Machine Learning for Stalling Detection

George Mencoff | Blaire Lee
Boston University
Boston, MA, 02215
gmencoff@bu.edu
bslee@bu.edu

Abstract

Capillary stalling is an event that occurs in the neurovasculature that increases with aging and neurodegenerative models such as Alzheimer's disease. Through a novel technique, stalling events can be detected using a fluorescent dye, Oxyphor, as a function of time vs. absolute photon counts per iteration. During experiments, thousands of iterations are collected and these data are manually classified as stalling or not. The motivation behind this project was to implement a machine learning algorithm to effectively detect stalling within an iteration. In this report, we attempt to solve this simple binary classification problem (stall or no stall) using various machine learning algorithms such as dynamic time warping, support vector machine, and k-nearest neighbors.

1. Introduction

Capillary stalling is a transient stoppage in blood flow that can last from seconds to minutes. The measurement of stalling in the brain is an active area of research because elevated levels of stalling are observed in Alzheimer's Disease models, and administration of an adhesion blocking antibody (Anti-LY6G) has been shown to reduce stalling and improve cognition in mice[1].

Blood flow through capillaries can be measured by measuring the intensity of a fluorescent dye which is present in the blood plasma and not contained in the red blood cell. Because the dye is not contained in the red blood cells (RBCs), when an RBC passes in front of the measurement device, a drop in the intensity of the dye is observed. This drop happens at a regular rate when blood flow is normal, and stops when stalling occurs. By observing abnormal signals which do not exhibit signs of typical blood flow, stalling can be detected. The figure below illustrates an example of normal blood flow and stalling. The top image shows an oscillating signal with a regular frequency in the range of 70 Hz, which is indicative of normal blood flow as RBCs continue to pass in front of the measurement equipment. The bottom figure however does not show signs of red blood cells passing in front of the measurement equipment and was therefore classified as a stall.

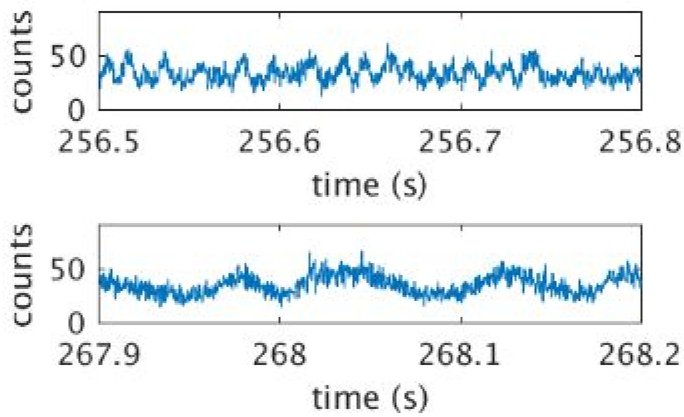

Figure 1: An example of normal blood flow (top) and an event that is believed to be a stall (bottom).

Currently, the process for stall detection is very manual. Once the data is collected, it is visually inspected for instances of stalling. This leads to slow processing times and potential misclassification of data. Our goal was to use previously classified data to build a machine learning model that could quickly and reliably detect stalling from these photon count signals.

2. Methods

The dataset was acquired by Jack Giblin, a PhD student in the Neurophotonics center at Boston University from *in vivo* experiments of wild-type mouse model (C57BL6) from Jackson Laboratory. The data consists of approximately 80,000 labeled and unlabeled data points that were used to train the ML models. Of these 80,000 data-points, 12,000 are labelled as non-stall events and ~400 are labelled as stall events. Each data point consists of a 1x1000 time series vector measuring the photon count over 0.3 seconds, a derived PO2 value, and a categorical classification of stall (1), no-stall (0), or unclassified (-1). The training to testing dataset ratio is 80:20.

2.1 Approach A

The first approach taken in classifying this data was directly comparing the data in the testing set to the data in the training set without feature extraction using a distance metric and a 1NN algorithm. This is a popular method for classifying time series because it is simple to implement and has proven very effective in many cases of time classification[2]. This approach is used to see if it would prove effective and to benchmark against other approaches.

