Fall prediction based on biomechanics equilibrium using Kinect

Xu Tao[1,2] and Zhou Yun[3]

Abstract

The fall is one of the most important research fields of solitary elder healthcare at home based on Internet of Things technology. Current studies mainly focus on the fall detection, which helps medical staffs bring a fallen elder out of danger in time. However, it neither predicts a fall nor provides an effective protection against a fall. This article studies the fall prediction based on human biomechanics equilibrium and body posture characteristics through analyzing three-dimensional skeleton joints data from the depth camera sensor Kinect. The research includes building a human bionic mass model using skeleton joints data from Kinect, determining human balance state, and proposing a fall prediction algorithm based on recurrent neural networks by unbalanced posture features. We evaluate the model and algorithm on an open database. The performance indicates that the fall prediction algorithm by studying human biomechanics can predict a fall (91.7%) and provide a certain amount of time (333 ms) before the elder injuring (hitting the floor). This work provides a technical basis and a data analytics approach for the fall protection.

Keywords

Fall prediction, recurrent neural networks, Internet of Things, data analytics, machine learning

Introduction

Fall is one of the most threats to elders' health.[1] Due to its bursting, diversity, and hard reproducibility, the research on fall has been a big challenge of eldercare. Rapid development of Internet of Things (IoT) technology provides a possibility to provide healthcare services at elders' home. Researchers attempt to record elders' behaviors through setting various sensors on the room or human's body and analyze the fall process happened at home. These data analytic techniques and cloud computing techniques enable the medical staffs to monitor elders' states at a distance away from elders' home.

Current researches are mainly focused on the fall detection. It can reduce the time that the elders remain lying on the floor after falling. This time is one of the key factors that determines the severity of a fall.[2] The works on the fall detection recognize the fall action after the elder hitting the floor. The features extracted include body shape change,[3] posture,[4] head change,[5] and inactivity.[6] The sensors used to detect a fall are accelerometers and cameras, which are limited to describe fall processing precisely. The fall detection provides the medical assistance after falling rather than avoid injuring. Therefore, if we accurately predict a fall and provide the effective protection, the risk of injury could be reduced effectively for the elder.

Either the fall detection or prediction can be considered as the research on sequential data processing. The

[1]School of Software and Microelectronics, Northwestern Polytechnical University, Xi'an, P.R. China
[2]State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, P.R. China
[3]School of Education, Shaanxi Normal University, Xi'an, P.R. China

Corresponding author:
Zhou Yun, School of Education, Shaanxi Normal University, 199, South Chang'an Road, Xi'an 710062, Shaanxi, P.R. China.
Email: zhouyun@snnu.edu.cn

recurrent neural networks (RNNs) as a sequential prediction model have a unique advantage and improve the prediction accuracy. It has achieved great success in sequential data analysis and has been widely used as in previous studies.[7–9] LeCun et al.[10] indicate that recurrent nets have shone light on sequential data. At same time, the advent of new sensors enhances the capacity for perception. Kinect,[11] designed by Microsoft, is a line of motion sensing input devices, which can locate 3D positions of body joints of the tracked users in space and track their movements over time. Overall, it provides more detailed information of human's movement and lays the foundation for elaborately analyzing fall processing.

In our article, we employ RNN to study the fall prediction based on human biomechanics equilibrium. According to analyzing the human action by biomechanics equilibrium, we study fall prediction by sequential data classification. The basic idea is to choose sequential data from before body imbalance to after body imbalance but not hitting the floor and then to classify it into activity of daily living (ADL) or fall, which makes falls predicted before injury. The overview of our method, as shown in Figure 1, includes three parts: building a human bionic dynamic mass model by skeleton joints data from Kinect and human mass distribution, determining body's balance by computing positions of center of mass (COM) and the region of support of base, and proposing a fall prediction algorithm based on the RNN by unbalanced posture features.

We evaluated our method using the data in the open database.[5] The results show that our fall prediction algorithm by studying human biomechanics equilibrium and posture recognition can predict 91.5% fall and classify 75% types of fall from others ADL before the elder hitting the floor in 333 ms in average. This method provides a technical basis for the fall protection.

The rest of the article is organized as follows. Related works are reviewed in next section, with the human balance is introduced, followed by the method to detect human balance from 3D skeleton joints from Kinect and human mass distribution. Then we provide a fall prediction method based on biomechanics equilibrium by long short-term memory (LSTM) networks. In the following section, we test and evaluate our method on the fall database. Finally, we conclude and discuss our future research direction.

