

**Research Article****A deep learning approach for human gait recognition from time-frequency analysis images of inertial measurement unit signal****Hacer KUDUZ^{a,b,*}** , **Firat KAÇAR^c** ^aDivision of Biomedical Engineering, Institute of Graduate Studies, Istanbul University-Cerrahpasa, 34320, Istanbul, Türkiye^bDepartment of Biomedical Engineering, Faculty of Engineering, Akdeniz University, 07058, Antalya, Türkiye^cDepartment of Electrical and Electronics Engineering, Faculty of Engineering, Istanbul University-Cerrahpasa, 34320, Istanbul, Türkiye

ARTICLE INFO

Article history:

Received 23 May 2023

Accepted 29 July 2023

Keywords:

Biomechanical data

CNN

Human activity recognition

IMU

Time-frequency analysis

ABSTRACT

Biomechanical analysis using deep learning has been increasingly used in recent studies to identify human activity. Wearable sensor data from inertial measurement units (IMUs) is widely used for recognizing human activity, but has several drawbacks owing to its high volume and diversity. To overcome these issues, the time-domain and power spectral characteristics of IMU data can be extracted using digital signal processing (DSP) methods. Our research aimed to investigate time-frequency analysis (TFA) methods for classifying the spatio-temporal gait characteristics of physical walking performed by healthy subjects. In this study, open-source biomechanical sensor signal dataset was used. The DSP step was first carried out by segmenting IMU data from the four body segments of 22 healthy subjects, and then by applying Continuous Wavelet Transform (CWT) and Short Time Fourier Transform (STFT) methods. Moreover, the resultants of linear accelerometer signals were applied in a similar manner. The image datasets obtained from this step were applied to a deep convolutional neural network (CNN) model to classify human walking speed (WS) into three classes: fast, normal, and slow. The performance of the 2D-CNN model and the impact of DSP methods using IMU data were evaluated. In conclusion, the highest test classification results demonstrated that STFT-all (85.9%), CWT-all (79.3%), and CWT-resAcc (76.3%) based CNN models present a remarkably precise and easy-to-implement classification problem, with the highest test accuracy, when all IMU channels are subjected to STFT. The classification accuracies of 2D-CNN models were compared to commonly used ML models. The Deep CNN model is a useful gait evaluation tool for healthy subjects. Furthermore, it can enable the diagnosis and phase assessment of gait abnormalities and detect gait biomarkers in rehabilitative wearables.

This is an open access article under the CC BY-SA 4.0 license.
(<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Introduction

Adult and geriatric gait disorders have emerged as major problems worldwide in the recent years. Declining step length, mean acceleration, and velocity can all be indications of an abnormal gait. Self-selected walking speed (WS), commonly referred to as gait velocity, is a simple screening approach that can be used to obtain information about human functional status and well-being [1, 2]. WS can be used as a functional, useful, and helpful vital sign for people, and can be assessed in a similar way to how physiotherapists routinely check heart rate,

ventilation, and skin temperature.

WS data were collected from two main biomechanical sources to recognize the human gait. Some researchers have used wearable inertial measurement unit (IMU) sensors [3], whereas others have utilized vision-based systems (cameras, videos, and images) [4, 5]. Although IMU-based devices, such as smartphones and smart watches, can be easily integrated and worn on the body to analyze human movement, vision-based systems have major limitations, such as the requirement for vast amounts of information storage and difficulties in camera configuration.

* Corresponding author. E-mail address: hacerkuduz@akdeniz.edu.tr
DOI: 10.58190/ijamec.2023.44

Recently, wearable sensor signals have been utilized in various applications, including sports science, rehabilitation, elderly care, biometrics, and the design of movement-assistive equipment to automatically recognize human activities of daily living [6-10]. Accelerometers, gyroscopes, magnetometers, IMUs, electromyography (EMG), and IoT-based wearables have been extensively employed in human activity recognition (HAR) applications to sense complex movement tasks. Walking patterns and environmental factors that differ on a daily basis affect the walking speed. IMUs, including accelerometers and gyroscopes, are widely used in long-term gait analysis studies. Although a single, low-power accelerometer may reduce simultaneous contextual information, it may enhance compliance. Additionally, the resultant acceleration of multiple IMU sensors worn on the body has been utilized in human activity recognition studies [11-13]. Huynh and Tran proposed that apply the power spectral density (PSD) approach, one of the TFA methods, for the resultants of IMU linear acceleration data in their study on human fall detection. They revealed utilizing daily activity data that the power spectral density (PSD) based methodology can detect fall with high sensitivity of 98.4% [13].

The time-varying frequency properties of a variety of human activities, which are essential for activity recognition, are well-represented by time-frequency domain spectrogram and scalogram. Because of its ease of implementation, the STFT is commonly utilized in the literature [14-16] reviewed to calculate the spectrogram with respect to data from EMG and wearable sensors. Additionally, Wavelet Transform [17-19] is an effective method for creating a power spectral density map from human physical activities. Too et al. extracted features of EMG signals by applying Wavelet Transform for hand gesture recognition and then classified them with ML algorithms such as SVM and KNN [17]. In another related work, Ozdemir et al. has utilized CNN and transfer learning approaches to classify spectral images they obtained through applying time-frequency analysis (TFA) to the EMG data [16].

