Use of Machine Learning and Michigan EHR Data

to assess the prognostic utility of radiomic features in predicting the survival of hospitalized patient with COVID-19

Stephen Salerno and Yuming Sun
on behalf of

Yi Li Lab

April 13, 2022

Michigan Biostatistics Collaborative Lunches

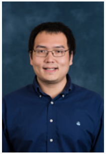
Table of Contents

Background

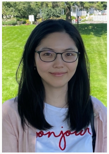
Our Team

Yi
Li

Peisong
Han



Jian
Kang



Jiyeon
Song

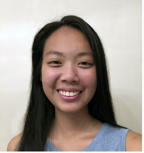
Chinakorn
Sujimongkol

Yuming
Sun

Stephen
Salerno

Xinwei
He

Ziyang
Pan



Eileen
Yang

Michigan Medicine & Precision Health

- Michigan Medicine (MM) is one of the primary regional centers managing the care of COVID-19 patients
- A dedicated team at MM has created, maintained, and updated an EMR database for COVID-19 patients treated since the outbreak
- The Precision Health Analytics Platform makes tools, services, and datasets available to UM researchers: precisionhealth.umich.edu/
- Access to this rich database enables us to conduct a comprehensive analysis of COVID-19 outcomes

Some Michigan Medicine COVID-19 Papers

BIOSTATISTICS

Original Investigation | Infectious Diseases

October 21, 2020

Characteristics Associated With Racial/Ethnic Disparities in COVID-19 Outcomes in an Academic Health Care System

Tian Gu, MS[1]; Jasmine A. Ma

Author Affiliations | Art

JAMA Netw Open. 2020;3(1

COVID-19 Resource Cer

PLOS ONE

advanced search

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Exposure and risk factors for COVID-19 and the impact of staying home on Michigan residents

Kuan-Han H. Wu, Whitney E. Hornsby, Bethany Klunder, Amelia Krause, Anisa Driscoll, John Kulka, Ryan Bickett-Hickok, Austin Fellows, Sarah Graham, Erin O. Kaleba, Salim S. Hayek, Xu Shi, Nadia R. Sutton, […], Cristen J. Willer [view all]

Published: February 8, 2021 • https://doi.org/10.1371/journal.pone.0246447

Article | Authors | Metrics | Comments | Media Coverage

Download PDF | Print | Share

PLOS ONE

advanced search

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Comprehensive evaluation of COVID-19 patient short- and long-term outcomes: Disparities in healthcare utilization and post-hospitalization outcomes

Stephen Salerno, Yuming Sun, Emily L. Morris, Xinwei He, Yajing Li, Ziyang Pan, Peisong Han, Jian Kang, Michael W. Sjoding, Yi Li

Published: October 6, 2021 • https://doi.org/10.1371/journal.pone.0258278

Article | Authors | Metrics | Comments | Media Coverage

Download PDF | Print | Share

medRx

THE PREPRINT SERVER FOR HEALTH

Estimating COVID-19 Va… of an Academic Medical Center in Michigan

-19 … a in Michigan Medicine

Emily K. Roberts, Tian Gu, Bhramar Mukherjee, Lars G. Fritsche

…ehal Patil, Thomas S. Valley, Lars G. Fritsche,

doi: https://doi.org/10.1101/2022.01.29.22269971

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

Abstract | Full Text | Info/History | Metrics | Preview PDF

Patterns of repeated diagnostic testing for COVID-19 in relation to patient characteristics and outcomes

S. Salerno[1,*], Z. Zhao[1,*], S. Prabhu Sankar[2,3], M. Salvatore[1], T. Gu[1], L. G. Fritsche[1,2,4], S. Lee[1,5], L. D. Lisabeth[6], T. S. Valley[7,8] & B. Mukherjee[1,2,6]

Results from Prior Literature

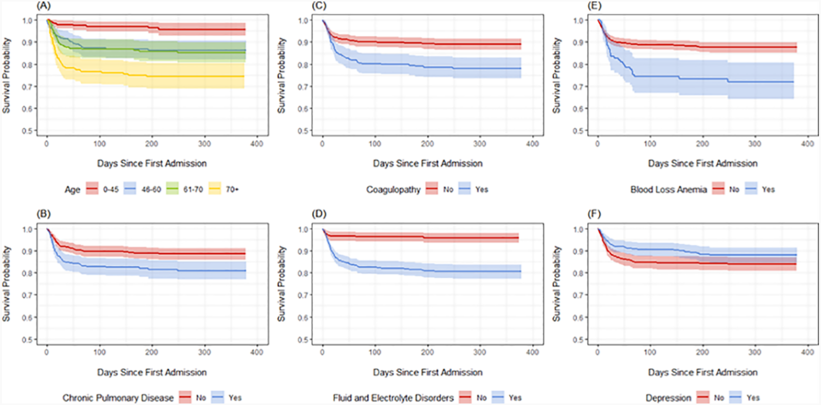
1. Risk factors identified for COVID-19 in previous literature:
    - Older age, male sex, higher body mass index
    - Racial, ethnic, and socioeconomic factors
    - Comorbidity conditions: cardiovascular disease, diabetes, chronic respiratory disease, hypertension, and cancer
2. Early work potentially limited by sample size or follow-up duration
3. Risk factors collected during the early phases of the pandemic may differ as the virus became more widespread