Related works

Recent years, one of most popular researches related to the ambient intelligent (AmI) is about the study of the fall. There are many attempts to provide protection against the fall, which mainly revolve around fall detection. According to the type of sensors used, the studies can be classified into two categories: accelerator based and vision based.

Acceleration data are collected during the falls using independent tri-axial accelerometers attached to different parts of the body.[2] In the beginning, most of the existing researches detect a fall by setting the threshold for acceleration data.[12] J Cheng et al.[13] improved it and proposed a framework for activity awareness using the surface electromyography and accelerometer (ACC) signals. This work used the angle of each ACC and surface to indicate body postures, the resultant ACC
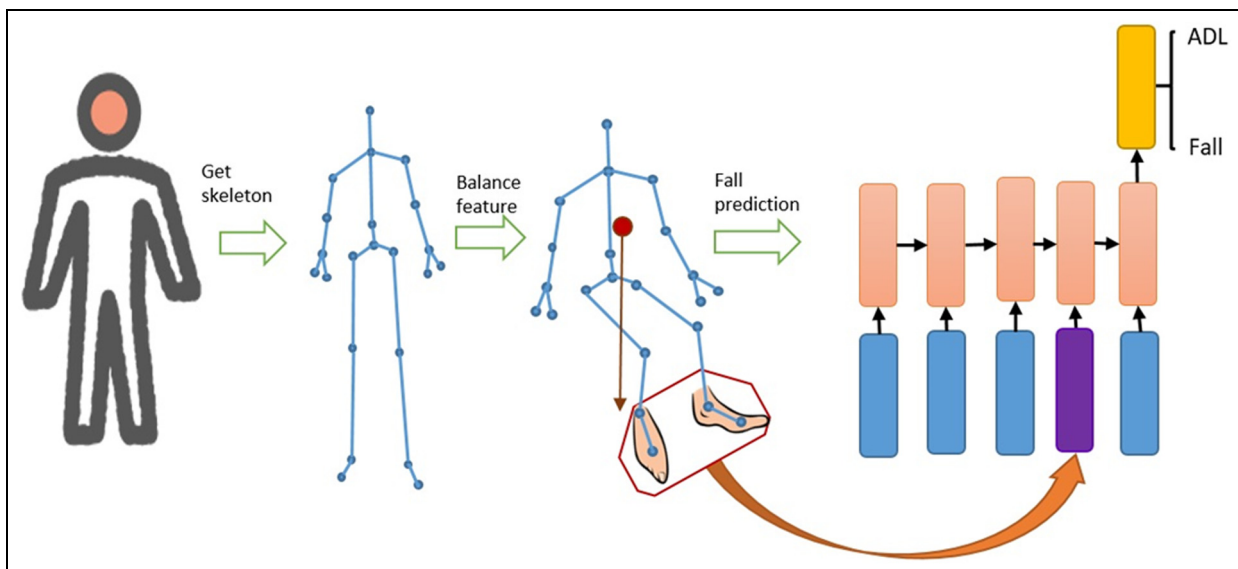

Figure 1. Overview of our fall predication method.

amplitude to distinguish between dynamic transition activities to normal transition activities and falls. Then increasing machine learning approaches have been employed. T Zhang et al.[14] used a wearable tri-axial accelerometer to capture the movement of human body and proposed a novel fall detection method based on one-class support vector machine (SVM). The preliminary results showed that this method could detect the falls effectively. However, these methods require the elder wearing some specific sensors on body, making the user feel inconvenient.

Vision-based methods for fall detection usually use camera to analyze elders' activities, which let elders avoid wearing extra sensors. Rougier et al.[15] proposed a method to detect falls by analyzing human shape deformation during a video sequence. Vishwakarma et al.[16] proposed a two-state finite state machine (FSM) to continuously monitor people and their activities. Increasing the new vision sensors is adopted in fall detection system, and the depth camera sensor Kinect is one of them with its economical price and good performance. Mastorakis and Makris[17] presented a novel fall detection system based on the Kinect sensor. Bevilacqua et al.[18] used Kinect to detect a fall by evaluating the contraction and the expansion speed 3D human bounding box. However, all of these methods mentioned above are on the fall detection rather than the fall prediction. Although the detection can help medical staffs to provide medical treatments in time, it cannot predict and prevent elders from fall injuries.

