Math 5545 Project 3 (Due: April 15, 2021)
AI Predictive Modeling to Identify COVID Positive Individuals Using Chest X-ray Images

Group # 1: X-ray Images 101-600

Ford, William Andrew waftx2@mail.umkc.edu
Tran, Thao Phuong ttcww@mail.umkc.edu
** Johnson, Reece LaVaughn rlj6pb@mail.umkc.edu
Whetsell, Torsten P torsten.whetsell@mail.umkc.edu
* Salas,J acob R jrsty8@mail.umkc.edu

Group # 2: X-ray Images 601-1000

Vaughn, Braeden bv2my@mail.umkc.edu
Stack, Caston A cas8y5@mail.umkc.edu
* Terry, Harrison Edward Lee het9t5@mail.umkc.edu
** Schaeffer, Alex as9nb@mail.umkc.edu

Group # 3: X-ray Images 1001-1500

Lim, Celine Shwu Ling csl9r3@mail.umkc.edu
Reesman, Grace Ellen gekhx5@mail.umkc.edu
Rosenblatt, Jennifer jrosenblatt@mail.umkc.edu
** Thomas, Micheal mctwcc@mail.umkc.edu
* Aljofei, Maha Mashan mmawv6@umkc.edu

*** team leader.** The team leader is responsible for organizing the zoom meetings, distributing the work between the team members coordinating with the project manager and making sure that the project is getting finished before the deadline. If you do not receive an email from your team leader by the end of this week, please let me know by Monday morning.

**** project manager**. the project manager is responsible for the quality of the project, making sure all calculations are correct, coordinating with the team leader and submitting all required work.

Collegiality and Group Work

The groups in this class are meant to imitate real-world research groups. Each group should regularly meet via zoom. Each group member should (1) maintain a friendly environment for the entire group; (2) facilitate collaboration and problem solving; (3) provide a vision of the main objectives and ensure discussions lead to conclusions and decisions; (4) motivate and inspire other group members; (5) contribute to the group by sharing his/her knowledge, expertise, and viewpoints; (6) participate in all meetings and discussions; (7) have productive suggestions.

AI Predictive Modeling to Identify COVID Positive Individuals Using Chest X-ray Images

Background & Objectives

One of the crucial steps in fighting COVID-19 is the ability to detect the infected patients early enough and put them under special care.

Detecting this disease from radiography and radiology images is perhaps one of the fastest ways to diagnose the patients.

Some of the early studies showed specific abnormalities in the chest radiograms of patients infected with COVID-19.

Inspired by earlier works, we study the application of machine learning models to predict COVID-19 patients from their chest radiography images.

