ECTI-CON 2023

INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING/ELECTRONICS, COMPUTER, TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY.

MAY 9 - 12, 2023

NAKHONPHANOM, THAILAND

Session: Signal, Image & Video Processing I

Room: Grand Ball Room
Session Chair:

14:45-15:00	CT	**1570871346** Rithea Sum, Chanon Khongprasongsiri, Watcharapan Suwansantisuk and Pinit Kumhom	Low Latency PDM-to-PCM Decoder
15:00-15:15	CT	**1570871363** Teerapon Chongpiphattanasiri, Pramod Wickramatilake, Jednipat Moonrinta, Mongkol Ekpanyapong and Matthew N. Dailey	Distilled Neural Network for Clothes Type Identification
15:15-15:30	CT	**1570875867** Chissanupong Jiamsuchon, Jakapan Suaboot and Norrathep Rattanavipanon	From Characters to Chaos: On the Feasibility of Attacking Thai OCR with Adversarial Examples
15:30-15:45	CT	**1570875972** Tomorn Soontornnapar and Tuchsanai Ploysuwan	Fall Detection Approach Using Variational Autoencoders with Self-Attention Features
15:45-16:00	CT	**1570876253** Antika Ditpattnaphan, Wichit Chaisuwan, Pongpat Vorasayan and Vera Sa-ing	Facial Stroke Classification from Face Features by using Machine Learning

Fall Detection Approach Using Variational Autoencoders with Self-Attention Features

Tomorn Soontornnapar[1] *and Tuchsanai Ploysuwan*[2] **CT**

[1]Department of Electrical Engineering, Siam University, Bangkok, Thailand
[2]Department of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

In this paper, we propose an alternative method for fall detection using variational autoencoders (VAEs) with an attention mechanism on an existing dataset. The dataset consists of 6 different fall cases from 21 people. For effective fall detection, we introduce the use of the magnitude of the acceleration vector (MAV) of wearable gyroscope data and apply fast-Fourier transform (FFT) to create new features. These FFT features are then passed through attention modules with self-combination to form attention features. Our experimental results show that the VAE with self-attention features achieved an average accuracy of 90.7% and an F1 score of 93.8% in fall detection, demonstrating the effectiveness of the proposed method in utilizing gyroscope sensors for fall detection in the context of threshold criteria.

Keywords: fall detection, wearable gyroscope, attention mechanism, variational autoencoder, fast-Fourier transform

Fall Detection Approach Using Variational Autoencoders with Self-Attention Features

Tomorn Soontornnapar
Department of Electrical Engineering
Siam University
Bangkok, Thailand
tomorn.soo@siam.edu

Tuchsanai Ploysuwan
Department of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
tuchsanai@it.kmitl.ac.th

***Abstract*— In this paper, we propose an alternative method for fall detection using variational autoencoders (VAEs) with an attention mechanism on an existing dataset. The dataset consists of 6 different fall cases from 21 people. For effective fall detection, we introduce the use of the magnitude of the acceleration vector (MAV) of wearable gyroscope data and apply fast-Fourier transform (FFT) to create new features. These FFT features are then passed through attention modules with self-combination to form attention features. Our experimental results show that the VAE with self-attention features achieved an average accuracy of 90.7% and an F1 score of 93.8% in fall detection, demonstrating the effectiveness of the proposed method in utilizing gyroscope sensors for fall detection in the context of threshold criteria.**

Keywords*— *fall detection, wearable gyroscope, attention mechanism, variational autoencoder, fast-Fourier transform

I. Introduction

According to the World Health Organization (WHO), the global aging population is increasing, with an estimated 1 in 6 people being 60 years or older by 2030. As people age, they are at a higher risk of falling and experiencing injuries. In fact, falls are a significant public health issue worldwide, with an estimated 684,000 fatal falls occurring annually [1]. This makes falls the second leading cause of unintentional injury death, following road traffic injuries.

The financial impact of fall-related injuries can be significant. In the Republic of Finland and Australia, the average health system cost per fall injury for people aged 65 years or older is $3611 and $1049, respectively [1]. Research from Canada suggests that implementing effective prevention strategies that result in a 20% reduction in the incidence of falls among children under 10 years of age could generate net savings of over $120 million annually [2].