In our work, we employ Microsoft Kinect sensor and RNN approach to build our system. Kinect can provide depth image to help analyze human actions more precisely. RNN brings a new dawn to deal with sequence data classification. Moreover, kinds of equipment for eldercare are invented to protect elders from fall injures, as shown in Figure 2. If a fall can be predicted before hitting the floor, this equipment would efficiently prevent elders from injuring.

Human balance determining

Fall can be considered as the beginning from an imbalance situation of the human body. Once the imbalance of the human body is detected, the fall is predicted. Imbalance is a key feature of imminent fall. The basic idea of the fall prediction is to determine whether the body balance through the relative position of line of gravity (LOG) and base of support (BOS). Most of the actions in daily life, such as standing and walking, belong to a state of equilibrium. Fall is an action of imbalance; sitting down and lying are not equilibrium either. If the imbalance happens, we will try to find whether he will fall or not.

Figure 2. An airbag created for elder by Prop, a Tokyo-based company, which can inflate in 100 ms, if the user starts moving quickly toward the ground.[19]

The elders usually move slowly and changes in a small range. So the elders' fall process, being different from the young's, can be approximately considered as the action from losing the static body balance to hitting the floor. The body balance refers to the capacity that the individual maintains the center of gravity (COM)'s the vertical projection (LOG) in the region of support area of the human body (BOS). Therefore, if we can get the COM and the BOS, respectively, the body's balance can be determined, as shown in Figure 3. Therefore, the body's balance state can be determined once getting the COM and the BOS.

The COM cannot directly get from sensors, but skeleton data from Kinect can help to estimate these positions. It is a line of motion sensing input devices by Microsoft, which provide the depth data and skeleton data. The COM can be computed by skeleton joints position and Dempster's body segment parameters.[20]

We improve the methods[21] to get the COM of human from two dimensions to three dimensions. Our methods include two steps, which we are going to present as follows. In the first step, we compute the COM of each body's segments by following formulas

$$x_{sCOM} = x_p l_p + x_d l_d \quad (1)$$

$$y_{sCOM} = y_p l_p + y_d l_d \quad (2)$$

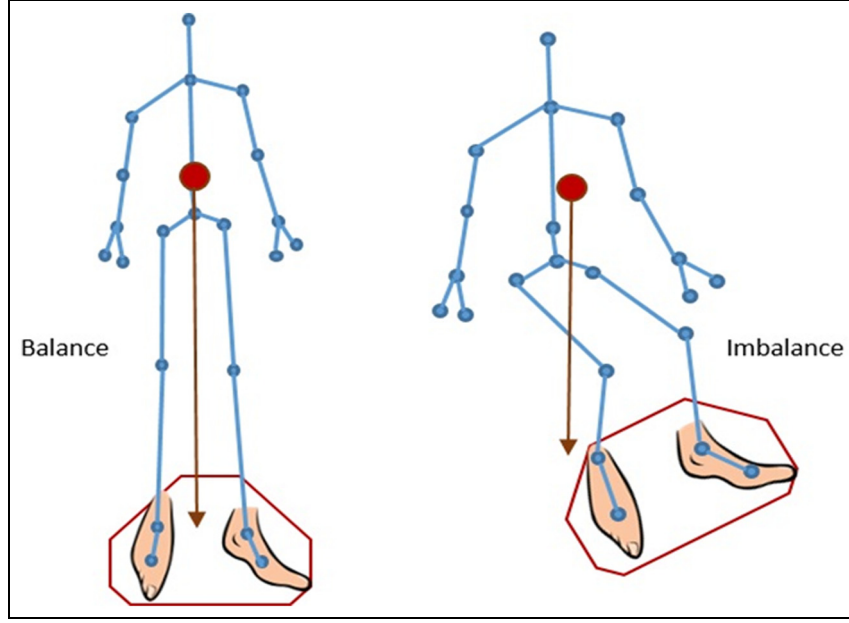$$z_{sCOM} = z_p l_p + z_d l_d \quad (3)$$

Figure 3. The method to determine human balance.

where x_{sCOM}, y_{sCOM}, and z_{sCOM} are coordinates of the COM of segmental body; x_p, y_p, and z_p are coordinates of proximal ends; x_d, y_d, and z_d are coordinates of distal ends; and the percentage of segmental length from the proximal and distal ends are represented by l_p and l_d.

