Using Bi-directional Long Short-Term Memory Recurrent Neural Network (Bi-LSTM RNN) to predict Elderly fall

Tai-Jung Chen

1. Background and Objective

Falling is much more serious for elderly people comparing to the young adults. Elderly falling could have a higher chance of causing bone, muscle, and joint injuries. Not to mention the elderlies have a higher chance of death after falling. On the other hand, the frequency of falling is found to be higher than the young adults as well. According to the reference, 30~60% of the community-dwelling elderly in the U.S. fall each year [1]. Furthermore, more than half of them experience multiple falls [1]. Owing to the severity of elderly fall stated above, preventing elderly falls become an important and urgent problem to deal with.

Poor balance had been cited to be a key casual factor of falling [1]. Fall assessments are a typical way to evaluate a test taker's ability of balance, gait, and fall risk. There are many fall assessment tests in the healthcare domain. These fall assessment tests include Timed Up and Go (TUG) test, Berg Balance Scale (BBS), Short Form Berg Balance Scale (SFBBS), etc. The details of how these fall assessment tests are carried out would be discuss in the Data Description section. While these fall assessment tests are proven to be highly correlated and accurate in terms of evaluating fall risk, some of these assessment tests actually require medical professionals to conduct. This fact makes evaluating fall risks not so timely, and also expensive in terms of time, money, and medical resources.

In order to alleviate the above-mentioned difficulties of accessing the fall assessment tests, the TUG test, which is a relatively simpler fall assessment test that could be conduct by non-medical professionals, is considered. If one can accurately map the outcome of the TUG test to the assessment evaluation given by the medical professionals, such as BBS or SFBBS, we could achieve the goal of providing timely assessment of fall risks, alleviate the medical resource burden, and save both participant and medical professionals' time at the same time. This idea is borrowed from blood glucose test. Traditional blood glucose test is an invasive test, which create small wounds on test takers [2] (figure 1) This nature makes test taker to be unwilling to test their blood glucose level sometimes, resulting problems to properly monitor their blood glucose level. An alternative for this traditional blood glucose test is to use a designed patch to collect sweats from the test taker, and estimate the blood glucose using the sweats, since research showed that certain contents in sweats are highly correlated to blood glucose [3] (figure 2) Another example is carrying out COVID test using saliva instead of using the horrible anterior nasal swab. These alternative tests are created hoping to allay the difficulties and unwillingness of testing.

With the high popularity of wearable devises, such as Fitbit and Apple Watch, nowadays, one opportunity point to solve this problem would be utilizing the data obtained by these devises, and cooperate the data with the knowledge of Data Science (DS).

Motivated by the above idea of developing a surrogate approach to evaluate fall risks for elderly adults, we plan to develop a predictive model that could take the sensor signal captured when the participant is carrying out a TUG test as input, and output a label which indicates whether the subject is classified as a potential faller or non-potential faller. We hope this predictive model could make the fall assessment more timely, cheaper, and simpler to carry out.

