Using Machine Learning on ICD-10 Data to Enhance an Expert Anaphylaxis Case Definition

Kamil Can Kural, Ilya Mazo, Mark Walderhaug, Lei Huang, Luis Santana-Quintero and Ravi Goud (CBER Office of Biostatistics and Epidemiology, CBER HIVE)

Identification of treatment related adverse events (AEs) is critical to FDA’s surveillance activities.

Case definitions are necessary to identify specific episodes that are potentially AEs from large collections of health care data.

Problem: Case definitions are time consuming and difficult to construct from administrative data.

Objective: Use machine learning to enhance a human constructed case definition for detection of true cases of incident anaphylaxis in the CMS database.

We suggest a novel way of using machine learning to help create case definitions and we apply it to identify cases of anaphylaxis, a serious, life-threatening allergic reaction.
Using Machine Learning on ICD-10 Data to Enhance an Expert Anaphylaxis Case Definition

Kamil Can Kural, Ilya Mazo, Mark Walderhaug, Lei Huang, Luis Santana-Quintero and Ravi Goud (CBER Office of Biostatistics and Epidemiology, CBER HIVE)

Original Workflow Without Chart Confirmed Results

Minimally curate 2500 claims into groups of 500 likely anaphylaxis, possible anaphylaxis, allergy, and 1000 controls using a subset of ICD-10 codes

Apply supervised and unsupervised machine learning models to the 2500 cases

Identify new important codes and construct models to improve the classification of anaphylaxis cases

Test machine learning models on chart reviewed data

Data Assumptions

• Anaphylaxis and allergy are similar.
• A classifier that discriminates anaphylaxis from allergy will also discriminate it from other acute health problems.
• Anaphylaxis is rare enough that the random control cases will not contain anaphylaxis episodes.

Methods Used

• Linear discriminant analysis (HIVE-RLDA). (Results not shown)
• T-distributed stochastic nearest neighbor embedding. (t-SNE)
• Decision Tree
• Random Forest
Using Machine Learning on ICD-10 Data to Enhance an Expert Anaphylaxis Case Definition

Kamil Can Kural, Ilya Mazo, Mark Walderhaug, Lei Huang, Luis Santana-Quintero and Ravi Goud (CBER Office of Biostatistics and Epidemiology, CBER HIVE)

Latest Workflow with Chart Confirmed Results

<table>
<thead>
<tr>
<th>Cohort 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consists of 2600 samples</td>
</tr>
<tr>
<td>Minimally curated CMS claims dataset</td>
</tr>
<tr>
<td>High, intermediate, and very low likelihood of Anaphylaxis, plus a random background claims dataset</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cohort 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consists of 530 samples</td>
</tr>
<tr>
<td>Specifically created to learn and extract information about vaccine-induced Anaphylaxis</td>
</tr>
<tr>
<td>Claims were identified through an Anaphylaxis Algorithm created by experts using patterns in claims data</td>
</tr>
</tbody>
</table>

Feature Selection

- Logistic Regression with High Regularization
- Chi Square Analysis
- Random Forest
- Light Gradient Boosting
- Logistic Regression with Low Regularization

Methods Used

- Random Forests
- Sammon Mapping
- T-distributed stochastic nearest neighbor embedding (t-SNE)
- Light Gradient Boosting
- Logistic Regression

Feature Selection

- Use statistical methods as well as machine learning models to identify most salient features between datasets.
- The features are color coded to check robustness between datasets.
- An ideal feature would pass at least 1 test in each dataset.

Table 1: Coloring scheme of features. Green represents features important for all three cohorts. Blue color represents features important for two out of three cohorts. Red represents features important only for that specific cohort.
Results: Feature Selection

90 Features were selected for classification task

<table>
<thead>
<tr>
<th>Tests Passed</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Chart Confirmed Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>38</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>33</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>118</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>82</td>
<td>60</td>
<td>93</td>
</tr>
<tr>
<td>1</td>
<td>127</td>
<td>80</td>
<td>91</td>
</tr>
<tr>
<td>0</td>
<td>3876</td>
<td>918</td>
<td>901</td>
</tr>
<tr>
<td>Total Features</td>
<td>4228</td>
<td>1197</td>
<td>1197</td>
</tr>
</tbody>
</table>

Table 2: Number of features that satisfy each threshold after running multiple feature selection algorithms. For example, there are 38 features that satisfied all 5 feature selection algorithms for Cohort 1.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Chart Confirmed Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.982120 +/- 0.010557</td>
<td>0.883137 +/- 0.067506</td>
<td>0.804325 +/- 0.080723</td>
</tr>
<tr>
<td>4</td>
<td>0.984580 +/- 0.007396</td>
<td>0.875336 +/- 0.071413</td>
<td>0.844127 +/- 0.085665</td>
</tr>
<tr>
<td>3</td>
<td>0.983860 +/- 0.007900</td>
<td>0.872437 +/- 0.071407</td>
<td>0.812116 +/- 0.090686</td>
</tr>
<tr>
<td>2</td>
<td>0.982180 +/- 0.008964</td>
<td>0.872619 +/- 0.075855</td>
<td>0.810952 +/- 0.103367</td>
</tr>
<tr>
<td>1</td>
<td>0.983120 +/- 0.008932</td>
<td>0.874118 +/- 0.075392</td>
<td>0.801905 +/- 0.088265</td>
</tr>
<tr>
<td>0</td>
<td>0.983740 +/- 0.008431</td>
<td>0.874118 +/- 0.075392</td>
<td>0.800053 +/- 0.081258</td>
</tr>
</tbody>
</table>

Table 3: Success metrics of each trained model in AUC for each individual dataset with different thresholds.