Salerno et al., 2021

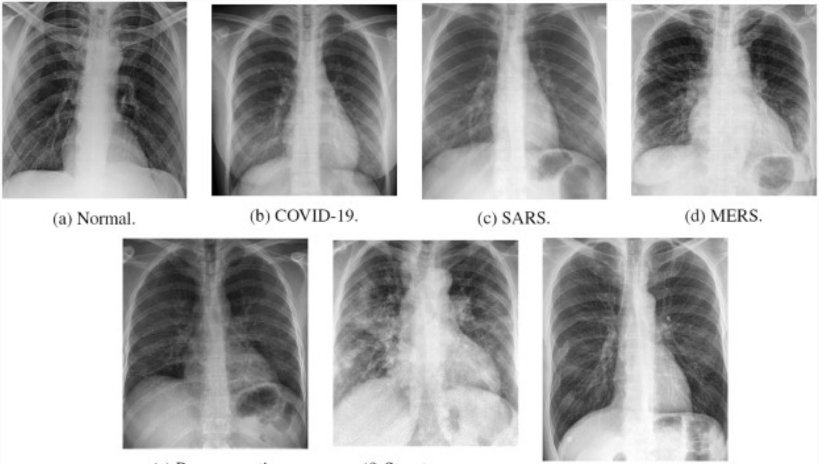
Our Prior Results on COVID-19 using the Michigan Medicine EHR and DataDirect found:

- Disparate outcomes associated with age, sex, and race, which have become particularly salient, were corroborated in this study
- Higher comorbidity burden was shown to be associated with higher healthcare utilization, hospitalization, and mortality rates
- Adjusting for other demographic and comorbid conditions, Black patients had lower utilization of medical resources, but higher mortality and higher hospitalization hazards

KM curves for post-admission mortality, stratified by (A) age quartiles, (B) chronic pulmonary disease, (C) coagulopathy, (D) fluid and electrolyte disorders,(E) blood loss anemia, and (F) depression:



The availability and easy of use has made portable CXR at MM a valuable tool for monitoring and guiding the care of patients with COVID-19:



(a) Normal. (b) COVID-19. (c) SARS. (d) MERS.

(e) Pneumocystis. (f) Streptococcus. (g) Varicella.

https://www.medrxiv.org/content/10.1101/2020.12.14.20248158v1.full.pdf

Motivation for Current Work

- While chest X-rays have utility in identifying COVID-19 infection, high variability in the irregular or patchy reticular opacifications present difficulties when grading patients
- Using X-ray data to predict long-term outcomes, though promising, is statistically challenging due the high-dimensional and heterogeneous nature of the relevant features
- A risk stratification project using chest X-ray imaging, led by Michael Sjoding, has collected over 100,000 images on patients at MM

Motivation for Current Work

In response to the unique challenges associated with chest X-ray data for COVID-19, that is, no available image segmentation information, we perform feature extraction and selection with the X-ray imaging data

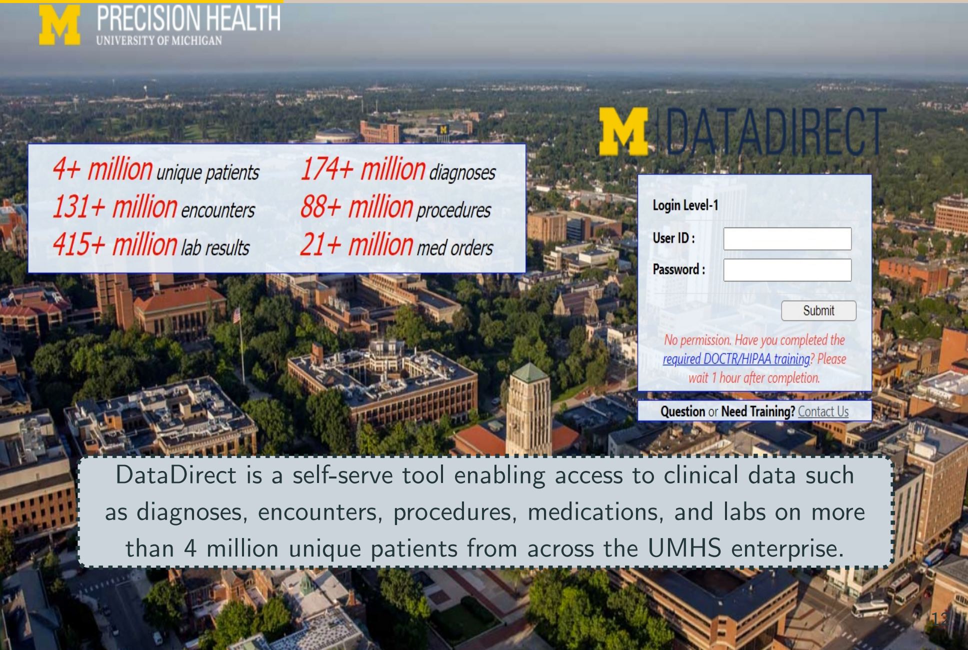
Motivating Question:

Making full use of patient demographic data, physiologic and laboratory measurements, and prevalent comorbidities available through DataDirect and the Michigan Medicine EHR, what is the incremental prognostic utility of radiomic features?

