Painless Prognosis of Myasthenia Gravis using Machine Learning

CS229
AUT 18

Abhishek Tapadar, atapadar@stanford.edu; Asherin George Anto George, asherin@stanford.edu

1. What Are We Predicting? Myasthenia Gravis Diagnosis

- Implement and compare machine learning algorithms to predict with high confidence, the presence of a chronic condition, Myasthenia Gravis which affect about 200,000 people in the US alone.
- 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 questionnaire.

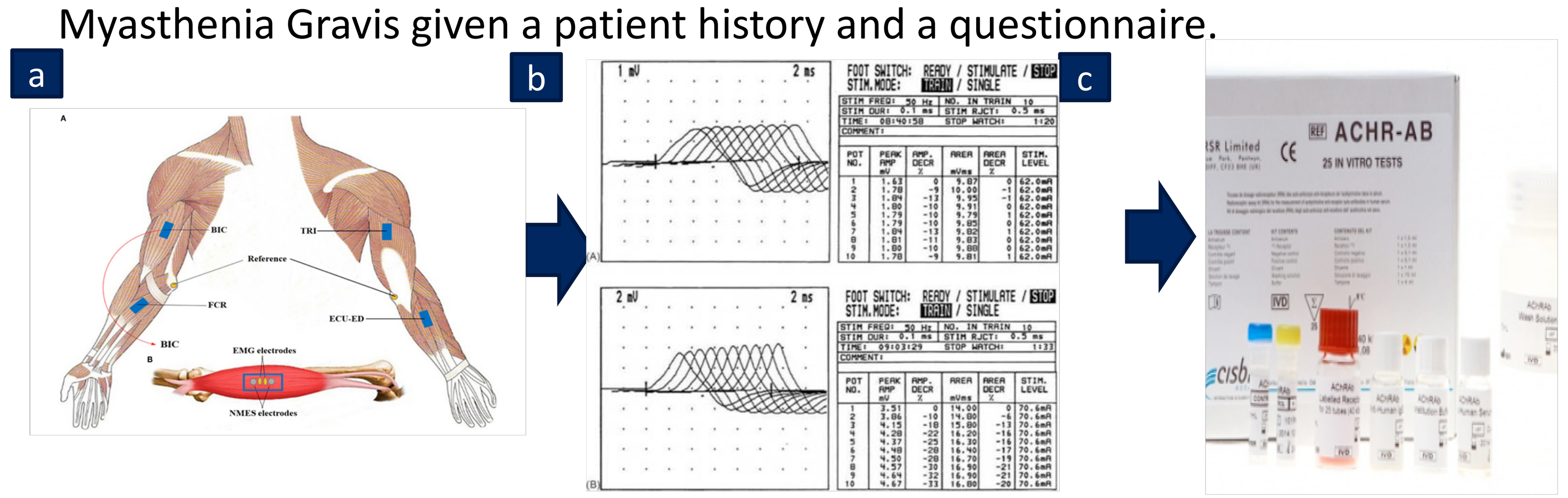


Figure 1. Traditional Myasthenia Gravis diagnosis: (a) Repeated Nerve Simulations, (b) RNS data showing positive Myasthenia Gravis, and (c) AcHR and MuSK tests.

2. How Are We Going To Do It? Workflow:

To ease the problem set above, we divided the problem into the following steps.
In addition, we have divided the problem into sub problems:

a) Transform the NIH dataset into metrics for the algorithms (Fig 2a).
b) Prediction Diagnosis on a test-set with Logistic regression as baseline
c) Use GDA, CNN's, Naïve Baye's and Random Forests and calculate F1 scores.

