

Machine Learning with Python to Diagnose Chiari Malformation Type I

Oliver Evans
Dr. Malena Español
University of Akron

May 13, 2016

1 Introduction

We have built on the work done by Urbizu et al. [2] in investigating the potential for machine learning in assisting in the diagnosis of Chiari Malformation Type I (CMI). MRI data has been shown to be very useful in diagnosing CMI. Currently, Tonsillar Descent is the primary feature used in diagnosis, although it does not always correctly characterize CMI. We seek to find a correlation between other MRI measurements and CMI diagnosis.

We have used Python and the machine learning algorithms included in the scikit-learn [1] package. We used the following algorithms in our investigation: Support Vector Machines (SVM), 4 Nearest Neighbors (4NN), Decision Tree (DT), Linear Discriminant Analysis (LDA), Naive-Bayesian (NB), Linear Regression (LR)

2 Data

We have MRI measurements from 100 CMI patients and 50 controls (non-CMI) for the following features from Urbizu et al. We also have data about symptoms for each of the CMI patients.

Feature #	Feature Name	Symptom Name	Symptom Name
0	TD	Asymptomatic	Fatigue
1	Tentorium	Occipital neck headace	Instability
2	Supraoccipital	Bifrontal headace	Sensory Loss
3	FM	Holocranial headace	Anxiety
4	Clivus	Hemicranial headace	Depression
5	FCP area	Any Headache	Motor Weakness
6	Osseus FCP area	Hydrocephalus	Dysphagia
7	Corpus Callosum to FM	Syringomyelia	Dysphonia
8	Height PCF	Pseudotumor Cerebri	Gait Disturbances
9	Width PCF	Complex Craniocervical Malformation	Paresthesia Upper Limbs
10	Fastigium to FM	Klippel-Feil Malformation	Paresthesia Lower Limbs
11	Pons to FM	Retrocurved Odontoid	Upper Limbs Pain
12	Basal angulation	Other CVJ Malformations	Lower Limbs Pain
13	Tentorium angle	Cough Headaches	Neck Pain
14	Wackenheim angle	Dizziness	Difficulty Swallowing
15	Basilar impression	Vertigo	Nystagmus
16	Odontoid angle	Visual Alterations	Kyphoscoliosis

Figure 1: List of Features and Symptoms

3 Method

Our primary goal was to effectively distinguish CMI patients from controls based on other MRI symptoms. To do this, we took all possible unordered combinations of 2-7 features and applied all of the above-listed algorithms in order to classify all samples as CMI or non-CMI, where a sample is either a CMI patient or a control. Let each of these combinations of features be called a tuple and each application of an algorithm to a tuple be called a trial. In each trial, the samples were divided into five sets each containing similar proportions of CMI and non-CMI samples. We used each set as a training dataset while the union of the other sets was used for testing. We calculated the average sensitivity and specificity obtained by using each of the five training sets, and reported these as the sensitivity and specificity for the trial. Then, for each algorithm, for each tuple size, we selected the tuple of that size which most successfully distinguished between CMI and non-CMI samples. In the following pages, we report the sensitivity, specificity, and tuple which we deemed most successful for each tuple size and algorithm.

Our next step was to attempt to isolate our CMI patients (ignore controls) and attempt to predict whether or not each patient experienced each of the above-listed symptoms. We performed a slightly simplified version of the previous procedure for each symptom, where we chose only the single combination of tuple and algorithm which most effectively distinguished between CMI patients and controls.

4 Defining Effective Classification

In seeking to choose the best among a set of trials, it becomes necessary to define a metric for comparison. For each trial, we know primarily the sensitivity and specificity, and so we will use these as the building blocks of our metrics. We use four metrics for comparing trials, each of which is a simple function of sensitivity and specificity. We call them “sum”, “prod” for product, “perf” for performance, and “alpha” for lack of a better term. Each of these metrics aims to favor trials with higher sensitivity and specificity. We should begin by defining the sensitivity and specificity of a trial. Let the events C and NC represent an individual having CMI and not having CMI respectively. Let the events TP (“test positive”) and TN (“test negative”) represent an individual being classified by the algorithm (perhaps incorrectly) as CMI and non-CMI respectively. We denote the sensitivity by “sens” and the specificity by “spec”.

