Predicting Respiratory Decompensation in Ventilated ICU Patients
Beini Hu1, Akanksha Girish1, Zina Kurian1, Yvette Tan1, Michael Young1, Dr. Joseph Greenstein1, Dr. Raimond Winslow1, Dr. Jules Bergmann2, Dr. Jim Fackler2
Johns Hopkins University | 1Whiting School of Engineering | 2Johns Hopkins Medicine, Anesthesia and Critical Care Medicine| Baltimore, MD

Annually, over 300,000 ICU patients receive life-saving mechanical ventilation. However, 5-10% of these patients develop a Ventilator-Associated Complication (VAC), which could increase mortality, length of ICU stay, and cost of treatment. A VAC is indicated by acute lung injury, infection, and pneumonia, as well as other vital signs as potential markers. With input from physiological time series and EHR data, we can predict a patient’s likelihood of developing a VAC. Physicians can then intervene in a critically early time window to mitigate the effects of the complication or prevent it.

**Objectives**
We aim to develop a statistical model that will predict each patient’s likelihood of developing a VAC. We aim to establish the concept for a real-time clinical model that will output an hourly risk score for VAC onset.

**Data Source**
We worked with the MIMIC III dataset, which includes about 46,000 patients in a Boston hospital from 2001-2012. From this dataset we took data only from adult patients who had been placed on a mechanical ventilator for at least 96 hours.

**Introduction**

Patients in the cohort were assigned a VAC or non-VAC label using the VAE framework (Figure 1). Then, we evaluated availability of data, identified errors and outliers, and managed gaps. We excluded patients who had over 50% of their data missing and used carry forward methods to impute outliers outside the physiologically valid ranges. This yielded 5740 subjects, of whom 1401 experienced a VAC. Categorical variables were one-hot encoded for input into our model.

To construct features for our model, we first selected expert-derived features and then used automated feature extraction packages (tsFresh) to generate additional statistical features as input to our model.

Different classifiers were constructed to predict VACs.

**Table 1: Patient Cohort**

<table>
<thead>
<tr>
<th>MIMIC-III</th>
<th>SUBJECTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique ICU admissions (n)</td>
<td>60840</td>
</tr>
<tr>
<td>Unique subjects (n)</td>
<td>46338</td>
</tr>
<tr>
<td>Age, years (mean, (s.d.))</td>
<td>74.93, (54.77)</td>
</tr>
<tr>
<td>Ventilated Adult patients (n)</td>
<td>21829</td>
</tr>
<tr>
<td>Ventilator Duration (mean days)</td>
<td>2.74</td>
</tr>
<tr>
<td>Male gender (n,(%))</td>
<td>26121, (56)</td>
</tr>
<tr>
<td>Hospital mortality (n,(%))</td>
<td>3130, (14)</td>
</tr>
<tr>
<td>Length of ICU stay (mean days)</td>
<td>4.18</td>
</tr>
</tbody>
</table>

**Data Preprocessing**

**Patient Cohort**

**Results**

Max. AUC = 0.719
Max. Accuracy = 0.67

Figure 2: ROC Curves using various classifier models. The random forest classifier gave the highest AUC at .719, with an accuracy of .67.

Figure 3: Ranked features by importance

**Conclusion**
We developed a statistical model that predicts the likelihood of a patient developing a VAC. To further improve the accuracy of this model, we can integrate high frequency physiological waveform data and consider data in different time windows of a patient's stay. The model can also be refined to output an hourly risk score for VAC prediction for clinical real-time prevention.

**References**