COVID-19 and the MSHS

THE MOUNT SINAI COVID INFORMATICS CENTER (MSCIC)
Girish N Nadkarni, MD, MPH
The Mount Sinai Health System (MSHS)
The Epicenter of the COVID-19 Pandemic in the United States

Total COVID-19 and PUI Hospitalized Cases by Day

- COVID-19 Positive
- PUI

Patients

0 500 1000 1500 2000 2500

BIG DATA IN A TIME OF BIG VIRUSES.

How the World is Using Data & Analytics to Fight COVID-19
Data contained in EHR

- Demographics
- Medical history
- Vital signs
- Diagnoses
- Medications
- Treatment plans
- Immunization history
- Radiology images
- Laboratory results

Wei-Qi, W. & Denny, J.C. Genome Medicine, 2015.
Mount Sinai COVID Informatics Center
Fighting Covid-19 with the Power of Data

Informatics Crisis Response Platform

**Critical Informatics Consultation Service**
provides MSHS clinicians and researchers easy-to-digest answers to pressing clinical questions within 24 hours

**Centralized Engineering Core**
A team of highly trained computer scientists, engineers, informaticists, and researchers who are dedicated to productizing this Informatics Crisis Response Platform to enable MSHS to be battle ready in this and in future crises

**Infrastructure** supported by Microsoft Azure cloud computing services, MSCIC has built and maintains a ground truth harmonized dataset that integrates data streams from MSHS clinical data (e.g. EHR, Imaging, Pathology) along with novel research data sets (e.g. –omics, digital health tracking, immune biomarkers)

**Rapid Clinical Intervention Toolkit**
facilitates the practice of evidence-based medicine in the MSHS by feeding insights from data science into the daily workflow via the electronic medical record
MSCIC Data Platform

Azure HIPAA subscription

All the VMs belong to same network

Each dataset is mounted individually in read-only on each VM. The Sys admin can decide which machine can have the full collection or a subset of it. The access also can be restrict a machine level with each user having access only to a specific area of filesystem.

Series of share disks in environment

Clarity
Registry
MSDW
IRW
Pathology
WebApp

Rsync from Daily snapshot to all the share drivers

Daily Snapshot of all datasets.

First QA/QC, de-identification phase. Prepare dataset to be shared with researchers VMs

VM ingestion machine

Unify datasets

Work in progress

Cosmos DB service

ETLs processes

OneDrive

MSDW
Pathology
Clarity
Oracle DB
IRW
Samba interface
Registry
Samba interface
Surveys webapp

semi-manual ingestion
script reads daily
rsync daily from the mounted disk
script reads daily
script reads daily

VM matrix
Each VM is based on a same image we created ad-hoc for the project
ASSOCIATION OF ANTICOAGULATION WITH MORTALITY

**A**

All Patients (N= 2773)

<table>
<thead>
<tr>
<th>In-hospital Anticoagulation</th>
<th>Number at Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>786 538 266 90 19 3</td>
</tr>
<tr>
<td>No</td>
<td>1987 977 296 71 13 1</td>
</tr>
</tbody>
</table>

**B**

Patients Requiring Mechanical Ventilation (N= 395)

<table>
<thead>
<tr>
<th>In-hospital Anticoagulation</th>
<th>Number at Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>234 197 137 65 14 3</td>
</tr>
<tr>
<td>No</td>
<td>161 100 54 25 7 1</td>
</tr>
</tbody>
</table>
Patient Inclusion Criteria

Patients from five MSHS hospitals: MSH, MSM, MSB, MSQ, MSW

- Inpatient?
  - Yes
  - No → Exclude

- Age ≥ 18?
  - Yes
  - No → Exclude

- Positive RT-PCR SARS-CoV-2 lab test?
  - Yes
  - No → Exclude

- COVID-19 lab order ≤ 48h after admit?
  - Yes
  - No → Exclude

- Intubation ≤ 48h from ICU admit?
  - Yes
  - No → Exclude

- Death time after admit time?
  - Yes
  - No → Exclude

- MSHS COVID-19 Cohort N = 3055

Time Horizons for Prediction

<table>
<thead>
<tr>
<th>Admission</th>
<th>t = 0</th>
<th>3 days</th>
<th>5 days</th>
<th>7 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>training data</td>
<td>!</td>
<td>!</td>
<td>!</td>
<td>!</td>
</tr>
</tbody>
</table>

