



Using a Stacked Autoencoder for Mobility and Fall Risk Assessment via Time–Frequency Representations of the Timed Up and Go Test

Shih-Hai Chen¹, Chia-Hsuan Lee^{2*}, Bernard C. Jiang² and Tien-Lung Sun¹

¹ Department of Industrial Engineering and Management, Yuan Ze University, Taoyuan, Taiwan, ² Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan

OPEN ACCESS

Edited by:

Robert Hristovski,
Saints Cyril and Methodius University
of Skopje, North Macedonia

Reviewed by:

Sergi Garcia-Retortillo,
Boston University, United States
Monika Petelczyc,
Warsaw University of Technology,
Poland

*Correspondence:

Chia-Hsuan Lee
sweat0430@mail.ntust.edu.tw

Specialty section:

This article was submitted to
Fractal Physiology,
a section of the journal
Frontiers in Physiology

Received: 17 February 2021

Accepted: 28 April 2021

Published: 28 May 2021

Citation:

Chen S-H, Lee C-H, Jiang BC
and Sun T-L (2021) Using a Stacked
Autoencoder for Mobility and Fall Risk
Assessment via Time–Frequency
Representations of the Timed Up
and Go Test.
Front. Physiol. 12:668350.
doi: 10.3389/fphys.2021.668350

Fall risk assessment is very important for the graying societies of developed countries. A major contributor to the fall risk of the elderly is mobility impairment. Timely detection of the fall risk can facilitate early intervention to avoid preventable falls. However, continuous fall risk monitoring requires extensive healthcare and clinical resources. Our objective is to develop a method suitable for remote and long-term health monitoring of the elderly for mobility impairment and fall risk without the need for an expert. We employed time–frequency analysis (TFA) and a stacked autoencoder (SAE), which is a deep neural network (DNN)-based learning algorithm, to assess the mobility and fall risk of the elderly according to the criteria of the timed up and go test (TUG). The time series signal of the triaxial accelerometer can be transformed by TFA to obtain richer image information. On the basis of the TUG criteria, the semi-supervised SAE model was able to achieve high predictive accuracies of 89.1, 93.4, and 94.1% for the vertical, mediolateral and anteroposterior axes, respectively. We believe that deep learning can be used to analyze triaxial acceleration data, and our work demonstrates its applicability to assessing the mobility and fall risk of the elderly.

Keywords: SAE, TFA, DNNs, wavelet transform, LDA

INTRODUCTION

Remote health monitoring has been gaining increased interest as a way to improve the quality and reduce the costs of healthcare, especially for the elderly (Seyfioğlu et al., 2017). According to the World Health Organization, a person aged 65 years and over has a fall risk of 28–35%, which increases to 32–42% for those aged over 70 years [World Health Organization [WHO], 2007]. According Letts et al. (2010), 33% of community-dwelling elderly have experienced a fall event, and 50% fall repeatedly. About one-third of elderly people fall every year, and the chance of falling increases with age [World Health Organization [WHO], 2007; Bergland, 2012]. Falling can have serious long-term consequences for the elderly, including hospitalization, decreased mobility, fear of falling and even death. Older people with gait, mobility or balance problems are at higher risk of falling in the future (Ganz et al., 2007; Cuevas-Trisan, 2017). To develop an effective fall prevention program, elderly people with a fall risk must first be identified.

Various factors drive the fall risk. Mitchell et al. (2012) showed sarcopenia, the typical age-related decline in skeletal muscle mass cause strength reduction as well as balance issue. Poor balance and

mobility have been validated as a key cause of falls among the elderly. Continuous monitoring could be a practical approach to reduce and prevent falls by providing early warnings to facilitate appropriate interventions (Shany et al., 2012). However, continuous monitoring of gait and postural stability requires extensive healthcare and clinical resources. Limited professional resources (e.g., physical therapists, nurses, and doctors) are insufficient for detecting balance deterioration in a timely fashion, especially as the aged population increases worldwide. This can result in many falls that could have been avoided through continuous monitoring and early intervention. To fill the gap between available resources and care needs, an approach is needed for assessing the balance and mobility of the elderly in a timely manner without involving healthcare professionals.

Wearable systems based on inertial sensors are light, portable, and cheap, and they can be used to quantify body motions. Previous research (Howcroft et al., 2013) on fall risk assessment focused on feature-based methods, in which many related features are derived with domain knowledge. This requires multiple feature engineering steps before the classification or discrimination results can be obtained. The timed up and go test (TUG) is commonly used to evaluate mobility and the fall risk of the elderly in hospital and community environments (Podsiadlo and Richardson, 1991; Barry et al., 2014). Tri-axial acceleration sensors can be used to obtain time-domain signals during TUG (Wu et al., 2019; Lee et al., 2020), which can be transformed through time–frequency analysis (TFA) to extract time-domain, frequency-domain, and spectral energy-related information. Since the past literature (Cardozo et al., 2011) and (Garcia-Retortillo et al., 2020) has shown investigating spectral power distribution of muscle (using accelerometer data or related physiological parameters, such as EMG) and its response to fatigue and aging in elderly subjects, we can use spectral energy-related information to assess fall risk of elderly subjects *via* TUG test.

Nweke et al. (2018) showed that deep neural network (DNN) methods are being adopted for automatic feature learning in diverse fields such as health, image classification, and recently, for the feature extraction and classification of simple and complex human activity recognition in mobile and wearable sensors. They also provided further insights on deep learning based on the decision fusion of human activity recognition for enhanced performance accuracy. Hossain et al. (2018) showed that deep learning architectures have been increasingly used in activity recognition problems that empower several application domains that require considerably less human supervision in the process. Moreover, they showed that such architectures are gaining increasing popularity for extracting meaningful information from these large volumes of data. DNNs are suitable for TFA owing to their excellent discrimination of images. The non-stationary nature of the TUG signal indicates that TFA can be used for motion identification in general and fall detection in particular (Jokanovic et al., 2016a, July). Deep learning can be used to capture the detailed and complex properties of the TF signature and feed the learned underlying features to the classifier (Jokanovic et al., 2016a, May). An autoencoder (AE) is a feed-forward neural network that aims to reconstruct the input

at the output under certain constraints. Seyfioğlu et al. (2018) proposed an unsupervised pre-training algorithm for initializing the AE weights and bias that is highly effective when only a small number of labeled training samples are available. The stacked autoencoder (SAE) is a DNN that can classify highly similar classes of aided and unaided walking, as might be encountered in assisted-living environments for the elderly, and it has been applied in recognizing 12 different gaits (Seyfioğlu et al., 2017) as well as in fall detection.