Study Design

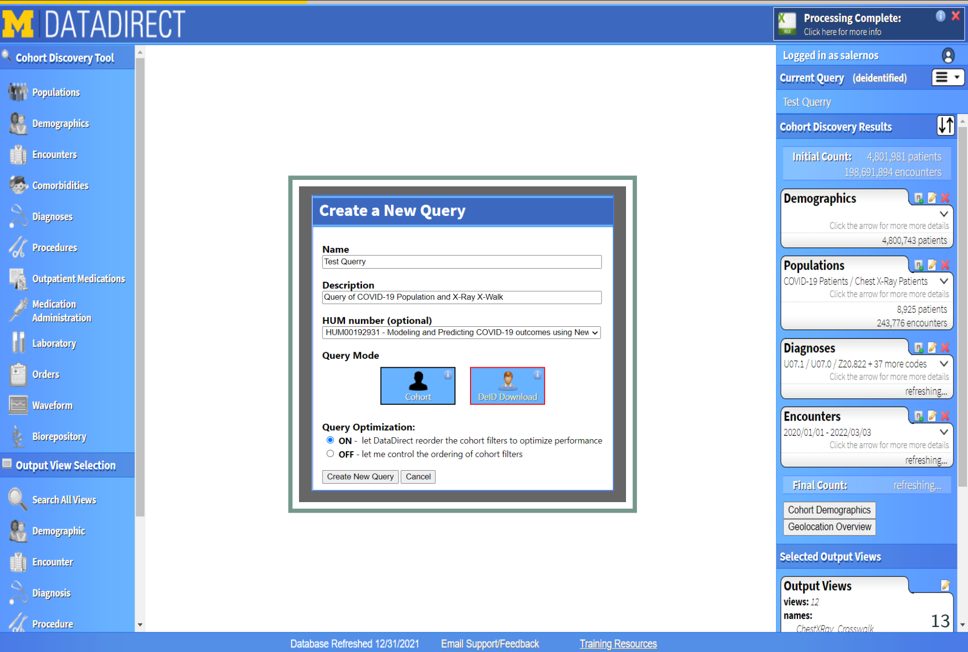
- Retrospective study of hospitalized COVID-19 patients with available data through DataDirect and at least one X-ray image
- First build survival prediction models from index COVID-19 hospitalization or ICU admission to death using clinical features
- Then re-train these models with additional radiomic texture features to assess their incremental prognostic utility

Data Acquisition



DataDirect is a self-serve tool enabling access to clinical data such as diagnoses, encounters, procedures, medications, and labs on more than 4 million unique patients from across the UMHS enterprise.

Querying DataDirect for Clinical Features



Data

Study Population

- COVID-19 patients who were treated at MM between March 10, 2020 and December 31, 2021, with at least one image taken
- These patients were tested positive either at Michigan Medicine or elsewhere and transferred in (diagnosis code of U07.1 or U07.2)
- We observe 1,403 hospitalized patients with COVID-19 having at least one X-ray image (1,378 with coronal views retained)

Outcome

- Patients were followed from first (index) hospitalization date following a positive COVID-19 diagnosis or from presentation to ED with subsequent diagnosis and admission
- Patients were followed until death (event), discharge (censored), or the end of follow-up (censored; December 31, 2021)
- Discharge to hospice was considered as an event with death (composite outcome; median survival post-discharge < 2 weeks)
- Among 1,378 patients in our analytic sample, we observe 299 in-hospital deaths and 7 patients discharged to hospice

Clinical Predictors

1. Patient age, sex assigned at birth (male vs. female), self-identified race (Black, non-Black, unknown), and body mass index (kg/m^2)

2. Quartile ranks for neighborhood affluence indicators (proxy for socioeconomic status)

3. Comorbidities (29 in total) coded as indicators, flagging if a patient carried any associated ICD-10 codes during the study frame:
   - All Elixhauser comorbidity index items, except for HIV/AIDS status.
   - Previously reported to be predictive of patient outcomes in numerous non-COVID and COVID-19 settings

4. Physiologic measurements (e.g., respiratory rate, oxygen saturation) taken within one week of admission

Select Patient Characteristics

Table 1: Descriptive characteristics for patient sample; Median (IQR) or N (%)

Characteristic	$N = 1,378$[1]
Age (yrs.)	61 (46, 72)
Female	549 (40%)
White	921 (68%)
Non-Hispanic	1,278 (94%)
Elixhauser Total Score	8.2 (5.0, 11.9)
Fluid & Electrolyte Disorders	1,089 (79%)
Metastatic Cancer	324 (24%)
Body Mass Index (kg/m^2)	29 (25, 35)
Oxygen Saturation (SpO2)	95.24 (93.77, 96.80)
Temperature ($^\circ F$)	98.30 (98.02, 98.76)
Respiratory Rate (BPM)	19.7 (17.9, 23.5)
Blood Pressure (Systolic)	122 (111, 134)
Blood Pressure (Diastolic)	67 (62, 72)

Image Pre-Processing

Medical Imaging

Anatomical Coordinate System: 3D space where image has been sampled, defined along the anterior-posterior (coronal), inferior-superior (axial), and left-right (sagittal) axes

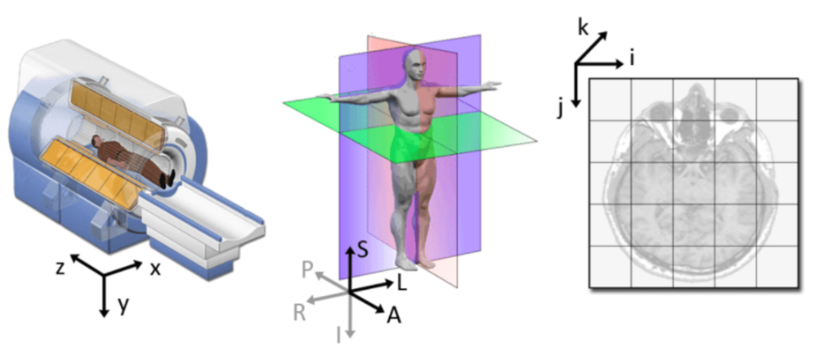
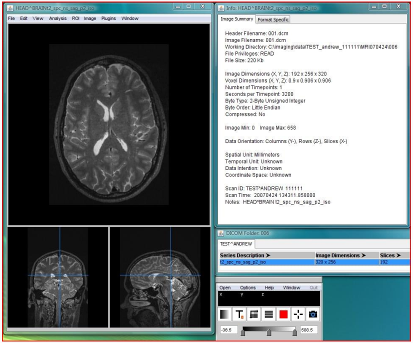


