Prediction Analytics for COPD and Sepsis Diagnosis using data analysis and machine learning

Manuchehr Aminian, Rituparna Basak, Elita Astrid Lobo, Jenna McDanold, Richard Moore, Ruqi Pei, Geneva Porter, Kosuke Sugita, Soheil Saghafi

New Jersey Institute of Technology
June 21, 2019
The Iterex healthcare app aims to make chronic disease management more accessible.

X-ray of COPD patient with emphysema\(^1\) (left), and sepsis blood sample photograph\(^2\) (right). The Iterex healthcare app aims to make chronic disease management more accessible.

\(^1\)Image taken from Cleveland Health Clinic
\(^2\)Image taken from Science Source
INTRODUCTION

Determine symptoms and disease variables → Generate clinical patient case scenarios → Collect and analyze triage data → Train machine learning models → Validate machine learning models

Iterex trials were shown to:
- Outperform Specialists
- Err in Favor of patient safety
- Help increase medication compliance
Machine Learning Methodology

- **Precision Score:** What proportion of positive identifications was actually correct?
- **Recall Score:** What proportion of actual positives was identified correctly?

**Figure 1:** Confusion Matrix

*aImage taken from Walber*
Machine Learning Methodology

Figure 2: Receiver operating characteristic (ROC) curve

3Image Taken from Sharpr
COPD Analysis and Results
**Question:** What set of patient signs, symptoms, and baseline health factors are indicative of a physician identified exacerbation for COPD patients?

We considered **over 30 health factors**, such as:

- General Stats like sex, age, weight
- Vitals like heart/respiratory rate and temperature
- Respiratory evaluations like FEV, inhaler use, or peak flow
- Medication compliance and symptom changes
The heat map describes correlations among all the features for COPD.

This shows there are **no clear correlation** observed among the features for predicting the COPD exacerbation result.

**Figure 3:** Features comparison for COPD data points
COPD Correlation and Relative Importance

### Feature and Rank

<table>
<thead>
<tr>
<th>Feature</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom 3</td>
<td>0.231</td>
</tr>
<tr>
<td>Symptom 2</td>
<td>0.182</td>
</tr>
<tr>
<td>Symptoms worse</td>
<td>0.172</td>
</tr>
<tr>
<td>Symptom 1</td>
<td>0.158</td>
</tr>
<tr>
<td>Symptom 6</td>
<td>0.150</td>
</tr>
<tr>
<td>FEV1 post-inhaler</td>
<td>0.107</td>
</tr>
</tbody>
</table>

**Table 1:** Top 6 features and their importance ranking

**Figure 4:** Bar graph with error for COPD features
COPD EXACERBATION CLASSIFICATION

We predicted exacerbation of COPD using the 6 most important features in order to avoid noise created by other features.

Figure 5: Optimal AUC Accuracy: 69.5%
SEPSIS ANALYSIS AND RESULTS
Question: Can we predict the onset of a septic infection using temporal sign and symptom data?

We considered over 40 data measurements such as:

- Vitals like heart rate, blood pressure, respiratory rate, and temperature
- Nutrient levels like calcium, potassium, and glucose
- Blood measurements like white blood cell and platelet counts, and hemoglobin level
- General stats like age, sex, and length of stay within the ICU
The heat map on the left shows that there are **no clear correlations** observed among the features sepsis prediction result.

**Figure 6:** Features comparison for sepsis data points
The clinical definition of SIRS (possibly indicating sepsis) is distinguished by two or more of the following:

- Heart rate > 90/min
- Temp ≥ 38 or < 36°C
- Respiratory rate > 20/min
- White blood cell count > 12 or < 4 cells/mL

This definition gives a 65% false positive rate in our data (2 of 3 healthy patients falsely diagnosed with sepsis!)
This algorithm is able to make predictions depending on the **current label of sepsis**.

Scores from the classification matrix:

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.86</td>
<td>0.85</td>
<td>0.86</td>
</tr>
<tr>
<td>1</td>
<td>0.85</td>
<td>0.86</td>
<td>0.85</td>
</tr>
</tbody>
</table>

**Figure 7:** ROC Curve (area under the curve: **0.91**
Using both **current and past** labels of sepsis, we applied **moving window** algorithm on this time series problem.

We use random forest classifier and sepsis label for prediction confusion matrix. (prediction row, true column)

\[
\begin{pmatrix}
\text{Predicted}/\text{True} & P & N \\
P & 2211 & 14 \\
N & 950 & 10628
\end{pmatrix}
\]

We notice the false positive cases and false negative cases are very small numbers, especially for false positive. We believe this is a good classifier.
All features in the data frame are used for Sepsis prediction:

**Figure 8:** Prediction Accuracy with all features: 0.93
Depending on past and future labels, we can predict the time until a patient get sepsis.

Root Mean squared error: 1.2 hours.

To validate our results, we excludes data points from the training set and that increases the mean squared error.
**Summary and Future Work**

What have we done?

1. Predicted exacerbations in COPD patients with an accuracy of roughly 70%
2. We can identify a collection of vitals as septic or not with an AUC of 0.91
3. We can predict *time until sepsis* in with a RMSE of 1.2 hours (!)

What do we need to do?

- For the regression model, Excluding data points from training sets increase the prediction of time to get sepsis, hence we need to find the optimal time / method to fix the problem.


Schembri, Stuart and Anderson, William and Morant, Steve and Winter, Janet and Thompson, Philip and Pettitt, Daniel and MacDonald, Thomas M. and Winter, and John H, A predictive model of hospitalisation and death from chronic obstructive pulmonary disease, Respiratory Medicine, vol. 103, pp.1461–1467, W.B. Saunders, oct 2009,


Marin, Jose M. and Carrizo, Santiago J. and Casanova, Ciro and Martinez-Cambor, Pablo and Soriano, Joan B. and Agusti, Alvar G.N. and Celli, and Bartolome R, Prediction of risk of COPD exacerbations by the BODE index, Respiratory Medicine, vol. 103, pp.373–378, W.B. Saunders, mar 2009

Giuliano, Karen K, Physiological monitoring for critically ill patients: testing a predictive model for the early detection of sepsis, American Journal of Critical Care, 16, pp.122–130, AACN, 2007,
Questions?