



**HAL**  
open science

# CNN-Based Self-Attention Weight Extraction for Fall Event Prediction Using Balance Test Score

Youness El Marhraoui, Stéphane Bouilland, Mehdi Boukallel, Margarita Anastassova, Mehdi Ammi

► **To cite this version:**

Youness El Marhraoui, Stéphane Bouilland, Mehdi Boukallel, Margarita Anastassova, Mehdi Ammi. CNN-Based Self-Attention Weight Extraction for Fall Event Prediction Using Balance Test Score. *Sensors*, 2023, 23 (22), pp.9194. 10.3390/s23229194 . hal-04586898

**HAL Id: hal-04586898**

**<https://hal.science/hal-04586898>**

Submitted on 24 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Article

# CNN-Based Self-Attention Weight Extraction for Fall Event Prediction Using Balance Test Score

Youness El Marhraoui <sup>1,2,\*</sup>, Stéphane Bouillard <sup>3</sup>, Mehdi Boukallel <sup>4</sup>, Margarita Anastassova <sup>4</sup> and Mehdi Ammi <sup>1</sup><sup>1</sup> CLI Department, University of Paris 8, 93200 Saint-Denis, France; mehdi.ammi@univ-paris8.fr<sup>2</sup> Laboratoire Analyse, Géométrie et Applications, University of Sorbonne Paris Nord, 93430 Villetaneuse, France<sup>3</sup> Fondation Hopale, 62608 Berck, France; stephane.bouillard@fondation-hopale.org<sup>4</sup> Laboratory for Integration of Systems and Technology, CEA, 91120 Palaiseau, France; mehdi.boukallel@cea.fr (M.B.); margarita.anastassova@cea.fr (M.A.)

\* Correspondence: youness.el-marhraoui@etud.univ-paris8.fr

**Abstract:** Injury, hospitalization, and even death are common consequences of falling for elderly people. Therefore, early and robust identification of people at risk of recurrent falling is crucial from a preventive point of view. This study aims to evaluate the effectiveness of an interpretable semi-supervised approach in identifying individuals at risk of falls by using the data provided by ankle-mounted IMU sensors. Our method benefits from the cause–effect link between a fall event and balance ability to pinpoint the moments with the highest fall probability. This framework also has the advantage of training on unlabeled data, and one can exploit its interpretation capacities to detect the target while only using patient metadata, especially those in relation to balance characteristics. This study shows that a visual-based self-attention model is able to infer the relationship between a fall event and loss of balance by attributing high values of weight to moments where the vertical acceleration component of the IMU sensors exceeds 5 m/s<sup>2</sup> during an especially short period. This semi-supervised approach uses interpretable features to highlight the moments of the recording that may explain the score of balance, thus revealing the moments with the highest risk of falling. Our model allows for the detection of 71% of the possible falling risk events in a window of 1 s (500 ms before and after the target) when compared with threshold-based approaches. This type of framework plays a paramount role in reducing the costs of annotation in the case of fall prevention when using wearable devices. Overall, this adaptive tool can provide valuable data to healthcare professionals, and it can assist them in enhancing fall prevention efforts on a larger scale with lower costs.

**Keywords:** fall risk detection; wearables; data-driven deep learning; interpretable artificial intelligence



**Citation:** El Marhraoui, Y.; Bouillard, S.; Boukallel, M.; Anastassova, M.; Ammi, M. CNN-Based Self-Attention Weight Extraction for Fall Event Prediction Using Balance Test Score. *Sensors* **2023**, *23*, 9194. <https://doi.org/10.3390/s23229194>

Academic Editor: Francesco E. Pontieri

Received: 11 August 2023

Revised: 16 October 2023

Accepted: 9 November 2023

Published: 15 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Fall event prediction is a task that aims to determine how likely it is that a person might experience a fall before a fall takes place. A variety of factors can influence fall occurrence [1,2], including balance and gait disorders, cognitive impairment, visual impairment, muscle weakness, and medication use [3]. In order to assist healthcare professionals in identifying individuals who are at risk of falling, automated fall detection tools have been developed to meet this need. These tools typically involve a combination of medical history, physical examination, and an assessment of functional abilities [4]. One of these fall event assessment tools is the Timed Up and Go (TUG) test [5]. The TUG test measures the time it takes an individual to perform a series of usual tasks that require balance and gait stability, and these tasks are to stand up from a chair, walk a short distance, turn around, walk back to the chair, and sit down again. Another tool, called the Berg balance scale [6], is used to assess an individual's ability to maintain their balance during various exercises. These tools can help identify individuals who may be at increased risk of falling.

However, some of these tools are limited by their broad-based assessment approach, which overlooks brief, fast, unstable, and irregular conduct; in addition, their accuracy in predicting fall events is not always high [7,8]. These tests may also be insufficient in terms of the number of factors that needs to be taken into account to detect this type of complex event. Furthermore, balance tests may lack the sensitivity and specificity needed to accurately identify individuals at high risk of falling. False positives and false negatives can occur, thereby leading to either unnecessary interventions for some or failure to identify fall events in others. In terms of monitoring, these tests are easily outdated, since they provide a snapshot of an individual's balance abilities at a specific point in time. Therefore, the continuous monitoring of balance and fall risk over time may prove difficult under these conditions.

With the recent advances in artificial intelligence (AI), machine learning and deep learning models play a role in the use of novel methods for conducting fall event detection on a variety of types of data [9]. For instance, AI models were used with the data provided by wearable sensors [10], such as Inertial Measurement Units (IMUs) and instrumented insoles. These models can be trained to predict an individual's risk of falling based on their gait patterns, and these are combined with the data on medication use, environmental hazards, and other such risk factors to provide a more accurate prediction of fall events. More specifically, wearable sensors can collect movement pattern-related data [11], and they can then detect, thanks to machine learning algorithms, the subtle changes that may indicate increased risk of falling. This can allow healthcare professionals to set up a personalized preventive rehabilitation protocol, thus preventing injury and improving the quality of life for adults and older adults.

In healthcare, there is a growing need for explainable AI (XAI). Indeed, explaining the predictions made by AI algorithms is mandatory for healthcare professionals and patients in order to check medical hypotheses and conduct assessments based on AI-powered clinical decision support systems [12]. This is particularly important in fall event prediction, where healthcare professionals need to understand the factors that contribute to an individual's risk of falling in order to develop effective prevention strategies [13].

This study is motivated by numerous purposes, including prevention and personalized monitoring. As a matter of fact, fall event prediction has several important implications for both healthcare providers and patients. Indeed, it helps healthcare providers identify patients who are at increased risk of falling. Once this is identified, appropriate preventive measures can be implemented. This may include education on fall prevention strategies, environmental modifications, and lifestyle changes. Through this project, we also aim to leverage the progress of AI to enhance the rehabilitation process, thus guiding rehabilitation efforts by helping therapists and clinicians target specific areas of concern and design rehabilitation programs tailored to patient needs. On another note, falls can sometimes be early indicators of underlying medical conditions, such as neurological disorders or cardiovascular problems. Fall event detection based on motion data can prompt further evaluation and diagnosis [14], potentially leading to earlier treatment and improved outcomes. These measures are the building blocks of this project, and they are used to conceive a proactive and patient-centered approach to assess the risk of falling and minimize their associated consequences.