Then for a given trial,

$$\text{sens} = P(TP|C) \quad (1)$$

$$\text{spec} = P(TN|NC) \quad (2)$$

We should also state two other useful probability relationships.

Firstly, Bayes’ Theorem:

$$P(A|B)P(B) = P(B|A)P(A) = P(A) \cap P(B) \quad (3)$$

Secondly, if $P(A) + P(B) = 1$, then

$$P(C|A)P(A) + P(C|B)P(B) = P(C) \quad (4)$$

Note that $P(TP) + P(TN) = 1$ and $P(C) + P(NC) = 1$.

4.1 sum

Our first metric is a simple sum. It fails to sufficiently penalize zero-values of sensitivity or specificity, which, in some cases results in inappropriately choosing a trial which has entirely failed, classifying the all of the samples as either CMI or non-CMI, yielding a sensitivity of 1 and specificity of 0 or vice versa.

$$\text{sum} = \text{sens} + \text{spec} \quad (5)$$

4.2 prod

Taking the product of the sensitivity and specificity successfully penalizes zero or almost-zero values for sensitivity or specificity, avoiding total failures which were problematic in the previous case.

$$\text{prod} = \text{sens} \cdot \text{spec} \quad (6)$$

4.3 perf

Another straightforward metric is simply the percentage of samples which were correctly classified, regardless of whether they are CMI patients or controls. This can be represented as

$$\begin{aligned} \text{perf} &= P(TP \cap C) + P(TN \cap NC) \\ &= P(TP|C)\tilde{P}(C) + P(TN|NC)\tilde{P}(NC) \end{aligned}$$

where \tilde{P} represents the occurrence within our data set, not within the general population. Then since we have 100 CMI patients and 50 controls, $\tilde{P}(C) = \frac{2}{3}$ and $\tilde{P}(NC) = \frac{1}{3}$. Therefore,

$$\text{perf} = \frac{2}{3} \text{sens} + \frac{1}{3} \text{spec} \quad (7)$$

4.4 alpha

Sensitivity and specificity are useful classifiers of a statistical test, but they are not so useful for a patient receiving a diagnosis. A patient is not interested in the probability that they will be correctly classified *if* they have the disease. Rather they are concerned with the probability that if they are classified, they *do* have the disease. Therefore, we would like to use the quantity $P(C|TP)$ as a metric. We call this quantity alpha ($\tilde{\alpha}$).

$$\begin{aligned}
\tilde{\alpha} &= P(C|TP) \\
&= \frac{P(TP|C)P(C)}{P(TP)} \\
&= \frac{P(TP|C)P(C)}{P(TP|C)P(C) + P(TP|NC)P(NC)} \\
&= \frac{P(TP|C)P(C)}{P(TP|C)P(C) + P(TP|NC)(1 - P(C))} \\
&= \frac{P(TP|C)P(C)}{P(TP|NC) + P(C)[P(TP|C) - P(TP|NC)]} \\
&= \frac{P(TP|C)}{\frac{1}{P(C)}P(TP|NC) + P(TP|C) - P(TP|NC)} \\
&= \frac{P(TP|C)}{P(TP|NC)(\frac{1}{P(C)} - 1) + P(TP|C)} \\
&= \frac{1}{\frac{P(TP|NC)}{P(TP|C)}(\frac{1}{P(C)} - 1) + 1} \\
&= \frac{1}{\frac{1 - P(TN|NC)}{P(TP|C)}(\frac{1}{P(C)} - 1) + 1} \\
&= \left[\frac{1 - \text{spec}}{\text{sens}} \left(\frac{1}{P(C)} - 1 \right) + 1 \right]^{-1}
\end{aligned}$$

Notice that we now have $P(C)$ in our equation, which is the prevalence of CMI among the whole population. We assume that this is unknown. Luckily, we can define a new quantity, α , and show that using α as a metric for comparison is equivalent to using $\tilde{\alpha}$.

