

Received October 22, 2021, accepted November 17, 2021, date of publication November 30, 2021, date of current version December 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3131613

# Deep Learning-Based Multimodal Abnormal Gait Classification Using a 3D Skeleton and Plantar Foot Pressure

KOOKSUNG JUN<sup>ID</sup>, SANGHYUB LEE<sup>ID</sup>, DEOK-WON LEE<sup>ID</sup>,  
AND MUN SANG KIM<sup>ID</sup>, (Member, IEEE)

School of Integrated Technology, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea

Corresponding author: Mun Sang Kim (munsang@gist.ac.kr)

This work was supported in part by the Ministry of Trade, Industry and Energy of Korea, under Grant 20003762, and in part by the Gwangju Institute of Science and Technology.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

**ABSTRACT** Classification of pathological gaits has an important role in finding a weakened body part and diagnosing a disease. Many machine learning-based approaches have been proposed that automatically classify abnormal gait patterns using various sensors, such as inertial sensors, depth cameras and foot pressure plates. In this paper, we present a deep learning-based abnormal gait classification method employing both a 3D skeleton (obtained with a depth camera) and plantar foot pressure. We collected skeleton and foot pressure data simultaneously for 1 normal and 5 pathological (antalgic, lurching, steppage, stiff-legged, and Trendelenburg) gaits and classified them by using a multimodal hybrid model fed both data types together. In the proposed method, we fed the sequential skeleton and average foot pressure data into recurrent neural network (RNN)-based encoding layers and convolutional neural network (CNN)-based encoding layers, respectively, to effectively extract features from different data types. Their output features were concatenated and fed to fully connected layers for classification. The pressure-based and skeleton-based single-modal models achieved classification accuracies of 68.82% and 93.40%, respectively. The proposed multimodal hybrid model showed improved performance with an accuracy of 95.66%. We fine-tuned the hybrid model by applying a 3-step training methodology and ultimately increased the accuracy to 97.60%. This study indicates that the integrated features of the skeleton and foot pressure data represent both the spatiotemporal motion information and weight distribution, so data fusion can generate a positive effect in pathological gait classification.

**INDEX TERMS** Abnormal gait classification, multimodal sensors, data fusion, neural networks, hybrid deep learning.

## I. INTRODUCTION

Gait is an important biomedical indicator that supports a doctor or a physician in determining which body function of a patient is weakened. Therefore, many methods for analyzing human gaits by using various sensors, such as inertial sensors [1], [2], planar foot pressure sensors [3]–[7], depth cameras [8]–[15] or motion capture systems [16], [28], have been proposed. Sensor data can be used to calculate gait parameters, such as stride length, velocity, or the durations

of gait phases. Furthermore, they can be used to classify pathological gait patterns, but doing so manually requires complicated processes and equations that are much harder than calculating gait parameters. Therefore, many automatic abnormal gait detection methods have been proposed; in particular, machine learning-based approaches have risen to prominence.

Sensors for analyzing human gaits are divided into two groups: wearable sensors and nonwearable sensors. An inertial sensor is a typical wearable sensor for analyzing gait patterns, and it can be attached to hips, knees, and ankles. Three-axis acceleration and gyro data regarding the

The associate editor coordinating the review of this manuscript and approving it for publication was Nadeem Iqbal.

sensor-attached joint are obtained by using the sensor [1], [2]. Motion capture systems, such as Vicon (Vicon Motion Systems Ltd., Oxford, UK) and OptiTrack (Natural Point, Inc., Oregon, USA), are also wearable sensors for collecting motion data from walkers for gait analysis; walkers are required to attach markers on their body [16], [28]. These wearable sensors guarantee accurate sensor data, but they require complicated sensor systems or walkers that involve attaching uncomfortable sensors or markers to the body. Therefore, it is difficult to collect sufficient gait data in real life by using these wearable sensors.

On the other hand, nonwearable sensors do not require walkers that involve attaching sensors or markers on the body, so it is possible to collect plentiful gaits in real life. A planar foot pressure sensor, also called a pressure plate, is a typical nonwearable sensor for gait analysis [3]–[7]. The vertical pressure value on each load cell can be measured while a person is walking on the sensor. Such a sensor can accurately measure foot pressure while walking, but it requires considerable money. Therefore, the sensor generally covers a short distance of approximately a meter in length and is sometimes used in combination with a treadmill [17], [18]. Pressure sensor modules can also be installed under insoles, so they are sometimes used as wearable sensors [19]–[21].

Depth cameras, such as Azure Kinect (Microsoft Corp., Redmond, WA, USA), Astra (Orbbec 3D Technology International, Inc., Troy, MI, USA), and Realsense (Intel Corp., Santa Clara, CA, USA) depth cameras, can also be used to analyze gait patterns, and they were developed most recently among the various sensors used for gait analysis. 3D skeleton data obtained by using a depth camera can be employed to analyze gait patterns. The accuracy of 3D skeleton data is lower than that of wearable sensors or foot pressure sensors because 3D skeletons are secondary data that are estimated by using depth values. However, a depth camera can cover a longer walkway and is less expensive than a force pressure sensor, exhibiting acceptable accuracy. Furthermore, a depth camera can measure the motion of the whole body, while a force pressure sensor only focuses on foot pressure. Therefore, until recently, many studies have been conducted regarding gait analysis using depth cameras [8]–[15].

In this study, we proposed a novel method to classify five pathological gaits (antalgic, lurching, steppage, stiff-legged, and Trendelenburg gaits) and normal gaits by using 3D skeleton data obtained with Azure Kinect and foot pressure data obtained with GW1100 (GHiWell, Korea). We utilized deep learning-based data fusion for the classification process. In particular, we applied recurrent neural networks (RNNs) and convolutional neural networks (CNNs) for the 3D skeleton data and foot pressure data, respectively. This study aimed to verify the effectiveness of using the two types of data together in abnormal gait classification and to maximize their positive synergy effect. This research helps improve existing single-modal systems for abnormal gait recognition by applying multimodality.

The key contributions of the paper can be summarized as follows:

- The skeleton and pressure data of pathological gaits, which were captured simultaneously, are shared throughout this paper. The datasets help to improve gait-related studies and other multimodal approaches.
- The proposed multimodal pathological gait classification model that applies RNN-based and CNN-based encoding layers to skeleton and foot pressure data, respectively, achieved better performance than that of single-modal models.
- We verified that data fusion of the skeleton and foot pressure can generate a positive effect in pathological gait classification by representing both the spatiotemporal motion information and weight distribution.
- We maximized the classification performance of the multimodal hybrid model by applying a 3-step training method. The model was fine-tuned and trained to effectively find the global minimum loss. The results help to improve the performances of other multimodal classification models in various fields.

## II. RELATED WORKS

Many machine learning-based approaches have been developed for gait analysis using various sensors, such as inertial measurement unit (IMU) sensors, pressure or force plates, and depth cameras. IMU sensor data can be used for walking activity classification by applying support vector machines (SVMs) [55] and the hidden Markov model [56]. 3D skeleton data obtained by using depth cameras and foot pressure or ground reaction force data obtained by using pressure or force plates have been used to classify pathologies, such as Parkinson's disease [9], [10], [19], [20], autism spectrum disorder (ASD) [22], [23], stroke [5], [24], diabetes [4], and functional gait disorders [25], [26]. Furthermore, 3D skeleton data can be used to assess the quality of rehabilitation exercises [53]. In particular, as high-performance graphics processing units (GPUs) and deep learning technologies have been developed, many deep neural network (DNN)-based approaches have been published recently [10]–[15], [19], [53].

Force and pressure sensors have been steadily used for pathological gait classification until now. The ground reaction forces (GRFs) obtained from a force plate can be used for the classification process, and foot pressure data obtained by using a pressure plate also can be utilized. Several machine learning-based approaches have been proposed that detect Parkinson's disease by using spatiotemporal GRFs. CNNs [19], linear discriminant analysis [20], and decision trees [21] have been fed raw GRFs or extracted feature vectors and have achieved accuracies of 95.5%, 95%, and 99.4%, respectively. Furthermore, functional impairments associated with the calcaneus, ankle, knee and hip can be classified with 62% accuracy by applying an SVM-based classifier to features extracted from GRFs and the center of pressure

by using principal component analysis (PCA) [25]. Various trials have been conducted to detect other abnormal gaits using GRFs. The Ensemble AdaBoost tree classifier was fed selected kinetic and temporal features extracted from GRF data to detect leg length discrepancies with an accuracy of 89.9% in [27], and the k-nearest neighbor (k-NN) classifier was used to classify ASD with an accuracy of 83.33% in [23].

Foot pressure data can also be used to classify pathological gaits. Diabetic gaits can be recognized with a sensitivity of 95% and specificity of 90% by applying multivariate logistic regression to the pedographic variables of a percentage mask or anatomic mask obtained from foot pressure data [4]. The gaits of stroke patients were identified with an accuracy of 91.4% by applying an SVM classifier on plantar pressure characteristic parameters obtained by using density-based spatial clustering in [5]. Furthermore, the classification of forefoot pain with an accuracy of 70.4% by using a neural network-based classifier with features extracted from the center of pressure and plantar pressure parameters by using PCA [6] and a method for identifying working toes in young children by using an SVM (accuracy of 94.36%) and a random forest classifier (accuracy of 97.5%) with features extracted through PCA and discriminative mapping [7] have been introduced.

