Using a Decision Tree Classifier with PCA Dimension Reduction as Adjunct to the Differential Diagnosis of GABHS

Thomas H Ehret

Department of Computer Science, New Mexico State University, tehret@nmsu.edu

Student Health and Well Being, California Polytechnic Humboldt, thomas.ehret@humboldt.edu

Applied ML Project Phase 4

ABSTRACT

Sore throat is one of the most common reasons for visits to primary care providers, with severe cases routinely presenting to Emergency Departments and Urgent Care centers. During the COVID19 pandemic clinicians were faced with logistical shortfalls of laboratory tests for common Upper Respiratory Illness (URI), Influenza Like Illness (ILI), and Group A beta-hemolytic Streptococcus (GABHS). A simple binary classification ML system can be used to assist differential diagnosis particularly when laboratory assessment is unavailable.

CCS CONCEPTS • Health Informatics• Health care information systems • Machine Learning approaches -> Classification

INTRODUCTION

Motivation: As of 2022 CDC data reports that approximately 12 million ambulatory care visits per year are due to complications from acute pharyngitis. These visits typically warrant a rapid strep test (RST); a rapid antigen detection test (RADT) that is widely used in clinics to assist in the diagnosis of bacterial pharyngitis caused by Group A beta-hemolytic Streptococcus, colloquially termed strep throat. This is often used as an alternative to the laboratory bacterial culture method – which can be cost prohibitive and has a 48-hour timeframe to produce a definitive culture. For much of 20201-2022, GABHS rapid tests were not available or had manufacturing deficiencies that placed further testing replication burdens on already overstretched clinical staff. As a "lessons learned" and proof of concept approach the team will develop a binary classification system that can be used to assist differential diagnosis in the event a positive GABHS laboratory diagnosis can't be made (i.e. no testing supplies). Furthermore, while telehealth medicine continues to revolutionize patient care, diagnostic tests for strep throat remain unavailable in the home setting. Therefore, there is a strong need to diagnose strep throat remotely or via telehealth consultations. This test would expand patient access to treatment and improve patient outcomes by decreasing the complications that inevitably arise from untreated strep throat.

Problem Definition

Given the signs and symptoms overlap of GABHS with other viral infectious agents, making a binary classification based solely on patient clinically labeled data will prove difficult. Therefore, it is imperative that the training data weighting be sufficiently vetted by clinicians as well as computer scientists. To that end I have interviewed medical colleagues to generate a set of clinical decision rules for the classification model to learn. Furthermore, no single element of a patient's history or physical examination can reliably confirm or exclude GABHS.

Solution

Training a binary classifier with modified class likelihood ratios on an anonymized dataset with the aim to reliably diagnose GABHS. The dataset dimensionality is 5 features with a positive or negative classification with likelihood ratio (table 2). The total number of dataset instances (95 as of this document) is not yet complete as more instances continue to be added. This is because I must wait for the electronic medical records database administrator to provide data access and server time. The goal is to have a dataset comprising the total number of cases from Cal Poly Humboldt student health center for the academic year 2021-2022; wherein the differential diagnosis is coded for GABHS and/or Pharyngitis.

During initial testing with a logistic regression model with an 80/20 split of 95 instances using a multiclass of multinomial the mean learning rate was 0.53 with 9 misclassifications. The confusion matrix is shown below in figure 1.