Figure 1. Invasive blood glucose test cause test taker to bleed

Figure 2. Using patch to monitor blood glucose level

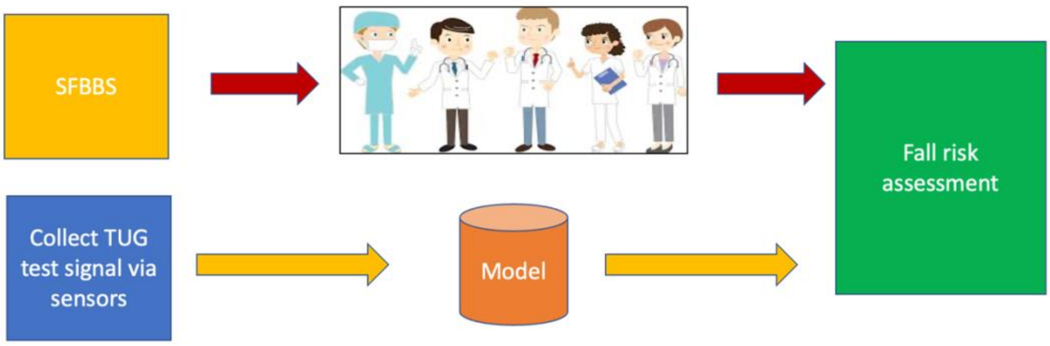

Figure 3. A graphical illustration of the objective for the research

2. Data Description

<u>2-1 TUG test</u>

The TUG test is a self-conductible test that could let the test takers to have a simple idea of how their balance and gait conditions are [4]. The test contains 5 movements. The test taker is asked to sit on a chair in the beginning, and perform the 5 movements (shown in figure 4), which are (1) stand up from the chair, (2) walk straightly to a targeted destination, (3) turn 180 degrees to face the starting point, (4) walk straightly back to the starting point, and (5) sit down on the chair again. The outcome of this test is the completion time. If the completion time is greater than 13.5 seconds, the test taker is classified as in high fall risk.

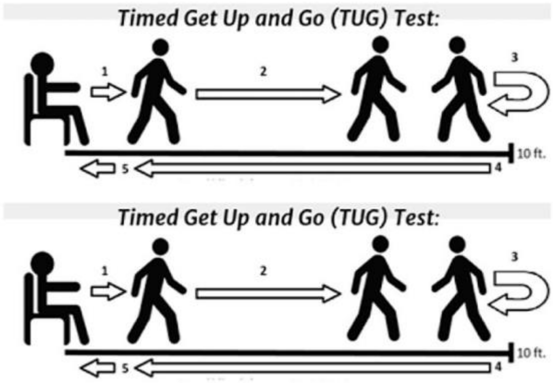

Figure 4. The 5 movements for TUG test

2-2 BBS and SFBBS

BBS is the fall assessment test that have to be carried out by a medical professional [5]. The subjects were asked to perform 14 movements (shown in table 1) individually under the supervision of a medical professional. A score of 0-4 will be assigned for each of the movement, resulting a total score ranged from 0-56. The larger the score is, the lower the participants' falling risks are. Although the criteria of different group of elderly adults are not the same, a score that is under 45 is classified as high fall risk for a general elderly adult, which is the target group of elderly people in this project.

Despite the widely used and recognitions of the BBS on assessing fall risk, some people argue that the procedure is time consuming. Alternatively, SFBBS, which is the simplified version of BBS was chosen in this project. SFBBS has been proven to be as effective as the original BBS [6]. Instead of 14 movements, the participants are asked to perform 7 movements (shown in table 2) individually under the supervision of medical professionals. Again, a score of 0-4 is assigned to each movement by the professionals. A score less than or equal to 23 is classified as at high falling risk for general elderly people.

Table 1. The 14 movements for BBS

Movement	Description
1	Move from a sitting to a standing position
2	Stand up unsupported
3	Sit unsupported
4	Move from a standing to a sitting position
5	Transfer from one chair to another
6	Stand up with your eyes closed
7	Stand with your feet together
8	Reach forward with an outstretched arm
9	Pick an object up off of the floor
10	Turn and look behind you
11	Turn around in a complete circle
12	Place each foot alternately on a stool in front of you
13	Stand unsupported with one food directly in front of the other
14	Stand on one leg for as long as you can

Table 2. The 14 movements for SFBBS

Movement	Description
1	Move from a sitting to a standing position
2	Stand up with your eyes closed
3	Reach forward with an outstretched arm
4	Pick an object up off of the floor
5	Turn and look behind you
6	Stand unsupported with one food directly in front of the other
7	Stand on one leg for as long as you can

2-3 Selected Dataset

The dataset collected by Feng-Yuan hospital in Taiwan was used in this project. These data are experimental data. The data were collected by the team from Yuan Ze University. [7] The target group of the experiment was general elderly adults with no specific chronic disease such as stroke or pulmonary disease. Note that the definition of an elderly adults are the adults whose age is over 65 years old.

