Painless Prognosis of Myasthenia Gravis using Machine Learning

Abhishek Tapadar (atapadar); Asherin George Anto George(asherin)

13 December 2018

Project category: Life-Sciences

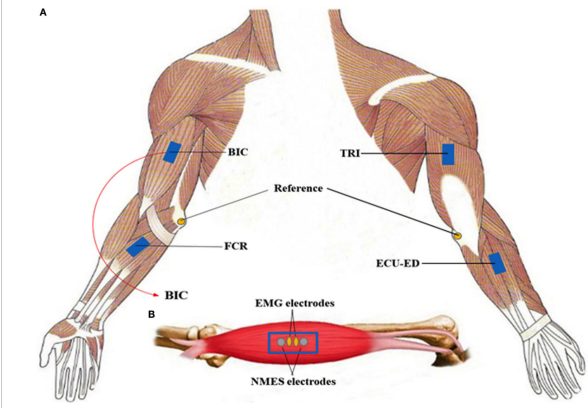
1 Abstract

We will implement and compare machine learning algorithms to predict with high confidence, the presence of a chronic condition, Myasthenia Gravis which affects about 200,000 people in the US alone. This should be helpful in eliminating the need for a painful and expensive Single-Fiber Electromyography(EMG) test and could potentially diagnose with a single anti–acetylcholine receptor (AChR) antibody (Ab) test. We have trained our algorithms on 22 co-factors/features commonly found with Myasthenia Gravis, and could also predict the probability of being afflicted with Myasthenia Gravis given a patient history and a questionnnaire.

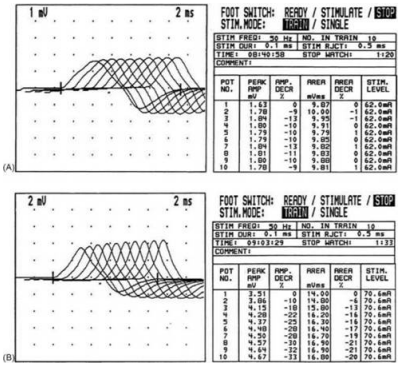
2 Introduction to Myasthenia Gravis

Myasthenia gravis (MG) is a neuromuscular disorder that causes weakness in the skeletal muscles, which are the muscles your body uses for movement. It occurs when communication between nerve cells and muscles becomes impaired. This impairment prevents crucial muscle contractions from occurring, resulting in muscle weakness.

Figure 1a illustrates a typical setup of a Single-Fiber Electromyography(EMG) test. Electromyography (EMG) measures muscle response or electrical activity in response to a nerve's stimulation of the muscle. The test is used to help detect neuromuscular abnormalities. An example electrical signal output from the repeated simulation, is shown in Figure 1b. Using many of these muscle simulations, it is possible to diagnose Myasthenia Gravis.



(a) Single Fiber Electromyography Schematic



(b) RNG Muscle Simulation Output

The input to our algorithm is anonymous medical data containing positive and negative labels for related medical tests, patient data such as age, gender and BMI and federally accepted metrics regarding the patient's quality of living and sleep. We then use logistic regression as our baseline model and compare the results with other algortihms such as GDA, CNN, Gradient Boosted Trees and Random Forests, with the ultimate goal of predicting if a patient has Myastenia Gravis or not.

We propose that with training set with the 22 cofactors and the Myasthenia diagnosis in existing patients, it would be possible for the system to correlate the occurrence of Myasthenia

with just a simple questionnaire and the AchR and MuSK phlebotomy tests and avoiding the painful EMG test. The computational cost is mainly in the form of training which is incurred only once upfront. Once the algorithm is trained, predictions can be produced in a fraction of the training time and would be beneficial to the patients.

3 Related Work

The state of art in medical history diagnosis using machine learning has primarily been applied to Imaging Diagnostics, in X-Rays and with Tumour Cell Imaging, It has not been studied on classification problems like diagnoses, and specifically not with Myasthenia Gravis. We have therefore assumed our baseline calibration with regression analysis which is standard in the field of medicine.

4 Dataset and Features

We propose that we can train the algorithms with a data set with 22 different co-factors related to Myasthnia Gravis and therefore, we can predict the occurance of Myasthenia Gravis using a combination of factors.

We thank Dr. Srikanth Muppidi (*muppids*) at the Stanford Neurosciences Hospital for his immense help in gathering anonymous patient records for us. We have a set of 10,056 data points with the above mentioned factors from a NIH repository for Myasthnia Gravis. We use a 95%:5% train:development set split to train and tune our model respectively. As for the test set, we use another separate set of 199 samples of anonymized data from patients only from the Menlo Park region, California.

The following are the various features each data sample holds.

1. Age Age of the candidate is a factor in diagnosing Myasthinia Gravis as it potentially affects people in advanced ages

2. Gender There are studies that show that more women are affected with Myasthnia Gravis than Men

3. BMI A higher BMI may correlate to a host of health problems and may be a feature in diagnosing Myasthenia

4. Years Diagnosed with MG This field is 0, if a patient does not have Myasthnia Gravis, but if diagnosed gives us an idea on the years they have been diagnosed

5. AChR Anti-bodies Approximately 85-90 percent of patients with myasthenia gravis (MG) express antibodies to the acetylcholine receptor (AChR)

6. MuSK Anti-bodies Useful for Diagnosis of autoimmune muscle-specific kinase (MuSK) myasthenia gravis. Second-order test to aid in the diagnosis of autoimmune myasthenia gravis when first-line serologic tests are negative

7. Presence of MuSK Ab and AChR Ab This is a field we have generated to account for presence of both antibodies and therefore might indicate a higher incidence of Myasthenia Gravis

8. Seronegative Around 10-20% of myasthenia gravis (MG) patients do not have acetylcholine receptor (AChR) antibodies (seronegative), of whom some have antibodies to a membrane-linked muscle specific kinase (MuSK). To examine MG severity and long-term prognosis in seronegative MG compared with seropositive MG, and to look specifically at anti-AChR antibody negative and anti-MuSK antibody negative patients.

9. Thymectomy Surgical removal of the thymus gland, may indicate a lowe incidence of Myasthnia Gravis

10. Sleep Apnea A potentially serious sleep disorder in which breathing repeatedly stops and starts.This is commonly incident with Myasthnia Gravis

11. Sleep Apnea Number This is an indication if the patient is aware of their affliction with Sleep Apnea

12. Non-Invasive Ventilation Support Noninvasive ventilation (NIV) refers to the provision of ventilatory support to the lungs, without the use of an endotracheal airway. It has emerged as an important tool in the treatment of acute respiratory failure, this field measure if the patient has ever been put on an NIV system

13. NIV number The number of times a candidate has been on NIV support

14. MG-QOL15 The MG-QOL15 is a brief survey, completed by the patient, that is designed to assess some aspects of "quality of life"

15. ESS The Epworth Sleepiness Scale (ESS) is a scale intended to measure daytime sleepiness that is measured by use of a very short questionnaire.

16. ESS is greater than 10 This indicates a strong affliction to daytime sleepiness and therefore increased chances of being prognosed with Myasthinia Gravis

17. PSQI The Pittsburgh Sleep Quality Index (PSQI) is a self-report questionnaire that assesses sleep quality over a 1-month time interval.

18. PSQI is greater than 5 This indicates a strong affliction to poor sleep rhythms due to fatigue and therefore increased chances of being prognosed with Myasthinia Gravis

19. FSS The Fatigue Severity Scale (FSS) is a method of evaluating the impact of fatigue on you. The FSS is a short questionnaire that requires you to rate your level of fatigue.

20. FSS is greater than 36 his indicates a strong affliction to increased fatigue and therefore increased chances of being prognosed with Myasthinia Gravis

21. MG ADL The MG-ADL profile provides a rapid assessment of MG symptom severity; it has been validated and shown to correlate with the QMG score.

22. MG ADL bulbar subset score A short questionnaire to find out the fatigue in the bulbar and throat region.

This data was pre-processed such that all text fields were transformed to take only numerical values (for instance, the gender field took '0' for 'male' patients and '1' for 'female' patients).