In this study, we present a data-driven approach to detect moments with the highest probability of falling by using unlabeled ankle-mounted-IMU data. Our model was trained to approach Berg balance scores, which are given metadata for all patients, in order to detect possible fall events. Since this approach relies mainly on data, we adopted the non-contrastive learning approach to allow the model to pre-train a data encoder that extracts the most informative patterns. Accordingly, the contributions of this article are summarized in the following points:

- Performing rare event detection with an interpretable deep learning architecture;
- Using a non-contrastive learning approach to improve the model's representation of data;

- Moments with high risk of falling are highlighted by the intermediate and interpretable layers of the model instead of the output layer.

The current paper is organized as follows: Section 2 presents the related works and use cases that use different approaches to identify possible fall events, and examples that tackle the interpretability aspect. Section 3 details the configuration of the project, as well as the approach that we used for this project. In Section 4, we analyze the results of our approach and highlight the interpretability aspect of this model. Then, Section 5 draws the conclusion about the use of interpretable models in detecting fall events indirectly by using a balance score. Finally, we discuss the importance of these contributions in terms of interpretability, data-based methods, and healthcare.

## 2. Background

Identifying individuals who are at risk of falling and implementing appropriate interventions to reduce fall events can help prevent falls and their associated injuries. Risk factors [1] are to be taken into account when treating this issue. Therefore, many risk assessment tools have been developed and continuously updated for this matter [2]. Fall risk assessment is increasingly needed for older adults, as it helps in identifying the risks that endanger them. This is because advanced age is a significant risk factor for falls due to age-related changes in strength, balance, and mobility [15]. Moreover, reduced muscle strength can impair balance and increase the likelihood of falls [16]. Balance and gait problems are also major causes of falls among older adults [17]. Furthermore, medication, chronic health conditions, and visual impairment may increase the risk of falling [18–20]. One other aspect related to fall occurrence is cognitive impairment. In fact, environment perception can affect judgment due to cognitive decline or dementia, thus increasing the probability of falling [21].

The automation and the reliability of such tools are important factors for health practitioners. By using these frameworks as decision support tools [22], practitioners can develop targeted interventions to prevent fall scenarios. Regular threshold-based algorithms have been designed in order to extract the specific values of the used input signals, and usually, these values are computed based on the streaming signals provided by IMU sensors. Both accelerometers and gyroscopes can be used for this calculus [23]. In other works, magnetometers, quaternions, and G-force sensors have been used for the same purpose [24–26]. These methods have certain limitations in terms of generalizability, performance, and the number of data required for processing [27]. With the growing importance of big data in contemporary experiments, new methods have taken the lead in this type of use cases.

Recently, machine learning techniques have shown promise in healthcare, particularly in the field of gait analysis. Machine learning algorithms can analyze large numbers of data to identify patterns and relationships that may be difficult for humans to detect [28,29]. Common machine learning algorithms are particularly efficient in the detection of fall occurrences. For example, the decision tree model, known for its simplicity and easy interpretation, was able to achieve good scores in the prediction of fall risk with high probability [30]. Furthermore, the Support Vector Machine (SVM) model was able to increase the accuracy of its predictions and could also generalize more effectively for instances of unseen data [31]. Probabilistic models such as the Hidden Markov Model (HMM) are also especially efficient due to their high accuracy in predicting human falls with accelerometer data [32].

In general, many technologies have been deployed in the service of healthcare, especially with respect to gait analysis. For instance, computer vision algorithms are used to analyze video data to assess an individual's gait, posture, and movements. These architectures can identify deviations from normal motion patterns that might indicate fall risk [33]. In the context of healthcare, data such as a patient's history, medication details, demographics, and physical assessments are usually recorded and stored. Hence, they can be combined and processed using machine learning models as relevant features for fall

assessment [34]. Additionally, deep learning models can fuse data from multiple sources, such as wearable sensors, video, and environmental sensors, to provide a comprehensive fall risk assessment [35].

More specifically, in the case of wearable sensors, deep learning architectures have also been used to solve the task of fall detection or assessment. In fact, Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and hybrid ConvLSTM are great assets to extract temporal as well as spatial features from data in order to predict falls for older adults [36]. These types of models are fairly known for their performance in different tasks that include time series. Therefore, they are convenient when processing the data obtained from wearable sensors. Yet, deep learning models show clear limitations when it comes to interpretability due to the complex calculus that they introduce and the depth of their layers [37]. As a result, explainable AI was introduced to address this issue [13]. Attention mechanism models, also known as self-attention (SA) models, utilize an architecture that creates similarity links inside an input sequence in order to assign importance weights to its components [38]. This method has rapidly spread in different domains thanks to its interpretability, especially in healthcare, which is a field with high demand for explanatory conclusions obtained by means of model predictions. Having said this, this type of method has not yet been properly exploited for assessing fall event probabilities.

In practice, two approaches have been proposed to determine the likelihood of experiencing a fall based on the protocol setting. The first one aims to assess if the person should be considered at risk of falling or not thanks to a classification process. On the other hand, other works have proposed time-wise fall event detection that can be applied in real time or at the end of the event [39]. In the first case, studies that used the accelerometer data provided by multiple sensors mounted on different parts of the body (accompanied with pressure insoles in certain cases) showed that accuracy in fall screening ranges from 57% to 90% [40,41]. Other methods, comparable to ours, use deep learning algorithms in order to identify the samples with the highest risk of falling. In such cases, results of around 73% accuracy were achieved [42,43].

Despite the quality and novelty of the utilized approaches that aim to assess the probability of falling among older adults, certain limitations still arise. This topic is often limited by the generalization capacities of AI models, as they usually struggle to maintain a robust prediction level across diverse populations [44]. In addition, the “black-box” nature of certain AI models can make it challenging to understand the reasoning behind fall event predictions, which may be critical for healthcare providers. Finally, AI models may misinterpret non-fall-related activities or environmental factors as falls, thus leading to false alarms. Through this study, we aim to mitigate these types of limitations by building an agnostic model that is generalizable to all participants. Furthermore, we also aim to exploit the interpretable aspect of self-attention architectures to avoid instances where there is no explanation due to lack of results and to—finally—benefit from the model’s interpretability by obtaining a probability for the occurrence of a fall event (instead of a binary classification) in order to adjust a threshold of confidence that could help avoid false fall alarms in falling events.

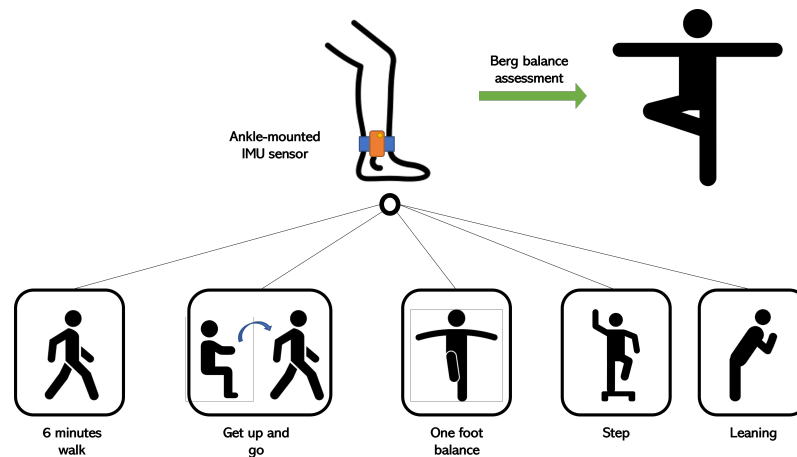
### 3. Method

In this section, we present our framework, which uses IMU sensor data to predict possible fall events. First, we analyze the data that were collected and used for this project. Then, we explore the deep learning architecture, as well as the training pipeline. Finally, we give details about the applications of this fall event assessment tool.

