Clock Drawing Test Project Paper

Table of Contents

Abstract

According to the World Health Organization, “There are more than 55 million individuals worldwide that are living with Dementia” (World Health Organization). The Clock Drawing Test is a simple test that screens for Dementia in which a patient is given a pre-drawn circle and is asked to create a clock within the circle showing the time as ten minutes past eleven. The patient’s clock is scored on a binary basis (either being at High-Risk for Dementia (a score of 0) or not being at High-Risk for Dementia (a score of 2)) manually by psychologists. The Mini-Cog test is another tool utilized to screen individuals for Dementia that involves a scored Clock Drawing Test and a 3-item recall test for memory (Mini-Cog). According to the Center for Disease Control, "Chronic Obstructive Pulmonary Disorder (COPD) refers to a group of diseases that cause airflow blockage and breathing-related problems” (Center). When someone has COPD, they are at an extremely elevated risk for Dementia (Faris). People with COPD have lower oxygen levels while having higher carbon dioxide levels which causes harm to the brain (Faris). COPDGene aims to find out why an individual is more likely to develop COPD than someone else (COPDGene). The COPDGene Study has collected data from over 10,000 participants that include their scores on the Clock Drawing Test along with their scores on the Mini-cog test (COPDGene); this data was in the Clock Drawing Test Project. A previous researcher has created a machine learning algorithm that can manually score the Clock Drawing Tests (either being at High-Risk for Dementia or not at High-Risk for Dementia). We will complete an in-depth study on the High-Risk for Dementia Clock Drawing Test Images. Currently, an individual’s score on the Clock Drawing Test is only an early indicator for dementia; not specifying which type of

dementia, or letting an individual know if they are at High-Risk for any other diseases linked to Dementia such as (COPD). There are 100 different types of Dementia (Adi - types of dementia). The Clock Drawing Test Project is a Machine Learning project designed to perform an in-depth study on the High-Risk for Dementia Clock Drawing Test images by pixelating the images to be put into a Convolution Neural Network to then perform Clustering Algorithms (K-Means and K-Medoids) on the convoluted pixelated image data. Furthermore, we will investigate the clusters obtained from the Clustering Algorithms to determine if the patterns within the clocks obtained by the Clustering Algorithms could be an early indicator for COPD outcomes. The importance of the Clock Drawing Test Project is that it could potentially allow for the Clock Drawing Test to not only be an early indicator of Dementia but also an early indicator of COPD.

Chapter 1: Introduction

"One in nine people age 65 and older (11.3%) has Alzheimer's dementia" (Facts and figures). "More than 6 million Americans are living with Alzheimer's. By 2050, this number is projected to rise to nearly 13 million people'' (Facts and Figures). This is a significant number of people that could be living with dementia by 2050. There are several studies suggesting the link between COPD and cognitive impairment such as the Pulmonology Advisor along with several academic journals on the topic. Of people who are diagnosed with COPD, 52% alongside COPD have a cognitive impairment (Rodriguez). More than 15 Million individuals in the United States are affected by COPD (Center). "More than 150,000 Americans die of COPD each year-that is 1 death every 4 minutes'' (Center). "COPD refers to a group of diseases that cause airflow

blockage and breathing-related problems'' (Center). Currently, COPD does not have a cure, only treatment plans that can help patients have a better quality of life. Both Dementia and COPD are terrible chronic conditions that affect millions of Americans each year. “COPD is a chronic and systemic inflammatory disease, and the inflammatory markers have been associated with cognitive impairment and dementia. COPD is also associated with an increased risk for cardiovascular disease (Liao).” Most of the people in the population we are working with have some type of COPD. We are looking for signs of worsening COPD. Thus, if we could diagnose people with dementia earlier then we could possibly diagnose people with COPD sooner since the link is prevalent.

As the elderly population is increasing rapidly so are the cases of cognitive impairment (Rodriguez). There is a test that screens for dementia called the Clock-Drawing test which is a simple test that involves the person being screened for dementia to create a clock with hands drawn to a specific time while being given a pre-drawn circle (Esther). The Clock-Drawing Test was created by Sir Henry Head to screen for signs of neurological problems (Alzheimer's and other types of dementia) (Esther). The Clock Drawing Test Is a flag for (elevated risk of) dementia. According to the Alzheimer’s Association, Alzheimer’s is the most prevalent type of dementia (Types). Furthermore, according to the Alzheimer’s Association, there are 11 different types (subtypes) of dementia. There is believed to be a plausible link between the errors on the clock drawing test to specific issues which could in turn be a plausible link between specific types of errors on the Clock Drawing Test and specific types of dementia. Rouleau is well known for describing the most popular used system of Clock Drawing test errors (Eknoyan). He was able to create five different categories of errors alongside the size of the clock listed as follows:

1.) Graphic Difficulties

2.) Stimulus-Bound Response

3.) Conceptual Deficit

4.) Spatial and/or Planning Deficit

5.) Perseveration

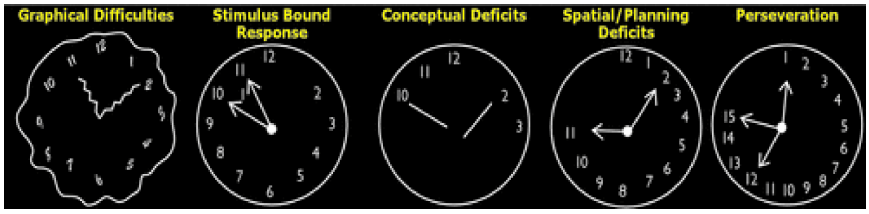

Figure 1- Clock Drawing Test Errors

Figure 1 (shown above) is a visual representation of the five different categories of errors that are associated with the Clock Drawing Test. Note that if the patient is given a pre-drawn circle, they may not show any issues with Graphical Difficulties as they do not have to draw the clock circle. The variations (different errors, different categories) exhibited in the Clock Drawing Tests may be linked to other conditions, such as COPD.

What is the Clock Drawing Test Actually Testing?

The Clock Drawing Test is a very useful and simple way to see if an individual is at high risk for Dementia. Several different cognitive functions are being tested by the Clock Drawing Test:

1. Verbal Understanding: Verbal understanding is being tested by the Clock Drawing Test as the test-taker is intaking words from the proctor and then performs actions

2. Visual Memory: Visual Memory is tested during the CDT as the test-taker must be able to recreate a clock based on a memory of what a clock looks like; individuals who are at high risk for Dementia often have a hard time remembering simple things such as what does a clock look like

3. Planning and Understanding: Planning and Understanding are tested by the CDT since to draw a correct clock an individual must have a plan (number in the correct order and then draw the hand to the correct time) and understand how to execute the plan

4. Abstract Thinking: The CDT evaluates Abstract thinking as the test-taker must abstractly think about the concept of how to correctly draw the hands to a specific time. (DementiaCareCentral).

All of the different things the Clock Drawing Test is evaluating are very important to evaluate and are related to screening the patient to see if they are at high risk for Dementia.

Chapter 2: Data

COPDGene Data

"The COPDGene Study is one of the largest studies ever to investigate the underlying genetic factors of Chronic Obstructive Pulmonary Disease or COPD" (Home COPDGene). The Clock Drawing Test Project utilizes COPDGene Data. This particular dataset has 5 variables of interest (SID, FEV1, Final GOLD Baseline, Self-Reported COPD, and Adjusted Clock Drawing Test Score) along with 1,717 study participants (data instances). The data was collected by COPDGene investigators from the University of Iowa. Now, the SID variable is a way to identify a particular patient while ensuring the privacy of the patient. The FEV1 variable is a measurement of the amount of air a patient can express from their lungs in a second. Furthermore, the COPD variable is a self-reported variable by the patients of whether or not they have COPD (1=Yes COPD, 0=No COPD). The Adjusted Clock Drawing Test Score variable is the patient's score on the Clocks Drawing Test (0=High Risk for Dementia, 2=NOT High Risk for Dementia). Finally, the Final Gold Baseline Variable is a measure of the disease severity of COPD. GOLD is standardized and well-known. Participants without spirometric evidence of airflow obstruction ($FEV_1/FVC \geq 0.70$ and $FEV_1 \geq 80\%$ predicted) were classified as GOLD 0. Subjects with $FEV_1/FVC \geq 0.70$ and FEV_1 <80% predicted were classified as Preserved Ratio Impaired Spirometry (PRISm) (Wan ES). The subtypes of COPD are discussed further in a published paper by co-investigator Dr. Greg Kinney (Kinney); this paper changed the idea that COPD was just one disease, that there are actually many layers (subtypes, severity) of COPD. related to airway disease Emph Airway & Gas

trapping. The main were Emphysema subtype, airway subtype, Gas-Trapping; these are the subtypes of COPD. These diseases were found using PCA of 26 variables that are related to: Pulmonary function, Inspiratory CT (CT scan of lungs), Expiratory CT, Airway Measurements (Kinney). Because they were found with PCA, the term factor is often used in this manuscript instead of subtype. COPD was grouped as spirometric grades 1-4 based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines (Rabe KF). These unique subtypes of COPD were utilized by co-investigator Dr. Kendra Young to identify patients who are at an increased risk for mortality (Young KA). Thus, the Final GOLD Baseline variable is comparing PRISM vs Control (healthy) vs GOLD 1-4 (COPD).

- Control: (FEV1 >= 80%, FEV1/FVC >= 0.7)
- GOLD 1: (FEV1 >= 80%, FEV1/FVC < 0.7)
- GOLD 2: (50% <= FEV1 < 80%, FEV1/FVC < 0.7)
- GOLD 3: (30% < = FEV1 < 50%, FEV1/FVC < 0.7)
- GOLD 4: (FEV1 < 30%, FEV1/FVC < 0.7)
- Prism: (Preserved Ratio Impaired Spirometry) (FEV1/FVC >= 0.7 but FEV1 < 80%)

The Clock Drawing Test images were one of the two elements from a Mini-Cog Analysis Test that was administered to COPDGene participants. The Mini-Cog test is a 3-minute test that can help increase the detection of cognitive impairment (Mini-Cog). The Mini Cognitive Assessment is a very brief test taking around three minutes for a patient to complete (DementiaCareCentral). . The Mini Cognitive Assessment consists of the following two testing elements: a test where the patient recalls three words and an easily scored Clock Drawing Test (DementiaCareCentral). The Clock Drawing Portion of the test is the same as described before where the participant is given a predawn circle and asked to recreate a clock at a specific time. Once the patient has completed the Clock Drawing Portion of the Mini Cognitive Assessment the individual is prompted to repeat the three words the test administrator stated at the beginning of the Assessment (DementiaCareCentral). The Mini-Cognition test is scored out of 5 points (DementiaCareCentral). For each word that the patients remember correctly, they score a point; thus, a patient can score up to three points on this portion of the assessment. Two points are given for a correctly drawn clock, and zero points are given for a clock that shows the patient is at high for Dementia (DementiaCareCentral). If the patient scores a total of three points or less they are considered to be at high risk for dementia, and should follow up with the necessary doctor. The current work focuses on the Clock Drawing Test element of the Mini-Cognition test.

COPDGene investigators spent time collecting and analyzing patients' Clock Drawing Tests along with their Mini-Cog Test. The data was comprised of nearly 2,500 different individuals. The clocks were scored manually by a team of psychologists based on a set of 54

rules. Currently, the clocks are scored on a binary basis with the patient either being at high risk for dementia or the patient not being at high risk for dementia; this will be the variable named Score in Python. Data is considered categorical when it can be classified based upon certain given labels (such as 0 (high risk for dementia) or 2 (not high risk for dementia)) (Categorical). Hence, by definition, the Score variable is a categorical variable as the different patients can be classified based upon their score of either a 0 (high risk for dementia) or 2 (not high risk for dementia). Out of the COPDGene study participants who took the clock drawing test, only around 565 of the clocks were scored manually by psychologists and were of sufficient quality to be used for this project. Out of the 565 scored clocks 500 people scored a two on the test (drawing the clock correctly), and 65 different people scored a zero on the test (having drawn the clock not exactly correct). COPDGene obtained the data following all review protocols. Each person consented to participation and was informed on how their data was going to be used.

The pictures of the patient's Clock Drawing Tests which consists of 565 PDF and JPEG images will be pixelated via Python (i.e., creating another dataset). Initially, we had 623 clock images. However, some of the clock images were corrupt files. A few files were deleted due to cropping issues. Once we have pixelated the image data, we will create a Convolutional Neural Network to Convolute the pixelated image data as a way to preprocess (clean) the pixelated image data.

Chapter 3: Foundational Work for Current Investigation

Convolutional Neural Networks (CNNs) are commonly implemented when trying to solve pattern recognition problems using image data (O'Shea). One example where CNNs are used is in self-driving cars utilizing Convolutional Neural Networks to help with detecting objects (Lecture 15 MATH 5388). Implementing a Convolutional Neural Network is a way of beginning to work with an Artificial Neural Network (ANN) (O'Shea). The first step of a CNN in our context is inputting the pixels of the Clock Drawing Test images via arrays into the *input layer* of the Neural Network (Lecture 15 MATH 5388, Dr Yaning Liu). The final layer of a Convolutional Neural Network is the *fully connected layer (Output)*.