2-4 Experimental design

The participants were asked to wear a tri-axis sensor on their waists, and perform the TUG test. (Figure 5) The sensor will then record the acceleration data along the three axes. Figure 6 is a snapshot of how the collected data would look like if we export the data into Microsoft Excel. Figure 7 also shows how the signal would look like after plotted via the software.

On the other hand, the participants were also asked to perform the SFBBS under the supervision of the medical professionals. The SFBBS total scores were reported. By thresholding at 23, we obtained the two types of labels (high fall risk and low fall risk), and pair them with the sensor data mentioned in the previous paragraph.

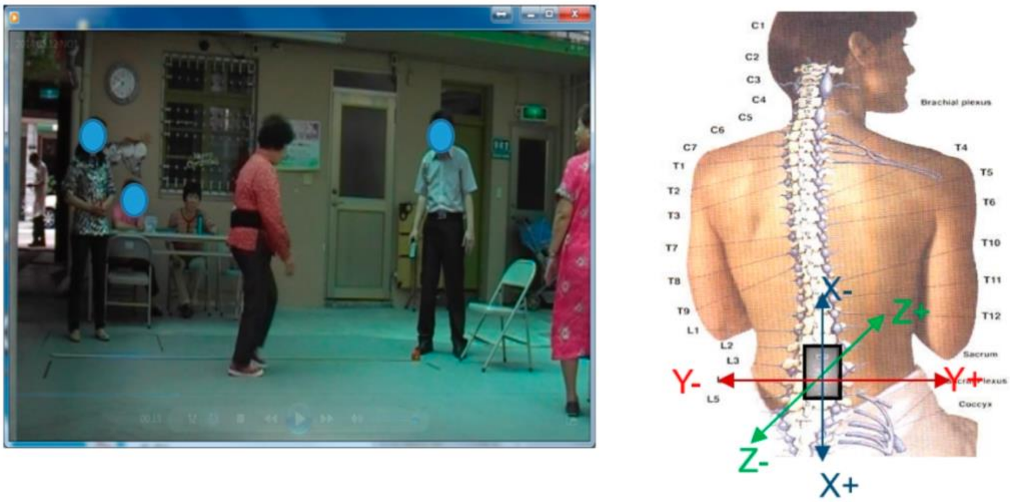

Figure 5. The sensors are placed on the waists of the participants. The direction of the three axes are defined as in the graph

X	Y	Z
11.61	1.88	-2.12
11.65	1.84	-2.12
11.65	1.84	-2.12
11.65	1.84	-2.16
11.65	1.84	-2.08
11.69	1.8	-2.12
11.69	1.84	-2.2
11.69	1.84	-2.16
11.69	1.84	-2.12
11.69	1.8	-2.16
11.69	1.8	-2.16
11.65	1.84	-2.12
11.65	1.84	-2.12
11.65	1.84	-2.12
11.69	1.77	-2.16
11.69	1.8	-2.16
11.69	1.84	-2.16
11.69	1.8	-2.16
11.65	1.84	-2.2
11.61	1.8	-2.12
11.69	1.84	-2.04
11.69	1.84	-2.16
11.61	1.84	-2.16
11.65	1.77	-2.16
11.77	1.8	-2.08

Figure 6. A snapshot of how the data would look like in Microsoft Excel. The three columns represent acceleration along the X-axis, Y-axis, and Z-axis, respectively.