Elderly individuals are hospitalized five times more frequently due to falls compared to other types of injuries. This makes falls the leading cause of accidental death for individuals over the age of 65. As a result, it is important for caregivers to be notified promptly in the event of an elderly person falling.

There is currently a significant interest in the development of intelligent fall detection systems due to the advancement of Internet of Things (IoT) technology, which allows for the creation of small, cost-effective devices that can utilize machine learning techniques to detect falls [3].

There are numerous research techniques for detecting falls using wearable device sensors. These approaches have gained popularity due to the personal nature of wearable devices. Traditional fall detection methods using wearable devices often rely on accelerometer signals [4]. Previous research has proposed threshold-based fall detection approaches using accelerometers [5].

There are several fall detection approaches that rely on analyzing the peak acceleration signal caused by a fall [6]. Other approaches use supervised learning to train the peak acceleration signal [7]. These approaches often utilize binary classifiers such as support vector machines (SVMs) [8], k-nearest neighbors [9], and feed forward neural networks (NNs) [10] for detecting fall signals.

While supervised learning can be used to optimize fall patterns and design an elaborate fall detection system, it can be difficult to obtain enough real fall data for training and to learn various fall patterns that occur under different circumstances such as slipping, tripping, or dropping in various directions [11]. As an alternative, the authors propose using FFT with unsupervised learning to extract significant components from fall signals and successfully detect falls because the authors found that the fall signal in the frequency domain was unique in each fall gesture significantly. Additionally, the use of variational autoencoders (VAEs) with self-attention features is suggested to improve the performance of the approach.

In this study, we describe an alternative method for fall detection using an existing dataset [11]. This dataset includes fall and non-fall cases from 21 individuals who wore a wrist-mounted gyroscope and acceleration sensor.

In the following sections, we will introduce the concepts of variational autoencoders (VAEs) and attention mechanisms. We will then describe the fall detection methods and algorithms used in this study, present the test results, evaluate the performance of the proposed model, and provide a final summary and discussion.

II. Variational Autoencoders and Attention Mechanisms

In this section, we will review the fundamental principles of variational autoencoders, attention mechanisms, and autoencoder-based anomaly detection. We will discuss the key concepts and how they can be applied in these areas.

A. Variational Autoencoders

The variational autoencoder [12] is grouped into a deep generative model that encodes input vectors x to be latent variables mean μ and variance σ^2 to create the latent code z. The VAEs use the conventional autoencoder to encode x with a prior distribution $p_\theta(z)$, usually a standard normal distribution, $Normal(0,I)$. The posterior $p_\theta(z \mid x)$ is difficult to create a continuous latent space z, so the variational inference technique is used to find a deterministic

979-8-3503-1046-7/23/$31.00 ©2023 IEEE

2023 20th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) | 979-8-3503-1046-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ECTI-CON58255.2023.10153189

approximation, $q_\varphi(z\mid x)$, of the intractable true posterior. The parameters of the approximate posterior $q_\varphi(z\mid x)$, often called the variational parameters, are derived using neural networks (e.g., mean μ_z and variance σ_z^2, in the case of a normal distribution). Hence, the training objective of the VAEs is to maximize the evidence lower bound (ELBO) on the training data loglikelihood. For a data point x, the evidence lower bound is given by the following equation, where θ and φ are the encoder and decoder parameters, respectively.

$$L_{\text{ELBO}}(\theta,\varphi;x) = \mathbb{E}_{q_\varphi(z|x)}[\log p_\theta(x\mid z)] - D_{\text{KL}}(q_\varphi(z\mid x)\,\|\,p_\theta(z)) \quad (1)$$

The expectation in the equation above can be approximated by Monte Carlo integration. The second term represents the Kullback-Leibler divergence (D_{KL}) between the approximate posterior and the prior. The distribution for the likelihood is usually a multivariate Normal or Bernoulli, depending on the type of data being continuous or binary, respectively.