In this paper, we propose the use of sensor and DNN-based technology, apply TFA to convert tri-axial accelerometer data and deep learning-based latent feature representation with a SAE to develop a surrogate approach for assessing the mobility function and fall risk detection of the elderly. And DNN-based analysis techniques will be an available approach for continuous monitoring in the future.

MATERIALS AND METHODS

We considered two evaluation methods for fall risk: feature-based and DNN-based evaluation. Feature-based evaluation, based on traditional statistical features and method for evaluation, combines feature extraction, feature selection and classifier, and it relies on heuristic handcrafted feature design. By contrast, DNN-based evaluation in this paper is based on the SAE and a softmax classifier layer, and it can automatically learn better feature representations than the handcrafted ones (Ng, 2011). Leave-one-out cross-validation was employed for both evaluation methods to ensure a robust classification accuracy. The results of the two evaluation methods were then compared.

Subjects

Our study took place at a hospital in central Taiwan between April 2014 and May 2015. We recruited and selected 44 elderly subjects dwelling in a community. A medical professional team that included rehabilitation physicians, physiotherapists and functional therapists performed TUG to evaluate the mobility function of the subjects. Prior to the evaluation, written consent was obtained from the subjects. The subjects were over 60 years of age, had no history of musculoskeletal injuries or central nervous system problems in the last 3 months and could walk independently without any help. Valid data were obtained for 44 elderly subjects with a mean age of 78.18 ± 7.97 years. There were 14 male subjects with an average age of 80.43 ± 5.60 years and 30 female subjects with an average age of 77.13 ± 8.74 years.

Sensor

As shown in **Figure 1**, a tri-axial accelerometer (RD3152MMA7260Q, Freescale Semiconductor-NXP, United States) with a sampling rate of 45 Hz was placed at vertebrae L3–L5 on a subject's back for the TUG experiments. L3–L5 correspond to the center of gravity of the human body and are used in most fall risk assessments (Howcroft et al., 2013). The X-, Y-, and Z-axes were aligned with the vertical (V; up: +, down: –), mediolateral (ML; right: +,

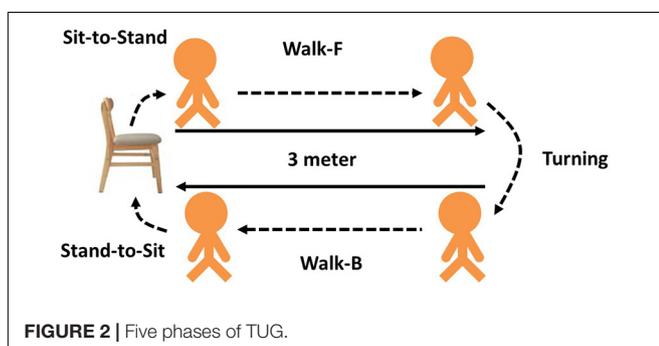
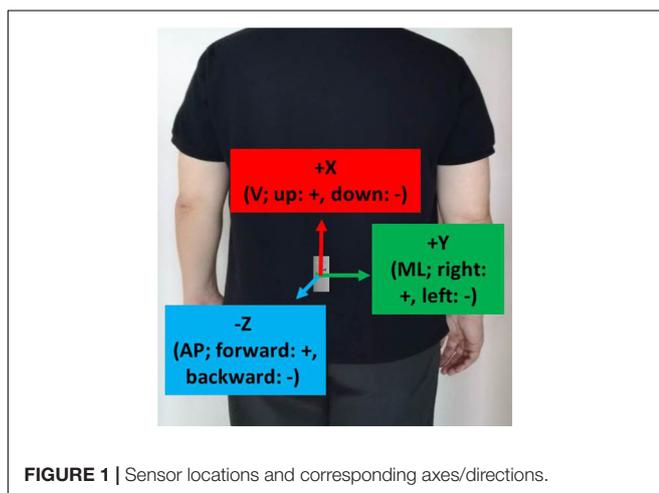
left: $-$), and anteroposterior (AP; forward: $+$, backward: $-$) directions, respectively.

Timed Up and Go Test

Each subject was asked to perform a TUG. The observer marked the start and end times. As shown in **Figure 2**, each TUG was divided into five phases or subtasks: from sitting to standing (sit-to-stand), walking forward (walk-F), reaching the 3-m mark and turning around (turning), walking backward (walk-B) and reaching the chair and returning to sitting (stand-to-sit). The TUG time was recorded, and a threshold time was determined to classify subjects as a fall risk or not a fall risk. Alexandre et al. (2012) recommend that it is considered a high fall risk if the time of community elderly for TUG is greater than 12.47 s.

Feature-Based Evaluation

For feature-based evaluation, the features of the axial signals were obtained by referring to past literature (Banos et al., 2014). The most widely used features include the mean, standard deviation, maximum, minimum, and mean crossing rate (MCR). The mean and standard deviation are used to express the average and variation of the force for each axial signal. The maximum and minimum express the largest and smallest values of the signal for the entire domain. The MCR is the rate at which data cross the average value, and it has been widely used in signal recognition



and physical activity recognition (Gao et al., 2014; Arivu et al., 2018; Bountourakis et al., 2019).

Features were selected for the feature-based evaluation according to their significance (Wu et al., 2019; Lee et al., 2020). The significance was obtained through Student's t -test. A feature was considered significant if $p \leq 0.05$. In addition, linear discriminant analysis (LDA) was performed to obtain a confusion matrix for evaluating the performance.

Deep Neural Network-Based Evaluation

Figure 3 shows the flowchart of the DNN-based evaluation. The input signal was the tri-axial data collected during the TUG experiments. TFA was applied to the data, and the SAE was applied in classifying the signal. Finally, the accuracy and confusion matrix were obtained.