Image Source: https://www.slicer.org/wiki/Coordinate$_s$*ystems*

DICOM Images

DICOM (Digital Imaging and Communications in Medicine):

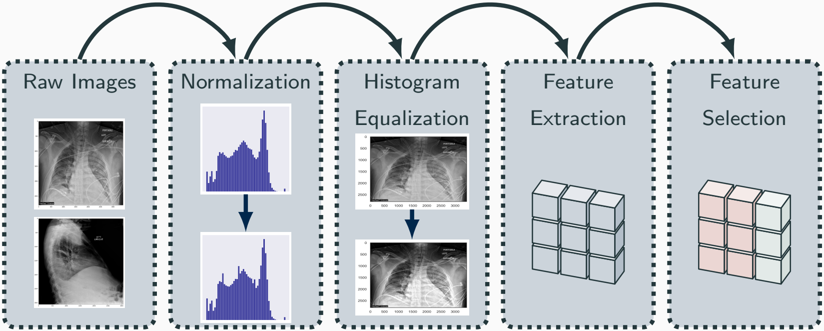
- International standard to transmit, store, retrieve, print, process, and display medical imaging information
- www.dicomstandard.org
- Contains pixel and metadata
- Header metadata: PHI and image acquisition parameters



https://technologyadvice.com/blog/healthcare/5-dicom-viewers/

Image Pre-Processing Workflow

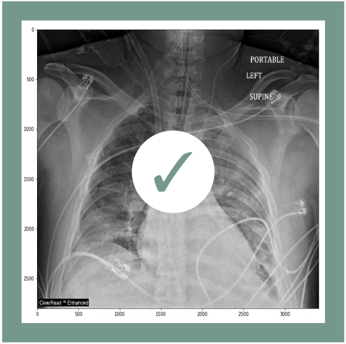
To extract useful information from X-ray images and predict the survival outcome for each patient, we follow the procedure shown below to process the images and extract features:

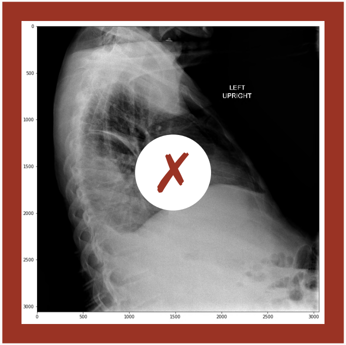


Raw Images

Raw Images ▶ Normalization ▶ Equalization ▶ Extraction ▶ Selection

For 1,403 hospitalized patients who have X-ray images taken within one month of the COVID diagnosis, we use 1,378 patients who have coronal view images to carry out the analysis:

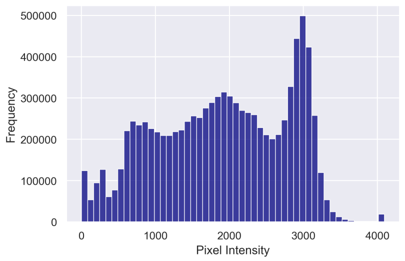




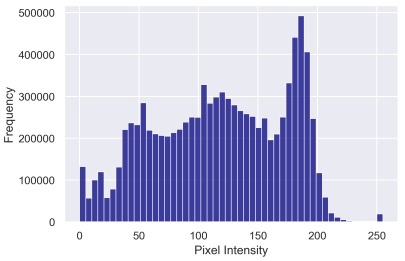
Normalization

Raw Images → **Normalization** → Equalization → Extraction → Selection

We normalize the pixel intensities of each image to the range of 0 - 255:



(a) Before Normalization

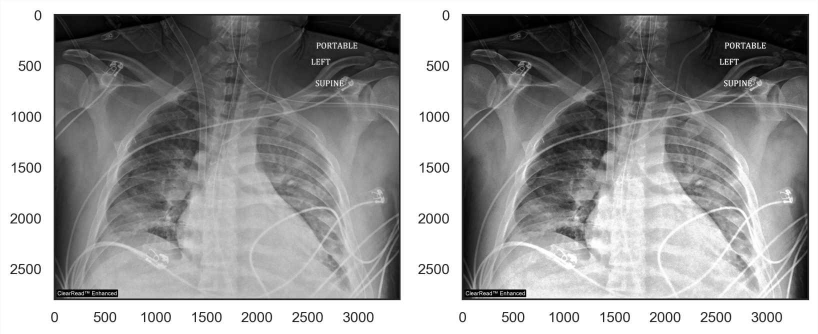


(b) After Normalization

Figure 1: Distribution of pixel intensity before and after normalization

Histogram Equalization

Histogram equalization enhances the contrast of the X-ray images:



(a) Before histogram equalization

(b) After histogram equalization

Figure 2: Enhance the contrast by histogram equalization

Feature Extraction

Raw Images → Normalization → Equalization → **Extraction** → Selection

- We extract 7 classes of features from each original image (e.g., first order features, gray level co-occurrence matrix)
- In addition, we apply 6 filters (transformations of pixel density) to the original images to derive different image types (e.g., Laplacian of Gaussian filter)
- Similar to the original image, we can extract seven classes of features from each of the image types which result in 1,311 features in total