#### 3.1. Experimental Configuration

The adopted setup for these experiments was based on accelerometer and gyroscope data. In order to acquire gait-related data, IMU sensors were mounted on the ankle of each foot. The reason behind this choice was to use the smallest number of IMU sensors as possible in places with the best relevance for gait analysis, such as the ankles. These

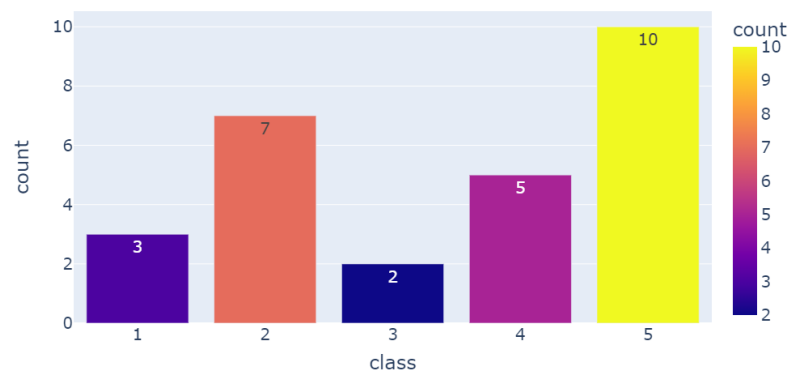
sensors have a sampling rate of 40 Hz and a full scale of  $\pm 160 \text{ m/s}^2$  for the accelerometer component (e.g.,  $\pm 2000 \text{ deg/s}$  for the gyroscope component). In our case, the observed ranges were  $\pm 40 \text{ m/s}^2$  (accelerometer) and  $\pm 260 \text{ deg/s}$  (gyroscope). Each patient performed a series of standardized exercises, including Get Up and Go, One Foot Balance, 6-Minute Walk, etc. Figure 1 shows the overall configuration for acquiring the data based on the sequence of the proposed exercises. The study population involved patients with a variety of parietic severities, and they were grouped into 5 categories of walking difficulty. For this project, the people who were recovering best were the most represented.



**Figure 1.** Activity descriptions.

### 3.2. Data Analysis

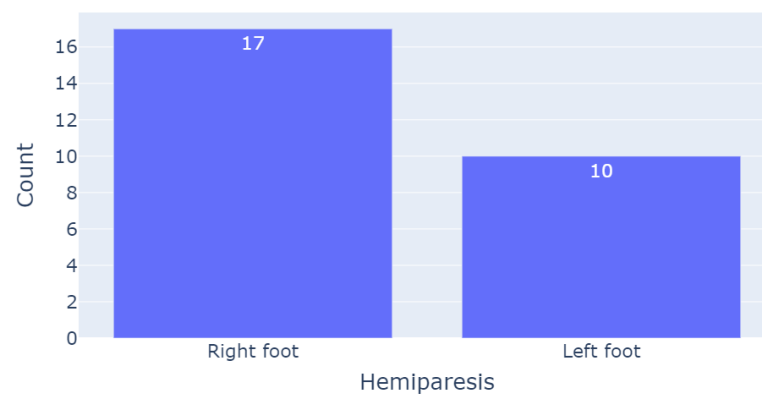
This study included 27 patients from Fondation Hopale in Berck-sur-Mer, France. The patients were stroke survivors undergoing rehabilitation. This dataset included patients with different motor disorders, especially hemiparesis. The sequelae of this pathology differ from one person to another, and a score of the quality of their gaits was given based on medical expertise and manual annotation: patients with severe gait disorders were given a score of 1, whereas healthy participants with a normal gait were given a score of 5. Figure 2 demonstrates the distribution of this population according to their given scores, which were based on a gait evaluation.



**Figure 2.** Number of patients per group of gait quality.

This population included participants aged from 35 years old to 84 years old, with a median age of 64 years old. There were 10 females and 16 males. The patients were being rehabilitated from 1 up to 160 months after stroke occurrence, thus making them quite varied in terms of their response to the rehabilitation process. Among the men, the mean height (in cm) and weight (in kg) were  $177.7 \pm 6.9$  and  $82.7 \pm 13.1$ , respectively. On the other hand, the mean height and weight among women were  $162.9 \pm 6.7$  and  $66.8 \pm 17.9$ ,

respectively. Regarding the medical condition of these participants, Figure 3 illustrates the number of people who were affected in their right/left foot.



**Figure 3.** Number of patients based on the side affected by hemiparesis.

Finally, some of the participants used different tools as means through which to support themselves while walking. Indeed, a support tool that had been advised based on their respective conditions was used to assist them during their rehabilitation sessions, as well as in their daily lives. Table 1 gives a detailed count of people using different types of support tools based on their paretic limb.

**Table 1.** Number of walking support users.

Type of Support	Paretic Side	
	Left Foot	Right Foot
Cane	7	2
Tripod cane	1	1
Molded lifter	1	3
Rubber band lifter	0	2

### 3.3. Data Preprocessing

A certain number of data are mandatory for an AI model when using the streams of data provided by IMUs for rare event detection tasks for it to be able to learn from different examples. Therefore, we clipped the recordings of the different exercises into short segments of 10 s with an overlapping sequence of 1.25 s. The duration of the recordings of each exercise is given by its mean and standard deviation in Table 2. In fact, a 10 s window of data can contain, on average, data corresponding to 6 steps, which are rich in a variety of types of behavior, repetition, and gait features. Overall, this technique helped us to acquire a consistent number of data, as well as to obtain a perfectly sized input for our model.

**Table 2.** Mean duration of the recording of each exercise.

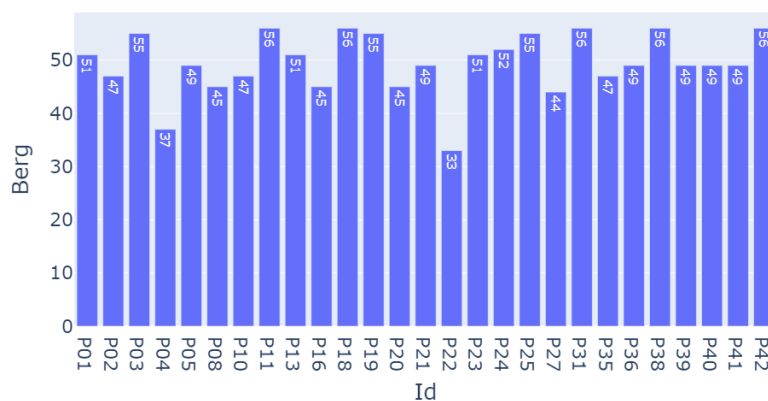
Activity	Duration	
	Mean	Std. Dev.
6-Minute Walk	6 min 31 s	13 s
Get Up and Go	50 s	29 s
One-Foot Balance	53 s	31 s
Step	46 s	27 s
Leaning	1 min 11 s	6 s

With respect to data representation, we propose a 3D data format in order to exploit the spatial correlation between the different axial components of the accelerometer signals

and the gyroscope signals. Therefore, we used 6 different channels for the x-, y-, and z-axes for both acceleration and gyroscope data. For each channel, we reformatted the signal into a  $20 \times 20$ -sized matrix by clipping it into 20 equally sized sequences of length 20. Then, we concatenated them as rows of a new 2D matrix. We next completed the same process for each signal component (Acc\_x, Acc\_y, Acc\_z, Gyr\_x, Gyr\_y, and Gyr\_z). Then, we superimposed the resulting 6 matrices into a single 3D matrix, where each channel corresponded to the 2D matrix of a signal component. At the end, each input observation for the model was of size  $20 \times 20 \times 6$ .

