

Fall Risk Classification Among Seniors

A Technical Report submitted to the Department of Biomedical Engineering

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Leyao Li

Spring, 2020.

Technical Project Team Member

Ziyang Guo

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

B. Eugene Parker, Jr., Barron Associates, Inc.

Timothy E. Allen, Department of Biomedical Engineering

Shannon Barker, Department of Biomedical Engineering

Fall Risk Classification Among Seniors

Ziyang Guo^a, Leyao Li^a, B. Eugene Parker, Jr.^{b,1}

^aBiomedical Engineering, University of Virginia

^bBarron Associates Inc.

¹Correspondence: B. Eugene Parker, Jr.

Email: parker@bainet.com

Address: 1410 Sagem Place, Suite 202, Charlottesville, VA 22901

Abstract

Falls in elderly people are the leading cause of visits to emergency departments and can lead to serious health problems. It is reported that nearly 28-35% of people aged 65 years and above fall each year and this percentage increases to 32-42% for those over 70 years of age¹. The aim of this study was to investigate the potential of neural network models to improve efficiency and accuracy of current methods for evaluating risk of falling. Two neural network models, Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), were developed using the MATLAB deep learning toolbox. The prediction accuracy was defined as the percentage of patients recognized as fallers both in the models we developed and in the results generated by Nithman et al. A total of 163 participants were included and their stabilogram data was analyzed. After adjusting for bias and optimization, the prediction accuracy for LSTM is 51.6% +/- 2.7% and for CNN is 51.2% +/- 2.4%. The results in the study indicated that the CNN and LSTM models developed were not advancing prediction accuracy for fall risk based on stabilogram data.

Keywords: Risk Analysis, Fall Risk Assessment, LSTM, CNN

Introduction

According to the U.S. Centers for Disease Control and Prevention, one in four Americans aged 65+ falls each year and 1800 falls directly result in death while 9500 deaths are associated with falling annually. Falling is also the major factor that contributes to bone fracture; fractures of the hip, forearm, humerus, and pelvis usually result from the combined effect of falls and osteoporosis, which lead to further limitation of activity to the extent of loss of mobility². It is reported that about 90% of hip fractures in the elderly are resulted from falling and the one year mortality rate due to hip fracture is 21.4%^{3,4}. Besides being a common cause of serious injuries, the cost of hospitalization associated with falling can be as high as \$30,000⁵. Falling might also induce future fear of falling, which is defined as a lasting concern about falling that can lead an individual to avoid activities that he/she remains capable of performing, leading to more sedentary activities⁶. A strategy works against falling that has proved successful is to select patients at high risk and target prevention strategies⁷. Therefore, developing a method for predicting risk of falling is critical in order to target high-risk individuals for preventive intervention. However, predicting fall risks can be challenging due to the difficulties in gathering a large set of well documented gait data and in adapting to the variations among individuals. Even with enough volunteers participating, designing an experiment capable of accurately recording sufficient gait features that contribute to the individual's falling can also be challenging. Much previous study has been done and over 400 clinical characteristics have been identified to be associated with an increased incidence of falls occurring at home or outdoors, while the clinical measurement of gait features is not standardized. Challenges in gait feature measurements include whether a sit-to-stand movement or only walk movement should be recorded, lack of data that includes a real fall since most of the falls do not occur during the measurement⁸. Furthermore, analysing the data and integrating multiple

gait features into the prediction can be complicated and requires a large amount of time.

Among the 400 risk factors identified, balance disorders have been discovered as critical factors for assessing fall risk⁹. Stabilogram is a developed tool that can be implemented to assess balance problems, and the procedure for stabilogram measurements is straightforward and is of low cost.

There have been studies investigating fall risk assessment with traditional statistical methods. However, such methods require intensive calculations by hand and require adjustment towards each individual included. Machine learning algorithms have been shown to have stronger prediction ability compared to traditional statistical methods because of their ability to adapt to the varying patterns of data. Long short-term memory (LSTM) recurrent neural networks have been successfully applied to predictions of future events based on time series data and classification tasks. Convolutional neural networks have also demonstrated promising results in application to motion and image recognition. These neural network models have not been applied to classification of fall risk based on stabilogram data. This study aims to explore the potential for such models to classify fall risk among seniors.

Results

Preliminary Results

Our first batch results from both CNN and LSTM networks showed 75% accuracy with 3% distribution deviation. However, due to the fact that our sample population has 74.6% non-fallers and only 25.4% fallers, our first batch accuracy rate was biased. After inspecting the end classifications of the networks given to the testing set, we discovered the networks were skewed towards non-fallers classification that there was only 1 faller classification in the first batch training with 3840 validation sets.