Feature Selection

We then conduct feature selection to reduce the feature dimension:

- We first fit univariate cox models for each of the features and exclude those features that are insignificant
- For the features that have correlation higher than 0.7, we only include the features with the lowest p-values
- Finally, we conduct forward selection according to the rank of prediction power for the remaining features and select those features that gives highest concordance statistic (C-stat)

Statistical Methods and Models

Statistical Methods and Models

We fit the following survival prediction models and studied the prognostic accuracy of these approaches using Harrell's C-Index, Uno's C-Index, and the area under the receiver operator characteristics curve (AUC):

- Cox proportional hazards model
- Survival Support Vector Machine
- Ensemble Models
  - Survival Gradient Boosting
  - Random Survival Forest

Assess the importance of selected features by

- feature importance: the decrease of prediction power (C-Index, accuracy ,R^2, etc.) when the feature is removed
- permutation feature importance: the decrease of prediction power when the feature is permuted
- splitting the data into training and testing, and permuting the feature values in testing data

Results

Prediction Performance

		Clinical	Clinical & Imaging
Cox	Harrell's C	0.75 (0.0028)	0.76 (0.0028)
	Uno's C	0.74 (0.0031)	0.75 (0.0032)
	AUC	0.76 (0.0035)	0.77 (0.0034)
SVM	Harrell's C	0.76 (0.0026)	0.77 (0.0024)
	Uno's C	0.75 (0.0029)	0.75 (0.0030)
	AUC	0.79 (0.0032)	0.80 (0.0027)
Random Forest	Harrell's C	0.76 (0.0029)	0.76 (0.0028)
	Uno's C	0.74 (0.0033)	0.74 (0.0035)
	AUC	0.79 (0.0036)	0.80 (0.0033)
Gradient Boosting	Harrell's C	0.76 (0.0029)	0.77 (0.0030)
	Uno's C	0.75 (0.0032)	0.76 (0.0034)
	AUC	0.79 (0.0034)	0.80 (0.0033)

Table 2: Prediction performance of different algorithms using clinical and clinical + imaging data

Kaplan–Meier Curve

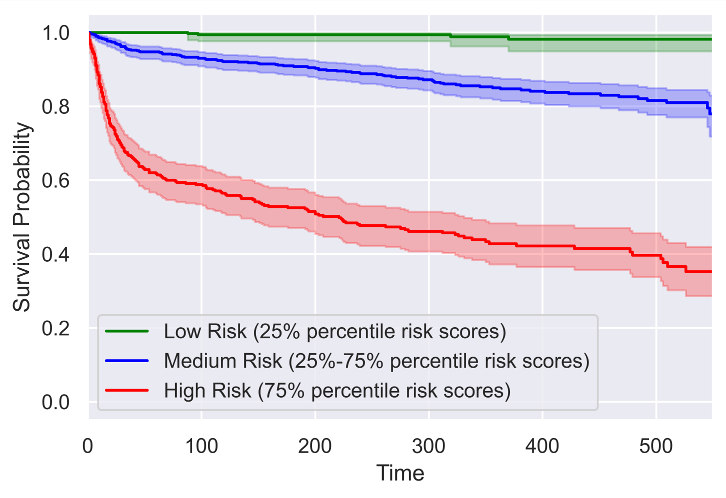
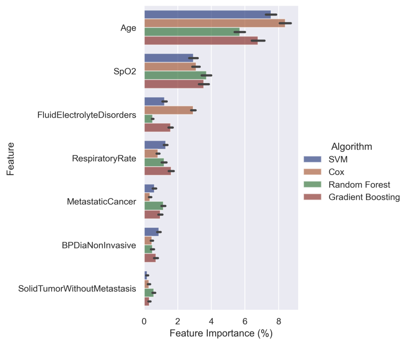
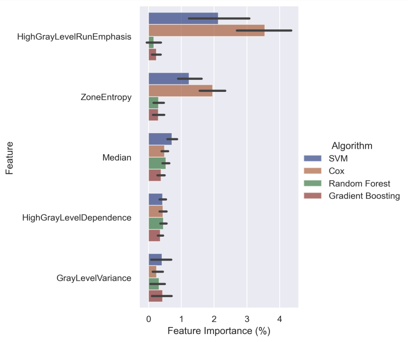


Figure 3: Kaplan–Meier curve using risk scores from SVM



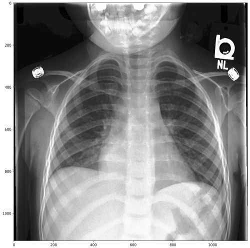
(a) Clinic Features



(b) Image Features

Figure 4: Feature importance of 100 experiments on testing data

Individual Texture Features for 2 Example Patients



Outcome	Alive	Dead
Survival Time (days)	451	225
High Gray Level Dependence	-1.70	0.84
High Gray Level Run Emphasis	-1.01	1.16
Median	-0.42	2.25
Zone Entropy	-4.49	1.13
Gray Level Variance	-1.01	1.60

Table 3: Important texture features of two selected patients

Texture Feature Distribution

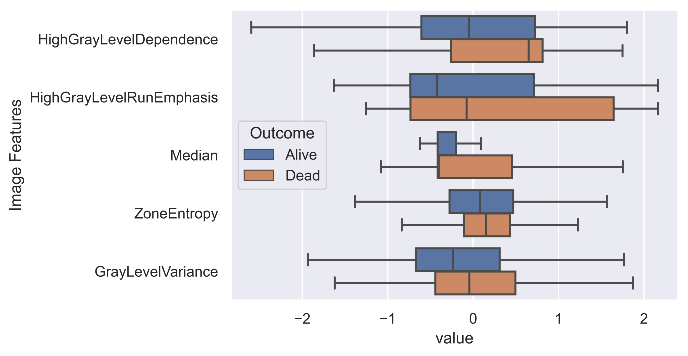


Figure 5: Distribution of important texture features

Adjusted Associations

Covariate	Hazard Ratio	95% CI	p value
Age	2.28	1.94 (2.68)	<0.005
Oxygen Saturation (SpO2)	0.67	0.60 (0.76)	<0.005
Fluid & Electrolyte Disorders	3.30	2.01 (5.42)	<0.005
Respiratory Rate	1.19	1.04 (1.37)	0.01
Metastatic Cancer	1.33	0.96 (1.85)	0.09
Diastolic Blood Pressure	0.83	0.74 (0.93)	<0.005
Solid Tumor Without Metastasis	1.32	0.95 (1.84)	0.10
Smoking Status			
Current	-	-	-
Former	0.66	0.38 (1.14)	0.14
Never	0.50	0.28 (0.87)	0.01
Unknown	0.93	0.53 (1.63)	0.80

Table 4: Adjusted associations of clinic and image features with COVID-19 outcomes among hospitalized COVID-19 patients

Adjusted Associations

- Age was the most important predictor, and older age associated with higher hazards of in-hospital death
- Higher oxygen saturation was associated with lower mortality hazards
- Fluid and electrolyte disorders on admission indicative of severely worsened outcomes

Adjusted Associations, cont.

- Patients had higher respiratory rates, lower blood pressure, on average, with tachypnea potentially indicative of respiratory distress
- Patients with cancer had a predisposition to being sicker – on immune-modifiers/chemotherapy
- Smoking status was also associated with worsened outcomes for hospitalized patients with COVID-19

*Difficult to interpret as many are on pressors coming from the emergency department

Adjusted Associations

Covariate	Hazard Ratio	95% CI	p value
High Gray Level Dependence	1.08	0.91 (1.27)	0.39
High Gray Level Run Emphasis	1.27	1.11 (1.46)	<0.005
Median	1.09	0.99 (1.21)	0.09
Zone Entropy	1.39	1.14 (1.68)	<0.005
Gray Level Variance	1.32	1.17 (1.48)	<0.005

Table 5: Adjusted associations of clinic and image features with COVID-19 outcomes among hospitalized COVID-19 patients (Continued)

- Lung consolidation and ground glass opacities are most common patterns on CXR and CT
- Irregular, patchy and widespread reticular opacities around regions of ground glass attenuation are more easily found on CXR

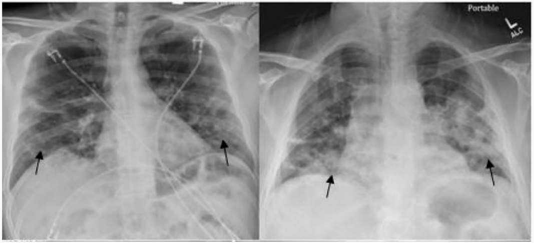


Figure 6: Two COVID-19 patients with reticular opacities (Jacobi et al., 2020)

Discussion

Limitations

- Only hospitalizations at Michigan Medicine were included in this analysis
- Diagnosis of comorbidities after positive COVID-19 test was not differentiated from those preceding infection
- Generalizability of these results to other populations was limited due to a focus on the Michigan Medicine patient population

Strengths

- Study had a large cohort of COVID-19 patients
- It had relatively long followup with a variety of outcomes observed, featuring a natural disease evolution trajectory
- Patient population presented a sizable portion of minority patients, facilitating study of racial disparities in COVID outcomes

Implications

BIOSTATISTICS

- The availability and easy of use has made portable CXR a valuable tool for monitoring and guiding the care of patients with COVID-19
- Patterns of COVID-19 lung disease that are identifiable on CXR are significantly associated with the survival outcomes of patients
- As the pandemic progresses, widely available CXR will be frequently used and able to help us predict the prognostic outcomes of patients

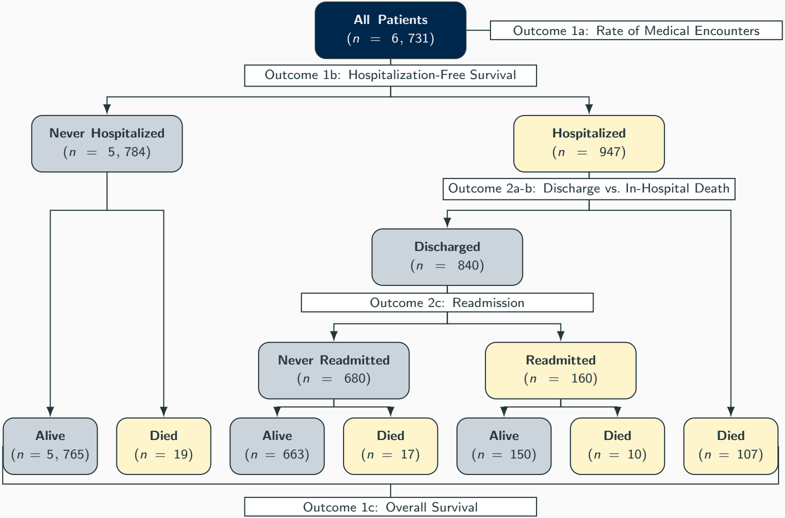
Thank You!

Thank you to all Michigan Medicine employees, frontline workers, researchers, and patients for your dedication to ending the pandemic!