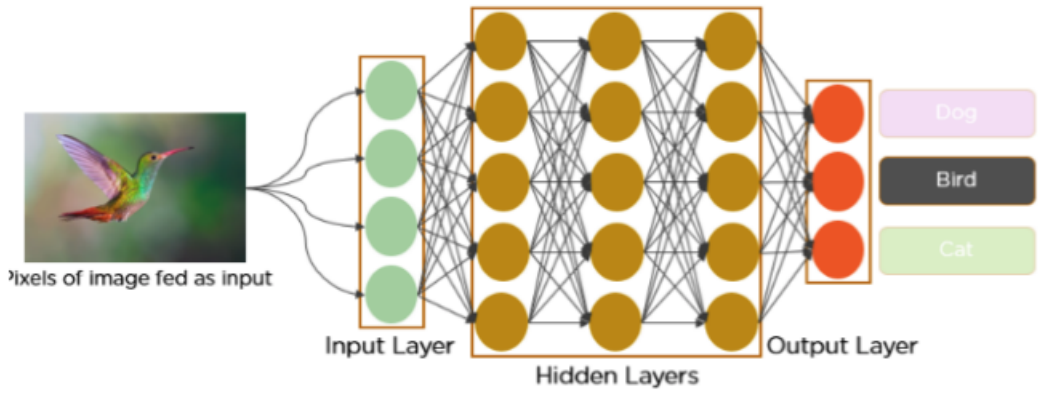

*Figure 2- Convolutional Neural Network Used to Identify an Image of a Hummingbird (*Lecture 15 MATH 5388, Dr Yaning Liu*)*

Note that a convolution is a mathematical operation (Lecture 15 MATH 5388, Dr Yaning Liu). Thus, when a Convolutional Neural Network is implemented to solve a machine learning problem the neural network is performing the mathematical operation called a convolution

(Lecture 15 MATH 5388, Dr Yaning Liu). We will implement a Convolutional Neural Network on the pixelated Clock Drawing images to obtain a cleaned/preprocessed version of the pixelated images by implementing a CNN with the following hidden layers; we will have the input layer and output layer:

- Convolutional Layer (An operation that passes filters over the image and transforms it into a Feature Map)
    - Can detect increasingly complex patterns in the pixel data

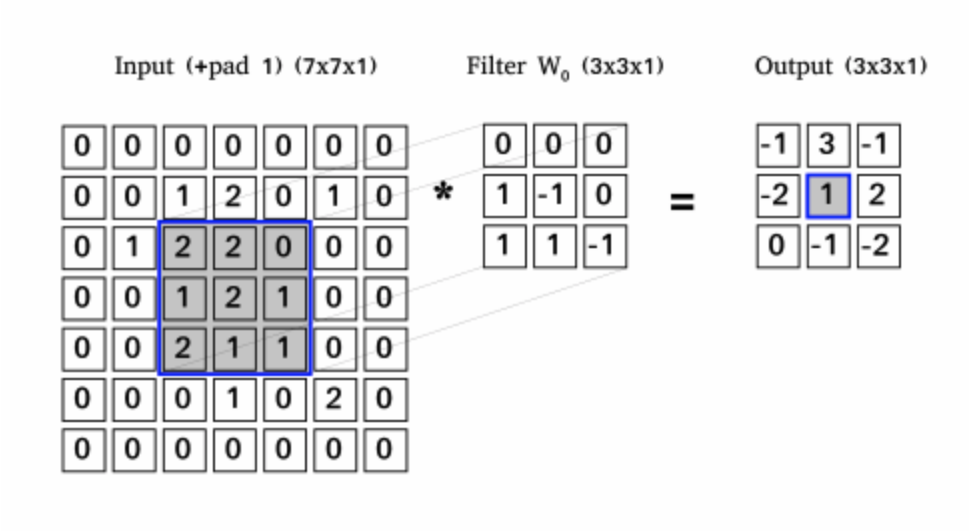

Figure 3- Visual Example of a Convolution (Navigation)

- Pooling Layer (Creates a second Feature Map by taking sections of the original feature map to take the Average or Maximum among the values)
- Flatten Layer (Converts the convoluted data from Matrix form into vector form)
- ReLU Layer
- Sigmoid Layer

Note that the output is not what is important in this instance, we are wanting to extract the convoluted data after it has been flattened (put into vector form) to then be able to perform clustering algorithms (K-Means and K-Medoids).

Previously, a fellow researcher, Dr. Mesbah Najafi, created a CNN-based machine learning algorithm that can manually score the clocks on a binary pass or fail basis (score the clock with either a 0 (High-Risk for Dementia) or 2 (not at High-Risk for Dementia)). By showing that machine learning algorithms could successfully classify the clocks, Dr. Najafi confirmed the potential for the algorithms to go further and uncover hidden patterns and clusters. These hidden classifications may, in turn, be predictors of other diseases such as COPD, which is the primary objective in our current follow-up study.

Correct Classification was achieved by creating a *Convolutional Neural Network (CNN)* model that was run on the images of the patient's Clock Drawing Tests. Then the performance of the *Convolutional Neural Network (CNN)* was assessed by obtaining the *accuracy* and *F1-score*. In the study approximately 25,000 patients had taken the Clock Drawing Test (CDT). However, as in our follow-up, many of the completed Clock Drawing Tests were not yet scored. Thus, the study was conducted on approximately 630 completed Clock Drawing Test images that were obtained from COPDGene investigators at the University of Iowa (where the Mini Cog analysis was conducted). These 630 Clock Drawing Test image files were in both PDF and JPEG format. Once the unscored CDTs were removed and the images were preprocessed, 496 CDT

images remained; including 64 abnormal scored CDT images (score of 0) and 432 normal scored CDT images (score of 2).

Specifically, the classification used the following steps and was performed with Python:

Data Pre-processing

1. The PDF Clock Drawing Test images were pixelated, cropped, grayscaled, resized, and then converted into a 288 by 288 matrix.
2. Similarly, the JPG Clock Drawing Test images were pixelated, cropped, grayscaled, resized, and then transformed into a 288 by 288 matrix via Python.
3. Once the PDF & JPEG images were translated into data matrices, the data was preprocessed via image rotation and noise removal. After the data was preprocessed, the PDF and JPEG matrices were concatenated together into one data matrix.
4. The labels for different Clock Drawing Test Scores were generated.
5. The Data was normalized and then split into test and training sets at a 80/20 ratio.

CNN Modeling

1. Training/Testing Loss Charts
2. Training/Testing Accuracy Charts
3. Obtained Classification Report
4. Randomized Search for CNN Hyperparameter Tuning
5. CNN Optimization-CPU Server
6. Ran the Baseline CNN Model

7. CNN- Dropout Effect (10, 25, & 40%)
8. CNN-Effect of Image Flipping
9. Flipped Abnormal Training Clocks
10. CNN Model- Effect of Resampling Method (Borderline SMOTE)
11. Generated new abnormal training clocks with Borderline SMOTE
12. CNN-Optimization Objective Function for final results

The best model was able to successfully classify with accuracy of 91%. The significant accuracy level demonstrated how the clocks possess sufficient information potentially related to participants' underlying conditions such as dementia.

There has been additional work completed by other researchers on the idea that the Clock Drawing Test can identify Subtypes of dementia. An article titled "Identifying Dementia Subtypes with the Clock Drawing Test '' addresses the research question Can the Clock Drawing Test differentiate between different forms of dementia, specifically Alzheimer's dementia and all other types of Dementia by conducting a Literature review (Rapposelli). The article states that with frontotemporal dementia, significant differences were found within Alzheimer Dementia patients Clock Drawing Test Scores aside from patients with other forms of dementia (Rapposelli). Subtypes were identified, but in this work machine learning was not used. I plan to improve upon this work by implementing machine learning to identify substructure within the clocks. If simpler methods found subtypes of dementia, we hypothesize that if machine learning could be implemented to not only differentiate between the different subtypes of dementia- but to find meaningful structures or patterns related to dementia and related disorders. Specifically clustering algorithms will be used to accomplish this pattern identification.

Chapter 4: Current Investigation

The Clock Drawing Test project will be completed by utilizing Machine Learning in the python environment. Previous research has produced a machine learning algorithm using Convolutional Neural Networks and Transfer Learning to be able to score the clocks on a pass/fail basis. I will be building on this work by applying machine learning to unsupervised data, focusing only on the high-risk dementia clocks (there is no pass/fail criteria to supervise the classification).

The main research question that I will be focusing on is the following: Can we create a machine learning algorithm (in particular implementing the K-Means and K-Medoids algorithm) that can find patterns among the High-Risk for Dementia Clock Drawing Test images in which we will investigate the clusters to determine if the patterns within the pixelated clocks could be an early indicator for Chronic Obstructive Pulmonary Disease (COPD) as defined in multiple ways.

Chapter 5: Machine Learning Background

In this chapter we introduce the key details needed for context on the results. We begin with an overview of the learning goal of our model training, and then provide the conceptual detail for the models.

What is Supervised Learning and Unsupervised Learning and How are they different?

Supervised and Unsupervised learning are two unique machine learning procedures. “Supervised Machine Learning is a machine learning method in which models are trained using labeled data (Supervised).” The supervised learning technique requires supervision in order for the model to be trained (Supervised). Supervised Machine Learning can be used in the classic problem of classification (Supervised). Previous work by Dr. Najafi used the supervised learning technique as the goal was to show there existed sufficient information in the clocks for accurate classification. In that phase of the project, supervised learning was used to classify the clocks on a pass/fail basis.

Unsupervised learning is a machine learning procedure in which the input data is unlabeled and patterns are inferred from the unlabeled data (Supervised). “The goal of unsupervised learning is to find the structure and patterns from the input data” (Supervised). “Unsupervised learning does not need any supervision. Instead, it finds patterns from the data on its own” (Supervised). In this current phase of the Clock Drawing Project we build upon Dr. Najafi’s results and will utilize unsupervised learning to try to find patterns within the incorrect

clocks along with the Mini-Cog data. If hidden patterns exist, we will test them for associations with COPD phenotypes.

Algorithms that can Categorize Unsupervised Data into Subtypes

For completeness in our discussion, there are two types of unsupervised algorithms: (1) Association Algorithms and (2) Clustering Algorithms (Roy). In the current work, only clustering algorithms are needed.

"Clustering Algorithms are a type of unsupervised learning algorithm where we try to find the features in our dataset on the similarity of which we can group or cluster the data together (Roy)." One type of clustering algorithm is the Partitioning Method (Roy). "The partitioning methods usually operate using a central tendency to represent a structure. They use a distance-based approach to create the clusters (Roy)." The most established, and thus the ones we explore, Partitioning algorithms are: (1) K-Means and (2) K-Medoids (Roy).

These two partitioning algorithms are described in more detail next.

K-means Algorithm

The goal of the K-Means Clustering Algorithm is to partition the unsupervised data so that the machine learning model created could train itself to divide the pixelated High-Risk for Dementia Clock Drawing Test images into individual groups based on the most similar

characteristics among them. “The K-means Algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group” (Dabbura). We implemented the K-Means algorithm to partition our unsupervised data set into K pre-defined unique non-overlapping subgroups (clusters). An ideal results when the K-Means Algorithm is implemented is shown below by Figure 3:

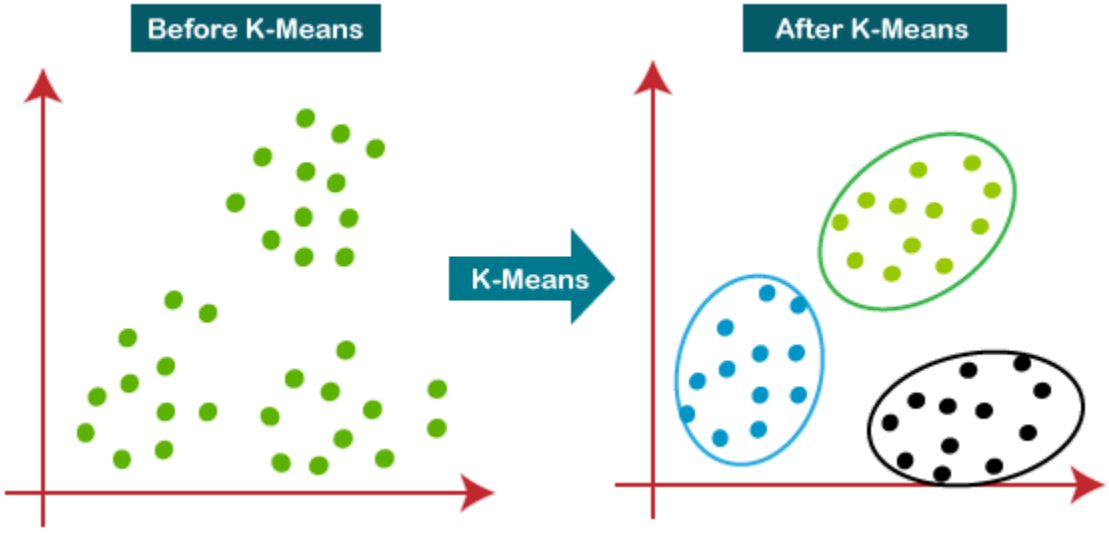

Figure 4 – Visual of K-Means Algorithm (Srivastava)

The K-means algorithm works as follows (using the description from Dabbura):

1. Specify the number of K clusters.

2. Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.