Many machine learning-based approaches have been proposed that use skeleton data obtained through depth cameras for pathological gait classification. An SVM with a Gaussian kernel was applied to gait parameters extracted from skeleton data to detect Alzheimer's disease with an accuracy of 92.31% in [8]. The k-NN classifier was applied to a covariance matrix converted from skeleton data to classify normal samples, hemiplegia, and Parkinson's disease with an accuracy of 79% in [9]. Decision trees, Bayesian networks, neural networks, and the k-NN classifier were compared to determine the approach with the best performance in the classification of Parkinson's disease stages when fed selected limb angles, bent angles, and the number of steps taken during a spin, and Bayesian networks showed the highest performance, with an accuracy of 93.40% in [10].

An RNN-based model recently received attention in skeleton-based pathological gait classification. Bidirectional long short-term memory (LSTM) was fed joint angle sequences to classify normal, in-toeing, out-toeing, drop-foot, pronation, and supination gaits with an accuracy of 88.90% in [11]. Normal, limping, and knee rigid gaits of different levels were classified with an accuracy of 95.90% by applying a gated recurrent unit (GRU) classifier to features extracted by using an LSTM-based autoencoder in [12]. Osteoarthritis can be detected with an accuracy of 98.77% by combining manually calculated features and automatically extracted features by using LSTM and applying an SVM classifier to the integrated features [13]. We previously proposed a method to classify normal, antalgic, stiff-legged, lurching, steppage, and Trendelenburg gaits with an accuracy of 93.67% by applying a GRU-based classifier to selected skeleton joints.

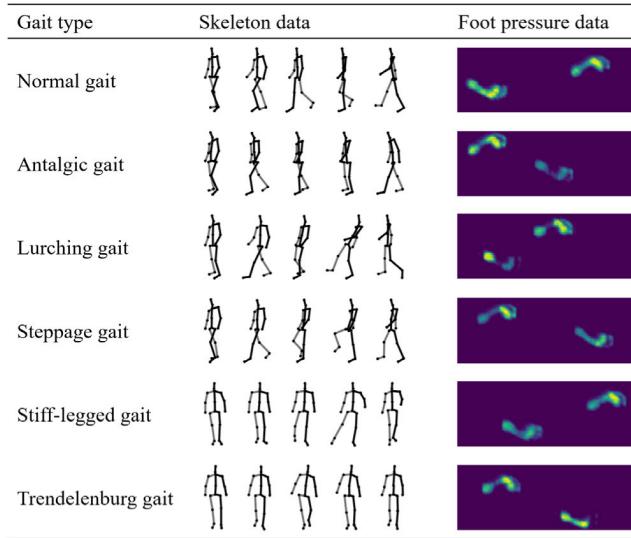
Multimodal data fusion can help to improve the gait analysis performance, and many studies have been introduced. Gait events, such as initial contact and toe off, were detected with an accuracy of 100% using IMUs and force sensors installed in shoes in [54]. A kernel SVM was fed features extracted from IMUs and plastic optical fiber-based floor sensors by using PCA and canonical correlation analysis to identify humans with an accuracy of 94% in [31]. Kinetic and kinematic gait features obtained from a 3D motion analysis system and two force plates were fed to linear discriminant analysis and quadratic discriminant analysis classifiers to classify ASD with an accuracy of 82.50% in [22]. An approach for identifying humans by using RGB images, depth images, and audio data was introduced in [34]. RGB and depth images were fed to PCA and linear discriminant analysis, and audio data were fed to an SVM with a linear kernel. Then, score-level fusion was conducted, and an accuracy of 99.4% was achieved.

Several approaches using neural network-based data fusion have also been proposed. Walking, fast walking, running, stair climbing/descending, and hill climbing/descending can be classified with an accuracy higher than 90% by using pressure, acceleration, and gyro data [32]. In this study, individual deep CNNs were used for each sensor dataset to extract features, and the concatenated features were fed to a DNN-based classifier. Data fusion of the scalograms and spectrograms obtained by using Wi-Fi sensing and radar sensors, respectively, was used to classify gait freezing episodes with an accuracy of 98.1% by applying a CNN-based autoencoder in [33].

Multimodal data fusion has been used in many gait-based classification approaches, but there are a few approaches for pathological gait classification using skeleton and foot pressure data together. Research analyzing the effectiveness of using them together for gait classification has not yet been proposed. However, some industrial studies have been introduced. In [46], a method to provide early warning for the posture of the human body using plantar pressure data and video information of the joints of feet, legs and shoulders was published. Furthermore, a gait analysis system using foot pressure and 3D skeleton data was introduced in [47]. Biologic age can be measured and abnormal gait can be detected in the system. However, the specific methods or analysis results of those industrial studies are not open to the public.

### III. MATERIALS

In this paper, we collected sequential skeleton datasets by using a single depth camera (Azure Kinect, Microsoft, USA), which is a recently released depth camera, and foot pressure data by using a pressure plate (GW1100, GHiWell, Korea). We obtained normal and five pathological (i.e., antalgic, lurching, steppage, stiff-legged, and Trendelenburg) gaits, which could be simulated by limiting physical functions and following guidelines [14]. Among various pathological gaits, we selected five pathological gaits by considering the



**FIGURE 1.** Skeleton and foot pressure data of normal and abnormal gaits for classification.

reproducibility in the simulation. It is difficult for subjects to simulate some gait patterns, such as gait of sensory problem or Parkinsonian gait, because they require stagger and random imbalance, which are difficult to define and generalize. Meanwhile, the selected five gait patterns can be clearly defined. Antalgic gait can be simulated by keeping the weight off the injured leg to minimize the pain; lurching gait can be simulated by lurching the trunk backward to compensate for abnormal hip extension at the heel strike of a damaged leg; steppage gait can be simulated by lifting the knee of the weakened leg higher than normal to avoid toe scraping on the ground; stiff-legged gait can be simulated by making a semicircle and not bending the problematic knee; Trendelenburg gait can be simulated by moving the weakened hip up and the opposite hip down because of problems in balancing the hip level during the stance phase. The subjects fully understood these definitions and causes of each gait before the data collection, so they could realistically simulate the pathological gaits.

Twelve healthy males participated in data collection. All of them were laboratory staff members and fully understood the data collection system. They agreed with the use of their data for research purposes. They had watched videos of each pathological gait and trained until they became familiar with simulating them. Data collection was conducted under strict supervision. When subjects walked through a 4-m walkway, sequential skeleton and planar foot pressure data were collected at the same time. Examples of the collected datasets are shown in Fig. 1. The datasets were shared on IEEE DataPort [59].

#### A. SKELETON DATA ACQUISITION

Azure Kinect is the latest Kinect sensor released. We obtained skeleton data by using the Azure Kinect sensor and the

corresponding Microsoft software development kit (SDK). We obtained the 3D XYZ coordinates of 33 joints, i.e., pelvis, spine\_naval, spine\_chest, neck, clavicles, shoulders, elbows, wrists, hands, hand tips, thumbs, hips, knees, ankles, feet, head, nose, eyes, and ears. We collected the skeletal gait data of the walkers on a 4-m walkway. Before data collection, we calibrated the sensor by using an ArUco [35] marker on the walkway. First, we detected the marker and measured the 3D XYZ coordinates of its corners. Then, we set the origin point and the XYZ directions as shown in Fig. 2 and transformed the original XYZ coordinates of the skeleton into the calibrated XYZ coordinates.

#### B. PLANTAR FOOT PRESSURE DATA ACQUISITION

GW1100 is a 1080 mm x 480 mm sized pressure plate. It has 6,144 high-voltage matrix sensors that can measure a maximum pressure of 100 N/cm<sup>2</sup>. In general, two steps of walking data can be collected by using the sensor. We set a 4-m walkway and installed a force plate on the middle of the walkway. We placed two pads next to the force plate to balance the height of the walkway and to allow subjects to walk naturally. In this paper, we used an average foot pressure that could be calculated by averaging the planar foot pressures of all time sequences. As shown in Fig. 1, the average foot pressure can be expressed as a 1-channel image.

We collected gait datasets for 12 people × 6 gait types × 20 instances, so a total of 1,440 skeleton and foot pressure instances were obtained. We augmented the datasets by reversing the left and right sides of the skeletons and foot pressures, so 2,880 instances could be used as the total datasets.

