

Article

Research on CNN-BiLSTM Fall Detection Algorithm Based on Improved Attention Mechanism

Congcong Li ^{1,2}, Minghao Liu ², Xinsheng Yan ² and Guifa Teng ^{1,2,*}¹ School of Information Science and Technology, Hebei Agricultural University, Baoding 071001, China² Hebei Key Laboratory of Agricultural Big Data, Baoding 071001, China

* Correspondence: tguifa@hebau.edu.cn

Abstract: Falls are one of the significant causes of accidental injuries to the elderly. With the rapid growth of the elderly population, fall detection has become a critical issue in the medical and healthcare fields. In this paper, we propose a model based on an improved attention mechanism, CBAM-IAM-CNN-BiLSTM, to detect falls of the elderly accurately and in time. The model includes a convolution layer, bidirectional LSTM layer, sampling layer and dense layer, and incorporates the improved convolutional attention block module (CBAM) into the network structure so that the one-dimensional convolution layer replaces the dense layer to aggregate the information from channels, which allows the model to accurately extract different behavior characteristics. The acceleration and angular velocity data of the human body, collected by wearable sensors, are respectively input into the convolution layer and bidirectional LSTM layer of the model and then classified and identified by softmax after feature fusion. Based on comparison with models such as CNN and CNN-BiLSTM, as well as with different attention mechanisms such as squeeze-and-excitation (SE), efficient channel attention (ECA) and the convolutional block attention module (CBAM), this model improves the accuracy, sensitivity and specificity to varying degrees. The experimental results showed that the accuracy, sensitivity and specificity of the CBAM-IAM-CNN-BiLSTM model proposed in this paper were 97.37%, 97.29% and 99.56%, respectively, which proves that the model has good practicability and strong generalization ability.



Citation: Li, C.; Liu, M.; Yan, X.; Teng, G. Research on CNN-BiLSTM Fall Detection Algorithm Based on Improved Attention Mechanism. *Appl. Sci.* **2022**, *12*, 9671. <https://doi.org/10.3390/app12199671>

Academic Editor: Dimitris Mourtzis

Received: 15 August 2022

Accepted: 22 September 2022

Published: 26 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

Keywords: fall detection; attention mechanism; CBAM; feature fusion; neural network

1. Introduction

The world is facing the severe challenge of population aging at present [1]. According to the latest statistics of the United Nations, the total number of elderly people in the world has reached 629 million, which accounts for 10% of the world population. It is estimated that the proportion of the aging population will increase to 15% around the world by 2050. According to the data of China's seventh national census, the number of people aged 60 and above is 264.02 million, which accounts for 18.70% of the total population. Among them, the number of people aged 65 and above is 190.64 million, which accounts for 13.50%. The proportion of people aged 60 and above increases by 5.44 percentage points, and the degree of population aging deepens further. With the rapid increase in the proportion of elderly people, the health problems of the elderly have attracted more and more attention from the whole society.

Falls are one of the major factors causing injury, disability and even death of the elderly. Decreased gait stability and impaired balance function are the main causes of falls in the elderly. Decreased central control, decreased responsiveness, prolonged reaction time, decreased balance and coordinated movement in the elderly all increase the risk of falls. Another point is that the structure, functional impairment and degeneration of bones, joints, ligaments and muscles in the elderly are common causes of falls. The occurrence of diseases in the elderly, as well as psychological factors, can increase the risk of falls in the

elderly. According to the relevant report of the World Health Organization (WHO), about 300,000 people worldwide die from falls every year, and more than half of them are elderly people over 60 years old [2]. According to the data from China's disease surveillance system, falls have become the leading cause of injury-related death among people over 65 years old in China [3]. According to the estimation, more than 40 million elderly people in China fall at least once every year [4]. According to the report of the World Health Organization, falls are one of the primary health-related problems that contribute to the disease burden of the elderly in China. After falling, the elderly are prone to serious physical damage, which may not only affect their psychology but also reduce their ability to live independently. It can be seen that falls have become a serious problem threatening the physical and mental health of the elderly, and this has also become an important issue of social concern. Therefore, the research on fall detection has great social significance.

In recent years, investigators have completed a lot of work on fall detection algorithms. According to the different devices and detection methods used, fall detection algorithms are mainly divided into methods based on computer vision, scene sensors and wearable devices [5].

The fall detection method based on computer vision can passively obtain human motion information from monitoring equipment and process the acquired video or image to detect if a fall occurs. However, this method infringes upon the privacy of users, and blockage of a large area of the human body easily leads to misjudgment, which affects the accuracy rate [6–8]. The method based on scene sensors uses the scene sensors installed in the monitored area to collect data, such as pressure, vibration and sound, to determine if a fall occurs. However, these kinds of sensors have disadvantages, such as high configuration costs, high sensitivity to noise information, high susceptibility to the surrounding environment and high false and missing alarm rates; there are also stringent requirements for equipment disposal in different environments. Therefore, they are not suitable for daily living scenes [9–16]. The method based on wearable devices usually uses sensors, such as accelerometers, gyroscopes and inertial measurement units (IMU), to automatically detect falls and send help-needed information to medical staff through communication devices such as WIFI, mobile networks and Bluetooth [17–19]. With the development of sensor technology and the posture algorithm, IMU with a smaller size, higher accuracy and stronger performance have been applied in motion analysis. At present, the accuracy of the motion capture technology of the inertial sensor with the highest accuracy is close to that of video motion capture technology based on multiple cameras. Additionally, the gait analysis system based on wearable sensors has become a feasible means by which to promote a continuous fall risk assessment in non-hospital environments. With the development of inertial sensors and the constant increase in their detection accuracy, the method based on wearable devices may detect human body falls anytime and anywhere, and it will not infringe upon the privacy of users. Considering its low equipment cost and good user experience, this method is more advantageous than the other two methods.

In this paper, we propose a CBAM-IAM-CNN-BiLSTM model and incorporate the improved convolutional attention block module (CBAM-IAM) into the network structure, so that the one-dimensional convolution layer replaces the dense layer to aggregate the information from the channels. Because of the parameter-sharing property of the convolution operation, the introduction of one-dimensional convolution can reduce the parameters of the channel attention module and improve the overall running efficiency of the model. Through experiments comparing SE, ECA, the CBAM and the CBAM-IAM, it is proved that the improved attention mechanism can effectively improve the performance of the model. At the same time, we designed the experimental scheme, collected fall and daily activity data and used these data to train and test the model, which proved that our proposed model has good reliability and practicability.

2. Related Work

Nowadays, with the improvement of computer computing power, deep learning neural network algorithms are slowly emerging in the field of artificial intelligence (AI). Deep learning, as a machine learning method and also an artificial neural network, can independently construct (train) basic rules according to the sample data during the learning process. With the continuous development of the deep learning model, it becomes increasingly important for inertial sensors to use deep learning to acquire human motion data via signals, analyze and process them and, thus, realize fall detection. Deep learning algorithms can automatically extract the most related features for evaluation, without the need to manually extract predetermined features from the sensor data, and they can provide better outcomes than the traditional machine learning algorithm.