Data-computing techniques, such as deep learning and machine learning algorithms, have demonstrated promise when applied to IMU generated gait data for managing complicated patterns from data in human movement, especially as access to higher processing machines has risen [10, 16, 20]. CNN deep learning algorithms are frequently employed in human gait recognition. Raw signals and time-frequency spectrum images of IMU sensors are widely used as CNN inputs [21-24]. The methodology of Jung et al. study is based on the study that provides reliable multi-classification using deep convolutional neural networks and spectrographic approach using IMU data to classify pathological gait phases without discernible differences [23]. In another

study, Jung et al. proposed an effective approach for gait recognition of three different groups using 2D-CNN based wearable IMU sensors, trained with gait spectrograms obtained from time-frequency domain analysis of raw acceleration and angular velocity data. This classifier showed an accuracy of 93.02% in classifying athlete, normal foot and deformed foot groups [24]. The main contribution on human gait recognition of this study is the development of a 2D-CNN model inspired by [23] for the use of IMU signal TFA images. Another contribution to this issue is to investigate the usability of TFA images obtained from IMU linear acceleration resultant, inspired by [16], in HAR studies by applying them to the developed 2D-CNN model.

In this study, our aims are (1) to classify the time domain (TD) and time-frequency domain (TFD) images of IMU sensor signals with 2D-CNN, (2) to classify the time domain and time-frequency domain images of the resultant linear accelerations of these signals with 2D-CNN, (3) evaluate the contribution of learning models to human gait recognition performance and (4) compare the accuracy of traditional machine learning models with the proposed 2D-CNN models. The result of our study indicates that the most accurate results may be obtained with a chosen image subset from STFT time-frequency domain representations when they are fed into a deep CNN classifier with IMU signals from all channels. The methodology used in this study can serve as a template for future important strategies for the comprehensive identification and phase assessment of gait disorders, in addition to identifying gait indicators for supportive and therapeutic wearable technologies. The findings of this study may also contribute to forecasting the progress of gait disorders by applying deep learning methods from the scientific field to the field of biomedicine.

2. Methods

In this study, we utilized the Camargo et al. open-source dataset [25], which consisted of lower-limb 3D biomechanical and wearable data. Time-frequency analysis methods were then used to classify the IMU sensor data. Finally, the 2D-CNN model utilizes IMU data to classify the obtained human gait features and predict the walking speed.

In this dataset, the human lower limbs are handled in level-ground locomotion mode at three distinct walking speed states employing data from IMU sensors. The dataset consists of 22 healthy subjects (age 21 ± 3.4 years, 3 females, and 19 males, mass 68.3 ± 10.83 kg, height 1.70 ± 0.07 m) participating in the study. Each subject performed a total of 30 level-ground walking trials for each of the three self-selected speeds—fast (F), normal (N), and slow (S)—in relation to each subject speed. The linear velocity of the pelvis was used as a reference to

calculate the average steady-state walking speed of each participant. According to the average pelvis velocity of all subjects, slow, normal, and walking speeds were 1.17 ± 0.21 m/s, 0.88 ± 0.19 m/s, and 1.45 ± 0.27 m/s, respectively. Four 6-axis IMUs were placed unilaterally on the right side of the subjects. The IMU sensor three-axis gyroscope and accelerometer were used to assess the rotational displacement and linear acceleration, respectively.

2.1. Biomedical Signal Processing

The IMU data were collected from the foot (Ft), shank (Sh), thigh (Th), and trunk (Tr) of the human lower limb at a sampling frequency of 200 Hz. Figure 1 shows the 24-channel IMU data consisting of the accelerometer (Acc_x , Acc_y , and Acc_z) and gyroscope ($Gyro_x$, $Gyro_y$, and $Gyro_z$) components.

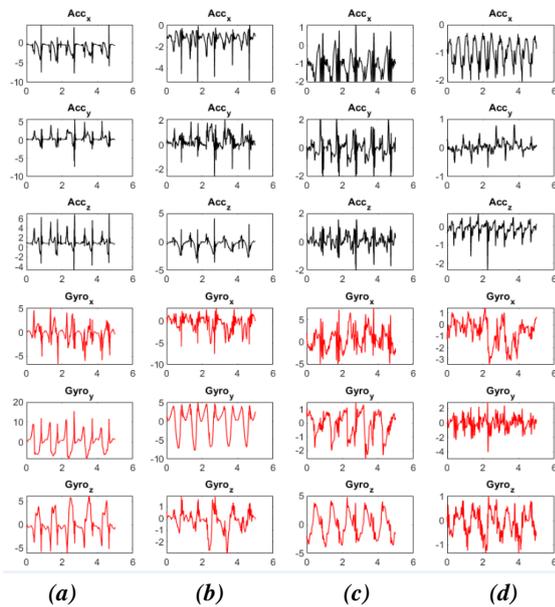


Figure 1. IMU time series signals of 5 sn, (a) Foot sensor, (b) Shank sensor, (c) Thigh sensor, (d) Trunk sensor.

A rectangular walking path was set up to collect the IMU signals. The signal is automatically split so that each signal's midpoint is equally spaced from its starting point and ending point to avoid the impact of interruptions and discontinuities at the turning points. These 24x500 signals, extracted from the raw IMU data, were converted into a data set. Additionally, a 1x12000 vector was generated by sequentially adding each channel, and the CWT and STFT transformations were used to obtain the signal power spectrum in the frequency-time domain. The CNN input datasets were created from the spectral images in Figure 2.