Dynamic time warping (DTW) refers to the alteration of two signals in order to minimize their Euclidian distance by matching points from one signal to points in the other signal. In the figure below, an example of a time warped signal is shown. As a distance measure, the Euclidean distance after warping was used.

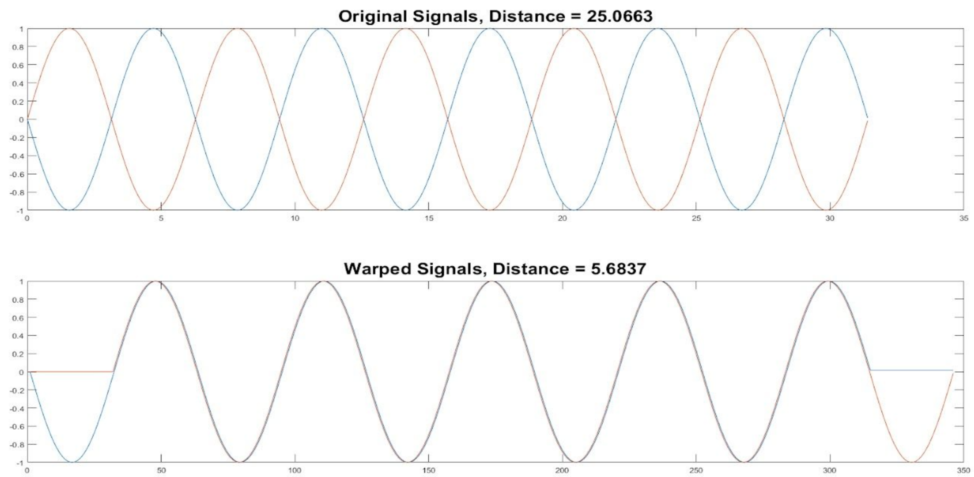

Figure 2: Distance between signals reduced using DTW

2.2 Approach B

Approach B extracting features that potentially capture key information about whether or not a signal was stalling. The table below shows a list of features that are extracted from the time signal, the information captured and the reasoning regarding why they were good features to use. Overall, the major differentiating characteristics between a stall and a no-stall case are signal frequency, randomness and deviation from the mean. Features were selected that were thought to capture this information.

Table 1: Features extracted from the signals

Main Features	**Definition**	**Significance**
Mean	Mean value of the time series data	Expect a lower mean for stalls
Standard Deviation	Standard deviation of the time series data	Expect a lower deviation for stalls
PO2 Value*	Partial pressure of oxygen in the vessel	Expect a lower PO2 value for stalls
Mean Spectral Entropy	Entropy of the signal in the frequency domain	Expect more randomness in stalls
Signal to Noise Ratio	Amount of power contained in the signal vs. noise	Expect more noise in stalls
Power Fraction 30-500 Hz	Power contained in frequency range of RBC passage	Expect less power in the RBC passage frequencies

DTW Distance to Non-Stalls	Dynamic time warping distance to 5 nearby non-stalls	Expect larger dtw distance between stall and non-stalls
DTW Distance to Stall	Dynamic time warping distance to 5 nearby stalls	Expect smaller dtw distance betweens stall and stalls
Peak to RMS	Ratio of peak amplitude to root mean square	Expected larger peaks in non-stalls
Signal Amplitude	Max signal value minus min signal value	Expected larger amplitude in non-stalls
Average Frequency	Average of signal in the frequency domain	Expect the mean frequency to be higher in stall cases
Median Frequency	Median of signal in frequency domain	Expect the median frequency to be higher in stall cases

Below is a figure which proves that many of the extracted features do indeed differ significantly between the stall and no-stall cases. Models were built using only features that were significantly different between the classes and using all features. The models which only used the significantly different features were found to outperform those which included all features, although this is likely due to overfitting and with further hyperparameter tuning would not be expected to occur.