Figure 1

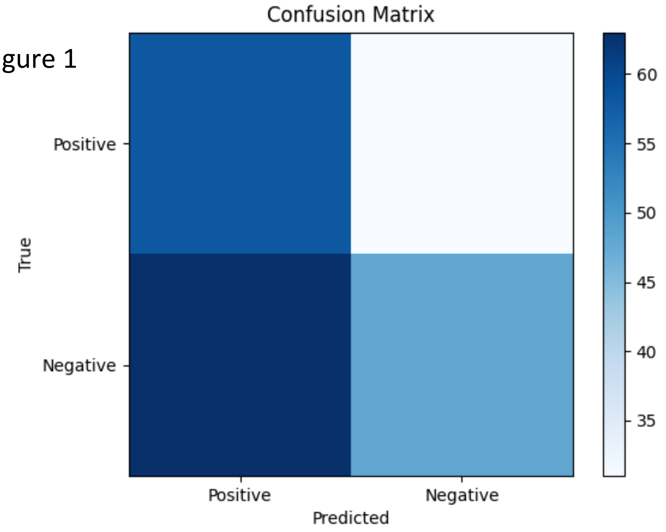
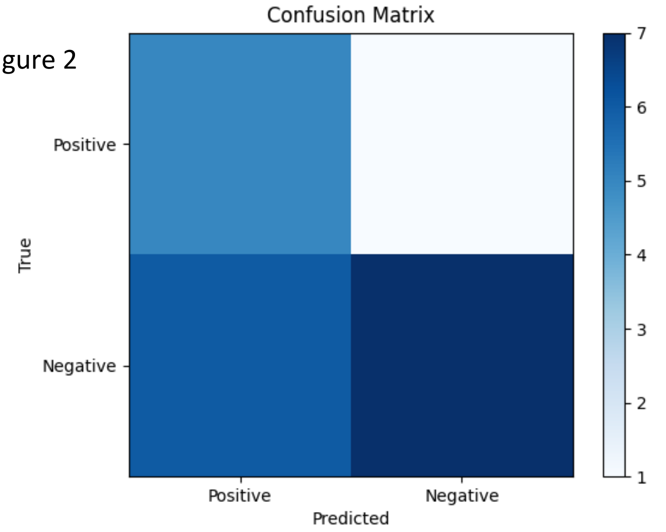


Figure 2



By changing the multiclass strategy to one-vs-rest (OVR) with the same testing / training split the results improved slightly: 0.63 learning rate with 7 misclassifications figure 2. As evidenced by the confusion matrix, the model is consistently erring on the side of False Positives. Given the results of initial testing there is still much work to be done with model fitting and weighting of the feature set.

In phase 4 of the project the classification model was switched to a Decision Tree. Initial testing shows that the tree model performs slightly better with default settings on the GABHS dataset. An interesting side effect of using this model is that the recursive splitting of the data by the decision tree is abstractly similar to how a clinician arrives at a preliminary diagnosis. This could be the unintended result of the diagnosis method and the collected data being unwittingly joined in a feedback loop. Wherein the phenomena observed is influenced by what datapoints clinicians think are important, and the datapoints themselves initially influence what clinicians observe. Furthermore, due to the change of model the threshold function was removed, and a tree structured classification ruleset was used to explore the decision tree heuristics.

Given the tight correlation of *Fever >= 100.9F* and *Headache*, further improvements in model accuracy were achieved by reducing the dimensions of the dataset with PCA -- with the *Fever* instances being retained and *Headache* discarded in the reduced dimension dataset -- similar to the clinician's diagnosis by exclusion. Similar *True Negative* predictions are retained, with a decrease in both *False Positive* and *True Positive* predictions (figure 4). In addition, there is an increase to 74% accuracy with only 5 misclassifications. A decision tree plot (figure 5) is included with this document.

The following parameters were used during the test runs:

PCA(n_components=4), DecisionTreeClassifier(criterion='gini',, splitter = "best", random_state=1)

