convolution kernel with shared weights across all channels, significantly reducing the parameter count to about k. As a result, the ECA mechanism is lightweight in terms of parameter overhead and complexity in floating point operations. Fig. 5 presents a diagram of the ECA mechanism, where GAP signifies the Global Average Pooling (GAP) layer. Initially, it adaptively determines convolution kernel size k based on channel dimension C. Subsequently, it executes a rapid 1D convolution of size k, followed by applying a sigmoid function

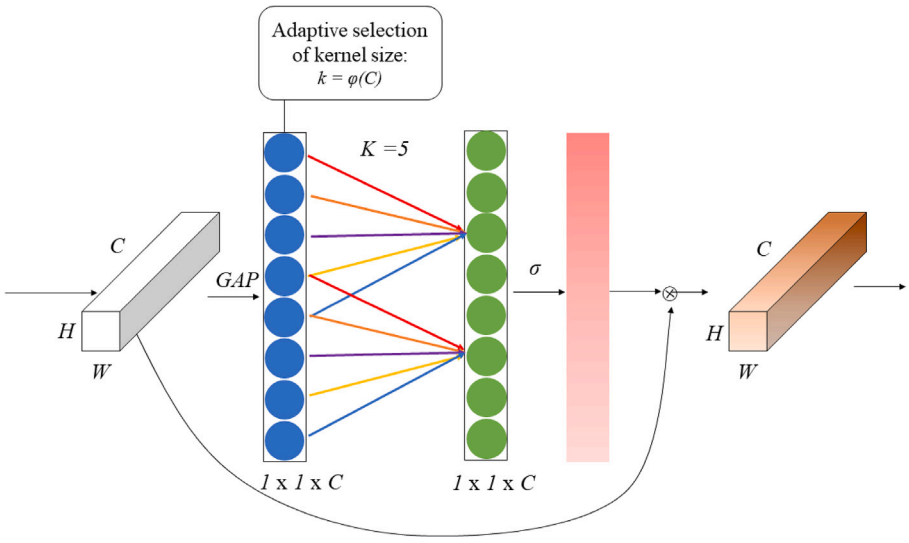

Fig. 5. Diagram of the ECA mechanism.

to facilitate learning of the channel attention. Convolution kernel size k is a crucial parameter that dictates the extent of interaction coverage and is associated with channel dimension C. The mapping relationship ϕ between k and C can be described as follows:

As is commonly known, the number of filters, or channel dimension C, is usually set to a power of two. Therefore, we suggest the following method to convert $\phi(k) = \gamma * k - b$, a linear function, into a non-linear one:

$$C = \phi(k) = 2^{(\gamma * k - b)}. \tag{5}$$

Then, kernel size k can be adaptively computed given channel dimensions C:

$$k = \psi(C) = \left| \frac{log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd}, \tag{6}$$

where the nearest odd number of t is indicated by $|t|_{odd}$.

$$C = \phi(k). \tag{7}$$

3.2.4. Residual attention mechanism

The core principle of the residual attention mechanism is to assign greater weights to important parts of the input data while suppressing less relevant parts, enabling efficient and robust feature representation learning. To boost the model's interpretability, we add an ATT mechanism [28] to the CNN–GRU as a baseline model by prioritizing critical features [28]. Its core idea is to help identify crucial parts of the input sequence to improve performance. It enhances training efficiency by focusing on vital information while filtering out less significant details, reducing the burden of excessive data. Set α_{t_i} comprises the weights each source hidden state should give to each output state. Attention weights α_{t_i} for hidden states h_i are computed by applying a softmax operation as follows:

$$\alpha_{t_i} = \frac{\exp\left(e_{t_i}^T \cdot e_s\right)}{\sum_t \exp\left(e_{t_i} e_s\right)} \tag{8}$$

$$c_i = \sum_t \alpha_{t_i} h_{t_i} \tag{9}$$

The context vector in Eq. (8) is derived by aggregating the relative weights from each time step. These weights, known as alignment scores e_{t_i}, are determined based on the hidden state generated in the preceding time step. Concurrently, e_s is assimilated into the network training process, ensuring the preservation of contextual information.

3.3. Pruning and dynamic quantization

Although the proposed DL model is effective, its size presents a challenge for implementation on wearable devices. To address this, we adopt two optimization techniques — pruning and dynamic quantization — to ensure compatibility with resource-constrained environments. Pruning and quantization offer significant advantages over knowledge distillation and weight clustering, as they simplify models and reduce memory usage without requiring additional training phases [45]. In contrast, knowledge distillation requires the simultaneous training of a large and a small model, while weight clustering involves extra steps for weight adjustment. Pruning and quantization generally cause less performance degradation by eliminating redundant weights or lowering precision, effectively maintaining accuracy while enhancing efficiency. Furthermore, quantization benefits from strong hardware support for integer operations (int8), whereas distillation and clustering may have limited compatibility, reducing their effectiveness in edge and wearable devices.

For pruning and quantization, we utilize the TensorFlow Model Optimization Toolkit [48]. Pruning is a compression technique that removes less essential weights from a trained model, producing a smaller and more efficient DL model. We begin with 20% sparsity and gradually increase it to 80%. Meanwhile, dynamic quantization reduces the computational cost of neural network operations by replacing high-precision data with lower-precision values, resulting in faster processing. It achieves this by substituting low-level floating-point calculations with efficient integer computations, thereby lowering processing costs. The following equation describes this mapping process:

$$F = S \times (N - Z) \tag{10}$$

where N, Z, and F are the fixed point representation, zero point, and floating point values. S is the scaling factor.

$$S = \frac{F_{max} - F_{min}}{N_{max} - N_{min}} \tag{11}$$

resulting in

$$F = \frac{F_{max} - F_{min}}{N_{max} - N_{min}} \times (N - Z) \tag{12}$$

We can calculate the quantized number range (N_{max}, N_{min}) where k is the quantization bits as follows.

$$\left[-2^{k-1}, 2^{k-1} - 1\right] \tag{13}$$

The dynamic quantization process only applies during inference, leaving the training phase unchanged. Before inference, the model's weights are converted to INT8 format, and the activations are quantized. This allows for efficient calculation of matrix multiplications using optimized INT8 functions. Since the quantization is performed on the fly, the model runs smoothly and efficiently, providing reassurance about its performance.

4. Experiments and results

This section presents the experimental results and analysis of the proposed FoG method. First, we will describe the datasets used to evaluate the proposed model, then explain the experimental setup and the performance measures employed. Next, we will discuss the hyperparameter tuning process, detailing how the parameters for the proposed DL model were determined. After that, we will compare the classification performance of the proposed method with current state-of-the-art techniques. Finally, we will conduct an ablation study to examine the impact of each component of the proposed DL model.

4.1. Dataset

We used the Daphnet FoG dataset [23] and the Multimodal dataset [14], both of which are widely recognized public datasets for FoG detection. The details of each dataset are provided in the following subsections. To facilitate seamless integration with wearable devices, the proposed method exclusively utilizes accelerometer and gyroscope data, ensuring efficient processing while maintaining high detection accuracy.

4.1.1. The Daphnet dataset

The Daphnet Freezing of Gait dataset is a widely used public dataset designed to support the development and evaluation of FoG detection algorithms. It provides critical acceleration data from wearable sensors placed on the legs and hips, enabling researchers to accurately identify FoG episodes. The dataset includes data from ten patients, eight of whom experienced FoG events during the test, while the remaining two exhibited normal walking patterns without any freezing episodes. Data acquisition involved placing sensors on the ankles, thighs, and trunk of each participant. Specifically, two three-dimensional accelerometers were attached to each leg, one positioned just above the ankle and the other above the knee. Another accelerometer was secured to a belt around the lower back to capture trunk movements. The sensors recorded accelerometric signals at a sampling rate of 64 Hz, capturing detailed motion patterns during different activities. The participants performed three specific tasks: linear walking, walking with multiple turns, and engaging in daily life activities such as navigating different rooms, fetching coffee, and interacting with everyday objects like doors.

