HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 3 (2023)

Freezing of Gait Classification in Parkinson's Patients Using Machine Learning

Anurag Jakkula[1] and Ron Yu#

[1]Dublin High School
#Advisor

ABSTRACT

This study aims to design personalized machine learning models for the classification of time-series data to detect freezing of gait (FoG) in Parkinsons' patients in short time intervals. FoG is the medical terminology for sudden episodes of an inability to move in patients that suffer from Parkinson's disease. Data collected experimentally by Xuanwu Hospital were used. Each 0.002-second interval is labeled as FoG positive or negative by physicians. Using information gain statistics, it was determined that out of 58 features of accelerometer, EEG, and EMG measurements, 35 measurements provide the most information about FoG presence. Features were normalized via z-score normalization. For feature vectors, data are grouped into 0.5-second batches with .002 second timeframes for LSTM; while data are grouped into 0.5-second intervals for other models. The FoG positive/negative classes were balanced through SMOTE. All models were hyperparameter trained through 10-fold cross-validation. The F-1 scores of LSTM, Random Forest, SVM, Decision Tree, and Logistic Regression are 89.71%, 89.69%, 87.00%, 74.44%, 67.21% respectively. Of the models analyzed, LSTM has the highest recall at 93.16%, while Random Forest has the highest precision at 94.34%. LSTM detects the most positive instances, while Random Forest has precise detection. LSTM has a higher F-1 score, indicating it is better at balancing precision and recall. These personalized short interval-input models can be implemented in wearable devices to detect freezing of gait to aid physicians' assessment of disease severity and treatment.

Methods

Data Description and Preparation

Table 1: Sensor information about the experimental study in Xuanwu Hospital.[10]

Sensing Type	Sensor system	Number of sensors	Sensor location
28D-EEG	The wireless	28	FPI. FP2. F3. F4. C3. C4. P3. P4. Q1, O2. F7,F8,1 P7. P8, Fz, Cz, Pz, FC1, FC2. CP1, CP2. FC5, FC6, CP5, CP6 TP9* TP10* TO**

3D-EMG	MOVE	3	Gastrocnemius muscle of right leg. Tibialis anterior muscle of left and right legs
3D-accelerometer 3D-Gyro	MPU6050	4	Lateral tibia of left and right legs Fifth lumbar spine Wrist
ID-SC	LM324	2	The second belly of the index finger and middle finger of the left hand

Publicly available data published by Xuanwu Hospital in Beijing, China were used to build classification models to detect freezing of gait. The data—consisting of EEG brain signals; EMG, ECG, and Electrooculogram signals; and acceleration data—were collected while the twelve patients accomplished tasks consisting of certain movements. During the process, two doctors labeled the data as either freezing of gait positive (1) or freezing of gait negative (0)[6].

	P-value
1	2.667494e-16
2	1.531298e-06
3	6.553385e-78
4	1.631823e-29
5	8.969280e-67

Figure 1: First five P-values for likelihood of Patient 1 and 2 features coming from the same distribution.

The Kolmogorov-Smirnov Test was used to determine whether the feature distributions of the patients vary significantly. By running the test between all patient combinations, it was determined that the features come from different distributions with respect to the patients since the probability of the data coming from the same distribution was less than 5% for all feature combinations across all patients. Since there are only 12 samples of patient data and feature distributions are distinct, a generalized detection model for all patients cannot be developed. Thus, personalized models are developed.

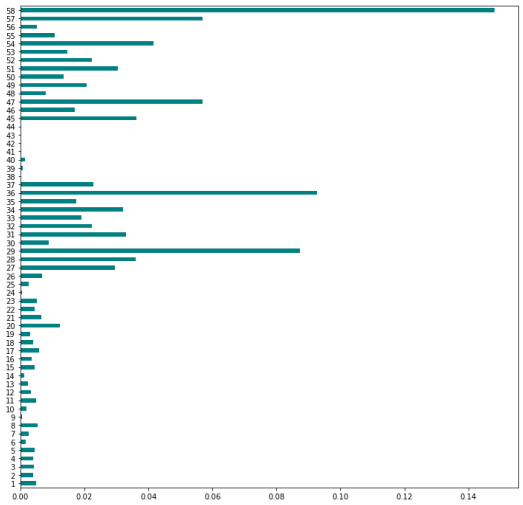

Figure 2: Graph of information gain across 58 features of Patient 1 (X-axis: decrease in entropy; Y-axis: features).