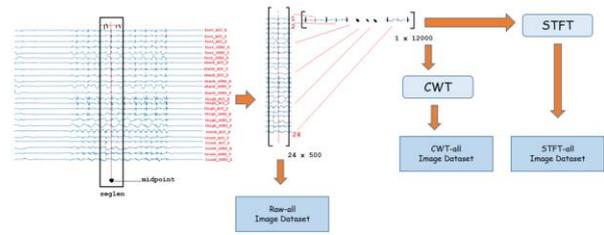


Figure 2. Raw IMU signal segmentation and image datasets.

In addition to the linear acceleration and angular velocity parameters, the resultant linear acceleration (Res-Acc), calculated by taking the square root of the linear accelerations along the x-, y-, and z-axes, is also utilized when machine learning and deep learning-based human motion pattern recognition studies conducted using IMU sensors are reviewed in the literature. As shown in Figure 3, the resulting linear accelerations of the foot, thighs, thighs, and trunk were recorded in the segmented signal, and an image dataset was created from the obtained 4×500 matrices. Furthermore, a 1×2000 vector was created using the end-to-end connection of the elements in each channel. Two separate datasets were generated by applying CWT and STFT transformations to obtain the power spectra of the resultant linear accelerations in the TFD.

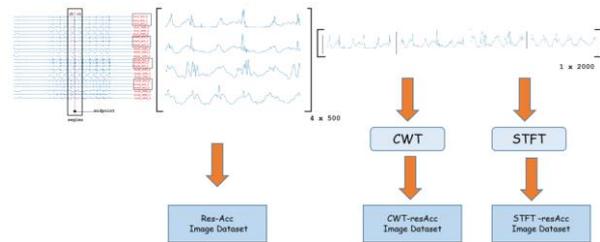


Figure 3. Res-Acc IMU signal segmentation and image datasets.

Figure 4 shows sample images of the datasets created after automatic segmentation and walking speed classes. In each dataset, there were 230, 221, and 229 images for the fast, normal, and slow walking-speed classes, respectively. The down-sampling method was applied in order to overcome unbalanced dataset problem. There were 220 samples in each class after down-sampling, for all classifier models included in the study.

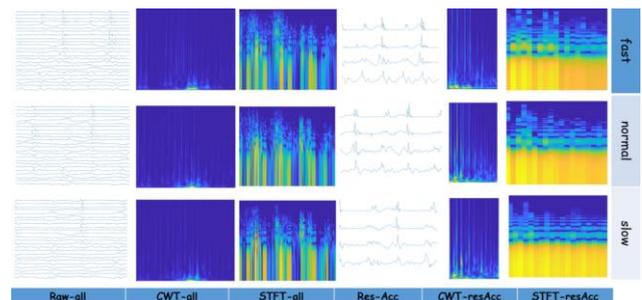


Figure 4. Sample images from CNN input datasets.

2.2. Deep Convolutional Neural Networks

The training and test datasets for the CNN model were randomly divided into two groups: 80% and 20% of the input dataset, respectively. The essential model parameters for the training phase of the CNN model had the greatest impact on the training and test accuracy. The model parameters were optimizer ‘adam,’ initial learning rate 10e-3, maximum epochs 10, and mini-batch size 64. The CNN model was developed using the parameters, and the highest test accuracy was achieved by experimenting with various parameter combinations.

Using the MATLAB (The MathWorks, R2021b) Deep Network Designer Toolbox, the CNN architecture is shown in Figure 5. The deep CNN architecture consisted of an input layer (Ly), three feature extraction layers, and an output layer. There were two convolution (Conv) layers, a linear rectification unit (ReLU) layer, and a batch normalization (Btch-Norm) layer for each feature extraction layer. A feature map is created by processing the images in the input layer, which consists of pixel values, with filter matrices comprising a series of weights in the convolution layer. The batch normalization layer forms a more structured model of neurons and improves performance during training, whereas the ReLU layer determines the activation function of the deep neural network.

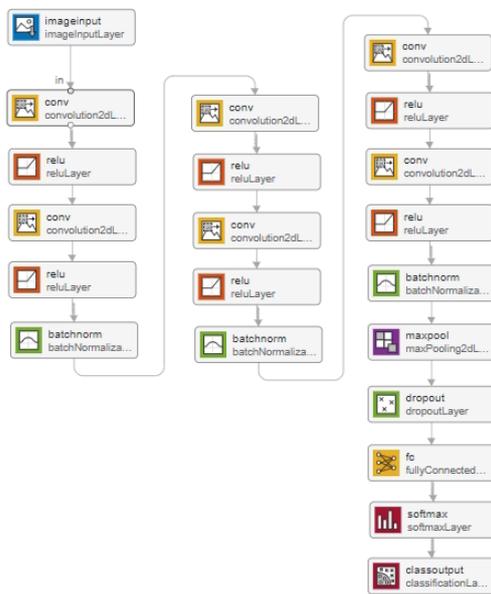


Figure 5. Deep-CNN model architecture.

The max pooling (Max-pool) layer decreases the weights in the deep CNN architecture and analyzes their applicability. Subsequently, to avoid over fitting during training, a dropout layer was implemented to eliminate excessive neurons. A fully connected layer was implemented to generate an input matrix resembling the inputs in a traditional neural network. The output layer, which is a classification layer, and the softmax layer, which has an activation function for categorical

classification, are the last two layers. The layers and weights of the CNN architecture are listed in detail in Table 1.

Table 1. Proposed Deep CNN model layers and weights.