At present, the most commonly used deep learning models in common fall detection include the convolutional neural network (CNN), long short-term memory networks (LSTM) and other network-based models. A fall detection model for the elderly was constructed by Lv et al. [20] On the basis of the CNN, which can directly extract critical information from a large amount of tagged data layer by layer through training and optimizing the multilayer convolutional neural network. When we use the CNN model to extract information from data, we usually use a convolution kernel to extract the data's local information, but each piece of local information has a different influence on whether the data can be correctly identified. In recent years, many researchers have integrated an attention mechanism into the convolution module, which proves that the attention mechanism has great potential for improving network performance. Hu et al. proposed the squeeze-excitation (SE) module [21], which can learn the correlation between channels in the feature graph and generate channel attention so that the information-rich channels more greatly concern the network, which brings an obvious performance improvement to the CNN. The convective block attention module (CBAM) [22] is a further extension of the SE module, which pools feature graphs globally according to channels to gain spatial attention. Attention mechanisms have become an increasingly common component of neural architecture and have been widely used in the field of behavior recognition [23–25].

However, when the CNN is used to extract data features and to detect and categorize the data, the correlation between the time series data will be ignored. In response to this issue, a recurrent neural network is designed to process the time series data. Using LSTM's ability to process the time-order correlation information, we may obtain the behavior pattern that leads to the fall by analyzing the sequence signals before the fall so as to determine whether or not the fall occurs. Musci et al. [26] designed a model architecture based on LSTM that can effectively detect falls and run on wearable devices. Duan et al. [27] used a bidirectional long-term and short-term memory neural network to detect falls. The results of the experiment showed that bidirectional LSTM better balanced the accuracy and detection delay, but the LSTM network had disadvantages in parallel processing and ignored the spatial characteristics of data. In [28], the authors argue that the LSTM cannot distinguish highly similar activities. In [29], the author considers using the BiLSTM instead of the LSTM for forecasting problems in the time series analysis. The final prediction result of the bidirectional LSTM is jointly determined by the forward layer and the reverse layer.

Based on the CNN and LSTM, Yang et al. [30] designed a CNN-LSTM model for fall detection that uses the CNN to extract features and then the LSTM layer to obtain a continuous time series. The method can effectively improve the accuracy of the algorithm, and the feasibility of the algorithm has been verified by experiments. Liang et al. [31] proposed the CBAM-CNN-LSTM model, and experiments proved that the collaboration between the LSTM, CNN and CBAM can enhance the modeling ability and improve the prediction accuracy. The CBAM-CNN-LSTM combines the advantages of the CNN and LSTM: the CNN is used to extract deep features of data, the LSTM is used to analyze temporal dependencies between data, and the CBAM can extract meaningful content and important information from data. Compared with the traditional CNN and LSTM models, this model is more reliable.

Therefore, in view of the problems of the CNN and LSTM deep learning networks mentioned above, a CBAM-IAM-CNN-BiLSTM fall detection algorithm, based on IMU combined with accelerometer and gyroscope to collect human motion data, is proposed in this paper in order to make full use of the data and features of effective fall detection. In this algorithm, the spatial features of the data are extracted by the CNN, the temporal features are extracted by the LSTM, and the features are fused. The CBAM is added to the CNN as an attention mechanism, and one-dimensional convolution is introduced instead of the dense layer in the CBAM's channel attention module to aggregate the information between channels, which can reduce redundant calculations and better extract features so as to improve the robustness and stability of the fall detection algorithm in complex environments.

3. Materials and Methods

This section mainly consists of three parts: human motion data acquisition, data preprocessing and the fall detection algorithm's design. The first part is mainly about designing the experimental scheme, collecting the data of human daily activities and falls, and constructing the data set. The second part mainly deals with data noise reduction and data segmentation and converts the data into a form suitable for the algorithm. The third part proposes the CBAM-IAM-CNN-BiLSTM algorithm and describes the automatic feature extraction and learning of data related to different types of falls and daily activities so as to finally realize accurate classification and identification of falls.

3.1. Experimental Scheme

A human body activity model is built first. When the sensor is placed vertically, the three-dimensional coordinates of the human body are shown in Figure 1, wherein the axis Y is perpendicular to the ground, representing the up–down direction of the human body; the axis X represents the front–back direction of the human body and the axis Z represents the left–right direction of the human body.

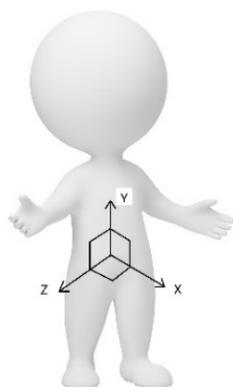


Figure 1. Human activity model.

At present, most of the fall detection research uses the data obtained from young people imitating the movements of the elderly in a laboratory environment. However, the research shows that the acceleration amplitude of the elderly is much smaller than that of the young people during walking, mainly because the physical function of the elderly declines, and their movements are sluggish. Therefore, in the absence of real fall data of the elderly, subjects wear an elderly life simulation experience suit, equipped with wrist guards at the knees, elbows and wrists to restrict joint activities and sandbags tied at the feet and wrists to simulate the state of the elderly, such as clumsy body, stiff limbs and sluggish movements, to make the young people imitate the movements of the elderly more realistically in the laboratory environment and to avoid unnecessary injuries. The LPMS-B2 posture sensor is used in the experiment, and the equipment parameters are shown in Table 1.

Table 1. LPMS-B2 main performance parameters.

Parameters	LPMS-B2
Bluetooth	Bluetooth 2.0 and BLE (Low Power/Bluetooth 4.1)
Communication range	<20 m
Output range of Euler angle	Roll: $\pm 180^\circ$; Pitch: $\pm 90^\circ$; Yaw: $\pm 180^\circ$
Resolution	<0.01°
Accuracy	<0.5° (Static), <2° RMS (Dynamic)
Accelerometer	3-axis, $\pm 2/\pm 4/\pm 8/\pm 16$ g, 16 bits
Gyroscope	3-axis, $\pm 125/\pm 245/\pm 500/\pm 1000/\pm 2000$ dps, 16 bits
Rate noise density	0.007 dps/ $\sqrt{\text{HZ}}$
Maximum sampling rate	400 Hz

The sensor communicates with the computer through Bluetooth and collects data at a frequency of 200 Hz, including acceleration data acquired in a range of ± 16 g and angular velocity data acquired in a range of ± 2000 dps. Based on the previous experimental analysis of different wearing positions of the sensor, it is found that the accuracy rate is the highest when the device is worn at the waist, so the sensor is fixed to the middle of the waist of the subject. The experimental environment is shown in Figure 2.