The second step is to compute the COM of all body by following formulas

$$x_{COM} = \frac{\sum m_i x_{s_i com}}{M} \qquad (4)$$

$$y_{COM} = \frac{\sum m_i y_{s_i com}}{M} \qquad (5)$$

$$z_{COM} = \frac{\sum m_i z_{s_i com}}{M} \qquad (6)$$

where x_{COM}, y_{COM}, and z_{COM} are coordinates of body; m_i is the mass of the ith segment; M is the whole mass of body. All the parameters mentioned above can be obtained from Table 1.

Kinect can extract 25 three-dimensional skeleton joints in 30 frames per second (fps) of the human actions. The data of proximal and distal ends of each body segment can be gathered from corresponding human skeleton joints. According to the formulas mentioned above, COM positions can be estimated. Moreover, the LOG can be computed by vertical projection of COM. The BOS is estimated by the polygon covering humans' feet skeleton. The situation of static balance is determined by whether the LOG in a BOS or not, as shown in Figure 4.

Table 1. The mass and coefficients of human segments.[20]

Body segment	Segmental mass/ total body mass (m/M)	COM/segment length	
		Proximal	Distal
Head and neck	0.081	1.000	N/A
Thorax and abdomen	0.355	0.500	0.500
Upper arms	0.028	0.436	0.564
Forearms	0.016	0.430	0.570
Hands	0.006	0.506	0.494
Pelvis	0.142	0.105	0.895
Thighs	0.100	0.433	0.567
Legs	0.0465	0.433	0.567
Feet	0.0145	0.500	0.500

COM: computing positions of center of mass.

Fall prediction based on LSTM

The fall can be seen as a sequence action from imbalance to hitting the floor. Therefore, we can study the fall prediction by supervised learning labeling sequence with RNNs. In this section, we first review RNNs and LSTM, and then we propose an approach of prediction using body imbalance features based on LSTM.

RNN and LSTM networks

The RNN is a very powerful method for dealing with classification for sequence data. The structure of RNN is illustrated in Figure 5.

In this structure, there are three main parts: the input of RNN is $x = (x_0, x_1, x_2, \ldots, x_t)$; the hidden state
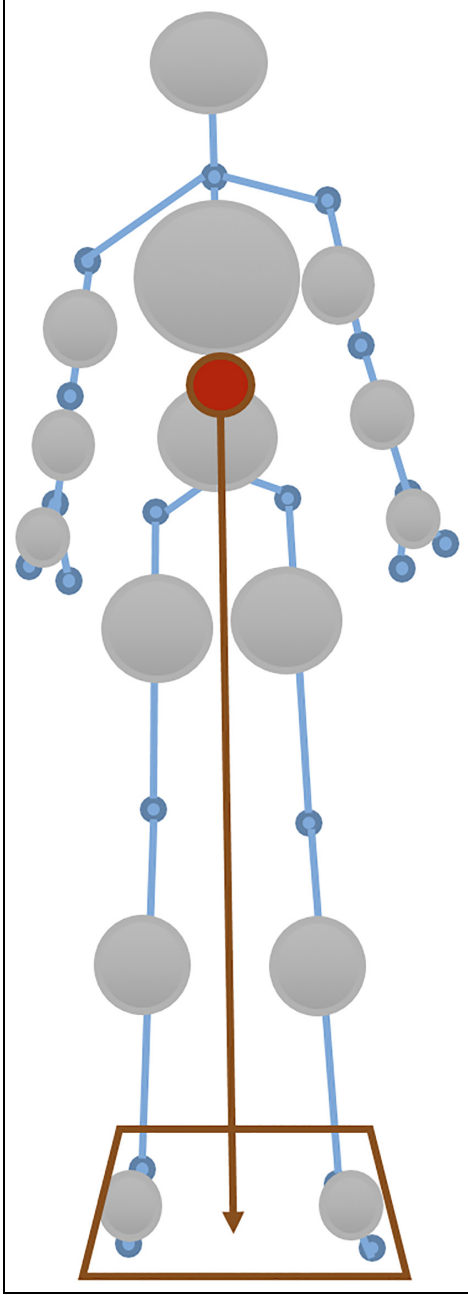
Figure 4. Computing COM, LOG, and BOS.