Ly	Type	Subtype	Configuration	Activation
1	Input layer	Image input	64 x 64 x 3 images with ‘zero-center’ normalization	64 x 64 x 3
2	Feature extraction layer-1	Conv	16 3 x 3 convolutions with stride [1 1] and padding ‘same’	64 x 64 x 16
3		ReLU	ReLU	64 x 64 x 16
4		Conv	16 3 x 3 convolutions with stride [1 1] and padding [1 1 1]	64 x 64 x 16
5		ReLU	ReLU	64 x 64 x 16
6		Btch-Norm	Batch normalization with 16 channels	64 x 64 x 16
7	Feature extraction layer -2	Conv	32 3 x 3 convolutions with stride [1 1] and padding ‘same’	64 x 64 x 32
8		ReLU	ReLU	64 x 64 x 32
9		Conv	32 3 x 3 convolutions with stride [1 1] and padding [1 1 1]	64 x 64 x 32
10		ReLU	ReLU	64 x 64 x 32
11		Btch-Norm	Batch normalization with 32 channels	64 x 64 x 32
12	Feature extraction layer -3	Conv	32 3 x 3 convolutions with stride [1 1] and padding ‘same’	64 x 64 x 32
13		ReLU	ReLU	64 x 64 x 32
14		Conv	32 3 x 3 convolutions with stride [1 1] and padding [1 1 1]	64 x 64 x 32
15		ReLU	ReLU	64 x 64 x 32
16		Btch-Norm	Batch normalization with 32 channels	64 x 64 x 32
17		Max-pool	2 x 2 max pooling with stride [1 1] padding ‘same’	64 x 64 x 32
18	Classification layer	Dropout	50% dropout	64 x 64 x 32
19		Fully connected	3 fully connected layer	1 x 1 x 3
20		Softmax	softmax	1 x 1 x 3
21	Output layer	Classification	‘fast’, ‘normal’ and ‘slow’	1 x 1 x 3

2.3. Machine Learning Algorithms

The time series and time-frequency domain features of the IMU signal are computed and used as input to the machine learning algorithms. The signal time series features employed in this case are the mean, maximum,

standard deviation, minimum, and area under the curve (AUC), whereas the TFD features are the maximum, skewness, kurtosis, and AUC that calculate the Fast Fourier Transform (FFT) of the signal. Three machine learning approaches that are often utilized to recognize human activity were applied using the MATLAB (The MathWorks, R2021b) Classification Learner Toolbox: Quadratic Support Vector Machine (QSVM), Wide Neural Networks (WNN), and Ensemble Subspace K-nearest Neighbours (ESKNN). All ML algorithms performed repeated 10-fold cross validation in order to prevent overfitting. Additionally, feature input dataset was randomly separated 80%-20% training and test split and analysed using these ML algorithms.

2.4. Classification Performance Metrics

Performance metrics, such as the confusion matrix (CM) and accuracy, were used to determine the performance of the analysed deep CNN classifiers on the training and test sets. True Negative (TN) refers to a sample that was correctly classified as negative (N), whereas True Positive (TP) refers to a sample that was correctly classified as positive (P). False-negative (FN) samples are those that are correctly classified as positive but are actually in the negative class, as opposed to false-positive (FP) samples, which are those that are correctly classified as positive but are actually in the negative class. The binary class confusion matrix is presented in Table 2(a), and the following quantitative performance metrics can be derived:

- *Accuracy*: Percentage of samples correctly classified among all samples in the test set.
- *Precision* (for the positive class): The number of samples that the model predicts will be in the positive class, which is included in that group.
- *Specificity*: The number of samples in the dataset where the negative class was correctly estimated to include all samples.
- *Recall* (for the positive class): A measure of the number of samples that actually belonged to the positive class was correctly predicted.
- *F1-Score* (for positive class): Harmonic mean of recall and precision scores for the positive class.

Table 2. CM and performance metrics of binary classification (a), multiple classification (b).

		True class		Accuracy $\frac{TP + TN}{TP + FP + TN + FN}$
		P	N	
Predicted class	P	TP	FP	Specificity $\frac{TN + FN}{TN + FP}$
	N	FN	TN	Precision $\frac{TP}{TP + FP}$
				Recall $\frac{TP}{TP + FN}$
				F-score $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

(a)

		True class			Class	TP	TN	FP	FN
		F	N	S					
Predicted class	F	a	b	c	F	a	e+f+h+j	b+c	d+g
	N	d	e	f	N	e	a+c+g+j	d+f	b+h
	S	g	h	j	S	j	a+b+d+e	g+h	c+f

(b)

A transformation is applied while calculating the multiple classification performance metrics, as indicated in Table 2(b). For the class-based evaluation, the TP, TN, FP, and FN values were computed. Multiple classifications are implemented using the performance metrics listed in Table 2(a).

3. Results

In this study, different image sets obtained from the proposed 2D-CNN classifier and TD and TFD analysis methods were compared in terms of classification performance metrics. Table 3 lists the training (a) and test (b) confusion matrices of the model with all the IMU signals, and the training (c) and test (d) confusion matrices of the model with the IMU linear acceleration resultant. The number of samples for each walking speed estimated by the 2D-CNN models is shown in the confusion matrices in Table 3. Actual walking speed labels for examples are also shown in the columns. The blue cells in each row of the table showed the TP value for each WS class. The FN value for a given WS class was the sum of the values in all cells in a row except the blue cell.