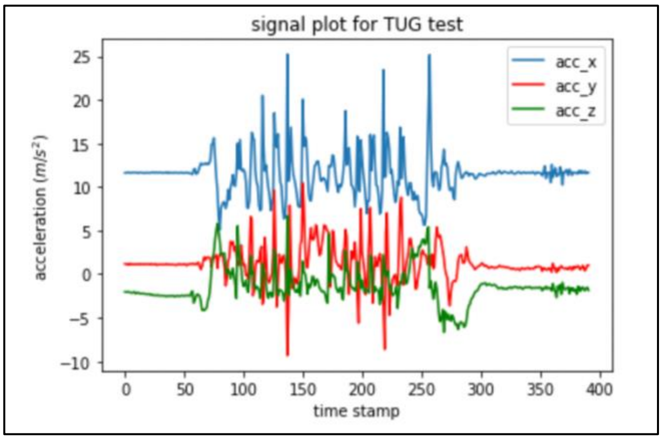

Figure 7. The signal plot for a particular subject.

2-5 Number of non-potential faller vs. potential faller

150 subjects participated in this experiment. 30 of them turned out to be classified as high-risk fallers; and 120 of them were classified as the low-risk fallers according to the SFBBS given by the medical professionals.

3. Modeling

3-1. Preprocessing

Since the data obtained from the tri-axis sensor are time series data, I decided to convert each time stamp into a 3-dimensional vector, which contains the information of the acceleration along x-axis, y-axis, and z-axis respectively. (see figure 8 for visualization).

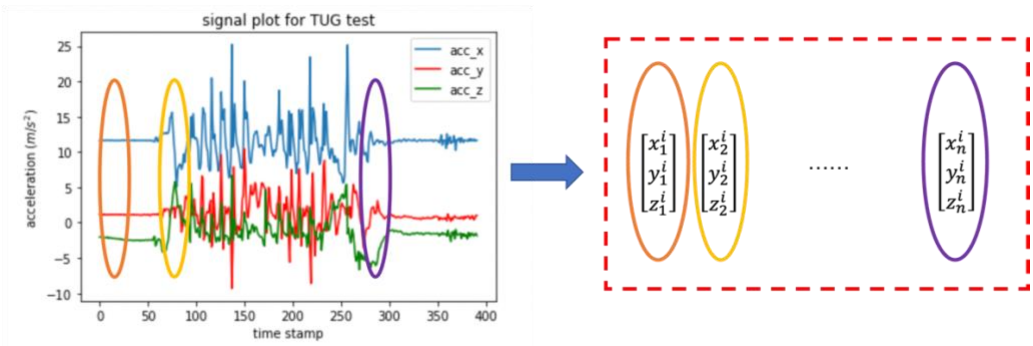

Figure 8. The visualization of preprocessing time series data to a sequence of 3-dimensional vectors.

In order to make the input sequence length the same, I went through all the subjects' experiments to see the lengths of time series for each of the subject. The distribution of these time series lengths was plotted in figure 9. The minimum, maximum, and average are 263, 2201, and 498.27, respectively. After obtaining this information, I truncated the subjects' data whose time series lengths are longer than 500 (obtained from the average time series length) to be 500. On the other hand, I performed a zero-padded to the subjects' data whose time series length are shorter than 500. In this case, all of the 150 subjects' data are in the same sequence length. (Illustration shown in figure 10 and 11)