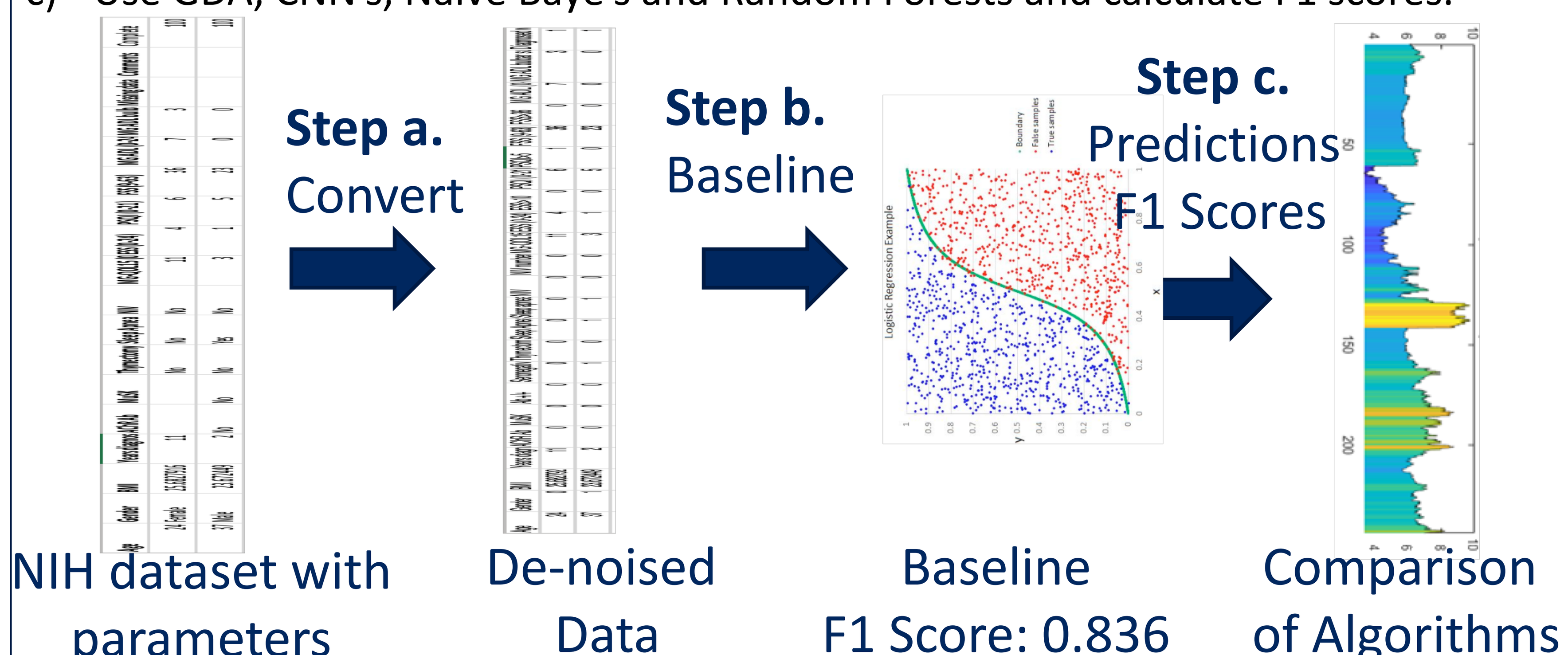


Figure 2. Workflow.

3. What Training Data Do We Have? 12056 patient data

- We have obtained the NIH patient dataset with 22 cofactors.
- Made possible by the contribution of Dr. Muppiddi at Stanford Neurosciences.
- We have 22 cofactors in the dataset to predict we trained our Algorithms on

- **Age**
- **Gender**
- **BMI**
- **Years diagnosed with MG**
- **AcHR Antibodies**
- **MuSK Antibodies**
- **MuSK and AcHR present**
- **Seronegative**
- **Thymectomy**
- **FSS is greater than 36**
- **Sleep Apnea**
- **Non-Invasive Ventilation Support**
- **NIV number**
- **MG-QoL 15**
- **ESS**
- **PSQI**
- **FSS**
- **MG-ADL**
- **MG- ADL Bulbar**

3b. Transform the data to metrics for the algorithm

Real Data Obtained:

- The obtained data-set contained scores and data which were not readily useable for training algorithms.
- We converted the raw data into trainable metrics for comparison.

4. Step a: Baseline Model, Logistic Regression

We have used Logistic Regression as the baseline model as used in several other Medical Literature. This is convenient and gives us a baseline to compare the other algorithms. The data set was shuffled and split into 10000 data points for training and 2056 data points for the dev-set. We then compute the recalls, and the F-! Scores on the dev-set. The following is a summary of the results