Table 3. CMs of the 2D-CNN model with all the IMU signals (a) training set, (b) test set. CMs of the 2D-CNN model with the IMU linear acceleration resultant (c) training set, (d) test set.

		True WS			True WS			True WS				
		F	N	S	F	N	S	F	N	S		
Predicted WS	F	157	5	14	F	158	11	7	F	170	6	0
	N	18	97	61	N	9	140	27	N	0	175	1
	S	3	4	169	S	1	1	174	S	1	13	162

(a)

		True WS			True WS			True WS				
		F	N	S	F	N	S	F	N	S		
Predicted WS	F	29	7	8	F	33	9	2	F	35	8	1
	N	12	10	22	N	6	19	19	N	5	34	5
	S	1	6	37	S	0	5	39	S	1	8	35

(b)

		True WS			True WS			True WS				
		F	N	S	F	N	S	F	N	S		
Predicted WS	F	163	4	9	F	132	31	13	F	174	0	2
	N	25	117	34	N	1	162	13	N	2	166	8
	S	3	3	170	S	2	9	165	S	0	0	176

(c)

		True WS			True WS			True WS				
		F	N	S	F	N	S	F	N	S		
Predicted WS	F	30	9	5	F	24	17	3	F	32	8	4
	N	14	9	21	N	5	30	9	N	15	18	11
	S	2	4	38	S	2	11	31	S	1	11	32

(d)

Table 3 (a-b) demonstrates that the STFT-all model is the highest in the case of including all IMU channels when TP values are evaluated on a class basis, followed by the CWT-all model, whereas Raw-all model has the lowest TP value. Table 3 (c-d) lists that the CWT-resAcc model is the highest for the resultant accelerometer channels of the wearable IMU sensors when TP values are evaluated on a class basis, followed by the STFT-resAcc model, whereas Res-Acc model has the lowest TP value.

Table 4. Classification performance metrics of 2D-CNN models.

	Accuracy	Specificity	Precision	Recall	F-score
fast	0.788	0.852	0.690	0.659	0.674
normal	0.644	0.852	0.435	0.227	0.299
slow	0.720	0.659	0.552	0.841	0.667
average	0.717	0.788	0.559	0.576	0.547
fast	0.871	0.932	0.846	0.750	0.795
normal	0.705	0.841	0.576	0.432	0.494
slow	0.803	0.761	0.650	0.886	0.750
average	0.793	0.845	0.691	0.689	0.680
fast	0.886	0.932	0.854	0.795	0.824
normal	0.803	0.818	0.680	0.773	0.723
slow	0.886	0.932	0.854	0.795	0.824
average	0.859	0.894	0.796	0.788	0.790
fast	0.773	0.818	0.652	0.682	0.667
normal	0.636	0.852	0.409	0.205	0.273
slow	0.758	0.705	0.594	0.864	0.704
average	0.722	0.792	0.552	0.583	0.548
fast	0.795	0.920	0.774	0.545	0.640
normal	0.682	0.682	0.517	0.682	0.588
slow	0.811	0.864	0.721	0.705	0.713
average	0.763	0.822	0.671	0.644	0.647
fast	0.788	0.818	0.667	0.727	0.696
normal	0.659	0.784	0.486	0.409	0.444
slow	0.795	0.830	0.681	0.727	0.703
average	0.747	0.811	0.611	0.621	0.614

The test sets of 2D-CNN models multiple classification performance metrics are listed in Table 4. On a class basis, the "normal class" has the lowest performance, and the "fast class" has the highest performance. The performance metrics of the IMU are listed as STFT-all, CWT-all, and Raw-all, from high to low when the macro average values are compared. The multiple classification performance metrics for the resultant linear accelerometer of the wearable IMU sensors, on a class basis, the "normal class" has the lowest performance, and the "fast class" has the highest performance. The performance metrics of these 2D-CNN models are listed as CWT-resAcc, STFT-resAcc, and Res-Acc, from high to low, when the macro-average values are compared.

Table 5. CMs of the ML model with all IMU signals (a) 10-fold cross-validation (CV), (b) 80%-20% random train-test split (RS).

		True WS			True WS			True WS		
		F	N	S	F	N	S	F	N	S
Predicted WS	F	174	45	1	164	50	6	180	37	3
	N	35	162	23	54	118	48	40	153	27
	S	2	44	174	7	42	171	2	31	187
		SVM-CV			KNN-CV			ANN-CV		
		(a)			(b)					
		True WS			True WS			True WS		
		F	N	S	F	N	S	F	N	S
Predicted WS	F	35	9	0	29	14	1	34	9	1
	N	8	31	5	11	22	11	6	34	4
	S	0	10	34	1	10	33	2	6	36
		SVM-RS			KNN-RS			ANN-RS		

Table 5 lists the classification confusion matrices of all IMU signals using conventional ML models. 10-fold cross validation (CV) and 80%-20% random train-test split (RS) methods were used in order to determine the training set in the ML model. When the confusion matrices to which the 10-fold CV is applied as listed in Table 5(a) are examined, it is seen that the highest TP value is the ANN-CV model and the lowest is the KNN-CV model. The ANN-RS model had the highest TP value, while the KNN-RS model had the lowest, depending on the evaluation of the confusion matrices in Table 5(b), when the 80%-20% random train-test split approach is used. Likewise, when the classification accuracies of the ML models listed in Table 6 are compared, the ANN model, in which the 10-fold cross validation approach was applied, reached the highest accuracy with 0.788, while the KNN model, in which the random split was applied, reached the lowest accuracy with 0.636.