Time–Frequency Analysis

A time–frequency representation (TFR) is a view of a signal, which is taken as a function of time, in both time and frequency domains. TFA can be applied to a time series signal to observe the time-domain, frequency-domain and spectral-energy information simultaneously. TFA based on wavelet transform (WT) is widely used in biomedical science for applications such as fall detection (Jokanovic et al., 2016a, July; Jokanovic et al., 2016b, May) and analysis of electroencephalography (Yordanova et al., 2013) and electromyography (Zia ur Rehman et al., 2018). In this study, the Morlet wavelet was used for TFA of the tri-axial acceleration signal from the TUG experiments. This method was described in previous literature (Tallon-Baudry et al., 1997). The complex Morlet wavelet $w(t, f_c)$ can be generated in the time-domain for different frequencies f as follows:

$$w(t, f_c) = A \exp(-t^2/2\sigma_t^2) \exp(i2\pi f_c t), \quad (1)$$

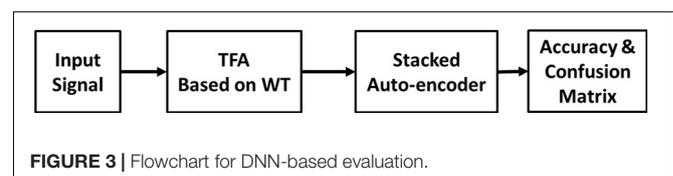
where t is the time, σ_t is the wavelet duration, normalization factor $A = (\sigma_t \sqrt{\pi})^{-1/2}$, a constant ratio of $f_c/\sigma_f = 7$ was used. f_c is the central frequency, and σ_f is the width of the Gaussian shape in the frequency-domain. For different f , the time and frequency resolutions can be calculated as $2\sigma_t$ and $2\sigma_f$, respectively, where $\sigma_t = 1/2\pi\sigma_f$. Finally, the time-varying energy $|E(t, f_c)|$ of the signal $[s(t)]$ is calculated by squaring the absolute value of the convolution of the signal with the complex Morlet wavelets:

$$E(t, f_c) = |w(t, f_c) \times s(t)|^2. \quad (2)$$

In this study, the frequency range was swept from 0.05 to 5 Hz, and a TF image was obtained for classification by the SAE.

Stacked Autoencoder Network Architecture

A neural network with multiple hidden layers can be used to solve classification problems with complex data such as images.



Each layer can learn features at a different level of abstraction. However, training a neural network with multiple hidden layers can be difficult. In this paper, we use the SAE structure, which is a DNN based on the AE concept. An AE is a neural network comprising an encoder, followed by a decoder, and it attempts to replicate its input at its output. We used an AE so that the hidden layers can be trained individually in an unsupervised fashion. No labeled data are required for training or learning. The encoder maps the input x to a new representation z , which is decoded back at the output to reconstruct the input \hat{x} : (Hinton and Salakhutdinov, 2006; Zia ur Rehman et al., 2018; MATLAB autoencoder, 2021).

$$z = h_1(W_1x + b_1), \tag{3}$$

$$\hat{x} = h_2(W_2z + b_2), \tag{4}$$

where h_1 and h_2 are activation functions, w_1 and w_2 are weight matrices and b_1 and b_2 are bias vectors for the encoder and decoder, respectively. Each layer can learn features with a different level of abstraction. If the number of hidden neurons is less than the number of input neurons, then the AE attempts to learn a sparse representation of the input data (Jokanovic et al., 2016a, July). Sparsity can be encouraged for an AE by adding a regulariser to the cost to prevent overfitting (Zia ur Rehman et al., 2018). In this study, the input was a color image with a resolution of 28×28 pixels and three channels ($28 \times 28 \times 3 = 2,352$ pixels). The AE had two hidden layers. The logistic sigmoid was used for both layers in the encoder and decoder.

In an SAE, the output of one AE is fed to the input of another AE, and sparsity is encouraged by adding regularization to the cost for neuron i . The average output activation for neuron i can be formulated as (MATLAB autoencoder, 2021):

$$\hat{p}_i = \frac{1}{n} \sum_{j=1}^n z_i(x_j), \tag{5}$$

where i is the i th neuron, n is the total number of training examples and j is the j th training example. A regulariser is introduced to the cost function using the Kullback–Leibler divergence: (Kullback, 1997; Zia ur Rehman et al., 2018).

$$\Omega_{\text{sparsity}} = \sum_{i=1}^d p \log\left(\frac{p}{\hat{p}_i}\right) + (1-p) \log\left(\frac{1-p}{1-\hat{p}_i}\right), \tag{6}$$

where d is the total number of neurons in a layer and p is the desired activation value (i.e., sparsity proportion). The L2 regularization term Ω_{weights} is also added to the cost function to control the weights:

$$\Omega_{\text{weights}} = \frac{1}{2} \sum_l^L \sum_j^N \sum_i^K (w_{ji}^l)^2, \tag{7}$$

where L is the number of hidden layers, N is the total number of observations and K is the number of features within an observation.

By inserting the regularization terms from Eqs 6, 7 into the mean squared error of the reconstruction, the cost function can be formulated as follows:

$$E = \frac{\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K (x_{kn} - \hat{x}_{kn})^2}{\text{mean square error}} + \lambda \cdot \frac{\Omega_{\text{weights}}}{\text{L2 Regularization}} + \beta \cdot \frac{\Omega_{\text{sparsity}}}{\text{Sparsity Regularization}}, \tag{8}$$

where λ is the coefficient for L2 regularization to prevent overfitting and β is the coefficient for sparsity regularization that controls the sparsity penalty term (MATLAB autoencoder, 2021).

Ju et al. (2015) and Coates et al. (2011) showed that the number of neurons in the hidden layer of a DNN may be more important than the feature-learning algorithm and model depth. In addition, the combinatorial space required to explore all possible combinations of hyperparameters is huge (Tsinalis et al.,

TABLE 1 | Demographic data of subjects at-risk of falling and not at-risk.

	Fall risk (n = 22)	Non-fall risk (n = 22)
Age	78.18 ± 7.97	76.59 ± 9.16
Gender		
Male	8	6
Female	14	16

TABLE 2 | Statistical features of TUG data for subjects.