3. Keep iterating until there is no change to the centroids. i.e. assignment of data points to clusters isn't changing.

4. Compute the sum of the squared distance between data points and all centroids

5. Assign each data point to the closest cluster (centroid).

6. Compute the centroids for the clusters by taking the average of all the data points that belong to each cluster

In order to solve the problem at hand the K-Means algorithm uses the Expectation-Maximization approach (Dabbura). The Expectation step is implemented such that the K-Means algorithm assigns each data point to the cluster it is closest to (Dabbura). The centroids of the individual clusters are computed in the Maximization step (Dabbura).
The objective function for the K-Means algorithm is as follows:

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} ||x^{i} - \mu_{k}||^{2}$$

such that $w_{ik} = 1$ when x^{i} is a member of cluster k; otherwise $w_{ik} = 0$ (Dabbura). This is a two-part minimization problem (Dabbura). First, J is minimized in respect to μ_{k} with w_{ik} fixed. The Expectation step is where we take the derivative of J in respect to w_{ik} while updating cluster assignments (Dabbura). The Minimization step is completed by taking the derivative of J

in respect to μ_k(Dabbura). Then the centroids are recomputed after clusters are assigned from the previous step (Dabburara).

The equation for the E- step is the following:

$$\frac{\delta J}{\delta w_{ik}} = \sum_{i=1}^{m}\sum_{k=1}^{K} w_{ik}\left\|x^i - \mu_k\right\|^2$$

$$\Rightarrow w_{ik} = \{1,\ if\ k = argmin_j\ ||x^i - \mu_k||\ 0,\ otherwise$$

(Dabbura).

The Expectation step is assigning each data point x^i to the nearest cluster based upon the data point's sum of squared distances from the centroid of the cluster (Dabbura).

The Maximization step is listed as follows:

$$\frac{\delta J}{\delta \mu_k} = 2\sum_{i=1}^{m} w_{ik}\left(x^i - \mu_k\right) = 0$$

$$\mu_k = \frac{\sum_{i=1}^{m} w_{ik}x^i}{\sum_{i=1}^{m} w_{ik}}$$

(Dabbura).

The Maximization step recomputes the centroids after the reassignment of the cluster from the expectation step (Dabbura). When using clustering algorithms such as the K-Means algorithm that use measurements based on distance to find similarities in the data it is good practice for the data to be standardized such that the data has a mean of zero with a standard deviation of 1 due to the fact that most of the time features in the data do not have the same

units of measurements (Dabbura). Note that since the K-Means algorithm is of the iterative nature and the centroids are randomly initialized at the beginning of the algorithm, different initialization of centroids could result in different clusters as the algorithm may not be converging to the global optimum as it is stuck in the local optimum (Dabbura). To solve this problem we can use different initializations of the centroids. Then we can pick the result that produces the lowest sum of squared distance (Dabbura).

Before implementing a specific clustering algorithm, in this case the K-means algorithm, that will partition the unsupervised data so that the machine learning model can learn how to divide the clocks with a score of 0 (at High-Risk for Dementia) into group based upon the most similar characteristics among them we want to make sure the algorithm we are looking at implementing has been successful in public health situations before. If it has not been too successful, we may want to consider another algorithm that can partition the unsupervised data.

In "Cluster Analysis and its Application to healthcare claims data: a study of end-stage renal disease patients who initiated hemodialysis" the "K-means method appeared to be the most appropriate in healthcare claims data with highly skewed cost information" (Liao). The authors of this academic journal claim that the cluster analysis method hasn't been a popular statistical method used in the public health sector when big healthcare claims data with skewed cost information is involved (Liao).

As a second example, the K-Means Clustering Algorithm was successfully used in a Public Health setting I reference the reader to "An Enhanced K-Means Clustering Algorithm for Pattern

Discovery in Healthcare Data '' (Ramzi). This work successfully implements an enhanced K-Means algorithm titled the G-means algorithm for data mining healthcare data (Ramzi). The G-Means Algorithm is another clustering Algorithm that could be used to cluster the pixelated High-Risk for Dementia Clock Drawing Test Images.

Overall, the K-Means algorithm has been shown to be effective as a good clustering algorithm in public health. Thus, it is reasonable to test if it can be implemented into my machine learning approach to find clusters within the High-Risk for Dementia Clock Drawing Test images.

K-Medoid Algorithm

The second partitioning method algorithm that we will implement to partition the unsupervised data so that the model created can train itself to put the High-Risk for Dementia Clock Drawing Test images into groups based on the most similar characteristics among them is the K-Medoid algorithm. The K-Medoid Algorithm is a clustering algorithm that highly resembles the K-Means algorithm. The K-Medoids clustering algorithm is an unsupervised machine learning algorithm (Shiledarbaxi). Similar to the K-Means algorithm the K-Medoids algorithm partitions data sets into a specific number of K clusters or groups (Gama). “In K-Medoids clustering, each cluster is represented by one of the data points in the cluster (Gama).” These specific points are called cluster medoids (Gama). “The term medoid refers to an object within a cluster for which average dissimilarity between it and all the other members of the cluster is minimal (Gama).” The medoids correspond to the point in the cluster that is the most centrally

located (Gama). The K-Medoids algorithm differs from the K-Means algorithm in the way that the algorithm selects the centers of the clusters (Shiledarbaxi). Recall that the K-Means algorithm takes an average of each data point that belongs to that specific cluster to compute the clusters' centers. Now, the K-Medoids algorithm selects the center of the cluster by choosing a data point to represent the cluster (the data point being called cluster medoids). The K-Medoid algorithm is going to be less sensitive than the K-Means algorithm to noise and outliers since the K-Medoid algorithm selects the centers of the clusters to be the medoids rather than computing an average as in the K-means algorithm (Gama). The fact the K-Medoid algorithm is very similar to the K-Means algorithm, but is less sensitive to noise and outliers, the K-Medoid may be a better algorithm to begin with than the K-Means algorithm. Just like the K-Means algorithm, the K-Medoid algorithm requires that the user must specify the number of K clusters. We must determine the optimal number of K clusters to specify; we can use the silhouette method to do this (Gama). The PAM algorithm (Partitioning Around Medoids) is the most popular K-medoids clustering method (Gama). To summarize the method, a flowchart of the K-Medoid Algorithm is shown below:

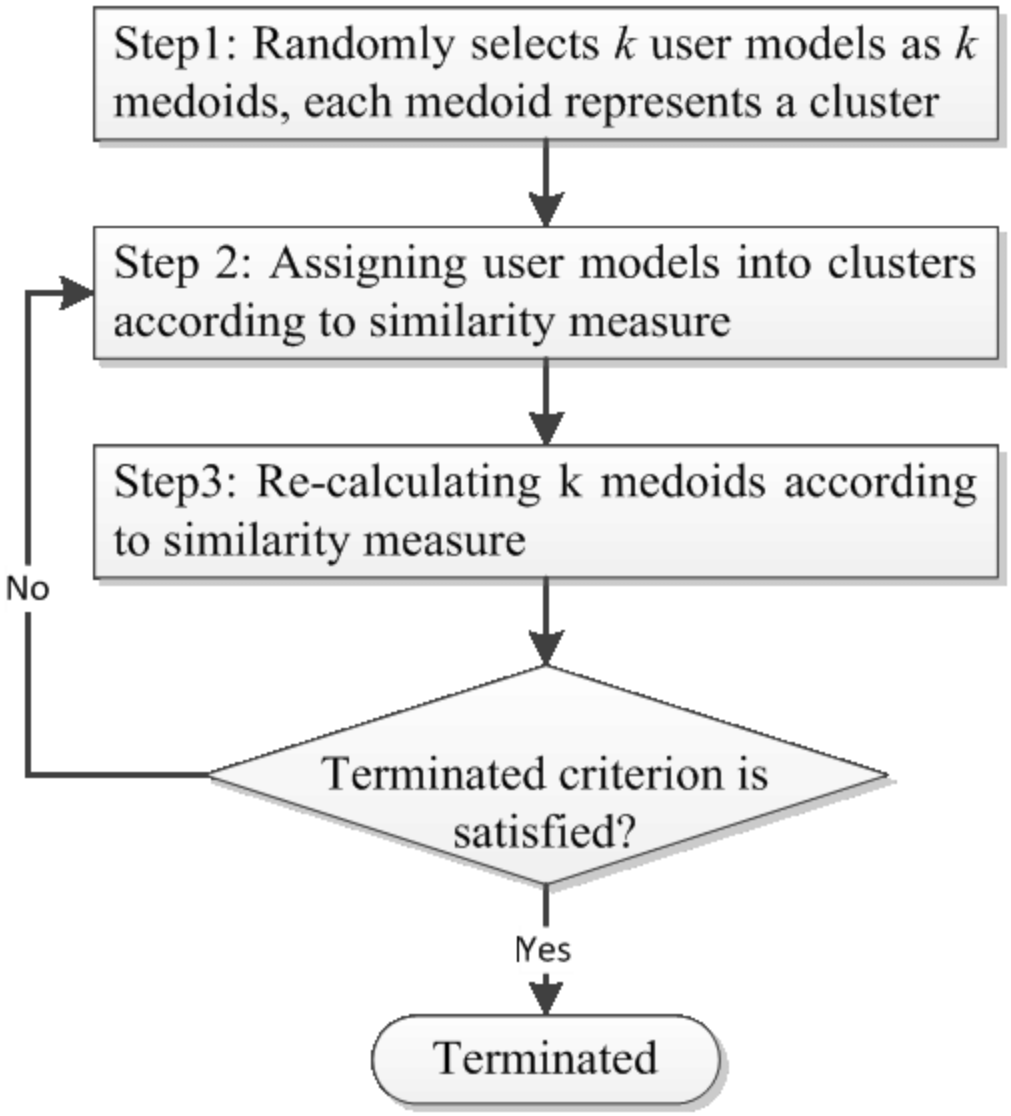

Figure 5 – Flowchart of K-Medoid Algorithm (Flow chart)

We will now provide a section on the K-Nearest Neighbors Algorithm (Supervised Learning Technique) as we will implement the K-Nearest Neighbors to solve several classification problems that will be shown below such as solving the classification problem of whether a patient has COPD or not.

K-Nearest Neighbors Algorithm

K-Nearest Neighbors is a commonly used algorithm for classification machine learning problems (Harrison). The *K-Nearest Neighbors* algorithm is an example of a supervised machine (Harrison). We recall that a *supervised* machine learning algorithm is an algorithm that learns off of labeled input data (Harrison). Note that a *classification* problem in machine learning will output discrete values such as 0 (high-risk for dementia) or 2 (NOT at high-risk for dementia) in our case (Harrison). In a typical classification problem, we will have a predictor variable (or several predictor variables) and a label (Harrison). The K-Nearest Neighbors algorithm takes on the assumption that data instances that are similar are in existence near each other.

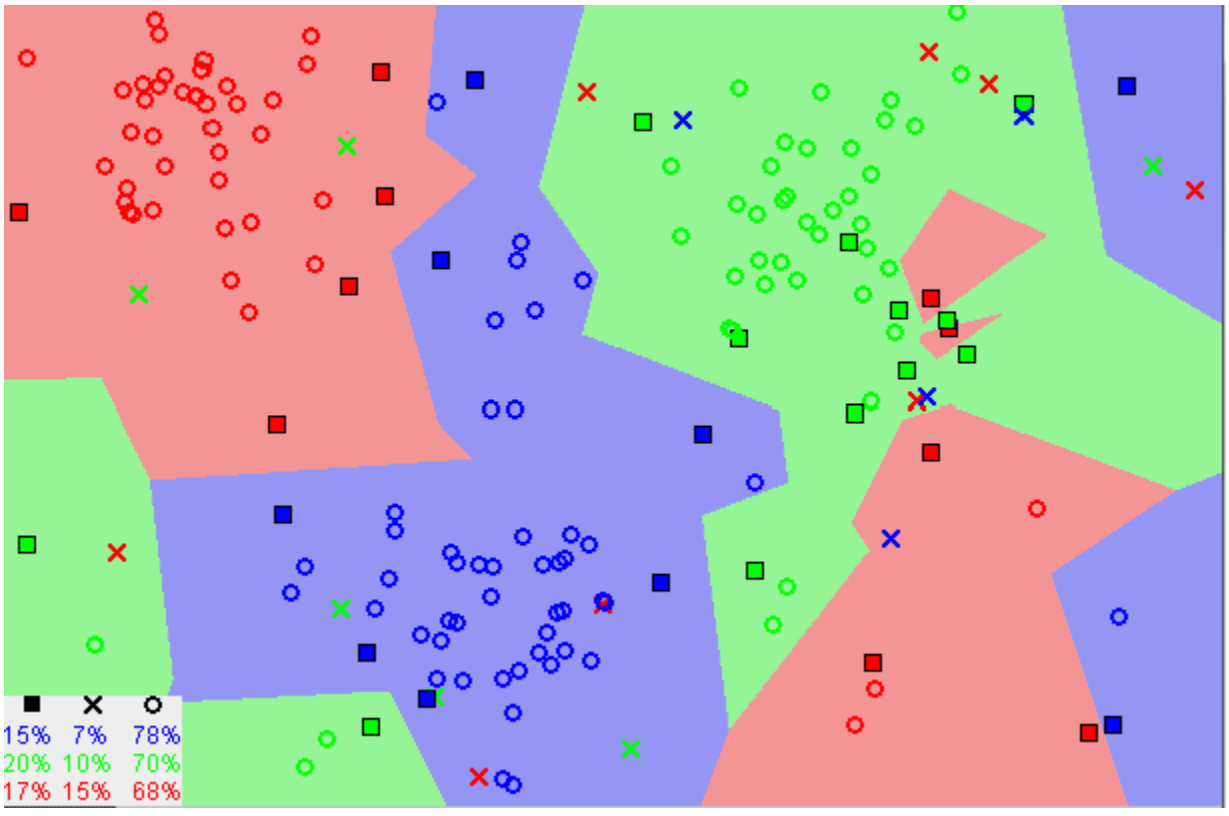

Figure 6– Visual Representation of Similar Data Instances existing nearby

(Harrison)

In Figure 6 (shown above) we can see that data instances that are of the same kind exist near each other. A K-Nearest Neighbors model will only sort the training data (Lec 4 Machine Learning Methods MATH 5388, Dr. Yaning Liu). The K-Nearest Neighbors algorithm considers the data instances that are the most near to the data point the algorithm is making a prediction for; hence the idea of "nearest neighbors" (Lec 4 Machine Learning Methods MATH 5388, Dr. Yaning Liu).