Table 6. Classification performance metrics of ML models.

	Accuracy	AUC-ROC
CV	SVM	0.773
	KNN	0.686
	ANN	0.788
RS	SVM	0.758
	KNN	0.636
	ANN	0.788

4. Discussion

This study proposed a new time-frequency analysis approach that can multi-classify the ground level gait of healthy individuals according to three different walking speeds. Without traditional kinematic gait assessment metrics such as spatial and temporal gait parameters, both raw and spectral images of the resultant of a 5 s duration gait signal and its linear accelerations were used as input data for efficient classification. Time-frequency analyses have been widely applied to ECG [26], radar [14], and speech signals [27] represented in the time-frequency domain have been trained with DCNNs, producing important results especially for speech recognition. However, its applications in gait classification are quite

limited. The study of Jung et al. is the first to use spectral analysis for biomechanical assessment of gait with the 6-DOF IMU and validate its utility for reliable classification of pathological gait with no observable difference [23]. It is noteworthy that their proposed 2D-CNN model, which is trained using spectrograms of IMU signals, provides 100% accuracy classification without requiring much manual labour, time-consuming and resource-intensive processes and specialized knowledge. In addition, their proposed 2D-CNN model has reduced training and testing time, since it is simpler than other existing deep learning models.

The classification performance metrics of each model were addressed independently to compare the time-frequency analysis techniques. First, a dataset was created using time-amplitude images of the signals, which included all the IMU sensor channels. In addition, by utilizing the STFT and CWT on these signals, datasets were generated from the spectral images. The dataset obtained from the images of the time-amplitude signals created by the resultant linear accelerations was then subjected to the same procedure. Then, by applying transformations via STFT and CWT to the resultant linear accelerations, datasets were produced from the spectral images. Besides, the features [10, 28, 29] that widely used HAR studies extracted from IMU time series signals were classified with ML models and compared with deep learning models.

The training and test set classification performance of each model is shown in Figure 6. With regard to these performance metrics, it can be seen that the models that apply STFT to models that include all IMU channels and those that apply CWT to models that use linear components perform well. For the training and test sets, the STFT-all model with the best classification accuracy was found to be 97.3% and 85.9%, respectively. The results for the other two models with high classification accuracy, CWT-all and CWT-resAcc, were 79.3% and 76.3%, respectively, for the test data.

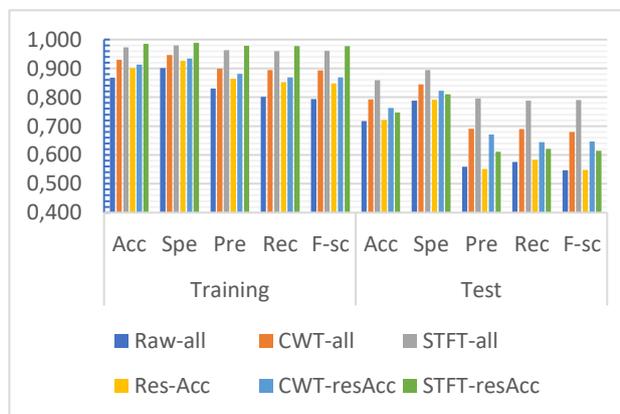


Figure 6. Comparison of the deep 2D-CNN model classification performance.

The classification performance of 2D-CNN and ML models was evaluated by using the confusion matrices listed in Tables 3 and 5. The STFT-all model, as shown, provides fewer false positive and false negative scores for both the training and test data. Tables 4 and 6 provide a list of each model's accuracy, precision, specificity, recall, and F1 scores for training and test data, which were calculated using the respective confusion matrices. Furthermore, the STFT-all model outperformed the CWT-all and CWT-resAcc models in terms of accuracy and specificity.

With the same dataset we used in the study, Camargo et al. classified according to walking speed with classical machine learning algorithms. In estimation of walking speed, they found the lowest RMSE error rate in classification of fast, slow and normal groups as 0.04 ± 0.01 m/s in ML-based SVM models. In addition, they showed in their study which sensor type and placement are important for different walking area conditions, mechanical sensors such as IMU and GON are more decisive for classification than EMG sensors, and shank IMU sensors have higher accuracy in estimating walking speed [28]. Similar to the gait parameters in this study, Bhakta et al. evaluated the effectiveness of multiple machine learning algorithms in determining the walking speed of subjects with transfemoral amputation in their study, which aimed to determine the walking speed on a robotic prosthesis in real time. They confirmed low error performance with 0.014 ± 0.001 m/s RMSE, especially in determining walking speed for dynamic situations [30].

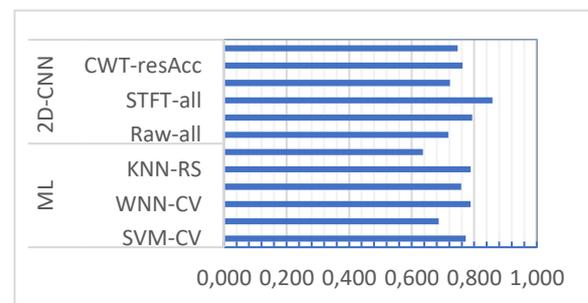


Figure 7. Comparison of 2D-CNN and ML models classification accuracies.