Statistic features		Fall risk (n = 22)	Non-fall risk (n = 22)	p-value
V-axis	Mean_V	0.949 ± 0.272	1.333 ± 0.308	0.000**
	Std_V	1.217 ± 0.360	1.709 ± 0.341	0.000**
	Max_V	4.063 ± 1.603	4.899 ± 1.061	0.049*
	Min_V	-3.504 ± 1.314	-4.122 ± 1.280	0.122
	MCR_V	0.098 ± 0.015	0.087 ± 0.010	0.008*
ML-axis	Mean_ML	0.994 ± 0.251	0.979 ± 0.205	0.826
	Std_ML	1.131 ± 0.297	1.247 ± 0.251	0.167
	Max_ML	2.846 ± 0.981	3.622 ± 1.241	0.027*
	Min_ML	-2.919 ± 0.833	-3.422 ± 0.959	0.07
	MCR_ML	0.060 ± 0.014	0.084 ± 0.018	0.000**
AP-axis	Mean_AP	2.075 ± 0.765	1.666 ± 0.597	0.055
	Std_AP	1.823 ± 0.382	1.892 ± 0.336	0.528
	Max_AP	1.352 ± 0.862	2.055 ± 0.834	0.009*
	Min_AP	-7.544 ± 1.352	-7.057 ± 1.494	0.263
	MCR_AP	0.050 ± 0.014	0.063 ± 0.014	0.005*

*Indicate $p < 0.05$ between two groups. **Indicate $p < 0.005$ between two groups.

TABLE 3 | Classification results for LDA classifiers.

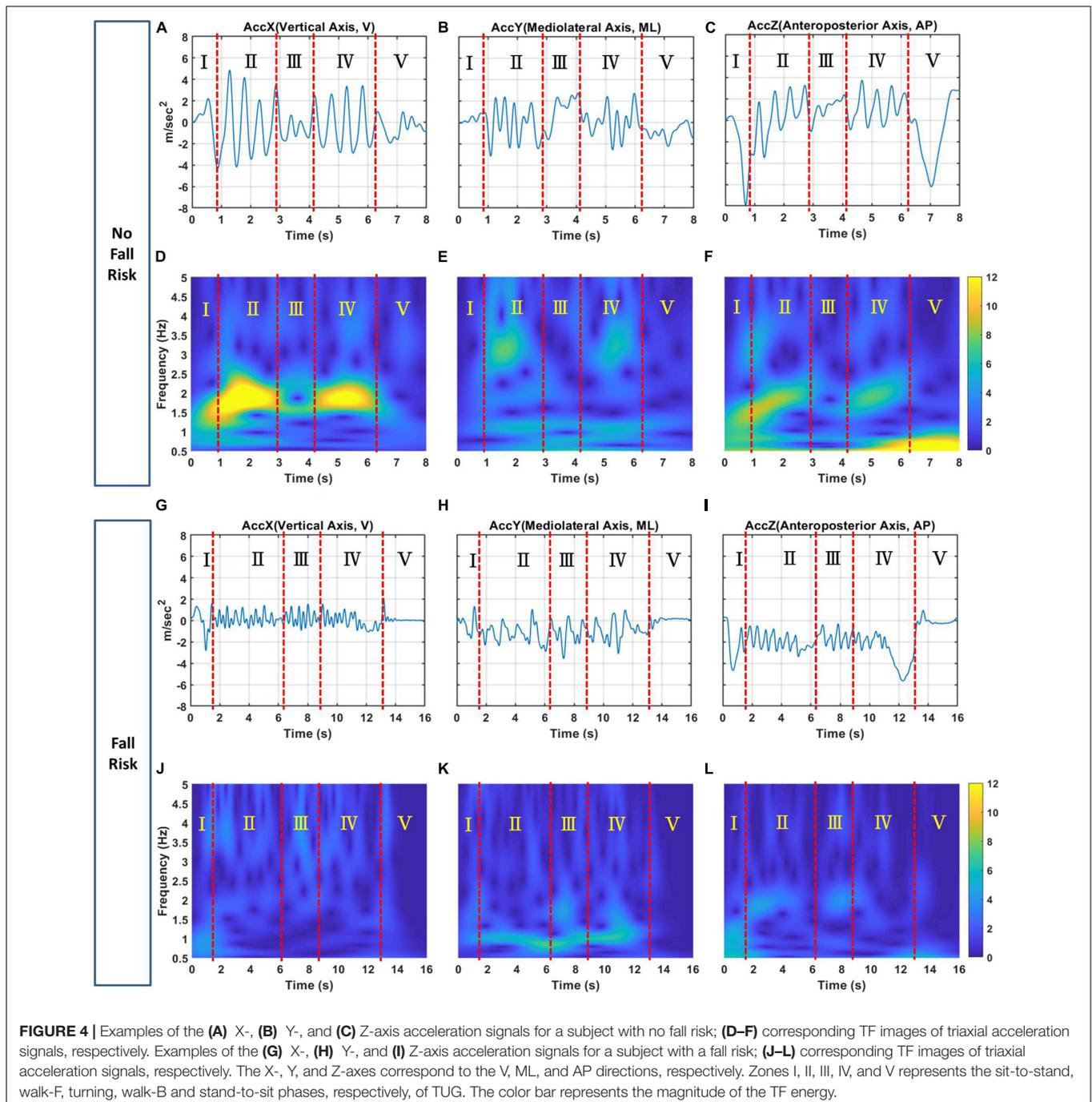
Axis	Acc.	Sen.	Spec.
V-axis	79.50%	72.70%	86.40%
ML-axis	81.80%	81.80%	81.80%
AP-axis	75.00%	72.70%	77.30%

2016). Therefore, we focused on locally optimizing the number of neurons for two layers and obtained the minimum mean squared error according to Eq. 8. The other parameters were taken from MATLAB: λ was set to 0.004 and 0.002 for the first and second hidden layers, respectively, $\beta = 4$ for both hidden layers and p was 0.015 and 0.01, respectively. After unsupervised training, the decoder was removed from the network, and the remaining encoder components were trained in a supervised manner by adding a softmax classifier with two neurons after the encoder. The softmax classifier is an advanced version of probability-based

logistic regression and is often used in the final layer of a neural network. Finally, the SAE was obtained.

RESULTS AND DISCUSSION

Subjects were considered a fall risk if their TUG time was greater than 12.47 s and not a fall risk if the TUG time was less than 12.47 s. **Table 1** lists the demographic data of the at-risk subjects ($n = 22$) and no-risk subjects ($n = 22$).