### IV. METHODS

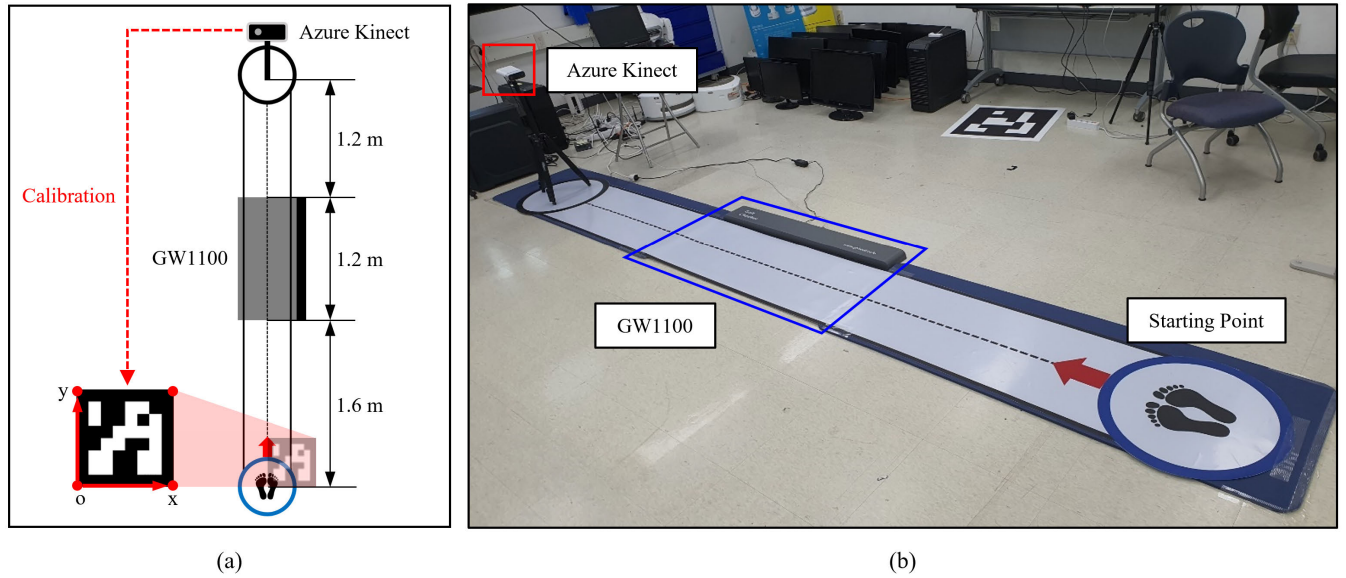
In this section, we describe a CNN encoder for extracting features from foot pressure data and an RNN encoder for extracting features from sequential skeleton data. Then, we introduce the proposed CNN-RNN hybrid model for pathological gait classification and a 3-step training method to improve the performance of the model.

#### A. CNN-BASED ENCODER FOR FOOT PRESSURE DATA

Average foot pressure data are similar to 1-channel image data. We therefore applied a CNN, which is known to achieve great performances on image processing and classification tasks, to extract features from the foot pressure data. We compared several CNN models to find the best model to process the foot pressure data. We evaluated the CNN models, i.e., the DenseNet [36], ResNet [37], Inception [38], Inception-ResNet [39], Xception [40] and MobileNet [41] architectures, by comparing their pathological gait classification performances using only foot pressure data. After we found the best CNN model to classify the gait patterns, we used the model to extract features from the foot pressure data for the hybrid classification model.

The foot pressure data  $p_{n,m}$ , where  $n \in \{1, \dots, N\}$  and  $m \in \{1, \dots, M\}$  denote the row and column of the pressure data,





**FIGURE 2.** Multimodal data collection system. (a) system specifications; (b) view of the system setup.

respectively, were fed to the CNN-based encoding layers. The pressure-based features  $\mathbf{f}_p$  were extracted using the following equation:

$$\mathbf{f}_p = \mathbf{W}_p \cdot \text{flatten}(E_{\text{CNN}}(\mathbf{p})) + \mathbf{b}_p \quad (1)$$

where  $\text{flatten}(\cdot)$  and  $E_{\text{CNN}}(\cdot)$  denote a neural calculation used to convert a 2-dimensional array matrix into a 1-dimensional array vector and a CNN-based calculation for processing the pressure data, respectively.  $\mathbf{W}_p$  and  $\mathbf{b}_p$  are the weight and bias used to convert the features to the desired size, respectively.

### B. RNN-BASED ENCODER FOR SKELETON DATA

Skeleton data are sequential data in a time series. Graph convolutional networks (GCNs) have recently risen to prominence in skeleton-based action recognition [48], [49]. RNNs are also known for their performance in skeleton-based action recognition [50], [51], and their performance in skeleton-based abnormal gait classification has been verified in many studies [11]–[15]. We used RNN-based encoding layers to extract features from the collected skeleton data. Among several RNN architectures, i.e., a simple RNN, LSTM [42] and a GRU [43], we found the best architecture for extracting features from the skeletal gait data. We evaluated the RNN architectures by comparing their classification accuracies for pathological gaits when using only skeletal gait data.

We did not use all joints as the input data for the RNN-based encoding layers. We excluded the joints of the face and arms (as they had bad influences on the classification performances [14]), so we used only 14 joints, i.e., pelvis, spine\_naval, spine\_chest, neck, clavicles, hips, knees, ankles, and feet. The selected joint data  $s_j^{(t)} = (x_j^{(t)}, y_j^{(t)}, z_j^{(t)})$ , where  $j \in \{1, \dots, J\}$  and  $t \in \{1, \dots, T\}$  are the indices of the joints and time steps, respectively, were fed to the RNN-based encoding layers. The skeleton-based features  $\mathbf{f}_s$  were

extracted as follows:

$$\mathbf{f}_s = \mathbf{W}_s \cdot E_{\text{RNN}}(\mathbf{s}) + \mathbf{b}_s \quad (2)$$

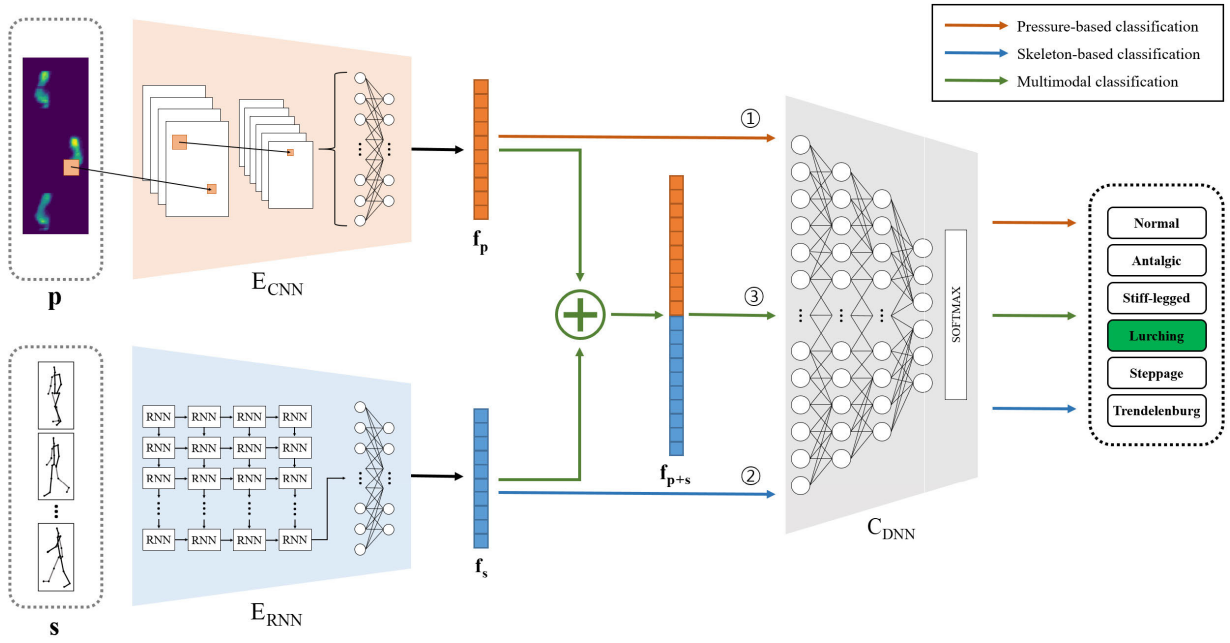
where  $E_{\text{RNN}}(\cdot)$  denotes the RNN calculation using a simple RNN, LSTM or GRU, and its output is the last element of the sequence resulting from the final RNN layer.  $\mathbf{W}_s$  and  $\mathbf{b}_s$  denote the weight and bias used to convert the features to the desired size, respectively. The simple RNN is known to have a vanishing gradient problem [44], [45], so LSTM and a GRU were designed to overcome this problem by applying a gated structure.

For LSTM, an input gate, a forget gate, an output gate, an input modulation gate and a cell state were added to reduce the long-term dependency problem of the simple RNN. The GRU has a similar gated structure to the LSTM, which consists of an update gate and a reset gate. The GRU achieves similar performance to the LSTM with fewer parameters. It is meaningful to compare the performances of LSTM and the GRU when they are used to construct the encoding layers for the sequential joint data. The best RNN architecture was used to extract features from the skeletal gait data for the hybrid classification model.

We constructed 4-layer RNN encoding layers, where the first two layers contained 256 hidden units and the other layers contained 128 hidden units. The activation function used for the recurrent step of the LSTM and GRU layer was sigmoidal. The first three layers return a sequence as output, but the last layer returns the last element of the sequence. We used 100 frames of the joint data as the inputs of the RNN-based encoding layers.