**Figure 2.** Subjects wore devices for simulated experiments.

This experiment collects data on four types of falling movements, including slipping when walking, falling in a faint, falling when sitting down and falling when getting up, and six types of daily movements, including walking, jogging, jumping, going up stairs, going down stairs and sitting down. A total of 12 subjects are included in the research, 10 males and 2 females aged between 20 and 25. The sizes of the samples are shown in Table 2. Among them, there are 560 samples of falls and 500 samples of each other daily activities, with a total of 3560 samples. The acceleration and angular velocity data of various falls and daily activities are shown graphically in Figures 3 and 4, respectively.

Table 2. Number of data collection samples.

Activity	Samples
Fall	560
Walk	500
Jog	500
Jump	500
Up stairs	500
Down stairs	500
Sit down	500
Total	3560

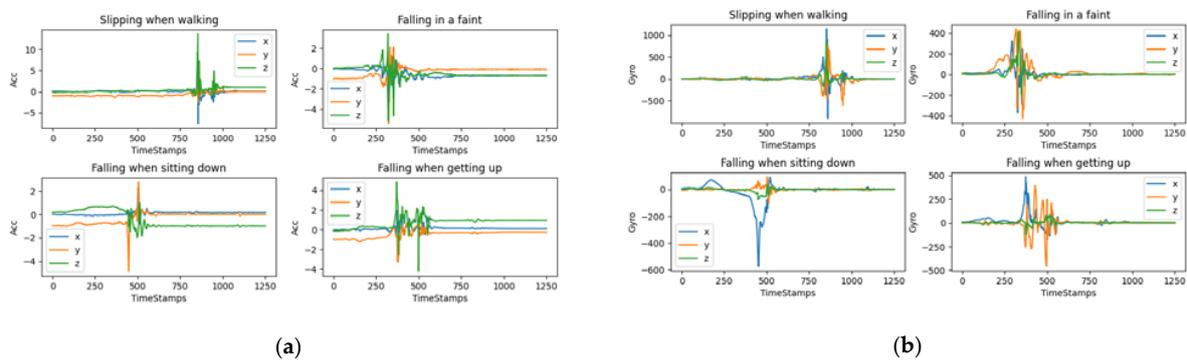


Figure 3. (a) Slipping when walking, falling in a faint, falling when sitting down and falling when getting up acceleration graphs; (b) slipping when walking, falling in a faint, falling when sitting down and falling when getting up angular velocity graphs.

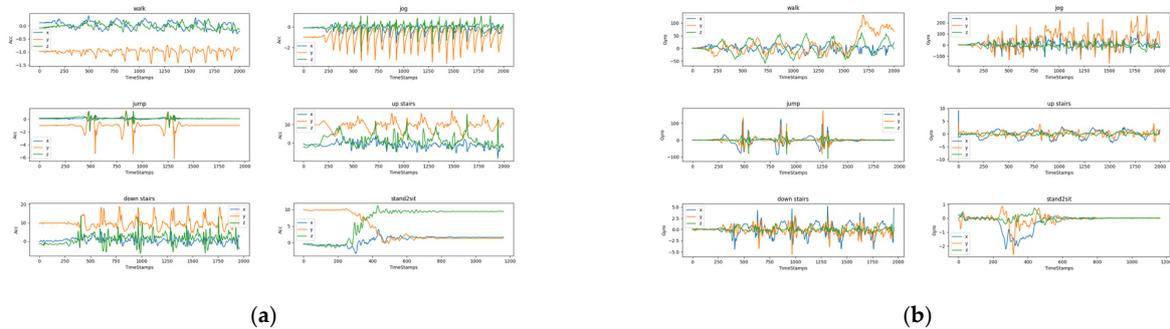


Figure 4. (a) Walking, jogging, jumping, going up stairs, going down stairs, sitting down acceleration graphs; (b) Walking, jogging, jumping, going up stairs, going down stairs, sitting down angular velocity graphs.

3.2. Data Preprocessing

3.2.1. Kalman Filtering

In 1960, R.E. Kalman published a paper on linear filtering of discrete data using recursive methods and proposed Kalman filtering [32]. Kalman filtering, an algorithm based on the linear system state equation, is the most widely used filtering method at present and can make the optimal estimation of the system state through inputting and outputting of the observation data of the system. Since the measured data include the influence of noise and interference from the system, the optimal estimation can also be regarded as a filtering process [33]. Zhu et al. [34] show a smoother waveform after Kalman filter processing of the experimental data.

In view of the sensors being easily influenced by various factors, such as vibration, temperature and electromagnetic interference, which affects the final classification accuracy of human activities [35], Kalman filtering is introduced to process the data and eliminate noise before extracting the features of the collected acceleration and angular velocity data.

The data before and after processing are shown in Figure 5. The main formulas of the Kalman filter are as follows:

$$X(k|k-1) = AX(k-1|k-1) + BU(k) \tag{1}$$

$$P(k|k-1) = AP(k-1|k-1)A^T + Q \tag{2}$$

$$K(k) = \frac{P(k|k-1)H^T}{HP(k|k-1)H^T + R} \tag{3}$$

$$X(k|k) = X(k|k-1) + K(k)[Z(k) - HX(k|k-1)] \tag{4}$$

$$P(k|k) = (1 - K(k)H)P(k|k-1) \tag{5}$$

wherein $X(k|k-1)$ is the priori state estimation value at time k , which is the result of the prediction at time k based on the best estimation at the last time point (time $k-1$); $X(k|k)$ and $X(k-1|k-1)$ represent the state estimation values at time k and time $k-1$, respectively; $P(k|k-1)$ is the priori covariance of estimation at time k ; $P(k|k)$ and $P(k-1|k-1)$ represent the posteriori covariance of the estimation at time k and time $k-1$, respectively; $Z(k)$ is the measured value; $U(k)$ is the observed noise; $K(k)$ is the Kalman gain; A is the state transition matrix; Q is the process excitation noise covariance; H is the transformation matrix from state variables to measurements; and $[Z(k) - HX(k|k-1)]$ is the residual of the actual observation and predicted observation, which are used together with the Kalman gain to correct the priori. Kalman filtering has two major calculation steps: In the first step, Formulas (1) and (2) are used to obtain the estimated value. In the second step, Formulas (3)–(5) are the state-updated equations of the Kalman filter, which are used to correct the previously estimated value to obtain the optimal estimated value at the current moment and to update the minimum mean square error matrix to prepare for the use of the Kalman algorithm at the next moment. Repeated updating of the equations can reduce the influence of noise on the sensor data, thus improving the accuracy of fall detection.