### 3.4. Pre-Training Target Variable

Our model consists of a pre-training phase where the model trains to approach a Berg balance assessment score. This score was the result of an evaluation of a series of equilibrium exercises that were performed by the participants. The values of this score for the participants varied from 33 to 56, as demonstrated in Figure 4.



**Figure 4.** Berg assessment score for each patient.

With respect to this, two possibilities were proposed for creating the training pipeline and defining the task type:

- Regression problem: By keeping the Berg score in its current format, the developed model could have been trained to predict the exact value of this score;
- Classification problem: Transforming the balance score into the three categories of the intervals 30–40, 41–50, and 51–60 could have allowed the model to be trained on a multi-label classification task.

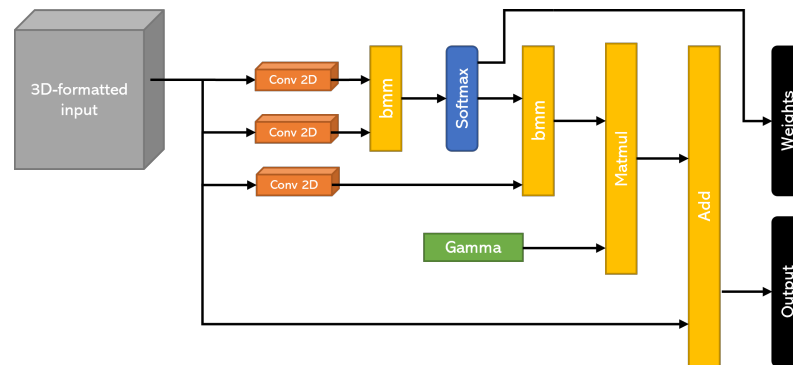
The second configuration was less effective in terms of training progress, and the loss function was especially unstable. In addition, the results of the classification were not sufficient. On another note, the first approach allowed for smoother training while also closing the gap between the model's predictions and the real values of the Berg assessment score. Consequently, the following sections involve the findings from the regression task alone.

### 3.5. Self-Attention Architecture

As detailed previously, this model uses a balance score for training and conceiving a compact and informative representation of data along with corresponding importance weights. In order to extract these weights, we used a spatial self-attention mechanism, which was implemented with convolutional layers. This type of architecture can highlight the correlations among the parts of the signal itself, as well as with other components. In our case, this was the most apt configuration, since we sought to pinpoint the instants of the signal that explained the most of the Berg score. In this way, we hoped to detect the instants with the highest probability of predicting a fall. Figure 5 illustrates the self-attention architecture that was used in the pre-training phase. This architecture consists



of three similar convolution layers, each of which computes key, query, and value. By computing the softmax of the dot product between the key and query, we obtained the similarity patterns between them, which were used as descriptive attention scores. On the other hand, we used the value representation to compute the weighted values of the input, which were multiplied by the attention scores. We used the gamma parameter in order to allow the model to give more or less importance to the self-attention module in creating the feature maps for the overall network. This meant that for high values of gamma, the outputs were more influenced by the self-attention module.



**Figure 5.** The spatial self-attention architecture.

### 3.6. Non-Contrastive Learning Approach

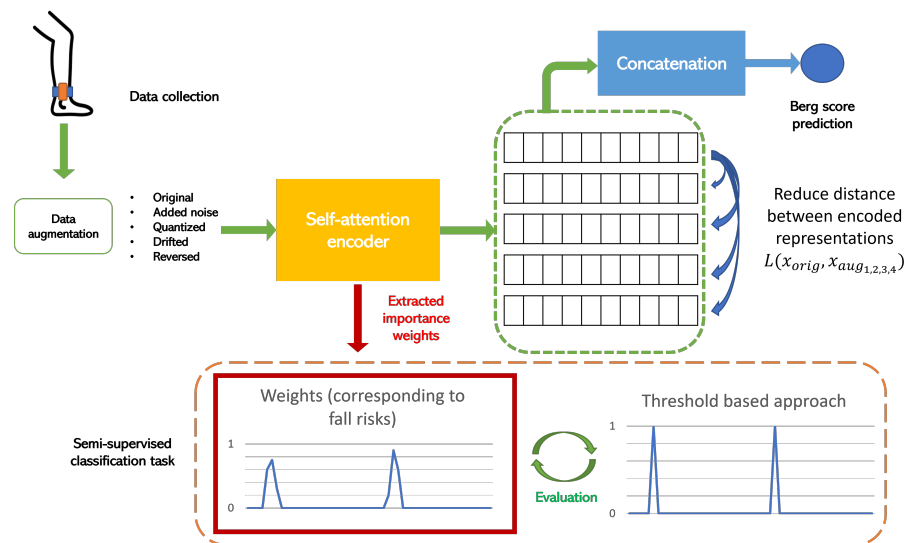
The self-attention mechanism was developed in order to be used for a non-contrastive learning approach. In Figure 6, we illustrate the steps of this approach. These consisted in creating different versions of the input data by using the following types of data augmentation:

- White noise: We added randomized white noise along the different time series, and this noise was simply a random Gaussian signal with a mean that was computed for each signal;
- Random quantization: This approach maps the values of the continuous signal into smaller sets of discrete values with a chosen step;
- Drift: In this case, the algorithm randomly and smoothly drifts the values of a time series from its original values;
- Reverse: This approach simply involves reversing the time series and rendering it backwards.

Therefore, we obtained the original time series and 4 more variants of it thanks to the augmentation techniques. After that, we extracted the data representation that was given by the self-attention encoder for the original input, as well as the four augmented versions. Then, we trained the model to learn the similarities among the rendered representations. Simultaneously, we concatenated the inputs' representations and provided them to a regression layer that predicted the Berg score assessment. In this way, we allowed the model to undergo a two-fold effect when learning data representation, which is meaningful in terms of equilibrium, and extracted the weights of correlation between the input signals and the loss of equilibrium, which is a direct cause of falling for older adults.

This method allows us to detect rare events, which are fall risk moments in our case, with unlabeled data. Furthermore, we only used the Berg score labels in the pre-training phase. By extracting the important sequences from the signal thanks to the self-attention module, we aimed to extract the moments that correlated with the Berg score, such that we could compare them with the phases with the highest risk of falling without any dedicated supervised training for this task. Then, we compared these results with a basic threshold-based approach for detecting the risk of falling based on accelerometer data. The combination of the non-contrastive learning approach and the self-attention mechanism is a novel way with which to solve classification problems in the detection of rare events and to avoid the problems of huge imbalances in data. Moreover, by means of interpretable

correlations, this model is able to create a link between the Berg score and the loss-of-equilibrium moments leading to fall possibilities.