Patient 1's data was imported. Information gain statistics were used to determine which features are most relevant to the classification problem. For this purpose, the features can be grouped into three categories: EEG signals (1-25), EMG/ECG/Electrooculogram signals (26-30), and acceleration data (31-58). It was determined that EEG signals decreased the entropy with respect to freezing of gait classification by the least amount, and thus those features were excluded in the training process.

In the experiment, data were collected in 0.002-second intervals. However, to make feature vectors, data were grouped into 0.5-second intervals to provide the model with more information for detection. 0.5-second intervals with one or more instances of freezing of gait were labeled freezing of gait positive and the rest were labeled freezing of gait negative. An exception is the LSTM neural network, for which the data was manipulated into a 3-D array of shape (number of 0.5-second intervals, 250, 35).

Every recorded feature was normalized through z-score normalization to scale the data as the number of standard deviations from the mean based on the means and standard deviations of the 35 measurements. This step is important for many reasons, including ensuring computationally friendly vectors for the gradient descent algorithm and consistency during regularization.

Synthetic minority oversampling was used to oversample the minority class, freezing of gait positive. This technique synthesizes data by iteratively selecting a random data point from the minority class and synthesizing a data point at a random distance between it and its nearest neighbor. This technique is effective for the dataset at hand since it consists of continuous numerical features. Synthesizing new data points rather than the more naive method of randomly duplicating existing ones reduces variability of the resulting models' classification decisions.

Metrics Description

The following metrics were utilized to train hyperparameters and compare models:

Equation 1: $Accuracy = (TP + TN)/all$

Accuracy is the proportion of data classified correctly. It is the proportion of 0.5-second intervals that are correctly classified.

Equation 2: $Precision = TP/(TP + FP)$

Precision is the proportion of data correctly classified as positive out of all data classified as positive. It is the proportion of 0.5-second intervals that are correctly classified as freezing of gait positive out of the 0.5-second intervals that are classified as freezing of gait positive.

Equation 3: $Recall = TP/(TP + FN)$

Recall is the proportion of data correctly classified as positive out of all real positives. It is the proportion of 0.5 second intervals that are correctly classified as freezing of gait positive out of the 0.5 second intervals in which the patient experienced freezing of gait.

Equation 4: $F1\ Score\ =\ (2\ x\ Precision\ x\ Recall)/(Precision\ +\ Recall)$

F1 Score is the harmonic mean of precision and recall.

An ROC curve displays the relationship between the true positive rate and false positive rate given a fluid decision boundary. An area under the curve that is closer to one indicates a model that is able to better differentiate between positive cases and negative cases, while an area under the curve that is closer to half indicates a model performance that is close to randomness.

Models Description

Logistic Regression

Equation 5: $p(x)\ =\ \frac{1}{1+e^{-\beta x}}$

Logistic regression utilizes the sigmoid function to map data points to their respective probabilities.

Equation 6: $Loss\ =\ l(p|Y)\ =\ -\sum_i \blacksquare\ y_i log(p_i) + (1-y_i) log(1-\ p_i)$

The negative-log likelihood function is defined based on the training data and minimized through the gradient descent algorithm. In this way, the curve is fit to the training data.

Equation 7: $Loss\ =\ l(p|Y)\ =\ -\sum_i \blacksquare\ y_i log(p_i) + (1-y_i) log(1-\ p_i) + \lambda \sum_{i=1}^{N} \blacksquare\ |w_i|$

L1 regularization can be utilized to penalize the magnitude of feature weights by adding the summation of the weights to the loss function. The intent is to produce a model that is generalizable to data beyond the training data. The λ value is selected through cross-validation.

Soft-Margin Support Vector Machine

A Soft-Margin Support Vector machine separates the classes by maximizing the margin between the support vector and decision boundary. Some error is tolerated for the sake of increasing model bias to reduce model variability.

Equation 8: $Loss\ =\ min\ (||w||^2 + \lambda \sum_{n=1}^{N} \blacksquare\ \xi_n)$

The loss function above is utilized to calculate the decision margin. λ is selected through cross validation. As λ increases, less error is tolerated. As λ decreases, more error is tolerated.

Equation 9: $\xi_n\ =\ max\ (0, 1-y_i(w^T x_i + b))$

Error (ε) is calculated according to the hinge-loss formula above.

Decision Tree

Decision Trees are trained through if/else conditions in a hierarchical structure. They are easily interpretable. Their decision boundaries are piecewise in nature.

Equation 9 (Gini impurity): $I(A)\ =\ p(1-p)$

Equation 10 (Entropy measure): $I(A)\ =\ -plog2(p) - (1-p) log2(1-p)$

Where A is the set of data and p is the proportion of misclassified records in a partition.

The splitting criteria, either Gini or Entropy, is chosen through cross-validation. The chosen criteria is used to determine the optimal split to maximize homogeneity in the resulting leafs.

Tree complexity can be controlled by defining the minimum samples in each leaf. A leaf is not split into portions smaller than the defined value. The value is determined through cross-validation.

Long-Short Term Memory Neural Network

An LSTM neural network consists of a forget gate, input gate, and output gate.

Forget Gate

The forget gate has two inputs, h_{t-1} and x_t, where h_{t-1} is the output from the previous cell and xt is the input at the current time-step. The inputs, multiplied by the weight matrices and having a bias added, are inputted into the sigmoid function to determine what percentage of the values to keep and discard. The cell state is multiplied by the resulting output[8].

Input Gate

The input gate has the inputs h_{t-1} and x_t. The sigmoid function is used to determine a vector consisting of values from 0 to 1 based on the inputs to regulate the information that is being added. The tanh function is used to determine an output vector consisting of values from -1 to 1 that is multiplied by the sigmoid function's output. The resulting values are added to the cell state[8].

Output Gate

The output gate has the inputs h_{t-1}, x_t, and the cell state. The tanh function is applied to the cell state, scaling its values from -1 to 1. The sigmoid function is used to determine values from 0 to 1 based on the inputs ht-1 and xt to regulate the output at the particular time-step. The resulting vector is inputted as the hidden state for the next cell state which evaluates for the next time step[8].

Adaptation to Classifying Freezing of Gait

Training features are inputted into the neural network as a 3-D array of the shape (batches, time-steps, features). Each batch is a 0.5 second interval, while each time-step is each discrete 0.002 second recording. The features are the 35 recorded metrics. To use LSTM for classification, the final layer in the neural network is a sigmoid function. After the LSTM layer, ReLu dense layers are added for further pattern recognition. The final neural network architecture consists of an LSTM layer, 5 ReLu dense layers, a final sigmoid dense layer, and binary cross-entropy as the loss function.

Results

Logistic Regression

A λ of 2 was selected for L1 regularization using 10-fold cross-validation. Accuracy was used as the evaluation metric for the hyperparameter training.

Results

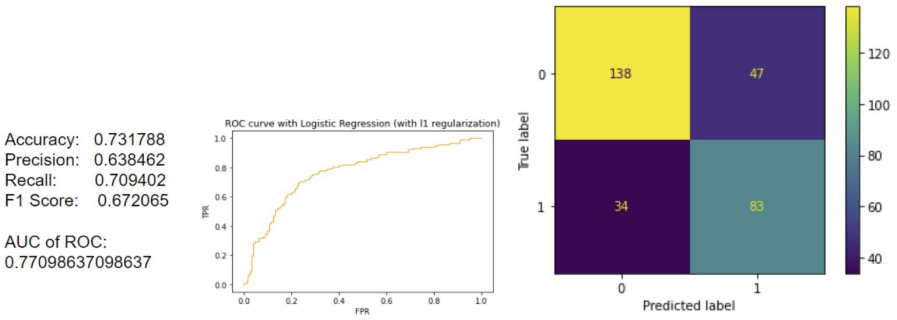

Figure 3: Performance of logistic regression model with L1 regularization. Accuracy, precision, recall, F-1 score, ROC metrics, and confusion matrix are displayed.

The results still do not indicate a reliable model. Both precision and recall are low, indicating that Logistic Regression is not suitable for freezing of gait classification.

Support Vector Machine

Out of the linear, polynomial, radial basis, and sigmoid kernels, the radial basis kernel was selected using 10-fold cross-validation. A λ of 0.05 was selected for soft-margin error tolerance using 10-fold cross-validation. Accuracy was used as the evaluation metric for the hyperparameter training.

Results

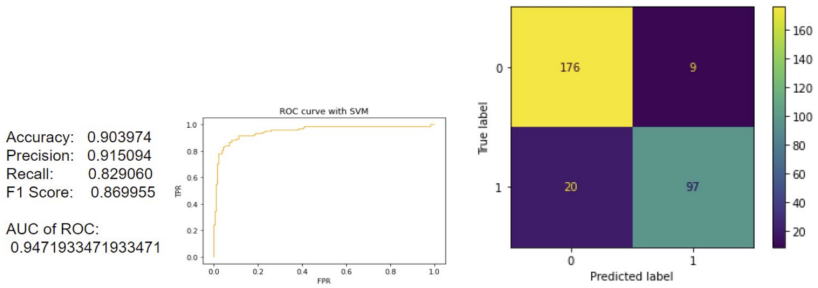

Figure 4: Performance of hyperparameter trained SVM model. Accuracy, precision, recall, F-1 score, ROC metrics, and confusion matrix are displayed.

The Support Vector Machine model performs noticeably better than the Logistic Regression model. However, its precision (91.5%) is greater than its recall (82.9%), indicating that its capability to avoid false positives comes with the tradeoff of outputting more false negatives. This indicates that the Support Vector Machine model is better at classifying freezing of gait negative instances than freezing of gait positive instances.

Decision Tree

Through 10-fold cross-validation, out of gini and entropy, gini was selected as the splitting criteria with accuracy as the evaluation metric. 3 samples were selected as the minimum samples per leaf using 10-fold cross-validation. Adding restrictions on minimum impurity decrease seemed to have a negative effect on accuracy as seen with the 10-fold cross-validation. Therefore, no restrictions were placed on the impurity decrease for splitting. Accuracy was used as the evaluation metric for hyperparameter training.

Results

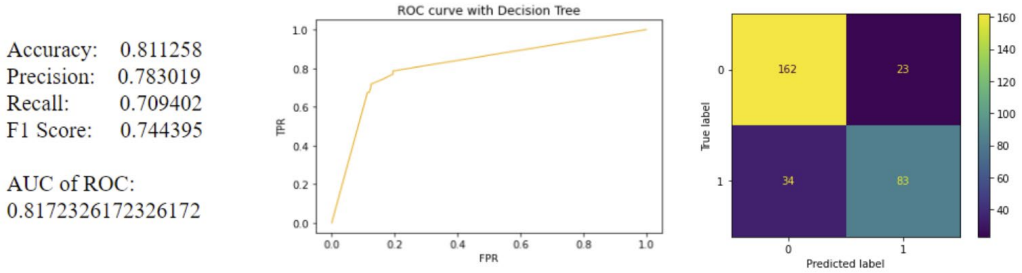

Figure 5: Performance of hyperparameter trained Decision Tree model. Accuracy, precision, recall, F-1 score, ROC metrics, and confusion matrix are displayed.

The Decision Tree model's ROC curve displays an ability to distinguish between positive and negative cases, but it is not as effective as the Support Vector Machine model. Similar to the Support Vector Machine model, its precision (78.3%) is greater than its recall (70.9%), indicating that its capability to avoid false positives comes with the tradeoff of outputting more false negatives. This indicates that the Decision Tree model is better at classifying freezing of gait negative instances than freezing of gait positive instances.

Random Forest

The random forest was constructed with 100 decision trees with the same hyperparameters as those selected for the standalone decision tree classification model through hyperparameter training.

Results

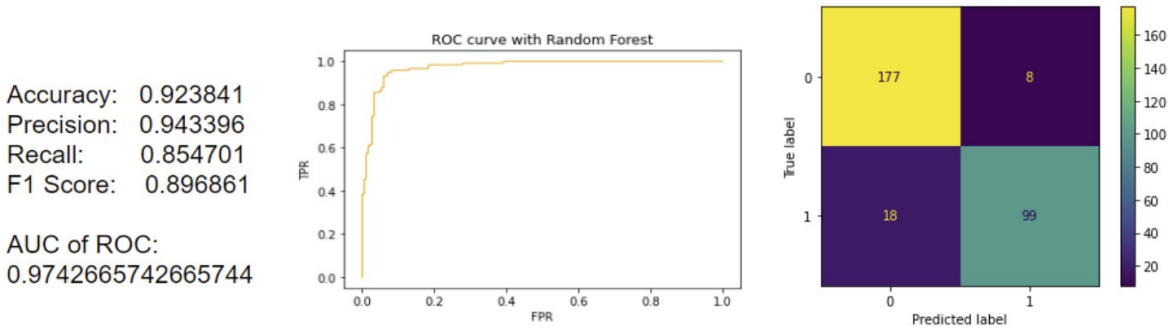

Figure 6: Performance of Random Forest model. Accuracy, precision, recall, F-1 score, ROC metrics, and confusion matrix are displayed.

The Random Forest model performs surprisingly well given that the Decision Tree model has modest results.very well. Like the previous two models, its precision (91.5%) is greater than its recall (82.9%), indicating that its capability to avoid false positives comes with the tradeoff of outputting more false negatives. This indicates that the Support Vector Machine model is better at classifying freezing of gait negative instances than freezing of gait positive instances.

LSTM Neural Network

The LSTM neural network consists of an LSTM layer, 5 ReLu dense layers, a final sigmoid dense layer, and binary cross-entropy as the loss function. Hyperparameters, such as L1 regularization lambda values and number of epochs, were selected through held-out validation data (split from the training data). Then, the model was trained with all the training data and evaluated on the testing data.

Results

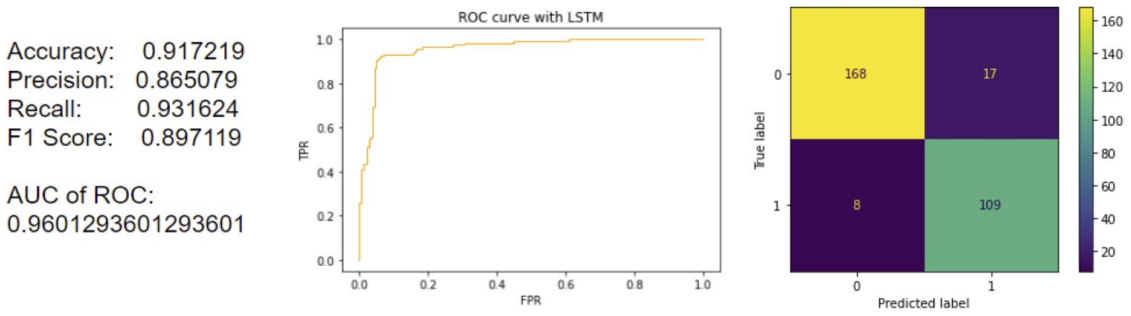

Figure 7: Performance of LSTM model. Accuracy, precision, recall, F-1 score, ROC metrics, and confusion matrix are displayed.

The LSTM neural network performs exceptionally well given the context of the problem at hand. The LSTM neural network has the best recall at 93.2%. Its recall is greater than its precision (86.5%), indicating that its capability to detect freezing of gait positive instances is at the tradeoff of outputting false positives. It also has a harmonic mean (89.7%) that is about the same as the Random Forest's harmonic mean (89.7%).

Discussion

Table 2: Comparison of model performance metrics on testing data. Boxes corresponding to the best model(s) for every metric are highlighted.

	Logistic Regression	**SVM**	**Decision Tree**	**Random Forest**	**LSTM**
Accuracy	0.731788	0.903974	0.811258	0.923841	0.917219
Precision	0.638462	0.915094	0.783019	0.943396	0.865079
Recall	0.709402	0.829060	0.709402	0.854701	0.931624
F1 Score	0.672065	0.869955	0.744395	0.896861	0.897119

The best algorithm in the context of detecting freezing of gait for this particular patient is the LSTM neural network, considering the high recall of 93.2%. Furthermore, LSTM is tied with Random Forest for best F-1 score,

indicating that both models are equally good at balancing precision and recall. However, LSTM tips the balance towards better recall, which is most important considering that an overestimate of a patient's amount of freezing of gait episodes is less harmful than an underestimate. The next best algorithms for the classification of freezing of gait were determined to be Random Forest and Support Vector Machine, with Random Forest having higher accuracy and recall.