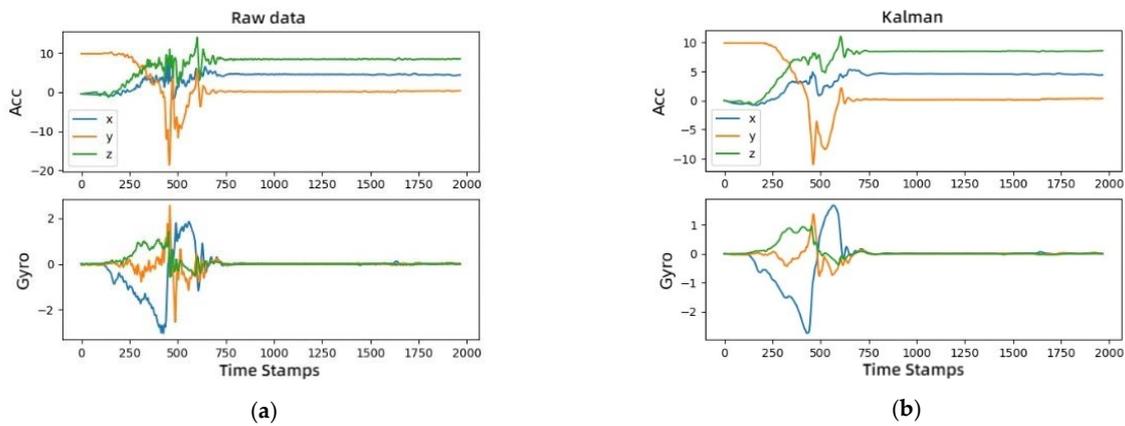


Figure 5. (a) Acceleration and angular velocity before Kalman filtering; (b) acceleration and angular velocity after Kalman filtering.

3.2.2. Data Segmentation

The experiment shows that the time from falling to touching the ground is generally less than 2 s. Since the sampling frequency is 200 Hz, the frequency of the falling activity is reduced to 100 Hz first so that a sliding window of 2 s is selected to intercept the three-axis acceleration and angular velocity data. Each group of data contains 1200 sampling points, which are, respectively, composed of triaxial acceleration and triaxial angular velocity data. The data format is shown in Formula (6).

$$X = [X_{1_{accx}}, X_{1_{accy}}, X_{1_{accz}}, X_{1_{gyrox}}, X_{1_{gyroy}}, X_{1_{gyroz}}, X_{2_{accx}}, \dots, X_{n_{gyroz}}] \tag{6}$$

wherein $X_{n_{accx}}$, $X_{n_{accy}}$, and $X_{n_{accz}}$ represent the accelerations of the three axes, respectively; $X_{n_{gyrox}}$, $X_{n_{gyroy}}$, and $X_{n_{gyroz}}$ represent the angular velocity of the three axes, respectively; and the value of n is 200.

3.3. Construction of Fall Detection Model

3.3.1. CBAM-IAM-CNN-BiLSTM Network Structure

A deep neural network model (the CBAM-IAM-CNN-BiLSTM model) using the CNN and bidirectional LSTM (BiLSTM), which integrates the improved CBAM attention mechanism, is proposed in this paper to classify human daily activities and fall behaviors to improve the accuracy of fall detection. In this model, the collected acceleration and angular velocity data are first sent to a CNN and bidirectional LSTM, respectively. The CNN is used to extract deep features of data; the LSTM is used to analyze temporal dependencies between data. The CBAM can extract meaningful content and important information from data, and then the spatial and temporal features of the human body movement data are fused and finally classified and identified. The overall structure of the model is shown in Figure 6.

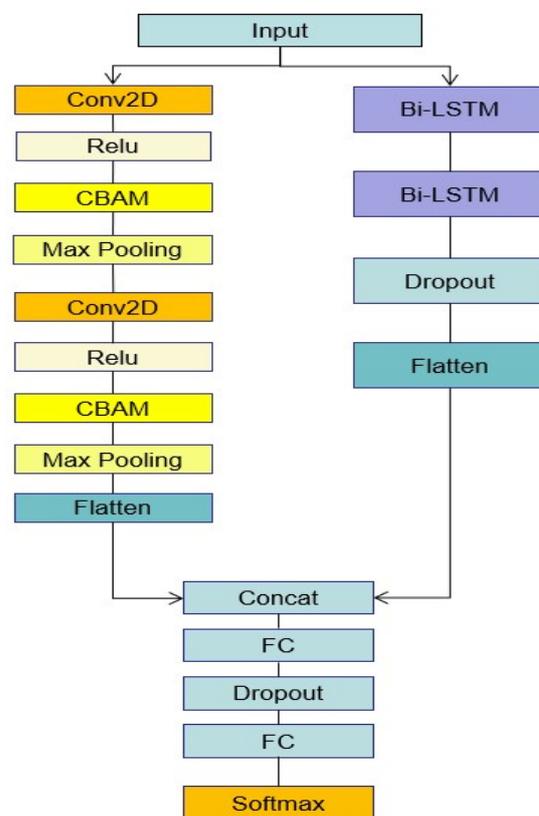


Figure 6. Schematic diagram of the CBAM-IAM-CNN-BiLSTM model.

The CBAM-IAM-CNN-BiLSTM model mainly consists of the following seven parts:

- 1 The input layer: This accepts data from the triaxial accelerometer and triaxial angular velocity meter and inputs them into the convolution layer and bidirectional LSTM layer, respectively.
- 2 The convolution layer: This accepts data from the input layer and performs a convolution operation. The size of the convolution kernel in the layer is $5 \times 5 \times 3$, and each unit is activated according to the equation's rectified linear unit (ReLU) activation function after convolution.
- 3 The attention mechanism: For the feature map generated by the convolutional neural network, the convolutional block attention module (CBAM) infers the attention map in two independent dimensions (channel and space), in turn, and then multiplies the

attention map by the input feature map for adaptive feature optimization. In this paper, the CBAMs are added in different positions of the model and improved in order to better extract the features of the human body’s daily behaviors and falling behaviors. Please refer to Section 3.3.2 for details.

- 4 The pool layer: The maximum pooling method is used for sub-sampling, which further reduces the dimension of the information extracted from the convolution layer, reduces the model size and computation burden and improves the robustness of the extracted features. The size of the kernel in the layer is 2×2 , the strides are two and the same padding is used to make the output size the same as the input size.
- 5 The bidirectional LSTM layer: This is composed of two layers of recurrent neural networks with the same inputs but different information transmission directions in two layers. The final prediction result is jointly determined by the forward layer and the reverse layer.
- 6 The dropout layer: This is added after the dense and bidirectional LSTM layers. During the training process, the neural network training unit is removed from the network according to a certain probability to prevent the model from over-fitting and improve the generalization ability of the model.
- 7 The output layer: Feature fusion is realized through splicing the feature vectors processed by convolution and the feature vectors processed by the bidirectional LSTM network. Then, the connection between all nodes of this layer and all corresponding nodes in the upper layer is realized via the dense layer, to synthesize the previously extracted features to obtain a specific value and, further, to obtain the final classification result according to the softmax classifier.

3.3.2. Convolutional Block Attention Module

The convolutional block attention module (CBAM) is a simple and effective attention module of the convolutional neural network. Given a feature map $F \in R^{C \times H \times W}$ as input, the CBAM derives a one-dimensional channel attention $M_C \in R^{C \times 1 \times 1}$ and a two-dimensional spatial attention $M_S \in R^{1 \times H \times W}$, in turn, as shown in Figure 7. The whole process can be represented by the following formulas:

$$F' = M_C(F) \otimes F \tag{7}$$

$$F'' = M_S(F') \otimes F' \tag{8}$$

wherein \otimes indicates multiplication of the corresponding elements. Before multiplication, the channel attention and space attention need to be broadcast according to the space dimension and channel dimension, respectively. F' is the feature map after the channel attention’s adjustment, and F'' is the final output feature map.

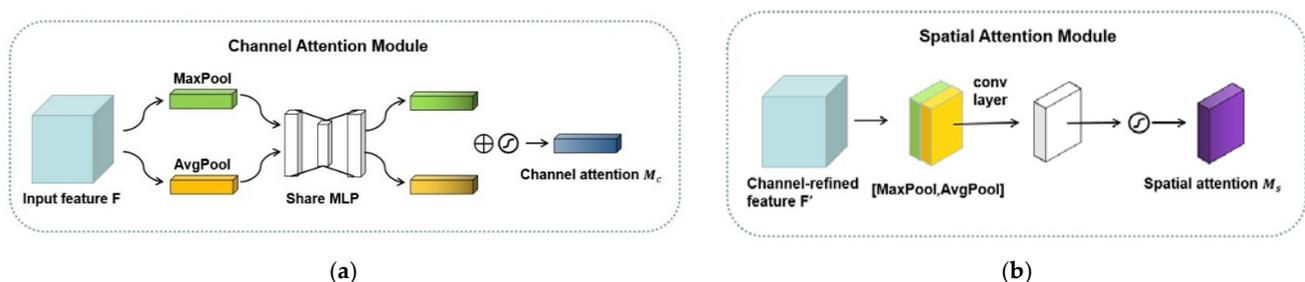


Figure 7. (a) Channel attention module in CBAM; (b) spatial attention module in CBAM.

For effective calculation of the channel attention, global mean pooling and maximum pooling are used firstly to aggregate the spatial information of feature mapping, thus generating two different channel descriptors, F_{avg}^C and F_{max}^C . The two descriptors are then forwarded to the shared network to generate the channel attention map $M_C \in R^{C \times 1 \times 1}$.

The shared network consists of a multilayer perceptron (MLP) with a hidden layer. The mathematical expression of the channel attention module is as follows:

$$M_C(F) = \sigma(MLP([AvgPool(F)] + MLP([MaxPool(F)])) = \sigma\left(W_1\left(W_0\left([F_{avg}^C]\right)\right) + W_1\left(W_0\left([F_{max}^C]\right)\right)\right) \quad (9)$$

wherein σ represents the Sigmoid function; $w_0 \in R^{\frac{c}{r} \times C}, W_1 \in R^{\frac{c \times C}{r}}$; the weights W_0 and W_1 of MLP are shared; and F_{avg}^C and F_{max}^C represent the mean and maximum pooling features, respectively.

The spatial attention module first performs global mean pooling and maximum pooling operations along the channel axis of the input feature map to generate two different spatial context descriptors F_{avg}^S and F_{max}^S and then makes connections and convolutions through the standard convolution layer to generate a two-dimensional spatial attention map. The mathematical expression of the attention module is as follows:

$$M_S(F) = \sigma\left(f^{7 \times 7}([AvgPool(F); MaxPool(F)])\right) = \sigma\left(f^{7 \times 7}\left([F_{avg}^S; F_{max}^S]\right)\right) \quad (10)$$

wherein σ represents the Sigmoid function, and $f^{7 \times 7}$ represents the convolution operation with a convolution kernel size of 7×7 .

When calculating and generating the channel attention, it is necessary to map features using the dense layer characterized by a heavy calculation burden. This takes up more resources and increases the parameters of the modules. Therefore, when a large number of the CBAMs are inserted into a convolutional neural network, the number of network parameters increases greatly, and the calculation time becomes longer, which does not meet the requirement that the fall detection needs a lower delay while ensuring accuracy. According to the literature [36], in any given intermediate feature map of the CNN, the mapping of channel features with the dense layer produces a large number of redundant calculations. Therefore, the one-dimensional convolution operation is designed in this paper to aggregate the channel features of the one-dimensional channel attention in order to solve this issue. Different from the original CBAM, which uses the dense layer to aggregate the channel features, the model proposed in this paper uses one-dimensional convolution with a convolution kernel length of k to aggregate the information of k channels in the neighborhood of this channel. Two convolved features are added by elements, and the channel attention $M_C \in R^{c \times 1 \times 1}$ is generated by the Sigmoid function operation. Then, the generated channel attention is broadcast and expanded to $R^{C \times H \times W}$ along two spatial dimensions and then multiplied with the input feature map by the corresponding elements to obtain the feature map after the channel attention is injected. The improved CBAM channel attention calculation process is shown in Formula (11) as follows:

$$M_C(F) = \sigma\left(f_{1D}^k(AvgPool(F)) + f_{1D}^k(MaxPool(F))\right) = \sigma\left(f_{1D}^k\left(F_{avg}^c\right) + f_{1D}^k\left(F_{max}^c\right)\right) \quad (11)$$

wherein σ represents the Sigmoid function, and f_{1D}^k represents the one-dimensional convolution operation with a convolution kernel size of k . The value of k is adaptively determined by the number of channels. The improved channel attention model is shown in Figure 8. Because of the parameter-sharing property of the convolution operation, the introduction of one-dimensional convolution can reduce the parameters of the channel attention module and improve the overall running efficiency of the model.

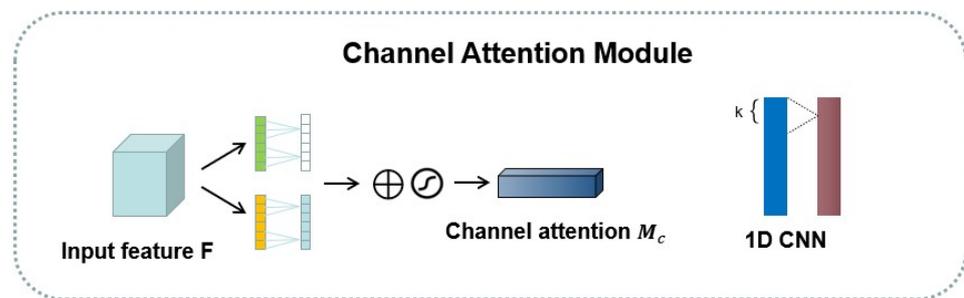


Figure 8. Improved CBAM channel attention mechanism model diagram.