These outcomes highlight the possible applicability of walking speed as biomechanical identifiers, this functionality could be broadened via using the STFT-CNN model recommended in this study. Gait recognition approaches comprised the use of spectrogram, time series signal, and scalogram. Figure 7 presents that when using spectrogram images from the TFA of all IMU channels and CNN to classify gaits, the accuracy was 85.9% however, when using machine learning to classify gaits, and the accuracy was 78.8%. This outcome can be attributed to the efficiency of the gait spectral image in accurately handling both the temporal and spatial gait parameters, as well as the feasibility of STFT-CNN based classification.

5. Conclusions

In conclusion, our deep CNN model for recognizing human gait offers different insights into the changes in temporal gait trends in healthy adults by executing three distinct walking speed experiments. This study contributes to the development of biomedical signal processing methods that can be used to recognize human lower limb gait using a variety of biomechanical parameters of wearable sensors through deep learning. We found that the accuracy of deep learning classification was barely affected by the time-frequency domain transformation when applied to the STFT and CWT models, including all IMU channels when compared to findings from other image datasets.

The method of study could be the basis for promising advances advancements in comprehensive gait abnormalities diagnosis and stage assessment, as well as the detection of gait biomarkers for wearable assistive and rehabilitation systems. In further research, the same lower limb biomechanics dataset would be used to develop deep learning models such as transfer learning and LSTM to recognize human activity including ramp and stair movement. To develop deep learning models, it is also intended to combine with the goniometer joint angles dataset and the IMU dataset.

Acknowledgements

We would like to thank Camargo J, et al. to provide open-source dataset of lower limb biomechanics in multiple conditions.

Authorship Contributions

HK, the deep learning analysis and model, biomedical signal processing, interpretation of the data, original draft preparation and revision of the manuscript. FK, supervision, the development of the deep learning model, and revising of the manuscript. All authors read and agreed to the published version of the manuscript.

Data Availability Statement

Raw data that used in this study are available from this article [25].

Conflict of 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.

Ethical Use of Images

There are no ethical issues with the publication of this manuscript.

Ethical Declarations

No ethical declarations have been needed because our study did not have human participation or involved experiments with animals.

References

- [1] I. Hagoort, N. Vuillerme, T. Hortobágyi, and C. J. Lamoth, "Outcome-dependent effects of walking speed and age on quantitative and qualitative gait measures," *Gait & Posture*, vol. 93, pp. 39-46, 2022. DOI: <https://doi.org/10.1016/j.gaitpost.2022.01.001>.
- [2] T. Scone, M. Saadat, H. Barton, and A. Rastegarpanah, "Effects of Variations in Hemiparetic Gait Patterns on Improvements in Walking Speed," *IRBM*, p. 100733, 2022. DOI: <https://doi.org/10.1016/j.irbm.2022.08.001>.
- [3] M. De Vos, J. Prince, T. Buchanan, J. J. FitzGerald, and C. A. Antoniadou, "Discriminating progressive supranuclear palsy from Parkinson's disease using wearable technology and machine learning," *Gait & Posture*, vol. 77, pp. 257-263, 2020. DOI: <https://doi.org/10.1016/j.gaitpost.2020.02.007>.
- [4] M. N. Favorskaya and V. V. Buryachenko, "Monocular depth-based visual tracker for gait recognition," *Procedia Computer Science*, vol. 207, pp. 205-214, 2022. DOI: <https://doi.org/10.1016/j.procs.2022.09.053>.
- [5] S. Gul, M. I. Malik, G. M. Khan, and F. Shafait, "Multi-view gait recognition system using spatio-temporal features and deep learning," *Expert Systems with Applications*, vol. 179, p. 115057, 2021. DOI: <https://doi.org/10.1016/j.procs.2022.09.053>.
- [6] B. Hu, S. Li, Y. Chen, R. Kavi, and S. Coppola, "Applying deep neural networks and inertial measurement unit in recognizing irregular walking differences in the real world," *Applied Ergonomics*, vol. 96, p. 103414, 2021. DOI: <https://doi.org/10.1016/j.apergo.2021.103414>.
- [7] C. F. Martindale, V. Christlein, P. Klumpp, and B. M. Eskofier, "Wearables-based multi-task gait and activity segmentation using recurrent neural networks," *Neurocomputing*, vol. 432, pp. 250-261, 2021. DOI: <https://doi.org/10.1016/j.neucom.2020.08.079>.
- [8] R. Soangra, "Multi-day longitudinal assessment of physical activity and sleep behavior among healthy young and older adults using wearable sensors," *Irbm*, vol. 41, no. 2, pp. 80-87, 2020. DOI: <https://doi.org/10.1016/j.irbm.2019.10.002>.
- [9] W. Teufel, M. Lorenz, M. Miezal, B. Taetz, M. Fröhlich, and G. Bleser, "Towards inertial sensor based mobile gait analysis: Event-detection and spatio-temporal parameters," *Sensors*, vol. 19, no. 1, p. 38, 2018. DOI: <https://doi.org/10.3390/s19010038>.
- [10] A. M. Nasrabadi, A. R. Eslaminia, P. R. Bakhshayesh, M. Ejtehadi, L. Alibiglou, and S. Behzadipour, "A new scheme for the development of IMU-based activity recognition systems for telerehabilitation," *Medical Engineering & Physics*, vol. 108, p. 103876, 2022. DOI: <https://doi.org/10.1016/j.medengphy.2022.103876>.
- [11] T. Wang et al., "Accelerometer-based human fall detection using sparrow search algorithm and back propagation neural network," *Measurement*, vol. 204, p. 112104, 2022. DOI: <https://doi.org/10.1016/j.measurement.2022.112104>.
- [12] M. C. Ruder, M. A. Hunt, J. M. Charlton, T. Calvin, and D. Kobsar, "Validity and reliability of gait metrics derived from researcher-placed and self-placed wearable inertial sensors," *Journal of Biomechanics*, vol. 142, p. 111263, 2022. DOI: <https://doi.org/10.1016/j.jbiomech.2022.111263>.
- [13] Q. T. Huynh and B. Q. Tran, "Time-Frequency Analysis of Daily Activities for Fall Detection," *Signals*, vol. 2, no. 1, pp. 1-12, 2021. DOI: <https://doi.org/10.3390/signals2010001>.
- [14] Y. Jia, Y. Guo, G. Wang, R. Song, G. Cui, and X. Zhong, "Multi-frequency and multi-domain human activity recognition based on SFCW radar using deep learning," *Neurocomputing*, vol. 444, pp. 274-287, 2021. DOI: <https://doi.org/10.1016/j.neucom.2020.07.136>.
- [15] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "A fully embedded adaptive real-time hand gesture classifier leveraging HD-sEMG and deep learning," *IEEE*