Recall	F1 Score
0.72	0.836

5. Step b: GDA, Naïve Bayes, CNN

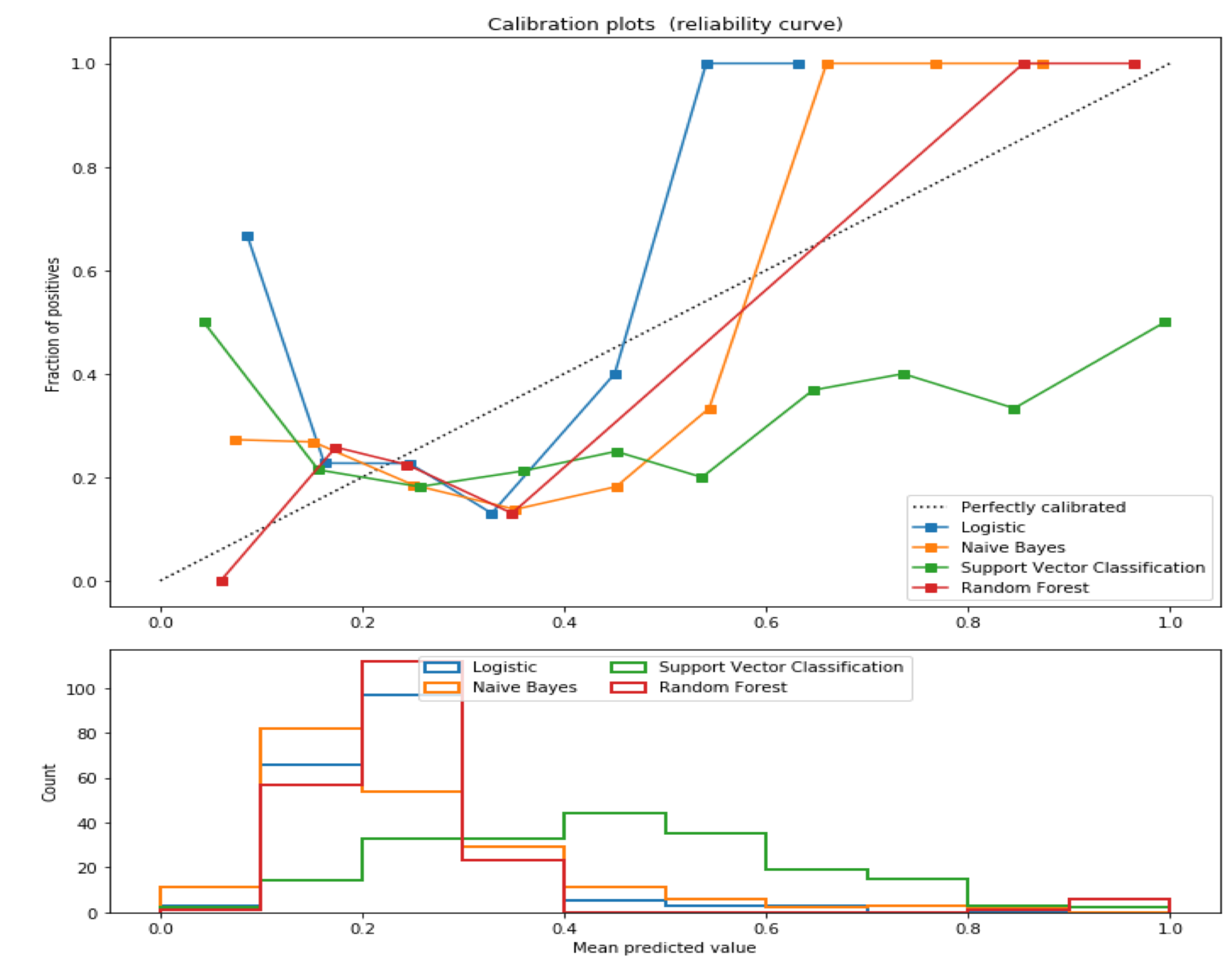
- We implemented GDA, Naïve Bayes, CNN using Keras, and Random Forests to obtain the recall and the F1 Scores

	Recall	F-1 Score
GDA	0.768	0.869
CNN	0.829	0.906
Naïve Bayes	0.742	0.845

6. CNN Architecture

```
Layer (type)                 Output Shape              Param #
=================================================================
input_40 (InputLayer)        (None, 22, 1)             0
conv0 (Conv1D)               (None, 8, 16)             128
bn0 (BatchNormalization)     (None, 8, 16)             64
activation_28 (Activation)   (None, 8, 16)             0
max_pool0 (MaxPooling1D)     (None, 4, 16)             0
flatten_13 (Flatten)         (None, 64)                0
fc (Dense)                   (None, 1)                 65
activation_29 (Activation)   (None, 1)                 0
=================================================================
Total params: 257
Trainable params: 225
Non-trainable params: 32
```

7. Random Forests and Comparisons



10. Summary Results

Algorithm	Recall	F-1 Score
Log-Reg(Baseline)	0.718	0.836
GDA	0.768	0.869
Naïve Bayes	0.742	0.845
CNN(with CE loss)	0.829	0.906
Random forests	0.846	0.897

11. Discussion

- The CNN model provided us with the highest F1 score, implying a good classifier model
- We see that, Random Forests perform well too as a classifier
- As CNN is a a good fit, this implies that some of the co-factors are better predictors than others

12. Future

- The model can be extended to predict the existence of other side effects of myasthenia gravis
- To achieve better calibration, multiple algorithms can be used simultaneously
- Improving the quality of life would be the final goal to any life-sciences project

13. 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