**Figure 6.** The full pipeline of the non-contrastive learning approach.

### 3.7. Training Configuration

Considering the fact that we trained the model on two tasks simultaneously, we used two different losses that were adapted for each case. For the score prediction, we used the simple mean absolute error, which is defined by the following equation:  $l_n = |x_n - y_n|$ . On the other hand, we used contrastive loss, given by  $y \times d + (1 - y) \times \max(\text{margin} - d, 0)$ , where  $d$  is the Euclidean distance between the two samples, demonstrated as follows:  $d = \|x_1 - x_2\|$ . Since our final task was the detection of potential falling moments without the use of annotated data, we had to visualize the weight maps for the samples of certain different configurations in order to have additional material to evaluate the performance of our model. The training phase included different experiments in order to find the most efficient hyperparameters for our models. Therefore, we used a grid search process to compare the performance of each configuration, as well as to extract the most fitting hyperparameters for our use case. On one hand, we varied the values of the learning rate, batch size, and number of epochs for the training process. On the other hand, we chose the kernel sizes in the CNN-based self-attention module to be evaluated. The search space for each of the different hyperparameters is given in Table 3.

**Table 3.** Search space for the model's hyperparameters.

Hyperparameters	Range of Search Space
Batch size	[1, 4, 8, 16, 32]
Learning rate	$[1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}, 1 \times 10^{-5}]$
Number of epochs	[5, 10, 15, 20, 30, 40, 50]
Self-attention kernel size	[1, 2, 4, 6, 8]

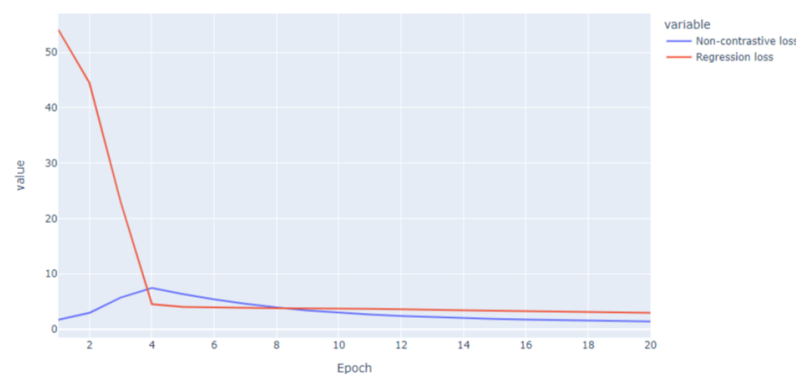
The series of experiments were launched on a DELL laptop with the following characteristics: 11th Gen Intel(R) Core(TM) i9 (2.60 Hz), 32Go on RAM, and NVIDIA RTX A4000 graphics card. We used the "Pytorch" library for the development of deep learning architectures, "sickit-learn" for model evaluation, and the "tsaug" package to generate the augmented data in the pre-training phase. As for the training session duration, depending on the combination of hyperparameters selected by the grid search algorithm, it varied between 17 min, 5 s and 3 h, 41 min.

### 3.8. Partitioning and Evaluation

To guarantee a diversified distribution of samples for both training and evaluation, we split the data per patient and per activity in an 80–20% ratio and then added them to the test and training sets, respectively. Accordingly, the overall size of the partitions was also split in an 80–20% ratio. Usually, 80% of the data are used for the training process. To demonstrate the potential of our method, we used only 20% of the data for the non-contrastive learning training phase; we then evaluated the model on the remaining 80%. The evaluation of the model required a ground-truth reference for comparison. Thus, we used one of the baselines of the fall event assessment tools to train the threshold-based approaches. Indeed, we compared our results with the predictions provided by an algorithm that extracts moments in which a sudden change in acceleration is detected in the vertical direction, which, in our case, was where the y-axis signal was concerned. The threshold of the values of acceleration was fixed at 5 m/s<sup>2</sup>, and this value was approximated at 0.5 g (where g is the gravitational constant and its value is 9.81 m/s<sup>2</sup>). Many works have used this value as a minimum threshold for the beginning of the free-fall phase; ref. [45] used 0.4 g in order to provide an alert during the pre-fall zone so as to avoid injury. Other approaches aim to find sequences where an abrupt change in acceleration is observed [46] (i.e., a variation from 5 m/s<sup>2</sup> to 15 m/s<sup>2</sup> in a short period of time, which—more specifically—is 300 ms).

## 4. Results

We chose our best model based on its loss evolution, its capacity to approach the Berg score, and the extracted weights for the detection of fall events. First, the loss functions (shown in Figure 7) were both stable and decreased smoothly through the epochs. In other cases, they had known fluctuations and unstable evolution in time. On rare occasions, the loss curves, especially the non-contrastive loss, diverged, and the model was not able to extract the pertinent weights. For this chosen model, the fixed parameters are detailed in Table 4.

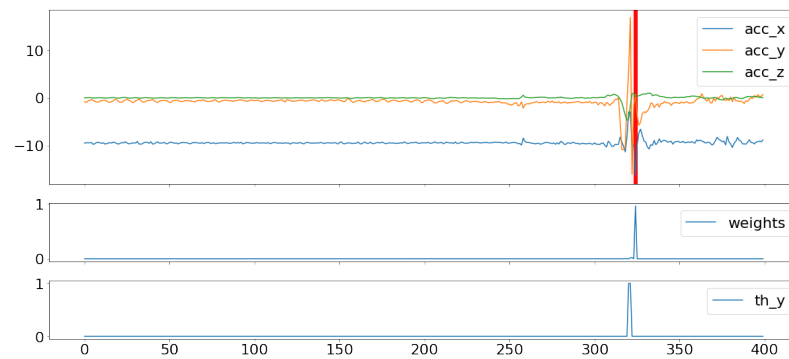


**Figure 7.** Loss functions for the non-contrastive learning and regression tasks.

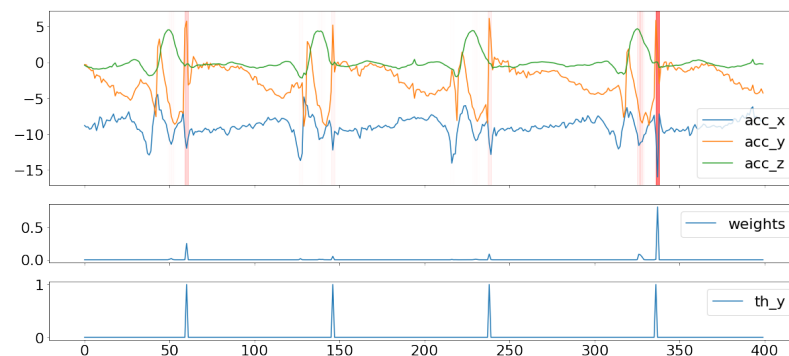
**Table 4.** Optimized hyperparameters.

Hyperparameter	Batch Size	Learning Rate	Epochs	Kernel Size
Optimized value	4	$1 \times 10^{-3}$	20	1

We used the second output of the model, which is the weight of each timestamp of the signal in explaining the Berg score in each trial. Figure 8 shows an example of the projection of these weights on the signal, and we superimposed this figure with the plot of the weights, as well as with the reference labels that indicate the instants where the axial components of the acceleration signal exceeded the value of 5 m/s<sup>2</sup>. In other cases, the model could also track a multitude of risk possibilities, although it was also able to accord more importance to potential falls with the highest risk, as demonstrated in Figure 9.

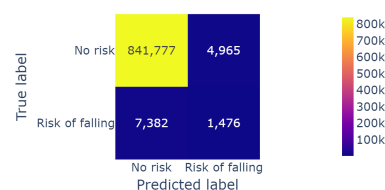


**Figure 8.** Weight projection on the IMU signals for one risk event. (The red line represents the projection of high values of weights).



**Figure 9.** Weight projection on the IMU signals for a multitude of risk events with different importance weights. (The red lines represent the projection of high values of weights).

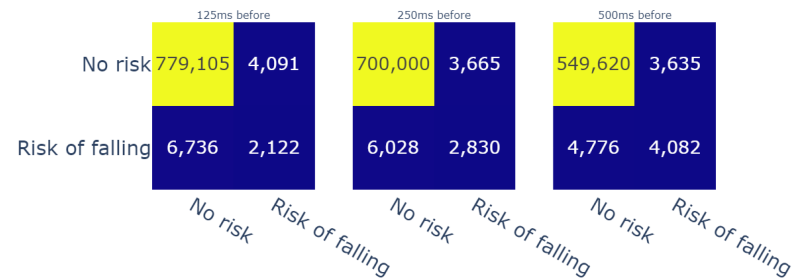
The model was able to perfectly predict 1476 cases of possible falls out of the 8858 cases that were computed by the threshold approach. The overall accuracy of the model reached 98.9% due to the imbalanced classes of the data. The confusion matrix (shown in Figure 10) listed the correctly labeled risk moments, as well as the mislabeled data. Therefore, we focused our analysis on the “fall risk” class.



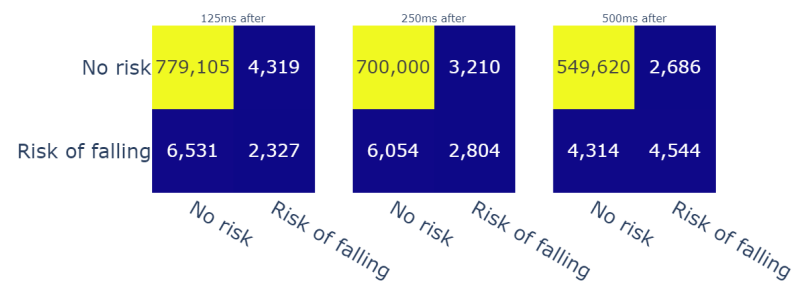
**Figure 10.** Confusion matrix for the fall risk instants.

Due to the fact that our model detects specific instants of fall risk on a temporal basis, our prediction might predict fall moments before or after the true labeled risk instant. In cases where the model’s prediction of risk is completed before the ground-truth prediction, the prediction is beneficial to the assessment tool from a preventive point of view. Moreover, if the model predicts the possibility of falling after its occurrence, certain methods suggest spreading the prediction on a small window, which we usually call a confidence interval. For this reason, we evaluated the True Positives (TPs), True Negatives (TNs), False Positives (FPs), and False Negatives (FNs) at a window level while comparing the different sizes of evaluation windows. We selected three modes of evaluation, and the following are the possible scenarios:

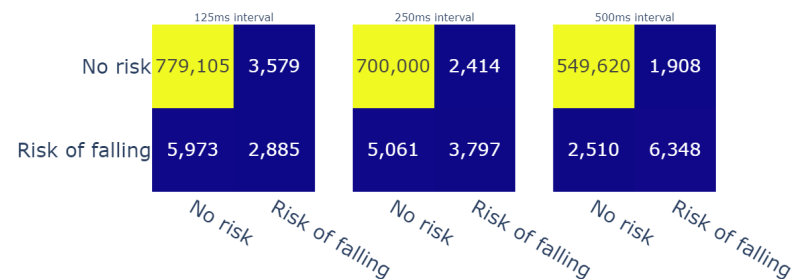
- Predicted value inside a window of time before the reference label (Figure 11);
- Predicted value inside a window of time after the reference label (Figure 12);
- A confidence interval, in which the reference label is in the middle of the window of time (Figure 13).



**Figure 11.** Confusion matrices for the different durations of windows of time (before reference).



**Figure 12.** Confusion matrices for the different durations of windows of time (after reference).



**Figure 13.** Confusion matrices for the different durations of window of time (confidence interval).

For therapists, detecting risk of falling in advance is particularly important from a preventive standpoint. Furthermore, intelligent fall prevention systems usually aim to launch alerts during the transition phase, i.e., between the “Activity of daily living” (ADL) phase and the free-fall zone. This transition phase can last an especially short time span, which varies from 150 ms [47] to 300–500 ms [48] (depending on mobility factors for each patient). Given these durations, we evaluated the predictions of our model with three temporal window sizes, i.e., 125 ms, 250 ms, and 500 ms, before the computed labels. The achieved results showed that the model detected up to 4082 fall risk moments (c.f. Figure 11). On another note, we used the same temporal windows to evaluate the risk events that were detected moments after the reference risk events. In this case, for the same temporal windows, the model detected up to 4544 risk moments (c.f. Figure 12). In addition, we chose to create a confidence interval that included the results before and after the references with respect to the same evaluation window sizes in order to have an overall idea about the performance of the model. In that regard, the model was able to detect 6348 out of 8858 fall (c.f. Figure 13) risk events, which represent 71.6% of the cases. Thanks to this evaluation process, we gained more than 55% in the model’s recall score. In addition, the precision of the model reached 77%. These scores were given for weights that were higher than 0.1 compared with recall of 9% and precision of 98%; however, this applied only if the weights higher than 0.5

were taken into account. This means, in that case, that the model detected fewer fall risk events (around 830 occurrences). On the other hand, it did produce fewer false negatives (only 11). Therefore, finding the best threshold was also mandatory for this study, and fixing it at 0.1 helped to obtain an acceptable ratio of correctness in the model's predictions.

## 5. Conclusions and Perspectives

In this paper, we propose a non-contrastive learning approach to predict, using the Berg balance assessment score, fall risk for both adults and older adults who are stroke survivors. Its advantages can be summarized in the following points:

- It can be trained on a small number of unlabeled data;
- It uses interpretable architectures on metadata to extract meaningful features;
- The model is used to exploit the cause–effect link between the balance score and the target variable.

The results obtained were particularly interesting in terms of the quality of detection. In fact, the model reached a recall score of 71% and a precision score of 77% in cases with a 1 s confidence interval (specifically 500 ms before and after the reference prediction). It is important to keep in mind that the model was not trained to detect these risk events in a direct and supervised way. Instead, it was trained to provide approximations of the Berg score and to use the interpretability assets of the model in order to find the link between the balance score and the risk of falling. This was achieved by extracting the importance weight of each timestamp, and it appeared that the model pinpointed instants with extreme values of acceleration and the sudden evolutions within them. Therefore, the model achieved decent scores, especially with respect to the fact that the model itself extracts the relation between the balance score and the values of acceleration in order to determine fall risk moments without the use of any labeled data for this purpose. Moreover, the threshold detection approach is well grounded and is shown to achieve good scores, yet it may show certain weaknesses in terms of generalizability and adaption to profiles, i.e., the reference labels are not especially reliable. In another vein, this study still has many directions in which there could be possible improvements. The proposed CNN-based self-attention model was only compared with the threshold approach. Other explainable architectures could be tested and adapted for this use case, and the model of this study could be compared with more reliable detection approaches or annotated data. Furthermore, the non-contrastive learning pipeline needs to be trained on large datasets in order to extract different types of patterns and information. Moreover, the contrastive self-supervised learning method could be tested for this problem, specifically in terms of allowing the model to distinguish between normal patterns of gait and abnormal sequences that may lead to falling. From a practical point of view, we observed that the model had a tendency to predict fall risk in the interval that preceded the reference labeling. Thus, our approach could serve as a prevention tool for fall risk assessments. In terms of utility, this project aims to conceive an interpretable tool for decision making to assist health practitioners in their daily tasks, especially for gait analysis and fall risk management in the elderly. Its varied purposes could evolve around profiling patients and suggesting adapted rehabilitation indicators, including the use of gait phases to improve the prevention of patient falls, or the use of a type of walking support and the adequate moment for changing it. Furthermore, this tool is bound to be used for monitoring patients in different configurations aside from clinical rehabilitation sessions. In order to meet these needs, we will work on building a more robust model that allows one to also have an interpretable output based on the inputs. By adding more metadata and information from patient records, the model could gain more efficiency and trustworthiness to reassure tool users. In addition, it could also help to enhance its interpretable outputs with indicators that are more relevant to the medical staff.