4.1.2. The multimodal dataset

The Multimodal dataset for FoG detection is another open dataset comprising data from 12 Parkinson's patients, and was designed to improve FoG identification through a blend of physical and physiological sensor data. In this dataset, subjects two and five did not experience FoG during the experiment, while the other participants did experience FoG episodes. The dataset integrates ACC, EEG, EMG, and SC measurements, offering a comprehensive view of the patient's condition. Recorded data includes 2 h 14 min of regular walking and 1 h 28 min of FoG events. A 32-channel MOVE system was used to record EEG and EMG data at 1000 Hz, and custom gear using TDK MPU6050 sensors and STM32 CPUs was used to record ACC and SC data. Four inertial sensors were placed on the legs and waist, with SC and ACC data sampled at 500 Hz and recorded onto TF memory cards. With the video recording, two sophisticated doctors would label the data to indicate whether FOG occurred.

Table 3
Split and sample Distribution for the Daphnet dataset.

Set	Subjects	Samples
Train	S1-1, S1-2, S3-1, S3-2, S6-1, S6-2, S7-1, S8-1, S9-1, S10-1	7,945
Test	S2-1, S4-1, S5–2	2,332
Validation	S2-2, S3-3, S5–1	1,602

4.2. Experimental setup and performance measures

This section presents the experimental setup and the performance measures used in the study. All experiments were conducted on a PC with an Intel Core i7-6700 CPU and 8 GB RAM. The model exclusively utilized ACC and gyroscope signals obtained from wearable devices with limited resources, excluding EEG, EMG, and SC readings to ensure practical applicability. To ensure fair and accurate model comparisons, different data splitting and validation strategies were applied to each dataset, aligning with the methodologies used in previous studies. This approach helps reduce performance variability caused by data partitioning, minimizes bias in small or imbalanced datasets, and enhances the objectivity and reproducibility of the evaluations. For the Daphnet dataset, the data was split into training, testing, and validation sets based on subject-wise experiments, following the approach in [12], as shown in Table 3. The Multimodal dataset was randomly divided into training and testing sets with a 4:1 ratio, following the method in [14]. Additionally, stratified five-fold cross-validation was performed on data from eight subjects, ensuring a robust evaluation of the model's performance.

Accuracy measures the overall effectiveness of a classification model, indicating the proportion of correct predictions, both true positives and true negatives, out of all predictions made. Precision focuses on the quality of positive predictions. It represents the fraction of instances predicted as positive that are actually positive. High precision implies a lower rate of false positives, making this metric particularly important when the cost of false positives is high, such as in medical diagnostics. Recall, also known as sensitivity, measures the model's ability to identify all positive instances correctly. A high recall indicates that the model has a low rate of false negatives, which is crucial in scenarios where missing positive cases can have serious consequences. On the other hand, specificity measures the model's ability to correctly identify negative instances. It indicates the proportion of actual negatives that were correctly classified. Specificity is particularly important when distinguishing negative cases, a key concern, such as in fraud detection, where false alarms can be costly. The F1 score is the harmonic mean of precision and recall, providing a balanced measure that considers false positives and false negatives. The evaluation metrics for benchmarking are Eqs. (14)–(18).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{14}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{Recall} = \text{Sensitivity} = \frac{TP}{TP + FN} \tag{16}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{17}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{18}$$

(TP : True Positive, TN : True Negative, FP : False Positive, FN : False Negative)

Also, we used area under the curve (AUC) as performance metric. AUC values range from 0 to 1, with a value of 0.5 denoting a model with no discriminative ability.

4.3. Hyperparameter tuning

This section describes the hyperparameter tuning process, explaining how the proposed DL model parameters were determined. We

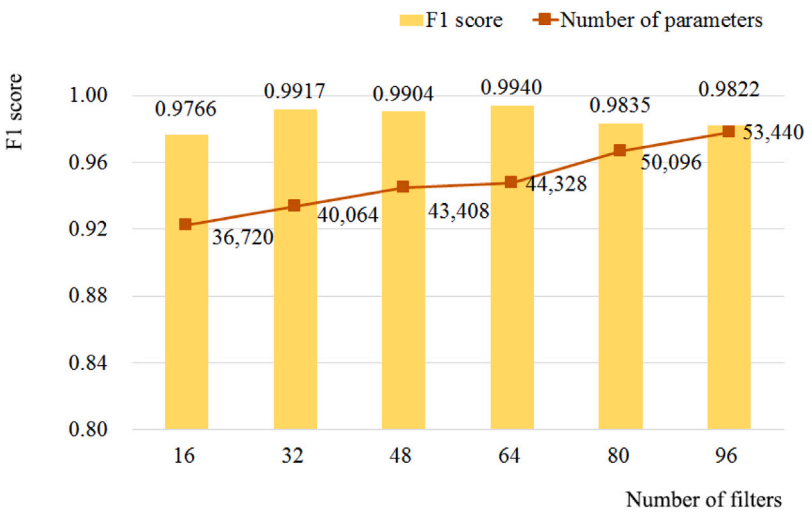

Fig. 6. The performance of the proposed DL model with increased convolutional layer filters on the Daphnet dataset.

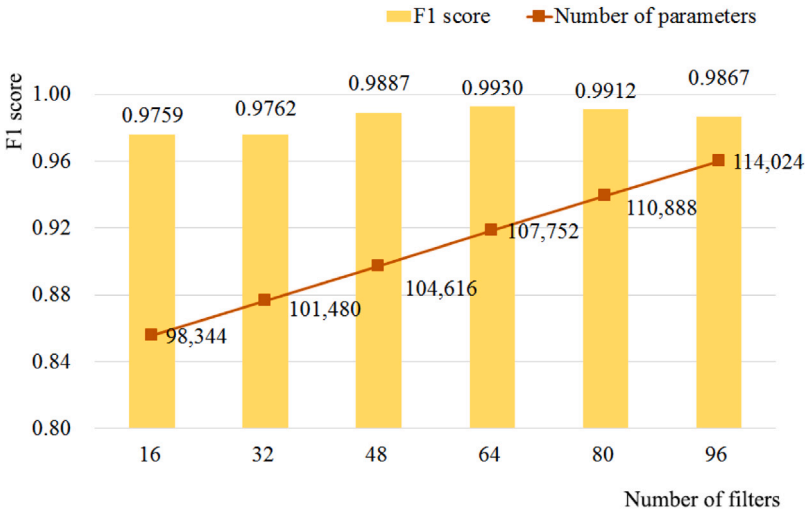

Fig. 7. The performance of the proposed DL model with increased convolutional layer filters on the Multimodal dataset.

focused on significant parameters such as the number of filters and kernel size in the convolutional layer, GRU neurons, and Dense neurons, which can be affected by the proposed DL model's performance. Our experiments were carried out using the Daphnet and Multimodal datasets, and we evaluated the model's performance using the F1 score.

4.3.1. Impact of the number of filters in the CNN layer

To achieve optimal performance in DL model, adjusting the number of filters in the convolutional layer is essential. The model can extract a broader range of features by increasing the filters. However, using more filters may result in overfitting. Figs. 6 and 7 illustrate the performance of the proposed deep learning model as the number of filters in the convolutional layer is varied. The x axis shows the number of parameters in the model, while the y axis indicates the F1 score related to the filter count. The highest F1 score was obtained with 64 filters in the convolutional layer for both the Daphnet and Multimodal datasets. When the number of filters exceeds 64, performance degrades due to overfitting.

4.3.2. Impact of the convolutional kernel size in the ECA module

Figs. 8 and 9 demonstrate the performance of the proposed DL model with an increased kernel size in the ECA module's convolutional layer. The F1 score plotted on the y-axis and the kernel size on the x-axis show a clear trade-off. For the Daphnet and Multimodal datasets, a kernel size of one resulted in the highest F1 scores, i.e., 0.9940 and 0.9930, respectively. In general, larger kernel sizes help capture broader contextual information. However, in scenarios with small and imbalanced datasets, complex attention mechanisms may lead to overfitting. Our experiments show that setting the kernel size of the ECA module to 1 yields the best performance by computing channel-wise attention weights independently, preserving the semantic integrity of handcrafted features and avoiding performance degradation from unnecessary inter-channel interactions.