Feature-Based Analysis of the Timed Up and Go Test Results

Table 2 details the *t*-test results for the significance of the 15 statistical features of the tri-axial acceleration data. Eight significant features were identified (Mean_V, Std_V, Max_V, MCR_V, Max_ML, MCR_ML, Max_AP, and MCR_AP), which are aligned with normality by using Kolmogorov-Smirnov test, and LDA was applied to each axis. **Table 3** presents the classification results. The classification accuracies along the X-axis (V), Y-axis (ML), and Z-axis (AP) were 79.5, 81.8, and 75.0%, respectively. The sensitivities were 72.7, 81.8, and 72.7%, respectively. The specificities were 86.4, 81.8, and 77.3%, respectively. These results were then used for comparison to the DNN-based evaluation.

Deep Neural Network-Based Analysis of Timed Up and Go Test Results

Analysis of TF Images

Figure 4 shows examples of tri-axial acceleration signals in the time-domain for subjects with and without a fall risk and their corresponding TF images.

(1) For the X-axis, this axis is the vertical acceleration signal in time domain for the no-risk and at-risk subjects showed as **Figures 4A,G**, respectively. **Figures 4A,G** can be transformed through TFA to obtain TF images showed **Figures 4D,J**. **Figure 4D** clearly shows that the no-risk subject had two regions of interest in zones II and IV of the TF image corresponding to the walk-F and walk-B phases. The TF energy was 10–12, and the frequency was 1.5–2.5 Hz. Similarly, **Figure 4J** shows that the at-risk subject had regions of interest in zones II and IV

corresponding to the walk-F and walk-B phases. The TF energy was 0.5–3, and the frequency was 1.5–2.5 Hz. Additionally, the turning phase showed obvious difference in Zone III between **Figures 4D,J**. The TF energy was 6–8, and the frequency was 1.5–2.0 Hz for no-risk subject. On the contrary, the TF energy was relatively low for no-risk subject. This is consistent with previous study (Drover et al., 2017; Wu et al., 2019), which noted that turn-based features are important predictors because they contain useful biomechanical information.

(2) For the Y-axis, this axis is the mediolateral acceleration signal in time domain for the no-risk and at-risk subjects showed as **Figures 4B,H**, respectively. **Figures 4B,H** can be transformed through TFA to obtain TF images showed **Figures 4E,K**. **Figure 4E** shows that the no-risk subject had two regions of interest in zones II and IV of the TF image corresponding to the walk-F and walk-B phases. The regions had high TF energies of 5–8 and 4–6, respectively, corresponding to frequencies of 2.5–3.5 and 1–1.3 Hz, respectively. Similarly, **Figure 4K** shows that the at-risk subject had regions of interest in zones II and IV corresponding to the walk-F and walk-B phases. Only one region had a high TF energy of 4–6 with a frequency of 1–1.3 Hz. In the Walk_F and Walk_B phases, the TF image showed the energy of mobility, which is supposedly related to the body and the arm swing when walking. Because of the walking duration, the arm swing is associated with postural stability (Meyns et al., 2013) can enhance gait stability (Bruijn et al., 2010).

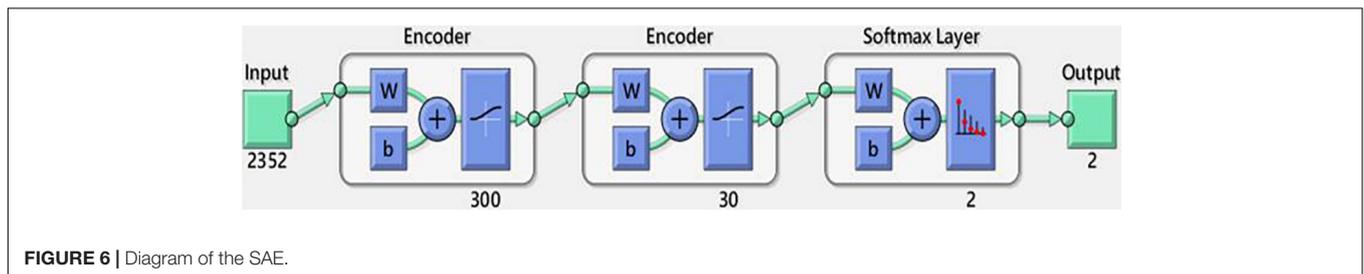
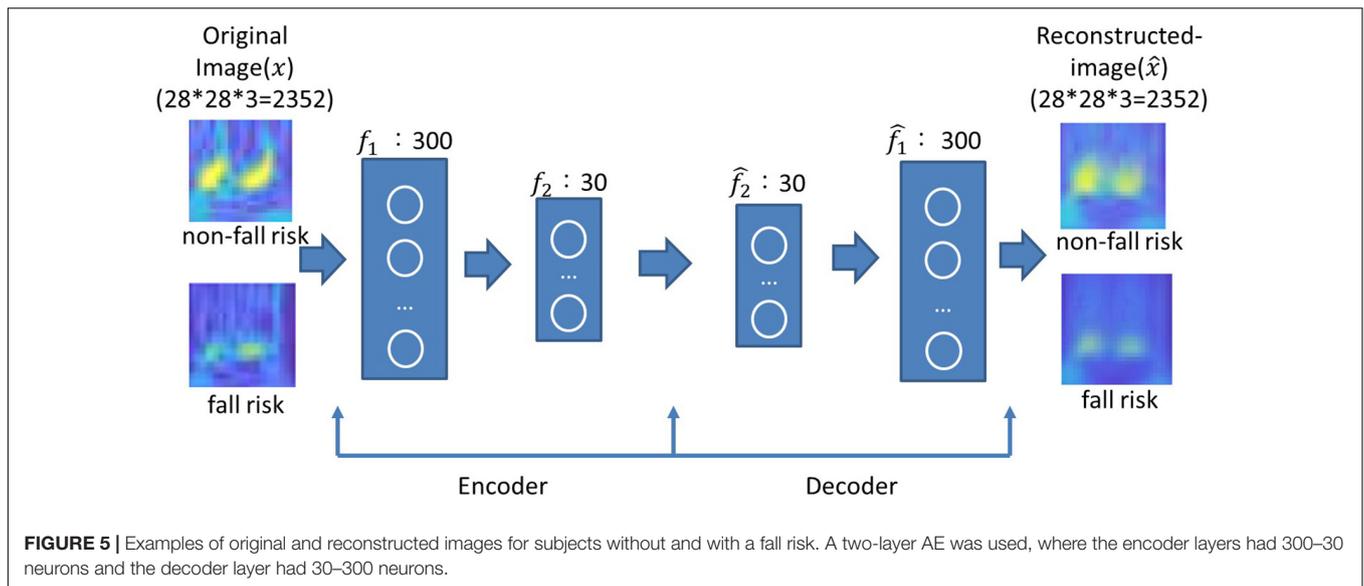
(3) For the Z-axis, this axis is the anteroposterior acceleration signal in time domain for the no-risk and at-risk subjects showed as **Figures 4C,I**. **Figures 4C,I** can be transformed through TFA to obtain TF images showed **Figures 4F,L**. **Figure 4F** shows that the no-risk subject had two regions of interest in zones II and IV of the TF image corresponding to the walk-F and walk-B phases. The TF energy was 6–9, and the frequency was 1.5–2.5 Hz. Similarly, **Figure 4L** shows that the at-risk subject had regions of interest in zones II and IV corresponding to the walk-F and walk-B phases. The TF energy was 2.5–4, and the frequency was 1.5–2.5 Hz. The body will move forward to maintain balance while walking, and the AP-axis is seemingly an important axis.