4. Experiment and Result Analysis

This experiment is designed to detect daily behaviors and falling behaviors in the data and verify the detection accuracy. In this experiment, the deep learning framework tensorflow 2.5.0 is adopted to build the fall detection model, and the hardware configuration includes Intel Core I7-10700 processor and NVIDIA Geforce GTX1060 graphics card.

The constructed data set is used to train and test the model. The data set is randomly divided into the training set and testing set proportionally, of which 80% of the data is used for model training and 20% for model testing. The training set is used to train the fall detection model, while the testing set is used to evaluate the generalization ability of the final model. The Adam algorithm is used to optimize the network, and the dropout method is added between the network layers to improve the generalization ability of the model. The softmax classifier is added to the last dense layer of the network to calculate the final output result. The model parameter settings are shown in Table 3.

Table 3. Network parameter setting.

Setting Items	Parameter Value
Learning rate	0.001
Batch size	64
Iteration number	800
Dropout rate	0.2
Number of hidden layer cells	400

The loss function adopts a multi-classification cross-entropy loss function, and its loss value expression is as follows:

$$L = -\frac{1}{K} \sum_{k=1}^K \sum_{i=1}^N y_i \cdot \log p_i \quad (12)$$

wherein y_i is the real tag corresponding to the i th sample; p_i is the predicted value of the model for training this movement; N is the total number of movement classifications; and K is the total number of samples.

4.1. Evaluation Indicators

According to the classified prediction results, the analysis is performed from three aspects:

1. The accuracy rate reflects the proportion of the correct classifications among the identification results in all classifications and reflects the training effect of the model on the data set. Its mathematical expression is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

2. The specificity reflects the proportion of identified negative samples in all negative samples, and its mathematical expression is as follows:

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (14)$$

3. The sensitivity reflects the proportion of identified positive samples in all positive samples, and its mathematical expression is:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (15)$$

wherein TP is a positive sample predicted by the model to be positive; TN is a negative sample predicted by the model to be negative; FP is a negative sample predicted by the model to be positive; and FN is a positive sample predicted by the model to be negative.

4.2. Comparison of Model Results

The comparison with the CNN, CNN-BiLSTM and the three most widely used attention mechanisms, namely, squeeze-and-excitation (SE), efficient channel attention (ECA) and the convolutional block attention module (CBAM), is made in this paper to verify the effectiveness of this research model and the improved effect. The training accuracy rate is shown in Figure 9.

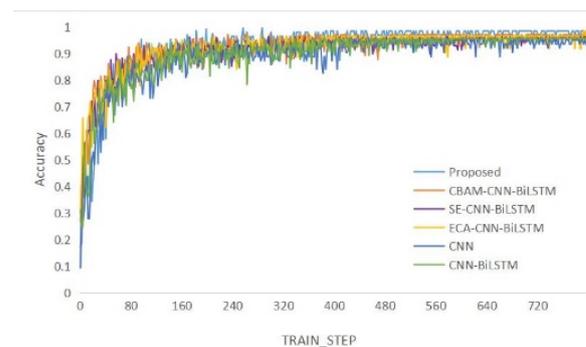


Figure 9. Training accuracy of different models.

The results are shown in Table 4. It can be seen intuitively that the accuracy, sensitivity and specificity of the model in this paper are the highest. Figure 9 shows that the accuracy of the model in this paper rises slowly, and there is no large-scale oscillation phenomenon. After 480 iterations, the accuracy of the model tends to be stable. The attention mechanism can assign different weights to the feature information to obtain more detailed information about the target needing attention, thus suppressing other useless information. Through the analysis of the CNN-BiLSTM model and the models incorporating the SE, ECA and CBAM attention mechanisms, it can be seen that the accuracy of the models has improved to different degrees after the introduction of attention mechanisms, among which the ECA attention mechanism module has brought the greatest improvement of the accuracy of the models (0.66%). Compared with the SE module, the ECA module uses a 1×1 convolution layer to replace the dense layer directly after the global mean pooling layer. This module avoids dimension reduction and effectively captures cross-channel interaction. SE and ECA pay more attention to the analysis of the channel domain and are limited to consideration of the interaction relationship between the feature map channels. Using the consideration of the channel and the action scope as the starting points, the CBAM introduces two analytical dimensions of spatial attention and channel attention and realizes the sequential attention structure from channel to space.

Table 4. Performance comparison of CBAM-IAM-CNN-BiLSTM model with other models.

Model	Accuracy	Sensitive	Specificity
CNN	95.65%	95.52%	99.27%
CNN-BiLSTM	95.79%	95.51%	99.30%
SE-CNN-BiLSTM	95.83%	96.08%	99.21%
ECA-CNN-BiLSTM	96.45%	95.22%	99.17%
CBAM-CNN-BiLSTM	96.18%	95.32%	99.21%
Proposed	97.37%	97.29%	99.56%

In view of the improvement of the ECA module compared with the SE module, one-dimensional convolution instead of the dense layer is introduced in this paper to aggregate the information between channels for the CBAM's channel attention module, so that the model can allocate attention to two dimensions and thus enhance the improvement effect of the attention mechanism on the model's performance. The results of the experiment show that, compared with the CBAM-CNN-BiLSTM network model, the accuracy, sensitivity and specificity of the improved model have increased by 1.19%, 1.97% and 0.35%, respectively, and that, compared with the CNN-BiLSTM model, the accuracy, sensitivity and specificity have improved by 1.58%, 1.78% and 0.26%, respectively.

The comparison results with the models proposed in other papers are shown in Table 5. The accuracy and specificity of the model in this paper are the highest. The results show that compared with an adapted RNN, all three indicators have been improved to a certain extent and that, compared with the CNN, the sensitivity and specificity have improved by 5.09% and 2.3%, respectively. Compared with the NT-FDS, the accuracy of the model is similar, but the specificity is 5% higher than that of the NT-FDS, and the false-positive rate is low, which does not cause a waste of resources.

Table 5. Performance comparison with models proposed in other papers.

Model	Accuracy	Sensitive	Specificity
CNN [37]	96.65%	92.20%	97.26%
Adapted RNN [26]	96.94%	96.73%	97.15%
NT-FDS [38]	97.21%	99.54%	94.56%
Proposed	97.37%	97.29%	99.56%

With the CNN parameters remaining constant, we have investigated the model's accuracy change in one, two and three layers of the LSTM in order to observe the influence of the number of LSTM layers on the classification accuracy. The results of the experiment are shown in Table 6. When the number of LSTM layers is two, the accuracy of the model is the highest, and its performance is better than that of one and three layers. Therefore, we finally set the number of LSTM layers to two.

Table 6. Accuracy of the model at different LSTM layers.

