Prediction Analytics for COPD and Sepsis Diagnosis

using data analysis and machine learning

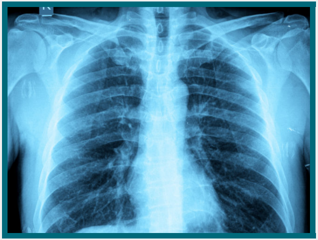
Manuchehr Aminian, Rituparna Basak, Elita Astrid Lobo, Jenna McDanold, Richard Moore, Ruqi Pei, Geneva Porter, Kosuke Sugita, Soheil Saghafi

New Jersey Institute of Technology

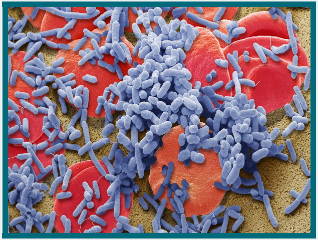
June 21, 2019

Introduction

COPD



Sepsis



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**?

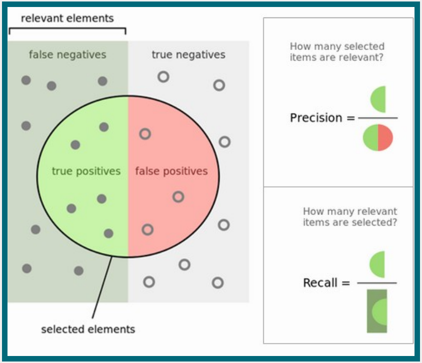


Figure 1: Confusion Matrix[a]

[a] Image taken from *Walber*

Machine Learning Methodology

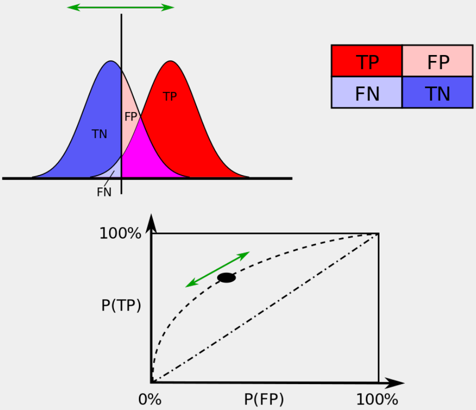


Figure 2: Receiver operating characteristic (ROC) curve[3]

[3] Image Taken from *Sharpr*

COPD Analysis and Results

Patient Symptoms in COPD

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

COPD Visualization and Trend Identification

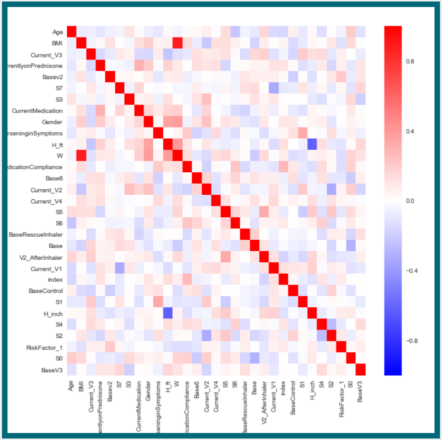


Figure 3: Features comparison for COPD data points

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.

COPD Correlation and Relative Importance

Feature	Rank
Symptom 3	0.231
Symptom 2	0.182
Symptoms worse	0.172
Symptom 1	0.158
Symptom 6	0.150
FEV1 post-inhaler	0.107

Table 1: Top 6 features and their importance ranking

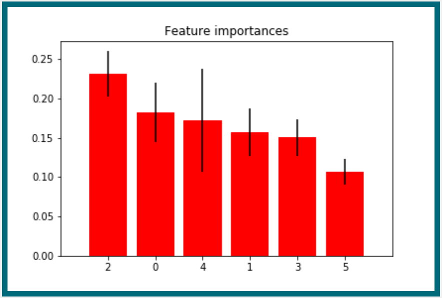


Figure 4: Bar graph with error for COPD features

COPD Exacerbation Classification

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

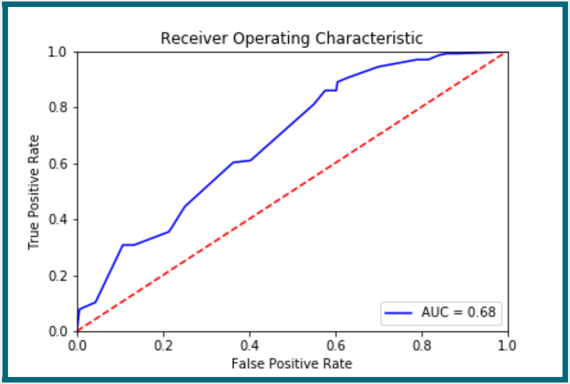


Figure 5: Optimal AUC Accuracy: 69.5̄%

Sepsis Analysis and Results

Predicting the Onset of Sepsis

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

Sepsis Visualization and Trend Identification

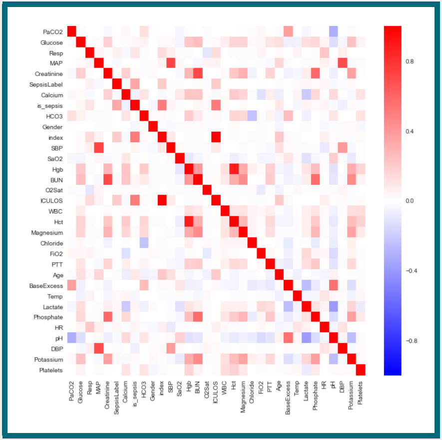


Figure 6: Features comparison for sepsis data points

The heat map on the left shows that there are **no clear correlations** observed among the features sepsis prediction result.