In summary, the no-risk subjects had higher TF energy than the at-risk subjects in zones II and IV corresponding to the walk-F and walk-B phases for all three axes. This is a reasonable assumption that no-risk subjects must have greater muscle strength or energy when walking than those with at-risk subjects. In addition, the no-risk subjects had obviously higher TF energy than the at-risk subjects did during the sit-to-stand (Zone I) and stand-to-sit (Zone V) phases in the Z-axis, referring to the transition subtask involving standing up and sitting down and these two abilities are largely related to strength and power of the lower extremities (Weiss et al., 2013). Moreover, the body must bend in forward–backward displacement. It is also reasonable to infer that the no-risk subjects had more energy to stand up or sit down than the at-risk subjects did. Regarding the differences located between 1 and 3 Hz approximately, the past literature (Schneider et al., 2010; Kline et al., 2016) have showed the frequency for movements along the longitudinal axis during running peaks at approximately 3 Hz, both in the activity

TABLE 4 | Mean squared errors for different combinations of neuron numbers in the first and second layers of the two-layer AE for the X-axis (V), Y-axis (ML), and Z-axis (AP).

Axis	The neuron number of 1st layer	The neuron number of 2nd layer				
		10	15	20	25	30
X-axis (V)	100	18.39	21.96	22.39	21.88	21.24
	200	23.07	23.52	19.93	19.77	22.06
	300	18.32	19.07	18.23	16.49	15.34*
	400	21.69	22.91	20.88	22.35	20.10
	500	21.62	20.58	18.65	20.65	18.52
Y-axis (ML)	100	16.13	14.02	15.01	13.47	15.11
	200	15.35	14.64	13.63	13.49	13.67
	300	14.11	13.59	13.43	12.46	12.03*
	400	14.15	13.01	12.69	12.43	13.09
	500	13.89	13.15	13.09	12.25	12.54
Z-axis (AP)	100	12.88	11.89	11.14	13.38	11.97
	200	15.37	14.05	10.63	14.14	10.47
	300	12.16	11.43	11.10	12.25	9.73*
	400	13.35	12.92	13.97	11.54	14.35
	500	14.23	12.66	12.21	11.21	10.30

*The minimum of mean square error.



and viewed movement conditions. They reported that a strong relationship exists between intrinsic and extrinsic oscillation patterns during exercise. A frequency of approximately 3 Hz seems to be dominant in different physiological systems (e.g., heart rate and brain cortical activity). Additionally, Robert C. et al. mentioned that when the step frequency fell in the range of 0.5–3 Hz, the activity was identified as walking (Wagenaar et al., 2011). Compare with these results, we assume TF images may be used as an auxiliary tool to support medical professionals for clinically assessing fall risk.

Parameter Optimization for AE and Reconstruction

The number of neurons was chosen according to the grid search strategy to minimize the mean squared error (Hinton and Salakhutdinov, 2006). The number of neurons in the first layer ranged from 100 to 500 in intervals of 100, and the number of neurons in the second layer ranged from 10 to 30

TABLE 5 | Classification results with the SAE.

Axis	Accuracy	Sensitivity	Specificity
X-axis(V)	89.1 ± 1.0%	85.5 ± 4.2%	92.7 ± 3.2%
Y-axis(ML)	93.4 ± 1.7%	94.1 ± 2.2%	92.7 ± 3.8%
Z-axis(AP)	94.1 ± 1.6%	94.6 ± 2.9%	93.6 ± 2.4%

The accuracy, sensitivity, and specificity are average ± SD values for ten runs.

in intervals of 5. The mean squared error was obtained by averaging ten runs. As presented in **Table 4**, the minimum mean squared errors for the X-, Y- and Z-axes were 15.34, 12.03, and 9.73, respectively. These corresponded to 300 and 30 neurons in the first and second layers, respectively, for all three axes. Image reconstruction was carried out with 300–30 neurons for the encoder and 30–300 neurons for the decoder. As shown in **Figure 5**, the reconstructed image successfully restored the original image. Unsurprisingly, the latent features were useful for object recognition and other visual tasks (Ng, 2011).

Analysis of the Stacked Autoencoder

As shown in **Figure 6**, the SAE had an input of 2,352 pixels with an encoder layer of 300–30 neurons and a softmax classifier layer with two classes. **Table 5** presents the classification results of the SAE. The classification accuracies were 89.1, 93.4, and 94.1% along the X-, Y-, and Z-axes, respectively. The sensitivities were 85.5, 94.1, and 94.6%, respectively. The specificities were 92.7, 92.7, and 93.6%, respectively. The SAE performed better along the Y- and Z-axes than along the X-axis. Thus, the latent features of the Y- and Z-axes may offer more predictive ability for DNN-based evaluation.

Tables 3, 5 indicate that the DNN-based evaluation performed much better than the feature-based evaluation. Thus, it is a

viable approach for fall detection. In addition, the Y- and Z-axes are both important for classification. With regard to the Y-axis, swinging arms are associated with postural stability and can enhance gait stability (Wu et al., 2019) and mobility function. With regard to the Z-axis, this is important to transitions involving standing up or sitting down, where the body must bend in forward–backward displacement. These results are similar to those of previous study (Wu et al., 2019), who identified features extracted along the Z-axis for TUG tasks as significant and Z-axis is seemingly an important axis.