Number of Layers	Accuracy	Sensitive	Specificity
1	95.31%	93.47%	98.96%
2	97.37%	97.29%	99.56%
3	96.18%	96.21%	99.36%

4.3. Model Test Results

In this paper, seven types of movements are identified, including falling, walking, jogging, jumping, going up stairs, going down stairs and sitting down, and a total of 760 data sets are used for testing. The test results are shown in Table 7, and the confusion matrix is shown in Figure 10. The confusion matrix includes the number of samples with correct and wrong predictions for the seven movements in the test set. In the confusion matrix, the numerical values on the principal diagonal are the numbers of correctly predicted samples,

while the numerical values at other positions are the numbers of incorrectly predicted samples. In the process of data collection, there are four types of falling movements, each of which also includes forward, backward and lateral falls. As shown in Table 7, the identification accuracy of falling movements is 99.74%, the sensitivity is 100%, and the specificity is 99.69%. As shown in Figure 10, one jump is identified incorrectly as falling, which may be due to the similarity in the forward/backward movement between jumping and falling and the large fluctuation in the Y-axis data. For daily behavior, the identification effect of sitting down is the best. During the experiment, the subjects wore the life simulation experience suit to imitate the elderly, which is intended to restrict their physical activities and result in a smaller movement range for jogging. Since the jogging movement is similar to walking behavior in their movement ranges due to physical restriction, the jogging behavior is predicted incorrectly as walking behavior in 11 groups. Through the analysis of the above experimental results, this model can effectively identify human falls and daily activities.

Table 7. CBAM-IAM-CNN-BiLSTM model test results.

Number of Layers	Accuracy	Sensitive	Specificity
Fall	99.74%	100%	99.69%
Walk	98.16%	97.94%	98.19%
Jog	97.89%	86.41%	99.70%
Jump	98.95%	95.87%	99.53%
Up stairs	99.11%	99.10%	99.08%
Down stairs	99.21%	95.54%	99.85%
Sit down	99.87%	100%	99.85%

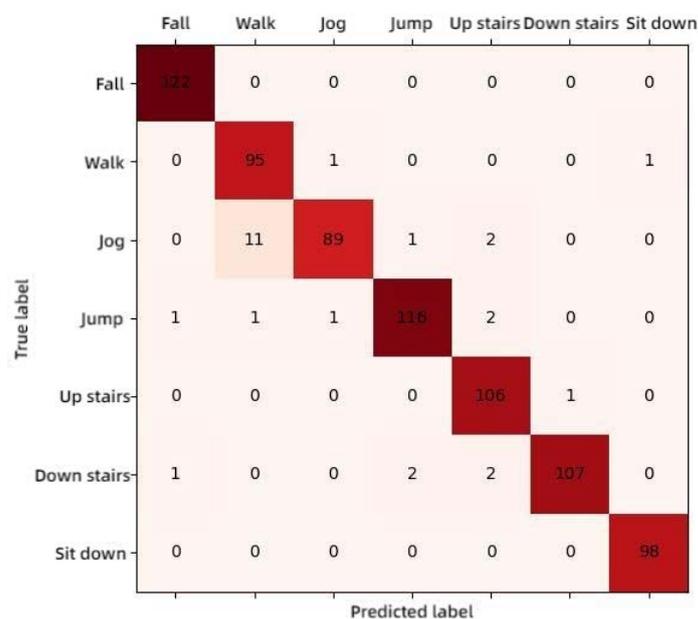


Figure 10. The test set’s confusion matrix.

5. Conclusions

First, the importance of fall detection is discussed in this paper. Then, combined with the application of deep learning in fall detection, the analysis is conducted on the shortcomings of existing fall detection methods. Through the discussion on several of the most widely used deep learning models and attention mechanisms, the CBAM-IAM-CNN-BiLSTM fall detection model is proposed, and the IMU sensor device is used to collect human body movement data to construct a fall data set. For the channel attention module in the CBAM, one-dimensional convolution is used instead of the dense layer to enhance the effect of the attention mechanism on the performance of the model. Combined with the feature fusion method, the accuracy of fall detection is improved by adding contextual

information. The experimental results show that the accuracy rate of the fall detection model proposed in this paper is 97.37%. Compared with other network models, this model is characterized by higher accuracy, specificity and sensitivity, as well as a stronger generalization ability and better practicability.

There are still limitations in this study. Xu et al. [39] argue that real falls are an unpredictable behavior, which makes it difficult to build data sets correlated with real data. Although this study took falls in different situations into account in the collection of data sets and tried to simulate the movement state of the elderly, there are still differences between the simulated data and the real data. Therefore, the high-precision fall detection algorithms implemented in the laboratory are not satisfactory in practice. In addition, the real-time nature of fall detection is particularly important. Compared with the CBAM-CNN-BiLSTM, the method proposed in this study reduces the number of parameters and takes less computation time, and, compared with the CNN and LSTM, the model has higher accuracy but a longer computation time.

In the future, we will further improve the data set and add more movements to verify the performance of the model. We will also continue to improve the model, optimize the real-time training process, shorten the training cycle, and reduce misjudgment in future research. Additionally, we will improve the spatial attention module in the CBAM and focus on multiple factors to improve the detection performance of the algorithm, such as changing the basic network.

Author Contributions: All authors contributed greatly to this article. C.L. designed the research protocol, paper framework, and feature fusion algorithm and wrote the manuscript. M.L. designed the data collection scheme, model construction and generation of results. X.Y. was responsible for experimental data collection, data processing and results analysis. G.T. proofread the article and revised and commented on the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the S&T program of Hebei, grant number 203777119D, 19227210D; the Scientific Research Projects of Universities in Hebei Province, grant number ZD2021056; and the National Natural Science Foundation of China's intention-driven network behavior measurement, analysis and system, grant number U20A20180.

Institutional Review Board Statement: Not applicable.

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

Data Availability Statement: The data presented in this study are available on request from the authors. The data are not publicly available due to the privacy concerns of subjects participating in the experiment.

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

References

1. United Nations Population Division. Available online: <https://www.un.org/development/desa/pd/content/World-Population-Prospect-2022> (accessed on 12 September 2022).
2. World Health Organization. Available online: <https://www.who.int/zh/news-room/fact-sheets/detail/falls> (accessed on 5 August 2022).
3. Zhang, Y.; Chen, W. Survey and progress of elderly fall research. *Chin. J. Gerontol.* **2008**, *28*, 929–931.
4. Ministry of Health. *Technical Guide for Intervention of Falls in the Elderly*; Ministry of Health: Beijing, China, 2011.
5. Mubashir, M.; Shao, L.; Seed, L. A survey on fall detection: Principles and approaches. *Neurocomputing* **2013**, *100*, 144–152. [[CrossRef](#)]
6. García, E.; Villar, M.; Fández, M.; Villar, J.R.; de la Cal, E.; Cho, S.-B. Towards effective detection of elderly falls with CNN-LSTM neural networks. *Neurocomputing* **2022**, *500*, 231–240. [[CrossRef](#)]
7. Taramasco, C.; Rodenas, T.; Martínez, F.; Fuentes, P.; Muñoz, R.; Olivares, R.; De Albuquerque, V.H.C.; Demongeot, J. A Novel Monitoring System for Fall Detection in Older People. *IEEE Access* **2018**, *6*, 43563–43574. [[CrossRef](#)]
8. Koshmak, G.; Linden, M.; Loutfi, A. Dynamic Bayesian Networks for Context-Aware Fall Risk Assessment. *Sensors* **2014**, *14*, 9330–9348. [[CrossRef](#)]

