Deep stair walking detection using wearable inertial sensor via long short-term memory network

Gan Wei Nie, Nurul Fathiah Ghazali, Norazman Shahar, Muhammad Amir As’ari
School of Biomedical Engineering and Health Sciences, Universiti Teknologi Malaysia, Malaysia
Sport Innovation and Technology Center (SITC), Institute of Human Centered Engineering (iHumEn), Universiti Teknologi Malaysia, Malaysia

ABSTRACT
This paper proposes a stair walking detection via Long Short-Term Memory (LSTM) network to prevent stair fall event happen by alerting caregiver for assistance as soon as possible. The tri-axial accelerometer and gyroscope data of five activities of daily living (ADLs) including stair walking is collected from 20 subjects with wearable inertial sensors on the left heel, right heel, chest, left wrist and right wrist. Several parameters which are window size, sensor deployment, number of hidden cell unit and LSTM architecture were varied in finding an optimized LSTM model for stair walking detection. As the result, the best model in detecting stair walking event that achieve 95.6% testing accuracy is double layered LSTM with 250 hidden cell units that is fed with data from all sensor locations with window size of 2 seconds. The result also shows that with similar detection model but fed with single sensor data, the model can achieve very good performance which is above 83.2%. It should be possible, therefore, to integrate the proposed detection model for fall prevention especially among patients or elderly in helping to alert the caregiver when stair walking event occur.

Keywords:
Confusion matrix
Deep learning
LSTM network
Stair walking
Wearable inertial sensor

1. INTRODUCTION
Among so many indoor activities of daily living (ADL), stair walk is the one that have a major potential and hazardous for people to falls especially elderly. It might cause significant injuries such as hip fracture, traumatic brain injuries (TBI) and deaths. The injury Facts 2015 [1] of national safety council's statistical report had reports that there are over one million injuries and 12000 deaths are cause by stairway accidents each year. Boyé et.al [2] also had investigated on falls rate of elderly populations in Netherland and 409 out of total of 5880 fall-related Emergency Department visit is due to walking up or down stairs. Among the indoor activities, falls during stair walking event have the highest percentage to sustain TBI which is 52% for women and 61% for men. Not only that, the study from Hwang, et.al [3] had shown that elderly are 3 times more likely to suffer from TBI after stair falls when compare to normal fall while walking. Another study in Malaysia by Sazlina et.al [4] also shows that 61% of elderly falls indoor and 57% of them experienced recurrent falls. The most common indoor places that elderly falls are stairs and bathroom which is 27% both. The factor that lead to stair falls can be classify into two which are host-related factors and environment-related factors. Host-related factor is factor that contributes by health condition...
of host [5] for example decline in muscle strength [6-7], ability of posture control [8], cognitive factor [9], visual condition [10], and obesity [11]. An environment-related factor is the factors that cause by environment of host which include stair architecture design, and stair obstacles such as absence of handrail, irregular riser height and object left on stairs. These factors will force the staircase user to use their maximal capabilities to walk up or down the stairs as greater body posture control effort is required.

Since there is high risk to fall and lead to serious injuries, human activity recognition (HAR) is developed as a part of a framework to automatically monitor elderly activities and reduce the burden of caregiver. Most of the previous studies [12-14] embedded common machine learning model that integrated with shallow and human crafted features extraction approaches which could only able to recognize low level activities. Recently, deep learning model has been formulated in HAR related studies [15-19] to overcome the limitation of common machine learning approaches. However, to our knowledge, there has been lack of research conducted on detecting stair walking event from other activities using wearable inertial sensor via deep learning approach as well as detect stair falls. Therefore, a stair walking detection is proposed to prevent stair fall by detect stair walking activities as well as other daily activities using inertial sensor and implement into the LSTM network. This can reduce the burden of caregiver by alerting caregiver as soon as any stair walk activity is detected before any stair fall happen.