Let

$$\alpha = \frac{\text{sens}}{1 - \text{spec}}. \quad (8)$$

Assume we have two trials, A and B which we seek to compare. Let x_A and y_A be sens and spec values for trial A respectively, and let x_B and y_B be the corresponding values for trial B. Assume it has been shown that by comparing $\tilde{\alpha}$ values, trial B has been deemed more effective. That is,

$$\begin{aligned}
& \tilde{\alpha}_A < \tilde{\alpha}_B \\
& \Leftrightarrow \\
& \left[\frac{1-y_A}{x_A} \left(\frac{1}{P(C)} - 1 \right) + 1 \right]^{-1} < \left[\frac{1-y_B}{x_B} \left(\frac{1}{P(C)} - 1 \right) + 1 \right]^{-1} \\
& \Leftrightarrow \\
& \frac{1-y_B}{x_B} \left(\frac{1}{P(C)} - 1 \right) + 1 < \frac{1-y_A}{x_A} \left(\frac{1}{P(C)} - 1 \right) + 1 \\
& \Leftrightarrow \\
& \frac{1-y_B}{x_B} < \frac{1-y_A}{x_A} \\
& \Leftrightarrow \\
& \frac{x_A}{1-y_A} < \frac{x_B}{1-y_B} \\
& \Leftrightarrow \\
& \alpha_A < \alpha_B
\end{aligned}$$

Since all of the above operations hold in both directions (assuming $0 < P(C)$, $\text{sens}, \text{spec} < 1$), α and $\tilde{\alpha}$ are equivalent metrics.

5 Results

In the following pages, we present the tables discussed previously: The sensitivity, specificity and feature numbers of the tuple for which each algorithm is most effective for each tuple size. These results are given for all of the previously discussed metrics. Also shown is a contour map of each metric as a function of sensitivity and specificity. A trial which receives a sensitivity and specificity rating pair which falls, for example, on a red contour line in the contour map will be judged by that metric as superior to a trial which results in a rating pair on a blue contour line.

On the final page of results is the outcome of our attempt to predict individual symptoms from MRI measurements. The single best trial according to the “prod” metric is shown in the table in all cases where classification was successful, along with the prod, sens, spec, and perf values for that trial. In some symptoms for which there was not enough data to successfully attempt classification, a blank line is shown. Above the table is a bar chart comparing the four previously-mentioned quantities for each symptom with at least one successful trial.

5.1 sum

$$\text{sum} = \text{sens} + \text{spec}$$

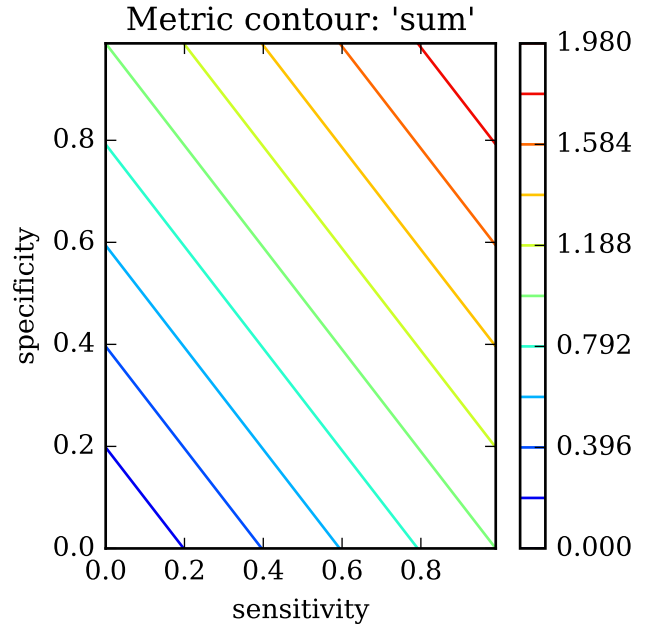
(9)

	SVM	4NN	DT	LDA	QDA	NB	LR
2	0.95	0.90	0.92	0.95	0.94	0.88	0.95
3	0.92	0.92	0.88	0.94	0.96	0.90	0.95
4	0.96	0.88	0.90	0.95	0.96	0.90	0.95
5	0.94	0.91	0.90	0.95	0.95	0.91	0.94
6	0.93	0.91	0.91	0.95	0.95	0.89	0.94
7	0.95	0.91	0.91	0.91	0.95	0.87	0.94

Figure 2: Sensitivity (sum)