		CNN	LSTM
Accuracy Rate	Faller	44% \pm 21.8%	62% \pm 18.8%
	Non-Faller	58.4% \pm 20.6%	41.3% \pm 19.4%
	Overall	51.2% \pm 2.4%	51.6% \pm 2.7%
Training Time		~7 minutes on PC /batch	~50 hours on Rivanna/batch

Table 1: Accuracy Rate for CNN and LSTM model after balancing number of data in each category (faller and non-faller). Both models showed accuracy rates of around 50%. The time required to train the CNN model was much less than that of a LSTM model.

Our second batch training with balanced and normalized data (results in Table 1) set showed an average 44% predictive accuracy rate in classifying fallers and an average 58.4% predictive accuracy rate in classifying non-fallers for CNN, and an average 62% predictive accuracy rate in classifying fallers and an average 41.3% predictive accuracy rate in classifying non-fallers for LSTM. Adjusted for sample population, the overall accuracy of predicting faller and non-faller for the CNN network is 51.2% and for the LSTM network is 51.6%. The standard deviations of CNN are 21.8% in predicting fallers and 20.6% in predicting non-fallers. The standard deviations of LSTM are 18.8% in predicting fallers and 19.4% in predicting non-fallers.

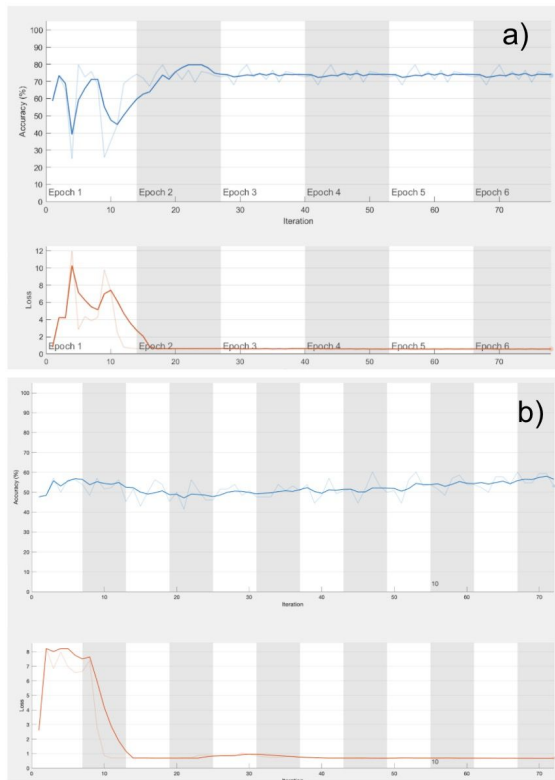


Figure 1: Comparison of Biased CNN Network Learning and Unbiased CNN Network Learning. Image a) shows that the biased learning model has a wide range of accuracy and loss fluctuation during training. Image b) shows a much restrained range of oscillation of model accuracy and loss.

Optimization

For CNN, we increased the max epoch numbers (number of times for each dataset to pass through the network) from 3 to 5 then to 12, and we adjusted the validation frequency from 20 to 35 then to 4. For LSTM, we trained the network with epoch numbers of 2, 3, 4, and 6, and we lowered the mini batch size from 887 to 2. For both networks, we included a layer of normalization of Z-score transformation to level out the different input impact to the weights and biases.

For the training dataset, we balanced the amount of faller and non-faller data by randomly selecting 400 datasets from each category. We randomized the order of all training data so that no strong correlations between each dataset would influence the output of the training. We also created independent validation trials of fallers and non-fallers for both networks so that we could properly assess the predictive accuracies for both faller and non-faller categories.

Discussion

Preliminary Result Implication

The first batch preliminary 75% accuracy rate closely resembles the number distributions of non-fallers in the entire dataset. Inspection shows that our networks only learned to classify all datasets as non-fallers to maximize its accuracy rate. Therefore, the 75% accuracy rate is a biased accuracy result; it demonstrates that our networks did not capture any patterns in neither time nor spatial dimensions to successfully identify fallers. This first batch preliminary result only presents the ability of our networks to develop strategies to reach the final training goal, which is maximizing the final accuracy rate in the given sample population.

The second batch results shows an unbiased predictive accuracy that is in the close reasonable range of random guess events for both networks. It is safe to conclude that our networks did not learn any time or spatial patterns to create a meaningful prediction. Instead, our networks rely on random guesses and random strategies including biased selection to produce results. In addition, the wide range of predictive accuracy and high standard deviation provide further confirmation that no significant patterns or logics were found during the training by both networks.