9. Wang, Z. Key Technologies for Human Abnormal Behavior Detection in Home Video Surveillance Environment. Master's Thesis, Nanjing University of Posts and Telecommunications, Nanjing, China, 2020.
10. Cai, W.; Zheng, X.; Guo, J.; Ruans, Z. Vision-aware fall detection algorithm based on SVM-MultiCNN model. *J. Hangzhou Dianzi Univ. Nat. Sci.* **2020**, *40*, 59–66. [[CrossRef](#)]
11. Yang, X.; Tang, X.; Zhang, G.; Huang, Y. Human fall detection method based on YOLO network. *J. Yangzhou Univ. Nat. Sci. Ed.* **2019**, *22*, 61–64+78. [[CrossRef](#)]
12. Wang, P.; Ding, H.; Li, L. A method of fall detection based on human posture in video. *Mod. Electron. Tech.* **2021**, *44*, 98–102. [[CrossRef](#)]
13. Zou, F. Research on Indoor Fall Detection and Behavior Analysis Based on Scene Context. Master's Thesis, Nanchang University, Nanchang, China, 2020.
14. Liu, F.; Xu, Z.; Gan, Z.; Liu, S. Fall detection algorithm based on temporal motion features in RGB-D videos. *J. Nanjing Univ. Posts Telecommun. Nat. Sci. Ed.* **2020**, *40*, 117–124. [[CrossRef](#)]
15. Zhu, Y.; Zhang, Y.; Li, S.; Li, W.; Liu, Y. Fall detection algorithm based on depth vision sensor and convolution neural network. *Opt. Tech.* **2021**, *47*, 56–61. [[CrossRef](#)]
16. Chuanbi, L.; Ziqian, D.; Weikai, K.; Yungfa, H. A Framework for Fall Detection Based on OpenPose Skeleton and LSTM/GRU Models. *Appl. Sci.* **2020**, *11*, 329.
17. Wang, X.; Li, D.; Zheng, X.; Lou, T.; Ding, G.; Jiao, Y.; Zhao, H. Research on fall detection algorithm based on RBF neural network. *J. Electron. Meas. Instrum.* **2019**, *33*, 185–191. [[CrossRef](#)]
18. Xue, Y.; Gao, X. Design of a fall monitoring system based on multi-sensor information fusion. *J. Wuhan Univ. Technol. Inf. Manag. Eng.* **2011**, *33*, 712–716.
19. Li, H.; Shrestha, A.; Fioranelli, F.; Kernec, J.L.; Spinsante, S. Multisensor data fusion for human activities classification and fall detection. In Proceedings of the 2017 IEEE Sensors, Glasgow, UK, 29 October–1 November 2017.
20. Lv, Y.; Zhang, M.; Jiang, W.; Ni, Y.; Qian, X. Design of elderly fall detection system using CNN. *J. Zhejiang Univ. Eng. Sci.* **2019**, *53*, 1130–1138.
21. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [[CrossRef](#)]
22. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. *arXiv* **2018**, arXiv:1807.06521.
23. Chen, T.; Ding, Z.; Li, B. Elderly Fall Detection Based on Improved YOLOv5s Network. *IEEE Access* **2022**, *10*, 91273–91282. [[CrossRef](#)]
24. Muhammad, K.; Ullah, A.; Imran, A.S.; Sajjad, M.; Kiran, M.S.; Sannino, G.; de Albuquerque, V.H.C. Human action recognition using attention based LSTM network with dilated CNN features. *Future Gener. Comput. Syst.* **2021**, *125*, 820–830. [[CrossRef](#)]
25. Chen, Z.; Zhang, L.; Jiang, C.; Cao, Z.; Cui, W. WiFi CSI based passive human activity recognition using attention based BLSTM. *IEEE Trans. Mob. Comput.* **2018**, *18*, 2714–2724. [[CrossRef](#)]
26. Musci, M.; Martini, D.D.; Blago, N.; Facchinetti, T.; Piastra, M. Online Fall Detection using Recurrent Neural Networks on Smart Wearable Devices. *IEEE Trans. Emerg. Top. Comput.* **2020**, *9*, 1276–1289. [[CrossRef](#)]
27. Duan, M.; Pan, J. Study on Wearable Fall Detection based on Bi-directional LSTM Neural Network. *J. Guangxi Norm. Univ. Nat. Sci. Ed.* **2022**, *40*, 141–150. [[CrossRef](#)]
28. Wu, X.; Zheng, Y.; Chu, C.-H.; Cheng, L.; Kim, J. Applying deep learning technology for automatic fall detection using mobile sensors. *Biomed. Signal Process. Control* **2022**, *72*, 103355. [[CrossRef](#)]
29. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The performance of LSTM and BiLSTM in forecasting time series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.
30. Yang, G.; Zhao, J.; Guo, J. Research on Fall Detection Algorithm Based on CNN and LSTM. In Proceedings of the 2021 3rd International Conference on Natural Language Processing (ICNLP), Beijing, China, 26–28 March 2021; pp. 190–195.
31. Liang, Y.; Lin, Y.; Lu, Q. Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. *Expert Syst. Appl.* **2022**, *206*, 117847. [[CrossRef](#)]
32. Kalman, R. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [[CrossRef](#)]
33. Tang, Y.; Dai, Y. Wireless strain synchronization acquisition method based on Kalman Filter. *J. Phys. Conf. Ser.* **2021**, *1754*, 012064. [[CrossRef](#)]
34. Zhu, X.; Qiu, T.; Qu, W.; Zhou, X.; Wu, D. BLS-Location: A Wireless Fingerprint Localization Algorithm Based on Broad Learning. *IEEE Trans. Mob. Comput.* **2021**. [[CrossRef](#)]
35. He, J.; Zhou, M.; Wang, X. Wearable Method for Fall Detection Based on Kalman Filter and k-NN Algorithm. *J. Electron. Inf. Technol.* **2017**, *39*, 2627–2634.
36. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
37. Han, T.; Kang, W.; Choi, G. IR-UWB sensor based fall detection method using CNN algorithm. *Sensors* **2020**, *20*, 5948. [[CrossRef](#)]
38. Waheed, M.; Afzal, H.; Mehmood, K. NT-FDS—A Noise Tolerant Fall Detection System Using Deep Learning on Wearable Devices. *Sensors* **2021**, *21*, 2006. [[CrossRef](#)]
39. Xu, T.; Zhou, Y.; Zhu, J. New advances and challenges of fall detection systems: A survey. *Appl. Sci.* **2018**, *8*, 418. [[CrossRef](#)]