### C. FUSION OF FOOT PRESSURE AND SKELETON DATA

In this paper, we propose a hybrid classification model using multimodal fusion of acquired skeleton and foot pressure



**FIGURE 3.** Processes of single-modal classification and multimodal classification. In multimodal classification, the average foot pressure  $\mathbf{p}$  is fed to the CNN-based encoding layers  $E_{CNN}$ , and the selected joint  $\mathbf{s}$  is fed to the RNN-based encoding layers  $E_{RNN}$ . Output features  $\mathbf{f}_p$  and  $\mathbf{f}_s$  are concatenated and fed to the DNN-based classification layers  $C_{DNN}$ . In the 3-step training method, single-modal classification processes ① and ② are first conducted, and the best trained variables are saved. Then, the trained variables in the  $E_{CNN}$  and  $E_{RNN}$  are loaded, and multimodal classification process ③ are conducted.

data. Multimodal data fusion could help to improve the performances of machine learning models, so many studies have been proposed in various fields, including medical applications [29], [30]. The best CNN and RNN architectures were used to extract features from the foot pressure and skeleton data, respectively, and these architectures composed the proposed hybrid model. We applied feature-level data fusion, as shown in Fig. 3. The extracted features  $\mathbf{f}_p = (f_p^{(1)}, \dots, f_p^{(N_p)})$  and  $\mathbf{f}_s = (f_s^{(1)}, \dots, f_s^{(N_s)})$ , where  $N_p$  and  $N_s$  denote the numbers of extracted features from the pressure-based and joint-based encoding layers, respectively, were integrated into the multimodal features  $\mathbf{f}_{p+s}$  as

$$\begin{aligned} \mathbf{f}_{p+s} &= \text{concatenate}(\mathbf{f}_p, \mathbf{f}_s) \\ &= (f_p^{(1)}, \dots, f_p^{(N_p)}, f_s^{(1)}, \dots, f_s^{(N_s)}). \end{aligned} \quad (3)$$

The sequential skeleton data involve spatiotemporal information of gait motion, which can be used for gait classification. However, some pathological gaits are spatiotemporally similar to each other, so it is hard to distinguish them by using only vision-based data. The foot pressure information could supplement the limitation of the skeleton-based classification by providing a weight distribution. For example, antalgic and Trendelenburg gaits are similar to each other in terms of spatiotemporal gait motion. However, in the case of the antalgic gait, the walker tries to minimize the weight on the problematic leg because of pain. Therefore, there is a difference between the weight distributions of antalgic and Trendelenburg gaits. We hypothesized that they can be more

effectively distinguished when skeleton data are fused with foot pressure data. In this research, we applied CNN and RNN encoding layers to foot pressure and skeleton data, respectively, to maximize the implication of their features. The features are integrated, and the fused features indicate both the spatiotemporal motion information and weight distribution, so data fusion can generate a positive effect in pathological gait classification.

#### D. DNN-BASED CLASSIFICATION LAYERS

The extracted features  $\mathbf{f} \in \{\mathbf{f}_p, \mathbf{f}_s, \mathbf{f}_{p+s}\}$  were fed to the same DNN-based classification layers  $C_{DNN}$ , which have a 4-layer structure, and each fully connected layer contained 128, 64, 32, and 6 units. Before the features were fed to fully connected layers, dropout (0.5) and batch normalization were applied. We activated the last layer by using a softmax function as follows:

$$\hat{\mathbf{y}} = \text{softmax}(C_{DNN}(\mathbf{f})) \quad (4)$$

where  $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_6)$  is the output vector of the 6 gait patterns. We applied a cross-entropy loss function to calculate the loss  $L_{CE}$  and applied l2 regularizations to avoid overfitting as follows:

$$L_{CE}(\mathbf{y}, \hat{\mathbf{y}}) = - \sum_{i=1}^6 y_i \log(\hat{y}_i) + \frac{\lambda}{2} \|\mathbf{W}\|^2 \quad (5)$$

where  $\lambda$  and  $\mathbf{W}$  are the regularization parameter and trainable weights, respectively.

We selected the best CNN and RNN architectures to extract features from the pressure and skeleton data, respectively,

by comparing the classification accuracies of all architectures. In the pressure-based classification model, we extracted features by feeding the acquired foot pressure data to CNN encoding layers, and the features  $\mathbf{f}_p$  then entered the DNN-based classification layers. Similar to the process of the pressure-based classification model, we fed the features  $\mathbf{f}_s$  extracted from the skeletal gait data into the DNN-based classification layers of the skeleton-based classification model. In the multimodal classification model, we fed the pressure and skeleton data into the CNN-based and RNN-based encoding layers, respectively, and the integrated multimodal features  $\mathbf{f}_{p+s}$  entered the DNN-based classification layers.

### E. 3-STEP TRAINING OF THE HYBRID MODEL

To improve the classification performance of the proposed multimodal hybrid model, we applied a 3-step training method. The 1-step training process is defined as training the model by feeding the pressure and skeleton data into untrained weights and biases. The 3-step training process is defined as training the end-to-end hybrid model after training the weights and biases of the RNN-based and CNN-based encoding layers. In the first step of the 3-step training process, we trained the CNN-based single-modal classification model using only the foot pressure data. We saved the best trained weights in the CNN-based encoding layers. Second, we trained the RNN-based single-modal classification model using only the skeleton data and saved the best trained weights in the RNN-based encoding layers. During the first and second steps, the encoding layers of the CNN and RNN focus on extracting the most meaningful features from the foot pressure and skeleton data, respectively. We assumed that training with the foot pressure and skeleton data separately would improve the feature extraction performances of the CNN-based and RNN-based encoding layers by focusing on each data type separately. Finally, we loaded the trained weights of the CNN-based and RNN-based encoding layers and trained the whole end-to-end hybrid model. The trainable weights were fine-tuned during the final step.

### F. TRAINING SPECIFICATIONS

We evaluated the single-modal and multimodal classification models by using leave-one-subject-out cross-validation. The data of one subject were used as the validation data, and the rest of the data were used as the training data. We repeated the procedure 12 times and calculated the average validation accuracies and other statistical indices. During the training process, we applied the early stopping method to obtain the best validation accuracy. In general, the training process was stopped before 200 training epochs. The weights of the CNN layers were randomly initialized. For RNN layers, the biases were zero-initialized, the weights used to update input data were initialized with `glorot_uniform` [57], and orthogonal initialization was applied to the weights to update the recurrent state. Similarly, for the DNN layers, the biases were zero-initialized, and the weights used to update input data were initialized with `glorot_uniform`. We trained all the neural

networks-based models by applying the Adam [58] optimizer with a learning rate of 0.0001 and set the batch size to 30.

The configuration of the computer used in the experiments consisted of an Intel®Core™ i7-7700K central processing unit (CPU), 64 GB of random-access memory (RAM), and an NVIDIA GeForce RTX 2080-Ti GPU. In this paper, we implemented the models using TensorFlow.

## V. RESULTS AND DISCUSSION

In this section, we evaluate the CNN and RNN architectures to extract features from the foot pressure and skeleton data, respectively. We classified the gait patterns by using the constructed single-modal classification models using either the foot pressure or skeleton data to find the best CNN and RNN architectures. The best CNN and RNN architectures were used to construct the multimodal hybrid model. We evaluated the hybrid model by determining the improvements induced in the validation accuracy and other statistical indices over those of the single-modal models. Finally, we evaluated the 3-step training methodology by comparing the output it with the result of the 1-step training process.

### A. CLASSIFICATION PERFORMANCES OF THE SINGLE-MODAL CLASSIFICATION MODELS

We compared the feature extraction performances of the CNN architectures by constructing pressure-based classification models using them and computing the corresponding validation accuracy rates. Table 1 shows the validation accuracies obtained using each CNN architecture for the encoding layers. The DenseNet [36] architectures achieved the best performances among the CNN architectures; in particular, DenseNet201 yielded the highest accuracy of 68.82%. Xception [40] and InceptionResNet [39] obtained the second and third highest performances by achieving 64.27% and 61.28% accuracy rates, respectively. MobileNet [41] achieved the lowest performance in terms of classifying gait patterns with a 47.15% accuracy rate. Therefore, we constructed the CNN-based encoding layers of the multimodal hybrid model by using the DenseNet201 architecture to extract the most discriminative features from the foot pressure data and to achieve the optimal model performance.

Similar to the previous experiment, we compared the simple RNN, LSTM, GRU, and DNN to find the best RNN architecture for the skeleton-based encoding layers, as shown in Table 1. Furthermore, we compared them with various machine learning algorithms used in other skeleton-based gait classification studies, i.e., random forest [14], support vector machine [8], [10], k-nearest neighborhood [9], [10], and logistic regression [52]. We fed the 100 frames of the selected skeleton joints to the skeleton-based single-modal classification model and computed the resulting validation accuracy. The machine learning-based classifiers showed a lower performance than the neural networks-based classifiers. Random forest, SVM, k-NN, and logistic regression methods achieved accuracies of 68.06%, 73.68%, 74.44%, and 75.42%, respectively. They had difficulty comprehending

**TABLE 1. Validation accuracies of the single-modal classification models.**

Modality	Model	Accuracy (%)
Pressure-based classification	DenseNet121	65.83
	DenseNet169	68.40
	DenseNet201	68.82
	ResNet50	58.75
	ResNet101	57.74
	ResNet152	58.82
	Inception	60.14
	InceptionResNet	61.28
	Xception	64.27
	MobileNet	47.15
Skeleton-based classification	Random forest	68.06
	SVM	73.68
	K-NN	74.44
	Logistic regression	75.42
	DNN	83.89
	Simple RNN	86.35
	LSTM	91.94
	GRU	93.40

the spatiotemporal information of sequential skeleton data for the five pathological gaits. The RNN-based classifiers showed much better performance than machine learning-based and DNN-based classifiers. In particular, the GRU exhibited the best performance, with a 93.40% accuracy. The LSTM, simple RNN, and DNN architectures achieved 91.94%, 86.35%, and 83.89% accuracy rates, respectively. Therefore, we decided that the GRU could extract the most discriminative features from the sequential skeleton data and lead to the best performance of the multimodal hybrid model when the RNN-based encoding layers were composed of GRUs.