2. LSTM NETWORK OVERVIEW

Generally, deep learning is an ideal approach for HAR as the property of deep learning able to solve the limitations of machine learning. It able to extract features automatically, recognize complex high-level activities and reduce computational cost. LSTM is a composition from Recurrent Neural Network (RNN) and it is capable of capturing long term dependencies with a lot of memory units called cells [20]. Figure 1 shows LSTM cell.

![LSTM cell](image)

Figure 1. LSTM cell [21]

LSTM network have lot of memory cell composed in it and this large stack of memory cell property enable it to learn complex input. Inside a memory cell unit consists an input gate \( i \), output gate \( o \) and a forget gate \( f \) as in Figure 1. All these gates unit will regulate the content of memory cell that flow in and out of the cell. A memory cell \( c \) will connected to another cell. The forget gate in memory cells make LSTM smart enough to decide what to erase from memory and keep only relevant data [22-23]. It removes unnecessary data memory from previous state by multiply with previous cell state as (1) where \( W \) is rectangular input weight matrices; \( b \) is the bias vector and \( x \) is the input vector.

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

(1)

Input gate is function to add new input to present cell state. (2) will decide which values to be updated and (3) will create vector for new candidate values.

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

(2)

\[
\hat{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)
\]

(3)
The present cell state is calculated by using (4)

\[ C_t = f_t * C_{t-1} + i_t * C_t \]  

A sigmoid layer is run to decide which parts of cell state is going to output. The cell state is tanh and multiplies with sigmoid gate output by using (5) and (6).

\[ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \]  

\[ h_t = o_t * \tanh(C_t) \]  

3. EXISTING WORKS IN WEARABLE SENSOR BASED HAR

There are many fall detector that were developed to solve the fall problem on elderly. The study of Ozdemir, et.al [24] had also successfully distinguished falls from ADLs that can cause high acceleration body part such as jumping, sitting down suddenly and running. 6 machine learning approach which are k-Nearest Neighbor (k-NN), Least Square Method (LSM), Bayesian Decision Making (BDM), Support Vector Machine (SVM), Artificial Neural Network (ANN) and Dynamic Time Warping (DTW) was use as the classifier and the performance of each classifier is compare. In terms of the required training time, the classifiers can be sorted as BDM, LSM, DTW, k-NN, SVM, and ANN in increasing order, whereas in term of the testing time, the order is ANN, SVM, LSM, BDM, k-NN, and DTW. The accuracy has achieved above 95% for all classifier.

Steven Eyobu et.al [18] had proposed a human activities recognition ranging from walking, sitting, falling, climbing and stair walking. LSTM neural network is proposed in the model to solve the issue that difficulty in discriminate amongst high similarity features. The proposed approach is Deep Recurrent Neural Network (DRNN). This approach has an advantage in high throughput which is short recognition time and able to discriminate activities that have almost similar features.

There is also few works that able to recognize complex instrumental activities daily life (ADL). For example, the A-Wristocracy wrist worn sensing recognition [20] that proposed by Vepakomma, et.al. The proposed A-Wristocracy recognition system are able to recognized fine-grained 22 indoor activities by multi-modal sensors which consists of accelerometer, gyroscope, ambient location context sensing and atmospheric environmental sensors. The 22 complex fine-grained activity is contexts into various classes which are locomotive, semantic, transitional and postural. All the test accuracy for various number of neurons in hidden layers had achieve testing accuracy above 84%. Panwar et. al [16] also had proposed a HAR recognition model using Convolutional Neural Network (CNN) to recognize 20 small actions in making a cup of tea. A single wrist worn tri-axial accelerometer is used in the study and detect with extension and flexion of forearm, rotation of forearm and rotation of the wrist about long axis of forearm. This study has achieved a performance accuracy of 99.8%.

4. RESEARCH METHOD

The workflow of this study is comprises of five steps which are data acquisition, data pre-processing, LSTM network architecture implementation, dataset training and testing, and performance evaluation. All the steps will explain further in the following subsections.