Questions?

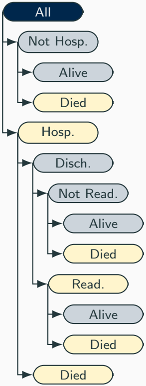
Yi Li Lab PLOS One Paper (2021)

Figure S1: Disease progression: patient outcomes, sub-populations, and models

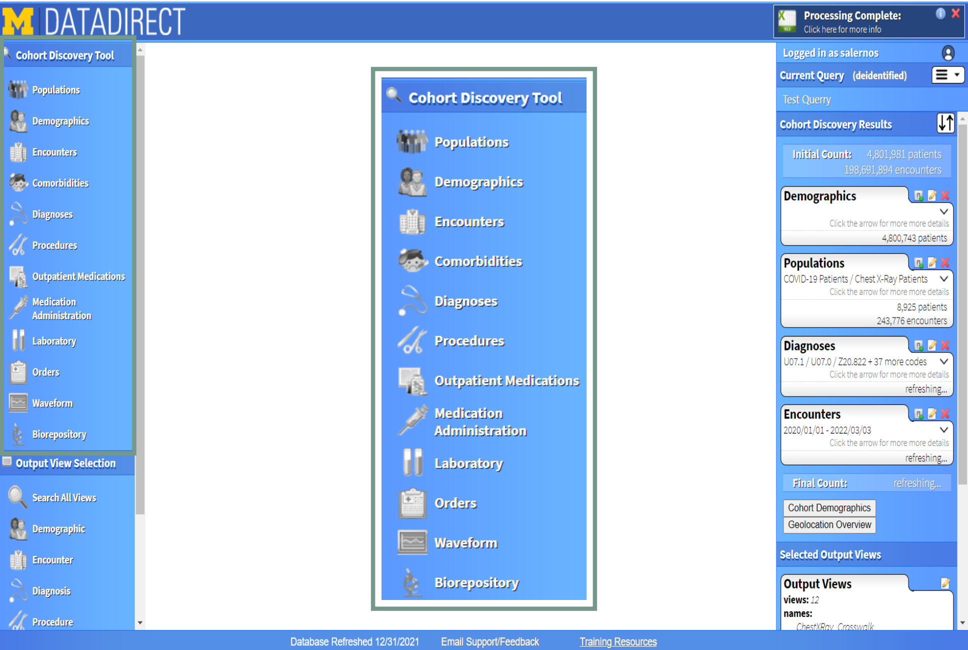


Yi Li Lab PLOS One Paper (2021)

- Among 6,731 COVID-19 patients, 974 (14%) were hospitalized and 153 died (2%)
- Among those hospitalized, 840 (89%) were discharged and 160 (17%) were readmitted
- 134 hospitalized deaths: 107 (81%) in-hospital, 17 (12%) after discharge, 10 (7%) after readmission

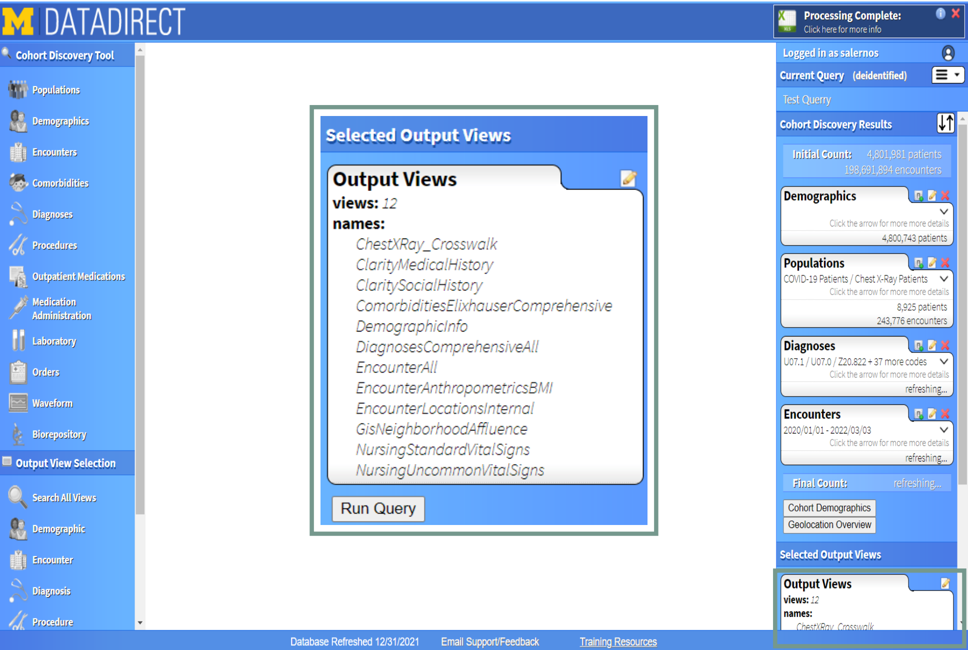


Options for Selecting Specific EMR Data

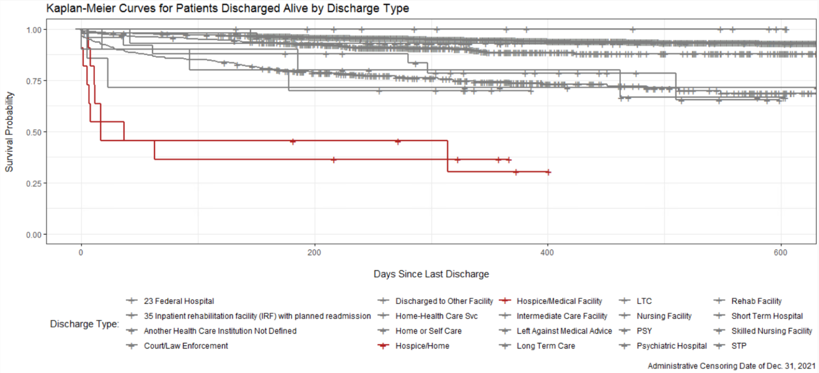


Subsetting on COVID-19 and X-Ray Populations

Output Clinical Measurements



Kaplan-Meier Estimated Survival Post-Discharge



Neighborhood Socioeconomic Status (NSES)