B. Attention Mechanisms

Attention mechanisms are a type of neural network architecture that are used to focus on specific parts of an input when processing it. They are commonly used in natural language processing tasks, such as machine translation or text summarization, but have also been applied to other domains such as image recognition.

The self-attention mechanism consists of three main components: queries, Q, keys, K, and values, V. The query in the attention mechanism is like the decoder output, while the values are like the encoded inputs. In the Bahdanau attention mechanism, the keys and values are the same vector.

The self-attention mechanism computes a score value, e by matching each query vector, q against a database of keys using the dot product of the query and each key vector, k_i:

$$e_{q,k_i} = q \bullet k_i \quad (2)$$

The scores are passed through a softmax operation to generate the weights:

$$\alpha_{q,k_i} = \text{softmax}\left(e_{q,k_i}\right) \quad (3)$$

The self-attention is calculated by taking a weighted sum of the value vectors, where each value vector is paired with its corresponding key:

$$\text{attention}\left(q,K,V\right) = \sum_i \alpha_{q,k_i} \times v_{k_i} \quad (4)$$

C. Autoencoder-based Anomaly Detection

Autoencoders are trained to reconstruct data that exhibits abnormal patterns by minimizing a loss function that measures the quality of the reconstructions. Once trained, the model is able to reconstruct data with abnormal patterns accurately, but it may struggle to reconstruct normal data because it was not exposed to it during training. Anomaly detection is performed by using reconstruction metrics, such as the reconstruction error, as an anomaly score. In other words, the model learns to recognize the abnormal data manifold, and the distance between a given observation and this abnormal data manifold is used to calculate anomaly scores, either in the latent space of representations, z, or in the reconstructions space, x'.

III. Fall Detection Methodology

To detect falls using VAEs, we begin by collecting data on both falls and non-falls and preparing the dataset. Next, we select and pre-process the features of the data. After that, we create and train the VAE model to optimize the reconstruction and probabilistic loss. Finally, we test the model by sliding windows of the input features to determine the threshold value for detecting falls.

A. Dataset preparation

The study utilized fall data from a particular source, which was cited as [11]. The data used in the study included two main cases, namely fall and non-fall. However, for the purposes of the research being conducted, only the fall data was selected and analyzed. This fall data was further classified into six distinct cases, each representing a specific type of fall. These six cases were Clockwise forward fall (case01), Clockwise backward fall (case02), Right to left lateral fall (case03), Counterclockwise forward fall (case04), Counterclockwise backward fall (case05), and Left to right lateral fall (case06). The data was divided into three phases: pre-fall, during fall, and post-fall, as shown in Fig. 1. Only the quaternion of acceleration signals, including the three parameters of x, y, and z, were selected because they are able to uniquely describe any three-dimensional rotation around an arbitrary axis and do not suffer from gimbal lock.

B. Features selection and pre-processing

After obtaining the fall dataset, we found that the quaternion xyz within the range of falling has time duration of 200 ms. Therefore, the size of the sliding window for the tested data was set to 200 points. Specifically, we discovered that the rising of the quaternion during the falling period can be used to detect falls using only 50 points. Therefore, we proposed using the rising fall of 50 points as input to the proposed model. To make it more effective and accurate in detecting falls, the magnitude of the acceleration vector (MAV) of the quaternion xyz is utilized and its value is subtracted from 1, as demonstrated in equation 5.