4.1. Data acquisition

Gait Up Physilog 5 Inertial measurement sensor unit as in Figure 2 is used in this study. This wearable inertial sensor sensing abilities includes 3D accelerometer, 3D gyroscope and barometric sensor. However, only 3D accelerometer and 3D gyroscope are used in this project. The data was collected by placing the inertial sensor on subject’s chest, wrists, and heels. The sampling frequency of Physilog 5 sensor is 128 Hz.

20 subjects were involved in this study and each subject was asked to perform few daily living activities. The daily living activities is including stair walking, walking, sitting, standing and laying down as in Figure 3. All the activities were performed continuously at a subject comfortable speed. The same activity set was repeated by each subject for 3 times at different stairs.
Deep stair walking detection using wearable inertial sensor via long short-term memory

4.2. Data pre-processing

Before the data is used for training, the collected raw data is required to undergo preprocessing. The collected raw data was first labelled. Walking is represented by “1”; stair walking is represented by “2”; sitting is represented by “3”; lying down is represented by “4”; and standing is represented by “5”. After that, a windowing technique is applied through the labelled data to take small subset through this large dataset. The window sizes applied to the dataset was 0.5 sec, 1 sec, 1.5 sec and 2 secs. The data after the windowing was then named as X and the labeling named as Y. Data X and data Y was divided into three parts which are 90% for training set and 10% for test set. Training set was used to fit the LSTM network model and testing set was used to evaluate the final LSTM network model.

4.3. LSTM network architecture implementation

LSTM network was implemented using deep learning toolbox of MATLAB 2018a. Figure 4 shows the framework of LSTM network. There will have several layers in the LSTM network which include an unknown n layer of LSTM hidden unit layers, fully connected layer, softmax layer and classification layer.

Figure 2. Gait up physilog 5 inertial measurement sensor unit

Figure 3. Daily living activities, (a) Walking, (b) Standing, (c) Stair walking, (d) Laying down, (e) Sitting

Figure 4. The framework of LSTM network
The LSTM network training setting is as Table 1. The optimizer solver used is Adam which decay rates can be specified. Gradient threshold is set to 1 so that can prevent gradient from exploding. The initial learning rate is set to 0.01 and drop in period of 20. This is to allow the network have large change in the beginning of the training and decrease the learning rate over the epochs to have smaller tuning changes at the training later. The regularization is set to L2 to control the model capacity and reduce overfitting of the model. The input signal dimension is 30 and the output dimension is 5.

<table>
<thead>
<tr>
<th>Settings</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output Layer Activation Function</td>
<td>Softmax</td>
</tr>
<tr>
<td>Optimizer Solver</td>
<td>Adam</td>
</tr>
<tr>
<td>Gradient Threshold</td>
<td>1</td>
</tr>
<tr>
<td>Max Epochs</td>
<td>30</td>
</tr>
<tr>
<td>Initial learning rate</td>
<td>0.01</td>
</tr>
<tr>
<td>Regularization</td>
<td>L2</td>
</tr>
<tr>
<td>Input signal dimension/ feature dimension</td>
<td>30</td>
</tr>
<tr>
<td>Output dimension/class</td>
<td>5</td>
</tr>
</tbody>
</table>

4.4. Data training and testing

As mention in the subtopic before, there were 5 sensors wear at each subject’s chest, wrists, and heels and each sensor will have 6 features (tri-axial accelerometer and tri-axial gyroscope). Thus, the total features dimension of the dataset was 30. The input data sequence will directly load into the layers for training purpose without feature extraction. This is due to the deep learning approach has automatic feature extraction property.

4.5. Performance evaluation

The performance will be evaluated using confusion matrix plot in Table 2 that consist of two dimensions which are actual and predicted. True positive, true negative, false positive and false negative can know from the confusion matrix.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>True Positive</td>
</tr>
<tr>
<td>False Positive</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

Sensitivity = \frac{TP}{TP + FN}
Specificity = \frac{TN}{TN + FP}
Accuracy = \frac{TP + TN}{TP + TN + FP + FN}

5. RESULT AND DISCUSSION

As mentioned in the previous chapter, there are several parameters need to be varied in searching an optimum walking detection model which produced highest accuracy value. The parameters are sliding window size, sensor deployment, number of hidden cell unit and LSTM architecture (either single layered or double layered LSTM). The evaluation process is divided into several stages.