The skeleton-based classification model achieved much higher accuracy than that of the foot pressure-based classification model. The skeleton-based model composed of GRUs achieved 93.4% accuracy, which was 24.58% higher than the accuracy of the pressure-based classification model composed of the DenseNet201 architecture. In Table 2 and Table 3, we provide the accuracy, sensitivity, specificity, and precision of the pressure-based classification and the skeleton-based classification results, respectively. In the pressure-based classification results, the accuracy and specificity of each gait type were higher than 80%. However, the sensitivities of the steppage and Trendelenburg gaits were less than 50% at 48.54% and 43.96%, respectively. Furthermore, the overall precision was much lower than the accuracy and specificity. Regarding skeleton-based classification, the

**TABLE 2. Accuracy (%), sensitivity (%), specificity (%) and precision (%) of the pressure-based classification model (DenseNet201).**

Gait	Accuracy	Sensitivity	Specificity	Precision
Normal	96.35	95.42	96.54	84.66
Antalgic	92.57	82.08	94.67	75.48
Stiff-legged	88.13	64.79	92.79	64.26
Lurching	86.70	78.13	88.42	57.43
Steppage	87.12	48.54	94.83	65.27
Trendelenburg	86.77	43.96	95.33	65.33

**TABLE 3. Accuracy (%), sensitivity (%), specificity (%) and precision (%) of the skeleton-based classification model (GRU).**

Gait	Accuracy	Sensitivity	Specificity	Precision
Normal	99.10	95.63	99.79	98.92
Antalgic	94.97	86.25	96.71	83.98
Stiff-legged	99.55	98.75	99.71	98.54
Lurching	99.48	98.75	99.63	98.14
Steppage	99.34	97.50	99.71	98.53
Trendelenburg	94.38	83.54	96.54	82.85

overall results were higher than those of pressure-based classification. However, the skeleton-based classification model achieved relatively low performances when classifying the antalgic and Trendelenburg gaits. The sensitivities of the antalgic and Trendelenburg gait results were 86.25% and 83.54%, respectively. Furthermore, the precision rates of the antalgic and Trendelenburg gaits were 83.98% and 82.85%, respectively. Compared to those of the pressure-based classification results, the sensitivities and precisions of the antalgic and Trendelenburg gait results were much higher, but they still need improvement.

## B. CLASSIFICATION PERFORMANCE OF THE MULTIMODAL CLASSIFICATION MODEL

To improve the abnormal gait classification performance, we developed a hybrid classification model using multimodal fusion of the acquired skeleton and foot pressure data. We hypothesized that the multimodal classification model would alleviate the shortcomings of the single-modal classification models using either the pressure or skeleton data. To verify the hypothesis, we fed the same pressure and skeleton data into the untrained hybrid model (1-step training) and evaluated the model by training it with the same learning options and applying the same evaluation method (leave-one-subject-out cross-validation) as before. As a result, the hybrid model with the 1-step training process achieved 95.66% validation accuracy, which was 26.84% and 2.26% higher than those of the pressure-based and skeleton-based single-modal classification models, respectively.

Table 4 shows the accuracy, sensitivity, specificity and precision of the results of the hybrid model with 1-step training. Compared to the pressure-based classification model,



**TABLE 4. Accuracy (%), sensitivity (%), specificity (%) and precision (%) of the multimodal classification model with 1-step training.**

Gait	Accuracy	Sensitivity	Specificity	Precision
Normal	99.17	99.79	99.04	95.42
Antalgic	97.74	97.29	97.83	89.98
Stiff-legged	99.44	99.17	99.50	97.54
Lurching	99.41	98.13	99.67	98.33
Steppage	99.27	98.75	99.38	96.93
Trendelenburg	96.28	80.83	99.38	96.28

the hybrid model achieved much higher abnormal gait classification performance. All the statistical indices of all gait types were further improved by the hybrid model with 1-step training. Compared to those yielded by the skeleton-based classification model, most of the statistical indices were improved; in particular, the sensitivity of the antalgic gait increased from 86.25% to 97.29%, and the precision of the Trendelenburg gait increased from 82.85% to 96.28%. However, the sensitivity of the Trendelenburg gait was reduced from 83.54% to 80.83%.

### C. EFFECTIVENESS OF THE 3-STEP TRAINING PROCESS

We applied 3-step training to improve the performance of the multimodal hybrid model. We trained the CNN-based and RNN-based classification models by feeding them pressure and skeleton data, respectively. Then, we loaded the trained weights and biases onto the  $E_{CNN}$  and  $E_{RNN}$  of the hybrid model and trained the model on whole variables by feeding it the pressure and skeleton data together. We hypothesized that the proposed 3-step training method would improve the classification performance of the model by effectively training  $E_{CNN}$  and  $E_{RNN}$ . The feature extraction performances of the  $E_{CNN}$  and  $E_{RNN}$  might be lower when both of them are trained at the same time than when they are trained separately in single-modal classification models, where they are each influenced by a single type of data and could focus on each of the pressure and skeleton data points.

The proposed hybrid model with 3-step training achieved 97.60% accuracy, which was 1.94% higher than the accuracy of the same model with 1-step training. We first separately trained  $E_{CNN}$  and  $E_{RNN}$  by feeding them the pressure and skeleton data, respectively. Then, we trained all weights and biases after loading the pretrained  $E_{CNN}$  and  $E_{RNN}$  layers effectively increased the abnormal gait classification performance of the model. The hybrid model can be fine-tuned and trained to find the global minimum loss by using the pretrained weights and biases of the  $E_{CNN}$  and  $E_{RNN}$  layers. Table 5 shows the accuracy, sensitivity, specificity and precision for each gait type when the hybrid model was trained in 3 steps. The overall statistical indices were higher than those of the 1-step training model; in particular, the sensitivity of the Trendelenburg gait was improved from 80.83% to 92.71%. Furthermore, the hybrid model with 3-step training

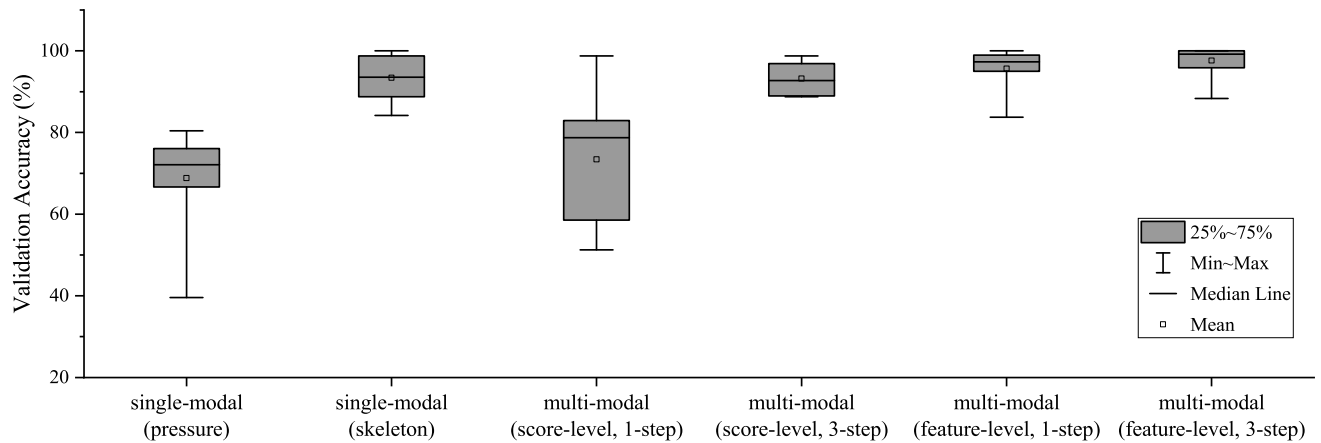
**TABLE 5. Accuracy (%), sensitivity (%), specificity (%) and precision (%) of the multimodal classification model with 3-step training.**

Gait	Accuracy	Sensitivity	Specificity	Precision
Normal	99.38	96.88	99.88	99.36
Antalgic	97.95	98.54	97.83	90.10
Stiff-legged	99.79	99.38	99.88	99.38
Lurching	99.93	99.58	100.00	100.00
Steppage	99.62	98.54	99.83	99.16
Trendelenburg	98.54	92.71	99.71	98.45

could achieve 100% specificity and precision for the lurching gait.

Feature-level-based fusion was applied to the proposed hybrid model. We conducted additional experiments by applying sum-ruled score-level fusion, which integrates the final score vectors of the skeleton-based classifier ( $\hat{y}_s$ ) and the foot pressure-based classifier ( $\hat{y}_p$ ), respectively. The score-level fusion model was trained with identical training specifications and showed 73.44% validation accuracy when 1-step training was applied, which was 4.62% higher than that of the pressure-based single-modal model and 19.96% lower than that of the skeleton-based single-modal model. Sum-ruled score-level fusion is effective when the models to combine show similar performances to each other because the data are fused in the most condensed shape. A large difference between the scores of the models makes it difficult for the fusion model to find the global minimum loss. Therefore, score-level fusion was not effective and the model was not well trained when 1-step training was applied since there was a large difference between the validation accuracies and losses of the pressure-based and the skeleton-based models. On the other hand, we achieved a 93.23% validation accuracy when applying 3-step training to the score-level fusion model. Since skeleton-based classification has much higher performance than pressure-based classification, the score-level fusion-based hybrid model selectively uses the score of skeleton-based classification while minimizing the impact of foot-pressure-based classification.

