

# Wearable Sensor Gait Analysis of Fall Detection using Attention Network

1<sup>st</sup> Haben Yhdego

*Computational Modeling and Simulation Engineering  
Old Dominion University  
Norfolk, USA  
hyhde001@odu.edu*

2<sup>nd</sup> Jiang Li

*Electrical and Computer Engineering  
Old Dominion University  
Norfolk, USA  
jli@odu.edu*

3<sup>rd</sup> Christopher Paolini

*Electrical and Computer Engineering  
San Diego State University  
San Diego, California  
cpaolini@sdsu.edu*

4<sup>th</sup> Michel Audette

*Computational Modeling and Simulation Engineering  
Old Dominion University  
Norfolk, USA  
maudette@odu.edu*

**Abstract**—The statistical data from the National Council on Aging indicates that a senior adult dies in the US from a fall every 19 minutes. The care of elderly people can be improved by enabling the detection of falling events, especially if it triggers the pneumatic actuation of a protective airbag. This work focuses on detecting impending fall risk of senior subjects within the geriatric population, towards a planned approach to mitigating fall injuries through pneumatic airbag deployment. With the widespread adoption of wearable sensors, there is an increased emphasis on fall prediction models that effectively cope with accelerometry signal data. Fall detection and gait classification are challenging tasks, especially in differentiating falls from near falls. We propose to apply attention to the deep neural network (DNN) analysis of acceleration data where a fall is known to have occurred. We take the maximum value of the sensor signals to define the observation window of the detector. Powered by a transformer DNN with word embedding, attention networks have achieved a state-of-the-art in natural language processing (NLP) tasks. Besides the success of the transformer for efficiently processing long sequences, it supports parallel computing with fast computation. In this paper, we propose a novel transformer attention network for gait analysis of fall detection modeling with Time2Vec positional encoding- founded on a Masked Transformer Network. Using our dataset, we demonstrate that the proposed approach achieves better specificity and sensitivity than the present models.

**Index Terms**—Attention network, Fall Detection, Gait Analysis, Transformer.

## I. INTRODUCTION

The data from WHO shows that there is around six hundred thousand death due to falls occurred yearly- which makes it the second-highest mortality rate following traffic accidents [1]. The death rate is high in the aging population (age 65+), which is half of the injury-related hospitalizations of senior subjects [1]. Consequently, almost 40% of the injury-related deaths are from a fall in the elderly population [1]. Furthermore, based on the CDC [2] data, an elder adult, age >65, in the USA suffers from fall

every second. Most fall detection systems use a wearable device with inertial sensors providing accelerometry as the main source of data, complemented with other sensors like gyroscopes and magnetometers. Adding extra parameters such as angular velocity helps to enhance the algorithm performance, but affects the computational demand at the same time. Independently of the device used, various fall detection methods have high false positives, false negatives, and computational time cost. We aim to reduce both false positives and false negatives coupled with low consumption methods. Our work uses an attention neural network which has the advantage of parallel computation while reducing the false positive and false negative rates by giving a high attention score to the important signal information.

Attention neural networks have recently achieved significant progress in natural language processing (NLP). They have enabled models like Transformer [3] and BERT [4] to form powerful language models that can be used for machine translation, sentiment analysis, and document detection. With their recent success in NLP, one would expect an adaptation of these methods for other time-series data to solve problems such as fall detection datasets - as both applications involve processing sequential data. However, attention networks for time series data are still inadequate [3].

This study proposes a new approach that uses attention-based networks for sensor signals to create a robust fall detection system. In his work on attention network applied to NLP, Cheng [5] emphasized that humans focus on key words when reading the sentence. It is not difficult for the reader to relate the word with the other words in the sentence, but a neural network would need to specifically design parts of it to replicate this attention to certain words. Similar to the NLP, in our case, first, we form input features that feed to the transformer by finding the maximum acceleration magnitude

vector and the fixed window size around this peak signal magnitude- to be used as input to our attention networks.

## II. RELATED WORKS

In the last decade, many research efforts have been devoted to the development of efficient and cost-effective fall detection systems (FDS). FDS help us to differentiate falls from activities of daily living (ADL), so that an alert system to a remote monitoring point is automatically emitted as soon as the patient falls [6]. Commonly, fall detection systems are categorized into two different classes depending on the deployed sensor technology: wearable sensors and ambient sensors (such as cameras, vibration, or infrared sensors). The ambient-based system, which is expensive, is restricted to a particular pre-defined area and also affects the privacy of patients caused by visual sensors. Considering those limitations, wearable FDS offer a cheaper alternative to detect falls based on one or several wearable sensors, which are attached to the user's body or clothes so that they can be ubiquitously transported.

Montesinos [7] and Klenk [8] have thoroughly compared the performance of basic thresholding techniques with a wide set of supervised-based learning solutions (mainly Support Vector Machine (SVM) and Convolutional Neural Network (CNNs)). Simple thresholding techniques are inapplicable because threshold techniques as we are deciding the average range of fall and non-fall activity for new test data. Although CNN networks were mainly used for image feature extraction, they have been also used in some recent works for fall detection systems of inertial sensor signals and action recognition using wearable sensor signals [9], [10]. Those papers mainly use raw inertial sensors as inputs. In contrast, the method of Yhdego [11] detects falls by converting the 1D sensor signals to 2D images consisting of spectrotemporal footprints and then applying CNN to classify those converted images.

Many fall detection methods take advantage of the Long Short-Term Memory (LSTM) architecture, which has the ability to encode and process sequential data. Other deep learning models like CNN and conventional machine learning algorithms like SVM process their data without any notion of sequential order. Therefore, LSTM models provide a promising result for modeling fall activities. Sensor fusion of accelerometer and gyroscope using CNN-LSTM method by Ruben [12], and LSTM-based activity recognition by Paolini [13] and Aicha [14], are used for fall recognition. Aicha [14] employs a single inertial sensor placed on the trunk and inputted to CNN and LSTM models. The limitation of Ruben's model is that he add a KNN last layer classifier with the CNN-LSTM method which is inefficient for real-time fall detection due to the computational cost. Moreover, the use of raw sensor signals for training CNN and LSTM models might contain noisy data and result in

low specificity, when we have near-fall activities such as a subject activity into a seated position in a chair, couch, or bed, which involves relatively large motion in the vertical direction.

Deep neural networks such as RNN and LSTM, which consider the temporal and spatial relationship across the data, have achieved better results in several sequential tasks. Regardless of the progress of these models in time series data, parallel processing for the layer outputs is not possible - thus makes it difficult to learn long-range dependencies. In the last decade researchers proposed attention-based models for NLP tasks and they obtained state-of-the-art results with less computation time [15]. However, these models have not been frequently used for sequential data such as fall detection. This paucity is likely due to several factors such as the difficulty of accurately encoding positional information; focus on point-wise values, and lack of research about handling multivariate input features. Additionally, outside of natural language processing many researchers are probably not familiar with attention neural networks and their potential. We use the attention neural network toward a fall detection system in order to achieve better specificity and sensitivity performance.

## III. METHODOLOGY

Attention models are networks that introduce a weighting of signals based on importance- which helps the model to emphasize important pieces in the feature space. To calculate the output of the attention layer, first score values of the input windows are calculated using the score function. Next, the Softmax activation function is used to find the attention weights for each observation windows computed. Finally, the output of the attention layer is the weighted sum of the values [16]. Hence, this weighted sum mechanism allows the model to focus and place more attention on the relevant parts of the input sequence.

Therefore, to use this advantage of attention network by giving the highest attention score to the determinant window of signal, we start sliding window feature engineering from the peak magnitude of the acceleration signal (as shown in algorithm 1) rather than using a conventional sliding window feature engineering. This kind of windowing helps us to include the important peak signal value in a single window, besides giving high attention to the maximum value signal. If we were using the conventional windowing based on the raw signal, the important information may be shared in different windows. For our proposed method, the maximum values of the acceleration and angular velocity simulated in the datasets of the accelerometer and gyroscope data, which consist of 3D acceleration data  $acc(t) = [acc_x(t), acc_y(t), acc_z(t)]$  and 3D angular velocity  $gyro(t) = [gyro_x(t), gyro_y(t), gyro_z(t)]$  [11], are calculated using the magnitude of these vectors for the  $i - th$  sample as followed:

$$\|acc_i(t)\| = \sqrt{acc_{xi}^2(t) + acc_{yi}^2(t) + acc_{zi}^2(t)}. \quad (1)$$

$$\|gyro_i(t)\| = \sqrt{gyro_{xi}^2(t) + gyro_{yi}^2(t) + gyro_{zi}^2(t)}. \quad (2)$$

Next, the maximum magnitude of the signal calculated above is determined as followed:

$$\|acc_{max}\| = \max(\{\|acc_i\| : i \in [1 : M]\}), \quad (3)$$

$$\|gyro_{max}\| = \max(\{\|gyro_i\| : i \in [1 : M]\}), \quad (4)$$

where M is the number of samples. Using the above two equations and the window size (which is 1 second in our case), the sliding window around the peak signal looks like as shown in the figure 1.

---

**Algorithm 1:** Calculating maximum value and labeling the window

---

**input :**  $acc(t) = [acc_x(t), acc_y(t), acc_z(t)]$  and  $gyro(t) = [gyro_x(t), gyro_y(t), gyro_z(t)]$   
**output:** Observation Windows and its label  
**for all data sequence activities of FallData do**  
    Calculate the magnitude using equation 1 and 2;  
    Calculate the maximum value using equation 3 and 4;  
    **if FallData is ADL then**  
        Label all the observation windows as ADL;  
    **else**  
        **if the observation window is maximum value window then**  
            label the window as Fall;  
        **else**  
            Label ADL;

---

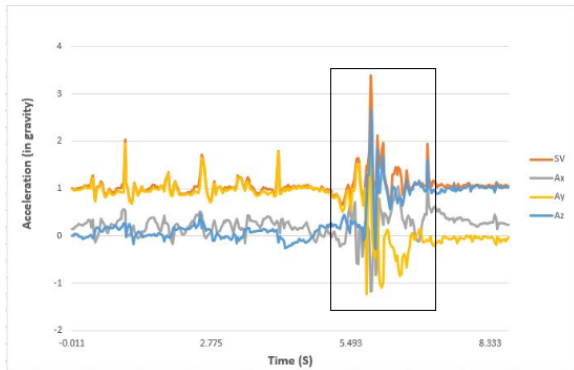


Fig. 1. Sliding window feature engineering based on the maximum signal (r0).

After we estimate the maximum acceleration and decide the sliding window size, the algorithm above gives

us the input features and its labels, which is formed by simply concatenating the six features of accelerometer and gyroscope,  $\{Acc_{xj}, Acc_{yj}, Acc_{zj}, Gyro_{xj}, Gyro_{yj}, Gyro_{zj} | j \in [r_o - \frac{w}{2}f_s, r_o + \frac{w}{2}f_s]\}$ , where  $w$  is the duration of observation window (1s) and  $f_s$  the sampling rate of the sensor (200 Hz) [17]. Finally, the size of input features ( $N_i$ ) which depends of the duration of the observation window ( $w$ ) is  $N_i = 6 * (w * f_s + 1)$ .

After we select the input features to our transformer model, we have to encode the sequence of time which is hidden in our signal data. We implement the existing method Time2Vec [18] for time embedding. This time embedding is a vector representation just like a normal embedding layer that can be added to a neural network architecture to improve a model's performance and to overcome a transformer's temporal indifferences. The mathematical representation of Time2Vec for the ideas of periodic and non-periodic patterns as well as the invariance to time re-scaling are presented by the following mathematical formula [18].

$$t2v(\tau)[i] = \begin{cases} \omega_i \tau + \varphi_i, & \text{if } i = 0. \\ \mathcal{F}(\omega_i \tau + \varphi_i), & \text{if } 1 \leq i \leq k. \end{cases}$$

Fig. 2. Mathematical formula of Time2Vec [18].

where  $t2v(\tau)[i]$  is the  $i$ th element of  $t2v(\tau)$ ,  $\mathcal{F}$  is a sin periodic activation function, and  $\varphi_i$ s and  $\phi_i$ s are learnable parameters of the frequency and the phase-shift of the sine function.

Now we are going to discuss the transformer architecture used for our fall detection method as shown in Figure 4. Since we don't need the decoder layers for such application, only the transformer encoder layers similar to the BERT [4] network, proposed by Devlin, with stacks of multi-head self-attention are used in this work [3]. Unlike the BERT network which supports 2D sequence input, our model can process 3D sequential data- where the dimensions are sequence length, feature size ( $N_i$  calculated above), and batch size. This transformer model with stacks of encoder layers contains time embedding, three attention blocks, and two fully connected layers. The three identical Attention block layers of our transformer encoder have two sublayers. The first is a multi-head self-attention with layer normalization and the second is a feed-forward neural network. In the self-attention encoder layer, the matrix of queries, keys, and values are taken from the outputs of the previous encoder layer. Additionally, between these sublayers, there is a residual connection followed by layer normalization. Next, we get a vector representation output from the transformer encoder model for each input sequence length. Finally, the last layer- Sigmoid layer, gets the vector representation output of the transformer encoder and provides us with a classification- fall or non-fall.

#### IV. EXPERIMENT RESULTS

Paolini (Co-author) conducted human subjects' fall experiments at the neuro-mechanics and neuro-plasticity lab

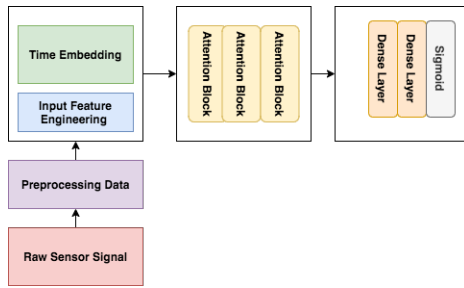


Fig. 3. Our Overall Model.

of San Diego State University. We collected the fall dataset, which contains near-fall activities with a sampling rate of 200 HZ [19] mounted in different parts of our body. The lab was set up with wireless 3D Motion capture cameras which record the human subject movements [20]. The data was collected from thirty subjects between the age of 20-50 years old. This dataset contains 538 thousand rows of data of which 435 thousand of them involve falls and 103 thousand of them consist of near-fall activities. The Noraxon myoMotion research inertial measurement unit (IMU) sensors measure features like orientation angles, joint angles, and linear acceleration of the subject's motion. These data points are preprocessed using median filtering with a window size of 10 to remove the high-frequency noise. The model used 10-fold cross-validation and is divided into training and testing with 80% and 20% respectively.

To evaluate our system accuracy, we calculated specificity, defined as the probability of near-fall activities occurred given that the classifier predicts non-fall activities. We also computed sensitivity, defined as the probability of fall activities occurred given that the classifier predicts fall. It can be seen that our models performed better in specificity and sensitivity than Aicha's CNN-LSTM.

TABLE I  
COMPARING THE RESULTS OF DIFFERENT METHODS.

	Specificity	Sensitivity	Accuracy
Our Methods(Transformer)	0.98	0.97	0.98
CNN-LSTM [14]	0.94	0.93	0.93

The table above shows the performance of our model with the existing method by Aicha [14]. We implemented the existing method [14] to compare the results on our dataset. We took the best candidate model to compare with our model considering the computational time to be used for a real-time fall detection system. It can be observed that, in almost all cases, our model obtained better results by taking advantage of the maximum signal information for windowing. Our model showed highest performance with 0.98 specificity, 0.98 sensitivity, and 0.97 accuracy. From the results, it can be concluded that attention-based network which is mostly used for NLP tasks can be used to detect falls with high specificity and sensitivity than Aicha's CNN+LSTM [14].

We also investigate the advantage of peak signal windowing by implementing the attention-based network with conventional windowing inputs. As we can see in the bar chart below, the transformer network with peak signal windowing gives a better distinction between falls and fall-like activities than the conventional windowing. Especially, the specificity, sensitivity, and F1-score results show that our method gets an advantage from both peak signal windowing and the attention-based network. Furthermore, an attention-based network has less computational complexity than CNN and RNN (LSTM). Such information could be useful for real-time fall prevention techniques in the future.

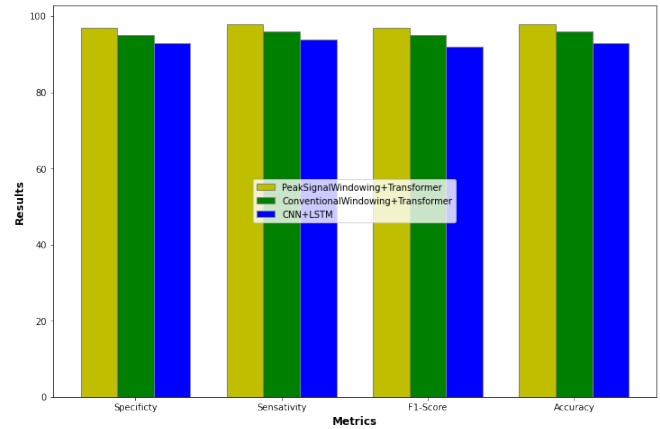


Fig. 4. Comparing the results of our peak signal based windowing-Transformer method, conventional windowing-Transformer Method and CNN-LSTM proposed by Aicha [14].

## V. CONCLUSION

In conclusion, this paper demonstrated the application of attention network for windows tuned to the peak signal to predict fall and near-fall events. To our best of knowledge, this work represents the first attempt to differentiate fall from near-fall activities based on the transformer encoder architecture and outperforms in specificity and sensitivity of previous attempts on wearable sensor signals. Future improvements will require a more sophisticated design of the attention network and test on different public datasets, collected using wearable sensors mounted at different parts of our body. Using different attention networks in different datasets with a few modifications and additional research will be essential to further support the findings to be used for a real-time fall detection system.

## REFERENCES

- [1] WHO, World Health Organization. <http://www.who.int/en/news-room/fact-sheets/detail/falls>, Falls, September 2019.
- [2] Nicole Blair Johnson, Locola D. Hayes, Kathryn Brown, Elizabeth C. Hoo, Kathleen A. Ethier, CDC National Health Report: Leading Causes of Morbidity and Mortality and Associated Behavioral Risk and Protective Factors—United States, 2005–2013
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, Attention Is All You Need, NIPS, 2017

- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL, 2019.
- [5] Jianpeng Cheng, Li Dong, and Mirella Lapata. Long short-term memory-networks for machine reading. EMNLP, 2016.
- [6] E Casilari , R Luque, and MJ Moron, Analysis of Android Device-Based Solutions for Fall Detection. *Sensors (Basel)*. 2015;15(8):17827-17894. Published 2015 Jul 23. doi:10.3390/s150817827
- [7] Montesinos L, Castaldo R, Pecchia L. Wearable Inertial Sensors for Fall Risk Assessment and Prediction in Older Adults: A Systematic Review and Meta- Analysis. *IEEE Trans Neural Syst Rehabil Eng*. 2018; 26(3): 573-582. doi:10.1109/TNSRE.2017.2771383
- [8] Klenk J., Becker C., Lieken F., Nicolai S., Maetzler W., Alt W., Zijlstra W., Hausdorff J., van Lummel R., Chiari L., et al. Comparison of acceleration signals of simulated and real-world backward falls. *Med. Eng. Phys.* 2011;33:368–373. doi: 10.1016/j.medengphy.2010.11.003.
- [9] N.Y. Hammerla, S. Halloran, T. Plotz, Deep, convolutional, and recurrent models for human activity recognition using wearables, *Proceedings of the 25th International Joint Conference on Artificial Intelligence, IJCAI'16*, AAAI Press (2016), pp. 1533-1540
- [10] G. Leoni, P. Endo, K. Monteiro, E. Rocha, I. Silva, T. Lynn, Accelerometer-based human fall detection using convolutional neural network, *Sensors*, 19 (2019), p. 1644
- [11] H. Yhdago, J. Li, S. Morrison, M. Audette, C. Paolini, M. Sarkar, H. Okhravi, "Towards Musculoskeletal Simulation-Aware Fall Injury Mitigation: TL with Deep CNN for Fall Detection," 2019 (SpringSim).
- [12] Ruben Delgado-Escano, Francisco M . Castro, Julian R. Cozar, Manuel J. Marin- Jimenez, Ni colas Guil, Eduardo Casilari, A cross-dataset deep learning-based classifier for people fall detection and identification, *Computer Methods and Programs in Biomedicine*, Volume 184, 2020, 105265, ISSN 0169-2607.
- [13] C. Paolini and D. Soselia and H. Baweja and M. Sarkar. Optimal Location for Fall Detection Edge Inferencing. *IEEE Global Communications Conference (GLOBECOM)*. 2019,8,1-6.
- [14] N Aicha, G. Englebiene, K. van Schooten, M. Pijnappels, and B. Krose, "Deep learning to predict falls in older adults based on daily-life trunk accelerometry," *Sensors*, vol. 18, no. 5, 2018, Art. no. 1654.
- [15] Gongbo Tang and M. Muller and Annette Rios Gonzales and Rico Sennrich, Why Self-Attention? A Targeted Evaluation of Neural Machine Translation Architectures, *Association for Computational Linguistics*, 2018
- [16] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola, "Dive into Deep Learning", <https://d2l.ai>, 2020, Release 0.7.0.
- [17] E Casilari , M Álvarez-Marco , F García-Lagos. A Study of the Use of Gyroscope Measurements in Wearable Fall Detection Systems, *Symmetry*, 2020; 12(4):649. <https://doi.org/10.3390/sym12040649>
- [18] Seyed Mehran Kazemi , Rishab Goel and Sepehr Eghbali and Janahan Ramanan and Jaspreet Sahota and Sanjay Thakur and Stella Wu and Cathal Smyth and Pascal Poupart and Marcus Brubaker, Time2Vec: Learning a Vector Representation of Time, *ICLR* , 2019
- [19] Noraxon 2019. "myoRESEARCH 3.14 User Manual". .
- [20] University, S. D. S. "Neuromechanics and Neuroplasticity Laboratory- ENS 216". <https://ens.sdsu.edu/dpt/research/faculty-research-interests/neuromechanics-and-neuroplasticity-lab/vol>. accessed: Apr. 29, 2020.