is $h = (h_0, h_1, h_2, \ldots, h_t)$; the output sequence is $y = (y_0, y_1, y_2, \ldots, y_t)$ with equations as follows

$$h_t = H(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \qquad (7)$$

$$y_t = O(W_{hy}x_t + b_o) \qquad (8)$$

where W_{xh}, W_{hh}, and W_{hy} denote, respectively, weight metrics from input x to hidden state h, hidden state h_t to next hidden state h_{t+1}, and hidden state h to output y, and b denotes bias vectors. It is difficult to train conventional RNN due to vanishing and exploding gradient problems.
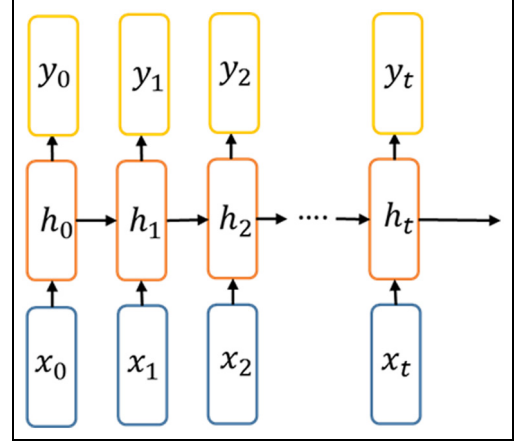

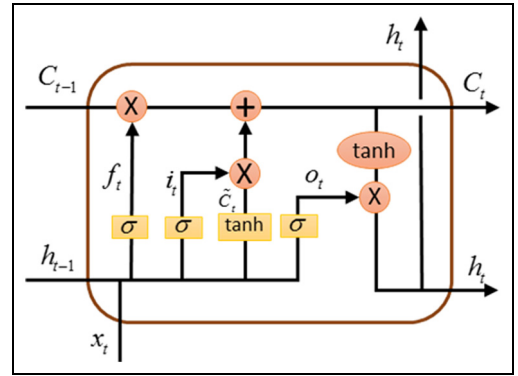**Figure 5.** The structure of RNN.

Figure 6. The structure of LSTM in sequence.[24]

The LSTM has been introduced by Hochreiter and Schmidhuber,[22] which have become a crucial ingredient in recent advances with recurrent networks since they are good at learning long-range dependencies[10] and not affected by vanishing and exploding gradient problems. It introduces a new structure called a memory cell, which is composed of four main elements: an input gate, a forget gate, an output gate, and cell activation vectors. The gates serve to modulate the interactions between the memory cell itself and its environment. The input gate allows incoming signal to alter the state of the memory cell or block it. The output gate allows the state of the memory cell to have an effect on other neurons or prevent it. The forget gate modulates the memory cell's self-recurrent connection, allowing the cell to remember or forget its previous state.[23]

Equations (9)–(14) describe how a model of LSTM works, as shown in Figure 6

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \qquad (9)$$

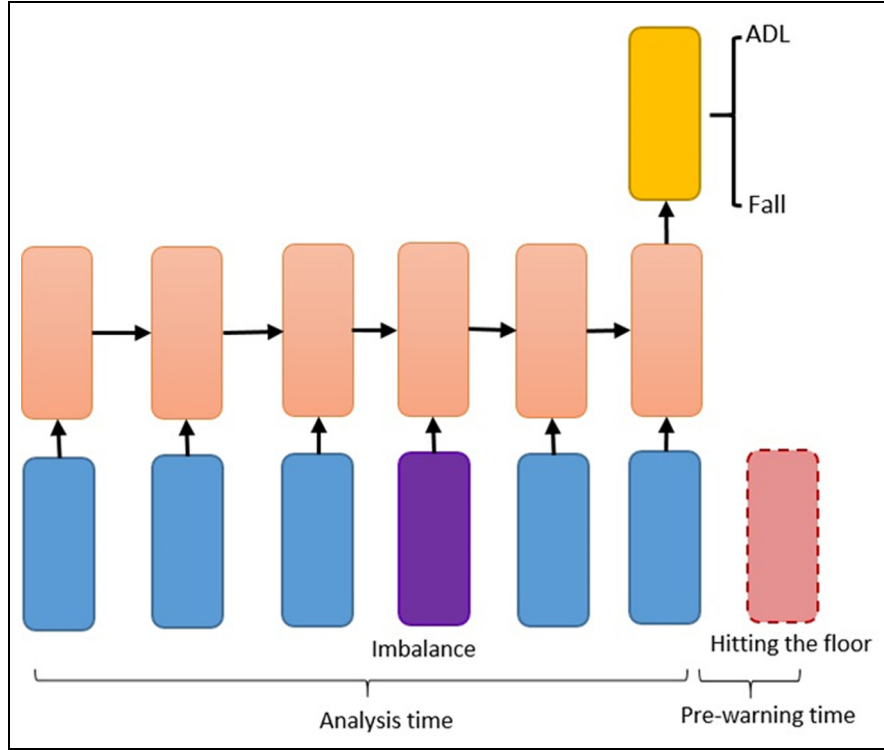$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \qquad (10)$$

Figure 7. Overview of the fall prediction based on LSTM.