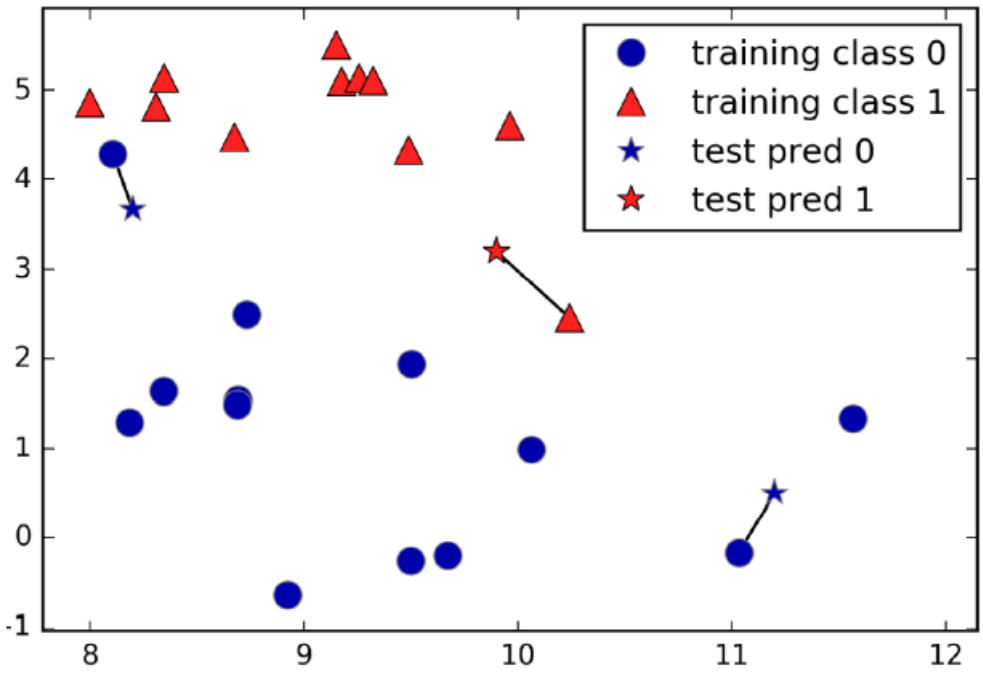

Figure 7- Visual Example of Implementing the K-Nearest Neighbors Algorithm with Precisely One Nearest Neighbor (Lec 4 Machine Learning Methods MATH 5388, Dr. Yaning Liu)

We can implement the K-Nearest Neighbors Algorithm on our data frames that obtain the 5 variables of interest (SID, FEV1, GOLD, COPD, and Score) to predict the Adjusted Clock Drawing Test Score.

Chapter 6: Final Data Processing

We complete data processing by implementing a Convolutional Neural Network on the pixelated Clock Drawing Test image data consisting of the following layers:

- Convolutinal
- Flatten
- ReLU
- Sigmoid

The Convolutional layer will process the pixelated Clock Drawing Test image data by finding increasing complex patterns (the most important parts) within the image. We will then extract the data after it has been convoluted and flattened (vectorized) to then perform clustering algorithms on the convoluted pixelated Clock Drawing Test image data such as the K-Means and K-Medoids algorithms that were previously described.

We will now pixelate the 565 Clock Drawing Test images via a for loop in Python to then convert them into a data matrix that we will be able to perform many Machine Learning Algorithms on. After the 565 Clock Drawing Test images were pixelated, we ended up with 485 viable Clock Drawing Test images. Some Clock Drawing Test images were deleted due to corrupted files. We also had to remove some “noise” files. There were also 4 images that needed to be rotated, we completed this step via Python. We then selected the pixelated images that correspond to a score of 0 (High-Risk for Dementia); leaving us with 65 pixelated Clock Drawing Test images with an associated Adjusted Clock Drawing Test Score of 0. Once the image files are pixelated, we assign this to be our data matrix that contains the values of each pixelated clock image. We then create the data matrix which will contain the Adjusted Clock Drawing Test Score of each pixelated clock.

Furthermore, we will perform a convolution on the data (the and data matrices) as a Convolutional Neural Network is one of the most popular deep learning techniques for image processing. We created the Training/Testing Loss Charts for the Convolutional Neural Network that was executed on the pixelated clock images with their associated Adjusted Clock Drawing Test Score. Note that the Convolutional Neural Network performed twenty epochs. The obtained Training/Testing Chart can be seen below:

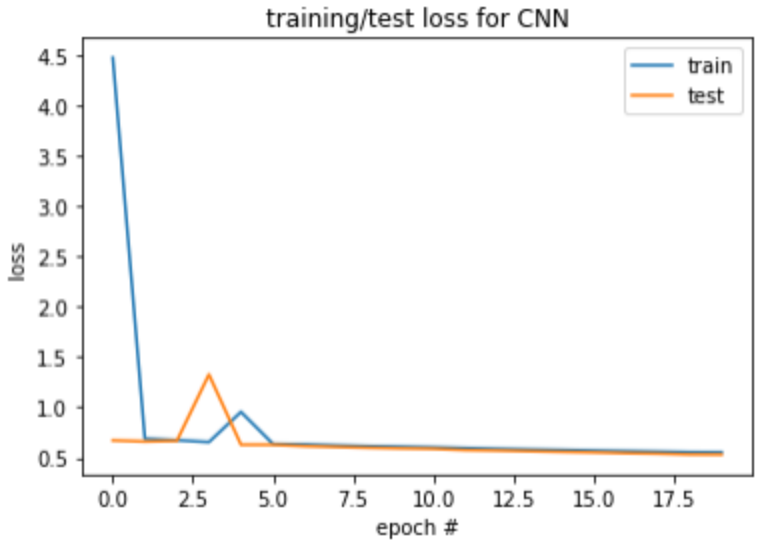

Figure 8- Training/Testing Loss Charts for the Convolutional Neural Network Executed on the Pixelated Clock Images with the Associated Adjusted Clock Drawing Test Scores

We are trying to minimize loss over the number of epochs. Studying Figure 8 we can see that we have achieved that goal with this CNN, which is optimal.

Another way to measure the performance of a Convolutional Neural Network is by the Training/Testing Accuracy Charts. The Training/Testing Accuracy Charts for the Convolutional Neural Network that was executed on the pixelated clock images with the associated adjusted Clock Drawing Test Scores can be seen below:

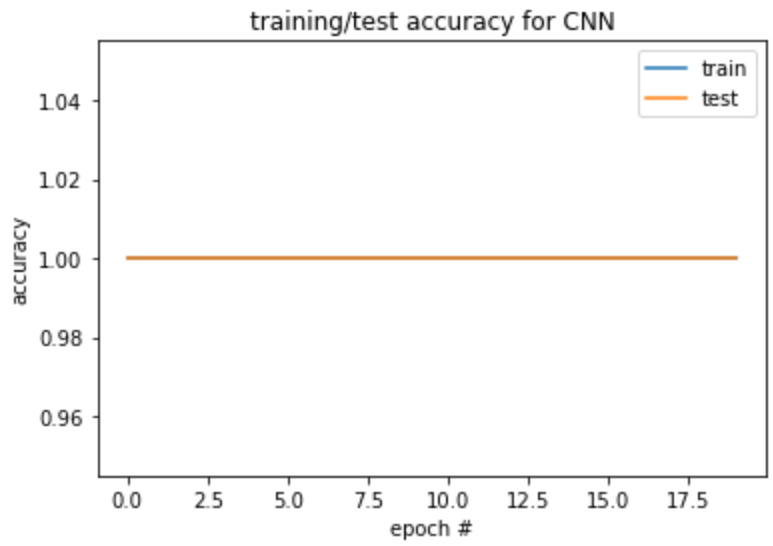

Figure 9- Training/Testing Accuracy Charts for the Convolutional Neural Network Executed on the Pixelated Clock Images with the Associated Adjusted Clock Drawing Test Scores

The higher the accuracy value is the better. We have obtained a perfect accuracy score here, which makes sense as we are working with only the High-Risk for Dementia Clocks (score of 0).

Then the best parameters to create the most optimal Convolutional Neural Network were found via a Turner Search. The summary of the best model parameters can be seen below:

Layer (type)	Output Shape	Param #
conv_layer_1 (Conv2D)	(None, 286, 286, 32)	320
maxpool_1 (MaxPooling2D)	(None, 143, 143, 32)	0
flatten_2 (Flatten)	(None, 654368)	0
dense_layer_1 (Dense)	(None, 32)	20939808
dense_2 (Dense)	(None, 2)	66

Total params: 20,940,194
Trainable params: 20,940,194
Non-trainable params: 0

Figure 10- Summary of the Best Parameters obtained from the Turner Search to Create the Most Optimal Convolutional Neural Network.

Chapter 7: Analysis & Results

The results will be presented in the following subsections:

(1) Preliminary Results

(2) Analysis of Association Between Scored Clocks and COPD Measures

(3) Analysis of the Factor Variables and the Adjusted Clocked Drawing Test Score

(4) K-Means/K-Medoids Clustering Algorithms (Unsupervised) on the pixelated Clock Drawing Test image data

(5) Testing COPD Subtypes versus New Clusters

Note that as there were a number of different methods used, I will provide the method used each time a result is presented.

(1) Preliminary Results

First, we will investigate the summary statistics of the 565 patient's in which their Clock Drawing Test image was successfully pixelated. Note that we are left with 244 patients in which we have their associate Age, Gender, Race, Self-Reported COPD, and Final GOLD Baseline Variable.

Our statistical analysis begins by investigating the continuous variable Age:

	Mean:	Standard Deviation:
Age:	67.84	8.31

Table 1. Summary Statistics for the Age Variable.

From Table 2 we can see that the average age of someone in the study is about 68 having a standard deviation of 8.31.

Our statistical analysis continued as we investigated the categorical variables (Gender, Race, Self-Reported COPD, and Final GOLD Baseline):

	Count:	Percent of Total:
Male:	121	49.59%
Female:	123	50.4%
Non-Hispanic White:	177	72.54%
Non-Hispanic Black:	67	27.46%
Self-Reported COPD:	67	27.56%

Table 2. Summary Statistics for the Gender, Race, and Self-Reported COPD.

Upon inspection of Table 2 we can see that 50.4% of our population are females. We can also see that the majority of our population consists of Non-Hispanic White patients. Only 67 patients reported having COPD.

Continuing on, we will fabricate a frequency table for the Final GOLD Baseline variable (Control, Prism, GOLD 1-4 (COPD)):

	Count:	Percent of the Total:
GOLD 1-4 (COPD):	130	53.28%
Control:	114	46.72
Prism:	0	0%

Table 3. Summary Statistics for the Final GOLD Baseline Variable.

Recall that patients corresponding to a Final GOLD Baseline value of 1-4 are considered to have COPD. We can see from Table 3 that the percentage of people who have COPD based on the Final GOLD Baseline variable is much higher than the Self-Reported COPD variable.

We created Boxplots of the means of the Subtypes of COPD, the Four Factor variables (F1compP2 (Factor 1), F2compP2 (Factor 2), F3compP2 (Factor 3), F4compP2 (Factor 4)) so that we can visually see the means of each factor variable. The Box Plot can be seen as follows:

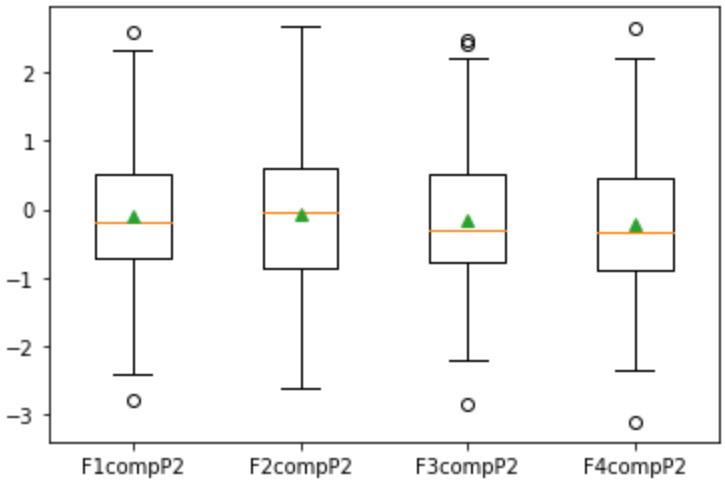

Figure 11- Box Plots of the Means of the Four Factor variables (F1compP2, F2compP2, F3compP2, F4compP2)

Note that the mean of each variable is denoted by the green triangle shown on the Box Plot for each variable. Notice that visually all of the means of the four variables appear fairly close in value. Specifically, the means of the F3compP2 (Factor 3) variable and the F4compP2 (Factor 4) variable appear very very similar in values according to the Box Plots. This would be an inclination we may not obtain a statistically significant value when we perform the One-Way ANOVA on this data. We can see that the F2compP2 (Factor 2) variable appears to obtain the largest mean value among the four factor variables. Also, notice that all the factor variables besides the F2compP2 (Factor 2) variable appear to contain several outliers.

(2) Analysis of Association Between Scored Clocks and COPD Measures

The K-Nearest Neighbors Algorithm was implemented on the dataset to solve the classification problem of whether or not a particular person has COPD or not.