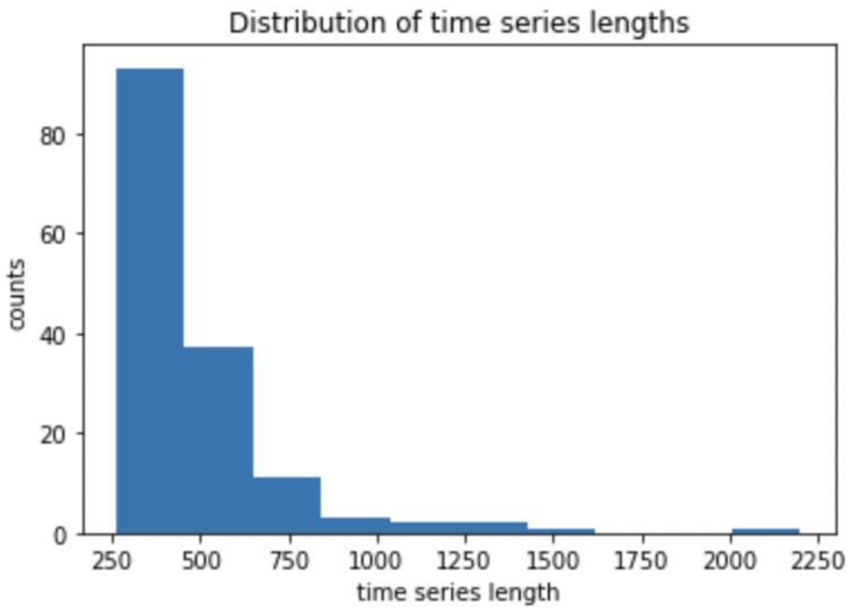

Figure 9. The time series lengths distribution among the 150 subjects.
$(min = 263, max = 2201, average = 498.27)$

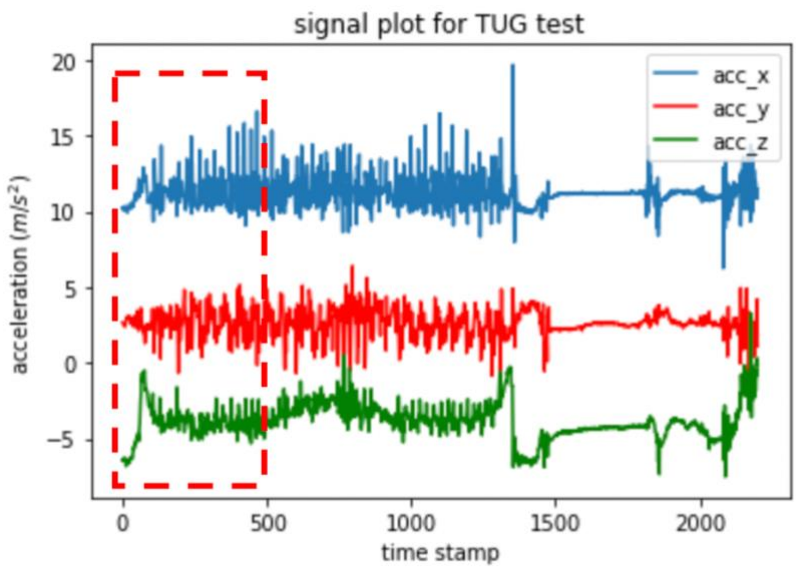

Figure 10. Illustration of truncation. The red dotted line represents the window where the input data came from. Any information after timestamp 500 were discarded in order to unify the input sequence length to the model.

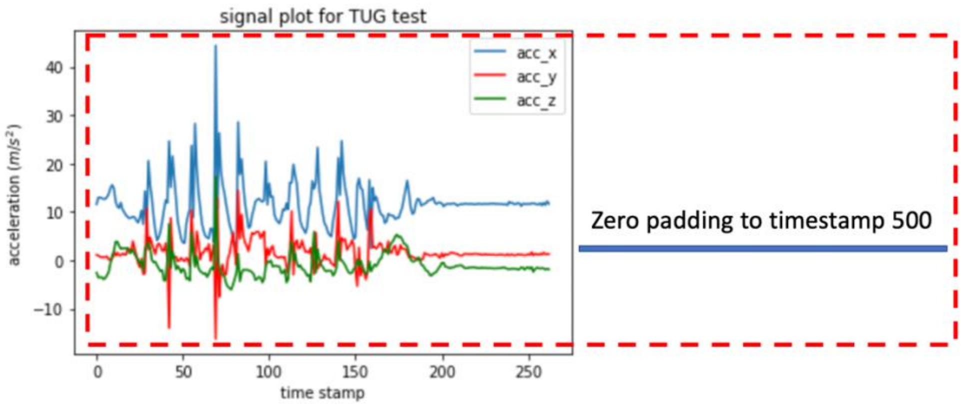

Figure 11. Illustration of zero-padding. The red dotted line represents the where the input data came from. After timestamp 250, this subject had ended the TUG test. In order to unify the input sequence length to the model, we had to pad zeros to this sequence.