Failure in this project indicates that the proposed method is inappropriate in identifying the fallers and in predicting the likelihood of falling. The failure of the CNN network in predicting either of the categories demonstrates that there might not be spatial correlations between each variable to time; thus, our innovation at turning time-series variables into images might have been a futile attempt. Moreover, the failure of the LSTM network in predicting either of the categories shows that there are no significant repeating patterns in the dataset that could potentially contribute to the calculation of the likelihood of falling. This lack of time-dependent patterns may be caused by the lack of movement data in the experimental dataset. Without movement data over a detectable distance, the data might not possess any recognizable repetitions.

In addition to the implications to the neural networks, this failure might also signal that the processing of input data was not sufficient. Due to the difference in walking and standing, each dataset might not have the same patterns or timestamps as we originally thought. It is possible that the dataset needed another layer of processing by cropping them into different phases and comparing them phase-wise.

Optimization

The modification of epoch number, validation frequency, and mini batch size did not exhibit any noticeable change to the networks' accuracy. This stagnation might indicate neither of the designed networks were not suitable or sophisticated enough for the task so that no change of parameters or variables could improve the accuracy of the system.

After adjusting the input to be an even dataset with a random distribution, the steep decrease in the accuracy rate from 75% to 51% shows that the first batch preliminary result was indeed biased, and the near randomly distributed 51% unbiased accuracy rate demonstrates that our networks were simply guessing the outcome classification like tossing a fair coin; there is no concrete evidence of any learning done by neither of the networks. The wide range of the accuracy spectrum in classifying fallers and non-fallers once again reinforces the claim that the results produced by the networks are completely random and no meaningful patterns were found.

Limitations and Future Research

There were several limitations in this experiment in finding a better way to identify fallers and to predict the likelihood of falling. First of all, the stabilogram dataset was not, by any means, a large diverse dataset. There are a total 163 participants in the experiment, which was not a number that could represent the wide variance in walking and mobility disorders. The number of combinations that can be generated and learned from the 8 channels in the data was relatively small compared to the sampling of 6000 timestamps. In addition, there are only 41 subjects in the study that were fallers. This relatively small amount of fallers could not adequately deliver features and patterns that might possibly exist in many other fallers. Secondly, the dataset only contains velocity, momentum, and positions of the center mass. The features and parameters introduced in this dataset are severely underrepresenting the complex movement of walking and standing.

Moreover, time, computing power, and the unfortunate lockdown also present considerable constraints on the project. Due to the sheer amount of training needed in the project, this was not a possible project on a personal computer. Running on the supercomputer *Rivanna*, LSTM training still took around 50 hours to complete; therefore, it took us 2 full days to wait and assess the results for each modification and iteration to the LSTM model. In addition, due to the limited experience in computer science, we most likely did not have the expertise to design the best algorithm possible for this project.

There are certainly a myriad of directions this project can expand in the future. First of all, this project could incorporate a lot more data including movement data and various features, such as medial/lateral swing angles, height and limb length, and changes in acceleration. Secondly, many different network architectures are waiting to be explored; these include different numbers and types of layers, different activation and loss functions, or even potential combinations of different networks. Better approaches to process data, such as constructing high-dimensional correlations between variables and dividing data into different phases of movement, can also be beneficial to investigate.

Materials and Methods

Data

From the study carried out by Muir et al.¹⁰, compared with young adults, the trace of center of pressure observed from the

stabilogram data for the elderly tend to be less stable and demonstrate multiple excursions from the center. This study suggested that patterns in stabilogram data might contain useful information in assessing fall risks. The data used to train and test the LSTM and CNN models was chosen to be the Human Balance Evaluation Database, retrieved from PhysioNet¹¹. There are a total of 1930 sets of data collected from 163 participants. Each dataset was obtained by recording the stabilogram reads for one minute and was sampled at a frequency of 100 Hz, consisting of 6000 timestamps. Each file contains 8 predictor variables: force (N) in x-axis, y-axis and z-axis, moment(N/m) in x-axis, y-axis and z-axis and center of pressure (CoP) (cm) in x-axis and y-axis. Each file also contains a response variable that recognizes whether the dataset comes from a faller (marked numerically as "0") or a non-faller (marked numerically as "1").

MATLAB Deep Learning Toolbox

Both CNN and LSTM models were designed using the MATLAB deep learning toolbox. The toolbox provided the framework for constructing the two neural network models. This toolbox provides built in neural network layers that can be used to combine and construct neural networks with different architectures by changing the layer types, number of layers included and number of hidden units as needed.

Rivanna

Initially, the estimated training time of the LSTM model on our own devices was more than 150 days, which far exceeded the time constraint of our project. Because of the size of our datasets and the computational requirements for a recurrent neural network model, we performed training and testing of LSTM at *Rivanna*, the high performance computing system at the *University of Virginia*. *Rivanna* is a multi-user environment and employs Simple Linux Utility for Resource Management (SLURM) scheduling and managing jobs. A slurm file was written and submitted for each training of the LSTM model. *Rivanna* has helped reduce the training time of a LSTM net from several months to an average of less than 50 hours and ensured that the project stayed on the planned timeline.

Preprocessing Data

Because the goal was to classify fallers and non-fallers, the response variables contained in the dataset were first transformed into categorical variables, with a "0" representing a non-faller and an "1" representing a faller. Because CNN is mostly used for image recognition and therefore the data input for the CNN model required a transformation of the predictor variables from an original 2-D array of 8*6000 doubles into a 4-D array of size 1*1*8*6000.

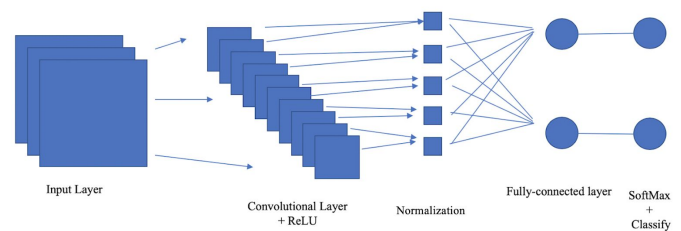


Figure 2. The CNN Model The CNN developed has 7 hidden layers: an input layer, a convolutional layer, a ReLU layer, a normalization layer, a fully connected layer, a softmax layer and a classification layer. Each of the input data has been transformed into the size of 1*1* 8*8000 (a 4-D array).

Convolutional Neural Network

Convolutional Neural Network is a type of deep learning algorithm that is mostly applied to for image recognition. A typical CNN model takes in an image as input, learns the patterns and features in the given image, assigns weights and biases to the features presented in the image and differentiates the target feature from the others. In this project, we explored the possibility of converting a time-series event into an image-like dataset and then implement CNN for recognizing patterns of falling in the data and eventually classifying out fallers. As shown in Figure 2, our CNN model contains 7 hidden layers: an input layer, a 2-D convolutional layer (8*25), a batch normalization layer, a ReLU layer, a fully connected layer, a softmax layer followed by a classification layer. Within each layer there is a different function that processes the data from different perspectives. The batch normalization layer in our CNN model standardized the data. Each data point was normalized by subtracting the mean of each channel and then divided by the standard deviation of each channel. In the classification layer, the function included in the layer was a binary cross entropy function (presented by equation (1)) that computes the cross entropy loss, heavily penalizes outputs that are highly inaccurate but assigns little penalty for correct classifications. Implementing the binary cross entropy function in the layer could help improve the classification accuracy given the imbalanced data.

$$bce = -t * \log(y) - (1 - t) * \log(1 - y) \tag{1}$$

Equation 1: the binary cross entropy equation This equation computes the difference between the predicted t is the binary indicator of class (“1” for faller and “0” for “non-faller”; y is the predicted probability of the data being classified as faller (“1”) for all data points.

Long Short-Term Memory Neural Network

LSTM is a specialized recurrent neural network model that takes in data that is organized in a sequence of time, and the output of a LSTM model at a time point t combines information from previous inputs. Similar to the CNN model, the LSTM model (Figure 3) has 5 hidden layers: an input layer, a BiLSTM layer, a fully connected layer, a softmax layer followed by a classification layer.

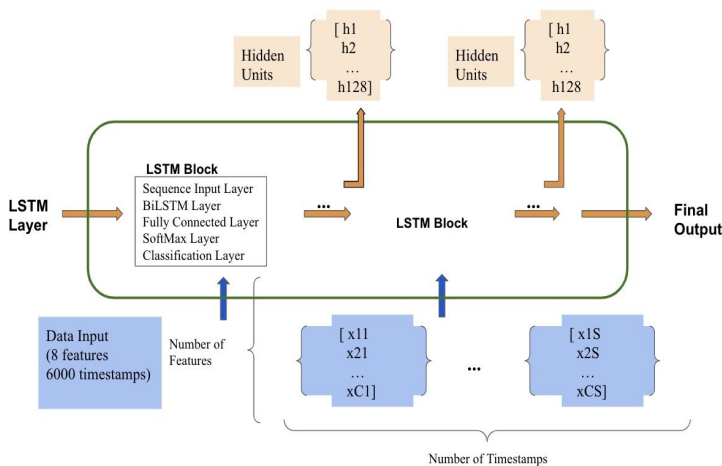


Figure 3. The LSTM Model The LSTM developed has 128 hidden units and 5 hidden layers: a sequence input layer, a BiLSTM layer, a fully connected layer, a softmax layer and a classification layer. Each of the sequence input has a size of 6*8000.

Comparison with Baseline Methods

To assess our model efficiency more comprehensively, we compared our model with other two methods applied to the risk

classification problem. The first method we used to compare is a study carried out by Nait Aicha et al.¹² In the study, the researchers also developed a CNN model, a LSTM model and a ConvLSTM model that combined CNN and LSTM, respectively. Their model achieved accuracy rates of 52% for the CNN model, 61% for the LSTM model and 60% for the ConvLSTM model. Although the data used by the researchers were different from what we used (theirs were accelerometer data and had 296 participants), the accuracy rates were not significantly different from our models (for CNN $p < 0.0001$, for LSTM $p < 0.0001$). The insignificant difference between the classification accuracy rates of our models and the models developed by Nait Aicha et al. suggests that the application of neural network models to fall risk classification needs further researches in terms of which factors should be included, how much participants’ data is required and more efforts are needed into further investigation of the architecture of the neural network models in recognizing patterns of falling.

The second baseline method was the GNOSIS tool developed at Barron Associates. The GNOSIS tool was a classification tool based on logistic regression. Two channels of the data recording center of pressure (CoP) in the Human Balance Evaluation database were applied to GNOSIS for training and testing. The classification accuracy was about 50%, which was very close to the accuracy rates of LSTM and CNN developed after balancing the number of faller and non-faller data. The comparison with GNOSIS indicated that the two neural network models developed demonstrated no significant advancement of classification power compared to that of a logistic regression classifier.

End Matter

Author Contributions and Notes

B.E.P. designed research, Z.G. and L.L. performed research, Z.G. and L.L. analyzed data; and Z.G. and L.L. wrote the paper. The authors declare no conflict of interest.

Acknowledgments

We would like to thank Dr. Eugene Parker, Dr. Brian R Clark, Dr. Christopher Wiles, Dr. Todd Summers at Barron Associates Inc. and Dr. Ed Hall at the University of Virginia Research Computing for advising.

References

- Fuller, G. F. Falls in the Elderly. *AFP* **61**, 2159 (2000).
- Prevention, I. of M. (US) D. of H. P. and D., Berg, R. L. & Cassells, J. S. *Falls in Older Persons: Risk Factors and Prevention*. (National Academies Press (US), 1992).
- Ensrud, K. et al. Group for the Study of Osteoporotic Fractures Research: Comparison of 2 frailty indexes for prediction of falls, disability, fractures, and death in older women. *Archives of internal medicine* **168**, 382–9 (2008).
- Schnell, S., Friedman, S. M., Mendelson, D. A., Bingham, K. W. & Kates, S. L. The 1-Year Mortality of Patients Treated in a Hip Fracture Program for Elders. *Geriatr Orthop Surg Rehabil* **1**, 6–14 (2010).
- Woolcott, J. C., Khan, K. M., Mitrovic, S., Anis, A. H. & Marra, C. A. The cost of fall related presentations to the ED: a prospective, in-person, patient-tracking analysis of health resource utilization. *Osteoporos Int* **23**, 1513–1519 (2012).
- Adelsberg, S., Pitman, M. & Alexander, H. Lower extremity fractures: relationship to reaction time and coordination time. *Arch Phys Med Rehabil* **70**, 737–739 (1989).

7. Oliver, D., Britton, M., Seed, P., Martin, F. C. & Hopper, A. H. Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: case-control and cohort studies. *BMJ* **315**, 1049–1053 (1997).
8. Oakley, A. *et al.* Preventing falls and subsequent injury in older people. *Qual Health Care* **5**, 243–249 (1996).
9. Deandrea, S. *et al.* Risk Factors for Falls in Community-dwelling Older People: A Systematic Review and Meta-analysis. *Epidemiology* **21**, 658–668 (2010).
10. Muir, J. W., Kiel, D. P., Hannan, M., Magaziner, J. & Rubin, C. T. Dynamic Parameters of Balance Which Correlate to Elderly Persons with a History of Falls. *PLOS ONE* **8**, e70566 (2013).
11. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals (2003). *Circulation*. 101(23):e215-e220.
12. Nait Aicha, A., Englebienne, G., van Schooten, K. S., Pijnappels, M. & Kröse, B. Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry. *Sensors (Basel)* **18**, (2018).