Sepsis Misdiagnosis

The clinical definition of SIRS (possibly indicating sepsis) is distinguished by two or more of the following:

- Heart rate > 90/min
- Temp ≥ 38 or $< 36^o$ Celsius
- 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!)

Sepsis Prediction Using Post-Septic Features

This algorithm is able to make predictions depending on the **current label of sepsis**.

Scores from the classification matrix:

	Precision	recall	F1
0	.86	.85	.86
1	.85	.86	.85

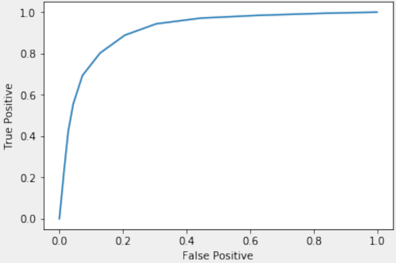


Figure 7: ROC Curve (area under the curve: **0.91**)

Sepsis Prediction Using Pre/Post-Septic Features

- 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)

$$\begin{pmatrix} \textit{Predicted/True} & P & N \\ P & \mathbf{2211} & 14 \\ N & 950 & \mathbf{10628} \end{pmatrix}$$

- 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.

Sepsis Prediction Using Pre/Post-Septic Features

All features in the data frame are used for Sepsis prediction:

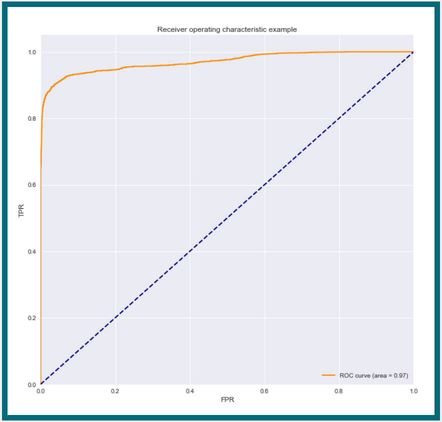
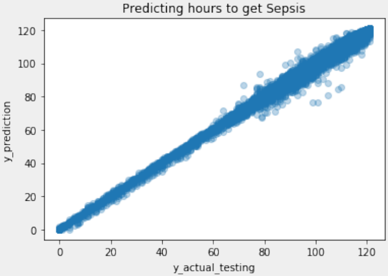


Figure 8: Prediction Accuracy with all features: **0.93**

Sepsis Prediction Using Pre/Post-Septic Features

- 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.

References

- Ko, Fanny W. and Chan, Ka Pang and Hui, David S. and Goddard, John R. and Shaw, Janet G. and Reid, David W. and Yang, and Ian A., Acute exacerbation of COPD, Respirology, vol. 21, pp.1152–1165, John Wiley & Sons, Ltd (10.1111), oct 2016.
- MacDonald, Martin and Korman, Tony and King, Paul and Hamza, Kais and Bardin, Philip, Exacerbation phenotyping in chronic obstructive pulmonary disease, Respirology, vol. 18, pp.1280–1281, John Wiley & Sons, Ltd (10.1111), nov, 2013.
- Donaldson, G C and Seemungal, T A R and Bhowmik, A and Wedzicha, J A, Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease, Thorax, vol. 57, pp.847–52, 2002.
- Quint, J K and Baghai-Ravary, R and Donaldson, G C and Wedzicha, J A, Relationship between depression and exacerbations in COPD, European Respiratory Journal, vol. 32, pp.53–60, 2008.
- 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,
- Make, Barry J and Eriksson, Göran and Calverley, Peter M and Jenkins, Christine R and Postma, Dirkje S and Peterson, Stefan and Östlund, Ollie and Anzueto, and Antonio. A score to predict short-term risk of COPD exacerbations (SCOPEX), International Journal of COPD, vol. 10, pp.201–209, Dove Press, 2015.
- Lode, H. Allewelt, M. Balk, S. De Roux, A. Mauch, H. Niederman, M. Schmidt-Ioanas, M, A prediction model for bacterial etiology in acute exacerbations of COPD, Infection, vol. 35, pp.143–149, 2007.
- Donaldson, G C and Wedzicha, J. A, COPD exacerbations · 1: Epidemiology, vol. 61, pp.164–168, 2006.
- Marin, Jose M. and Carrizo, Santiago J. and Casanova, Ciro and Martinez-Camblor, 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?