3-2. Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM)

To produce a simple preliminary result, I constructed a simple Bi-LSTM Recurrent Neural Network (RNN) with one-hidden-layer (Figure 12). The reason this model was picked was because RNN is known to be having great effect in time series data [8]. It is designed to have the ability to take the information of the previous timestamp into account when doing propagation. Moreover, LSTM makes this feature more powerful (figure 13). LSTM is formed by three gates, which are input gate, forget gate, and output gate (figure 14). These gates enable the RNN to decide whether the memory will be clear out or not in a particular timestamp. The weights to control these gates are learned by the machine itself. We could implement this technique by substituting the original vanilla neural of the RNN to be the LSTM unit. Lastly, Bi-LSTM means that the network is trained based on two directions, the forward direction, and the backward direction. In this case, comparing to the original RNN that only consider the forward direction, Bi-LSTM RNN could capture the time series data in both original order and reverse order.

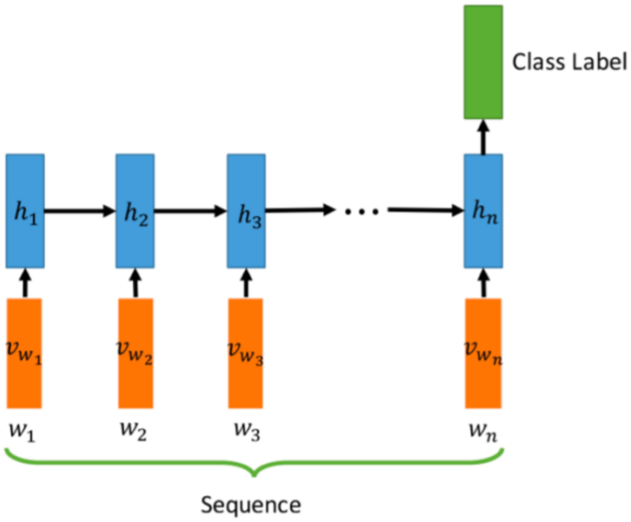

Figure 12. The RNN structure in this project

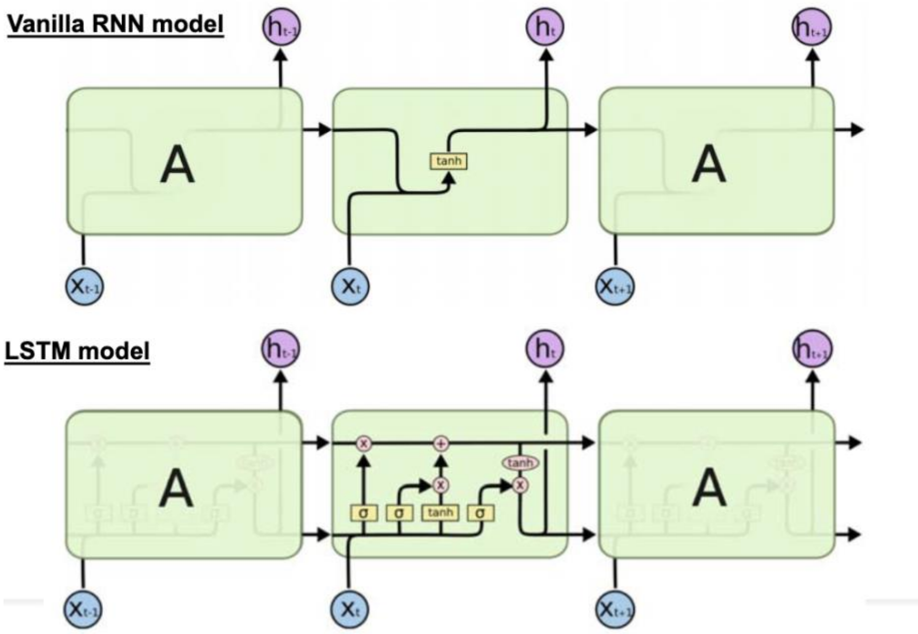

Figure 13. Comparison of vanilla RNN and RNN with LSTM

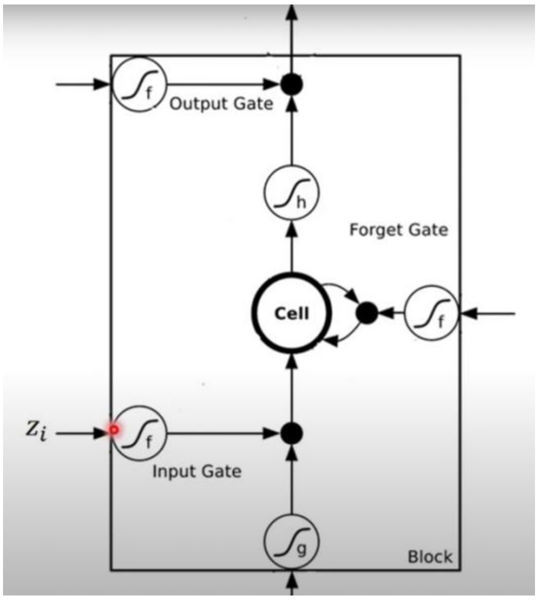

Figure 14. LSTM structure

4. Preliminary Results

4-1. Metrics

Since the problem is a classification problem, the four metrics used in this project would be accuracy, precision, recall, and confusion matrix. The formulas for these metrics are summarized below:

A \ P	Positive	Negative
Positive	Number of TP cases	Number of FN cases
Negative	Number of FP cases	Number of TN cases

A: Actual, P: predicted,

TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