The next step is to shift the probability-based decision boundary to below 50% to increase recall at the expense of specificity for the sake of favoring freezing of gait's detection. Furthermore, using a similar training methodology, a predictive machine learning model may be able to be built by offsetting the feature and target arrays to predict freezing of gait before it occurs.

Conclusion

This study has many insights with regard to detecting freezing of gait in Parkinson's disease patients with machine learning. This study finds that the feature distributions of accelerometer data, EMG, and EEG signals are significantly different across patients ($p<0.05$ for all features between all patients). This means that unless a large representative experiment is run to gather the data, models should be made on the personalized level to ensure effectiveness. Furthermore, it was identified that EEG signals provide very little information with regard to detecting freezing of gait. Therefore, it is suggested that the data is not used to train models as it adds noise. This study constructed hyperparameter-trained personalized detection models of logistic regression, support vector machine, random forest, and long-short term memory. Though LSTM performed the best with regard to the question at hand for the particular patient, it is suggested that both LSTM and random forest are compared for other patients since their results are close. A future study could build an ensemble of these two models.

Multimodal Data for the Detection of Freezing of Gait in Parkinson's Disease, published in October, 2022, also built classification models for this data, but in 3-second intervals[10]. Using 1/2-second intervals is beneficial because it allows for estimating freezing of gait episode's duration based on the number of sequential detections, given that a typical freezing of gait episode only lasts 1-2 seconds[5]. A 3-second interval would not be able to distinguish between a normal and abnormal freezing of gait episode duration. This study distinguishes itself from previous studies by both using short time intervals and rigorously defining the rationale behind the features chosen and the models' applicability.

Limitations

The models trained in this study are personalized for individual patients. Therefore, this study demonstrates that models can be trained on individual patients with good accuracy. To design models that are generalizable to all Parkinsons' patients, an experiment must be conducted to collect data from a large sample of Parkinsons' patients from varying ages, ethnicities, and disease severities, since those are likely confounding variables.

Acknowledgements

The data used in this study are from Xuanwu Hospital.[10]

References

[1] Brownlee, J. (2018, September 23). LSTMs for Human Activity Recognition Time Series Classification. Machine Learning Mastery. https://machinelearningmastery.com/how-to-develop-rnn-models-for-human-activity-recognition-time-series-classification/
[2] Brownlee, J. (2019, April 26). Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras. Machine Learning Mastery. https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/
[3] Bruce, P. C., Bruce, A., & Gedeck, P. (2020). Practical statistics for data scientists : 50+ essential concepts using R and Python. O'reilly Media, Inc.
[4] Chollet, F. (2018). Deep Learning with Python. Manning, Cop.
[5] Gilbert, R. (2019, November 19). Freezing of Gait. American Parkinson Disease Association. https://www.apdaparkinson.org/article/freezing-gait-and-parkinsons-disease/
[6] Li, Hantao (2021), "Multimodal Dataset of Freezing of Gait in Parkinson's Disease", Mendeley Data, V3, doi: 10.17632/r8gmbtv7w2.3
[7] Müller, A. C., & Guido, S. (2017). Introduction to machine learning with Python : a guide for data scientists. O'reilly.
[8] Srivastava, P. (2020, May 18). Essentials of Deep Learning : Introduction to Long Short Term Memory. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/
[9] Time Series Classification Tutorial: Combining Static and Sequential Feature Modeling using Recurrent Neural Networks. (2022, May 5). Omdena | Building AI Solutions for Real-World Problems. https://omdena.com/blog/time-series-classification-model-tutorial/
[10] Zhang, W., Yang, Z., Li, H. et al. Multimodal Data for the Detection of Freezing of Gait in Parkinson's Disease. Sci Data 9, 606 (2022). https://doi.org/10.1038/s41597-022-01713-8