CONCLUSION

In this paper, tri-axial accelerometer data were collected from a cheap wearable sensor, and TFA was used to convert the data into TFRs. These TF images offered abundant and discriminative information such as time, frequency and spectral energy-related power in five phases of TUG, which clarified specific TUG aspects or subtasks were impaired in mobility. High energy-related power of no-risk subjects in both walk phases (walk-F and walk-B) and transition phases (sit-to-stand and stand-to-sit) phases can be observed obviously from TF images for all three axes and AP axis, respectively. We also applied SAE model, DNN-based evaluation, to classify TFRs of elderly subjects for assessing the mobility and fall risk. Experimental results show that the DNN-based evaluation offers much considerably accuracy, sensitivity and specificity rates. Moreover, the results indicated the superior performance of DNN-based evaluation over feature-based evaluation. Further, the discrimination analysis of Y and Z axes seems to be more important than that of X axis.

In the future, we will continuously work on DNN-based evaluation of fall risk for the elderly. This innovative method based on the artificial intelligence technology, i.e., DNN-based evaluation, can be widely used in wearable sensing technology, smart home development and continuous monitoring technologies for real-time measurement and recording of various physiological signals. We trust it will improve the accessibility and convenience of people's medical care.

REFERENCES

- Alexandre, T. S., Meira, D. M., Rico, N. C., and Mizuta, S. K. (2012). Accuracy of Timed Up and Go Test for screening risk of falls among community-dwelling elderly. *Braz. J. Phys. Ther.* 16, 381–388. doi: 10.1590/S1413-35552012005000041
- Arivu, S. A., Amutha, R., Muthumeenakshi, K., and Edna Elizabeth, N. (2018). "Design of smart vest to monitor physical activities of children," in *Fourth International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, (Piscataway: IEEE), doi: 10.1109/AEEICB.2018.8480993
- Banos, O., Galvez, J. M., Damas, M., Pomares, H., and Rojas, I. (2014). Window size impact in human activity recognition. *Sensors* 14, 6474–6499. doi: 10.3390/s140406474
- Barry, E., Galvin, R., Keogh, C., Horgan, F., and Fahey, T. (2014). Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta-analysis. *BMC Geriatr.* 14:14. doi: 10.1186/1471-2318-14-14
- Bergland, A. (2012). Fall risk factors in community-dwelling elderly people. *Nor. Epidemiol.* 22, 151–164. doi: 10.5324/nje.v22i2.1561

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: we have signed a confidentiality agreement with the hospital. Requests to access these datasets should be directed to C-HL, sweat0430@mail.ntust.edu.tw.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Tsaotun Psychiatric Center, Ministry of Health and Welfare (IRB No. 104013). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

S-HC, C-HL, BJ, and T-LS: conceptualization and validation. C-HL, S-HC, and T-LS: data curation. S-HC: formal analysis. C-HL, BJ, and T-LS: investigation. S-HC and C-HL: methodology, software, and writing – original draft. BJ and T-LS: resources and writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Ministry of Science and Technology under projects MOST 103-2221-E-155-044-MY3, MOST 105-2218-E-011-010-MY3, MOST 106-2221-E-155-023-MY3, and MOST 109-2223-E-011-001-MY3 for which we are especially grateful.

ACKNOWLEDGMENTS

We would like to thank the staff of the Feng-Yuan Hospital in Taiwan.

- Bountourakis, V., Vrysis, L., Konstantoudakis, K., and Vryzas, N. (2019). An enhanced temporal feature integration method for environmental sound recognition. *Acoustics* 1, 410–422. doi: 10.3390/acoustics1020023
- Bruijn, S. M., Meijer, O. G., Beek, P. J., and van Dieën, J. H. (2010). The effects of arm swing on human gait stability. *J. Exp. Biol.* 213, 3945–3952. doi: 10.1242/jeb.045112
- Cardozo, A. C., Gonçalves, M., and Dolan, P. (2011). Back extensor muscle fatigue at submaximal workloads assessed using frequency banding of the electro-myographic signal. *Clin. Biomech. (Bristol, Avon)* 26, 971–976. doi: 10.1016/j.clinbiomech.2011.06.001
- Coates, A., Lee, H., and Ng, A. Y. (2011). "An analysis of single-layer networks in unsupervised feature learning," in *Appearing in Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) 14, vol. 15*, (Fort Lauderdale, FL: JMLR). (Access by Google Scholar).
- Cuevas-Trisan, R. (2017). Balance problems and fall risks in the elderly. *Phys. Med. Rehabil. Clin. N. Am.* 28, 727–737. doi: 10.1016/j.pmr.2017.06.006
- Drover, D., Howcroft, J., Kofman, J., and Lemaire, E. D. (2017). Faller classification in older adults using wearable sensors based on turn and straight-walking accelerometer-based features. *Sensors* 17:1321. doi: 10.3390/s17061321