RMSE (Training Data):	0.3074
RMSE (Test):	0.4985
Training Accuracy Score:	0.593
Test Accuracy Score:	-0.132

Table 4. K-Nearest Neighbors Algorithm.

The training accuracy score is good, however the negative accuracy score is concerning.

Next we report the results of the Chi-Squared Test for Independence or Fisher's-Exact Test as appropriate after checking all model assumptions, for the following:

- Adjusted Clock Drawing Test Score by the Self-Reported COPD Score variable
- Adjusted Clock Drawing Test Score by the Final GOLD Baseline variable

Statistical Analysis began by performing the Chi-Squared Test for Independence on the Adjusted Clock Drawing Test Score by the Self-Reported COPD Score variable to determine if a statistically significant relationship exists between the two variables of interest. The results obtained from the Chi-Squared Test for Independence are shown as follows:

Test:	**Degrees of Freedom:**	**Chi-Squared:**	**P-Value:**
Pearson:	1	4.43385	**0.0352329**
Log-Likelihood:	1	4.25976	**0.0390255**

Table 5. Chi-Squared Test for Independence on the Adjusted Clock Drawing Test Score by the Self-Reported COPD Score Variable.

When observing the results obtained from the Chi-Squared Test for Independence we can see that we have obtained some interesting results as the Pearson-Test obtained a statistically significant relationship. Thus, implying that the Adjusted Clock Drawing Test Score variable and the Self-Reported COPD variable have a statistically significant relationship. We can also see that the P-Value obtained from the Log-Likelihood test is statistically significant as well.

Moreover, the Chi-Squared Test for Independence was executed on the Adjusted Clock Drawing Test Score variable by the Final GOLD Baseline variable to determine if a statistically

significant relationship exists among the variables of interest. The results obtained from implementing the Chi-Squared Test for Independence are shown below:

Test:	**Degrees of Freedom:**	**Chi-Squared:**	**P-Value:**
Pearson:	2	2.06987	0.282192
Log-Likelihood:	2	2.06174	0.282827

Table 6. Chi-Squared Test for Independence Performed on the Adjusted Clock Drawing Test Score Variable by the Final GOLD Baseline Variable.

When we observe the results obtained from performing the Chi-Squared Test for Independence on the Adjusted Clock Drawing Test Score variable by the Final GOLD Baseline variable, we can see that we have not obtained statistically significant results as the calculated P-Values are not smaller than the set alpha value of 0.05. Hence, we do not have statistical evidence to conclude that the variables of interest obtain a statistically significant relationship.

Finally, we complete the statistical analysis based on the clock score by performing the K-Nearest Neighbors Algorithm to help solve the classification problem of classifying a patient's Adjusted Clock Drawing Test Score as either High-Risk for Dementia or not at High-Risk for Dementia based upon the FEV1, Final GOLD Baseline, and Self-Reported COPD variables. As a reminder, this was done prior to applying machine learning algorithms to establish a link between the clocks and COPD phenotypes. We can see the calculated Root Mean Squared Error (RMSE) and Accuracy Score for the Training and Test Data as follows:

RMSE (Training Data):	0.2396506466347054
RMSE (Test):	0.5575040903982775
Training Accuracy Score:	0.891
Test Accuracy Score:	0.428

Table 7. K-Nearest Neighbors Algorithm (Classifying Adjusted Clock Drawing Test Score):

Upon observation of the results obtained from implementing the K-Nearest Neighbors Algorithm to help solve the classification problem of a patient's Score on the Clock Drawing Test we can conclude that the KNN ran fairly well as we have a good Training Accuracy Score and an okay Test Accuracy Score.

(3) Analysis of the Factor Variables and the Adjusted Clock Drawing Test Score

With some evidence that the clocks, at least the summary score level of a clock, may be related to COPD, we will steer our statistical analysis to our specific COPD sub diseases of interest (known as the COPD subtypes or factors, such as Emphysema or Airway Disease). That is, we test for associations between our four different factors and the manually assigned score of the clocks. In the following tables Factor 1 (F1compP2) is the Emphysema variable. Now, the Factor 2 variable (F2compP2) is the Airway Factor. Furthermore, Factor 3 (F3compP2) is known as the Gas-Trapping Factor. Factor 4 (F4compP2) is known as the Total Lung Capacity Factor. Smaller values reflect more advanced (worse) disease. The factor variables were also normalized over all COPDGene participants (not just those with clocks). We use the factor data in which we have an associated pixelated clock and associated Adjusted Clock Drawing Score to explore associations between the scored version of the clock and the factors.

First, we will now perform four different T-Tests for each Factor variable and the Adjusted Clock Drawing Test Scores variable.

The calculated Statistic and P-Values and statistics from performing the T-Test on the Factor Variables and the the Adjusted Clock Drawing Test Score variable can be seen below:

	Statistic:	P-Value:
Factor 1 (Emphysema):	17.993	**2.9525947380683684e-59**
Factor 2 (Airway):	16.599	**5.094635135350925e-52**
Factor 3 (Gas-Trapping):	18.669	**7.720186926657116e-63**
Factor 4 (Total Lung Capacity):	18.664	**8.121916090616809e-63**

Table 8. T-Test on the Four Factor Variables and the Adjusted Clock Drawing Test Score Variable.

Upon observation of the results obtained from the T-Test executed on the Emphysema Factor (Factor 1) and the Adjusted Clock Drawing Test Score variable we can see that we have obtained a significant P-Value here as the calculated P-Value is much smaller than the set alpha value of 0.05.

When observing the results obtained from implementing the T-Test on the Airway Factor and the Adjusted Clock Drawing Test Score variable we see that we have obtained a statistically significant P-Value here as the obtained P-Value is a lot smaller than the chosen alpha value of 0.05.

Upon observation of the results obtained from the T-Test executed on the Gas-Trapping Factor and the Adjusted Clock Drawing Test Score variable we can see that we have obtained a significant P-Value here as the calculated P-Value is much smaller than the set alpha value of 0.05.

When studying the results obtained from implementing the T-Test on the Total Lung Capacity Factor (Factor 4) and the Adjusted Clock Drawing Test Score variable we see that we have obtained a statistically significant P-Value as the calculated P-Value is much smaller than the chosen alpha value of 0.05.

Without considering any other variables, there is substantial evidence of a difference in subtype values (mean strength of the subdisease of COPD being studied) for clocks that show

dementia risk versus those that don't. As there are additional variables known to affect the mini cog score, we now make adjustments for the four factor variables of interest (F1compP2, F2compP2, F3compP2, F4compP2) by incorporating the following variables into the statistical analysis: age, gender, race, education, Income, and kidney disease (Yes or No).

The ANOVA Table for the Emphysema Factor (Factor 1) as a function of the Adjusted Clock Drawing Test Score plus Age plus Gender plus Race plus Education plus Income plus Kidney Disease (Yes or No) is shown below:

	Degrees of Freedom:	Sum of Squares:	Mean Squared Error:	F-Value:	P-Value:
Adjusted Clock Drawing Test Score:	1	7.08	7.08	10.137	**0.001668**
Gender:	1	2.616	2.616	3.745	0.54273
Race:	1	1.154	1.54	1.652	0.2
Education:	2	3	1.5	2.147	0.119
Income:	5	3.66	0.73	1.03	0.391
Kidney Disease:	1	3.19	3.19	4.56	**0.033856**
Age:	1	12.58	12.58	18	**0.000033**
Residual:	215	150.19	0.699	NaN	NaN

Table 9. ANOVA Table for Factor 1 (Emphysema Factor) as a function of the Adjusted Clock Drawing Test Score plus Age plus Gender plus Race plus Education plus Income plus Kidney Disease (Yes or No).

When observing the ANOVA Table, we can see we have obtained three statistically significant P-Values corresponding to the Adjusted Clock Drawing Test Score, Kidney Disease, and Age variables. Note that the statistically significant P-Values are bolded. We can use the P-Value to decide if the difference between group means is statistically significant or not (Zach). We know that these P-Values are statistically significant as they are all smaller than the chosen alpha value of 0.05. Hence, we can imply that the means of the groups of the Adjusted Clock Drawing Test Score, Kidney Disease, and Age Variables is statistically significant. Thus implying that an individual's score on the Clock Drawing Test can impact the severity of that individual's Emphysema. Likewise, this implies that whether or not someone has kidney disease can affect the severity of the individual's Emphysema. This also implies that an individual's age can affect the severity of their Emphysema. "The larger the F-statistic, the greater the variation between sample means relative to the variation within the samples (Zach)." Notice that the F-Values corresponding to the variables that obtain a statistically significant P-Value are larger.

The ANOVA Table for the Airway Disease Factor (Factor 2) as a function of the Adjusted Clock Drawing Test Score plus Age plus Gender plus Race plus Education plus Income plus Kidney Disease (Yes or No) is shown below:

	Degrees of Freedom:	Sum of Squares:	Mean Squared Error:	F-Value:	P-Value:
Adjusted Clock Drawing Test Score:	1	1.25	1.25	1.488	0.223790
Gender:	1	1.714	1.714	2.033	0.155328
Race:	1	0.338	0.338	0.401	0.527495
Education:	2	2.65	1.325	1.571	0.210128
Income:	5	8.45	1.68	2.004	0.079221
Kidney Disease:	1	0.012	0.012	0.014	0.904801
Age:	1	1.27	1.27	1.507	0.220883
Residual:	215	181.274	0.843	NaN	NaN

Table 10. ANOVA Table for Factor 2 (Emphysema Factor) as a function of the Adjusted Clock Drawing Test Score plus Age plus Gender plus Race plus Education plus Income plus Kidney Disease (Yes or No).

Upon observation of the created ANOVA Table, we see that there were not any statistically significant P-Values obtained as all of the calculated P-Values here are not smaller than the set alpha value of 0.05. Recall that the P-Values help us determine if the means of the groups of a variable are statistically significant or not (Zach). Thus, we can conclude that in this setting, none of the means of the groups of the variables are statistically significant. Let us notice that the calculated F-Values are not very large here.

The ANOVA Table for the Gas-Trapping Factor (Factor 3) as a function of the Adjusted Clock Drawing Test Score plus Age plus Gender plus Race plus Education plus Income plus Kidney Disease (Yes or No) is shown below:

	Degrees of Freedom:	Sum of Squares:	Mean Squared Error:	F-Value:	P-Value:
Adjusted Clock Drawing Test Score:	1	7.43	7.43	10.70	**0.001246**
Gender:	1	0.7	0.7	1	0.316237
Race:	1	0.000024	0.000024	0.000034	0.995331
Education:	2	6.05	3.03	4.36	**0.013956**
Income:	5	4.37	0.87	1.26	0.283208
Kidney Disease:	1	3.22	3.22	4.64	**0.032338**
Age:	1	16.42	16.42	23.64	**0.000002**
Residual:	215	149.31	0.69	NaN	NaN

Table 11. ANOVA Table for Factor 3 (Gas-Trapping Factor) as a function of the Adjusted Clock Drawing Test Score plus Age plus Gender plus Race plus Education plus Income plus Kidney Disease (Yes or No).

When studying the results obtained in the ANOVA Table above (Table 14), we can observe that we have obtained four statistically significant P-Values corresponding to the

Adjusted Clock Drawing Test Score, Education, Kidney Disease, and Age. Note that the statistically significant P-Values are bolded. P-Values help us conclude if the means of the groups of a variable are statistically significant or not (Zach). The P-Values corresponding to the Adjusted Clock Drawing Test Score (0.001246), Education (0.013956), Kidney Disease (0.032338), and Age (0.000002) are statistically significant as they are all smaller than the chosen alpha value of 0.05. Notice that the smallest statistically significant P-Value corresponds to the Age variable. Hence, we can imply that the means of the groups of the Adjusted Clock Drawing Test Score, Education, Kidney Disease, and Age Variables is statistically significant. Thus implying that an individual's score on the Clock Drawing Test can impact the amount of Gas Trapped in someone's lungs. Likewise, this implies that whether or not someone has kidney disease can affect the amount of gas that is trapped in that person's lungs. This also implies that an individual's age and education can affect the amount of gas that is trapped in the individual's lungs. We can see that the variables that correspond to a significant P-Value obtain the largest F-Values.

The ANOVA Table for the Total Lung Capacity Factor (Factor 4) as a function of the Adjusted Clock Drawing Test Score plus Age plus Gender plus Race plus Education plus Income plus Kidney Disease (Yes or No) is shown below:

	Degrees of Freedom:	Sum of Squares:	Mean Squared Error:	F-Value:	P-Value:

Adjusted Clock Drawing Test Score:	1	3.51	3.51	4.19	0.41846
Gender:	1	0.003	0.003	0.003	0.956145
Race:	1	0.26	0.26	0.32	0.574331
Education:	2	14.66	7.33	8.76	**0.000220**
Income:	5	6.21	1.24	1.48	0.196206
Kidney Disease:	1	4.27	4.27	5.10	**0.024903**
Age:	1	6.71	6.71	8.02	0.005065
Residual:	215	179.89	0.84	NaN	NaN

Table 12. ANOVA Table for Factor 4 (Total Lung Capacity Factor) as a function of the Adjusted Clock Drawing Test Score plus Age plus Gender plus Race plus Education plus Income plus Kidney Disease (Yes or No).

Upon observation of Table 44, we observe that we have obtained three statistically significant P-Values corresponding to the Education and Kidney Disease variables. Note that the statistically significant P-Values are bolded. P-Values are useful statistical tools in the sense that can help us determine if the means of the groups of a variable are statistically different or not (Zach). The P-Values corresponding to Education (0.000220) and Kidney Disease (0.024903) are statistically significant as they are all smaller than the chosen alpha value of 0.05. Notice that the smallest statistically significant P-Value corresponds to the Education variable.This implies that whether or not someone has kidney disease can affect the person's Total Lung Capacity.

This also implies that an individual's education can affect their Total Lung Capacity. We can see that the variables that correspond to a significant P-Value obtain the largest F-Values.

Even after adjustment for known confounders, emphysema and gas trapping still showed an association with the clock score. That is, there is evidence of a relationship between the clocks and our main COPD phenotypes of interest: the subtypes. In the next section we employ the clustering algorithms to find a possible alternative classification of the clocks than the summary 0 or 2 from the adjudication (manual review) process. Following that, we assess if this new clustering scheme finds differences in the subtypes that are only hinted at by the comparison of factor scores across clock scores.

(4) K-Means/K-Medoids Clustering Algorithms (Unsupervised) on the pixelated Clock Drawing Test image data

The Convolutional Neural Network served the purpose of cleaning the image data by zoning in on the more important pixels. We can see that we have a Flatten layer in our Convolutional Neural Network which flattens the pixels into a singular number; this is exactly what we need to perform the K-Means as the K-Means expects a 2-Dimensional input.

Now, we will extract the convoluted data from the flatten layer to then perform K-Means on the convoluted data. We are performing K-Means on the pixelated clock data to investigate the clusters obtained from implementing the K-Means. For example, we want to determine the percentage of participants that have Self-Reported Chronic Obstructive Pulmonary Disease in each cluster obtained from the K-Means Algorithm.

Note that the clustering algorithms are performed on the 65 High-Risk for Dementia (Score of 0).

We begin by using 2 clusters when implementing the K-Means Algorithm. To ensure that the optimal number of clusters is being used when implementing the K-Means Algorithm we check via the Elbow Method which is shown graphically below:

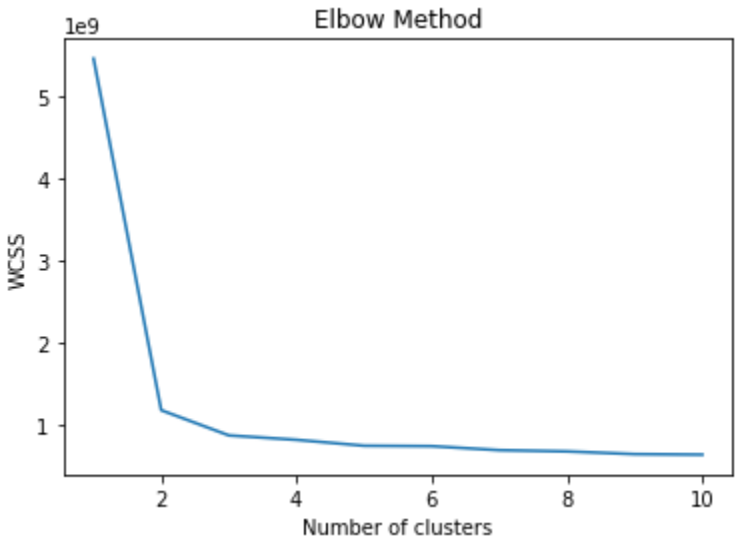

Figure 12- Ensuring the Optimal Number of Clusters are Being Used when Implementing the K-Means Clustering Algorithm via the Elbow Method.

Upon observation of Figure 12 we can see that the optimal number of clusters is 2 when implementing the K-Means Clustering Algorithm as that is when we see the WCSS (Within-Cluster Sum of Square) drop significantly creating the infamous elbow shape.

We obtained the following Confusion Matrix when performing the K-Means Clustering Algorithm on the Convoluted Pixelated High-Risk for Dementia Clock Drawing Test Images:

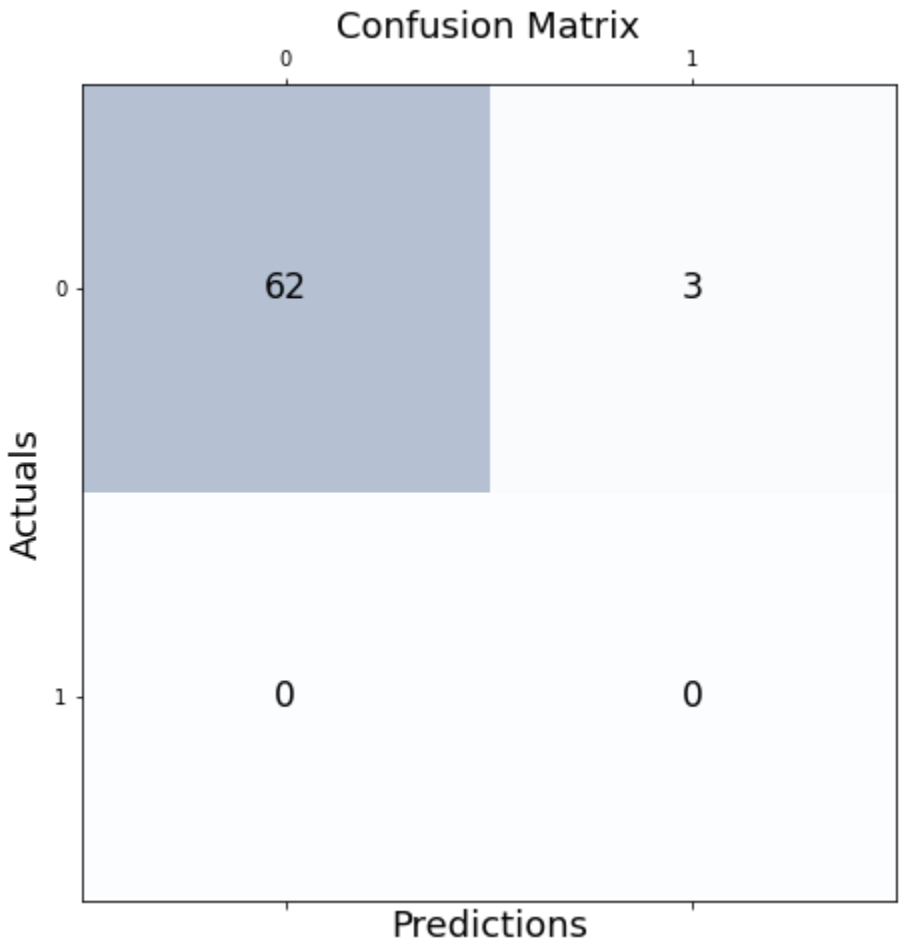

Figure 13- Confusion Matrix Obtained from Performing the K-Means Clustering Algorithm on the Convoluted Pixelated High-Risk for Dementia Clock Drawing Test Images.

From Figure 13 we can see that the K-Means Clustering Algorithm has grouped 62 of the High-Risk for Dementia Clock Drawing Test images in a cluster (Cluster A) while the remaining 3 High-Risk for Dementia Clock Drawing Test images were grouped together in a separate cluster (Cluster B). This is not very useful as most of the clocks (95%) are in the same cluster (Cluster A).

Furthermore, we will investigate images from the clusters. We can see a sample of 9 Clock Drawing Test images that were clustered together in Cluster A shown below:

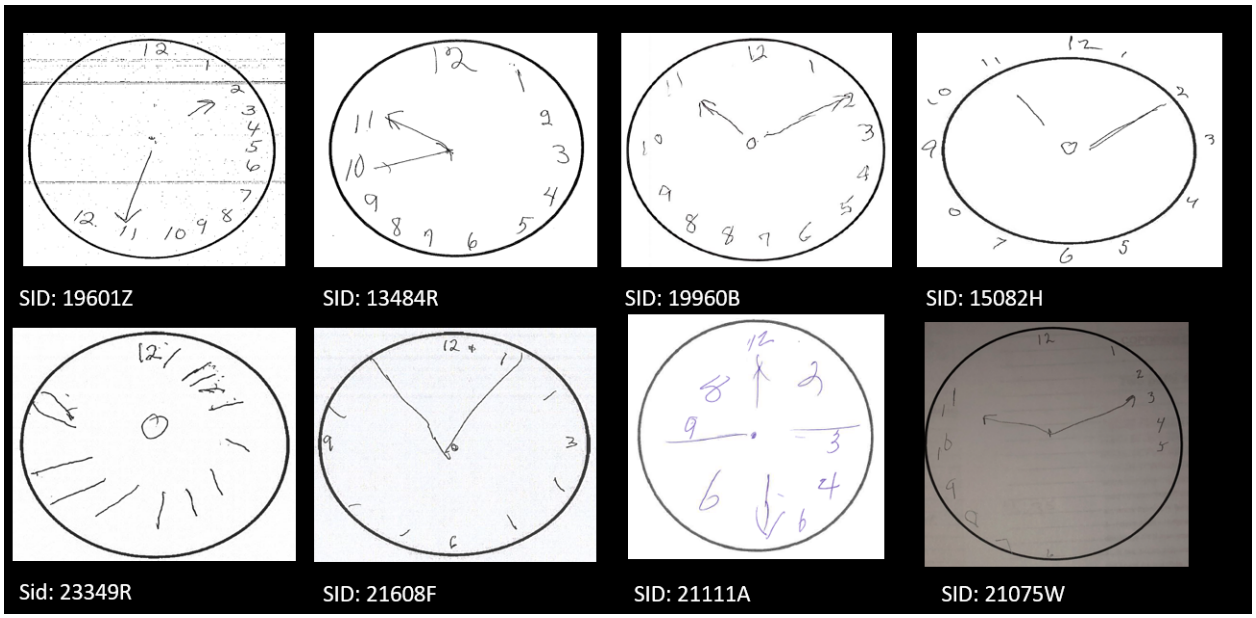

Figure 14- Sample of Clock Drawing Test Images from Cluster A (Cluster of 62 Clock Drawing Test Images).

Next, we investigate the Clock Drawing Test images within Cluster B (the second cluster). Recall that there were only three Clock Drawing Test images in Cluster B. The SID, Age, Gender, Final GOLD Baseline, Kidney Disease, and Self-Reported COPD variables for each Clock Drawing Test in Cluster B is listed above the image. The three Clock Drawing Test images contained in Cluster B are shown below:

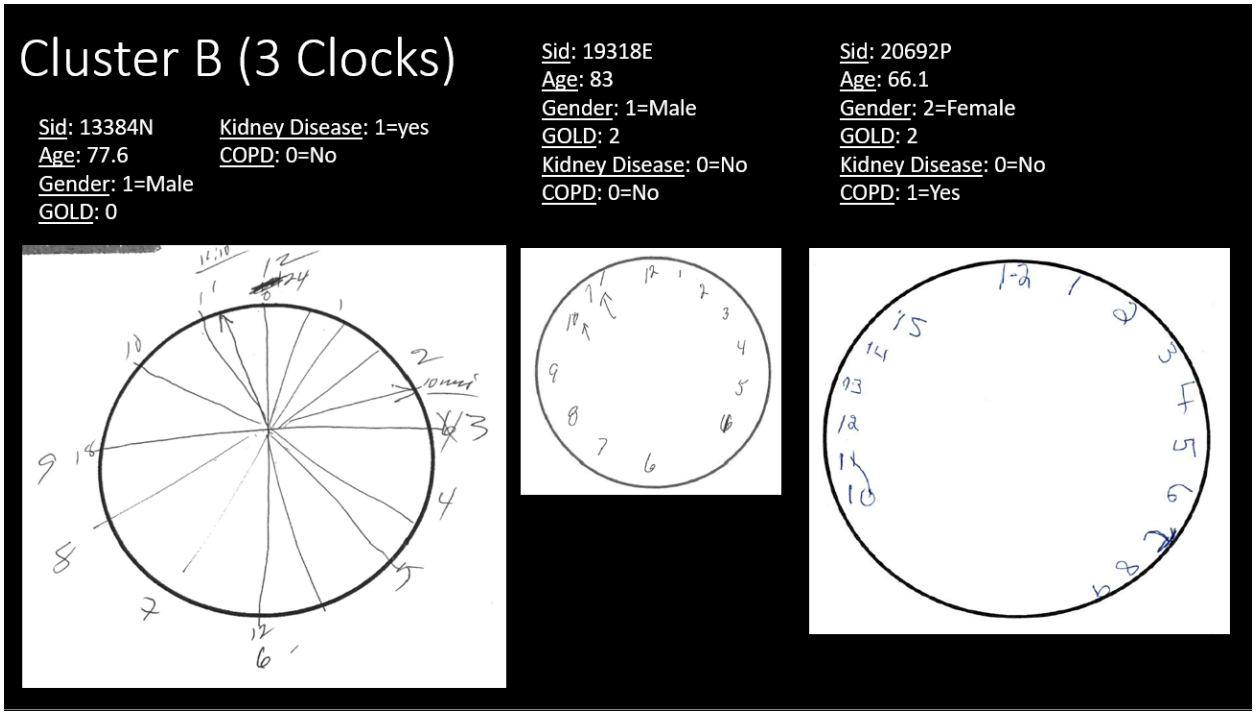

Figure 15- Sample of Clock Drawing Test Images from Cluster B (Cluster of 3 Clock Drawing Test Images).

The K-Means Inertia value is a performance measure of how well the K-Means Algorithms clustered the data (Learn). The calculated K-Means Inertia value can be seen below:

K-Means Inerita Value:	769,336,887.67

Table 13. Inertia Value (Lowest Sum of Squares Error Value) for the K-Means Algorithm Implemented on the Convoluted High-Risk for Dementia Pixelated Clock Drawing Test Images.