Intuitively, accuracy measures the how many cases are correctly classified by the model among all the cases. On the other hand, precision measures the ratio of correctly predicted positive cases to the total predicted positive observations. High precision relates to the low false positive rate. Finally, recall is the ratio of correctly predicted positive cases over all cases in actual class. Recall measures that among all the actual positive cases, how many did the model label?

4-2 Train test splitting

In order to evaluate whether the model is generalized enough, I split the original dataset into training set, and testing set in a ratio of 90:10. The stratified sampling technique was used to maintain the ratio of positive cases to negative cases. Table 3 summarized the number of cases in training and testing.

Table 3. Summary of number of cases among the training set and testing set

	Faller	Non-faller	Total cases
Training set	27	108	135
Testing set	3	12	15

4-3. Results

The results are summarized in table 4-6. Note that although the accuracy looks good, the model actually was still really naïve to predict all cases to be non-potential fallers due to the imbalanced cases (number of non-potential fallers: number of potential fallers = 120 : 30). This could also be seen if we look at the precision and recall.

Table 4. Summary of accuracy, precision, and recall for both training set and testing set

	Accuracy (%)	Precision (%)	Recall (%)
Training set	80	80	100
Testing set	80	80	100

Table 5. Confusion matrix for training data

A \ P	non-potential faller	Faller
non-potential faller	108	0
Faller	27	0

Table 6. Confusion matrix for training data

A \ P	non-potential faller	Faller
non-potential faller	12	0
Faller	3	0

5. Future Works

Based on the preliminary results, there are still a long way to make this model more predictive.