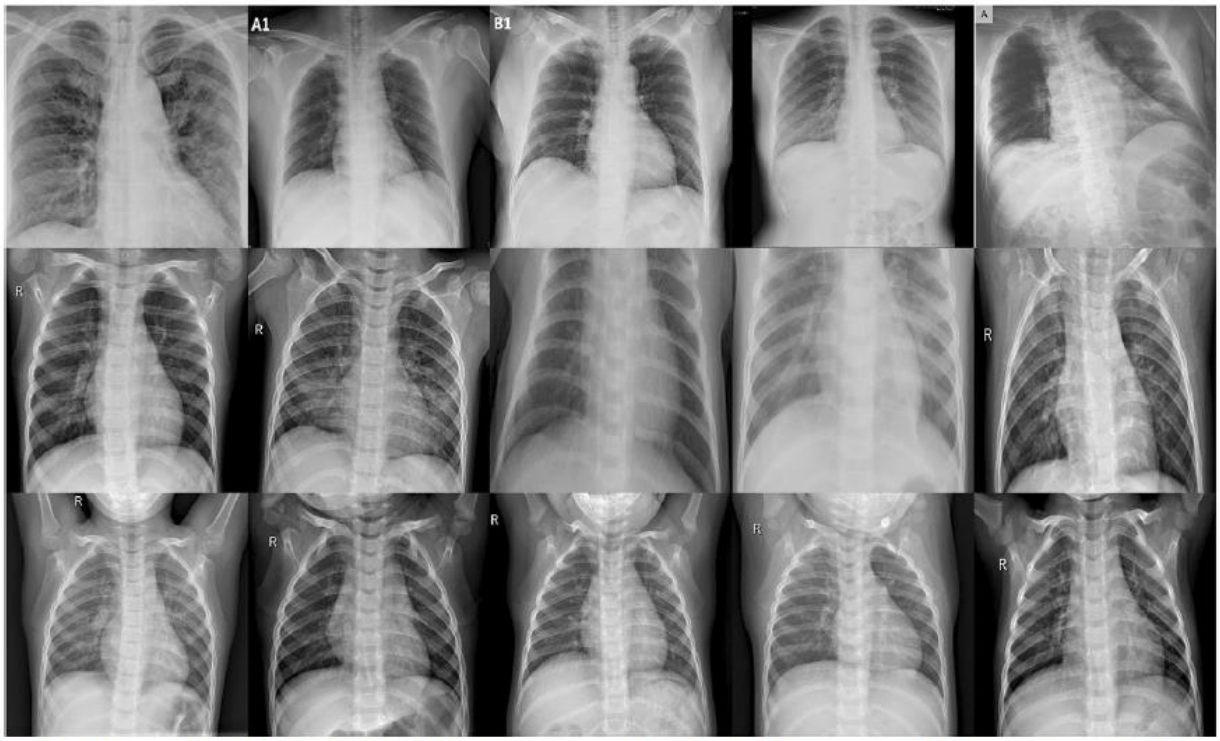


Fig. 1 Chest X-rays from COVID-19 Radiography Database: COVID-19 (Row 1), viral pneumonia (Row 2), and normal (Row 3)

Procedure:

Step 1: Download and unzip X-ray images (740 MB)

We first prepare a dataset of 200 Chest X-rays from the publicly available datasets. Images exhibiting COVID-19 disease presence were identified by board-certified radiologist.

DATABASE

A public open dataset of chest X-ray images. This database contains chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images.

https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

About this directory *COVID-19 CHEST X-RAY DATABASE

A team of researchers have created a database of chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images. This COVID-19, normal and other lung infection dataset is released in stages. In the first release we have released 219 COVID-19, 1341 normal and 1345 viral pneumonia chest X-ray (CXR) images. In the first update, we have increased the COVID-19 class to 1200 CXR images. In the 2nd update, we have increased the database to 3616 COVID-19 positive cases along with 10,192 Normal, 6012 Lung Opacity (Non-COVID lung infection) and 1345 Viral Pneumonia images. We will continue to update this database as soon as we have new x-ray images for COVID-19 pneumonia patients.

Step 2: Data Visualization and interpretations using PCA

1. Create a folder of 200 images assigned to you (see the following table).

Important note: The folder name must be "COVIDData"

Group # 1: X-ray Images 101-600

- Ford, Andrew (from COVID-101.png to COVID-200.png & from Normal-101.png to Normal-200.png)
- Tran, Thao (from COVID-201.png to COVID-300.png & from Normal-201.png to Normal-300.png)

- Johnson, Reece (from COVID-301.png to COVID-400.png & from Normal-301.png to Normal-400.png)
- Whetsell, Torsten (from COVID-401.png to COVID-500.png & from Normal-401.png to Normal-500.png)
- Salas, Jacob (from COVID-501.png to COVID-600.png & from Normal-501.png to Normal-600.png)