Figure 4

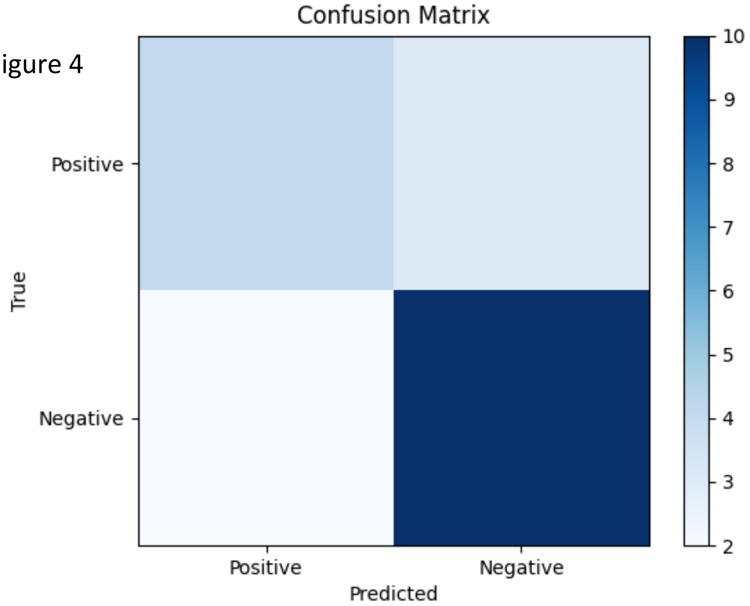
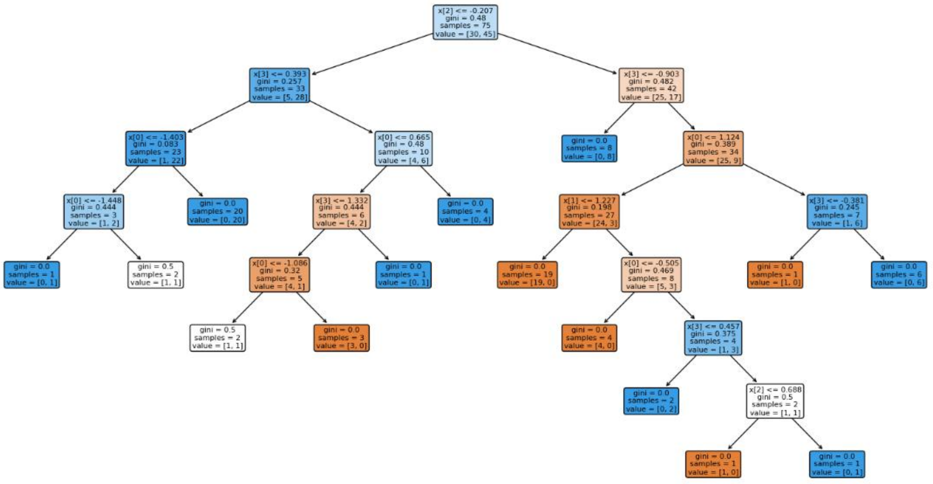