$$MAV = 1 - \sqrt{x^2 + y^2 + z^2} \quad (5)$$

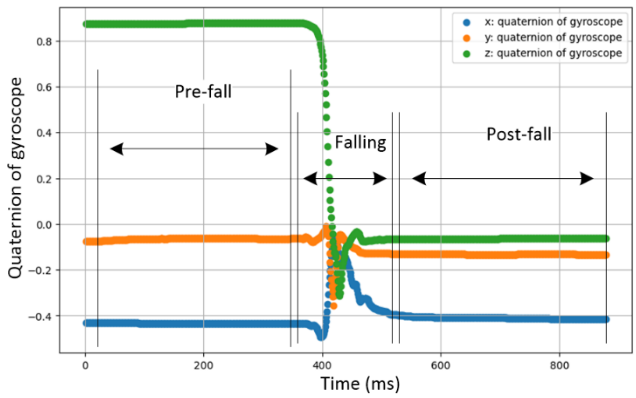

Fig. 1. Quaternion of acceleration signals of falling case01 person#14

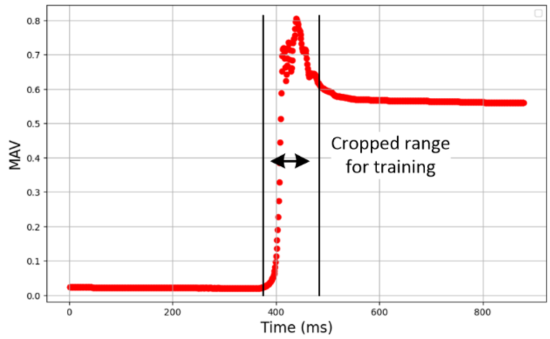

Fig. 2. MAV of quaternion xyz of falling case01 person#14

As seen in Fig. 2, the MAV of the quaternion appears to have an increase in the front-wave, with an average of all cases in the range of 50 points. Therefore, we crop it to 50 points for training. Next, we apply a FFT to a 50-point MAV and reduce it to 25 points in the FFT domain due to the similarity of the bands.

C. VAE Model with self-attention features

The goal of fall detection is to identify unusual behavior and determine whether the detected signal is abnormal or not. One approach that can be effective with a limited amount of data is using representative learning methods, such as the autoencoder.

The proposed VAE with self-attention features is shown in Fig. 3. The system takes in quaternion xyz as input and passes it through the MAV, FFT, and attention modules. The inputs fed to the neural network encoder are the product of the attention and FFT, referred to as attention features. The encoder of the VAE uses a neural network to encode the input into statistical latent matrix parameters, mean and variance, to create the latent z. The z is then decoded by the decoder neural network to reconstruct the quaternion xyz'. The mean and variance matrix is optimized using gradient backpropagation with the Adam optimization algorithm and the loss function is mean squared error (MSE).

D. Training and Testing

The reference input that we use to train the VAE model is manually selected in each case, totaling 6 references. The attention mechanism is used to create attention features that improve the input FFT and enhance the performance of the model. Fig. 4 shows the attention features of the input FFT and the attention maps. Fig. 4a-c and g-i depict the attention features used as references in training, while Fig. 4d-f and j-l display the attention maps for each reference case.

The attention features for each case have a distinct behavior when transformed from the time to the frequency domain, which is the key aspect of our proposed model. For testing, we use a sliding window of size 1x25, which refers to 25 points of FFT input data, and feed it to the trained model for testing.

IV. RESULTS AND DISCUSSION

To evaluate the performance of the proposed model, a total of 104 falling waveforms were tested. The criterion used for this testing was the threshold of the total loss from the VAE model. The training loss was below 1.0 and we set