Firstly, for data preprocessing, the acquired sequence length could be harvested more sensibly. Right now, I just simply obtained the time series information from timestamp 1 to timestamp 500 (which is based on the average time series length for all 150 subjects). For the subject that had overall timestamp length longer than 500, the information after timestamp 500 will be discarded due to truncation. On the other hand, for the subject whose overall timestamp length that is shorter than 500, a lot of zeros will be included in the processed data, making a large portion of the input data to the model not so meaningful. A reasonable way will be to evenly spread out/ condense the 500 sample points to the subject data, this will pretty much let us have a higher quality input data, because we will no longer discard meaningful data or including a lot of zeros in this case. (See figure 10 and 11 for graphical expression)

In terms of splitting strategy, due to the limited number of cases overall, I would like to give bootstrapping, and leave-one-out a try to see if the model could be trained better.

For model structure, a single hidden layer RNN is definitely not enough for the machine to do a good job in classification. I will try to fine-tune the structure, for example, add more layers, and tune hyper-parameter to see how the model could be improved.

Last but not least, another direction is to consider interpretable feature models. The input of this model will be interpretable features such as range of the acceleration data, slope of the acceleration data, mean of the acceleration data (figure 15). Demographic features such as age, gender, walking assistant used or not could also be included, since some of these features were proven to be useful in the literature [9].

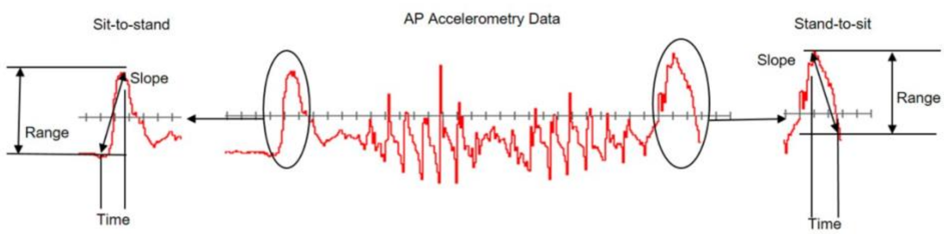

Figure 15. Extracting features from the time series data

Reference

[1] L. Yu, Y. Zhao, H. Wang, T.-L. Sun, T. E. Murphy, and K.-L. Tsui, “Assessing elderly’s functional balance and mobility via analyzing data from waist-mounted tri-axial wearable accelerometers in timed up and go tests,” BMC Medical Informatics and Decision Making, vol. 21, no. 1, 2021.

[2] W. Villena Gonzales, A. Mobashsher, and A. Abbosh, “The progress of glucose monitoring—a review of invasive to minimally and non-invasive techniques, devices and sensors,” Sensors, vol. 19, no. 4, p. 800, 2019.

[3] P.-H. Lin, S.-C. Sheu, C.-W. Chen, S.-C. Huang, and B.-R. Li, “Wearable hydrogel patch with noninvasive, electrochemical glucose sensor for natural sweat detection,” Talanta, vol. 241, p. 123187, 2022.

[4] K. Dunning, “Timed up and go test,” Encyclopedia of Clinical Neuropsychology, pp. 1–2, 2017.

[5] “Berg balance scale,” Physiopedia. [Online]. Available: https://www.physio-pedia.com/Berg_Balance_Scale. [Accessed: 05-May-2022].

[6] Karthikeyan G, Sheikh S, Chippala P: Test-retest reliability of short form of berg balance scale in elderly people. Glo Adv Res J Med Med Sci 2012, 1(6):139-144.

[7] C.-H. Lee, T.-L. Sun, B. Jiang, and V. Choi, “Using wearable accelerometers in a community service context to categorize falling behavior,” Entropy, vol. 18, no. 7, p. 257, 2016.

[8] A. Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and long short-term memory (LSTM) network,” Physica D: Nonlinear Phenomena, vol. 404, p. 132306, 2020.

[9] Y.-C. Hsu, Y. Zhao, K.-H. Huang, Y.-T. Wu, J. Cabrera, T.-L. Sun, and K.-L. Tsui, “A novel approach for fall risk prediction using the inertial sensor data from the timed-up-and-go test in a community setting,” IEEE Sensors Journal, vol. 20, no. 16, pp. 9339–9350, 2020.