**Author Contributions:** Conceptualization, Y.E.M., S.B., M.B., M.A. (Margarita Anastassova) and M.A. (Mehdi Ammi); methodology, Y.E.M., M.B. and M.A. (Margarita Anastassova); validation, M.A. (Mehdi Ammi); formal analysis, Y.E.M.; investigation, S.B.; data curation, S.B., M.B. and M.A. (Margarita Anastassova); writing—original draft preparation, Y.E.M. and S.B.; writing—review and editing, Y.E.M., S.B., M.B., M.A. (Margarita Anastassova) and M.A. (Mehdi Ammi); visualization, Y.E.M.; supervision, M.A. (Mehdi Ammi); project administration, M.A. (Mehdi Ammi). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the guidelines in the Declaration of Helsinki and was approved by the Ethics Committee of COMITE DE PROTECTION DES PERSONNES Sud Est V (IRB: 6705, 22 March 2018).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data available on request.

**Acknowledgments:** We would like to thank the medical personnel at Fondation Hopale who helped with data curation and preparation for this project.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Pfortmueller, C.; Lindner, A.G.; Exadaktylos, A.K. Reducing fall risk in the elderly: Risk factors and fall prevention, a systematic review. *Minerva Med.* **2014**, *105*, 275–281.
- Natalie, C. Falls prevention: Identification of predictive fall risk factors. *Appl. Nurs. Res.* **2016**, *29*, 53–58.
- Centre for Clinical Practice at NICE (UK). Falls: Assessment and Prevention of Falls in Older People. 2013. Available online: <https://www.nice.org.uk/guidance/cg161> (accessed on 5 June 2023).
- Stapleton, C.; Hough, P.; Oldmeadow, L.; Bull, K.; Hill, K.; Greenwood, K. Four-item fall risk screening tool for subacute and residential aged care: The first step in fall prevention. *Australas. J. Ageing* **2009**, *28*, 139–143. [[CrossRef](#)]
- Kear, B.M.; Guck, T.P.; McGaha, A.L. Timed Up and Go (TUG) Test: Normative Reference Values for Ages 20 to 59 Years and Relationships with Physical and Mental Health Risk Factors. *J. Prim. Care Commu. Health* **2017**, *8*, 9–13. [[CrossRef](#)]
- Miranda-Cantellops, N.; Tiu, T.K. Berg Balance Testing. Updated 17 February 2023. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island, FL, USA, 2023. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK574518/> (accessed on 5 June 2023).
- Helen, M. Hospital fall risk assessment tools: A critique of the literature. *Int. J. Nurs. Pract.* **2003**, *9*, 223–235.
- Chapman, J.; Bachand, D.; Hyrkaes, K. Testing the sensitivity, specificity and feasibility of four falls risk assessment tools in a clinical setting. *J. Nurs. Manag.* **2011**, *19*, 133–142. [[CrossRef](#)]
- Subramaniam, S.; Faisal, A.I.; Deen, M.J. Wearable Sensor Systems for Fall Risk Assessment: A Review. *Front. Digit. Health* **2022**, *4*, 921506. [[CrossRef](#)] [[PubMed](#)]
- Ferreira, R.N.; Ribeiro, N.F.; Santos, C.P. Fall Risk Assessment Using Wearable Sensors: A Narrative Review. *Sensors* **2022**, *22*, 984. [[CrossRef](#)]
- Chen, B.; Chen, C.; Hu, J.; Sayeed, Z.; Qi, J.; Darwiche, H.F.; Little, B.E.; Lou, S.; Darwish, M.; Foote, C.; et al. Computer Vision and Machine Learning-Based Gait Pattern Recognition for Flat Fall Prediction. *Sensors* **2022**, *22*, 7960. [[CrossRef](#)]
- Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V.I.; on behalf of the Precise4Q consortium. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med. Inf. Decis Mak.* **2020**, *20*, 310. [[CrossRef](#)]
- Tang, Y.T.; Romero-Ortuno, R. Using Explainable AI (XAI) for the Prediction of Falls in the Older Population. *Algorithms* **2022**, *15*, 353. [[CrossRef](#)]
- Mengarelli, A.; Tigrini, A.; Fioretti, S.; Verdini, F. Identification of Neurodegenerative Diseases from Gait Rhythm through Time Domain and Time-Dependent Spectral Descriptors. *IEEE J. Biomed. Health Inf.* **2022**, *26*, 5974–5982. [[CrossRef](#)] [[PubMed](#)]
- Moylan, K.C.; Binder, E.F. Falls in older adults: Risk assessment, management and prevention. *Am. J. Med.* **2007**, *120*, 493.E1–493.E6. [[CrossRef](#)] [[PubMed](#)]
- Peterson, M.D.; Rhea, M.R.; Sen, A.; Gordon, P.M. Resistance exercise for muscular strength in older adults: A meta-analysis. *Ageing Res. Rev.* **2010**, *9*, 226–237. [[CrossRef](#)] [[PubMed](#)]
- Li, Y.; Hou, L.; Zhao, H.; Xie, R.; Yi, Y.; Ding, X. Risk factors for falls among community-dwelling older adults: A systematic review and meta-analysis. *Front. Med.* **2023**, *9*, 1019094. [[CrossRef](#)]
- Seppala, L.J.; Glind, E.M.M.v.; Daams, J.G.; Ploegmakers, K.J.; de Vries, M.; Wermelink, A.M.A.T.; der Velde, N.V.; Blain, H.; Bousquet, J.; Bucht, G.; et al. Fall-Risk-Increasing Drugs: A Systematic Review and Meta-analysis: III. Others. *J. Am. Med. Dir. Assoc.* **2018**, *19*, 372.e1–372.e8. [[CrossRef](#)]
- Williams, G.R.; Deal, A.M.; Nyrop, K.A.; Pergolotti, M.; Guerard, E.J.; Jolly, T.A.; Muss, H.B. Geriatric assessment as an aide to understanding falls in older adults with cancer. *Sup. Care Cancer* **2015**, *23*, 2273–2280. [[CrossRef](#)]