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{11}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{12}$$

$$o_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \tag{13}$$

$$h_t = o_t * \tanh(C_t) \tag{14}$$

where $i, f, o,$ and C denote input gate, forget gate, output gate, and cell activation vectors, respectively, and σ denotes the logistic sigmoid function.

Fall prediction based on biomechanics equilibrium

As mentioned above, we estimate the balance situation of human body by calculating the LOG and the BOS. When becoming imbalance, human is likely to fall down.

We use a method to convert the fall prediction to the classification problem. The imbalance situation is seen as the trigger. When the body is in the state of imbalance, several frames before and after the imbalance frame are gathered as the input for the LSTM. The duration from the normal state to several frames before hitting the floor is the process as shown in Figure 7. Each frame of three-dimensional skeleton joints data is regarded as one input vector. Choosing frames around imbalance frame, the LSTM classifies the fall from other states. Due to leaving a pre-warning time, this approach can predict the fall before hitting the floor.

Table 2. The survey of the duration of the fall.

Types of the fall	Mean duration (ms)
Back	767
Front	778
Side	533
Other	767
Mean duration of all types	746

The next step is to set the length of the input. The fall is a process of serial actions, from the imbalance to hitting the floor. We choose frames during before and after imbalance frame imbalance as the input. At first, it requires knowing the length of the duration of the fall. We surveyed existing fall databases[15,25] and computed the mean duration from the imbalance to different types of the fall, as shown in Table 2. The mean duration of all types of the fall is 746 ms. Because Kinect can record skeleton data by 30 frames per second (fps), in other words, the fall lasts about 22.4 frames in average. Then we chose the period from stable state to 10 frames and 5 frames before hitting the floor as the input to evaluate our method.

Results and discussion

We test our fall prediction algorithm on database build by Rougier et al.[5] It involves 11 healthy volunteers from

22 to 39 years old with different heights and weights. Kinect is employed to record human's actions, which are grouped into two main categories: ADLs and the fall. ADLs has four types of actions: sit, grasp, walk, and lay, and fall has four types of actions depending on direction: front, back, side, and end-up-sit, as shown in Table 3. Every action is recorded in database three times from 11 health volunteers, so it contains 264 different actions.

The skeleton data of actions are used to test our algorithm. We randomly choose 224 actions as training set and the remaining are testing set. We test algorithm by actions sequential data including 10 frames, 5 frames before hitting the floor, hitting the floor and 5 frames after hitting the floor, as shown in Figure 8.

Table 3. Database protocol.

Type	Action	Description
ADLs	Sit	Sit on a chair
	Grasp	Walk and grasps an object from the floor
	Walk	Walk back and forth
	Lay	Lie down on the mattress
Fall	Front	Fall from the front and ends up lying
	Back	Fall backward and ends up lying
	Side	Fall to the side and ends up lying
	End-up-sit	Fall backward and ends up sitting

ADLs: activity of daily living.

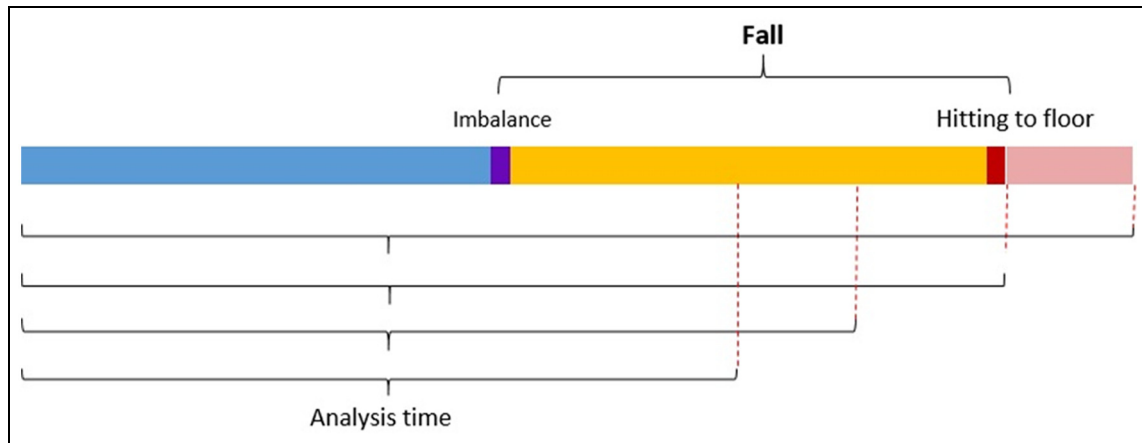The performances of method based on human biomechanics equilibrium are shown in Figure 9. Our