- Ganz, D. A., Bao, Y., Shekelle, P. G., and Rubenstein, L. Z. (2007). Will my patient fall?. *JAMA* 297, 77–86. doi: 10.1001/jama.297.1.77
- Gao, L., Bourke, A. K., and Nelson, J. (2014). Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Med. Eng. Phys.* 36, 779–785. doi: 10.1016/j.medengphy.2014.02.012
- Garcia-Retortillo, S., Rizzo, R., Wang, J. W. J. L., Sitges, C., and Ivanov, PCH (2020). Universal spectral profile and dynamic evolution of muscle activation: a hallmark of muscle type and physiological state. *J. Appl. Physiol.* 129, 419–441. doi: 10.1152/jappphysiol.00385.2020
- Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647
- Howcroft, J., Kofman, J., and Lemaire, E. D. (2013). Review of fall risk assessment in geriatric populations using inertial sensors. *J. Neuroeng. Rehabil.* 10:91. doi: 10.1186/1743-0003-10-91
- Jokanovic, B., Amin, M., and Ahmad, F. (2016a). “Radar fall motion detection using deep learning,” in *IEEE radar conference (RadarConf), 2016*, (Piscataway: IEEE), 1–6. doi: 10.1109/RADAR.2016.7485147
- Jokanovic, B., Amin, M. G., and Ahmad, F. (2016b). “Effect of data representations on deep learning in fall detection,” in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 2016*, (Piscataway: IEEE), 1–5. doi: 10.1109/SAM.2016.7569734
- Ju, Y., Guo, J., and Liu, S. (2015). “A deep learning method combined sparse autoencoder with SVM,” in *2015 international conference on cyber-enabled distributed computing and knowledge discovery*, (Piscataway: IEEE), doi: 10.1109/CyberC.2015.39
- Kline, J. E., Huang, H. J., Snyder, K. L., and Ferris, D. P. (2016). Cortical spectral activity and connectivity during active and viewed arm and leg movement. *Front. Neurosci.* 10:91. doi: 10.3389/fnins.2016.0009
- Kullback, S. (1997). *Information theory and statistics*. Courier Corporation. Hoboken: Wiley Online Library, doi: 10.1002/9781118445112.stat01635
- Lee, C. H., Chen, S. H., Jiang, B. C., and Sun, T. L. (2020). Estimating postural stability using improved permutation entropy via TUG accelerometer data for community-dwelling elderly people. *Entropy* 22:1097. doi: 10.3390/e22101097
- Letts, L., Moreland, J., Richardson, J., Coman, L., Edwards, M., Ginis, K. M., et al. (2010). The physical environment as a fall risk factor in older adults: systematic review and meta-analysis of cross-sectional and cohort studies. *Aust. Occup. Ther. J.* 57, 51–64. doi: 10.1111/j.1440-1630.2009.00787.x
- MATLAB autoencoder (2021). MATLAB autoencoder. Available online at: <https://www.mathworks.com/help/deeplearning/ref/trainautoencoder.html;jsessionid=e99dbcbcfdbbafaf24af307fb6e2> (accessed Jan 30, 2021).
- Meyns, P., Bruijn, S. M., and Duysens, J. (2013). The how and why of arm swing during human walking. *Gait Posture* 38, 555–562. doi: 10.1016/j.gaitpost.2013.02.006
- Mitchell, W. K., Williams, J., Atherton, P., Larvin, M., Lund, J., and Narici, M. (2012). Sarcopenia, dynapenia, and the impact of advancing age on human skeletal muscle size and strength; a quantitative review. *Front. Physiol.* 3:260. doi: 10.3389/fphys.2012.00260
- Ng, A. (2011). Sparse autoencoder. *CS294A Lecture notes*. 72, 1–19. (Access by Google Scholar),
- Podsiadlo, D., and Richardson, S. (1991). The timed “up & go”: a test of basic functional mobility for frail elderly persons. *J. Am. Geriatr. Soc.* 39, 142–148. doi: 10.1111/j.1532-5415.1991.tb01616.x
- Hossain, H. M. S., Abdullah Al Haiz Khan, M. D., and Roy, N. (2018). DeActive: scaling activity recognition with active deep learning. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1–23. doi: 10.1145/3214269
- Schneider, S., Askew, C. D., Abel, T., and Strüder, H. K. (2010). Exercise, music, and the brain: is there a central pattern generator? *J. Sports Sci.* 28, 1337–1343. doi: 10.1080/02640414.2010.507252
- Seyfioglu, M. S., Gürbüz, S. Z., Özbayoglu, A. M., and Yüksel, M. (2017). “Deep learning of micro-Doppler features for aided and unaided gait recognition,” in *IEEE Radar Conference (RadarConf)*, (Piscataway: IEEE), doi: 10.1109/RADAR.2017.7944373
- Seyfioglu, M. S., Ozbayoglu, A. M., and Gurbuz, S. Z. (2018). Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities. *IEEE Trans. Aerosp. Electron. Syst.* 54, 1709–1723. doi: 10.1109/TAES.2018.2799758
- Shany, T., Redmond, S. J., Narayanan, M. R., and Lovell, N. H. (2012). Sensors-based wearable systems for monitoring of human movement and falls. *IEEE Sens. J.* 12, 658–670. doi: 10.1109/JSEN.2011.2146246
- Tallon-Baudry, C., Bertrand, O., Delpeuch, C., and Permier, J. (1997). Oscillatory γ -band (30–70 Hz) activity induced by a visual search task in humans. *J. Neurosci.* 17, 722–734. doi: 10.1523/JNEUROSCI.17-02-00722.1997
- Tsinalis, O., Matthews, P. M., and Guo, Y. (2016). Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. *Ann. Biomed. Eng.* 44, 1587–1597. doi: 10.1007/s10439-015-1444-y
- Wagenaar, R. C., Sapir, I., Zhang, Y., Markovic, S., Vaina, L. M., and Little, T. D. (2011). Continuous monitoring of functional activities using wearable, wireless gyroscope and accelerometer technology. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2011, 4844–4847. doi: 10.1109/IEMBS.2011.6091200
- Nweke, H. F., Wah, T. Y., Al-Garadi, M. A., and Alo, U. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. *Expert Syst. Appl.* 105, 233–261. doi: 10.1016/j.eswa.2018.03.056
- Weiss, A., Mirelman, A., Buchman, A. S., Bennett, D. A., and Hausdorff, J. M. (2013). Using a body-fixed sensor to identify subclinical gait difficulties in older adults with IADL disability: maximizing the output of the timed up and go. *PLoS One* 8:e68885. doi: 10.1371/journal.pone.0068885
- World Health Organization [WHO]. (2007). *Global Age-Friendly Cities: A Guide*. Geneva: World Health Organization, doi: 10.1080/17441692.2011.652972
- Wu, C. H., Lee, C. H., Jiang, B. C., and Sun, T. L. (2019). Multiscale entropy analysis of postural stability for estimating fall risk via domain knowledge of timed-up-and-go accelerometer data for elderly people living in a community. *Entropy* 21:1076. doi: 10.3390/e21111076
- Yordanova, J., Kolev, V., and Rothenberger, A. (2013). Event-related oscillations reflect functional asymmetry in children with attention deficit/hyperactivity disorder. *Suppl. Clin. Neurophysiol.* 62, 289–301. doi: 10.1016/B978-0-7020-5307-8.00018-1
- Zia ur Rehman, M., Gilani, S., Waris, A., Niazi, I., Slabaugh, G., Farina, D., et al. (2018). Stacked sparse autoencoders for EMG-based classification of hand motions: a comparative multi day analyses between surface and intramuscular EMG. *Appl. Sci.* 8:1126. doi: 10.3390/app8071126 doi: 10.3390/app8071126

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Chen, Lee, Jiang and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.