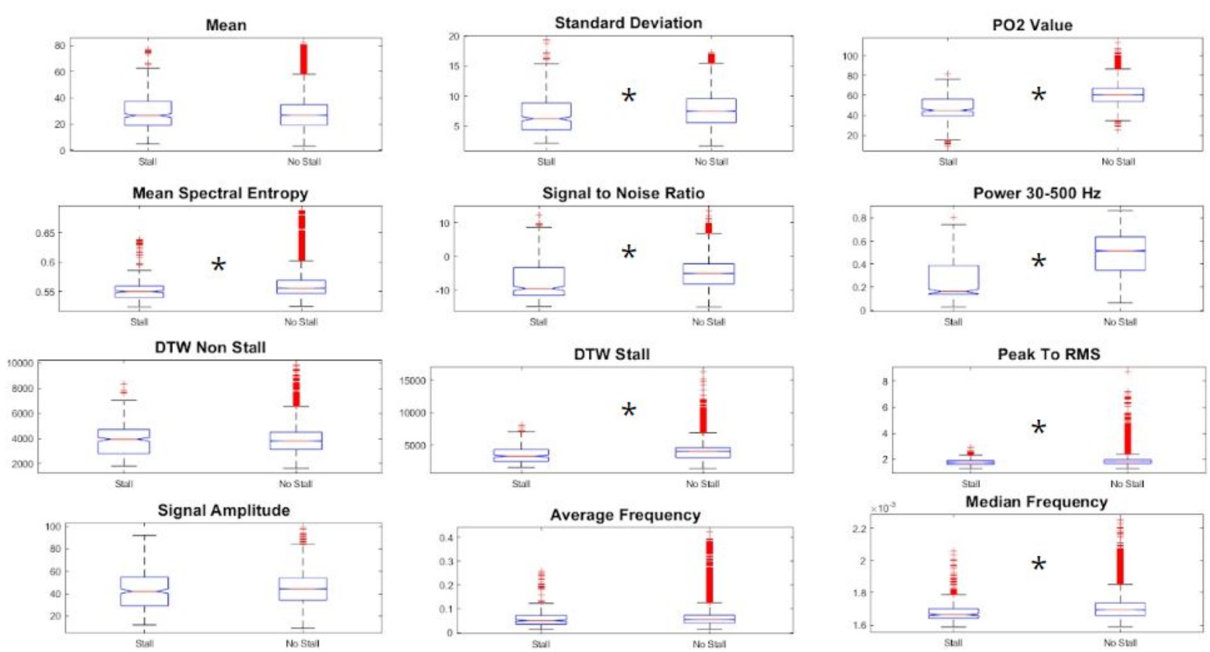

Figure 3: Variation of features between the two classes. An asterisk represents a statistically significant difference (P<.001)

With these newly extracted features, a few different classes of machine learning algorithms were tested to attempt to classify stalling events. The algorithms tested were SVM with different kernels, variations of KNN, and a few tree based approaches. SVM with a simple linear kernel was found to have the best performance of the algorithms tested.

2.21 Undersampling:

One of the major difficulties with this project is that stalling is a very rare event, so there is a large data imbalance between the stall case and the non-stall case. This data imbalance lead to models in which the accuracy was high because there are a large number of non-stall events to classify correctly, but stall events go largely undetected. In order to improve the accuracy rate in predicting stalls, models were developed using randomly undersampled training sets with a ratio of 1:1 and 1:4 stall:no stalls, whereas the full data set has about 3:100 stall:no stalls. It was found that using this simple undersampling technique not only significantly improved the accuracy of stall detection, but also increased the overall accuracy of the model.

3. Results

The results obtained from the DTW-KNN approach are shown in the figure below. This algorithm was tested on the 1:1 undersampled dataset to compensate for computational efficiency:

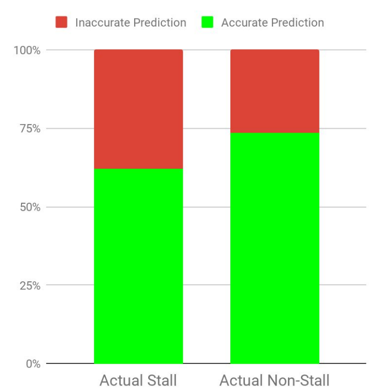

Figure 4: Dynamic Time Warping 1NN (Undersampled 1:1) Accuracy Results.

The feature extraction method proved to be much more effective than the DTW method. SVM was found to be the most effective classification algorithm, with only small differences occurring based on the kernel used. As mentioned previously, the most effective models were constructed based on undersampled data because of the large data imbalance in the training set. Results from all of the models based on these extracted features are shown in the tables below.

Table 2: Model Results with all extracted features as inputs

Sample Ratio (Stall:No Stall)	Model	True +	False +	True -	False -	Total Accuracy	Stall Accuracy
1:1	Cubic SVM	60	599	2294	18	0.79	0.77
1:1	Weighted KNN	57	465	2428	21	0.84	0.73
1:4	Cubic SVM	34	140	2753	44	0.94	0.43
1:4	Weighted kNN	36	178	2715	42	0.93	0.46
Full	Cubic SVM	23	45	2848	55	0.97	0.29
Full	Weighted kNN	9	5	2888	69	0.97	0.11

Table 3: Model Results with only features shown to be significantly different between classes shown as features

Sample Ratio (Stall:No Stall)	Model	True +	False +	True -	False -	Total Accuracy	Stall Accuracy
1:1	Linear SVM	50	155	2738	28	0.94	0.64
1:1	Cubic SVM	46	447	2446	32	0.84	0.59
1:1	Weighted KNN	58	420	2473	20	0.85	0.74
1:4	Weighted KNN	35	153	2740	43	0.93	0.45
1:4	Cubic SVM	29	99	2794	49	0.95	0.37
1:4	Linear SVM	40	8	2885	38	0.98	0.51
Real	Cubic SVM	15	2	2891	63	0.98	0.19
Real	Linear SVM	1	0	2893	77	0.97	0.01
Real	Weighted KNN	8	0	2893	70	0.98	0.10

4. Discussion and Conclusion

SVM provides a good foundation for building a useful machine learning algorithm for binary classification of stall detection. Using undersampled data, various models correctly predicted 65-75% of stall events as shown above in Table 1. As the stall versus no stall ratio became closer to the real dataset, the accuracy of true positive stalling events decreases while true negatives increases . There seems to be a trade off between using under sampled and the entire dataset. Furthermore, this data has been subjectively classified by a person so there is a possibility of human error in our initial dataset.

Our model has a lot of room for improvement. Capturing stalling information in the brain vasculature is a novel method that has been piloted in the BOAS lab in recent years. With more accumulation of quality data, we expect significant improvement in our SVM model as we will have better training data. One important thing to mention is that stalling events are fundamentally hard to catch and proportionately rare in healthy animal models (about ~3% of the whole dataset). Thus, even with more data collection, the data imbalance may continue to pose challenges and more sophisticated sampling techniques should be considered. Our current models were built on basic signal classification features that are commonly used for data analysis. We acknowledge that there are more intelligent feature extraction algorithms to be explored such as autoregression and optimizing hyperparameters.

5. References

[1] Hernandez, J.C.C, Bracko, O., Schaffer, C.B. Neutrophil adhesion in brain capillaries reduces cortical blood flow and impairs memory function in Alzheimer's disease mouse models. Nature Neuroscience, 22, 413-420, 2019.

[2] Jeong, Y-S., Jeong, M.K., Omitaomu, O. A. Weighted dynamic time warping for time series classification. Pattern Recognition, 2011. Volume 44, 9, 2231-2240.

6. Additional Resources and Acknowledgements

- https://www.mathworks.com/help/wavelet/examples/ecg-classification-using-wavelet-features.html
- https://arxiv.org/abs/1809.04356
- https://arxiv.org/pdf/1512.06747.pdf
- Thank you Sreekanth Kura (in-lab ML expert) for helping with coding
- Dataset was provided by the B.O.A.S lab of the Neurophotonics Center at Boston University, Boston, MA.