Defined four composite measures of NSES at the US census tract-level based on patient residences, derived from the National Neighborhood Data Archive:

1. *Affluence Quartile*: the quartile in which the average proportion of households with income greater than $75K, proportion of the population aged 16+ employed in professional or managerial occupations, and proportion of adults with Bachelor's degrees or higher falls under
2. *Disadvantage Quartile*: the quartile in which the average proportion of non-Hispanic Black, proportion of female-headed families with children, proportion of households with public assistance income or food stamps, proportion of families with income below the federal poverty level, and proportion of the population aged 16+ unemployed falls under
3. *Ethnic Immigrant Concentration Quartile*: the quartile in which the average proportion of Hispanic and proportion of foreign born falls under
4. *Education Quartile*: the quartile in which average proportion of adults with less than a high school diploma falls under

Elixhauser Comorbidity Conditions i

Comorbidity Condition	ICD-10 Codes
Congestive Heart Failure	I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0
Cardiac Arrhythmias	I44.1-I44.3, I45.6, I45.9, I47.x-I49.x, R00.0, R00.1, R00.8, T82.1, Z45.0, Z95.0
Valvular Disease	A52.0, I05.x-I08.x, I09.1, I09.8, I34.x-I39.x, Q23.0-Q23.3, Z95.2-Z95.4
Pulmonary Circulation Disorders	I26.x, I27.x, I28.0, I28.8, I28.9
Peripheral Vascular Disorders	I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9
Hypertension, Uncomplicated	I10.x
Hypertension, Complicated	I11.x-I13.x, I15.x
Paralysis	G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0-G83.4, G83.9
Other Neurological Disorders	G10.x-G13.x, G20.x-G22.x, G25.4, G25.5, G31.2, G31.8, G31.9, G32.x, G35.x-G37.x, G40.x, G41.x, G93.1, G93.4, R47.0, R56.x
Chronic Pulmonary Disease	I27.8, I27.9, J40.x-J47.x, J60.x-J67.x, J68.4, J70.1, J70.3
Diabetes, Uncomplicated	E10.0, E10.1, E10.9, E11.0, E11.1, E11.9, E12.0, E12.1, E12.9, E13.0, E13.1, E13.9, E14.0, E14.1, E14.9
Diabetes, Complicated	E10.2-E10.8, E11.2-E11.8, E12.2-E12.8, E13.2-E13.8, E14.2-E14.8

Elixhauser Comorbidity Conditions ii

Comorbidity Condition	ICD-10 Codes
Hypothyroidism	E00.x-E03.x, E89.0
Renal Failure	I12.0, I13.1, N18.x, N19.x, N25.0, Z49.0-Z49.2, Z94.0, Z99.2
Liver Disease	B18.x, I85.x, I86.4, I98.2, K70.x, K71.1, K71.3-K71.5, K71.7, K72.x-K74.x, K76.0, K76.2-K76.9, Z94.4
Peptic Ulcer Disease, Excluding Bleeding	K25.7, K25.9, K26.7, K26.9, K27.7, K27.9, K28.7, K28.9
Lymphoma	C81.x-C85.x, C88.x, C96.x, C90.0, C90.2
Metastatic Cancer	C77.x-C80.x
Solid Tumour without Metastasis	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-C58.x, C60.x-C76.x, C97.x
Rheumatoid Arthritis/Collagen Vascular Diseases	L94.0, L94.1, L94.3, M05.x, M06.x, M08.x, M12.0, M12.3, M30.x, M31.0-M31.3, M32.x-M35.x, M45.x, M46.1, M46.8, M46.9
Coagulopathy	D65-D68.x, D69.1, D69.3-D69.6
Obesity	E66.x
Weight Loss	E40.x-E46.x, R63.4, R64
Fluid and Electrolyte Disorders	E22.2, E86.x, E87.x
Blood Loss Anaemia	D50.0

Elixhauser Comorbidity Conditions iii

Comorbidity Condition	ICD-10 Codes
Deficiency Anaemia	D50.8, D50.9, D51.x-D53.x
Alcohol Abuse	F10, E52, G62.1, I42.6, K29.2, K70.0, K70.3, K70.9, T51.x, Z50.2, Z71.4, Z72.1
Drug Abuse	F11.x-F16.x, F18.x, F19.x, Z71.5, Z72.2
Psychoses	F20.x, F22.x-F25.x, F28.x, F29.x, F30.2, F31.2, F31.5
Depression	F20.4, F31.3-F31.5, F32.x, F33.x, F34.1, F41.2, F43.2

Fluid & Electrolyte Disorders

The indicator for fluid and electrolyte disorders is present if the patient has any of the following comorbid conditions:

- Syndrome of inappropriate secretion of antidiuretic hormone
- Volume depletion (e.g., dehydration, hypovolemia, or unspecified volume depletion)
- Other disorders of fluid, electrolyte, and acid-base balance

References i