4.3.3. Impact of the number of neurons in the GRU layer

The number of GRU neurons is crucial to the model's ability to capture complex patterns in sequential data. A higher number of neurons allows the model to encode intricate relationships, potentially leading to overfitting and increased computational demands, which can slow down training. Conversely, too few neurons may limit the model's ability to capture data diversity effectively. Therefore, selecting an optimal number of neurons is a delicate balance crucial for maintaining the model's learning capability and efficiency.

In the proposed DL model, the impacts of adding more neurons to the GRU layer are depicted in Figs. 10 and 11. The F1 score and the number of parameters in the proposed DL model are indicated by the y axis, while the x axis shows the number of neurons at the GRU layer.

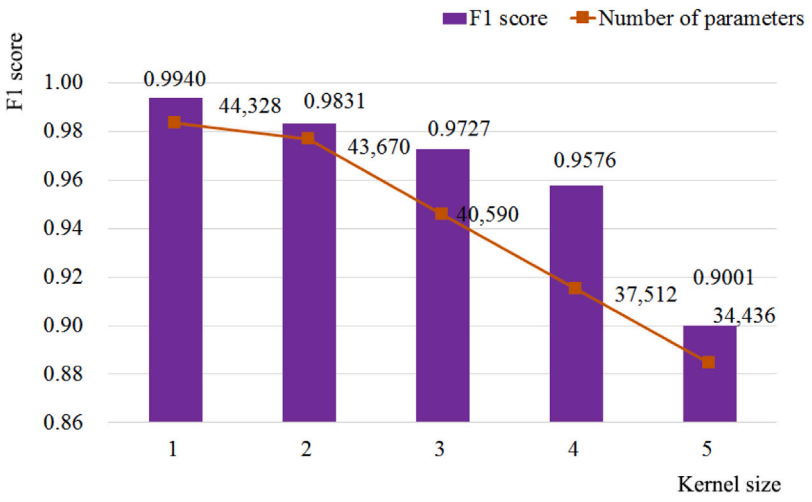

Fig. 8. The performance of the proposed DL model with an increased kernel size in the convolutional layer of the ECA module on the Daphnet dataset.

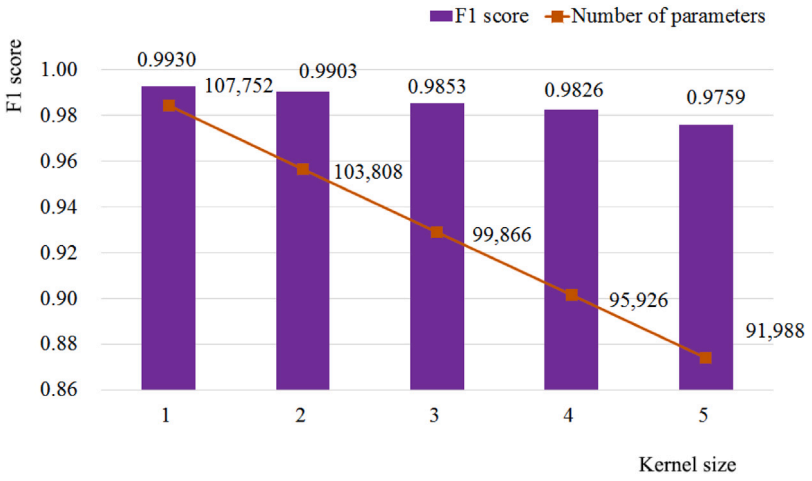

Fig. 9. The performance of the proposed DL model with an increased kernel size in the convolutional layer of the ECA module on the Multimodal dataset.

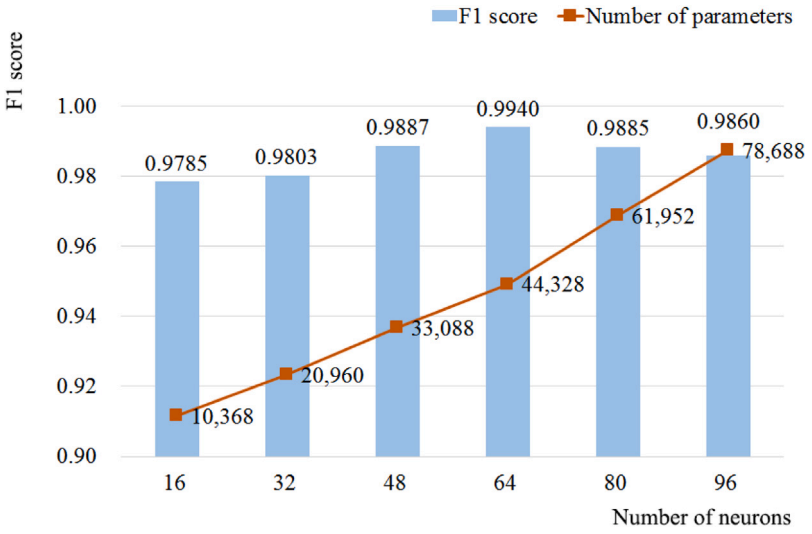

Fig. 10. The proposed DL model performance with increased the GRU layer's neuron size on the Daphnet dataset.

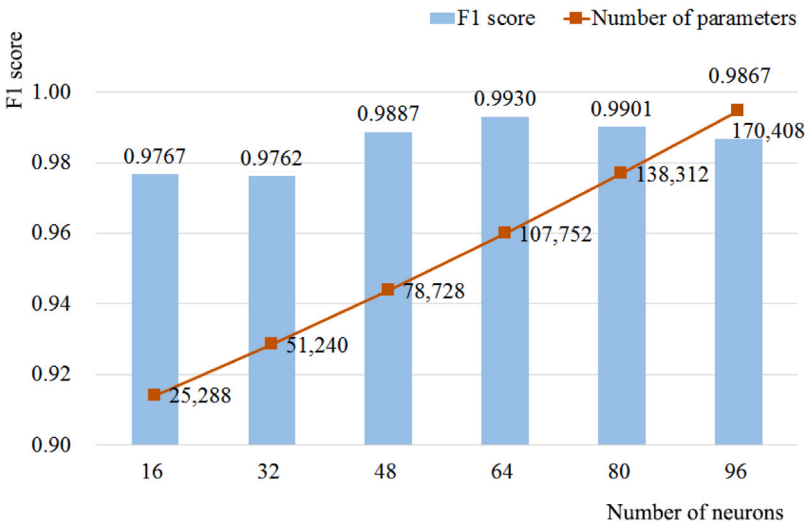

Fig. 11. The proposed DL model performance with increased the GRU layer's neuron size on the Multimodal dataset.

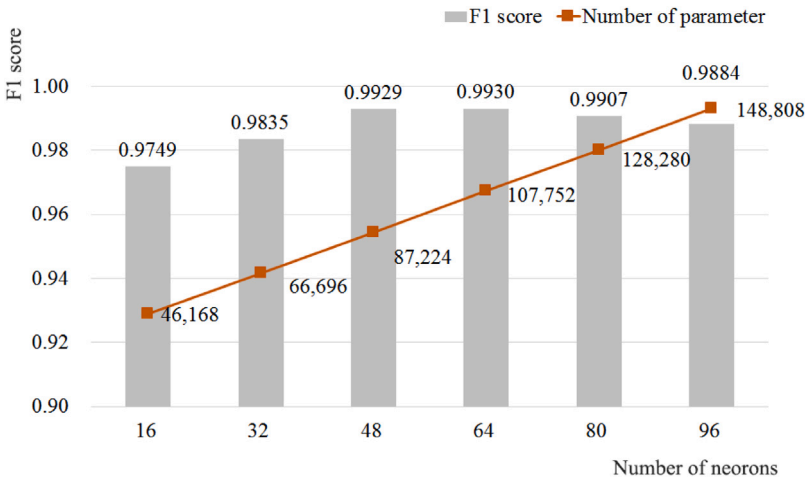

Fig. 12. The proposed DL model performance on Multimodal dataset with increased Dense layer neurons.