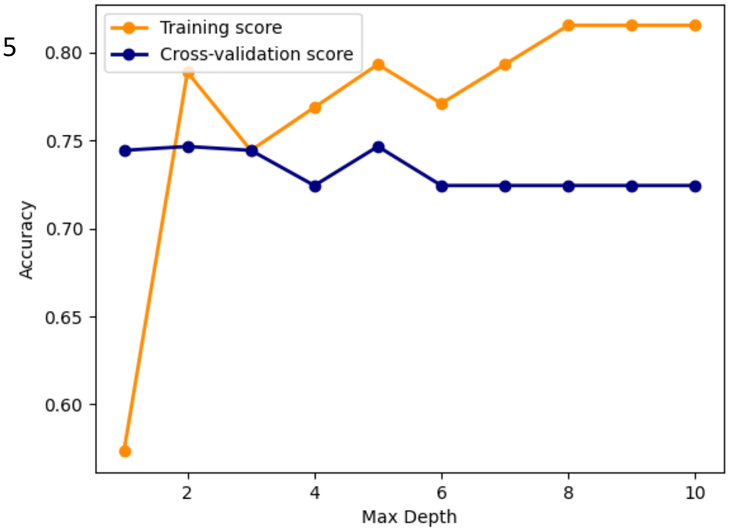
Figure 5


x[2] <= -0.207 gini = 0.48 samples = 75 value = [30, 45]
x[3] <= 0.393 gini = 0.257 samples = 33 value = [5, 28]
x[3] <= -0.903 gini = 0.482 samples = 42 value = [25, 17]
x[0] <= -1.403 gini = 0.083 samples = 23 value = [1, 22]
x[0] <= 0.665 gini = 0.48 samples = 10 value = [4, 6]
gini = 0.0 samples = 8 value = [0, 8]
x[0] <= 1.124 gini = 0.389 samples = 34 value = [25, 9]
x[0] <= -1.448 gini = 0.444 samples = 3 value = [1, 2]
gini = 0.0 samples = 20 value = [0, 20]
x[3] <= 1.332 gini = 0.444 samples = 6 value = [4, 2]
gini = 0.0 samples = 4 value = [0, 4]
x[1] <= 1.227 gini = 0.198 samples = 27 value = [24, 3]
x[3] <= -0.381 gini = 0.245 samples = 7 value = [1, 6]
gini = 0.0 samples = 1 value = [0, 1]
gini = 0.5 samples = 2 value = [1, 1]
x[0] <= -1.086 gini = 0.32 samples = 5 value = [4, 1]
gini = 0.0 samples = 1 value = [0, 1]
gini = 0.0 samples = 19 value = [19, 0]
x[0] <= -0.505 gini = 0.469 samples = 8 value = [5, 3]
gini = 0.0 samples = 1 value = [1, 0]
gini = 0.0 samples = 6 value = [0, 6]
gini = 0.5 samples = 2 value = [1, 1]
gini = 0.0 samples = 3 value = [3, 0]
gini = 0.0 samples = 4 value = [4, 0]
x[3] <= 0.457 gini = 0.375 samples = 4 value = [1, 3]
gini = 0.0 samples = 2 value = [0, 2]
x[2] <= 0.688 gini = 0.5 samples = 2 value = [1, 1]
gini = 0.0 samples = 1 value = [1, 0]
gini = 0.0 samples = 1 value = [0, 1]

The number of components used by PCA had diminishing returns with settings *n_components* >= 4. With the average over 20 test runs being a classification accuracy of 71% with 4-5 misclassifications. This contrasts with 3 PCA components which produced an average accuracy of 68% and >=5 misclassifications. Thus, a dimension reduction of at least 1 but not more than 2 features delivered statistically similar results. This might be mitigated in the future with the addition of additional features such as : recent known exposure to GABHS (ICD-10 Z20.818), and streptococcal pharyngitis breath – (a heuristic used by clinicians to denote bacterial infection without laboratory findings). However, those features are not included in the dataset at this time.

There is also the concern that the decision tree model may be overfitting to this dataset. Initial testing of validation curve is shown in figure 6.

Figure 5



Given how the model is performing I am beginning to suspect that the dataset might have validation issues. I have requested verification that the ICD coding of the symptoms is correct, and that they are appropriately mapped to corresponding features in the dataset.

The weighting of features will be based upon the specifics listed in table 1. This is the common clinical decision-making intervals as suggested by the American Academy of Family Physicians.

Table 1: History and Physical Examination Findings Suggesting GABHS Pharyngitis

Factor	Sensitivity (%)	Specificity (%)	Positive Ratio	Negative Ratio
Absence of cough	51 to 79	36 to 68	1.1 to 1.7	0.53 to 0.89
Anterior cervical node swelling	55 to 82	34 to 73	0.47 to 2.9	0.58 to 0.92
Headache	48	50 to 80	0.81 to 2.6	0.55 to 1.1
Myalgia	49	60	1.2	0.84
Palatine petechiae	7	95	1.4	0.98
Pharyngeal exudates	26	88	2	0.85
Streptococcal exposure	19	91	2	0.9
Temperature ≥ 100.9° F (38.3° C)	22 to 58	53 to 92	0.68 to 3.9	0.54 to 1.3
Tonsillar exudates	36	75	1.8	0.74

Adapted from Ebell MH, Smith MA, Barry HC, Ives K, Carey M. The rational clinical examination JAMA. 2000;284(22):2915.

However, there was a need for normalizing the feature set and reducing factors to align with a *Modified Centor* score. A cumulative score greater or equal to 4 is the threshold where the risk of GABHS is considerd to be 51 to 53% and is therefore warranted for empiric treatment with antibiotic therapy.

Based upon the aformentioned critieris, the current decision tree model performs within the range of recommended empriical treatment. It is possible that the model is overfitting;.therefore further testing and dataset expansion is needed.

References

[1] Centor, R. M., Witherspoon, J. M., Dalton, H. R., Brody, C. J., & Link, K. (1981). The Diagnosis of Strep Throat in Adults in the Emergency Room. *Medical Decision Making*, *1*(3), 239–246. https://doi.org/10.1177/0272989x8100100304

[2] Hing E, Cherry DK, Woodwell DA. National Ambulatory Medical Care Survey: 2003 Summary. Adv Data. 2005;365:1-48

[3] McIsaac WJ, White D, Tannenbaum D, Low DE. A clinical score to reduce unnecessary antibiotic use. CMAJ. 1998;158(1):79.