The validation accuracies of the single-modal classifications and multimodal classifications with changes in the level of data fusion and the training method are shown in Fig. 4. The pressure-based classification model achieved the lowest performance with a mean accuracy of 68.82%. The skeleton-based classification model achieved 93.40% mean accuracy, which was 24.58% higher than that of the pressure-based classification model. Therefore, it is much easier for the skeleton-based model to classify the given abnormal gait patterns than it is for the pressure-based model. However, the pressure-based model could assist the skeleton-based model in achieving better performance when feature-level fusion was applied. The multimodal hybrid model with feature-level fusion achieved higher accuracy than the skeleton-based model. In particular, the performance



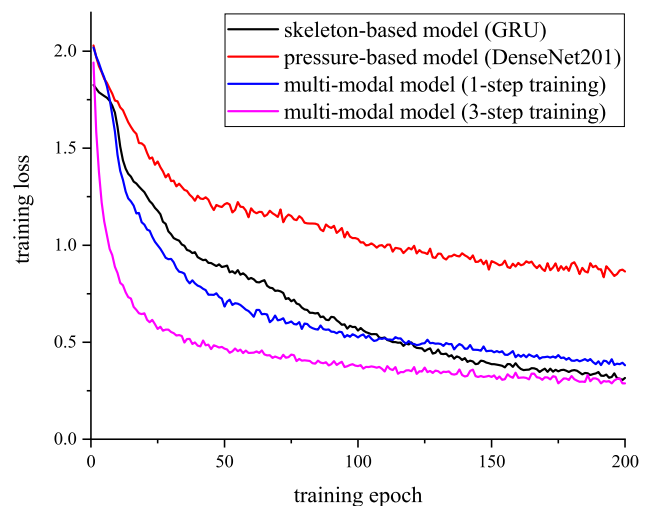
**FIGURE 4.** Validation accuracies of the single-modal classification models and multimodal classification models with the change in the level of data fusion and the training method.

of the hybrid model could be maximized by applying feature-level fusion and the 3-step training methodology.

The training losses of the pressure-based model (DenseNet201), the skeleton-based model (GRU), and the proposed multimodal hybrid models with 1-step and 3-step training are shown in Fig. 5. All the models were trained with same learning rate. The training loss of the multimodal model with 3-step training converged to the global minimum with the fastest speed, which means the model was trained most effectively. The multimodal model with 1-step training showed the second fastest convergence speed in the early phase. The final training loss of the skeleton-based model was lower than that of the multimodal model with 1-step training, but its validation loss was higher because of overfitting. Since the input data and the trainable parameters were much smaller in the skeleton-based model than in the multimodal models, overfitting started at an earlier point. The training loss of the pressure-based model converged the slowest, and its final value was also the highest among those of the models.

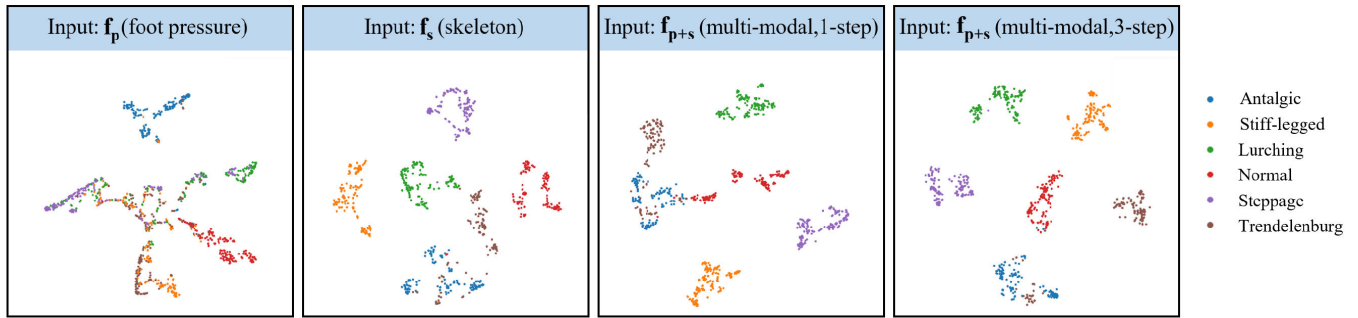
We applied t-distributed stochastic neighbor embedding (t-SNE) to show that the multimodal features  $\mathbf{f}_{p+s}$  are more recognizable than the single-modal features  $\mathbf{f}_p$  and  $\mathbf{f}_s$ , as shown in Fig. 6. The t-SNE was used to reduce the dimensionality of the input data, so the input data could be visualized in 2D or 3D space. We used the data of eight subjects for the training and the remainder for the validation and visualization. We extracted features from each well-trained model and fed them to the t-SNE model. The results show that the multimodal features of the hybrid model are more recognizable than the features of the foot pressure-based and skeleton-based single-modal classification models. In particular, the multimodal features with 3-step training showed the best performance. The features of the six gait patterns could be clearly distinguished by grouping, except some cases confused the Trendelenburg gait and the antalgic gait.

The confusion matrices of the single-modal classifications and multimodal classifications with 1-step and 3-step training

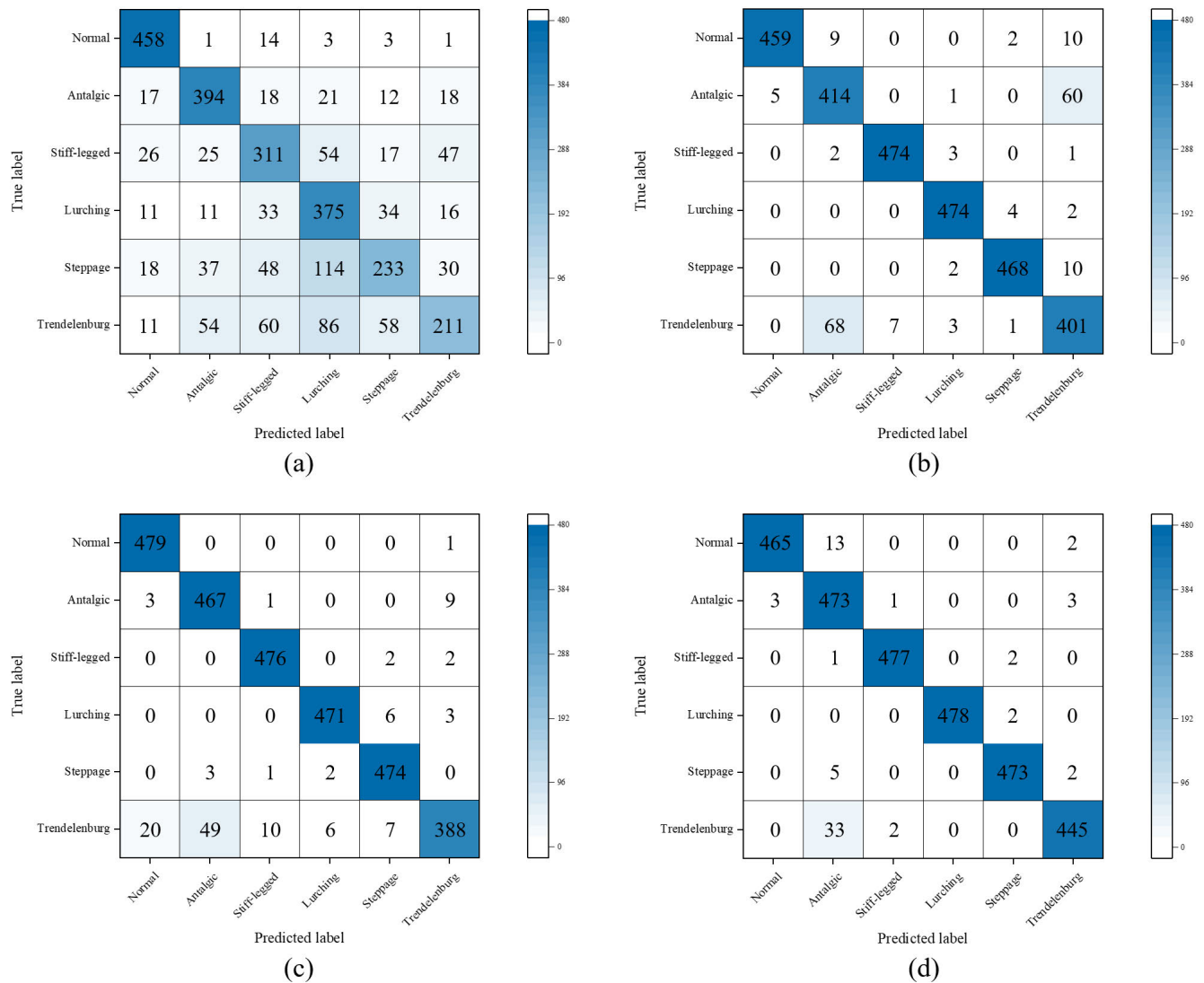


**FIGURE 5.** Training loss of the single-modal classification models and multimodal classification models with the 1-step and 3-step training processes.

are shown in Fig. 7. The pressure-based classification model could classify the normal gait better than the other gait types. The overall performances were low; in particular, the step-gait and Trendelenburg gaits were not classified well. The skeleton-based classification model exhibited much better performance than the pressure-based classification for the given gait types. However, the antalgic gait and Trendelenburg gaits were sometimes misclassified as each other. The hybrid model with 1-step training achieved improved performance over that of the skeleton-based model. In particular, the number of antalgic gaits misclassified as Trendelenburg gaits decreased from 60 to 9. However, there were more misclassifications of the Trendelenburg gait. Among the 480 Trendelenburg gait instances, only 388 instances were correctly classified in the hybrid model with 1-step training, while 401 instances were correctly classified by the skeleton-based model. The  $E_{CNN}$  and  $E_{RNN}$  layers might not have been



**FIGURE 6.** t-SNE visualization for different features.  $f_p$  and  $f_s$  are the output features of  $E_{CNN}$  and  $E_{RNN}$  for the foot pressure-based and the skeleton-based classification models, respectively.  $f_{p+s}$  is the integrated output feature of  $E_{CNN}$  and  $E_{RNN}$  for the multimodal hybrid classification model.  $f_{p+s}$  was extracted by using different training methods, i.e., 1-step training and 3-step training.



**FIGURE 7.** Confusion matrices of the single-modal classification model and multimodal classification model results with the 1-step and 3-step training processes. (a) pressure-based single-modal model (DenseNet201); (b) skeleton-based single-modal model (GRU); (c) multimodal model with 1-step training; (d) multimodal model with 3-step training.

perfectly trained to extract features from the pressure and skeleton data, respectively, during the 1-step training process. The hybrid model with 3-step training achieved the best

performance in terms of abnormal gait classification. In particular, the number of correctly classified antalgic gaits was remarkably increased compared to that of the skeleton-based

model. Furthermore, the number of correctly classified Trendelenburg gaits was largely increased compared to those of the skeleton-based model and the hybrid model with 1-step training. Therefore, the 3-step training process could effectively train the  $E_{CNN}$  and  $E_{RNN}$  layers and enable them to extract more discriminative features.

Our results are comparable to the performances in existing studies about machine learning-based gait classification. For foot-pressure-based classification, a sensitivity of 95% and specificity of 90% in identifying a diabetic foot with an ulcer were achieved by applying multivariate logistic regression in [4], where 90 nondiabetic patients and 120 diabetic patients participated; 91.4% accuracy (leave-one-subject-out) in recognizing stroke patients was achieved by using the SVM classifier in [5], where 9 healthy subjects and 17 stroke patients participated; 70.4% accuracy (training:test = 8:2) in recognizing forefoot pain was achieved by using neural networks in [6], where 297 subjects participated. For the skeleton-based classification, 92.31% accuracy, 96.33% sensitivity, 88.62% precision, and 90.81% specificity (training:test = 8:2) to detect Alzheimer's disease was achieved by using SVM with a Gaussian kernel in [8], where 30 healthy controls and 30 patients with Alzheimer participated. The classification of patients with Parkinson's disease into three stages achieved 93.40% accuracy (10-fold cross-validation) by applying Bayesian networks to selected features in [10], where 30 Parkinson's patients participated. Bidirectional LSTM achieved 88.90% accuracy (leave-one-subject-out) in classifying normal, in-toeing, out-toeing, drop-foot, pronation, and supination gaits in [11], where 16 healthy subjects simulated the gaits, and 768 data points were collected. Our proposed hybrid model achieved a 97.60% accuracy (leave-one-subject-out) to classify normal, antalgic, lurching, step-page, stiff-legged, and Trendelenburg gaits, where 12 healthy subjects simulated the gaits, and 2,880 data points were collected.

We achieved a relatively higher accuracy than other researches, although we classified more various gait patterns with larger data points. However, there were relatively few subjects, and the datasets were collected by simulation in our research. To compensate for the lower number of subjects, we increased the data points by repeating the simulation and applied the leave-one-subject-out cross-validation, which is appropriate to validate a model with few subjects. Furthermore, to compensate for not using data of real patients in the experiments, we selected the five pathological gaits that can be reproduced by following clearly defined manuals, and the subjects simulated them under strict management to collect as similar gait datasets to real patients as possible.

## VI. CONCLUSION

In this paper, we proposed a novel method to classify abnormal gaits by using a multimodal hybrid model that received the skeleton and foot pressure data. The proposed hybrid model showed improved classification performance relative to single-modal models fed with either skeleton or foot

pressure data. Furthermore, we applied a 3-step training method to maximize the performance of the model. This paper indicates that the deep learning-based fusion of the skeleton and foot pressure data could create a positive synergy effect for abnormal gait classification. The different encoding layers could effectively extract features from each type of input data. The 3-step training process could improve the performances of other multimodal classification models in various fields. These results will help improve the existing gait analysis applications by applying multimodality, so they can provide more accurate result to doctors or physicians in gait classification. In future work, we will collect datasets of real patients by collaborating with orthopedic, otolaryngology, and rehabilitation medical centers. We will evaluate the proposed hybrid model with real patient datasets and verify its potential for application in the real world.

## ACKNOWLEDGMENT

The authors are grateful to all staff and participants involved in the data collection process. The foot pressure plate (GW1100) used in this study was supported by P&S Mechanics, South Korea, and GHiWell, South Korea.

## REFERENCES

- [1] R. Caldas, M. Mundt, W. Potthast, F. B. de Lima Neto, and B. Markert, "A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms," *Gait Posture*, vol. 57, pp. 204–210, Sep. 2017.
- [2] A. Mannini, D. Trojaniello, A. Cereatti, and A. M. Sabatini, "A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and Huntington's disease patients," *Sensors*, vol. 16, no. 1, p. 134, 2016.
- [3] K. Deschamps, G. A. Matricali, P. Roosen, K. Desloovere, H. Bruyninckx, P. Spaepen, F. Nobels, J. Tits, M. Flour, and F. Staes, "Classification of forefoot plantar pressure distribution in persons with diabetes: A novel perspective for the mechanical management of diabetic foot?" *PLoS ONE*, vol. 8, no. 11, Nov. 2013, Art. no. e79924.
- [4] U. Waldecker, "Pedographic classification and ulcer detection in the diabetic foot," *Foot Ankle Surg.*, vol. 18, no. 1, pp. 42–49, Mar. 2012.
- [5] M. Wang, S. Yong, C. He, H. Chen, S. Zhang, C. Peng, and X. Wang, "Research on abnormal gait recognition algorithms for stroke patients based on array pressure sensing system," in *Proc. IEEE 3rd Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, Mar. 2019, pp. 1560–1563.
- [6] N. L. W. Keijsers, N. M. Stolwijk, J. W. K. Louwerens, and J. Duysens, "Classification of forefoot pain based on plantar pressure measurements," *Clin. Biomech.*, vol. 28, no. 3, pp. 350–356, 2013.
- [7] G. Guo, K. Guffey, W. Chen, and P. Pergami, "Classification of normal and pathological gait in young children based on foot pressure data," *Neuroinformatics*, vol. 15, no. 1, pp. 13–24, Jan. 2017.
- [8] M. Seifollahi, H. Soltanizadeh, A. H. Mehraban, and F. Khamseh, "Alzheimer's disease detection using skeleton data recorded with Kinect camera," *Cluster Comput.*, vol. 23, pp. 1469–1481, Nov. 2019.
- [9] Q. Li, Y. Wang, A. Sharf, Y. Cao, C. Tu, B. Chen, and S. Yu, "Classification of gait anomalies from Kinect," *Vis. Comput.*, vol. 34, no. 2, pp. 229–241, Feb. 2018.
- [10] L. Dranca, L. de Abetxuko Ruiz de Mendarozketa, A. Goñi, A. Illarramendi, I. N. Gomez, M. D. Alvarado, and M. C. Rodríguez-Oroz, "Using Kinect to classify Parkinson's disease stages related to severity of gait impairment," *BMC Bioinf.*, vol. 19, no. 1, p. 471, 2018.
- [11] Y. Guo, F. Deligianni, X. Gu, and G.-Z. Yang, "3-D canonical pose estimation and abnormal gait recognition with a single RGB-D camera," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3617–3624, Oct. 2019.
- [12] K. Jun, D.-W. Lee, K. Lee, S. Lee, and M. S. Kim, "Feature extraction using an RNN autoencoder for skeleton-based abnormal gait recognition," *IEEE Access*, vol. 8, pp. 19196–19207, 2020.