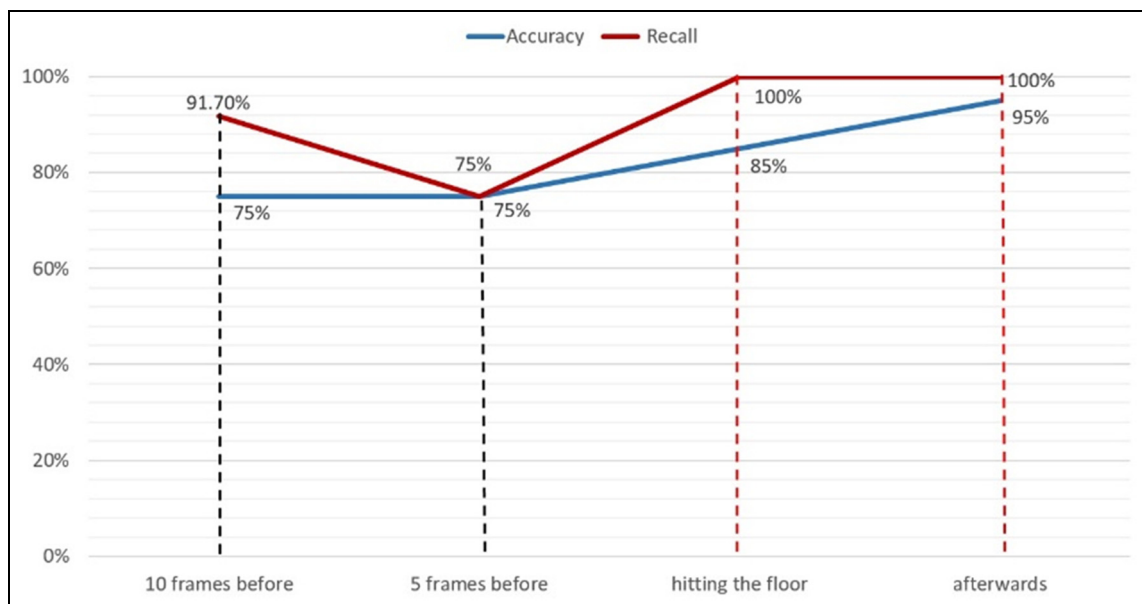**Figure 8.** Input duration analysis.

Figure 9. Performance of fall prediction.

method can achieve 75% for action classification and 91.7% fall prediction in 10 frames before hitting to the floor, which can provide 333 ms for pre-warning time. If input sequential data including frames after hitting the floor, that is fall detection, our method can get achieve 95% for action classification and detect 100% fall.

Conclusion

The fall is a major public health problem among older people.[2] Microsoft Kinect device can get depth image and provide skeleton data of human, which can be used to monitor elders' activities at home. Through Kinect, we obtained skeleton data of human and described human postures by 25 3D joints data. In this article, we employed skeleton data to analyze the whole process of the fall based on human biomechanics equilibrium and proposed a fall predication method based on LSTM by Kinect. Different from previous studies to detect the fall by human postures after falling, we choose the data before and after human imbalance state as the input and employed LSTM to distinguish the state from a fall to ADLs before hitting the floor. The performances show that this method could predict 91.7% fall with a pre-warning time of 333 ms before hitting the floor. This work lays the foundation for the studies on the fall protection.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

1. Noury N, Fleury A, Rumeau P, et al. Fall detection—principles and methods. In: *Proceedings of the 29th annual international conference of the IEEE engineering in medicine and biology society (EMBS 2007)*, Lyon, 23–26 August 2007, pp.1663–1666, ftp://ftp.prip.tuwien.ac.at/pub/outgoing/zamba/fallpapers/Noury07.pdf

2. Igual R, Medrano C and Plaza I. Challenges, issues and trends in fall detection systems. *Biomed Eng Online* 2013; 12: 66.