5.1. Window size varied (stage 1)

At the first stage, the window size is varied while fixing the sensor deployment which used input data from all attached sensors on single layered LSTM networks with 100 hidden cell units. Table 3 summarizes the LSTM network models accuracy of varying window size while fixing other parameters. The results obtained shows that window size of 2 seconds give the best performance when compared to the window size of 0.5 seconds, 1 second and 1.5 seconds. Thus, window size of 2 seconds was used for the next stage.

<table>
<thead>
<tr>
<th>Window Size (seconds)</th>
<th>Testing Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>74.0</td>
</tr>
<tr>
<td>1.0</td>
<td>76.1</td>
</tr>
<tr>
<td>1.5</td>
<td>77.7</td>
</tr>
<tr>
<td>2.0</td>
<td>79.4</td>
</tr>
</tbody>
</table>
Deep stair walking detection using wearable inertial sensor via long short-term memory... (Gan Wei Nie)

5.2. LSTM architecture varied (stage 2)

In the second stage, the LSTM network architecture is varied while fixing the sensor deployment which used input data from all attached sensors and also fixing the window size at 2 seconds. Table 4 summarized the LSTM network model accuracy of 2 seconds window size for all sensors data deployed in different architecture. The results show that the testing accuracy has increase with the increase number of LSTM hidden cell unit. Also, double layered LSTM model have higher accuracy than single layered LSTM as demonstrated in [25]. This is because the capacity of LSTM network is increase and train in a deeper way when the number of LSTM layer is increase. Among the trained LSTM architecture, double layered LSTM network with 250 hidden cell units per layer produced the best performance. Thus, double layered LSTM network with 250 hidden cell units per layer is used in the next stage.

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Number of LSTM Hidden Cell Unit per Layer</th>
<th>Single</th>
<th>Double</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Sensors</td>
<td>100</td>
<td>79.4%</td>
<td>80.6%</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>92.6%</td>
<td>96.5%</td>
</tr>
</tbody>
</table>

5.3. Varying input sensor data (stage 3)

In the third stage, the sensor input data used is varied while fixing the window size at 2 seconds and LSTM network architecture at double layered with 250 hidden cell units per layer. Table 5 summarized the LSTM network models accuracy of different sensor at 2 seconds window size with double layered architecture of 250 hidden cell units per layer. The result also shows that the LSTM network model that feed with all sensor data have greater performance than single sensor data. It can be interpreted that the performance of LSTM network for stair walking detection is affected by the number of sensor data that fed into network. The more input data, the better the LSTM performance which is same as stated in [26]. It can also be interpreted that sensor data from chest, right heel and left heel majorly contribute in producing great accuracy for stair walking detection while right wrist and left wrist provide the least performance. This might due to the hand movement is unpredictable and in randomize direction while performing the activities of daily living. Thus, the produced sensor data from both wrists is very difficult to discriminate between an activity to another activity.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Chest</th>
<th>Right Heel</th>
<th>Left Heel</th>
<th>Right Wrist</th>
<th>Left Wrist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing Accuracy (%)</td>
<td>87.8%</td>
<td>87.7%</td>
<td>89.7%</td>
<td>86.6%</td>
<td>83.2%</td>
</tr>
</tbody>
</table>

5.4. The best LSTM network

The best LSTM network model is the double layered LSTM model with 250 hidden units per layer which showed the best testing accuracy for dataset with 2 seconds window size. The testing accuracy obtained is 95.6% as in Figure 5 and the error rate is 3.5%. From the same confusion matrix obtained, we also can know that the sensitivity to detect stair walking event (class 1) correctly is 97.9%. Only 1 out of 424 and 8 out of 424 is classifying wrongly as standing (class 4) and walking (class 0) respectively. The positive predicted value is 97% which have 1 out of 188 from standing event and 12 out of 442 from walking event had wrongly detected as stair walking event. There is none of the Laying down event (class 3) and sitting event (class 2) wrongly classify into it. In terms of specificity, activities lying down, sitting, standing and walking have 97.3%, 95.1%, 94.1% and 96.6% respectively. This means that, there is also have a good performance in detect activities other than stair walking correctly. For negative predicted value, activities lying down, sitting, standing and walking have 96.3%, 96.6%, 95.2% and 96.6% respectively. All the activities that other than stair walking do not have much wrong classify from other classes into it.
Figure 5. Confusion matrix that obtained from the best LSTM network model