The size of the model grows as the number of neurons increases. The highest F1 scores (0.994 and 0.993, respectively) were obtained when the GRU layer's neuron count was 64 while classifying the Daphnet and Multimodal datasets. More features can be extracted if more neurons are at the GRU layer in the proposed DL model. The proposed DL model's F1 score can be increased, but overfitting with more than 64 neurons cannot be prevented.

4.3.4. *Impact of the number of neurons in the Dense layer*

The DL model's output can be tailored to different tasks by adjusting the number of neurons in the dense layer. A high number of neurons in the Dense layer allows for a more detailed understanding of complex relationships, but it also increases the risk of overfitting and requires more computational resources, potentially prolonging training times. Conversely, a small number of neurons could limit the model's ability to appreciate the diversity of the data. Therefore, it is crucial to identify the ideal number of neurons to ensure the model's success and efficiency in its learning process.

Figs. 12 and 13 show the impact of increasing the neuron count in the Dense layer of the proposed DL model. The y axis represents the F1 score, and the number of parameters in the model, and the x axis shows the number of neurons in the Dense layer. The complexity of the model increases with the number of neurons. For the Daphnet and Multimodal datasets, the best F1 scores of 0.994 and 0.993 were achieved with 64 neurons in the Dense layer, respectively. Although it does not significantly affect performance compared to other hyperparameters, increasing the number of neurons allows for more complex feature extraction, which could improve the DL model's F1 score. However, going beyond 64 neurons leads to performance degradation due to overfitting.

Table 4
List of selected hyperparameters.

Stage	Hyper-parameters		Value
Data Processing	Overlapping Window		50%
Model Architecture	Convolution	Kernel Size	1
		Stride	1
		Filters	64
		Activation	ReLU
	GRU	Neurons	64
	Dense	Neurons	64
Training	Optimizer		Adam
	Batch Size		32
	Learning Rate		0.001
	Number of Epochs		200

Based on the above results, we have selected the hyper-parameter values as shown in Table 4.

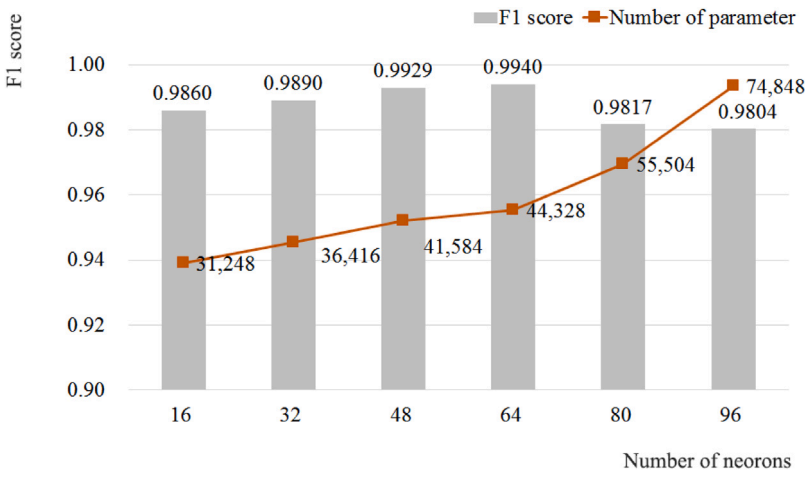

Fig. 13. The proposed DL model performance on Daphnet dataset with increased Dense layer neurons.

Table 5
Comparison of various models' performance using the Daphnet dataset.

Model	F1 score	Sensitivity	Specificity	AUC	No. of parameters
RF [49]	–	0.951	0.620	–	–
XGBoost [24]	0.970	0.960	–	–	–
CNN [39]	0.856	0.888	0.827	–	–
CNN [12]	0.930	–	–	–	628,538
CNN-LSTM [12]	0.930	–	–	–	4,025,714
LSTM [50]	–	0.906	0.693	–	–
BiLSTM [12]	0.920	–	–	–	238,669
iSPLInception [12]	0.940	–	–	–	1,326,726
Ours	0.994	0.996	0.992	0.994	44,328 (173.16 KB)

Table 6
Comparison of various models' performance using the Multimodal dataset.

Model	F1 score	Sensitivity	Specificity	AUC	No. of parameters
SVM [51]	0.907	0.830	0.900	–	–
Ensemble ML [21]	0.965	0.968	0.975	–	–
Ensemble DL [15]	0.920	0.810	0.850	–	–
CNN-BiLSTM [25]	0.958	–	–	–	–
Ours	0.993	0.892	0.996	0.992	107,752 (420.91 KB)

4.4. Experimental results

This section presents experimental results that are compared with those of current state-of-the-art techniques.

Table 5 compares various models' performance on the Daphnet dataset. Using this dataset, the proposed DL model performed better than the others. We exceeded the iSPLInception model [12], which got the best F1 value to date on the Daphnet dataset by 0.054 with an F1 score of 0.994. Notice that the proposed method relies exclusively on accelerometer and gyroscope data to simplify integration with wearable devices. Additionally, there are 29.9 times fewer parameters in our proposed DL model than in the iSPLInception model. The memory needed for the parameters is only 173.16 KB. The results of the proposed DL model on the Multimodal dataset are summarized in Table 6. Table 6 summarizes the results from training the proposed model on the Multimodal dataset. With an F1 score of 0.993, it performed better on this dataset than the other models—the parameters required only a tiny amount of 420.91 KB of memory. From the above results, the proposed model can achieve an F1 score of 0.994 with fewer parameters than state-of-the-art supervised learning methods. On the Daphnet and Multimodal datasets, the proposed FoG detection method yielded the best results, showing a notable decline in true positive rate concurrent with a reduction in the false positive rate.

Tables 7 and 8 show the pruning and dynamic quantization results for the proposed DL model on the Daphnet and Multimodal datasets, respectively. Table 7 shows that on the Daphnet dataset, quantization and pruning had no negative impact on the F1 score. However, the size of the DL model was reduced by 7.84 times. Similarly, the proposed DL model's F1 score on the Multimodal dataset (Table 8) was only 0.002 lower than the baseline DL model's, which was nearly the same. However, the model's size decreased by 10.2 times. These findings show that, compared to the baseline DL model, the proposed DL technique can considerably reduce the DL model size by more than

Table 7
The proposed DL model performance on Daphnet dataset after pruning and quantization.

Method	F1 score	Model Size
Baseline Model	0.994	173.16 KB
Dynamic Quantization (DQ)	0.994	55.02 KB
Pruning + DQ	**0.994**	**22.09 KB**

Table 8
The proposed DL model performance on Multimodal dataset after pruning and quantization.

Method	F1 score	Model size
Baseline Model	0.993	420.91 KB
Dynamic Quantization (DQ)	0.992	96.92 KB
Pruning + DQ	**0.991**	**44.04 KB**

Table 9
Ablation study on the Daphnet dataset with statistical significance (paired t-test, baseline = CNN+GRU, $N = 10$).

Model	F1 score (Mean ± SD)	Sensitivity	p-value
CNN+GRU	0.817 ± 0.012	0.830	–
CGNET	0.976 ± 0.010	0.980	0.0153
CGNET + ATT	0.990 ± 0.009	0.983	0.0224
CGNET + ECA	0.991 ± 0.008	0.994	0.0191
CGNET + ATT + ECA	0.994 ± 0.007	0.996	0.0105

Table 10
Ablation study on the Multimodal dataset with statistical significance (paired t-test, baseline = CNN+GRU, $N = 10$).