An optimal model obtains a small inertia value and a small number of clusters (Learn). In this case, the most optimal number of clusters is 2 (which is small). However, we have a very large inertia meaning that the K-Means algorithm did not perform very well when clustering the convoluted pixelated High-Risk for Dementia Clock Drawing Test images.

Since the K-Means Algorithm did not perform very well on clustering the data (pixelated High-Risk for Dementia Clock Drawing Test images), we will proceed to our next partitioning algorithm; the K-Medoids Clustering Algorithm. As with the K-Means, we will be implementing the K-Medoids Algorithm on the convoluted, pixelated High-Risk for Dementia Clock Drawing Test images.

We start by using 2 clusters when implementing the K-Medoids Algorithm. We ensure that the optimal number of clusters is being used via the Elbow Method, which is shown graphically below:

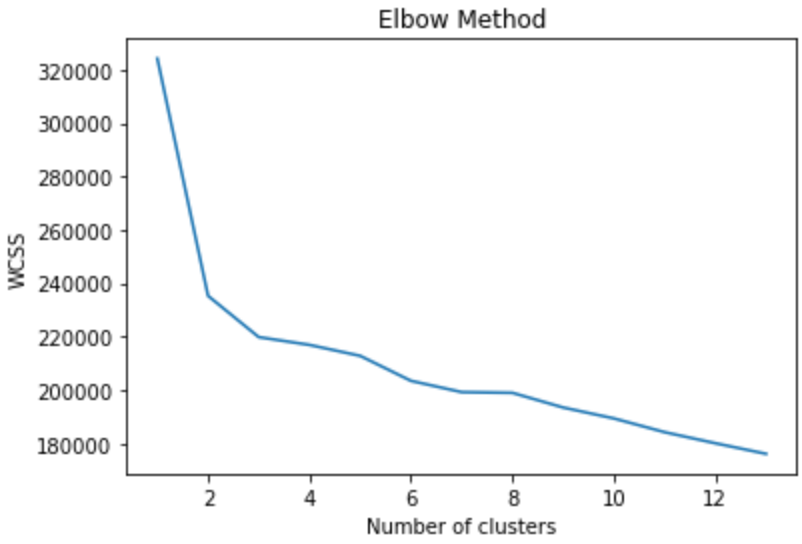

Figure 16. Ensuring the Optimal Number of Clusters are Being Used when Implementing the K-Medoids Clustering Algorithm via the Elbow Method.

Upon study of Figure 16 we can see that the choice of 2 clusters is reasonable, but we may do better choosing more clusters.

We obtained the following Confusion Matrix when performing the K-Medoids Clustering Algorithm on the Convoluted Pixelated High-Risk for Dementia Clock Drawing Test Images:

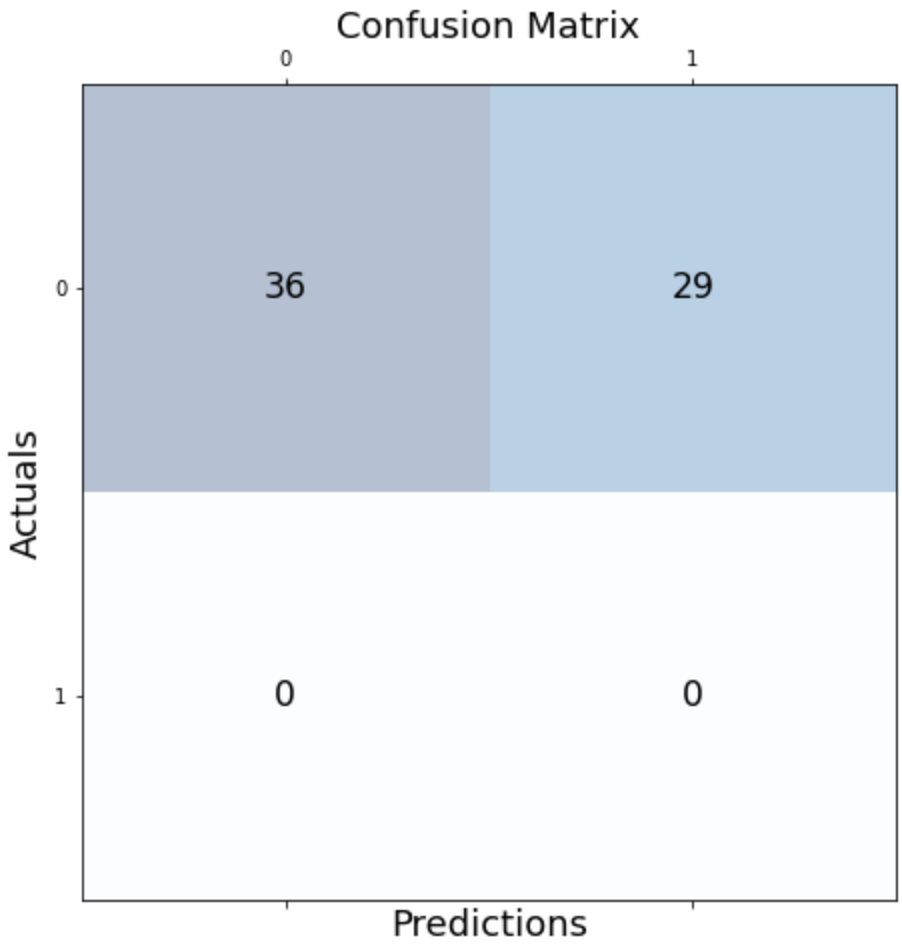

Figure 17- Confusion Matrix Obtained from Performing the K-Medoids Clustering Algorithm on the Convoluted Pixelated High-Risk for Dementia Clock Drawing Test Images.

From Figure 17 we can see that the K-Medoids Clustering Algorithm has grouped 36 (55.38%) of the High-Risk for Dementia Clock Drawing Test images in a cluster (Cluster A) while the remaining 29 High-Risk for Dementia Clock Drawing Test images were grouped together in a separate cluster (Cluster B). The clusterings obtained here are potentially more meaningful than the ones obtained by the K-Means algorithm.

We will now investigate the Clock Drawing Test Images within each cluster obtained by the K-Medoids Algorithm in Figures 18 and 19.

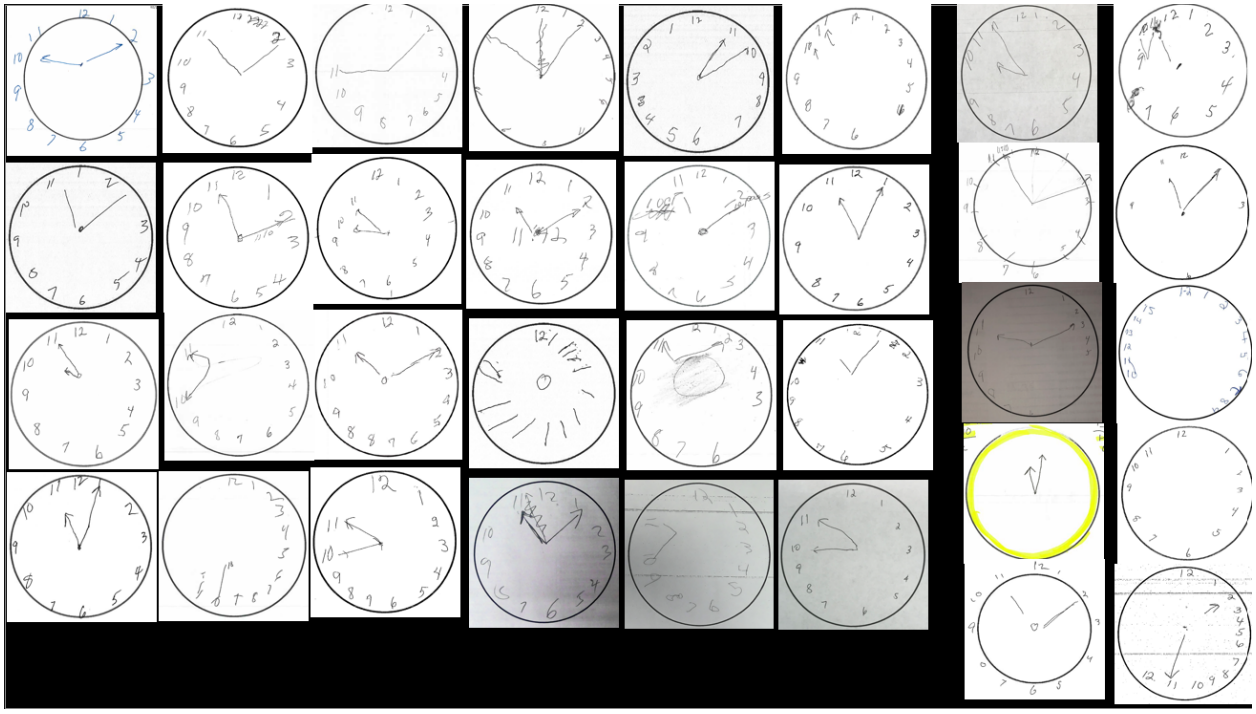

Figure 18- Clock Drawing Test Images from Cluster A (Cluster of 36 Clock Drawing Test Images).

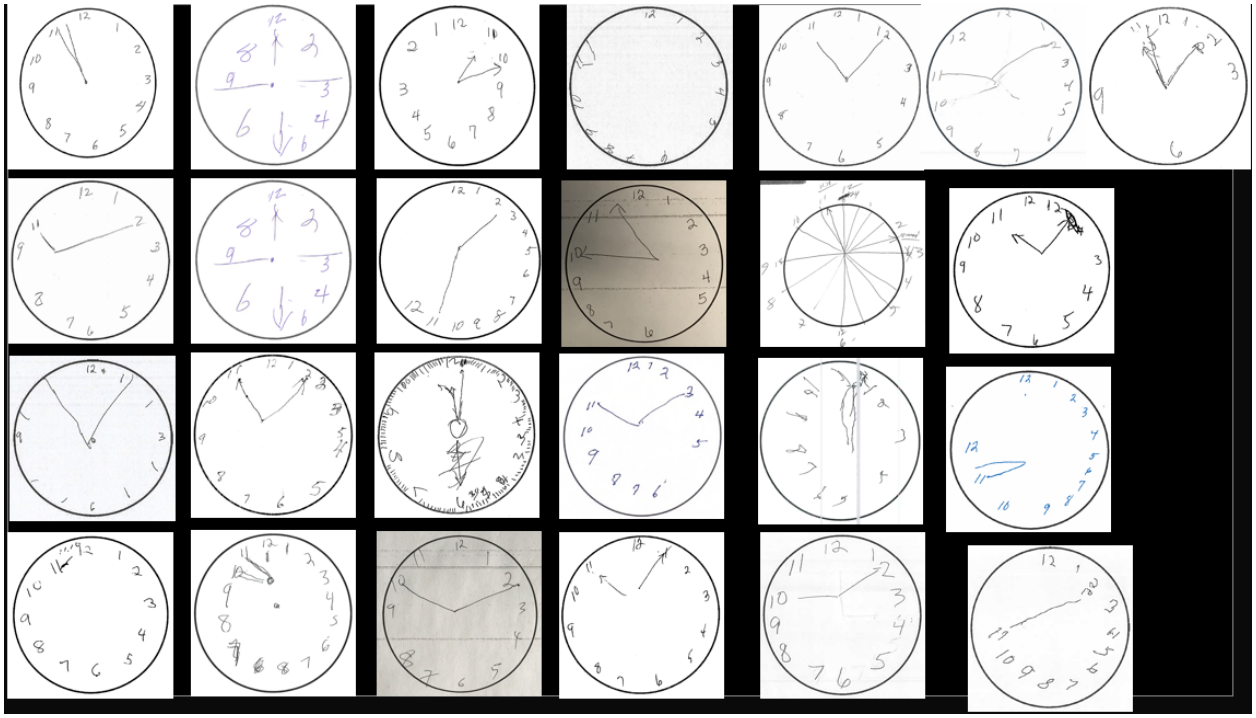

Figure 19- Sample of Clock Drawing Test Images from Cluster B (Cluster of 29 Clock Drawing Test Images).

Figures 18 and 19 are observed to see if we can observe any obvious patterns within the clustered clocks. We cannot see any obvious patterns from looking at the Clock Drawing Test Images within Clusters A & B. While we do not notice anything by inspection, this is the purpose of the K-Medoids cluster algorithm. We will determine if these hidden clusters are related to COPD traits (which will be tested in the next section).

The K-Medoids Inertia value is a performance measure of how well the K-Medoids Algorithms clustered the data (Learn). The calculated K-Medoids Inertia value can be seen below:

K-Medoids Inertia Value:	320,033.72

Table 14. Inertia Value (Lowest Sum of Squares Error Value) for the K-Medoids Algorithm Implemented on the Convoluted High-Risk for Dementia Pixelated Clock Drawing Test Images.

An optimal model obtains a small inertia value and a small number of clusters (Learn). In this case, the most optimal number of clusters is 2 (which is small). Here, the calculated inertia is not super small, but it is much smaller than the calculated inertia value for the K-Means Algorithm.

(5) Testing COPD Subtypes versus New Clusters

First, we will find the mean of the four factor variables of the patients from Cluster A:

Factor 1 (Emphysema Factor):	0.439810
Factor 2 (Airway Factor):	0.074058
Factor 3 (Gas-Trapping):	0.449269
Factor 4 (Total Lung Capacity Factor):	0.488264

Table 15. The Average of the Four Factor Variables for Cluster A Obtained from K-Medoids Algorithm Implemented on the Convoluted High-Risk for Dementia Pixelated Clock Drawing Test Images.