6. CONCLUSION

In conclusion, a set of tri-axial accelerometer and tri-axial gyroscope data was collected and new activities dataset was created for model training and testing. Deep structured LSTM network models was implemented to detecting stair walking event as well as other activities of daily activities. Based on the results obtained, it shows that the window size of 2 seconds gives the best performance when compared to 0.5, 1, and 1.5 seconds. In the second stage, the result shows that the testing accuracy increase with number of hidden units and double layered LSTM give better performance than single layered. The best accuracy is at double layered LSTM with 250 hidden units per layer which is 96.5%. From the third stage, the testing accuracy of each single sensor have achieved above 83.2%. It also shows that the stair walking event has higher dependency on left heel, chest and right heel. This LSTM model can be further implemented into an automated stair walking detection system that can detect and alert the caregivers when stair walking event occurs on elderly or patient. The burden of caregivers can be reduced and stair falls on elderly or patient can be prevented by using this trained model.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Universiti Teknologi Malaysia (UTM) and the Minister of Education (MOE), Malaysia for supporting this research work under FRGS Research Grant No. R.J130000.7851.5F108.

REFERENCES


**BIOGRAPHIES OF AUTHORS**

Gan Wei Nie currently finished her study in Bachelor of Engineering (Bio-medical) at Universiti Teknologi Malaysia. Her research interest is about human activity recognition using machine learning and deep learning.

Nurul Fathiah Ghazali received Diploma in Electrical Engineering (Power) from Universiti Teknologi Malaysia, Kuala Lumpur (2013) and Bachelor of Engineering in Biomedical from Universiti Teknologi Malaysia, Johor Bahru (2017). Her research interest includes activity recognition studies, signal processing, and machine learning. She is currently working on her PhD in Biomedical Engineering at Universiti Teknologi Malaysia, Johor Bahru and her current research is about activity recognition study in sport using inertial sensor which collaborated with Sport Innovation and Technology Centre (SITC) and Institute of Human Centered Engineering (iHumEn) from UTM JB.

Deep stair walking detection using wearable inertial sensor via long short-term memory... (Gan Wei Nie)
Norazman Shahar received his Bachelor of Engineering (Bio-Medical) from Universiti Teknologi Malaysia, Johor Bahru (2017). His research interest includes accelerometer sensor, activity recognition and signal processing. Currently, he is working on his Doctor of Philosophy (PhD) in Bio-Medical Engineering Universiti Teknologi Malaysia, collaborated with Sport Innovation and Technology Centre (SITC) and Institute of Human Centered Engineering (iHumEn). His current research is concentrating on the inertial sensor, activity recognition, signal processing and machine learning.

Muhammad Amir bin As’ari holds a PhD in Biomedical Engineering from Universiti Teknologi Malaysia. His PhD’s work is in the field of assistive technology, computer vision, and image processing and his work focused on developing a novel 3D shape descriptor for recognizing the activities of daily living (ADLs) based on Kinect-like depth image. He pursued his master and bachelor degree at Faculty of Electrical Engineering in UTM. Currently, he is working on development of signal and image processing with intelligent intervention and human action recognition based on wearable sensor and context-aware modality for assistive and sport technology. Currently, he is a senior lecturer in Universiti Teknologi Malaysia specialized in Biomedical Engineering area. He is also member of Biomedical Instrumentation & Electronics (BMiE) Research Group and Associate fellow researcher of Sport Innovation & Technology Centre (SITC).