	SVM	4NN	DT	LDA	QDA	NB	LR
2	0.78	0.84	0.72	0.76	0.80	0.88	0.76
3	0.84	0.86	0.76	0.82	0.82	0.90	0.78
4	0.82	0.92	0.78	0.84	0.82	0.90	0.84
5	0.84	0.92	0.78	0.84	0.86	0.88	0.86
6	0.88	0.92	0.78	0.86	0.84	0.90	0.86
7	0.88	0.92	0.78	0.90	0.84	0.92	0.86

Figure 3: Specificity (sum)



	SVM	4NN	DT
2	(12, 15)	(11, 12)	(8, 11)
3	(11, 12, 13)	(12, 15, 16)	(8, 10, 11)
4	(6, 12, 13, 15)	(4, 8, 9, 12)	(5, 6, 8, 16)
5	(7, 8, 12, 13, 15)	(6, 7, 8, 11, 12)	(8, 9, 12, 13, 16)
6	(7, 10, 11, 12, 13, 15)	(5, 6, 7, 8, 11, 12)	(2, 5, 8, 11, 13, 16)
7	(4, 8, 9, 11, 12, 13, 15)	(3, 6, 8, 11, 12, 15, 16)	(3, 5, 6, 7, 8, 11, 16)

	LDA	QDA	NB
2	(12, 15)	(12, 15)	(11, 12)
3	(11, 12, 15)	(11, 12, 15)	(11, 12, 13)
4	(11, 12, 13, 15)	(4, 11, 12, 15)	(11, 12, 13, 14)
5	(4, 6, 12, 13, 15)	(8, 9, 11, 13, 15)	(3, 4, 11, 12, 15)
6	(8, 9, 11, 12, 13, 15)	(2, 3, 7, 8, 11, 15)	(2, 10, 11, 12, 13, 16)
7	(2, 8, 11, 12, 13, 15, 16)	(4, 5, 8, 9, 11, 13, 15)	(7, 8, 9, 11, 12, 13, 14)

	LR
2	(12, 15)
3	(4, 12, 15)
4	(11, 12, 13, 15)
5	(7, 11, 12, 13, 15)
6	(5, 7, 11, 12, 13, 15)
7	(4, 6, 7, 11, 12, 13, 15)

Figure 4: Best Tuples (sum)

5.2 prod

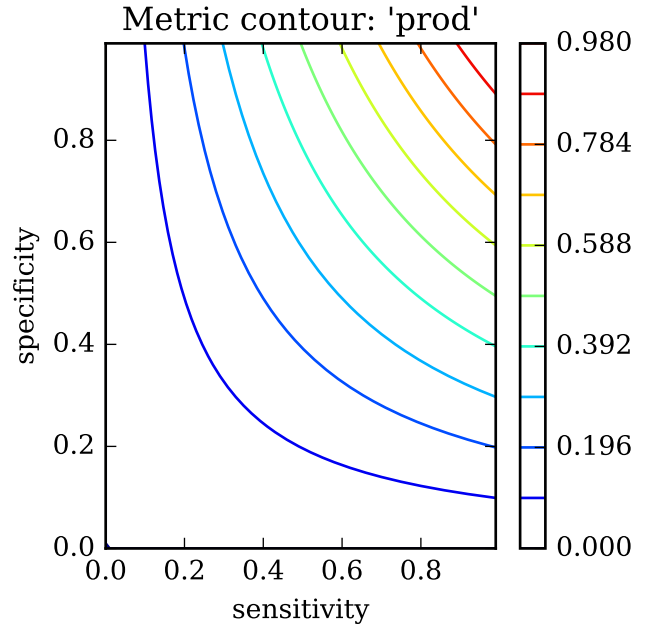
$$\text{prod} = \text{sens} \cdot \text{spec} \quad (10)$$

	SVM	4NN	DT	LDA	QDA	NB	LR
2	0.95	0.90	0.92	0.95	0.94	0.88	0.95
3	0.92	0.92	0.88	0.94	0.96	0.90	0.95
4	0.96	0.88	0.90	0.95	0.96	0.90	0.95
5	0.90	0.91	0.90	0.95	0.95	0.91	0.94
6	0.93	0.91	0.91	0.95	0.95	0.89	0.94
7	0.95	0.91	0.91	0.91	0.95	0.87	0.94