Next, we will investigate the average of the four factor variables from cluster B:

Factor 1 (Emphysema Factor):	0.340643
Factor 2 (Airway Factor):	-0.056599
Factor 3 (Gas-Trapping):	0.310559
Factor 4 (Total Lung Capacity Factor):	0.347953

Table 16. The Average of the Four Factor Variables for Cluster B Obtained from K-Medoids Algorithm Implemented on the Convoluted High-Risk for Dementia Pixelated Clock Drawing Test Images.

Comparing Tables 15 and 16 we can see that the averages of the factors is much smaller in Cluster B, even obtaining a negative average for the Airway Factor (Factor 2).

We will not obtain the Percentage of patients in clusters A and B who obtained a positive Self-Reported COPD diagnosis as well as patients who obtain a Final GOLD Baseline Score of 1-4:

Percentage of Patients in Cluster A that Obtained a Positive Self-Reported COPD Diagnosis:	22.22%
Percentage of Patients in Cluster A that Obtain a Final GOLD Baseline Score of 1-4:	44.44%

Percentage of Patients in Cluster B that Obtained a Positive Self-Reported COPD Diagnosis:	27.59%
Percentage of Patients in Cluster B that Obtain a Final GOLD Baseline Score of 1-4:	44.83%

Table 17. Percentage of COPD for the Clusters Obtained from K-Medoids Algorithm Implemented on the Convoluted High-Risk for Dementia Pixelated Clock Drawing Test Images.

When we study Table 17 we can see that in Cluster B there was a higher percentage of patients that obtained a positive self-reported COPD diagnosis as well as patients who obtained a final GOLD baseline score of 1-4.

Chapter 8: Conclusions

In the Clock Drawing Test Project we had one major finding: when utilizing the K-Medoids clustering algorithm we found that overall Cluster A is less healthy on all Four Factor variables (the subtypes of COPD) than Cluster B. However the percentage of patients that obtain a Final GOLD Baseline Score of 1-4 is same for both clusters, implying that GOLD stage did not catch the sicker cluster. That is, within the dementia group, machine learning methods could use the same dementia screening tool (the Clock Drawing Test) to find people with more extreme COPD (sicker patients).

For further research I think that it would be interesting to find Factor (subtypes of COPD) threshold values (i.e. if someone is below the threshold for Emphysema they are sicker). I also think it would be beneficial for this work to continue on experimenting with more clustering algorithms or maybe a different CNN. Another idea is to incorporate more data as we only had 65 High-Risk for dementia Clock Drawing Tests.

Some limitations in this study are the mini-cog as it is a very limited tool. The mini-cog is fast and inexpensive to implement, but it is limited in what it can show us. However, the clustering algorithms used in the project did find something in showing that one cluster of patients was sicker. Since the obtained interia value for the K-Medoid clustering algorithm is large, there could be a more optimal clustering algorithm that could improve the inertia.

Chapter 9: Python Code (via Github)

I have attached a link that will take you to the python code that was posted to Github that was used to create the results seen above:

hopehaygood/Clock-Drawing-Test-Project (github.com)

Once clicking on the link, click on the .ipynb file titled 'CDT_Project_Code', and this will take you to the Python code used in this Clock Drawing Test Project.

Chapter 10: Acknowledgments

I would like to thank my Master's Committee:

- Erin Austin, PhD, University of Colorado Denver (Chair, Advisor)
- Kendra A. Young, PhD, Colorado School of Public Health
- Greg Kinney, PhD, Colorado School of Public Health
- Yaning Liu, PhD, University of Colorado Denver

along with COPDGene; Karin Hoth,PhD, and Cari from the University of Iowa; and the LACE team:

- John E. Hokanson, PhD, Colorado School of Public Health
- Erin Austin, PhD, University of Colorado Denver
- Rebecca Conway, PhD, Colorado School of Public Health
- Negar Janani, MS, University of Colorado Denver
- Greg Kinney, PhD, Colorado School of Public Health
- Kendra A. Young, PhD, Colorado School of Public Health
- Yisha Li, MPH, Colorado School of Public Health
- Danielle Sansone-Poe
- Nichole Bulpett

Chapter 11: References

Adi - types of dementia. Alzheimer's Disease International (ADI). (n.d.). Retrieved November 22, 2022, from https://www.alzint.org/about/dementia-facts-figures/types-of-dementia/#:~:text=There%20are%20over%20100%20forms,of%20dementia%20being%20Alzheimer's%20disease.

Brownlee, J. (2020, August 18). *Transfer learning in Keras with computer vision models*. Machine Learning Mastery. Retrieved October 16, 2021, from https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/.

Categorical data vs Numerical Data. Voxco. (2022, August 11). Retrieved September 12, 2022, from https://www.voxco.com/blog/categorical-data-vs-numerical-data/#:~:text=Definition-,Categorical%20data%20refers%20to%20a%20data%20type%20that%20can%20be,any%20language%20or%20descriptive%20form.&text=Also%20known%20as%20qualitative%20data%20as%20it%20qualifies%20data%20before%20classifying%20it.

Centers for Disease Control and Prevention. (2021, July 29). *Copd*. Centers for Disease Control and Prevention. https://www.cdc.gov/dotw/copd/index.html.

Dabbura, I. (2020, August 10). *K-means clustering: Algorithm, applications, evaluation methods, and drawbacks*. Medium. Retrieved November 6, 2021, from https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a.

DementiaCareCentral.com. (2020, May 28). *Mini-cog test for dementia and alzheimer's: Accuracy, Administration & Scoring*. Dementia Care Central. Retrieved October 15, 2021, from https://www.dementiacarecentral.com/mini-cog-test-for-dementia/.

De-Yu, C. (2022, March 4). *Chi-square test, with python*. Medium. Retrieved October 2, 2022, from https://towardsdatascience.com/chi-square-test-with-python-d8ba98117626

Eknoyan, D., Eknoyan, D., Hurley, R. A., Taber, K. H., Address correspondence to Dr. Robin Hurley, I, R., Al., E., T, I., P, L., E, K., KI, S., E, P., M, C., H, G., N, E., PJ, W., BW, W., Y, K., M, B., … Kao, H.-J. (2012, July 1). *The clock drawing task: Common errors and functional neuroanatomy*. The Journal of Neuropsychiatry and Clinical Neurosciences. Retrieved

December 2, 2021, from https://neuro.psychiatryonline.org/doi/10.1176/appi.neuropsych.12070180.

Ellison, J. M., & Swank Center for Memory Care and Geriatric Consultation. (2020, November 9). *What is dementia?* BrightFocus Foundation. https://www.brightfocus.org/alzheimers/article/what-dementia?gclid=Cj0KCQjw-NaJBhDsARIsAAja6dOH0pP2D37tI4XzWZPh15btl0Io6HWXG3rWI6HUCr6h-XVBemHMB4oaAkGOEALw_wcB.

Esther Heerema, M. S. W. (2021, July 30). *Scoring and interpreting the clock DRAWING test for dementia*. Verywell Health. https://www.verywellhealth.com/the-clock-drawing-test-98619.

Facts and figures. Alzheimer's Disease and Dementia. (n.d.). https://www.alz.org/alzheimers-dementia/facts-figures.

Faris, S. (2020, February 26). *Recognizing serious COPD complications*. Healthline. Retrieved November 30, 2021, from https://www.healthline.com/health/copd/serious-complications#dementia.

Flow chart of K-medoid algorithm. | download scientific diagram. (n.d.). Retrieved September 15, 2022, from

https://www.researchgate.net/figure/Flow-chart-of-k-medoid-algorithm_fig2_276496042

Gama, G. (2018, October 20). *K-Medoids in R: Algorithm and practical examples*. Datanovia. Retrieved November 23, 2021, from https://www.datanovia.com/en/lessons/k-medoids-in-r-algorithm-and-practical-examples/.

Harrison, O. (2019, July 14). *Machine learning basics with the K-nearest neighbors algorithm*. Medium. Retrieved September 17, 2022, from https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761

Home. COPDGene. (n.d.). Retrieved November 22, 2022, from http://www.copdgene.org/

Kinney GL, Santorico SA, **Young KA**, Cho MH, Castaldi P, San Jose Estapar R, Ross J, Dy J, Make B, Regan E, Lynch D, Everett DC, Lutz SM, Silverman EK, Washko G, Crapo J, Hokanson JE. Identification of Chronic Obstructive Pulmonary Disease Axes That Predict All-Cause Mortality: The COPDGene Study. *American Journal of Epidemiology.* 2018 Oct 1;187(10):2109-2116. doi.org/10.1093/aje/kwy087. PubMed PMID: 29771274; PubMed Central PMCID: PMC6166205.

Liao, M., Li, Y., Kianifard, F., Obi, E., & Arcona, S. (2016, March 2). *Cluster analysis and its application to healthcare claims data: A study of end-stage renal disease patients who initiated hemodialysis*. BMC nephrology. Retrieved November 16, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4776444/.

Liao, K.-M., Ho, C.-H., Ko, S.-C., & Li, C.-Y. (2015, June). *Increased risk of dementia in patients with chronic obstructive pulmonary disease*. Medicine. Retrieved September 27, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4616489/.

Navigation. Cosmos. (n.d.). Retrieved November 29, 2022, from https://www.cosmos.esa.int/web/machine-learning-group/convolutional-neural-networks-introduction

O'Shea, Keiron and Ryan Nash. "An Introduction to Convolutional Neural Networks." *ArXiv* abs/1511.08458 (2015): n. pag.

Rabe KF;Halpin DMG;Han MK;Miravitlles M;Singh D;Grönke L;Voß F;Martinez FJ; (n.d.). *Composite endpoints in COPD: Clinically important deterioration in the uplift trial*. Respiratory research. Retrieved November 29, 2022, from https://pubmed.ncbi.nlm.nih.gov/32646424/

Radhakrishnan, P. (2019, October 26). *What is transfer learning?* Medium. Retrieved December 14, 2021, from https://towardsdatascience.com/what-is-transfer-learning-8b1a0fa42b4

Ramzi A. Haraty, M. D. (2012, July 1). *An enhanced K-means clustering algorithm for pattern discovery in healthcare data - Ramzi A. Haraty, Mohamad Dimishkieh, Mehedi Masud, 2015*. SAGE Journals. Retrieved November 16, 2021, from https://journals.sagepub.com/doi/10.1155/2015/615740.

Rapposelli, D. (2020, November 16). *Identifying dementia subtypes with the clock drawing test*. Psychiatric Times. Retrieved September 13, 2021, from https://www.psychiatrictimes.com/view/identifying-dementia-subtypes-clock-drawing-test.

Real Python. (2022, September 1). *Logistic regression in python*. Real Python. Retrieved November 5, 2022, from https://realpython.com/logistic-regression-python/

Rodriguez, T. (2019, July 30). *Copd-related cognitive impairment unlike other forms of dementia*. Pulmonology Advisor. Retrieved September 11, 2021, from https://www.pulmonologyadvisor.com/home/topics/copd/copd-related-cognitive-impairment-unlike-other-forms-of-dementia/.

Roy, A. (2020, November 23). *A dive into unsupervised learning*. Medium. Retrieved October 24, 2021, from https://towardsdatascience.com/a-dive-into-unsupervised-learning-bf1d6b5f02a7.

Scipy.stats.chisquare#. scipy.stats.chisquare - SciPy v1.9.1 Manual. (n.d.). Retrieved October 2, 2022, from https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chisquare.html

Shiledarbaxi, N. (2021, April 9). *Comprehensive guide to K-medoids clustering algorithm*. Analytics India Magazine. Retrieved November 23, 2021, from https://analyticsindiamag.com/comprehensive-guide-to-k-medoids-clustering-algorithm/#:~:text=K%2DMedoids%20is%20a%20clustering,it%20selects%20the%20clusters%27%20centres.

Srivastava, A. (2021, January 18). *K-means clustering*. Medium. Retrieved September 15, 2022, from https://medium.com/analytics-vidhya/k-means-clustering-3ce2456db7f1

Supervised vs unsupervised learning - javatpoint. www.javatpoint.com. (n.d.). Retrieved October 24, 2021, from https://www.javatpoint.com/difference-between-supervised-and-unsupervised-learning.

Types of dementia. Alzheimer's Disease and Dementia. (n.d.). Retrieved September 27, 2021, from https://www.alz.org/alzheimers-dementia/what-is-dementia/types-of-dementia.

Young KA, Strand M, Ragland MF, Kinney GL, Austin EE, Regan EA, Lowe KE, Make BJ, Silverman EK, Crapo JD, Hokanson JE; COPDGene® Investigators. Pulmonary Subtypes Exhibit Differential Global Initiative for Chronic Obstructive Lung Disease Spirometry Stage Progression: The COPDGene® Study. *Chronic Obstructive Pulmonary Disease*. 2019 Nov;6(5):414-429. doi: 10.15326/jcopdf.6.5.2019.0155. PubMed PMID: 31710796; PubMed Central PMCID: PMC7020848.

Zach. (2020, April 17). *How to calculate a P-value from a T-test by hand*. Statology. Retrieved June 14, 2022, from https://www.statology.org/how-to-calculate-a-p-value-from-a-t-test-by-hand/

Zach. (2021, August 25). *How to perform Fisher's exact test in Python*. Statology. Retrieved October 2, 2022, from https://www.statology.org/fishers-exact-test-python/

World Health Organization. (n.d.). *Dementia*. World Health Organization. Retrieved November 22, 2022, from https://www.who.int/news-room/fact-sheets/detail/dementia