20. Lord, S.R.; Smith, S.T.; Menant, J.C. Vision and falls in older people: Risk factors and intervention strategies. *Clin. Geriatr. Med.* **2010**, *26*, 569–581. [[CrossRef](#)] [[PubMed](#)]
21. Allali, G.; Launay, C.P.; Blumen, H.M.; Callisaya, M.L.; De Cock, A.M.; Kressig, R.W.; Srikanth, V.; Steinmetz, J.P.; Verghese, J.; Beauchet, O.; et al. Falls, Cognitive Impairment, and Gait Performance: Results from the GOOD Initiative. *J. Am. Med. Dir. Assoc.* **2017**, *18*, 335–340. [[CrossRef](#)]
22. Mozaffari, N.; Rezazadeh, J.; Farahbakhsh, R.; Yazdani, S.; Sandrasegaran, K. Practical fall detection based on IoT technologies: A survey. *Internet Things* **2019**, *8*, 100124. [[CrossRef](#)]
23. Razum, D.; Seketa, G.; Vugrin, J.; Lackovic, I. Optimal threshold selection for threshold-based fall detection algorithms with multiple features. In Proceedings of the 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 21–25 May 2018; pp. 1513–1516. [[CrossRef](#)]
24. Wu, F.; Zhao, H.; Zhao, Y.; Zhong, H. Development of a wearable-sensor-based fall detection system. *Int. J. Telemed. Appl.* **2015**, *2015*, 576364. [[CrossRef](#)]
25. Mao, A.; Ma, X.; He, Y.; Luo, J. Highly Portable, Sensor-Based System for Human Fall Monitoring. *Sensors* **2017**, *17*, 2096. [[CrossRef](#)]
26. Chaitep, T.; Chawachat, J. A 3-phase threshold algorithm for smartphone-based fall detection. In Proceedings of the 2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Phuket, Thailand, 27–30 June 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 183–186.
27. Xefteris, V.-R.; Tsanoua, A.; Meditskos, G.; Vrochidis, S.; Kompatsiaris, I. Performance, challenges, and limitations in multimodal fall detection systems: A review. *IEEE Sens. J.* **2021**, *21*, 18398–18409. [[CrossRef](#)]
28. Saboor, A.; Kask, T.; Kuusik, A.; Alam, M.M.; Moullec, Y.L.; Niazi, I.K.; Zoha, A.; Ahmad, R. Latest Research Trends in Gait Analysis Using Wearable Sensors and Machine Learning: A Systematic Review. *IEEE Access* **2020**, *8*, 167830–167864. [[CrossRef](#)]
29. Altilio, R.; Rossetti, A.; Fang, Q.; Gu, X.; Panella, M. A comparison of machine learning classifiers for smartphone-based gait analysis. *Med. Biol. Eng. Comput.* **2021**, *59*, 535–546. [[CrossRef](#)]
30. Makino, K.; Lee, S.; Bae, S.; Chiba, I.; Harada, K.; Katayama, O.; Tomida, K.; Morikawa, M.; Shimada, H. Simplified Decision-Tree Algorithm to Predict Falls for Community-Dwelling Older Adults. *J. Clin. Med.* **2021**, *10*, 5184. [[CrossRef](#)] [[PubMed](#)]
31. Alizadeh, J.; Bogdan, M.; Classen, J.; Fricke, C. Support Vector Machine Classifiers Show High Generalizability in Automatic Fall Detection in Older Adults. *Sensors* **2021**, *21*, 7166. [[CrossRef](#)]
32. Tong, L.; Song, Q.; Ge, Y.; Liu, M. HMM-Based Human Fall Detection and Prediction Method Using Tri-Axial Accelerometer. *Sens. J. IEEE* **2013**, *13*, 1849–1856. [[CrossRef](#)]
33. Gutiérrez, J.; Rodríguez, V.; Martín, S. Comprehensive Review of Vision-Based Fall Detection Systems. *Sensors* **2021**, *21*, 947. [[CrossRef](#)]
34. Cuaya-Simbros, G.; Perez-Sanpablo, A.I.; Morales, E.F.; Quiñones, U.I.; Nuñez-Carrera, L. Comparing Machine Learning Methods to Improve Fall Risk Detection in Elderly with Osteoporosis from Balance Data. *J. Heal. Eng.* **2021**, *2021*, 8697805. [[CrossRef](#)]
35. Galvão, Y.M.; Ferreira, J.; Albuquerque, V.A.; Barros, P.; Fernandes, B.J.T. A multimodal approach using deep learning for fall detection. *Expert Syst. Appl.* **2021**, *168*, 114226. [[CrossRef](#)]
36. Nait Aicha, A.; Englebienne, G.; Van Schooten, K.S.; Pijnappels, M.; Kröse, B. Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry. *Sensors* **2018**, *18*, 1654. [[CrossRef](#)] [[PubMed](#)]
37. Shickel, B.; Rashidi, P. Sequential interpretability: Methods, applications, and future direction for understanding deep learning models in the context of sequential data. *arXiv* **2020**, arXiv:2004.12524.
38. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762v7.
39. Bet, P.; Castro, P.; Ponti, M. Fall detection and fall risk assessment in older person using wearable sensors: A systematic review. *Int. J. Med Inform.* **2019**, *130*, 103946. [[CrossRef](#)]
40. Howcroft, J.; Kofman, J.; Lemaire, E.D. Feature selection for elderly faller classification based on wearable sensors. *J. Neuroeng. Rehabil.* **2017**, *14*, 47. [[CrossRef](#)]
41. Marscholke, M.; Nemitz, G.; Gietzelt, M.; Wolf, K.H.; Meyer, Z.; Schwabedissen, H.; Haux, R. Predicting in-patient falls in a geriatric clinic: A clinical study combining assessment data and simple sensory gait measurements. *Z. Gerontol. Geriatr.* **2009**, *42*, 317–321. [[CrossRef](#)]
42. Meyer, B.M.; Tulipani, L.J.; Gurchiek, R.D.; Allen, D.A.; Adamowicz, L.; Larie, D.; Solomon, A.J.; Cheney, N.; McGinnis, R.S. Wearables and Deep Learning Classify Fall Risk from Gait in Multiple Sclerosis. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1824–1831. [[CrossRef](#)]
43. Wu, S.; Ou, J.; Shu, L.; Hu, G.; Song, Z.; Xu, X.; Chen, Z. MhNet: Multi-scale spatio-temporal hierarchical network for real-time wearable fall risk assessment of the elderly. *Comput. Biol. Med.* **2022**, *144*, 105355. [[CrossRef](#)]
44. Carrara, F.; Pasco, L.; Gennaro, C.; Falchi, F. Learning to Detect Fallen People in Virtual Worlds. In Proceedings of the 19th International Conference on Content-Based Multimedia Indexing (CBMI'22), Graz, Austria, 14–16 September 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 126–130. [[CrossRef](#)]
45. de Sousa, F.A.S.F.; Escriba, C.; Bravo, E.G.A.; Brossa, V.; Fourniols, J.-Y.; Rossi, C. Wearable Pre-Impact Fall Detection System Based on 3D Accelerometer and Subject's Height. *IEEE Sens. J.* **2022**, *22*, 1738–1745. [[CrossRef](#)]



46. Lo, P.; Aziz, A. A FPGA threshold-based fall detection algorithm for elderly fall monitoring with verilog. *Bull. Electr. Eng. Inform.* **2021**, *10*, 2477–2487. [[CrossRef](#)]
47. Thella, A.K.; Suryadevara, V.K.; Rizkalla, M.; Hossain, G. Smart unit care for pre fall detection and prevention. In Proceedings of the 2016 IEEE National Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS), Dayton, OH, USA, 25–29 July 2016; pp. 220–224. [[CrossRef](#)]
48. Noury, N.; Rumeau, P.; Bourke, A.K.; ÓLaighin, G.; Lundy, J.E. A proposal for the classification and evaluation of fall detectors. *Irbm* **2008**, *29*, 340–349. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.