Figure 5: Sensitivity (prod)

	SVM	4NN	DT	LDA	QDA	NB	LR
2	0.78	0.84	0.72	0.76	0.80	0.88	0.76
3	0.84	0.86	0.76	0.82	0.82	0.90	0.78
4	0.82	0.92	0.78	0.84	0.82	0.90	0.84
5	0.88	0.92	0.78	0.84	0.86	0.88	0.86
6	0.88	0.92	0.78	0.86	0.84	0.90	0.86
7	0.88	0.92	0.78	0.90	0.84	0.92	0.86

Figure 6: Specificity (prod)



	SVM	4NN	DT
2	(12, 15)	(11, 12)	(8, 11)
3	(11, 12, 13)	(12, 15, 16)	(8, 10, 11)
4	(6, 12, 13, 15)	(4, 8, 9, 12)	(5, 6, 8, 16)
5	(6, 10, 11, 12, 16)	(6, 7, 8, 11, 12)	(8, 9, 12, 13, 16)
6	(7, 10, 11, 12, 13, 15)	(5, 6, 7, 8, 11, 12)	(2, 5, 8, 11, 13, 16)
7	(4, 8, 9, 11, 12, 13, 15)	(3, 6, 8, 11, 12, 15, 16)	(3, 5, 6, 7, 8, 11, 16)

	LDA	QDA	NB
2	(12, 15)	(12, 15)	(11, 12)
3	(11, 12, 15)	(11, 12, 15)	(11, 12, 13)
4	(11, 12, 13, 15)	(4, 11, 12, 15)	(11, 12, 13, 14)
5	(4, 6, 12, 13, 15)	(8, 9, 11, 13, 15)	(3, 4, 11, 12, 15)
6	(8, 9, 11, 12, 13, 15)	(2, 3, 7, 8, 11, 15)	(2, 10, 11, 12, 13, 16)
7	(2, 8, 11, 12, 13, 15, 16)	(4, 5, 8, 9, 11, 13, 15)	(7, 8, 9, 11, 12, 13, 14)

	LR
2	(12, 15)
3	(4, 12, 15)
4	(11, 12, 13, 15)
5	(7, 11, 12, 13, 15)
6	(5, 7, 11, 12, 13, 15)
7	(4, 6, 7, 11, 12, 13, 15)

Figure 7: Best Tuples (prod)

5.3 perf

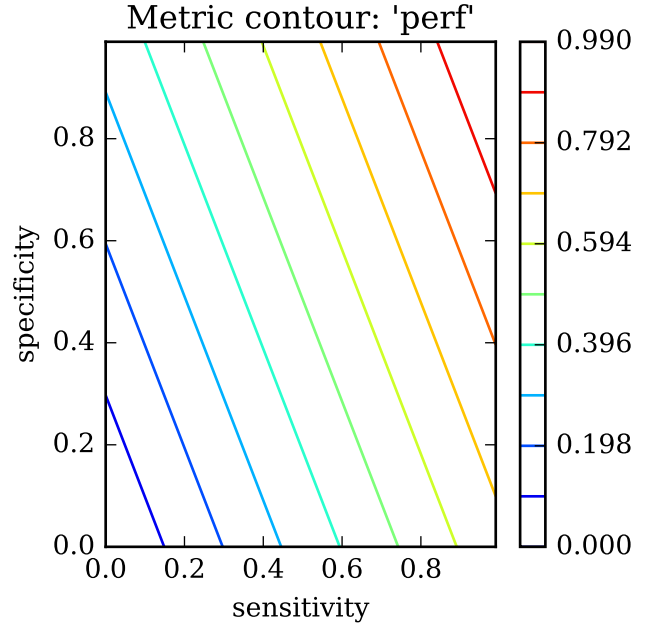
$$\text{perf} = \frac{2}{3} \text{sens} + \frac{1}{3} \text{spec} \quad (11)$$

	SVM	4NN	DT	LDA	QDA	NB	LR
2	0.95	0.90	0.92	0.95	0.94	0.91	0.95
3	0.96	0.92	0.88	0.96	0.96	0.92	0.95
4	0.96	0.91	0.90	0.95	0.96	0.92	0.95
5	0.96	0.91	0.90	0.95	0.95	0.91	0.95
6	0.95	0.91	0.91	0.95	0.95	0.92	0.94
7	0.95	0.95	0.91	0.95	0.96	0.93	0.94

Figure 8: Sensitivity (perf)

	SVM	4NN	DT	LDA	QDA	NB	LR
2	0.78	0.84	0.72	0.76	0.80	0.84	0.76
3	0.80	0.86	0.76	0.78	0.82	0.86	0.78
4	0.82	0.88	0.78	0.84	0.82	0.86	0.84
5	0.82	0.92	0.78	0.84	0.86	0.88	0.84
6	0.86	0.92	0.78	0.86	0.84	0.86	0.86
7	0.88	0.88	0.78	0.86	0.82	0.86	0.86

Figure 9: Specificity (perf)



	SVM	4NN	DT
2	(12, 15)	(11, 12)	(8, 11)
3	(12, 13, 15)	(12, 15, 16)	(8, 10, 11)
4	(6, 12, 13, 15)	(7, 8, 12, 15)	(5, 6, 8, 16)
5	(8, 9, 11, 12, 15)	(6, 7, 8, 11, 12)	(8, 9, 12, 13, 16)
6	(5, 7, 8, 12, 13, 15)	(5, 6, 7, 8, 11, 12)	(2, 5, 8, 11, 13, 16)
7	(4, 8, 9, 11, 12, 13, 15)	(3, 5, 11, 12, 13, 15, 16)	(3, 5, 6, 7, 8, 11, 16)

	LDA	QDA	NB
2	(12, 15)	(12, 15)	(8, 12)
3	(12, 13, 15)	(11, 12, 15)	(11, 12, 15)
4	(11, 12, 13, 15)	(4, 11, 12, 15)	(3, 11, 12, 15)
5	(4, 6, 12, 13, 15)	(8, 9, 11, 13, 15)	(3, 4, 11, 12, 15)
6	(8, 9, 11, 12, 13, 15)	(2, 3, 7, 8, 11, 15)	(3, 4, 10, 11, 12, 15)
7	(4, 6, 8, 11, 12, 13, 15)	(2, 4, 10, 11, 12, 13, 15)	(3, 4, 10, 11, 12, 14, 15)

	LR
2	(12, 15)
3	(4, 12, 15)
4	(11, 12, 13, 15)
5	(3, 11, 12, 13, 15)
6	(5, 7, 11, 12, 13, 15)
7	(4, 6, 7, 11, 12, 13, 15)

Figure 10: Best Tuples (perf)

5.4 alpha

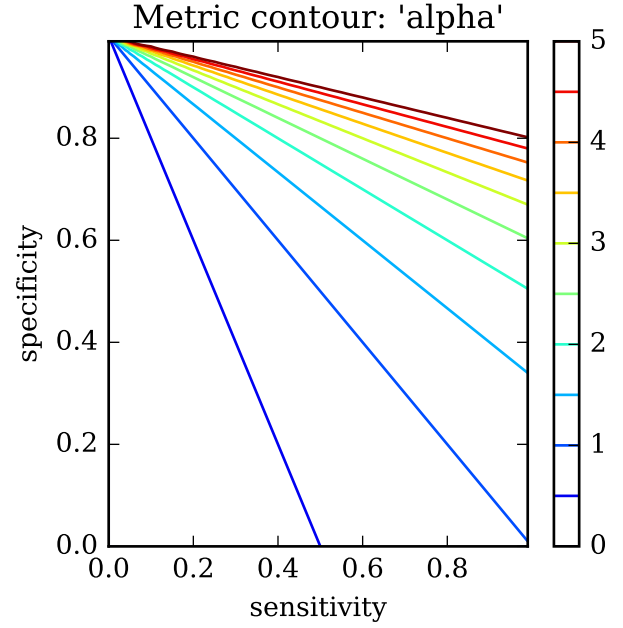
$$\alpha = \frac{\text{sens}}{1 - \text{spec}} \quad (12)$$

	SVM	4NN	DT	LDA	QDA	NB	LR
2	0.95	0.90	0.92	0.95	0.94	0.88	0.95
3	0.92	0.90	0.88	0.94	0.96	0.90	0.95
4	0.93	0.88	0.90	0.95	0.93	0.90	0.95
5	0.90	0.88	0.87	0.95	0.95	0.89	0.94
6	0.93	0.88	0.84	0.92	0.95	0.89	0.94
7	0.95	0.88	0.82	0.91	0.95	0.87	0.94

Figure 11: Sensitivity (alpha)

	SVM	4NN	DT	LDA	QDA	NB	LR
2	0.78	0.84	0.72	0.76	0.80	0.88	0.76
3	0.84	0.88	0.76	0.82	0.82	0.90	0.78
4	0.84	0.92	0.78	0.84	0.84	0.90	0.84
5	0.88	0.94	0.80	0.84	0.86	0.90	0.86
6	0.88	0.94	0.82	0.88	0.84	0.90	0.86
7	0.88	0.94	0.82	0.90	0.84	0.92	0.86

Figure 12: Specificity (alpha)

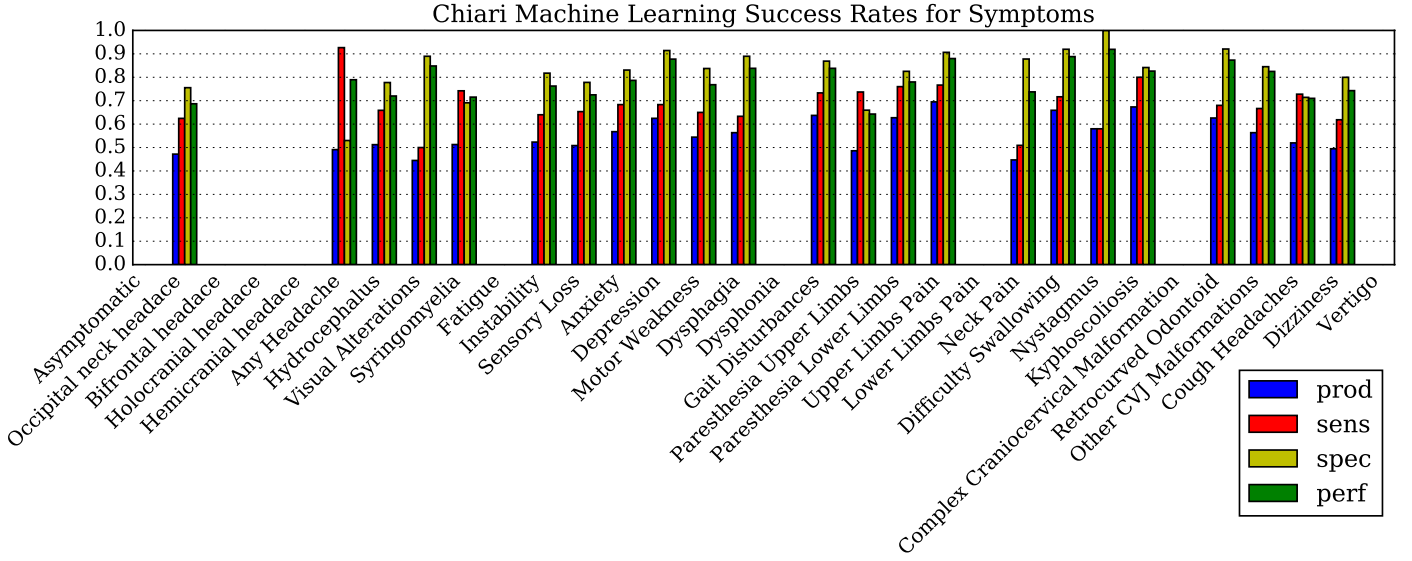


	SVM	4NN	DT
2	(12, 15)	(11, 12)	(8, 11)
3	(11, 12, 13)	(2, 8, 12)	(8, 10, 11)
4	(12, 13, 15, 16)	(4, 8, 9, 12)	(5, 6, 8, 16)
5	(6, 10, 11, 12, 16)	(4, 7, 8, 9, 12)	(5, 6, 8, 11, 15)
6	(7, 10, 11, 12, 13, 15)	(6, 7, 8, 10, 11, 12)	(3, 4, 7, 8, 13, 16)
7	(4, 8, 9, 11, 12, 13, 15)	(4, 5, 8, 9, 10, 11, 12)	(3, 4, 7, 8, 13, 14, 16)

	LDA	QDA	NB
2	(12, 15)	(12, 15)	(11, 12)
3	(11, 12, 15)	(11, 12, 15)	(11, 12, 13)
4	(11, 12, 13, 15)	(12, 13, 15, 16)	(11, 12, 13, 14)
5	(4, 6, 12, 13, 15)	(8, 9, 11, 13, 15)	(2, 10, 11, 12, 13)
6	(8, 11, 12, 13, 15, 16)	(2, 3, 7, 8, 11, 15)	(2, 10, 11, 12, 13, 16)
7	(2, 8, 11, 12, 13, 15, 16)	(4, 5, 8, 9, 11, 13, 15)	(7, 8, 9, 11, 12, 13, 14)

	LR
2	(12, 15)
3	(4, 12, 15)
4	(11, 12, 13, 15)
5	(7, 11, 12, 13, 15)
6	(5, 7, 11, 12, 13, 15)
7	(4, 6, 7, 11, 12, 13, 15)

Figure 13: Best Tuples (alpha)



Symptom	Best Tuple	Algorithm	prod	sens	spec	perf
Asymptomatic						
Occipital neck headace	(8, 10, 11, 15)	DT	0.47	0.62	0.76	0.69
Bifrontal headace						
Holocranial headace						
Hemicranial headace						
Any Headache	(4, 6, 13, 15, 16)	LDA	0.49	0.93	0.53	0.79
Hydrocephalus	(6, 7, 8, 10, 11, 12, 13)	DT	0.51	0.66	0.78	0.72
Visual Alterations	(3, 5, 8, 11, 12, 13, 14)	DT	0.44	0.50	0.89	0.85
Syringomyelia	(2, 12)	DT	0.51	0.74	0.69	0.72
Fatigue						
Instability	(12, 14, 15)	DT	0.52	0.64	0.82	0.76
Sensory Loss	(4, 7, 9, 11, 12, 13, 15)	DT	0.51	0.65	0.78	0.72
Anxiety	(2, 5, 9, 12)	DT	0.57	0.68	0.83	0.79
Depression	(2, 5, 6, 8, 9, 15, 16)	DT	0.62	0.68	0.91	0.88
Motor Weakness	(4, 6, 7, 8, 11)	DT	0.54	0.65	0.84	0.77
Dysphagia	(4, 5, 6, 10, 11)	DT	0.56	0.63	0.89	0.84
Dysphonia						
Gait Disturbances	(4, 5, 10, 13, 15)	DT	0.64	0.73	0.87	0.84
Paresthesia Upper Limbs	(2, 8, 10, 13, 15, 16)	DT	0.49	0.74	0.66	0.64
Paresthesia Lower Limbs	(3, 7, 9, 13, 14, 16)	DT	0.63	0.76	0.83	0.78
Upper Limbs Pain	(4, 7, 8, 11, 12, 13)	DT	0.69	0.77	0.91	0.88
Lower Limbs Pain						
Neck Pain	(2, 3, 6, 12, 13, 14)	QDA	0.45	0.51	0.88	0.74
Difficulty Swallowing	(4, 8, 11, 16)	DT	0.66	0.72	0.92	0.89
Nystagmus	(2, 4, 6, 10, 11)	QDA	0.58	0.58	1.00	0.92
Kyphoscoliosis	(6, 9, 11, 12, 13, 14)	DT	0.67	0.80	0.84	0.83
Complex Craniocervical Malformation						
Retrocurved Odontoid	(2, 3, 4, 11, 12, 15)	DT	0.63	0.68	0.92	0.87
Other CVJ Malformations	(4, 6, 14, 16)	DT	0.56	0.67	0.85	0.82
Cough Headaches	(2, 5, 8, 9, 10, 12, 15)	QDA	0.52	0.73	0.71	0.71
Dizziness	(4, 6, 12, 14)	DT	0.49	0.62	0.80	0.74
Vertigo						

Figure 14: Classification results for symptoms based on the product of sensitivity and specificity

References

- [1] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [2] Aintzane Urbizu. Mri-based morphometric analysis of posterior cranial fossa in the diagnosis of chiari malformation type i. *American Society of Neuroimaging*, 2013.