- [13] F. Chen, X. Cui, Z. Zhao, D. Zhang, C. Ma, X. Zhang, and H. Liao, "Gait acquisition and analysis system for osteoarthritis based on hybrid prediction model," *Computerized Med. Imag. Graph.*, vol. 85, Oct. 2020, Art. no. 101782.
- [14] K. Jun, Y. Lee, S. Lee, D.-W. Lee, and M. S. Kim, "Pathological gait classification using Kinect V2 and gated recurrent neural networks," *IEEE Access*, vol. 8, pp. 139881–139891, 2020.
- [15] D.-W. Lee, K. Jun, S. Lee, J.-K. Ko, and M. S. Kim, "Abnormal gait recognition using 3D joint information of multiple kinects system and RNN-LSTM," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 542–545.
- [16] T.-N. Nguyen, H.-H. Huynh, and J. Meunier, "Skeleton-based abnormal gait detection," *Sensors*, vol. 16, no. 11, p. 1792, Oct. 2016.
- [17] C. Nüesch, J.-A. Overberg, H. Schwameder, G. Pagenstert, and A. Mündermann, "Repeatability of spatiotemporal, plantar pressure and force parameters during treadmill walking and running," *Gait Posture*, vol. 62, pp. 117–123, May 2018.
- [18] U. Rashid, N. Kumari, D. Taylor, T. David, and N. Signal, "Gait event anomaly detection and correction during a split-belt treadmill task," *IEEE Access*, vol. 7, pp. 68469–68478, 2019.
- [19] A. S. Alharthi, A. J. Casson, and K. B. Ozanyan, "Gait spatiotemporal signal analysis for Parkinson's disease detection and severity rating," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1838–1848, Jan. 2021.
- [20] R. Alkhatib, M. O. Diab, C. Corbier, and M. E. Badaoui, "Machine learning algorithm for gait analysis and classification on early detection of Parkinson," *IEEE Sensors Lett.*, vol. 4, no. 6, pp. 1–4, Jun. 2020.
- [21] E. Balaji, D. Brindha, and R. Balakrishnan, "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease," *Appl. Soft Comput.*, vol. 94, Sep. 2020, Art. no. 106494.
- [22] C. Z. C. Hasan, R. Jailani, N. M. Tahir, I. M. Yassin, and Z. I. Rizman, "Automated classification of autism spectrum disorders gait patterns using discriminant analysis based on kinematic and kinetic gait features," *J. Appl. Environ. Biol. Sci.*, vol. 7, no. 1, pp. 150–156, 2017.
- [23] C. Z. C. Hasan, R. Jailani, N. Md Tahir, and S. Ilias, "The analysis of three-dimensional ground reaction forces during gait in children with autism spectrum disorders," *Res. Develop. Disabilities*, vol. 66, pp. 55–63, Jul. 2017.
- [24] S. J. Park, I. Hussain, S. Hong, D. Kim, H. Park, and H. C. M. Benjamin, "Real-time gait monitoring system for consumer stroke prediction service," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2020, pp. 1–4.
- [25] D. Slijepcevic, M. Zeppezauer, C. Schwab, A.-M. Raberger, C. Breiteneder, and B. Horsak, "Input representations and classification strategies for automated human gait analysis," *Gait Posture*, vol. 76, pp. 198–203, Feb. 2020.
- [26] D. Slijepcevic, M. Zeppezauer, A.-M. Gorgas, C. Schwab, M. Schuller, A. Baca, C. Breiteneder, and B. Horsak, "Automatic classification of functional gait disorders," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1653–1661, Sep. 2018.
- [27] S. Marquez J, R. Atri, M. R. Siddiquee, C. Leung, and O. Bai, "A mobile, smart gait assessment system for asymmetry detection using machine learning-based classification," *J. Biomed. Eng. Med. Devices*, vol. 3, no. 2, p. 135, 2018.
- [28] R. Rucco, V. Agosti, F. Jacini, P. Sorrentino, P. Varriale, M. D. Stefano, G. Milan, P. Montella, and G. Sorrentino, "Spatio-temporal and kinematic gait analysis in patients with Frontotemporal dementia and Alzheimer's disease through 3D motion capture," *Gait Posture*, vol. 52, pp. 312–317, Feb. 2017.
- [29] N. Srivastava and R. R. Salakhutdinov, "Multimodal learning with deep Boltzmann machines," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 2949–2980, Oct. 2014.
- [30] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Ng, "Multimodal deep learning," in *Proc. 28th Int. Conf. Mach. Learn.*, Bellevue, WA, USA, 2011.
- [31] S. U. Yunus, A. Alharthi, and K. B. Ozanyan, "Multi-modality fusion of floor and ambulatory sensors for gait classification," in *Proc. IEEE 28th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2019, pp. 1467–1472.
- [32] S.-S. Lee, S. T. Choi, and S.-I. Choi, "Classification of gait type based on deep learning using various sensors with smart insole," *Sensors*, vol. 19, no. 8, p. 1757, Apr. 2019.
- [33] S. A. Shah, A. Tahir, J. Ahmad, A. Zahid, H. Pervaiz, S. Y. Shah, A. M. Abdulhadi Ashleibta, A. Hasanali, S. Khattak, and Q. H. Abbasi, "Sensor fusion for identification of freezing of gait episodes using Wi-Fi and radar imaging," *IEEE Sensors J.*, vol. 20, no. 23, pp. 14410–14422, Dec. 2020.
- [34] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll, "The TUM gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits," *J. Vis. Commun. Image Represent.*, vol. 25, no. 1, pp. 195–206, Jan. 2014.
- [35] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, 2014.
- [36] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
- [39] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-V4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1.
- [40] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
- [41] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [42] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [43] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.
- [44] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600.
- [45] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 6, no. 2, pp. 107–116, 1998.
- [46] L. Shuguang, T. Yulong, and G. Fengrui, "Device and method for body posture detection," Chinese Patent 11 135 847 1A, Jul. 3, 2020.
- [47] X. Tu, "Intimate technology helps elderly people: Walk all over the place," (in Chinese), *Ind. Technol.*, vol. 341, pp. 32–35, Jun. 2020. [Online]. Available: [https://www.itri.org.tw/ListStyle.aspx?DisplayStyle=18\\_content&SiteID=1&MmmID=1036452026061075714&MGID=1072136747670106070](https://www.itri.org.tw/ListStyle.aspx?DisplayStyle=18_content&SiteID=1&MmmID=1036452026061075714&MGID=1072136747670106070)
- [48] K. Cheng, Y. Zhang, X. He, W. Chen, J. Cheng, and H. Lu, "Skeleton-based action recognition with shift graph convolutional network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 183–192.
- [49] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Skeleton-based action recognition with multi-stream adaptive graph convolutional networks," *IEEE Trans. Image Process.*, vol. 29, pp. 9532–9545, 2020.
- [50] Y. Du, W. Wang, and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1110–1118.
- [51] S. Zhang, X. Liu, and J. Xiao, "On geometric features for skeleton-based action recognition using multilayer LSTM networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 148–157.
- [52] S. Chakraborty, S. Jain, A. Nandy, and G. Venture, "Pathological gait detection based on multiple regression models using unobtrusive sensing technology," *J. Signal Process. Syst.*, vol. 93, no. 1, pp. 1–10, Jan. 2021.
- [53] Y. Liao, A. Vakanski, and M. Xian, "A deep learning framework for assessing physical rehabilitation exercises," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 468–477, Feb. 2020.
- [54] H. F. Maqbool, M. A. B. Husman, M. I. Awad, A. Abouhossein, N. Iqbal, and A. A. Dehghani-Sanij, "A real-time gait event detection for lower limb prosthesis control and evaluation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 9, pp. 1500–1509, Sep. 2017.

- [55] T. Hussain, H. F. Maqbool, N. Iqbal, M. Khan, and A. A. Dehghani-Sani, "Computational model for the recognition of lower limb movement using wearable gyroscope sensor," *Int. J. Sensor Netw.*, vol. 30, no. 1, pp. 35–45, 2019.
- [56] G. Panahandeh, N. Mohammadiha, A. Leijon, and P. Händel, "Continuous hidden Markov model for pedestrian activity classification and gait analysis," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 1073–1083, May 2013.
- [57] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
- [58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [59] K. Jun, S. Lee, D. Lee, and M. S. Kim, "Azure Kinect 3D skeleton and foot pressure data for pathological gaits," *IEEE Dataport*, to be published, doi: [10.21227/ev8a-wr16](https://doi.org/10.21227/ev8a-wr16).



**KOOKSUNG JUN** received the B.S. degree in mechanical engineering and the M.S. degree in intelligent robotics from the Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2018 and 2019, respectively, where he is currently pursuing the Ph.D. degree. His current research interests include healthcare robotics, artificial intelligence, and pattern recognition.



**SANGHYUB LEE** received the B.S. degree in biomedical engineering from the University of Ulsan, in 2017, and the M.S. degree in intelligent robotics from the Gwangju Institute of Science and Technology, in 2019, where he is currently pursuing the Ph.D. degree. His current research interests include image processing, healthcare robotics, and pattern recognition.



**DEOK-WON LEE** received the B.S. degree in electronics and electric wave and information engineering from Chungnam National University, in 2009, and the M.S. degree in electrical and electronics engineering from Yonsei University, in 2013. He is currently pursuing the Ph.D. degree with the Gwangju Institute of Science and Technology. His current research interests include healthcare robotics, artificial intelligence, and pattern recognition.



**MUN SANG KIM** (Member, IEEE) received the B.S. and M.S. degrees in mechanical engineering from Seoul National University, Seoul, South Korea, in 1980 and 1982, respectively, and the Dr.-Ing. degree in robotics from the Technical University of Berlin, Berlin, Germany, in 1987. From 1987 to 2016, he was a Research Scientist with the Korea Institute of Science and Technology, Seoul. He led the Advanced Robotics Research Center, in 2000, where he became the Director of the "Intelligent Robot—The Frontier 21 Program," in October 2003, which is one of the most challenging research programs in South Korea. He is currently a Professor with the School of Integrated Technology, Gwangju Institute of Science and Technology. His current research interests include healthcare robotics, UWB-based indoor localization systems, and culture technology.

...