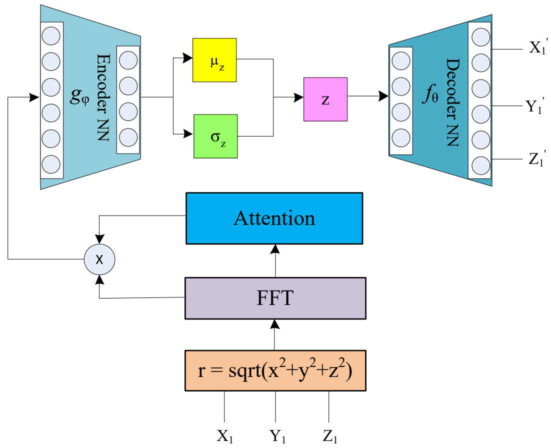

Fig. 3. Proposed VAE with self-attention features framework

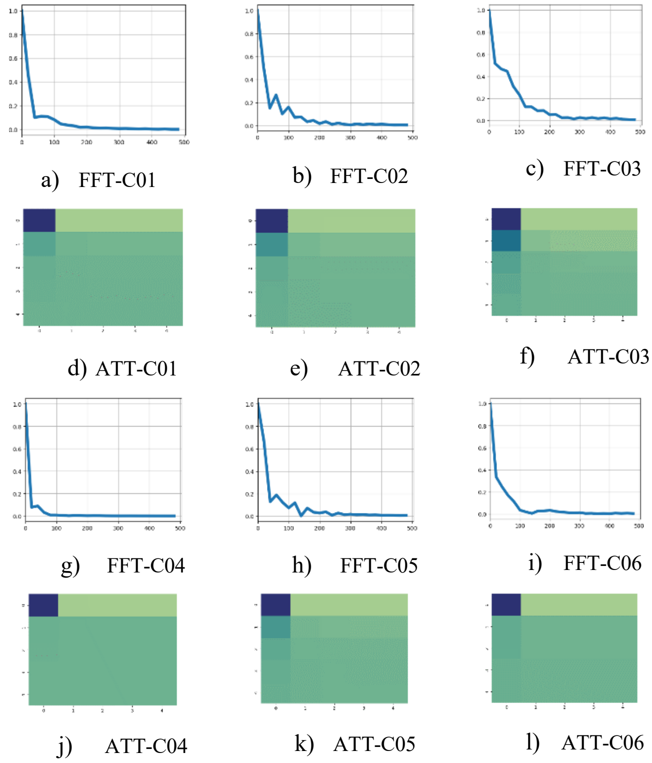

Fig. 4. Attention features case01-03: a) – c), case04-06: g) – i) and Attention maps for reference inputs case01-03: d) –f) , case04-06: j) –l)

the threshold for successful fall detection to be below 1.8 for testing loss. The results are shown in Table I. The proposed model has an average accuracy of 90.7% and an F1 Score of 93.8%. Its performance is quite good and it can accurately detect falls at a rate of 90%.

In our preliminary study, we discovered a potential method for detecting real-time falls when they have unique frequency components in a cohesive manner, as seen in the mean and variance matrices in Fig. 5. The density of the mean and variance matrices for 6 cases is shown in Figure 6. If we choose a model that encompasses the mean and variance for all cases, we may be able to use only one reference to detect falls. This could be a future research direction for finding a universal reference for fall detection using VAEs.

TABLE I. PERFORMANCE TESTING OF THE PROPOSED MODEL

Fall Cases	Model Performance			
	Accuracy	*Precision*	*Recall*	*F1 Score*
Case01	0.9545	0.8571	0.9474	0.9000
Case02	0.9524	0.9524	1.0000	0.9756
Case03	0.9048	0.9048	1.0000	0.9500
Case04	0.8333	0.8333	1.0000	0.9091
Case05	0.9231	0.9231	1.0000	0.9600
Case06	0.8750	0.8750	1.0000	0.9333

With more data from the fall data, we would be able to more accurately find a model that covers most fall cases. As shown in Fig. 5, which corresponds to Fig. 6, the mean and variance are consistent when considering the proposed process. It was found that the mean was in the range (-0.001, 0.001), while the log variance was in the range (-0.002, 0.000).

In terms of the study results, incorporating attention mechanisms into the FFT signal improved the accuracy, with a focus on the frequency profile values. It is anticipated that in the future, this attention will be added to the hidden layers outside of the encoder, which may further improve the model's performance.

From Fig. 5, it can be seen that classification can be achieved through the use of cluster-based machine learning techniques. This can be beneficial for caregivers, as it allows them to quickly and accurately identify fall and non-fall cases in real-time and take appropriate actions accordingly. By using cluster-based machine learning, the caregiver is able to leverage the data presented in Fig. 5 to more effectively distinguish between fall and non-fall cases, enabling them to provide more timely and appropriate care to individuals in need.

One advantage of the proposed approach is that it requires a minimal amount of time to interpret fall situations, as it does not require a large dataset for training. Furthermore, the accuracy rate could be improved by using more sensors attached in proper locations, which would make it scalable for use with patients or elderly individuals. Despite this limitation, the proposed approach is still useful due to its quick interpretation time and potential for use with cost-effective sensors.