Model	F1 score (Mean ± SD)	Sensitivity	p-value
CNN+GRU	0.888 ± 0.011	0.896	–
CGNET	0.952 ± 0.012	0.969	0.0172
CGNET + ATT	0.975 ± 0.010	0.980	0.0216
CGNET + ECA	0.978 ± 0.009	0.987	0.0198
CGNET + ATT + ECA	0.993 ± 0.008	0.988	0.0093

7.84 times without sacrificing the F1 score. After pruning and dynamic quantization, the proposed model size was decreased by 234.42 times while maintaining high accuracy compared to state-of-the-art supervised learning methods.

4.5. Ablation study

This section presents the ablation studies conducted to evaluate the contributions of different components within the proposed deep learning model, tested on the Daphnet and Multimodal datasets. This study aims to analyze how each element affects overall performance and to determine the impact of specific components or modules. Initially, we developed a baseline DL model using CNN and GRU layers without optimized data preprocessing. To enhance performance, we implemented an optimized data preprocessing method that included feature selection, resulting in an improved model called CGNET. Further improvements were made to CGNET by integrating the ATT and ECA modules. To maximize performance, the final model incorporated the optimized data preprocessing and the ATT and ECA modules. Tables 9 and 10 display the results of five distinct model variants, designed to assess these enhancements' individual and combined effects on detecting FoG in the Daphnet and Multimodal datasets. The configurations tested include the baseline model (CNN+GRU), the CGNET (CNN+GRU with optimized data preprocessing), and CGNET versions augmented with the ATT and ECA modules.

- CNN+GRU: Combining a CNN and a GRU as the baseline DL model without optimized data preprocessing
- CGNET: The baseline DL model after optimized data preprocessing
- CGNET+ATT: CGNET with only the ATT module
- CGNET+ECA: CGNET with only the ECA module
- CGNET+ATT+ECA: CGNET with ATT and ECA modules

To assess the statistical significance of the performance improvements introduced by each module, we conducted a paired t-test using the F1 scores obtained from 10-fold cross-validation ($N = 10$). This determined the statistical significance ($p < 0.05$) of model improvements compared to the baseline CNN+GRU. As shown in Tables 9 and 10, each variant of the model incorporating the ATT and/or ECA modules demonstrated statistically significant improvements in F1 score compared to the baseline CNN+GRU model ($p < 0.05$). The complete model (CNN+GRU + ATT + ECA) delivered the best performance with the lowest p-values, reaffirming the statistical significance of the performance improvements and the model's robustness in enhancing detection performance.

As shown in Tables 9 and 10, we can observe that the baseline DL model, which uses CNN and GRU, significantly improved the detection of FoG episodes when combined with the optimized data preprocessing technique and the proposed feature selection mechanism. On the Daphnet dataset, the baseline CNN+GRU model achieved an F1 score of 0.817. The CGNET model, with optimized data processing, significantly improved the performance, reaching an F1 score of 0.976. Furthermore, the CGNET + ATT model, which incorporated the ATT module, achieved an F1 score of 0.990. The CGNET + ECA model, which included the ECA module, attained an F1 score of 0.991. Notably, the CGNET + ATT + ECA configuration, which comprised all components, achieved the highest F1 score of 0.994. The baseline CNN+GRU model reached an F1 score of 0.888 on the Multimodal dataset. The CGNET model achieved an F1 score of 0.952, the CGNET + ATT model achieved 0.975, and the CGNET + ECA model reached 0.978. Finally, the integrated CGNET + ATT + ECA configuration attained an F1 score of 0.993. The findings from this ablation study provide several important insights. The CGNET model's significant improvement over the baseline model emphasizes the critical importance of optimized data preprocessing in enhancing model performance. Furthermore, incorporating the ATT module into the model configuration improves performance by enabling the model to concentrate on the most relevant features, thereby increasing predictive accuracy. Additionally, the ECA module enhances performance by effectively capturing channel-wise relationships and refining the feature extraction process. Consequently, combining the ATT and ECA modules in a baseline DL model that includes CNN and GRU layers and optimized data processing can maximize model performance. This demonstrates a synergistic effect, where the strengths of each module complement one another.

Table 11
Performance evaluation of the proposed DL model based on different feature selection methods.

Dataset	Feature selection	F1 score
Daphnet	Correlation Coefficient	0.927
	Chi-square Test	0.972
	Information Gain	0.874
	Proposed Method	**0.994**
Multimodal	Correlation Coefficient	0.941
	Chi-square Test	0.980
	Information Gain	0.971
	Proposed Method	**0.993**

Table 11 shows the performance of the proposed DL model using various feature selection methods on the Dalphnet and Multimodal datasets. We can see that the proposed feature selection method outperforms commonly used techniques. While the proposed feature selection method achieved the highest F1 scores, the performance gap compared to the Chi-square Test was only about 0.01 0.02. Given the simplicity and interpretability of the Chi-square Test, its use could be justified in some instances. However, the proposed method offers added value by considering feature interactions and nonlinear relationships through model-based evaluation. It also showed consistent performance across datasets and maintained robustness in noisy or imbalanced conditions. Therefore, despite the small margin, the method provides practical benefits regarding generalization and adaptability for wearable deployment.

4.6. Discussion and future work

4.6.1. Discussion

This section discusses major considerations for deploying the proposed model on wearable devices, focusing on inference time and power consumption. The trained deep learning model is converted into TensorFlow Lite format, applying pruning and dynamic quantization to reduce its size and computational cost. This optimized model is suitable for deployment on development boards like the Arduino Nano

33 Sense, commonly used in wearable devices. To facilitate DL model deployment into wearable devices, libraries such as TensorFlow Lite for Microcontrollers must be installed, and the inference code must be uploaded to the device.

Reducing the size of DL models is a common strategy for wearable deployment; however, it often challenges the balance between model size and predictive accuracy. Parameter reduction and quantization can lower computational demands but may degrade accuracy. The proposed model overcomes this issue, achieving high performance without accuracy loss, even after pruning and dynamic quantization, as demonstrated in Tables 7 and 8. Nevertheless, implementing the model on actual hardware is necessary to validate its performance comprehensively in real-world scenarios.

Reducing the size of a model alone does not ensure real-time inference. Inference time and battery consumption depend on various factors, including the complexity of the input data, the number of sensors, and hardware limitations. Therefore, balancing memory usage, power consumption, and actual inference performance is crucial while reducing model size. Fanariotis et al. [52] demonstrated that Quantization-Aware (QA) training significantly decreased memory usage, inference time, and power consumption for the LeNet-5 model, which had a size of 240 KB on an STM32H7 board. With QA training, the LeNet-5 model size was reduced to 144 KB, and the Post Quantization (PQ) model size was further compressed to 120 KB. The inference times for the uncompressed, the QA, and the PQ models were 13 μs, 12 μs, and 12.5 μs, respectively. The power consumption of each model was measured as 526.2 mW, 525.2 mW, and 524 mW, showing no significant differences. Similarly, Ghuman et al. [38] deployed a CNN converted to TensorFlow Lite on an Arduino Nano 33 Sense, achieving inference times of under 5 s with only 0.3 MB of flash memory. Lin et al. [39] developed a lightweight model utilizing 478 KB of memory, which achieved a 0.21-second inference time on a 32-bit microcontroller with an ARM Cortex-M4 CPU. Based on these results, it can be anticipated that the proposed DL model, with a smaller size of 44.04 KB, is well-suited for deployment on wearable devices. However, further validation is necessary to evaluate its real-world performance concerning inference time, accuracy, and energy consumption. This additional assessment will determine whether the model is efficient and practically viable for real-time applications.

Dynamic adaptation is another critical challenge for wearable devices, as varying user activities and environmental conditions require real-time flexibility. While lightweight models offer some efficiency, they often struggle with such variations. To address this, DL model implementations should incorporate network architecture optimization, on-device learning, and energy-efficient design strategies. These enhancements enable better adaptability, improve performance, and maximize overall efficiency, including extended battery life and enhanced inference accuracy.

Lastly, comprehensive energy consumption analysis is vital for FoG detection in wearable environments. Most studies, including this one, focus on the inference phase and propose post-processing techniques to optimize energy use. However, other components like data acquisition and preprocessing are often overlooked. Future research should evaluate energy consumption across all operational phases. To address these challenges, the proposed DL model will be implemented on wearable devices and tested in real-world scenarios to provide deeper insights into practical energy usage and facilitate the development of optimized solutions.

4.6.2. Future prospects and limitations

The proposed lightweight DL model has not only demonstrated high accuracy and computational efficiency for detecting freezing of gait (FoG) in patients with Parkinson's disease, but it also holds significant promise for broader applications in age-related health monitoring and early disease prediction. First, the model's core contributions — including its compact neural architecture, pruning and dynamic quantization strategies, and feature selection approach — can be extended to other clinical conditions such as physical frailty, mild cognitive impairment (MCI), and early-stage dementia. Recent large-scale studies [53–58] have shown that wearable accelerometers' physical activity (PA) data can be transformed into digital biomarkers by extracting high-level features such as activity rhythm, variability, periodicity, and entropy. These approaches share strong similarities with the model proposed in this study, indicating promising generalizability to other neurodegenerative or functional decline-related conditions. Second, the model could evolve into a multi-task framework capable of monitoring multiple health indicators simultaneously. This includes detecting general mobility decline, increased fall risk, or abnormal gait patterns in real-time. As a core module in an integrated, lightweight health monitoring system, such a model could enable continuous tracking of aging-related changes and support early clinical interventions to improve the quality of life in older adults. However, it is important to note that the current model was evaluated only on publicly available benchmark datasets, and its applicability to real-world scenarios is yet to be confirmed. It is crucial that we conduct further studies involving the long-term deployment of wearable devices in naturalistic environments, across diverse populations, disease stages, and activity contexts, to ensure robust validation. Finally, while this study employed a supervised learning approach, the rarity of FoG events and the high cost of manual labeling present significant scalability challenges. It is imperative that future research explores innovative strategies, such as contrastive learning or diffusion-based pretraining, to enhance the model's generalization capabilities while reducing dependence on labeled data.

5. Conclusion

FoG is a common symptom of Parkinson's disease that significantly increases the risk of falls and reduces patients' quality of life. Effective real-time monitoring and management of this symptom on wearable devices is essential. However, existing DL models for FoG detection often rely on complex architectures with numerous parameters and sensors to achieve high accuracy. This complexity results in high computational demands and significant memory usage, making these models unsuitable for deployment on wearable devices.

To address these limitations, we developed a lightweight DL model that integrates CNN and GRU with ECA and ATT mechanisms. The model's performance was optimized through hyperparameter tuning, systematically identifying the best parameter values. Additionally, pruning and dynamic quantization techniques were applied to reduce model size while maintaining computational efficiency. To further enhance performance, we introduced a new feature selection algorithm that prioritizes the most critical features, ensuring robust performance even with limited resources.

Performance results on the Daphnet dataset demonstrated an F1 score of 0.994, surpassing the previous state-of-the-art F1 score of 0.94 reported in [12]. The proposed model achieved this performance while using 29.9 times fewer parameters and reducing memory usage to just 173.16 KB, making it highly suitable for wearable deployment. On the Multimodal dataset, the model achieved an F1 score of 0.993, with a memory requirement of only 420.91 KB, further demonstrating its feasibility for wearable environments. Moreover, pruning and dynamic quantization reduced the model size by 7.84 times and 10.2 times, respectively, with minimal performance degradation. These findings indicate that the proposed lightweight DL model enables real-time FoG detection on wearable devices, facilitating early medical interventions, reducing healthcare costs, and ultimately improving patients' quality of life.

Despite these promising results, the generalization performance and practical applicability of the proposed DL model have yet to be validated in real-world environments. In future work, we plan to implement the model on an actual wearable device to assess its performance under realistic constraints, including computational efficiency, latency,

and scalability across diverse scenarios. Additionally, data imbalance remains a challenge, as FoG episodes occur far less frequently than non-FoG episodes. To address this, an unsupervised learning approach for FoG detection is necessary. Moving forward, we aim to extend our work by incorporating unsupervised learning techniques using a diffusion model to improve FoG detection on wearable devices.

CRediT authorship contribution statement

Myung-Kyu Yi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Seong Oun Hwang:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) through the Korean Government (Ministry of Science and ICT) under Grant RS-2024-00340882.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.110138.

References

[1] Minming Gu, Zhixiang Chen, Kaiyu Chen, Haipeng Pan, IR-ST: A lightweight transformer network for human fall detection based on FMCW radar, IEEE Sens. J. 23 (2023) 25128–25135, http://dx.doi.org/10.1109/JSEN.2023.3314407.

[2] Sunil Kumar Bansal, Bijit Basumatary, Rajinder Bansal, Ashish Kumar Sahani, Techniques for the detection and management of freezing of gait in Parkinson's disease – A systematic review and future perspectives, MethodsX 10 (2023) 10210610, http://dx.doi.org/10.1016/j.mex.2023.102106.

[3] Rebecca J. Tarbert, Wamis Singhatat, Real world evidence of wearable smartbelt for mitigation of fall impact in older adult care, IEEE J. Transl. Eng. Heal. Med. 11 (2023) 247–251, http://dx.doi.org/10.1109/JTEHM.2023.3256893.

[4] Roongroj Bhidayasiri, Jirada Sringean, Saisamorn Phumphid, Chanawat Anan, Chusak Thanawattano, Suwijak Deoisres, Pattamon Panyakaew, Onanong Phokaewvarangkul, Suppata Maytharakcheep, Vijittra Buranasrikul, Tittaya Prasertpan, Rotjana Khontong, Priya Jagota, Araya Chaisongkram, Worawit Jankate, Jeeranun Meesri, Araya Chantadunga, Piyaporn Rattanajun, Phantakarn Sutaphan, Weerachai Jitpugdee, Marisa Chokpatcharavate, Yingyos Avihingsanon, Chanchai Sittipunt, Werasit Sittitrai, Grisada Boonrach, Aekamorn Phonsrithong, Pichit Suvanprakorn, Janprapa Vichitcholchai, Tej Bunnag, The rise of Parkinson's disease is a global challenge, but efforts to tackle this must begin at a national level: a protocol for national digital screening and eat, move, sleep lifestyle interventions to prevent or slow the rise of non-communicable diseases in thailand, Sec. Neurotechnol. 15 (2024) http://dx.doi.org/10.3389/fneur.2024.1386608.

[5] Frédéric Moisan, Sofiane Kab, Fatima Mohamed, Marianne Canonico, Morgane Le Guern, Cécile Quintin, Laure Carcaillon, Javier Nicolau, Nicolas Duport, Archana Singh-Manoux, Marjorie Boussac-Zarebska5, Alexis Elbaz, Parkinson disease male-to-female ratios increase with age: French nationwide study and meta-analysis, movement disorders, Neurol. Neurosurg. Psychiatry 87, 952–957, http://dx.doi.org/10.1136/jnnp-2015-312283.

[6] Mohd Halim Mohd Noor, Amril Nazir, Mohd Nadhir Ab Wahab, Jodene Ooi Yen Ling, Detection of freezing of gait using unsupervised convolutional denoising autoencoder, IEEE Access 9 (2021) 115700-115709, http://dx.doi.org/10.1109/ACCESS.2021.3104975.

[7] N. Naghavi, A. Miller, E. Wade, Towards real-time prediction of freezing of gait in patients with Parkinson's disease: Addressing the class imbalance problem, Sensors 19 (2019) 3898, http://dx.doi.org/10.3390/s19183898.

[8] Lloyd L.Y. Chan, Song Yang, Mira Aswani, Lauren Kark, Emily Henderson, Stephen R. Lord, Development, validation, and limits of freezing of gait detection using a single waist-worn device, IEEE Trans. Biomed. Eng. 71 (10) (2024) 3024–3031, http://dx.doi.org/10.1109/TBME.2024.3407059.

[9] Navleen Kour, Sunanda, Sakshi Arora, Computer-vision based diagnosis of Parkinson's disease via gait: A survey, IEEE Access 7 (2019) 156620-156645, http://dx.doi.org/10.1109/ACCESS.2019.2949744.

[10] Scott Pardoel, Jonathan Kofman, Julie Nantel, Edward D. Lemaire, Wearable-sensor-based detection and prediction of freezing of gait in Parkinson's disease: A review, Sensors 19 (23) (2019) 5141, http://dx.doi.org/10.3390/s19235141.

[11] Minglong Sun, Amanda Watson, Gang Zhou, Wearable computing of freezing of gait in Parkinson's disease: A survey, Smart Heal. 18 (2020) 100143, http://dx.doi.org/10.1016/j.smhl.2020.100143.

[12] Mutegeki Ronald, Alwin Poulose, Dong Seog Han, iSPLInception: An inception-ResNet deep learning architecture for human activity recognition, IEEE Access 9 (2021) 68985–69001, http://dx.doi.org/10.1109/ACCESS.2021.3078184.

[13] Y. Zhang, W. Yan, Y. Yao, J. Bint Ahmed, Y. Tan, D. Gu, Prediction of freezing of gait in patients with Parkinson's disease by identifying impaired gait patterns, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2022) 591–600, http://dx.doi.org/10.1109/TNSRE.2020.2969649.

[14] Wei Zhang, Zhuokun Yang, Hantao Li, Debin Huang, Lipeng Wang, Yanzhao Wei, Lei Zhang, Lin Ma, Huanhuan Feng, Jing Pan, Yuzhu Guo, Piu Chan, Multimodal data for the detection of freezing of gait in Parkinson's disease, Sci. Data 9 (2022) 1–10, http://dx.doi.org/10.1038/s41597-022-01713-8.

[15] Rishabh Bajpai, Suyash Khare, Deepak Joshi, A multimodal model-fusion approach for improved prediction of freezing of gait in Parkinson's disease, IEEE Sensors J. 23 (2023) 16168–16175, http://dx.doi.org/10.1109/JSEN.2023.3284656.

[16] Xia Yi, et al., A dual-modal attention-enhanced deep learning network for quantification of Parkinson's disease characteristics, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2019) 42–51, http://dx.doi.org/10.1109/TNSRE.2019.2946194.

[17] Kang Ren, Zhonglue Chen, Yun Ling, Jin Zhao, Recognition of freezing of gait in Parkinson's disease based on combined wearable sensors, BMC Neurol. 22 (2022) 229, http://dx.doi.org/10.1186/s12883-022-02732-z.

[18] Scott Pardoel, Julie Nantel, Jonathan Kofman, Edward D. Lemaire, Prediction of freezing of gait in Parkinson's disease using unilateral and bilateral plantar-pressure data, Front. Neurol. 13 (2022) http://dx.doi.org/10.3389/fneur.2022.831063.

[19] Kun Hu, Zhiyong Wang, Kaylena A. Ehgoetz Martensm, Markus Hagenbuchner, Mohammed Bennamoun, Ah Chung Tsoi, Simon J.G. Lewis, Graph fusion network-based multimodal learning for freezing of gait detection, IEEE Trans. Neural Netw. Learn. Syst. 34 (2023) 1588–1600, http://dx.doi.org/10.1109/TNNLS.2021.3105602.

[20] Dimitris Dimoudis, Nikos Tsolakis, Christoniki Magga-Nteve, Georgios Meditskos, Stefanos Vrochidis, Ioannis Kompatsiaris, InSEption: A robust mechanism for predicting FoG episodes in PD patients, Electronics 12 (2023) 2088, http://dx.doi.org/10.3390/electronics12092088.

[21] Kshitij Goel, Neetu Sood, Indu Saini, Ensemble technique based parkison's disease detection from FOG and EEG signals, in: Proc. of 2023 World Conference on Communication & Computing, WCONF, 14-16 July, 2023, http://dx.doi.org/10.1109/WCONF58270.2023.10235004.

[22] B. Shi, A. Tay, W.L. Au, D.M.L. Tan, N.S.Y. Chia, S.-C. Yen, Detection of freezing of gait using convolutional neural networks and data from lower limb motion sensors, IEEE Trans. Biomed. Eng. 69 (7) (2022) 2256–2267, http://dx.doi.org/10.1109/TNNLS.2021.3105602.

[23] Marc Bächlin, Meir Plotnik, Daniel Roggen, Inbal Maidan, Jeffrey M. Hausdorff, Nir Giladi, Gerhard Tröster, Wearable assistant for Parkinson's disease patients with the freezing of gait symptom, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 436–446, http://dx.doi.org/10.1109/TITB.2009.2036165.

[24] Hagar Elbatanouny, Natasa Kleanthous, Sundus Alusi, Soliman Mahmoud, Abir Hussain, Navigating the freeze: A machine learning approach to detect freezing of gait in Parkinson's patients, in: Proc 2024 IEEE International Conference on Omni-Layer Intelligent Systems, COINS, London, United Kingdom, 29-31 July, 2024, http://dx.doi.org/10.1109/COINS61597.2024.10622129.

[25] B. Lisha Preethi, V.E. Jayanthi, M. Indhumathy, Multimodel sensor based fog prediction in Parkinson's disease: A deep learning approach, in: Proc. of 2023 Innovations in Power and Advanced Computing Technologies, I-PACT, 08-10 December, 2023, http://dx.doi.org/10.1109/i-PACT58649.2023.10434791.

[26] STM32 Nucleo-144 development board with STM32H755ZI MCU, 2025, [Online]. Available: https://www.st.com/en/microcontrollers-microprocessors/stm32h755zi.html. (Accessed January 2025).

[27] Myung-Kyu Yi, Wai-Kong Lee, Seong Oun Hwang, A human activity recognition method based on lightweight feature extraction combined with pruned and quantized CNN for wearable device, IEEE Trans. Consum. Electron. 69 (2023) 657–670, http://dx.doi.org/10.1109/TCE.2023.3266506.

[28] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2014, arXiv:1409.0473.

[29] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, Qinghua Hu, ECA-net: Efficient channel attention for deep convolutional neural networks, in: Proceedings of the Computer Vision and Pattern Recognition, CVPR, June 13 2020 - June 19 2020. Seattle, WA, USA, https://arxiv.org/pdf/1910.03151.

[30] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation, in: Proc. of 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, Doha, Qatar, 2014, pp. 1724–1734, http://dx.doi.org/10.3115/v1/D14-1179.

[31] S.T. Moore, D.A. Yungher, T.R. Morris, V. Dilda, H.G. MacDougall, J.M. Shine, Naismith, S.J. Lewis, Autonomous identification of freezing of gait in Parkinson's disease from lower-body segmental accelerometry, J. Neuroeng. Rehabil. 10 (2013) 19:1–19:10.

[32] S. Mazilu, U. Blanke, M. Hardegger, G. Troster, E. Gazit, M. Dorfman, J.M. Hausdorff, GaitAssist: A wearable assistant for gait training and rehabilitation in Parkinson's disease, in: Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communication Workshops, PERCOM WORKSHOPS, Budapest, Hungary, 24–28 March, IEEE, Piscataway, NJ, USA, 2014, pp. 135–137.

[33] Benoît Sijobert, Jennifer Denys, Christine Azevedo Coste, Christian Geny, IMU based detection of freezing of gait and festination in Parkinson's disease, in: 2014 IEEE 19th International Functional Electrical Stimulation Society Annual Conference, IFESS, 2014, http://dx.doi.org/10.1109/IFESS.2014.7036751.

[34] L. Pepa, L. Ciabattoni, F. Verdini, M. Capecci, M.G. Ceravolo, Smartphone based fuzzy logic freezing of gait detection in Parkinson's disease, in: 2014 IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications, MESA, http://dx.doi.org/10.1109/MESA.2014.6935630.

[35] A.H. Esfahani, Z. Dyka, S. Ortmann, P. Langendörfer, Impact of data preparation in freezing of gait detection using feature-less recurrent neural network, IEEE Access 9 (2021) 138120–138131, http://dx.doi.org/10.1109/ACCESS.2021.3117543.

[36] X. Li, H. Xiong, X. Li, X. Wu, X. Zhang, J. Liu, et al., Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond, Knowl. Inf. Syst. 64 (12) (2022) 3197–3234, http://dx.doi.org/10.1007/s10115022-01756-8.

[37] N. Naghavi, E. Wade, Towards real-time prediction of freezing of gait in patients with Parkinson's disease: A novel deep one-class classifier, IEEE J. Biomed. Heal. Inform. 26 (4) (2022) 1726–1736, http://dx.doi.org/10.1109/JBHI.2021.3103071.

[38] Purnoor Ghuman, Tyama Lyall, Usama Mahboob, Alia Aamir, Xilin Liu, ECE496Y final report: Edge machine learning for detecting freezing of gait in Parkinson's patients, 2022, http://dx.doi.org/10.48550/arXiv.2211.06951.

[39] Ourong Lin, Tian Yu, Yuhan Hou, Yi Zhu, Xilin Liu, Edge deep learning enabled freezing of gait detection in Parkinson's patients, 2022, http://dx.doi.org/10.48550/arXiv.2212.00729.

[40] Scott Pardoel, Jonathan Kofman, Julie Nantel, Edward D. Lemaire, Wearable-sensor-based detection and prediction of freezing of gait in Parkinson's disease: A review, Sensors 19 (23) (2019) 5141, http://dx.doi.org/10.3390/s19235141.

[41] Arduino Nano 33 IoT, 2025, [Online]. Available: https://store.arduino.cc/products/arduino-nano-33-iot?srsltid=AfmBOop5kLPqa6rOLtPDqYIlVGra-6QnW7BKbR9uZRQ0JoXyTTvcToom. (Accessed January 2025).

[42] Arm Cortex-M4 in a nutshell, 2025, [Online] Available: https://www.st.com/content/st_com/en/arm-32-bit-microcontrollers/arm-cortex-m4.html. (Accessed January 2025).

[43] Jing Zhang, Jia Li, Weibing Wang, A class-imbalanced deep learning fall detection algorithm using wearable sensors, Sensor 21 (2021) 6511, http://dx.doi.org/10.3390/s21196511.

[44] Bruce Ratner, The correlation coefficient: Its values range between +1 and −1, or do they? J. Target. Meas. Anal. Mark. 17 (2009) 139–142, http://dx.doi.org/10.1057/jt.2009.5.

[45] Myung-Kyu Yi, Kyung hyun han and seong oun hwang, fall detection of the elderly using denoising LSTM-based convolutional variant autoencoder, IEEE Sensors J. 24 (11) (2024) 18556–18567.

[46] Mohamed Hamada, Jesse Jeremiah Tanimu, Mohammed Hassan, Habeebah Adamu Kakudi, Patience Robert, Evaluation of recursive feature elimination and LASSO regularization-based optimized feature selection approaches for cervical cancer prediction, in: Proc. of 2021 IEEE 14th International Symposium on Embedded Multicore/Many-Core Systems-on-Chip, MCSoC, 20-23 December, 2021, http://dx.doi.org/10.1109/MCSoC51149.2021.00056.

[47] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu, Squeeze-and-excitation networks, Comput. Vis. Pattern Recognit. (2018) http://dx.doi.org/10.48550/arXiv.1709.01507.

[48] F. Chollet, Keras: The python deep learning library, 2013, [Online]. Available: https://keras.io/. (Accessed Octorber 1 2024).

[49] S. Mazilu, M. Hardegger, Z. Zhu, D. Roggen, G. Troester, M. Plotnik, J. Hausdorff, Online detection of freezing of gait with smartphones and machine learning techniques, in: Proceedings of the 6th Int. Conf. Pervas. Comput. Technol. for Healthcare, San Diego, CA, USA, 2012, pp. 123–130, http://dx.doi.org/10.4108/icst.pervasivehealth.2012.248680.

[50] Y. Xia, J. Zhang, Q. Ye, N. Cheng, Y. Lu, D. Zhang, Evaluation of deep convolutional neural networks for detection of freezing of gait in Parkinson's disease patients, Biomed. Signal Process. Control. 46 (2018) 221–230, http://dx.doi.org/10.1016/j.bspc.2018.07.015.

[51] B. Lisha Preethi, V.E. Jayanthi, M. Indhumathy, Multimodal data for the detection of freezing of gait in Parkinson's disease, Sci. Data 9 (2022) 606, http://dx.doi.org/10.1038/s41597-022-01713-8.

[52] Anastasios Fanariotis, Theofanis Orphanoudakis, Konstantinos Kotrotsios, Vassilis Fotopoulos, George Keramidas, Panagiotis Karkazis, Power efficient machine learning models deployment on edge IoT devices, Sensors 23 (3) (2023).

[53] A. Mañas, J. Mota, P. De Souto Barreto, I. Rodríguez-Gómez, L. Rodríguez-Mañas, F.J. García-García, Longitudinal associations of physical activity and cognitive functioning: A systematic review and meta-analysis of prospective cohort studies, Ageing Res. Rev. 70 (2021) 101401, http://dx.doi.org/10.1016/j.arr.2021.101401.

[54] S. Belleville, S. Gauthier, É. Lepage, M.J. Kergoat, B. Gilbert, H. Chertkow, Consortium for the Early Identification of Alzheimer's Disease–Quebec (CIMA-Q), Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: A systematic review and meta-analysis, Neuropsychol. Rev. 27 (4) (2017) 328–353, http://dx.doi.org/10.1007/s11065-017-9361-5.

[55] R.P. Troiano, E. Stamatakis, F.C. Bull, How can global physical activity surveillance adapt to evolving physical activity guidelines? Needs, challenges and future directions, Br. J. Sports Med. 54 (24) (2020) 1468–1473, http://dx.doi.org/10.1136/bjsports-2020-102621.

[56] L. Fan, H. Ding, X. Chen, Y. Zhang, X. Wang, From physical activity patterns to cognitive status: Development and validation of novel digital biomarkers for cognitive assessment in older adults, NPJ Digit. Med. 5 (2022) 120, http://dx.doi.org/10.1038/s41746-022-00665-9.

[57] L. Fan, Y. Zhang, H. Ding, X. Huang, X. Wang, Predicting physical functioning status in older adults: Insights from wrist accelerometer sensors and derived digital biomarkers of physical activity, J. Biomed. Inform. 139 (2023) 104349, http://dx.doi.org/10.1016/j.jbi.2023.104349.

[58] L. Fan, Y. Liu, Y. Zhang, H. Ding, X. Wang, PFIMPA: A multimodal approach to predict physical function impairment in older adults using physical activity from wrist-worn accelerometers, Comput. Biol. Med. 169 (2024) 107671, http://dx.doi.org/10.1016/j.compbiomed.2024.107671.

Myung-Kyu Yi received a Ph.D. in Computer Science and Engineering from Korea University in 2005. He worked as a Research Professor with the Department of Computer Engineering, Gachon University, from 2005 to 2024. He is currently a Research Professor with Hanyang University, and is a member of the personal health information standardization task force of the TTA U-Health project group. His research interests include healthcare, security, machine learning, deep learning, and human activity recognition.

Seong Oun Hwang received the B.S. degree in mathematics from Seoul National University, in 1993, the M.S.degree in information and communications engineering from the Pohang University of Science and Technology, in 1998, and the Ph.D. degree in computer science from the Korea Advanced Institute of Science and Technology, in 2004, South Korea. He worked as a Software Engineer with LG-CNS Systems, Inc., from 1994 to 1996. He also worked as a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI), from 1998 to 2007. He worked as a Professor with the Department of Software and Communications Engineering, Hongik University, from 2008 to 2019. He is currently a Professor with the Department of Computer Engineering, Gachon University. His research interests include cryptography, cybersecurity, and artificial intelligence. He is an Editor of ETRI Journal.