- Fit individual models for each dataset to see how features that satisfied multiple criteria contribute to the models’ success.
- Some features were selected even though they did not pass at least 1 test for each Cohort.
- Create the final models using features without expert curated codes. (Remove codes while finding important features)
Using Machine Learning on ICD-10 Data to Enhance an Expert Anaphylaxis Case Definition
Kamil Can Kural, Ilya Mazo, Mark Walderhaug, Lei Huang, Luis Santana-Quintero and Ravi Goud (CBER Office of Biostatistics and Epidemiology, CBER HIVE)

Results: Unsupervised Machine Learning Models (T-SNE)

• T-SNE analysis overall separates Anaphylaxis episodes (Orange) from Allergy. (blue)

• Further, it shows we might have different influencers for each different dataset based on 3 different clusters.

• 90 features are enough for clear separation.

• Classifications likely had inaccuracies, thus need for improved case definitions.

• After Chart Confirmation, there were multiple changes in Cohort 2 data. (Allergy Cases went up by 52 samples)

• Even the expert definition has less than desired success in identifying Anaphylaxis cases correctly.

Points are colored by group, anaphylaxis vs allergy

Anaphylaxis – 1
Allergy - 0
Results: Unsupervised Machine Learning Models: Sammon Mapping

- Sammon Mapping performs much better in Classifying CMS Data.
- Again, we can see some of the labels which might be misclassified with expert definitions. (Cohort 2)
- Emphasizes our earlier point which states that the expert definition is not very successful for identifying Anaphylaxis cases.
Results: Supervised Machine Learning Models: Light Gradient Boosting (Easy)

- Supervised classification with Allergy and Anaphylaxis.
- 80% in training set, 20% in testing set stratified and shuffled randomly.
- A separate portion 20% of the data was used for Hyperparameter tuning.
- Construct ‘Easy’ decision trees by including codes used to create the data in the lists.
- Train decision trees until one with high discrimination is found.
- Extract classifiers and construct final tree.

Machine learning finds the codes experts used to construct the allergy and anaphylaxis groups in the data set. This is expected, since the outcomes and codes the experts used are highly correlated by design of the dataset.

Epinephrine, Allergy Diagnosis, Anaphylaxis Diagnosis, Urticaria are key features.
Results: Supervised Machine Learning Models: Light Gradient Boosting (Hard)

- **Construct ‘Hard’ decision trees:** remove codes used to manually select the cases and look for other factors embedded in the data.

- This is necessary to eliminate very strong and highly correlated signals.

- Instead of seeking codes which define cases directly, identifying the patterns this way would be much more likely.

- **Validation:** compare to the codes used in manual case definition built independently by group of experts.

Acute Kidney Failure, Acute Respiratory Failure, Injection Codes and Emergency, Saline administration and Department visit codes were very significant.

Models identified the importance of billing for injections. Treatment setting is important in identifying Anaphylaxis.
Results: Light Gradient Boosting Model Performance

- **Model accuracy** is 96.7% with considerable information removed.
- “Easy” model setting has around 1200 features while hard setting has only 90 features.
- Information that seemed nonspecific can be used by machine learning models to identify Anaphylaxis cases from ICD-10 codes.

### Confusion Matrices

<table>
<thead>
<tr>
<th></th>
<th>Predicted as Allergy</th>
<th>Predicted as Anaphylaxis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Easy</strong></td>
<td>122</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>180</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Predicted as Allergy</th>
<th>Predicted as Anaphylaxis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hard</strong></td>
<td>119</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>177</td>
</tr>
</tbody>
</table>

- Confusion matrices for both settings points at a lower model performance for hard model setting.
- Lower model performance for hard model is expected, since the data used ‘easy’ model codes for determining Anaphylaxis and Allergy cases. As a result, some of the cases are misclassified.
- Our aim is to identify a pattern instead of trying to associate a specific code to Anaphylaxis, which might mislead researchers with or without its absence.
Conclusions

This work demonstrates that the combination of machine learning and a minimally curated claims data set could help to quickly identify traits associated with cases of interest, thereby potentially improving the efficiency and accuracy of case definition construction and foster public health innovation.

Validation:
ML methods on the minimally curated dataset succeeded in independently identifying treatment codes (e.g. diphenhydramine, methylprednisolone as well as injection codes) that were used in the manual case definition built by experts.

Python code developed in this project can be made available HIVE for CBER scientists interested in using machine learning.
Acknowledgements

We would like to thank Acumen and CMS for providing the data.

“This presentation is an informal communication and represents the best judgment of authors. These comments do not bind or obligate FDA.”