V. CONCLUSION

The proposed model for detecting falls has demonstrated high accuracy, with a rate of 90.7%. We made the detection of falls more effective and accurate by using a technique called MAV and subtracting 1 to produce a rising signal. We then employed FFT to identify the unique characteristics of fall signals, which is a key aspect of our model. In an effort to improve the overall performance of the system, we also incorporated VAEs with attention features. The mean and variance obtained from the trained models are not only grouped and cohesive but also serve as a solid foundation for further research on classification. In addition, the models can be used to develop alarms for caregivers, who can rely on them to make informed decisions.

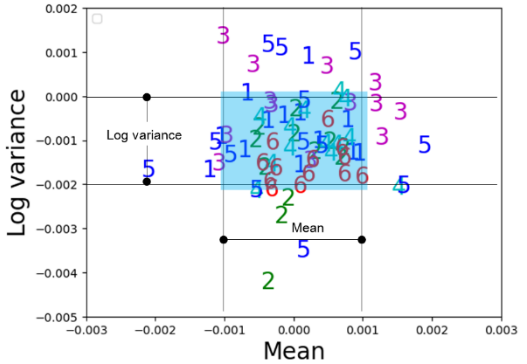

Fig. 5. Mean and log variance of six fall cases, with the digit number indicating the label for each case

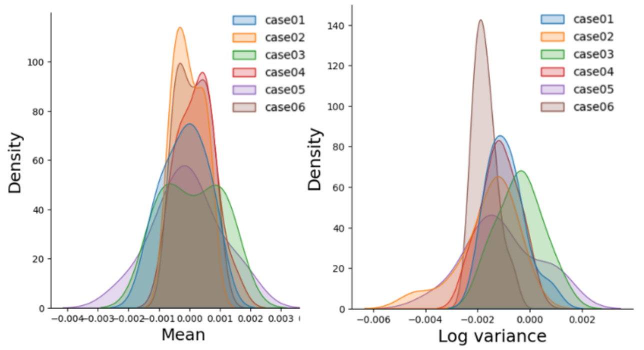

Fig. 6. Mean and log variance density of six fall cases

REFERENCES

[1] World Health Organization., 2019. Falls. Retrieved from https://www.who.int/news-room/fact-sheets/detail/falls

[2] Public Health Agency of Canada., 2017. The economic burden of falls in Canada. Retrieved from https://www.canada.ca/en/public-health/services/publications/healthy-living/economic-burden-falls-canada.html

[3] Zhang, Y., He, Y., & Sun, Y., "Internet of things for fall detection," A survey. IEEE Access, vol. 7, pp. 607-622, 2019

[4] Raza, S., & Mehmood, K., "A review on fall detection systems using wearable sensors," Journal of Medical Systems, 41(12), 181, 2017

[5] Bhatia, K., & Goel, P., "A review of fall detection techniques using wearable sensors," Current Science, 113(3), pp. 437-447, 2017

[6] Chen, C., & Chen, Y., "A review of fall detection using wearable devices," Journal of Medical Systems, 41(11), 169, 2017

[7] Ayesha, Butt, et al., "Fall Detection Algorithm Based on Inertial Sensor and Hierarchical Decision," Sensors, 23, 107, 2022

[8] J. Alizadeh, M. Bogdan, Joseph C., and Christopher F., "Support Vector Machine Classifiers Show High Generalizability in Automatic Fall Detection in Older Adults," Sensors, 21, 7166, 2021

[9] C. Ding, Z. Ding, L. Wang and Y. Jia, "A Fall Detection Method Based on K-Nearest Neighbor Algorithm with MIMO Millimeter-Wave Radar," 2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP), pp. 624-628, 2021

[10] G. Wang, Z. Liu and Q. Li, "Fall Detection with Neural Networks," 2019 IEEE International Flexible Electronics Technology Conference (IFETC), pp. 1-7, 2019

[11] Y. -H. Nho, J. G. Lim and D. -S. Kwon, "Cluster-Analysis-Based User-Adaptive Fall Detection Using Fusion of Heart Rate Sensor and Accelerometer in a Wearable Device," in IEEE Access, vol. 8, pp. 40389-40401, 2020

[12] J. Pereira and M. Silveira, "Unsupervised Anomaly Detection in Energy Time Series Data Using Variational Recurrent Autoencoders with Attention," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1275-1282, 2018