Critical Event or Death

= Benjamin Glicksberg

Model Training and Evaluation

Features

- Demographics
- Labs
- Medical History
- Vitals

Location Split

- MSH: N = 1225
- MSM: N = 560
- MSB: N = 460
- MSQ: N = 459
- MSW: N = 351

Other Hospitals

Temporal Split

- Retrospective
- Prospective

May
Performance at MSH (train + CV)

Model Performance at Training

Mortality (ROC)

Mortality (PRC)

Critical Event (ROC)

Critical Event (PRC)

Sulaiman Somani, BS
Akhil Vaid, MD
External and temporal performance (validation)

XGBoost Performance on Validation Sets

Mortality (ROC)

Mortality (PRC)

Critical Event (ROC)

Critical Event (PRC)
What did the model learn?

Sulaiman Somani, BS
Allan Just, PhD
Interactions between features can further reveal what the model learned.
SARS-CoV-2 is Devastating to Numerous Organ Systems

1. Lungs
A cross-section shows immune cells crowding an inflamed alveolus, whose walls break down during attack by the virus, diminishing oxygen uptake. Patients cough, fever rises, and it takes more and more effort to breathe.

2. Liver
Up to half of hospitalized patients have enzyme levels that signal a struggling liver. An immune system in overdrive and drugs given to fight the virus may be causing the damage.

3. Kidneys
Kidney damage is common in severe cases and makes death more likely. The virus may attack the kidneys directly, or kidney failure may be part of whole-body events like plummeting blood pressure.

4. Intestines
Patient reports and biopsy data suggest the virus can infect the lower gastrointestinal tract, which is rich in ACE2 receptors. Some 20% or more of patients have diarrhea.

5. Brain
Some COVID-19 patients have strokes, seizures, mental confusion, and brain inflammation. Doctors are trying to understand which are directly caused by the virus.

6. Eyes
Conjunctivitis, inflammation of the membrane that lines the front of the eye and inner eyelid, is more common in the sickest patients.

7. Nose
Some patients lose their sense of smell. Scientists speculate that the virus may move up the nose’s nerve endings and damage cells.

8. Heart and blood vessels
The virus (green) enters cells, likely including those lining blood vessels, by binding to ACE2 receptors on the cell surface. Infection can also promote blood clots, heart attacks, and cardiac inflammation.
AKI Stages: Overall and ICU Admissions

20% of AKI had acute dialysis
Objective: To Evaluate Approaches for predicting the Need for Acute Hemodialysis over a variety of time horizons using data from <24 hours of admission
Study Workflow

A. Facilities

B. Data

COVID - 19 Patients
(Demographics, Comorbidities, Vitals, Labs)

- Inpatient
- Age >= 18
- Positive COVID - 19 RT-PCR

Inclusion Criteria

- Death prior to one session of dialysis
- End Stage Renal Disease
- Features / Patients with >30% missingness
- Admitted for >48 hours before positive COVID test

Exclusion Criteria

C. Timeline

1 Day 3 Days 5 Days 7 Days
## Performance Characteristics of Models over Time Horizons

<table>
<thead>
<tr>
<th></th>
<th>Internal Validation</th>
<th></th>
<th>External Validation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUROC</td>
<td>AUPRC</td>
<td>AUROC</td>
<td>AUPRC</td>
</tr>
<tr>
<td><strong>Horizon: 1 day</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASSO</td>
<td>0.85</td>
<td>0.24</td>
<td>0.82</td>
<td>0.17</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.88</td>
<td>0.29</td>
<td>0.81</td>
<td>0.13</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.91</td>
<td>0.30</td>
<td>0.89</td>
<td>0.23</td>
</tr>
<tr>
<td>XGBoost (imputed)</td>
<td>0.93</td>
<td>0.34</td>
<td>0.91</td>
<td>0.30</td>
</tr>
<tr>
<td><strong>XGBoost (not-imputed)</strong></td>
<td>0.96</td>
<td>0.55</td>
<td>0.96</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>Horizon: 3 days</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASSO</td>
<td>0.86</td>
<td>0.28</td>
<td>0.84</td>
<td>0.25</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.86</td>
<td>0.30</td>
<td>0.82</td>
<td>0.19</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.89</td>
<td>0.39</td>
<td>0.83</td>
<td>0.26</td>
</tr>
<tr>
<td>XGBoost (imputed)</td>
<td>0.92</td>
<td>0.42</td>
<td>0.87</td>
<td>0.33</td>
</tr>
<tr>
<td><strong>XGBoost (not-imputed)</strong></td>
<td>0.94</td>
<td>0.57</td>
<td>0.89</td>
<td>0.44</td>
</tr>
</tbody>
</table>
## Performance Characteristics of Models over Time Horizons