5 Methods

Various algorithms were used on the training data with **Logistic Regression** as the **baseline model** since it is the most widely used machine learning algorithm to classify and predict medical data.

5.1 Logistic Regression

We will focus on the binary classification problem in which y can take on only two values, 0 and 1. The logistic model is a widely used statistical model that, in its basic form, uses a logistic function to model a binary dependent variable. Cross-entropy loss is used to measure the performance of the model. Logistic regression models $p(y|x;\theta)$ as

$$h_\theta(x) = g(\theta^T x)$$

where g is the sigmoid function. By making significantly weaker assumptions, logistic regression is more robust and less sensitive to incorrect modeling assumptions.

5.2 Gaussian Discriminant Analysis

When we have a classification problem in which the input features x are continuous-valued random variables, we can then use the Gaussian Discriminant Analysis (GDA) model, which models p(x—y) using a multivariate normal distribution. The GDA model makes strong modelling assumptions and when these assumptions are correct, informally, there is no other algorithm that performs better.

5.3 Convolutional Neural Network

The advantage of using a CNN is that they often require only very little pre-processing. A 1D convolution was used across the 22 features of the input data. The following CNN architecture was used.

Layer (type)	Output Shape	Param #
$input_1$ (InputLayer)	(None, 22, 1)	0
conv0 (Conv1D)	(None, 8, 16)	128
bn0 (BatchNormalization)	(None, 8, 16)	64
$activation_1$ (Activation)	(None, 8, 16)	0
max $Pool_0$ (MaxPooling1D)	(None, 4, 16)	0
$flatten_1$ (Flatten)	(None, 64)	0
fc (Dense)	(None, 1)	65
$activation_2$ (Activation)	(None, 1)	0

Total params: 257 Trainable params: 225 Non-trainable params: 32

5.4 Random Forests

Trees have the capacity to learn highly complex data patterns depending on the depth of the trees. This will always however lead to over-fitting and hence increases the variance on the model. Thus random forests can be used, which can be imagined as a method of averaging across these deep decision trees by applying feature bagging. The bootstrapping procedure ensures that the variance of the final model is less. Moreover, if a certain feature is a strong predictor of the final response, then that feature will be selected in many of the decision trees.

6 Results

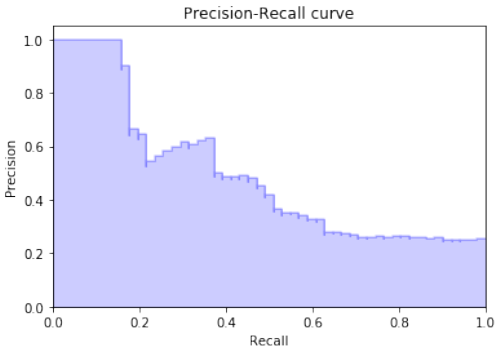
6.1 Metrics

We define three metrics that are very useful in calculating if our model performs well. They are: 1. Precision, 2. Recall and 3. F1 score.

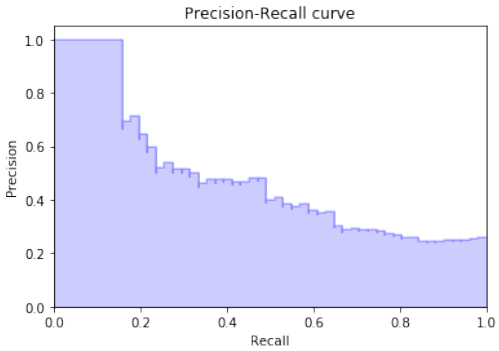
Metric Name	Formula #
Precision	$\frac{TruePositive}{TotalPredictedPositive}$
Recall	$\frac{TruePositive}{TotalActualPositive}$
F1 score	$2 * \frac{Precision*Recall}{Precision+Recall}$

6.2 Results from the models

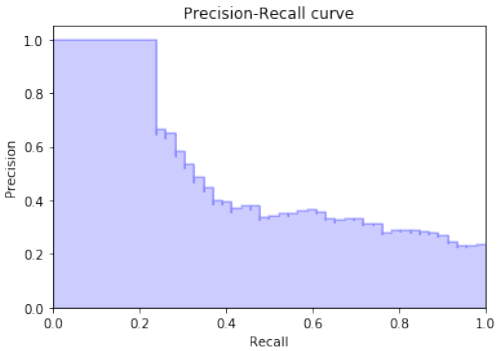
Model	Recall	F1 score #
Logistic Regression	0.72	0.837
GDA	0.745	0.854
CNN	0.829	0.906