3. Foroughi H, Aski BS and Pourreza H. Intelligent video surveillance for monitoring fall detection of elderly in home environments. In: *Proceedings of the 11th international conference on computer and information technology*, Khulna, Bangladesh, 24–27 December 2008, pp.219–224. New York: IEEE.

4. Cucchiara R, Grana C, Prati A, et al. Probabilistic posture classification for human-behavior analysis. *IEEE T Syst Man Cy* 2005; 35: 42–54.

5. Rougier C, Opérationnelle ER and Meunier J. Demo: fall detection using 3D head trajectory extracted from a single camera video sequence. *Int J* 2007; 2–3, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.482.9227&rep=rep1&type=pdf

6. Sixsmith A and Johnson N. A smart sensor to detect the falls of the elderly. *IEEE Pervas Comput* 2004; 3(2): 42–47.

7. Mohamed A, Dahl GE and Hinton G. Acoustic modeling using deep belief networks. *IEEE T Audio Speech* 2012; 20(1): 14–22.

8. Mikolov T and Zweig G. Context dependent recurrent neural network language model. In: *Proceedings of the 2012 IEEE spoken language technology workshop (SLT)*, Miami, FL, 2–5 December 2012, pp.234–239. New York: IEEE.

9. Shi Y, Zhang WQ, Liu J, et al. RNN language model with word clustering and class-based output layer. *EURASIP J Audio Speech Music Process* 2013; 2013(1): 22.

10. LeCun Y, Bengio Y and Hinton G. Deep learning. *Nature* 2015; 521(7553): 436–444.

11. Kinect, https://en.wikipedia.org/w/index.php?title=Kinect&oldid=754197218 (Wikipedia)

12. Tamura T, Yoshimura T, Sekine M, et al. A wearable airbag to prevent fall injuries. *IEEE T Inf Technol B* 2009; 13(6): 910–914.

13. Cheng J, Chen X and Shen M. A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals. *IEEE J Biomed Health Inform* 2013; 17(1): 38–45.

14. Zhang T, Wang J, Xu L, et al. Fall detection by wearable sensor and one-class SVM algorithm. In: Huang D-S, Li K and Irwin GW (eds) *Intelligent computing in signal processing and pattern recognition*. Berlin, Heidelberg: Springer, 2006, pp.858–863.

15. Rougier C, Meunier J, St-Arnaud A, et al. Robust video surveillance for fall detection based on human shape deformation. *IEEE T Circ Syst Vid* 2011; 21(5): 611–622.

16. Vishwakarma V, Mandal C and Sural S. Automatic detection of human fall in video. In: *Proceedings of the 2nd international conference on pattern recognition and machine intelligence*, Kolkata, India, 18–22 December 2007, pp.616–623. Berlin: Springer.

17. Mastorakis G and Makris D. Fall detection system using Kinect's infrared sensor. *J Real Time Image Process* 2014; 9(4): 635–646.

18. Bevilacqua V, Nuzzolese N, Barone D, et al. Fall detection in indoor environment with kinect sensor. In: *Proceedings of the 2014 IEEE international symposium on innovations in intelligent systems and applications (INISTA)*, Alberobello, 23–25 June 2014, pp.319–324. New York: IEEE.

19. Japanese company makes airbags for the elderly, http://www.telegraph.co.uk/news/newstopics/howaboutthat/

3076551/Japanese-company-makes-airbags-for-the-elderly. html

20. Winter DA. *Biomechanics and motor control of human movement*. New York: John Wiley & Sons, 2009.

21. Yeung LF, Cheng KC, Fong CH, et al. Evaluation of the Microsoft Kinect as a clinical assessment tool of body sway. *Gait Posture* 2014; 40(4): 532–538.

22. Hochreiter S and Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9(8): 1735–1780.

23. LSTM networks for sentiment analysis—DeepLearning 0.1 documentation, http://deeplearning.net/tutorial/lstm. html

24. Understanding LSTM networks—colah's blog, http:// colah.github.io/posts/2015-08-Understanding-LSTMs/

25. Gasparrini S, Cippitelli E, Gambi E, et al. Proposal and experimental evaluation of fall detection solution based on wearable and depth data fusion. In: Loshkovska S and Koceski S (eds) *ICT innovations 2015*. Cham: Springer, 2016, pp.99–108.