<table>
<thead>
<tr>
<th></th>
<th>Internal Validation</th>
<th></th>
<th>External Validation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUROC (95% CI)</td>
<td>AUPRC (95% CI)</td>
<td>AUROC</td>
<td>AUPRC</td>
</tr>
<tr>
<td><strong>Horizon: 5 days</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASSO</td>
<td>0.86</td>
<td>0.38</td>
<td>0.83</td>
<td>0.26</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.86</td>
<td>0.33</td>
<td>0.81</td>
<td>0.21</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.87</td>
<td>0.40</td>
<td>0.80</td>
<td>0.26</td>
</tr>
<tr>
<td>XGBoost (imputed)</td>
<td>0.87</td>
<td>0.43</td>
<td>0.86</td>
<td>0.32</td>
</tr>
<tr>
<td>XGBoost (not-imputed)</td>
<td>0.89</td>
<td>0.52</td>
<td>0.89</td>
<td>0.46</td>
</tr>
<tr>
<td><strong>Horizon: 3 days</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASSO</td>
<td>0.84</td>
<td>0.39</td>
<td>0.84</td>
<td>0.27</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.84</td>
<td>0.35</td>
<td>0.81</td>
<td>0.22</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.85</td>
<td>0.37</td>
<td>0.81</td>
<td>0.25</td>
</tr>
<tr>
<td>XGBoost (imputed)</td>
<td>0.85</td>
<td>0.40</td>
<td>0.87</td>
<td>0.31</td>
</tr>
<tr>
<td>XGBoost (not-imputed)</td>
<td>0.89</td>
<td>0.54</td>
<td>0.89</td>
<td>0.43</td>
</tr>
</tbody>
</table>
Model Explainability and Features

DIALYSIS (1 DAY)

Creatinine, pH, Anion Gap, Monocyte #, Glucose, Systolic BP, BMI, C Reactive Protein, Calcium, Diastolic BP

DIALYSIS (3 DAYS)

Creatinine, Calcium, pH, BMI, Glucose, C Reactive Protein, Total Bilirubin, Age, Lymphocyte #, Diastolic BP

DIALYSIS (5 DAYS)

Creatinine, pH, C Reactive Protein, Glucose, BMI, Calcium, Platelet Count, Anion Gap, Diastolic BP, Lactate Dehydrogenase

DIALYSIS (7 DAYS)

Creatinine, C Reactive Protein, pH, BMI, Glucose, Calcium, Age, Lymphocyte #, Lactate Dehydrogenase, Potassium
Acknowledgements

Adam Bernheim
Akhil Vaid
Alexander Charney
Allan Just
Andrew Kasarskis
Andrew Leader
Arash Kia
Arden Moscati
Ashish Atreja
Benjamin Glicksberg
Bethany Percha
Carlos Cordon
Carol Horowitz
Christie Chang
Christopher Bellaire
Cristoforo Terrasi
Daniel Carrion
Daniel Robins
Dara Meyer
Eddye Golden
Erick Scott
Erwin Bottinger
Farah Fasihuddin
Fayzan Chaudhry
Girish Nadkarni
Huei Hsun Wen
Ishan Paranjpe
Ismail Nabeel
Jason Bobe
Jason Rogers
Jennifer Gutierrez
Jessica DeFreitas
Johnathan Rush
Kathryn Twyman
Keith Sigel
Kodi Arfer
Krishna Vemuri
Kumardeep Chaudhary
Laura Huckins
Manbir Singh
Marco Pereanez
Matteo Danielello
Melissa Nieves
Mesude Bicak
Michael Chung
Micol Zweig
Nancy Yi
Nicholas Defelice
Noam Beckmann
Paul O'Reilly

MSCIC

▶ Ale
▶ Zah
▶ Sav
▶ Erw
▶ Jos
▶ Pati
▶ Rot