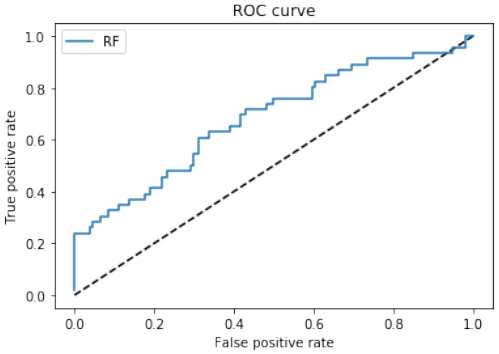
(a) Logistic Regression



(b) GDA



(c) Random Forests - PRC



(d) Random Forests - ROC

6.3 Discussion of Results

From our results, we can see that the CNN model performed really well since it has the highest F1 score amongst all the other models. The GDA model performs slightly better than the logistic regression. This can be expected since most of the features of the data are scores from questionnaires filled out by the patients themselves and hence, the scores should be treated to contain some sort of noise associated with them. Most natural processes tend to be normally distributed and hence the reason why GDA performs better than logistic regression. Random Forests overfit the data and so the depth of the tree was limited to 2, the maximum features to 3 and the maximum leaf nodes = 2 to prevent overfitting. The CNN model however, allowed for picking the features that strongly control the prediction since it used a convolution and max pooling layer.

7 Future Work Scope

Some input features might affect our final prediction more strongly than others. More often that not, in the medical sphere, medical records of patients are documents with very few of these features available. If we can narrow down the features that do not affect the final prediction severely, we can successfully predict if a person has myasthenia gravis simply by looking at existing medical records and open up new doors for more tests that will help with a more certain diagnosis.

It can be noticed that the input features excluding the AChR, MuSK and seronegativity tests, are side effects and related effects caused in a person suffering from Myasthenia gravis. Thus, if we knew the age, gender and BMI of a person suffering from Myasthenia gravis and the remaining features, we could extend our model to predict the existence of other effects such as sleep apnea or if a thymectomy could help lessen the symptoms.

8 Contributions

The team consisted of two members: 1. Abhishek Tapadar and 2. Asherin George Anto George. Both the team members helped reach out to the professionals at the Stanford Neurosciences Hospital in order to obtain the data required. Most of the discussions and decisions regarding this project were also done together.

Abhishek Tapadar contributed in writing code and deriving results from the Logistic Regression, GDA and CNN models. He also helped in deciding upon the hyperparameters for the above models and the gradient boosted and random forests model. In addition, he helped in putting together the report for the various intermediate milestones and the final project report.

Asherin George Anto George contributed in performing the gradient boosted and random forests models. He also helped in hyperparameter selection for the models Abhishek Tapadar was involved in. He contributed in putting together the write-up for the project proposal, milestone, poster presentation and the final project report.

9 References

[1] Chiou-Tan, F. Y. and Gilchrist, J. M. (2015), Repetitive nerve stimulation and single-fiber electromyography in the evaluation of patients with suspected myasthenia gravis or Lambert–Eaton myasthenic syndrome: Review of recent literature. Muscle Nerve, 52: 455-462. doi:10.1002/mus.24745

[2]Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Liu, P.J., Liu, X., Marcus, J., Sun, M. and Sundberg, P., 2018. Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1), p.18.

[3]H. Yu, Towards answering biological questions with experimental evidence: automatically identifying text that summarize image content in full-text articles, in: Proceedings of the AMIA 2006 Symposium, AMIA, 2006, pp. 834–838.

[4]Python libraries that were used: Numpy, Scikit Learn and Keras

10 Code Repository

Link to our code: `https://drive.google.com/file/d/1PsVVrVfs-g3Yp0kjofUi_MNkyd_mS0My/view`