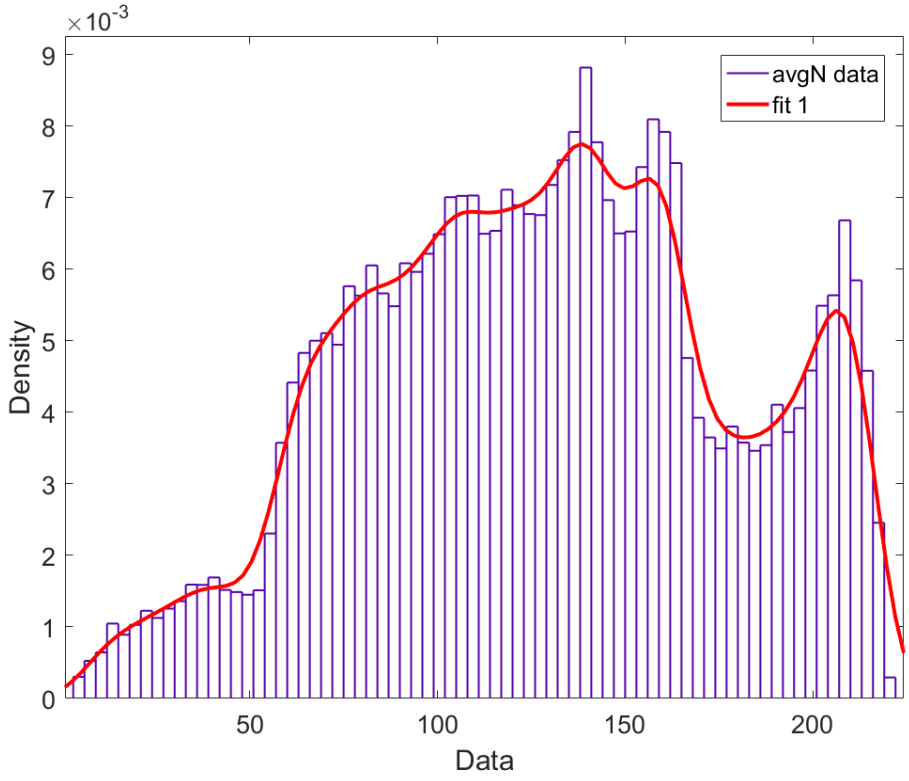
Group # 2: X-ray Images 601-1000

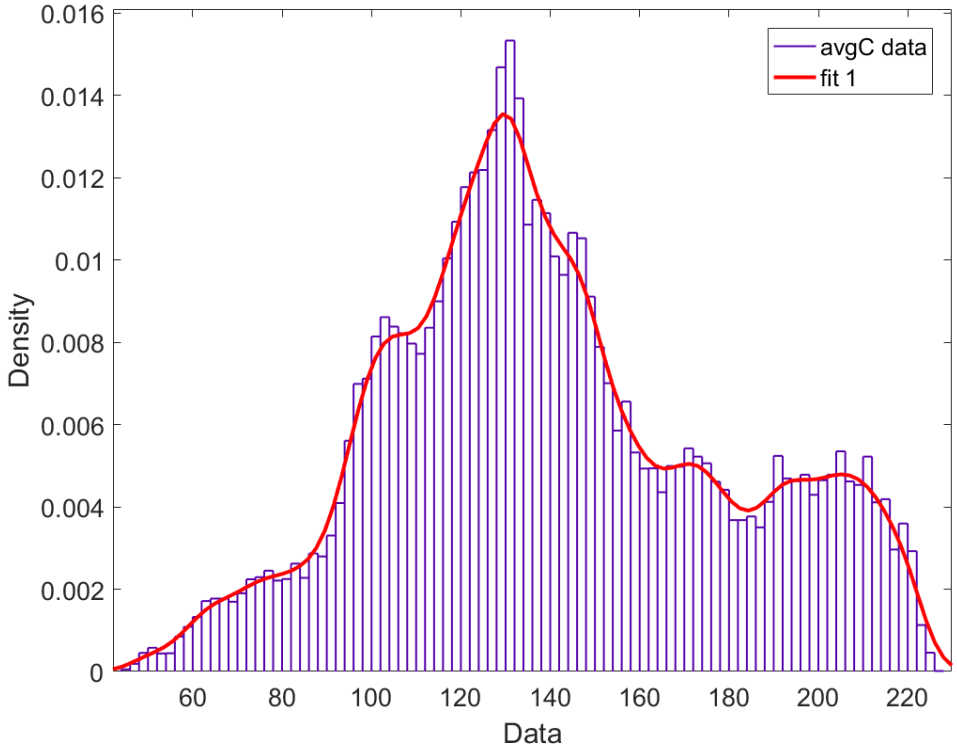
- Vaughn, Braeden (from COVID-601.png to COVID-700.png & from Normal-601.png to Normal-700.png)
- Stack, Caston (from COVID-701.png to COVID-800.png & from Normal-701.png to Normal-800.png)
- Terry, Harrison (from COVID-801.png to COVID-900.png & from Normal-801.png to Normal-900.png)
- Schaeffer, Alex (from COVID-901.png to COVID-1000.png & from Normal-901.png to Normal-1000.png)