- transactions on biomedical circuits and systems, vol. 14, no. 2, pp. 232-243, 2019. DOI: <https://doi.org/10.1109/tbcas.2019.2955641>.
- [16] M. A. Ozdemir, D. H. Kisa, O. Guren, and A. Akan, "Hand gesture classification using time–frequency images and transfer learning based on CNN," *Biomedical Signal Processing and Control*, vol. 77, p. 103787, 2022. DOI: <https://doi.org/10.1016/j.bspc.2022.103787>
- [17] J. Too, A. R. Abdullah, and N. M. Saad, "Classification of hand movements based on discrete wavelet transform and enhanced feature extraction," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019. DOI: <https://doi.org/10.14569/ijacsa.2019.0100612>.
- [18] N. Nahid, A. Rahman, and M. A. Ahad, "Deep learning based surface EMG hand gesture classification for low-cost myoelectric prosthetic hand," in *2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2020: IEEE, pp. 1-8. DOI: <https://doi.org/10.1109/icievicivpr48672.2020.9306613>.
- [19] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand gesture recognition using compact CNN via surface electromyography signals," *Sensors*, vol. 20, no. 3, p. 672, 2020. DOI: <https://doi.org/10.3390/s20030672>.
- [20] N. U. Ahamed et al., "Using machine learning and wearable inertial sensor data for the classification of fractal gait patterns in women and men during load carriage," *Procedia Computer Science*, vol. 185, pp. 282-291, 2021. DOI: <https://doi.org/10.1016/j.procs.2021.05.030>.
- [21] C. Benegui and R. T. Ionescu, "Convolutional neural networks for user identification based on motion sensors represented as images," *IEEE Access*, vol. 8, pp. 61255-61266, 2020. DOI: <https://doi.org/10.1109/access.2020.2984214>.
- [22] C.-T. Yen, J.-X. Liao, and Y.-K. Huang, "Human daily activity recognition performed using wearable inertial sensors combined with deep learning algorithms," *Ieee Access*, vol. 8, pp. 174105-174114, 2020. DOI: <https://doi.org/10.1109/access.2020.3025938>.
- [23] D. Jung, M. D. Nguyen, M. Park, J. Kim, and K.-R. Mun, "Multiple classification of gait using time-frequency representations and deep convolutional neural networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 4, pp. 997-1005, 2020. DOI: <https://doi.org/10.1109/tnsre.2020.2977049>
- [24] D. Jung et al., "Deep neural network-based gait classification using wearable inertial sensor data," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019: IEEE, pp. 3624-3628. DOI: <https://doi.org/10.1109/embc.2019.8857872>
- [25] J. Camargo, A. Ramanathan, W. Flanagan, and A. Young, "A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions," *Journal of Biomechanics*, vol. 119, p. 110320, 2021. DOI: <https://doi.org/10.1016/j.jbiomech.2021.110320>.
- [26] J. Huang, B. Chen, B. Yao, and W. He, "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," *IEEE access*, vol. 7, pp. 92871-92880, 2019. DOI: <https://doi.org/10.1109/access.2019.2928017>
- [27] M. Kaushik, N. Baghel, R. Burget, C. M. Travieso, and M. K. Dutta, "SLINet: Dysphasia detection in children using deep neural network," *Biomedical Signal Processing and Control*, vol. 68, p. 102798, 2021.
- [28] J. Camargo, W. Flanagan, N. Csomay-Shanklin, B. Kanwar, and A. Young, "A machine learning strategy for locomotion classification and parameter estimation using fusion of wearable sensors," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1569-1578, 2021. DOI: <https://doi.org/10.1109/tbme.2021.3065809>
- [29] I. Boukhennoufa, X. Zhai, V. Utti, J. Jackson, and K. D. McDonald-Maier, "Wearable sensors and machine learning in post-stroke rehabilitation assessment: A systematic review," *Biomedical Signal Processing and Control*, vol. 71, p. 103197, 2022. DOI: <https://doi.org/10.1016/j.bspc.2021.103197>
- [30] K. Bhakta, J. Camargo, W. Compton, K. Herrin, and A. Young, "Evaluation of continuous walking speed determination algorithms and embedded sensors for a powered knee & ankle prosthesis," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4820-4826, 2021. DOI: <https://doi.org/10.1109/lra.2021.3068711>