Group # 3: X-ray Images 1001-1500

- Lim, Celine (from COVID-1001.png to COVID-1100.png & from Normal-1001.png to Normal-1100.png)
- Reesman, Grace (from COVID-1101.png to COVID-1200.png & from Normal-1101.png to Normal-1200.png)
- Rosenblatt, Jennifer (from COVID-1201.png to COVID-1300.png & from Normal-1201.png to Normal-1300.png)
- Thomas, Micheal (from COVID-1301.png to COVID-1400.png & from Normal-1301.png to Normal-1400.png)
- Aljofei, Maha (from COVID-1401.png to COVID-1500.png & from Normal-1401.png to Normal-1500.png)

2. Use and modify the code “COVID_PCA_project.m” to create the 3-dim PCA clusters and other supporting images and statistics (e.g. distributions) to compare COVID versus Normal cases.


×10^-3
9
8
7
6
5
4
3
2
1
0
Density
50
100
150
200
Data
avgN data
fit 1


0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
Density
60
80
100
120
140
160
180
200
220
Data
avgC data
fit 1

Important note: you will get no credit or minimum credit if you just copy and paste images without providing any explanation of your observations. you must articulate your observations and characterize any differences that you note between covid and normal data.

Watch Part (a) video to learn how:

Math 5590:
https://umkc.instructure.com/courses/62852/modules/items/966719

Math 490:
https://umkc.instructure.com/courses/67839/modules/items/966994

Step 3: Data Extraction for Predictive Modeling

1. Use the code "CRTNNW.m" to extract new variables and data for predictive modeling done by machine learning methods.

Watch Part (b) video to learn how:

Math 5590:
https://umkc.instructure.com/courses/62852/modules/items/966720

Math 490:
https://umkc.instructure.com/courses/67839/modules/items/966995

Step 4: Decision Tree & Neural Network Modeling

Use the data obtained in Step 3 and the IBM SPSS 26 available through UMKC remotelab to build the following two models

a) Decision tree model (classification & regression tree)
b) Neural network model (multilayer perceptron & Radial Basis Function)

Watch Part (c) video to learn how:

Math 5590:
https://umkc.instructure.com/courses/62852/modules/items/966721

Math 490:
https://umkc.instructure.com/courses/67839/modules/items/966996

See also https://www.youtube.com/watch?v=EoDMyjIKpFs

Items to be submitted by each student:

1. **PowerPoint slides of each student's individual work**
2. **Excel file obtained by in Step 3**

Item to be submitted by each group:

1. **PowerPoint slides of the entire group, representing the best predictive models and the best PCA**

Important notes:

1. In your slides, you should provide a brief description of each of the above-mentioned models (see the attached PDF files for more information)
2. You will get no credit or minimum credit if you just copy and paste figures and tables without providing any explanation of your observations.
3. You should provide the pros and cons of each method and limitation of the study.
4. Suppose that you would like to implement this method using AI, in a few sentences explain how this could be done.

References

**COVID-19 data:

COVID data are collected from different publicly accessible dataset, online sources and published papers.

-2473 CXR images are collected from padchest dataset [1].

-183 CXR images from a Germany medical school[2].

-559 CXR image from SIRM, Github, Kaggle & Tweeter [3,4,5,6]

-400 CXR images from another Github source[7].

***Normal images:

10192 Normal data are collected from from three different dataset.
-8851 RSNA [8]
-1341 Kaggle [9]

***Lung opacity images:

6012 Lung opacity CXR images are collected from Radiological Society of North America (RSNA) CXR dataset [8]

[1]https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711
[2]https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png
[3]https://sirm.org/category/senza-categoria/covid-19/
[4]https://eurorad.org
[5]https://github.com/ieee8023/covid-chestxray-dataset
[6]https://figshare.com/articles/COVID-19_Chest_X-Ray_Image_Repository/12580328
[7]https://github.com/armiro/COVID-CXNet
[8]https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data
[9] https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia