

# BME Design-Fall 2023 - Richard YANG

## Complete Notebook

PDF Version generated by

DHRUV NADKARNI

on

Dec 13, 2023 @02:33 PM CST

### Table of Contents

Project Information	2
Team contact Information	2
Project description	3
projectTimeline	4
Team activities	5
Client Meetings	5
2023/9/15 Client Meeting #1	5
Advisor Meetings	8
2023/9/15 Advisor Meeting #1	8
progressReports	9
template	9
Progress Report 9/14/2023	10
Aran Viswanath	11
Research Notes	11
Biology and Physiology	11
09/15/2023 - Sjögren's Syndrome Diagnosis Research	11
9/22/2023 - Research On Other Diagnostic Tests	12
9/29/2023 - Image classification research VGG-19	13
9/29/2023 - BPAG Meeting	14
Competing Designs	15
9/22/2023 - Machine-aided workflow in ultrasound imaging	15
9/22/2023 - Method for developing a machine learning model of a neural network for classifying medical images	16
Design Ideas	17
10/17/2023 - Optical Character Recognition - Pytesseract	17
10/17/2023 - DICOM File Reading	18
10/17/2023 - File Location and Name Manipulation	19
10/24/2023 - Pytorch Documentation: Converting Grayscale Images Into Tensors	20
Brandon Drew	21
Research Notes	21
Biology and Physiology	21
2023/09/15-Sjogren syndrome overview	21
2023/09/15-Current grading system of salivary gland ultrasounds	22
2023/09/21-Machine Learning Intro	23
2023/09/28 - KNN research	24
2023/09/29-VGG-19 Research	25
Training Documentation	26
CITI Training	26
Dhruv Nadkarni	27
Research Notes	27
Biology and Physiology	27
2023/09/15_SalvaryGlandUltrasound	27
2023/09/15_MachineLearningForMedicalUltrasound	28
2023/09/15_EndocrinepSS	29
2023/09/20_CTScanSialolithiasis	30
2023/09/19_BilateralMultiSialWithSjor	31
Algorithms	32

2023/09/26 - DesignMatrixResnet50	33
Design Ideas	34
kNN Notes	34
2023/09/26 - DesignMatrixDNN	34
2023/10/05 - kNNSciKitDocumentation	35
2023/10/11 - CompleteKNNGuide	36
Dhruv's Past Colab Notebooks To Reference	37
2023/10/17 - DataProcessingKNN	38
KNN Data Imports (VGG19)	39
10/30-11/2_KNN/VGG19 Data Import Examples	39
2023/11/07-kNNforImageClassification	41
Training Documentation	42
CITI Training Certificate	42
Meeting Notes	43
10/12/2023_ClientMeeting	43
2023/11/10 - Tong Lecture	44
Siya Mahajan	45
Research Notes	45
Biology and Physiology	45
9/15/23 Sjogren's Syndrome	45
9/15/23 Machine Learning	46
Competing Designs	47
9/29/23 Nearest Neighbor	47
Yousef Gadalla	48
Research Notes	48
Biology and Physiology	48
2023/9/15 Ultrasound as Diagnosis in Sjogren CasesResearch	48
2023/9/21 Machine Learning in Cardiovascular disease	51
Competing Designs	55
2023/9/29 Machine Learning in Medicine Research	55
2023/10/9 UNet Research	58
2023/10/9 VGG-19 For COVID-19 paper	59
2023/12/9 CNN Overview Gradient	60
2023/12/9 CNN Overview (Backpropagation)	62
Training Documentation	63
2023/9/24 CITI Training	63
2023/9/27 PyTorch Lessons Tensors	64
2023/9/29 PyTorch Data Loading	65
Meeting Notes	66
2023/9/15 Team Meeting #2	66
2023/9/8 Team Meeting #1	67
2023/11/10 Tong Lecture	68
Richard Yang	69
Research Notes	69
Biology and Physiology	69
2023/09/10_mortalityRiskFactorsInSJS	69
2023/09/10_transferLearningLatentFactors	70
2023/09/12_classification	71
2023/09/12_OMERACTGradingSystem	72
2023/09/15_OMERACTStudy	73
2023/09/15_indexForTestRetest	74
Standards and Regulations	75
2023/09/21_dataAcquisition	75
2023/09/21_de-identification	77
Design Ideas	79
2023/09/22_data structure	79
2023/09/26_importVectorMachine	80
2023/09/27_KNN	82
2023/09/29_plansForward	83
2023/10/02_VGG19	84
2023/10/06VGG19Continued	85

2023/10/07_ML for US .....	89
2023/10/11_loadingLocalImages .....	91
2023/10/13_loadingImages .....	94
2023/10/25_plansForward .....	95
2023/10/16_thirdClassifier .....	96
2023/10/18_parotidClassifier .....	97
2023/10/20_OMERACTRegressor .....	99
2023/10/21_goalsForShowNTell .....	100
2023/10/23_finalClassifier .....	101
2023/10/23_ideas .....	102
2023/11/01_showAndTell .....	103
2023/11/01_GUIprogress .....	105
2023/11/08_plansForward .....	106
2023/11/10_savingWeights .....	107
2023/11/13_modelResults .....	108
2023/11/15_data imbalance .....	109
2023/11/27_batchSizeAndLearningRate .....	110
2023/11/29_savingWeights .....	111
2023/12/12_futureWork .....	112
Training Documentation .....	113
2023/09/21_citiTraining .....	113
CITI training cert .....	114
2014/11/03-Entry guidelines .....	115
2014/11/03-Template .....	116

**Team contact Information**

DHRUV NADKARNI - Oct 11, 2023, 10:50 PM CDT

Last Name	First Name	Role	E-mail	Phone	Office Room/Building
Casey	Cameron	Advisor	cpcasey3@wisc.edu		
McCoy	Sara	Client	ssmccoy@medicine.wisc.edu		
Yang	Richard	Leader	tyang296@wisc.edu	9783871970	
Gadalla	Yousef	Communicator	ygadalla@wisc.edu	6086928106	
Drew	Brandon	BSAC	bsdrew2@wisc.edu	5038161514	
Nadkarni	Dhruv	BWIG	dnadkarni@wisc.edu	9787935560	
Mahajan	Siya	BWIG	mahajan24@wisc.edu	2624209266	
Viswanath	Aran	BPAG	viswanath3@wisc.edu	8053992726	



## Project description

---

DHRUV NADKARNI - Oct 11, 2023, 11:01 PM CDT

**Course Number:** BME 200/300

**Project Name:** Machine learning for salivary gland ultrasound scoring

**Short Name:** MLSalivaryGlandUS

**Project description/problem statement:**

Sjögren's syndrome (SjS) is a systemic autoimmune disease (SAD) that causes dysfunction of the exocrine glands (mainly the salivary and lacrimal glands) with patients often showing persistent dryness of the mouth and eyes [1, 2]. According to estimations, two to four million people in the United States have SjS; however, only one million have been diagnosed, likely due to the nonspecific diagnostic guidelines and the heterogeneous nature of the disease [3]. The current standard of care of the client is to perform at least baseline salivary gland ultrasounds (of the submandibular and parotid glands) in patients who potentially have SjS. For some higher-risk individuals, regularly scheduled salivary gland ultrasounds are performed.

The problem arises within the current Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) ultrasound grading system, which requires subjective opinions and lacks nuance. As a result, a machine learning approach is proposed to reduce inter-reader variability and to provide a more exact prognosis. The proposed algorithm takes ultrasound grayscale images as input and outputs SjS positive or SjS negative.

**About the client:**

Dr. Sara McCoy is a faculty member in the Division of Rheumatology within the Department of Medicine. She is a clinical rheumatologist, the director of the UW Health Sjögren's clinic, and a core member of the University of Wisconsin Carbone Cancer Center.



Richard YANG - Sep 08, 2023, 4:23 PM CDT



[Download](#)

**projectGANTT\_-\_Sheet1.pdf (77.1 kB)**



## 2023/9/15 Client Meeting #1

---

**Title: Client Meeting #1****Date:** 9/15/2023**Content by:** Yousef**Present:** Yousef, Richard, Siya, Brandon, Dr. McCoy**Goals:**

1. Introduce ourselves to the client
2. Confirm the team's understanding of the project with Dr. McCoy to ensure her goal matches the team's understanding
3. Ask Dr. McCoy questions to learn more about Dr. McCoy's goals and resources for the project

**Content:**

1. Introduction
  1. The team went around introducing ourselves to Dr. McCoy. Each member shared their name and grade.
  2. Dr. McCoy introduced herself as well, supplying information about her reason for coming to the BME department with this problem.
    1. There currently is no scoring system that does not suffer from user variance when diagnosing Sjogren's. This is a problem that other institutions are currently facing as well, which is what prompted Dr. McCoy to approach the BME department with this project.
2. Clarifying project goal
  1. Dr. McCoy shared her desire for the project to initially be made to diagnose a patient with Sjogren's or not
    1. The data for the project will be at least 156 unique patients with confirmed Sjogren's disease
      1. This does not include ultrasounds that Dr. McCoy has of non-Sjogren patients that can be used as a negative control for the learning process
    2. Dr. McCoy mentioned utilizing potentially 1/3 of the images as a testing/validation phase while 2/3 be used for training the algorithm.
      1. As Dr. McCoy works with a group at UCLA, there is potential for more ultrasounds in the future (likely after this semester)
    3. As the data is patient data, there will need to be a way for the team to access these images while following various institutional policies.
      1. Dr. McCoy will add the team to her IRB protocol to help with this process. She will also send a follow-up email with other trainings that the team should complete before gaining access to the de-identified ultrasounds
3. The team asking questions to the client
  1. Question 1: Will the client like the program to be completely software, or potentially for the team to add hardware aspect?
    1. Dr. McCoy said this was up to the team. The ultrasound machines would not be able to have direct access to the internet, but they do have USB ports so it could be possible for hardware aspects. Ultimately Dr. McCoy left it for the team to decide.
  2. Question 2: Would the analysis of the ultrasound images be done in real time, or post patient visit?



1. Dr. McCoy also left this up to the team. If it can be done in real time, there is a slight preference for this, however this is not a deal breaker for Dr. McCoy.
3. Question 3: How often with Dr. McCoy hope to meet with the team?
  1. Dr. McCoy would like to meet with the team again later in the semester. Based on Dr. McCoy having a busy schedule, there is no date set in stone at the moment, but this will be decided upon later.
4. Future Goals
  1. During the discussion of the program, Dr. McCoy mentioned that this project could lead to potential publications and/or IPs.
    1. This also raised questions on who would be listed as first author. This was not a question that Dr. McCoy wanted an answer for, but raised as a hypothetical for the team to consider.
    2. Dr. McCoy also urged the team to look into IP information within the BME department to understand more about what ownership will look like.
5. Takeaways
  1. As the meeting was coming to a closing, Dr. McCoy went over bullet items for what needs to be completed
    1. City training for the team to access the ultrasounds
    2. Dr. McCoy will add the team to her IRB
    3. To research and/or ask PI's how the team can access Pacs (database for ultrasounds)
6. Closing
  1. To close the meeting Dr. McCoy reiterated how exciting having the team work on this machine learning algorithm is. Rheumatology and many other medical areas really could improve with this algorithm
    1. The program will be very broadly applicable and have many important usages in the medical field.

**Conclusions/action items:**

This meeting went extremely well. Dr. McCoy is very enthusiastic about this project and genuinely is looking to help the team succeed. With the large amount of data that Dr. McCoy has, the training process for the program should go relatively well. The only issue will be gaining access to these ultrasound images.

With all the information that the team has gathered, it seems like this project will likely just be a program that will take the ultrasound grey-scale data and analyze it to determine if a patient has Sjogren's or not. This was an important distinction to make, especially because it helps focus the team's focus away from a scoring system and towards solely determining a diagnosis or not.



## 2023/9/15 Advisor Meeting #1

---

YOUSEF GADALLA - Sep 21, 2023, 8:43 PM CDT

**Title:** Advisor Meeting #1

**Date:** 9/22/2023

**Content by:** Yousef

**Present:** Whole team and Dr. Casey

**Goals:**

Meet with advisor and receive any feedback that Dr. Casey has. If the team has any questions to ask them as well.

**Content:**

1. Feedback

1. Dr. Casey mentioned to that the team should continually work on lab archives and to ensure that lab archives are updated individually once a week to show that work is getting completed.
2. The team also shared the list of questions that the team had for the client meeting later that day. Dr. Casey gave feedback on questions to add.

2. Questions

1. The team asked some clarifying questions regarding the PDS
  1. Question asked: Areas of the PDS requirements will not be applying to the software program that the team will be creating. How should the team handle these aspects of the PDS?
    1. Dr. Casey said to answer questions as best as the team can and areas that do not apply to the program can be marked N/A.

**Conclusions/action items:**

This was a very good advisor meeting that allowed the team to get a better understanding about Dr. Casey's expectations for the semester along with how the team can go about meeting those expectations. As the team will meet with Dr. Casey weekly, it will a very good way to continually stay in touch with our advisor and receive support when necessary.



---

Richard YANG - Sep 08, 2023, 4:25 PM CDT

<https://docs.google.com/document/d/1YnjMz-JBiUMdCzwLSay7uplPOHCBam4WCaxRt5XrRfg/edit?usp=sharing>



# Progress Report 9/14/2023

YOUSEF GADALLA - Sep 15, 2023, 11:58 AM CDT

BME Design 200/300

ML: Salivary Gland Ultrasound Scoring

## Machine Learning for Salivary Gland Ultrasound Scoring

9/14/2023

### Team:

Richard Yang ( <a href="mailto:ryang256@uiowa.edu">ryang256@uiowa.edu</a> )	Team Lead
Yousef Gadalla ( <a href="mailto:ygadalla@uiowa.edu">ygadalla@uiowa.edu</a> )	Communicator
Brendon Derw ( <a href="mailto:bderw2@uiowa.edu">bderw2@uiowa.edu</a> )	BSAC
Elmer Nadjari ( <a href="mailto:enadjari@uiowa.edu">enadjari@uiowa.edu</a> )	BWIG
Siva Makujan ( <a href="mailto:smakujan24@uiowa.edu">smakujan24@uiowa.edu</a> )	BWIG
Araa Viswanath ( <a href="mailto:aviswanath3@uiowa.edu">aviswanath3@uiowa.edu</a> )	BPAG

### Problem Statement

Sjogren's syndrome (SS) is a systemic autoimmune disease (SAD) that causes the dysfunction of the exocrine glands (mainly the salivary and lacrimal glands) with patients often showing persistent dryness of the mouth and eyes [1, 2]. According to estimation, two to four million persons in the United States have SS; however, only one million have been diagnosed, likely due to the nonspecific diagnostic guidelines and the heterogeneity nature of the disease [3].

The current standard of care of the client is to perform at least baseline salivary gland ultrasonography (of the submandibular and parotid glands) in SS patients. For some higher risk individuals, regularly scheduled salivary gland ultrasonography is performed. The problem arises within the current COMEFACT ultrasound grading system, which requires subjective opinions and lacks accuracy. As a result, a machine learning approach is proposed to eliminate inter-reader variability and to strive to provide more exact prognostication.

1

[Download](#)**ML\_salivaryGlandUltrasound\_progressReport\_9\_14\_2023.docx (11.1 kB)**



## 09/15/2023 - Sjögren's Syndrome Diagnosis Research

---

Aran Viswanath - Sep 15, 2023, 2:42 PM CDT

**Title:** Salivary gland ultrasonography in primary Sjögren's syndrome from diagnosis to clinical stratification: a multicentre study

**Date:** 9/15/2023

**Content by:** Aran Viswanath

**Present:** N/A

**Goals:** Learn more about salivary gland ultrasonography (SGUS) and the OMERACT scoring system.

**Content:**

OMERACT is a standardized system which is used in rheumatology. This system has been applied to primary Sjogren's syndrome (pSS) and is being used to score ultrasound scans on a scale of 0-3 based on the symptoms of the patient.

According to the article when pSS and non-pSS scans were done of the left and right sides of the parotid and sub mandibular glands, patients with pSS tended to score higher than those with non-pSS conditions. This form of diagnosing pSS is also seen as a less invasive and effective alternative to performing a biopsy

[Article](#)

Zhang, X., Feng, R., Zhao, J. *et al.* Salivary gland ultrasonography in primary Sjögren's syndrome from diagnosis to clinical stratification: a multicentre study. *Arthritis Res Ther* **23**, 305 (2021). <https://doi.org/10.1186/s13075-021-02689-3>

**Conclusions/action items:**

goals:

-Learn about machine learning models

-Questions for client meeting



**Title:** Research on other diagnostic tools for SjS

**Date:** 9/22/2023

**Content by:** Aran Viswanath

**Present:** N/A

**Goals:** Understand current diagnostic tools for SjS to Understand what can be improved.

**Content:** Currently the tests for SjS include Ultrasound exams, blood and urine tests, Schirmer tear test, sialography, salivary scintigraphy, and biopsy. Compared to the other tests, Ultrasound is significantly less invasive as it does not require any incisions or injections. The current downside to ultrasound is that the diagnosis is subjective and prone to human error. In order for our project to provide accurate results as to avoid the need for a patient to go through the other time consuming and invasive procedures.

1. "Blood and urine tests," Johns Hopkins Sjögren's Center, <https://www.hopkinssjogrens.org/disease-information/diagnosis-sjogrens-syndrome/blood-and-urine-tests/> (accessed Sep. 19, 2023).
2. A.-L. Stefanski et al., "The diagnosis and treatment of Sjögren's syndrome," Deutsches Ärzteblatt international, 2017. doi:10.3238/arztebl.2017.0354
3. N. Ohbayashi, I. Yamada, N. Yoshino, and T. Sasaki, "Sjögren syndrome: Comparison of assessments with mr sialography and conventional sialography.," Radiology, vol. 209, no. 3, pp. 683–688, 1998. doi:10.1148/radiology.209.3.9844659
4. I. Umehara, I. Yamada, Y. Murata, Y. Takahashi, N. Okada, and H. Shibuya, "Quantitative evaluation of salivary gland scintigraphy in Sjögren's syndrome," Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine, vol. 40, no. 1, pp. 64–69, Jan. 1999, Accessed: Sep. 22, 2023. [Online]. Available: [https://www.google.com/url?q=https://pubmed.ncbi.nlm.nih.gov/9935059/&sa=D&source=docs&ust=1695404911950832&usg=AOvVaw1wYoNw1\\_pR1FXK9wno3T62](https://www.google.com/url?q=https://pubmed.ncbi.nlm.nih.gov/9935059/&sa=D&source=docs&ust=1695404911950832&usg=AOvVaw1wYoNw1_pR1FXK9wno3T62)
5. "Diagnosing sjogren's syndrome," Patient Care at NYU Langone Health, <https://nyulangone.org/conditions/sjogrens-syndrome/diagnosis> (accessed Sep. 19, 2023).

**Conclusions/action items:**



## 9/29/2023 - Image classification research VGG-19

---

Aran Viswanath - Sep 29, 2023, 11:15 AM CDT

**Title:** Research on VGG-19

**Date:** 9/29/2023

**Content by:** Aran Viswanath

**Present:** N/A

**Goals:** Understand VGG-19 Architecture

**Content:**

Before learning about the specific VGG-19 architecture it was important to understand convolutional neural networks. A convolutional neural network works by first filtering an image. This allows for sharp contrast between different features and better edge detection. Then it is max pooled to reduce image size while retaining important features. Then to prepare it for the neural network, the feature map of the image is flattened into a one-dimensional vector. Through processes such as backpropagation and gradient descent, the neural network is trained, and the filter weights and biases are adjusted.

VGG-19 is a specific CNN architecture which consists of 16 convolution layers using a 3x3 filter and 3 fully connected layers, hence the number 19 in the name. In addition to the convolution and connected layers, there are also max pooling layers in between which have the same function as mentioned above.

This architecture can be applicable to our project as it is very good for feature detection and image classification. The model we design must be able to recognize features that determine Sjogren's syndrome and classify the ultrasound scans.

**Conclusions/action items:**

“Convolutional Neural Networks (CNN),” OpenGenus IQ: Learn Computer Science, Feb. 04, 2019.

<https://iq.opengenus.org/convolutional-neural-networks/>

A. Kaushik, “Understanding the VGG19 Architecture,” OpenGenus IQ: Computing Expertise & Legacy, Feb. 26, 2020.

<https://iq.opengenus.org/vgg19-architecture/>

Action Items:

Understand the application of VGG-19 and how it can be implemented



## 9/29/2023 - BPAG Meeting

---

Aran Viswanath - Sep 29, 2023, 1:10 PM CDT

**Title:** BPAG Meeting

**Date:** 9/2/2023

**Content by:** Aran Viswanath

**Present:** N/A

**Goals:** Learn about purchasing and accounting

**Content:**

This meeting covered how records of purchases must be kept and how reimbursement must take place. Although the information covered in this meeting is not currently applicable to our project, if we do need to purchase something in the future, we will need to discuss with our client whether she will make the purchase, or if I as the BPAG will make the purchase and get reimbursed later. Due to our project being a machine learning algorithm, we do not have any foreseen purchases.

**Conclusions/action items:**







# 9/22/2023 - Method for developing a machine learning model of a neural network for classifying medical images

Aran Viswanath - Sep 22, 2023, 7:12 AM CDT

**Title:** Method for developing a machine learning model of a neural network for classifying medical images

**Date:** 9/22/23

**Content by:** Aran Viswanath

**Present:** N/A

**Goals:** Understand this method and how aspects can be applied to our design


**Content:** This patent describes a method to develop machine learning models to classify medical images including ultrasound. The system comprises obtaining medical images, analyzing them for specific characteristics, classifying them, splitting the dataset into training and validation portions, and conducting model training and validation. The model is stored upon meeting the established performance benchmarks. The method of developing a model can definitely be used in our project as we are developing a model to classify ultrasound images as positive or negative for Sjogren's.

[link](#)

**Conclusions/action items:**

Continue research into machine learning and Sjogren's

Aran Viswanath - Sep 22, 2023, 7:13 AM CDT



United States Patent  
Barry et al.

Patent No.: US 11,017,695 B2  
Date of Patent: May 23, 2021

**ABSTRACT**  
A method for classifying medical images, including: receiving a set of medical images; analyzing the set of medical images to identify a set of features; and classifying the set of medical images based on the set of features.

**BACKGROUND**  
The present disclosure relates to a method for classifying medical images, including: receiving a set of medical images; analyzing the set of medical images to identify a set of features; and classifying the set of medical images based on the set of features.

**BRIEF DESCRIPTION OF THE DRAWINGS**  
FIG. 1 is a block diagram of a system for classifying medical images. FIG. 2 is a flowchart of a method for classifying medical images. FIG. 3 is a block diagram of a method for classifying medical images.

**DETAILED DESCRIPTION**  
The present disclosure relates to a method for classifying medical images, including: receiving a set of medical images; analyzing the set of medical images to identify a set of features; and classifying the set of medical images based on the set of features.

[Download](#)



## 10/17/2023 - Optical Character Recognition - Pytesseract

---

Aran Viswanath - Oct 17, 2023, 8:01 AM CDT

**Title:** Optical Character Recognition

**Date:** 10/17/2023

**Content by:** Aran Viswanath

**Present:** N/A

**Goals:** Find a way to sort the ultrasound images into groups based on anatomical location.

**Content:**

A problem that we encountered was that the images we received are not in the correct format for a machine learning model. In order to focus the learning, we need to split the images into Parotid(P) and Submandibular (SM) gland subgroups. This way a patient's SM scans are graded differently to their P scans. The easiest way that I found to do this was to do an OCR of each image and change the file name to specify the patient, the image number, and the gland of which the image was taken.

One resource I found that could help was pytesseract, an OCR tool to convert pixel data, into the characters that they represent. This way if the image had certain keywords such as "Parotid" or "SM" we could change the file name to organize the data better.

<https://pypi.org/project/pytesseract/>

**Conclusions/action items:**

Figure out how to change the field of view to only display the ultrasound scan.



## 10/17/2023 - DICOM File Reading

---

Aran Viswanath - Oct 17, 2023, 8:18 AM CDT

**Title:** DICOM File Reading

**Date:** 10/17/2023

**Content by:** Aran Viswanath

**Present:** N/A

**Goals:** Figure out how to read DICOM files and convert them into pixel arrays

**Content:**

One problem that I ran into was converting the DICOM file into a useable pixel array. It took very long for me to convert the files into JPEG format so in order to improve the client's ease of use I decided to write a program to do the conversion for them. After looking into the DICOM file tags I realized that a DICOM file is simply a key-value system where a tag is associated with a value such as patient gender, age, and most importantly the pixel array of the scan.

Using the pydicom API I was able to use the `dcmread("filename")` function to convert the DICOM file into a dictionary-like structure. After that finding the pixel array converting the ultrasound scan to a useable image was very simple as it only required the use of the `.pixel_array` function

<https://pydicom.github.io/pydicom/stable/reference/index.html>

**Conclusions/action items:**

- Research box/edge detection as a tool to crop the ultrasound images
- Look into DICOM file to find any information about scan field of view



## 10/17/2023 - File Location and Name Manipulation

---

Aran Viswanath - Oct 17, 2023, 8:27 AM CDT

**Title:** File Location and Name Manipulation

**Date:** 10/17/2023

**Content by:** Aran Viswanath

**Present:** N/A

**Goals:** Learn how to distinguish files by patient and by the gland.

**Content:**

Although it is possible to rename and drag and drop each file into a different folder it is a very menial time-consuming process as there are thousands of images to be classified. Using the Python OS and Shutil module allows for the automation of this process making it much faster and less demanding of me.

OS automates the file renaming as I can provide a few text strings and in integer values to discern the files by patient number, image number, and the gland

Shutil is useful for creating paths for the files to be sent to. I used it to create Parotid and Submandibular folders for the classified images to be sent to.

<https://docs.python.org/3/library/os.html>

<https://www.geeksforgeeks.org/shutil-module-in-python/>

**Conclusions/action items:**

Continue research into automating the cropping of images



## 10/24/2023 - Pytorch Documentation: Converting Grayscale Images Into Tensors

---

Aran Viswanath - Oct 24, 2023, 10:19 AM CDT

**Title:** Converting Grayscale Images to Tensors

**Date:** 10/24/2023

**Content by:** Aran Viswanath

**Present:** N/A

**Goals:** Learn how to convert the sorted and cropped images into tensors

**Content:** In order for the machine learning models to analyze the images, they must be converted into 2d tensors. Normally when using images, the tensor would be 3d with x, y, and rgb dimensions. However due our images being grayscale with each pixel having equal rgb values it is unnecessary to create a 3d tensor. The images, however, are naturally displayed in 3 color channels, so we must convert the image to a one channel grayscale image and then convert the images into a matrix of black/white values. These tensors can then be saved as a comma separated values(csv) file.

[PyTorch documentation — PyTorch 2.1 documentation](#)

**Conclusions/action items:**

1. Save tensors as CSV files
2. - Write a program to take diagnostic data from an excel sheet and match it to its respective tensor.
  1. pyexcel or just csv reading



## 2023/09/15-Sjogren syndrome overview

---

BRANDON DREW - Sep 15, 2023, 2:22 PM CDT

**Title:** Sjogren Syndrome Overview

**Date:** 9/15/23

**Content by:** Brandon Drew

**Present:** N/a

**Goals:** Gain more knowledge about the syndrome

**Content:**

<https://www.mayoclinic.org/diseases-conditions/sjogrens-syndrome/symptoms-causes/syc-20353216>

Sjogren Syndrome is a disorder that affects a person's immune system. Typically this disorder occurs in women over 40. Sjogren syndrome causes the person's immune system to attack the body's cells and tissues. Since Sjogren Syndrome typically targets the glands that make tears and saliva, a person with this disorder might get dry eyes or mouth. However, damage to other parts of the body will have differing effects on a person.

**Conclusions/action items:** Continue to research



## 2023/09/15-Current grading system of salivary gland ultrasounds

---

BRANDON DREW - Sep 15, 2023, 2:34 PM CDT

**Title:** Current Grading System of Salivary Gland Ultrasounds

**Date:** 9/15/23

**Content by:** Brandon Drew

**Present:** N/a

**Goals:** To understand how salivary gland ultrasounds are graded and what this grade tells us

**Content:** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9444027/>

To detect Sjogren Syndrome salivary gland ultrasounds (SGUS) are typically used. The Outcome Measures in Rheumatology Clinical Trials (OMERACT) has created a grading system that ranges from 0 to 3 using SGUS. In this system, 0 represents normal salivary glands. As the score gets higher, the likelihood of Sjogren Syndrome increases.

**Conclusions/action items:** Continue researching





## 2023/09/21-Machine Learning Intro

---

BRANDON DREW - Sep 21, 2023, 6:51 PM CDT

**Title:** Machine Learning Intro

**Date:** 9/21/23

**Content by:** Brandon Drew

**Present:** N/a

**Goals:** To research more about how machine learning works

**Content:**

[https://d2l.ai/chapter\\_introduction/index.html](https://d2l.ai/chapter_introduction/index.html)

This textbook explains the basics of how computer learning works in the introductory chapter. This is to help me get a better understanding of how machine learning works so I can be as helpful as possible for my team.

- As the name suggests, a machine-learning algorithm will learn from experience and improve as time goes on
- The more data the algorithm sees, the more accurate becomes
- Machine learning relies heavily on statistics
- First, machine learning algorithms should be given a large sample of data where the results are known (for us this will be the pre-marked ultrasounds from Dr. McCoy)
- Then, the algorithm should be given new unseen data to test how well the algorithm is predicting the data
- After each set of new data the algorithm gets, we need to adjust the algorithm so it is more accurate in the future
- Machine learning focuses on how to automatically represent data as accurately as possible

**Conclusions/action items:** Continue reading this textbook to help gain a better understanding of how machine learning works



## 2023/09/28 - KNN research

---

BRANDON DREW - Sep 28, 2023, 6:46 PM CDT

**Title:** KNN Algorithm

**Date:** 9/28/23

**Content by:** Brandon Drew

**Present:** N/a

**Goals:** Learn about the KNN algorithm

**Content:**

<https://www.geeksforgeeks.org/k-nearest-neighbours/>

- One of the more basic algorithms
- Works well at recognizing patterns
- Does not make assumptions about data distribution
- Works well at determining which group is most related to a new data point (what our algorithm needs to do)
- Easy to implement (not complex)
- Adapts easily (stores all data in memory which could be a problem for patient confidentiality)
- Has a hard time classifying data when the dimensionality is too high

**Conclusions/action items:** Continue to research and begin to become familiar with python



## 2023/09/29-VGG-19 Research

---

BRANDON DREW - Sep 29, 2023, 2:34 PM CDT

**Title:** VGG-19 Research

**Date:** 9/29/23

**Content by:** Brandon Drew

**Present:** N/a

**Goals:** Learning more about the VGG-19 algorithm

**Content:**

<https://iq.opengenus.org/vgg19-architecture/>

- Consists of 19 layers
- Has 16 convolutional layers
- Has 19.6 billion FLOPS (floating point operations)
- Uses convolutional neural layers to help improve accuracy
- A deep CNN to classify images
- Overall is a good classification algorithm for different types of datasets

**Conclusions/action items:** Begin working on the baseline model with KNN



---

BRANDON DREW - Dec 12, 2023, 10:04 PM CST

**Title:** Citi Training

**Date:** 9/20/23

**Content by:** Brandon Drew

**Goals:**

Complete CITI training so I can gain access to the ultrasounds fro Dr. McCoy

**Content:**

Worked to complete CITI training for Dr. McCoy so I am able to access the ultrasounds

**Conclusions/action items:**

Have completed CITI training



## 2023/09/15\_SalvaryGlandUltrasound

---

DHRUV NADKARNI - Sep 15, 2023, 4:02 PM CDT

**Title:** Salivary Gland Ultrasound in Primary Sjögren's Syndrome: Current and Future Perspectives

**Date:** 09/15/2023

**Content by:** Dhruv

**Present:** N/A

**Goals:** Research Sjögren's Syndrome and the effects on Salivary glands.

**Content:**

[paper link](#)

The paper discusses the usage of Ultrasound image modeling with respect to primary Sjögren's Syndrome in salivary glands. Additionally, the paper uses the OMERACT scoring system, something we will be drifting from.

Useful takeaways from this paper: Understanding the disease a little more and how ultrasound scores the disease.

Imaging is available, and useful for us as we understand how to identify scars and other symptoms of the syndrome on salivary glands.

**Conclusions/action items:**

Ask questions during the client meeting regarding OMERACT scoring system and whether it is what the client wants to continue forwards with.



## 2023/09/15\_MachineLearningForMedicalUltrasound

---

DHRUV NADKARNI - Sep 15, 2023, 4:09 PM CDT

**Title:** Machine Learning For Medical Ultrasound

**Date:** 09/15/2023

**Content by:** Dhruv

**Present:** N/A

**Goals:** Further understand how to relate machine learning with ultrasounds

**Content:**

[paper link](#)

The article provides a step by step approach to how we take the data, label and organize it, and run it through a model to get viable data.

KEYNOTE: Few types of US mentioned. The first being the US imaging, second being US Elastography (strain elastography vs SWE), and third being contrast enhanced US. When building a machine learning model, we need to keep in mind the type of US imaging we will get. We could also provide a labeling system where we can accept all three types and the machine determines which type of US we are looking at prior to salvary gland scoring.

**Conclusions/action items:**

Continue researching Ultrasound + Machine learning and start looking at prior models with Ultrasounds.



**Title:** Endocrine alterations in primary Sjogren's syndrome: an overview

**Date:** 09/15/2023

**Content by:** Dhruv

**Present:** N/A

**Goals:** Research Sjogren's syndrome at it's base within the endocrine system

**Content:**

[paper link](#)

Hypo-activity detected within the endocrine systems for those with pSS. Certain chemicals are low within individuals with pSS --> Could be supplementary with our project as we'll get additional information to help the machine score.

Conversely, some chemicals have been heightened. Can use this information still.

**Conclusions/action items:** Focus more on salivary glands as that is what is requested by client, but relate info from this article to that.

**Title: CT Scan as an Essential Tool in Diagnosis of Non-radiopaque Sialoliths****Date:** 09/20/2023**Content by:** Dhruv**Present:** N/A**Goals:** Understanding CT Scans (Different views of Salivary Glands)**Content:**[paper link](#)

Though not focused on Sjögren's Syndrome, the paper linked provides images of another disease in Salivary glands, Sialolithiasis. It is important cause we need to be able to distinguish diseases of data from Sjögren's scarring (useful with training and testing data for our ML model).

Sialolithiasis is the most common Salivary gland disease and there are cases where it can indicate Sjögren's whilst in others it completely differs from Sjorens.

**Conclusions/action items:**

If we can get access to other disease, then we must also train model with these "false positives" and "false negatives". Once we get access to CT scans, we can begin distinguishing Sjögren's from other diseases.





**Title:** Bilateral multiple sialolithiasis of the parotid gland in a patient with Sjögren's syndrome

**Date:** 09/19/2023

**Content by:** Dhruv

**Present:** N/A

**Goals:** Relating Sjögren's to other diseases within CT scans, and how to distinguish them if any differences.

**Content:**

[paper link](#)

Sjögren's can also display instances of Sialolithiasis, and CT scans show a difference between symptoms of Sialolithiasis vs the common scarring of Sjögren.

White spots indicate instances of Sialolithiasis vs curvyish darker lines indicate the scarring from Sjögren's.

**Conclusions/action items:**

Use this information with machine learning, as we need to also include images with Sialolithiasis as to not incorrectly train the model.



## 2023/09/26 - DesignMatrixResnet50

---

DHRUV NADKARNI - Sep 29, 2023, 1:51 PM CDT

**Title:** Resnet-50 Convolutional Neural Network

**Date:** 9/26/19

**Content by:** Dhruv Nadkarni

**Present:** N/A

**Goals:** Understand Resnet-50, one of our top choices for algorithms.

**Content:**

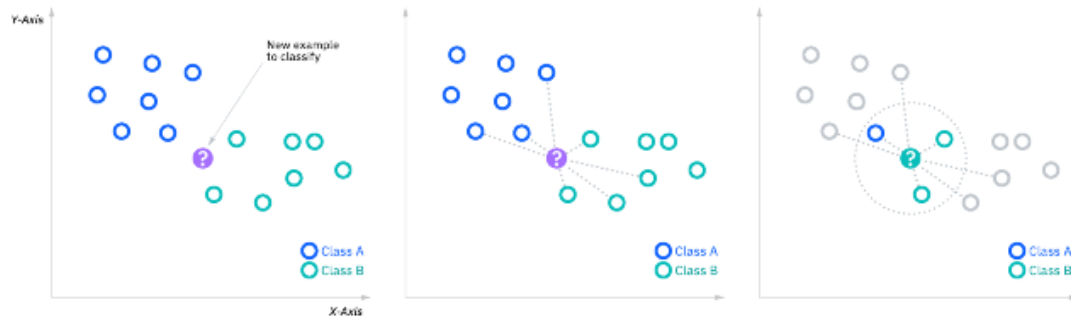
[article](#)

- 50 layers deep (Hence the name)
- Contains pretrained models on their website (we can use as a baseline for our model and work around with it)

**Conclusions/action items:**

Resnet 50 is viable for us, we need to research other algorithms. I personally believe this is the best choice for us in terms of our design matrix factors.

(Edit 9/29 - After reviewing design matrix responses, we are changing final model to VGG19. Will bring it up, but I feel as if accuracy is the only limiting feature of this.. by 1% comparably)

**Title: What is the k-nearest neighbors algorithm?****Date:** 09/26/2023**Content by:** Dhruv**Present:** N/A**Goals:** Understand kNN classifier.**Content:**[article link](#)

- Basic break down of how kNN classifies.

- Various distance metrics and formulas to determine how close a test is to the trained. Could be our go to.

**Conclusions/action items:**

kNN is like the "cliche" for machine learning, but it is still very efficient. I believe this is a good choice for us use in Baseline model.



**Title:** Sklearn KNN

**Date:** 10/05/23

**Content by:** Dhruv Nadkarni

**Present:** None

**Goals:** Learn more about KNN documentation

**Content:**

[article link](#)

The documentation is provided here. It discusses the parameters and attributes of KNN models and also gives some examples as to how to define a model in scikit.

**Conclusions/action items:**

We plan on using Pytorch for KNN, which means we will have to translate Scikit's built-in kNN into Pytorch. This will require familiarity with pytorch documentation as well as a thorough understanding of the distance calculations use to predict in KNN. I have a KNN model for an astronomy project I did a few years ago, so I plan on revisiting that to relearn some aspects of KNN.



## 2023/10/11 - CompleteKNNGuide

---

DHRUV NADKARNI - Oct 12, 2023, 11:36 PM CDT

**Title:** A Complete Guide to KNN

**Date:** 10/10/23

**Content by:** Dhruv Nadkarni

**Present:** N/A

**Goals:** Understand KNN Basics

**Content:**

[article link](#)

Analytics by Vidhya is a great source for beginner machine learning. I have looked at KNN models in the past, as well as created some, but that was with guidance. The contents of this article discuss how to get data, create a model, understand parameters, and much more for KNN.

**Conclusions/action items:**

When actually building the model, refer to SciKit tools documentation for literal code examples but use this article to translate it to a more "human" approach.



## Dhruv's Past Colab Notebooks To Reference

---

DHRUV NADKARNI - Oct 12, 2023, 11:41 PM CDT



[Download](#)

**Astrohunters\_Colab\_Notebooks.zip (1.77 MB)**

---

DHRUV NADKARNI - Oct 12, 2023, 11:42 PM CDT

Content/Goals: A little bit unorthodox, but we can use these files to reference how we formatted code in the past.

Credit: Dhruv Nadkarni & Inspirit AI



## 2023/10/17 - DataProcessingKNN

---

DHRUV NADKARNI - Oct 19, 2023, 10:56 PM CDT

**Title:** Building a KNN model with Sci-Kit

**Date:** 10/17/2023

**Content by:** Dhruv Nadkarni

**Present:** N/A

**Goals:** Understand data processing

**Content:**

[article link](#)

The information in the article demonstrate how to split data accordingly for training and testing data. key note is that we need a "y" set for training as well, which will store our results. This way we can measure accuracy properly.

**Conclusions/action items:**

I will begin splitting the dummy data, and if possible, start splitting the regular data as well.



## 10/30-11/2\_KNN/VGG19 Data Import Examples

DHRUV NADKARNI - Nov 02, 2023, 4:13 PM CDT

#4 Is used for KNN model

#6 is used for KNN/VGG19 models

DHRUV NADKARNI - Nov 02, 2023, 4:04 PM CDT

### MAKE A COPY OF THIS NOTEBOOK SO YOUR EDITS ARE SAVED

#### Instructor Led Discussion

##### Introduction to Breast Cancer Biopsy Classification

In this project, imagine that your colleague, an oncologist (cancer doctor), is working in a major hospital that specializes in treating breast cancers. Breast cancer tumors are very complicated at the cellular level, and this makes determining whether a patient's tumor is malignant (dangerous) or benign (not dangerous) a challenge. Your task will be to build a classifier that can determine whether a sample is malignant or benign to help your colleague!

Every patient that arrives at the hospital undergoes a biopsy of their tumor. This means that a small sample of the tumor is taken from the patient and various metrics are recorded about it, including: radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension.

Using a large dataset of labeled biopsy samples from breast cancer tumors, you will build your binary classification model to determine whether a tumor is malignant or benign based on these features. Then, this model can help you to better determine diagnoses for new patients who arrive at the hospital.

#### Today...

We will explore together the steps that you could take to help your friend solve this problem!

##### Background and data exploration

- Exploring the data
- Visualizing the data

##### Predicting Diagnosis: Working up to Logistic Regression

- Approach 1: Linear Regression classifier
- Approach 2: Simple boundary classifier
- Approach 3: Modifying with logistic regression
- Approach 4: Multiple feature logistic regression

##### Bonus Discussion: What makes a separation good?

Optional: Decision trees walkthrough

#### Background and data exploration

Diagnosing cancer with k-NN

[Download](#)

[4\\_Copy\\_of\\_StudentCopy\\_Beginner\\_Master\\_DoNotEdit\\_Logistic.ipynb \(180 kB\)](#)



## MAKE A COPY OF THIS NOTEBOOK SO YOUR EDITS ARE SAVED

We work for CC: ConscientiousCars, where we help self-driving vehicles be more conscientious of their surroundings. Our cars have been very good at recognizing and avoiding humans. They haven't, however, been capable of recognizing dogs. Since dogs are man's best friend and will always be where we humans are, we want our cars to know if a dog is on the road in front of them and avoid the dog!

The first step to avoiding these cute puppies is knowing if a pupper is in front of the car. So today we will build a detector that can tell when our car sees a dog or not!

In this notebook, you'll

- Explore the cars vs. roads dataset
- Train a simple K-neighbor classifier for computer vision
- Train neural nets to tell dogs from roads
- Improve your model with convolutional neural networks

```

In [ ]:
# Import fun stuff to load some packages and data! (display-mode: "raw")
from sklearn.neural_network import MLPClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn import model_selection
import numpy as np
from sklearn.metrics import accuracy_score
from collections import Counter
import keras
from keras.models import Sequential
from keras.layers import Dense, Conv2D
from keras.layers import Activation, MaxPooling2D, Dropout, Flatten, Reshape
from keras.wrappers.scikit_learn import KNeighborsClassifier
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import cross_val_score
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings("ignore", category=UserWarning)

def load_data():
    # Run this cell to download our data into a file called 'cars_data'
    import os
    # given download https://drive.google.com/uc?id=1d4mXUw0Qd6d6UdAm
    wget -O cars_data https://storage.googleapis.com/aspire-it-ai-data-bucket-1/Data/A

    # now load the data from our cloud computer
    import pickle
    data_file = pickle.load(open("cars_data", "rb"))

    data = data_dict["data"]
    labels = data_dict["labels"]

    return data, labels

```

[Download](#)

[6\\_Copy\\_of\\_Beginner\\_StudentCopy\\_Master\\_DoNotEdit\\_NN\\_ConscientiousCars.ipynb \(142 kB\)](#)



## 2023/11/07-kNNforImageClassification

---

DHRUV NADKARNI - Nov 09, 2023, 2:08 PM CST

**Title:** kNN for Image Classification

**Date:** 11/03/2023

**Content by:** Dhruv Nadkarni

**Present:** None

**Goals:** Import images into models directly with pixel data

**Content:**

[article link](#)

Reading through, it makes sense as to why we chose KNN as our baseline model. The accuracy level is going to be VERY low because kNN image classification is greater with bigger variations (ie Cats vs Dogs... Roads vs Dogs etc...).

Split images into the following:

- Labeled --> Training Testing
- Unlabeled

**Conclusions/action items:**

Need to organize images into separate folders first, and then we can continue with the rest of the model! Actual classification is relatively simple once organized properly.



DHRUV NADKARNI - Oct 11, 2023, 12:21 AM CDT



[Download](#)

**citiCompletionCertificate\_12660810\_58603167.pdf (77.3 kB)**



## 10/12/2023\_ClientMeeting

---

DHRUV NADKARNI - Oct 19, 2023, 7:50 PM CDT

### Objectives

- Give updates to client
- Talk about data
  - Ask her to explain the different types of data?

### Questions

- What do the 'boxed' heatmap images represent?
- Where are the classifications for our data?
- How is the data collected during the trial?
  - Is there text on the screen while an image is taken?
  - Images differ with heatmap sizes, boxes, numbers/text all over the screen...

### Notes

- For images of individuals with a negative for Spores, Dr McCoy has an alternative
  - Will ask her fellows to go through the list and ask to get some negative values.
  - Other option would be to use other US images
- We can use a slide of data via powerpoint
- We can crop the data to only show the image part, ignoring the text.
- She wants at least two images to be fed into the machine
- Don't include the highlighted/boxed images in the module

[Download](#)

**clientMeeting\_10\_13.pdf (55.8 kB)**



## 2023/11/10 - Tong Lecture

---

DHRUV NADKARNI - Nov 10, 2023, 12:26 PM CST

**Title:** Tong Lecture Notes

**Date:** 11/10/2023

**Content by:** Dhruv

**Present:** All Members

**Goals:**

**Content:**

**Conclusions/action items:**



## 9/15/23 Sjogren's Syndrome

---

SIYA MAHAJAN - Sep 15, 2023, 9:04 PM CDT

**Title:** Sjögren's Syndrome

**Date:** 09/15/2023

**Content by:** Siya Mahajan

**Present:** N/A

**Goals:** Research what Sjögren's Syndrome is and how it affects the body

**Content:**

<https://my.clevelandclinic.org/health/diseases/4929-sjogrens-syndrome>

Sjögren's Syndrome is an autoimmune disorder that affects the moisture production in the eyes and the mouth. It can develop in two ways, primary and secondary the former of which is developed independently and the latter of which is developed alongside another autoimmune disorder. There are also listed symptoms and ways to identify this disease that may be helpful further along in our project.

**Conclusions/action items:**

Move on to researching the scoring system/guidelines and how machine learning can be worked into this.



## 9/15/23 Machine Learning

SIYA MAHAJAN - Sep 15, 2023, 9:08 PM CDT

**Title:** Machine Learning

**Date:** 09/15/2023

**Content by:** Siya Mahajan

**Present:** N/A

**Goals:** Research what Machine Learning is and how we can use it in our project

**Content:**

[https://d2l.ai/chapter\\_introduction/index.html](https://d2l.ai/chapter_introduction/index.html)

This textbook goes into a lot of detail about machine learning including the very basics of what it is and the more detailed such as how to implement. At a glance it gave me more context into what exactly a machine learning algorithm is and how it is created.

**Conclusions/action items:**

Research models related to our project that we can use and improve on.



## 9/29/23 Nearest Neighbor

---

SIYA MAHAJAN - Sep 29, 2023, 10:18 PM CDT

**Title:** Nearest Neighbor

**Date:** 09/29/2023

**Content by:** Siya Mahajan

**Present:** N/A

**Goals:** Research different designs that can be implemented in our project

**Content:**

[https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/02\\_knn\\_notes.pdf](https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/02_knn_notes.pdf)

This is a set of notes from UW-Madison's statistics department which goes more in depth about nearest neighbor methods, what they are useful for and helped me both understand what it is and whether it is applicable to our project. This method is useful for pattern classification which is a large part of our project. This method picks a classification based on its nearest "neighbors" and postpones training until it is actually making decisions.

**Conclusions/action items:**

Work on preliminary presentation.

Start working on algorithm for base line model.





# 2023/9/15 Ultrasound as Diagnosis in Sjogren CasesResearch

---

## Title: Research into Ultrasounds as Sjogren's Syndrome Diagnosis

Date: 9/15/2023

Present: Yousef

### Goals:

Learn more about how ultrasounds are typically used in diagnosing Sjogren to better understand how our algorithm/program can be used in clinical settings and potential limitations or considerations that we will need to address.

### Content:

Lorenzon, Michele et al. "Salivary Gland Ultrasound in Primary Sjögren's Syndrome: Current and Future Perspectives." *Open access rheumatology : research and reviews* vol. 14 147-160. 1 Sep. 2022, doi:10.2147/OARRR.S284763

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9444027/>

#### 1. Introduction

1. Primary Sjogren's Syndrome (pSS) is autoimmune disease that is known for causing dryness of the eyes and mouth, and causes the immune system to attack other organs/tissues
2. pSS patients have an increased risk of "lymphoproliferative diseases"
  1. These are diseases in which lymphocytes are uncontrollably produced
3. Diagnosis in clinical settings has been using salivary gland ultrasound with increased usage
  1. Due to noninvasiveness and inexpensiveness
4. Ultrasound able to see abnormalities in salivary glands
  1. Scoring might involve fatty infiltration, visible posterior border
5. Scoring systems mentioned American College of Rheumatology European League Against Rheumatism (ACR-EULAR) and OMERACT
  1. OMERACT is the one mentioned by the client via project description
6. OMERACT scoring breakdown
  1. scoring based on hypoechoic areas in salivary glands.
  2. Scoring is on scale of 0-3 (0 being no hypoechoic areas and 3 being entire gland surface covered).
7. Ultrasounds of salivary glands of pSS patients might have similar symptoms to other diseases.
  1. Will need to clarify with client whether we need to consider other diseases with other symptoms

#### 2. Lymphomas

1. serious complication of pSS
  1. typically non-Hodgkin's lymphoma (NHL) specifically
  2. NHL risk is 4-40 fold in pSS patients
    1. OMERACT score of 3 is more highly associated with NHL

#### 3. Monitoring Disease Activity with Ultrasounds

1. Scores are used as a way to view disease progression through comparison to earlier sonographs
  1. this process, as the time this paper was published, is still under investigation
2. Ultrasound scores and pSS organ damage association detected.
3. Ultrasounds can also detect physical changes in salivary glands





# 2023/9/21 Machine Learning in Cardiovascular disease

---

## Title: Machine Learning in Cardiovascular Disease research

Date: 9/21/2023

Content by: Yousef Gadalla

### Goals:

Learn more about existing machine learning usages in the medical field to have a better understanding of how it can be utilized in rheumatology

### Content:

Al'Aref, Subhi J et al. "Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging." *European heart journal* vol. 40,24 (2019): 1975-1986. doi:10.1093/eurheartj/ehy404

#### Introduction

1. ML (machine learning) has been used in the radiology and found to be as effective/accurate as radiologists
2. In cardiology settings, ML found to be "more adept in the prediction of either cardiovascular or all-cause mortality than clinical or imaging modalities used separately"
  1. Means that ML is having better success at predicting mortality than typical medical schemes when separately looked at

#### Machine Learning Background

1. ML's basis will allow it to predict outcomes based on previous data, but only within a given problem/field
  1. No capability of general intelligence
2. ML used to start with recognizing written numbers
3. Common ML algorithms can be supervised, unsupervised, or reinforcement
  1. supervised utilizes learning from pre-labeled data (similar to what the team's program will do)
  2. unsupervised has the program analyze unlabeled data and group similar data
  3. reinforcement learning is much like operant conditioning (ie reward based learning)

#### Data

1. The data sets typically have many subsets of information within them
  1. example given is medical imaging and clinical outcome
2. Data can be binarized to make the data easier to understand for the algorithm
  1. During this step, some data transformation occurs like removal of outliers'

#### Algorithms

1. With supervised algorithms, they have to main capabilities
  1. Classifying something by some given label
  2. Predicting some valued output
2. These algorithms are relatively similar in coding and don't have much difference between them
3. Current ML algorithms have at least one of the following
  1. linear/log regression
  2. artificial neural networks
  3. support vector machines
  4. tree based methods

4. Training of ML algorithms usually uses most of the data, minus some to be used later on for testing and slightly adjusting the learned parameters
5. Another ML algorithm piece that is really good for image analysis is convolutional neural networks
  1. uses many filters to find patterns between images/data
  2. become a "standard" for most cardiovascular imaging ML
  3. Has been issues in the past of mislabeling a given image that was pixley

#### Performance metrics and model refinement

1. It is important to set a benchmark goal for the program early on in order to gauge success/failure
  1. Will allow for easier time improving the program
2. It can be hard to set a good benchmark/goal in some cases where the program is looking for a very rare case
  1. Having such a rare reading will not necessarily mean the program is working as it could be misclassifying data
  2. medical papers use "receiver-operating characteristic curve" as their way of validating use of ML

#### Domains of Application

1. Despite literature looking at ML in cardiology, there isn't an agreed upon area of cardiology that ML should be used. Proposed areas by the paper include
  1. ECG
  2. 2DE
  3. CCTA
  4. SPECT
2. No mention of ultrasound (could be ultrasounds not typically used in cardiology imaging)

#### Conclusions/action items:

This paper highlighted some really good algorithms that we can utilize in our Design Matrix and potentially within the actual program. It seems like the best algorithm to use would be the convolutional neural network, as this is the algorithm that is used in image based ML. That being said, more research should be conducted into the CNN to make sure that it will actually be good for our program.

The idea of setting a benchmark early on in the process will be good for training/updating the algorithm. It should be relatively easy to set a benchmark for how accurate the ML should be in the training/testing phase. This will allow us to easily decipher if the algorithm we are using is working within our desired boundaries of accuracy or if we need to make adjustments to the algorithm to improve its accuracy.

## Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging

Sahbi J. Al'Aref<sup>1†</sup>, Khalil Anchochea<sup>1†</sup>, Gurpreet Singh<sup>1</sup>, Piotr J. Skonka<sup>2</sup>,  
 Kranthi K. Kollu<sup>3</sup>, Amit Kumar<sup>4</sup>, Mohit Pardey<sup>5</sup>, Gabriel M. Sakala<sup>6</sup>,  
 Alexander R. van Rosendaal<sup>7</sup>, Ashley N. Beece<sup>8</sup>, Daniel S. Berman<sup>9</sup>,  
 Jonathan Leipaic<sup>10</sup>, Koen Nieman<sup>11</sup>, Daniele Andreini<sup>12</sup>, Gianluca Pontone<sup>13</sup>,  
 U. Joseph Schoepf<sup>14</sup>, Leslie J. Shaw<sup>15</sup>, Hyun-jae Chang<sup>16</sup>, Jagat Narula<sup>17</sup>, Jeroen J. Bax<sup>18</sup>,  
 Yuanfang Guan<sup>19</sup>, and James K. Min<sup>19\*</sup>

<sup>1</sup>Department of Radiology, Mount Sinai Hospital and Mount Sinai Medical Center, New York, NY, USA; <sup>2</sup>Department of Integrated Medicine and Biomedical Sciences, California State University Long Beach, California, USA; <sup>3</sup>Department of Nuclear and Biomedical Engineering, University of North Carolina, Charlotte, NC, USA; <sup>4</sup>Department of Cardiology and Imaging, Stanford University School of Medicine and Cardiovascular Institute, Stanford, CA, USA; <sup>5</sup>Cardio-Catheterization, Mount Sinai, Mount Sinai Health System, Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>6</sup>Department of Cardiology, University of California, San Diego, San Diego, CA, USA; <sup>7</sup>Department of Cardiology, University of California, San Diego, San Diego, CA, USA; <sup>8</sup>Department of Cardiology, University of California, San Diego, San Diego, CA, USA; <sup>9</sup>Department of Cardiology, University of California, San Diego, San Diego, CA, USA; <sup>10</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>11</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>12</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>13</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>14</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>15</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>16</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>17</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>18</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA; <sup>19</sup>Department of Cardiology, Mount Sinai Hospital, New York, NY, USA

Received 21 March 2023; revised 21 May 2023; editorial decision 22 June 2023; accepted 14 July 2023; online publication 17 July 2023

Artificial intelligence (AI) has been shown to have a wide range of applications in medicine. Machine learning (ML), which is a subset of AI where machines use data to learn from, has been used to analyze large amounts of data to identify patterns and predict outcomes. In this review, we present a brief overview of ML methodologies that are used for the construction of inferential and predictive data-driven models. We highlight several domains of ML application such as electrocardiography, echocardiography, and recently developed non-invasive imaging modalities such as coronary artery calcium scoring and coronary computed tomography angiography. We conclude by reviewing the limitations associated with contemporary application of ML algorithms within the cardiovascular disease field.

**Keywords:** Machine learning • Cardiovascular disease • Coronary computed tomography angiography • Echocardiography

### Introduction

Machine learning (ML), an extension of the increasingly long quest for artificial intelligence (AI), has altered our collective conception of information and its seemingly boundless potential for guiding change. Machine learning is broadly defined as the ability of a system to autonomously acquire knowledge by analyzing patterns from large data sets. This field has spurred tremendous innovation in all sectors of the technology industry, from speech recognition and sentiment analysis to spam filters, chessbots and autonomous

driving. While the adoption of ML in the information technology sector is nearly ubiquitous, its introduction into the medical field has been much more subdued. The landscape, however, is rapidly changing. Equipped with novel ML frameworks, increasing computational power and the availability of big data, the ML community is now concentrating its efforts squarely at complex tasks in the healthcare sector. These efforts have borne their fruits, for example, in radiology, where an ML platform has been demonstrated to be as effective as a human radiologist in radiating pneumonia diagnosis<sup>1</sup> in pathology, ML has outperformed existing

\*Correspondence: James K. Min, MD, PhD, jkmin@mskcc.org  
 †The two authors contributed equally to the content of this manuscript.  
 Received editorial of the European Society of Cardiology 10/10/2023. For permissions, please email: journals.permissions@oup.com

[Download](#)

**ML\_CardiologyPaper.pdf (788 kB)**



# 2023/9/29 Machine Learning in Medicine Research

---



## Title: Machine Learning in Medicine

Date: 9/29/23

Content by: Yousef

### Goals:

Learn more about machine learning's applications in the medical field and learn more about algorithms used in the medical field.

### Content:

#### 1. Supervised Learning

1. Learning is through analyzing known outputs/target values
  1. Examples listed are classifying objects, recognizing handwriting, etc)
2. As these are tasks that people can do, the machine is mimicking human performance.
3. SL focuses on classification
4. Medical examples
  1. EKG interpretation
    1. Pattern recognition is used to compare observed data to list of possible diagnoses
    2. Automatic chest x-ray detecting lung nodules
5. SL also used to estimate risk
  1. coronary heart disease risk score calculated by some ML algorithms.

#### 2. Learning Problem

1. learning problem exists that has limited the presence of ML in the clinical field
  1. Typically can be hard to distinguish features that are relevant as they depend on many other factors/features to decide if they warrant a diagnosis
  2. How do we connect measured values to predicted outcomes
    1. Linear regression, logistic, many different relation types
  3. High performing models need many attributes for success
    1. descriptions to differentiate between similar illnesses/cases

#### 3. Supervised learning - learning from forests and trees

1. "one of the best 'off the shelf' algorithms" for baseline readings
  1. Uses decision trees to go through each individual value until a diagnosis is obtained
    1. problem is that each tree has only limited training examples

#### 4. C path

1. uses automated image processing
2. this algorithm was first used in analyzing breast images to determine whether the breast tissue was more epithelial vs stromal tumor sections
  1. along with 6000 other predictors

#### 5. Attractor metagenes in cancer and bake offs in machine learning

1. ML "bake offs" became a big thing for the generation of different algorithms
  1. teams would be given the same dataset and tasked with coming up with an algorithm
2. One such bake off was the Sage Bionetworks -DREAM Breast Cancer Prognosis Challenge
  1. The winning model used a mix of unsupervised and supervised learning
  2. Utilized cancer biomarkers that had been seen across several cancer types to predict potential cancer cases

## Conclusions/action items:

This paper gave a lot of valuable information about various competing designs to ML in the medical field. This also shows many other algorithm types that have been used with relatively good amounts of success. The C path algorithm should be researched more as it uses automated image processing, which is likely something that our machine learning algorithm will do. Learning more about how C path does this might help us gain a better understanding of the process.

YUSEF GADALLA - Sep 29, 2023, 12:50 AM CDT

Author Manuscript



**HHS Public Access**  
 Author Manuscript  
 Circulation. Author manuscript; available in PMC from March 01, 2024.

Published in final edited form as:  
 Circulation. 2017 November 17; 135(20):1920-1930. doi:10.1161/CIRCULATIONAHA.115.061593.

### Machine Learning in Medicine

**Rehal C. Des, MD, PhD**  
 Cardiovascular Research Institute, Department of Medicine and Institute for Human Genetics, University of California, San Francisco, and California Institute for Quantitative Biosciences, San Francisco, CA

**Abstract**

Spurred by advances in processing power, memory storage, and an unprecedented wealth of data, computers are being asked to tackle increasingly complex learning tasks, often with astounding success. Computers have now mastered a popular variant of poker, beaten the game of Go, played from experimental data, and become experts in video games – tasks which would have been deemed impossible not too long ago. In parallel, the number of computers contained on applying complex data analysis to varying industries has exploded, and it is thus surprising that more analytic computers are solving numerous problems in healthcare. The purpose of this review is to explore what problems in medicine might benefit from such learning approaches and use examples from the literature to introduce basic concepts in machine learning. It is important to note that, so far, the majority of large medical data sets and adequate learning algorithms have been available for many decades – and yet, although there are thousands of papers applying machine learning algorithms to medical data, very few have contributed meaningfully to clinical care. This lack of impact stands in stark contrast to the common relevance of machine learning to many other industries. Thus part of my effort will be to identify what obstacles there may be to changing the practice of medicine through statistical learning approaches, and discuss how these might be overcome.

**Keywords**

computers; statistics; risk factor; prognostic machine learning

Machine learning is the scientific discipline that focuses on how computers learn from data.<sup>1,2</sup> It arises at the intersection of statistics, which seeks to learn relationships from data, and computer science, with its emphasis on efficient computing algorithms. This marriage between mathematics and computer science is driven by the unique computational challenges of learning: instead of models from massive data sets, which can take billions or trillions of data points. The types of learning used by computers are conventionally subdivided into categories such as supervised learning and unsupervised learning. However, I find, in addition, that another vision can be useful when considering how machine learning might inform the practice of medicine: distinguishing between these tasks

Address for Correspondence: Rehal C. Des, MD, PhD, South Cardiovascular Research Building, 101 Mission Bay Boulevard, Room 4125, San Francisco, California, 94133, USA. Tel: 415-476-0760; Fax: 415-753-3228; rehal@card.ucsf.edu.  
 Conflict of Interest Declaration: None.

[Download](#)

**ML\_inMedicin.pdf (1.21 MB)**



## 2023/10/9 UNet Research

---

YOUSSEF GADALLA - Oct 11, 2023, 9:37 PM CDT

**Title:** Unet Algorithm Research

**Date:** 10/9/2023

**Content by:** Yousef

**Goals:** Have a really good understanding of UNet for the preliminary report and find research to add to said report

**Content:**

- UNet is type of CNN
  - first applied in 2015 for Biomedical images
  - dedicated to finding area of abnormality for disease distinguishing/diagnosis
    - analyzes every pixel
- Two major parts for architecture
  - 1st - contracting path
    - general convolution processes
  - 2nd - expansive path
    - 2d con. layers
- Follows kernel strategy
  - data transformation
- Contracting path
  - has three rounds of 2 layered con. processes followed by "max pooling process"
    - max pooling essentially resizes image to be 1/2x smaller
- Expansive Path
  - Image will be resized upwards while analyzing each small image/pixels

**Conclusions/action items:**

UNet seems to be a really good algorithm for biomedical image analysis. The fact that it can locate the abnormalities makes it seem like it could work really well within our project. I think this is a good second algorithm to consider if we finish the VGG-19 algorithm relatively quickly and we want to analyze a second algorithm. Further research should focus on UNet and potential limitations that are present in the algorithm. This paper seems very supportive of UNet and doesn't display any negatives.



## 2023/10/9 VGG-19 For COVID-19 paper

YOUSEF GADALLA - Oct 11, 2023, 9:55 PM CDT

**Title:** VGG-19 analysis for COVID identification using X-ray images

**Date:** 10/9/2023

**Content by:** Yousef

**Goals:**

Gather a better understanding of VGG-19 for writing about it in the preliminary report and have sources to cite.

**Content:**

### 1. Intro

1. Due to the COVID breakout, there has continued research into the illness
2. X-rays and CT scans are very important in "global effort against Covid.
3. To counter limitations in data, a triple classification method is considered.

1. Batch normalization + dropout layers+VGG-19

### 2. Materials and Methods

#### 1. Batch Normalization (BN)

1. reduces data variation within nodes of deep network
2. speeds training process
3. BN calculates sample mean and variance
  1. These values used for scaling, translation which allowed for training at higher rates by looking at feature distribution

#### 2. Dropout

1. strategy for limiting over fitting
2. works by slimming the current network

### 3. Data set

1. 200 positive cases
2. 1000 negative cases

### 4. Results

1. VG-19 with BN + D preformed better than VGG-16 with BN + D
2. VGG-19 scored 83% with COVID cases, 95% with negative cases, and 90% with pneumonia
  1. pneumonia used to test VGG-19's ability to look at several things at once.

**Conclusions/action items:**

VGG-19 seems like a really good algorithm for medical imaging and disease classification (at least with X-rays). If I can find further research that talks about VGG-19 with ultrasounds that would be really good because we can learn how those papers utilized different layers to improve the accuracy of their algorithm.

I think further research should be done to look at Batch normalization and dropout because they seemed to really improve the accuracy of their algorithm.



# 2023/12/9 CNN Overview Gradient

---

**Title: CNN Parameter overview (Gradient)****Date:** 12/9/23**Content by:** Yousef**Goals:**

To have a better understanding of the specific parameters for explanation in the final report (specifically gradients)

**Content:**

N. Cui, "Applying Gradient Descent in Convolutional Neural Networks" *IOP Conf. Series: Journal of Physics: Conf. Series* 1004 2018. doi: 10.1088/1743-6596/1004/1/012027

1. Loss function
  1. used to determine the loss in relationship to the specific weights
2. Optimizer and Gradient
  1. The optimizer is used to minimize the loss function. The specific optimizer we'll use has the gradient method.
  2. The gradient is the partial derivative of the error divided by the weights. By looking at the graph of this gradient, we are able to understand how the weights correlate with the error we see in the machine learning algorithm
  3. Checking every combination of weights would take a very large computational time and hardware. Instead, there is a pre-set learning rate that will determine the adjustment of the weights.
    1. learning rate \* |dError/dWeight| = step size
    2. The step size is how much we change the weight in order to the change of the gradient
    3. If the gradient is negative, we much increase the weights, and if the gradient is positive we decrease the weights
  4. Determining the best learning rate is important because it effects the overall speed and accuracy of the algorithm
    1. The goal is to find the minima of the loss function. This will be when the gradient is zero.
    2. Having too small of a learning rate will require a large amount of calculations that could take a long time
    3. Having too large of a learning rate would cause the gradient to step over its zero value. This could cause oscillations of the gradient from going positive to negative without actually getting closer to the zero value.
  5. For classification, the most used used loss function is the cross entropy that is used for classification between several possible classifiers.

**Conclusions/action items:**

The gradient and learning rate will be essential aspects of the algorithm because they can not only determine accuracy, but also speed. Through various testing of different learning rates, we can look at how the accuracy changes to see the best learning rate.



## 2023/12/9 CNN Overview (Backpropagation)

---

YOUSEF GADALLA - Dec 13, 2023, 11:44 AM CST

**Title:** CNN Parameter overview (Backpropagation)

**Date:** 12/9/23

**Content by:** Yousef

**Goals:**

Understand how backpropagations work in machine learning

**Content:**

A. Roy, "An introduction to Gradient Descent and Backpropagation," Towards Data Science, <https://towardsdatascience.com/an-introduction-to-gradient-descent-and-backpropagation-81648bdb19b2> (accessed 12/9/2023)

**Conclusions/action items:**



## 2023/9/24 CITI Training

---

YOUSEF GADALLA - Sep 29, 2023, 12:02 AM CDT

**Title:** Citi Training

**Date:** 9/24/23

**Content by:** Yousef Gadalla

**Goals:**

Complete required CITI training in order to gain access to de-identified ultrasound images

**Content:**

Worked on the CITI training modules as per Dr. McCoy's request. These modules revolved around protocols and policies that surround any Human Subject Research. Modules completed were Human Subjects Research Infrastructure, investigator Responsibilities, Defining Research with Human Subjects, History and Ethical Principles, Informed Consent, Assessing Risk, Populations in Research, Research with Decisionally Impaired Subjects, Consent with Subject Who Do Not Speak English, and a module on UW Madison policies.

**Conclusions/action items:**

Now that the trainings have been completed, the next steps can be taken to get added to the IRB protocol and move the project one step closer to gaining access to the data. Follow up goals should be to communicate with the team to ensure that they have also finished Citi Training and then email Dr. McCoy to inform her that the team has completed trainings.





## 2023/9/27 PyTorch Lessons Tensors

---

YOUSEF GADALLA - Sep 29, 2023, 12:12 AM CDT

**Title:** PyTorch Lessons

**Date:** 9/27/23

**Content by:** Yousef Gadalla

**Goals:**

Learn more about Machine Learning by completing PyTorch lessons/modules

**Content:**

1. Unit 1: Tensors
  1. Arrays that are able to store data (data can be nonspecific data type that the tensor will infer)
  2. import statements
    1. import torch
    2. import numpy as np
  3. tensor objects hold three properties based on their shape, datatype, and storage device
    1. similar to objects based coding in java and how the objects can have their own specific variables
  4. tensor operations
    1. tensors can have many different operation types done on them
      1. Switching to GPU will make these operations be done at faster speeds
        1. "if torch.cuda.is\_available(): tensor = tensor.to("cuda")
    2. Operations include indexing (and changing specific indices), addition, combining tensors, etc

**Conclusions/action items:**

**With this knowledge of tensors, I think I have a greater understanding of how they can be used within machine learning. We might be able to use tensors to store the available data such that we have tensors within tensors. The nested tensor will hold the ultrasound image and the scoring of said image, while the out tensor holds this list of patients/tensors of data.**

**Next step will be to continue learning pytorch information**



## 2023/9/29 PyTorch Data Loading

---

YOUSEF GADALLA - Sep 29, 2023, 12:25 AM CDT

### **Title: PyTorch Data Loading**

**Date:** 9/29/2023

**Content by:** Yousef

### **Goals:**

Learn how to load in data

### **Content:**

1. Loading datasets
  1. the code for loading in data takes four inputs
    1. root gives file location of data
    2. train is a boolean for whether this is test data or training data
    3. download is a boolean that downloads data from the internet if root error occurs
    4. tranform is the function/method to be used on the data
      1. their example is to use ToTensor, which would load the data into a tensor
  2. Creating your own dataset
    1. requires three functions
      1. `_init_`
        1. This function will be run to initiate the dataset object/tensor
      2. `_len_`
        1. returns sample number in dataset
      3. `_getitem_`
        1. grabs an item from an inputed index
  3. Preping Data for training
    1. use DataLoader method to iterate through and retrieve one sample at a time. This will be good for training.
      1. allows for "minibatches"/sets of data to train the machine

### **Conclusions/action items:**

This data in loading data and some of the code used to actually go through one sample at a time will be very beneficial to have when it comes time to test the algorithm and train it. This will be a good lesson to review as we get closer to the coding of the algorithm and the testing phase of the design process.

Next step will be to learn more about how we can access the data and how to get that into the machine learning algorithm. Another step will be to continue with PyTorch lessons



## 2023/9/15 Team Meeting #2

---

YOUSEF GADALLA - Sep 15, 2023, 12:50 PM CDT

**Title:** Team Meeting #2

**Date:** 9/15/2023

**Content by:** Yousef Gadalla

**Present:** Whole team

**Goals:**

1. Prepare document for client questions
2. Overview of PDS
  1. Divide work for the PDS

**Content:**

1. Client Meeting overview
  1. Created a list of questions for the client
  2. Everyone added any questions they can think of that will be beneficial to asking the client
2. Machine learning discussion
  1. Should focus on how machine learning works for research
  2. Research could also look at data augmentation/transforming to extrapolate data
  3. dtl.ai is a good website to learn more about machine learning and get practice with it in smaller projects
    1. similar to khan academy
3. PDS overview
  1. Went over PDS as a document, and what we need to do for the assignment
    1. Discussed guideline document
    2. Plan to divide up work based on information gathered from client meeting
      1. If there are no hardware aspects and the project is just coding, there will be several PDS sections that are not applicable.
  2. Plan to divide work post client meeting

**Conclusions/action items:**

This was a pretty helpful team meeting for us to get together and go over expectations for the next week. We were also able to generate a document of client questions that will be very helpful in gaining more information about the project and the client's expectations.



## 2023/9/8 Team Meeting #1

---

YOUSEF GADALLA - Sep 15, 2023, 1:01 PM CDT

**Title:** Team Meeting #1

**Date:** 9/8/2023

**Content by:** Yousef

**Present:** Whole team

**Goals:**

1. Introductions
2. Decide team roles
3. Take team picture
4. Plan meeting times for next week/future

**Content:**

**1. Introductions**

1. **The team went around introducing themselves to each other and learning more about their background knowledge in the field**

**2. Decide team roles**

1. **The team decided on our team roles and put them into the project information document.**
2. **At this point, the team shared contact information**

**3. Team Picture**

1. **The team picture was taken and was uploaded to team website**

**4. Plan Meeting times for future**

1. The team decided to meet at the scheduled BME 200/300 time of 12:00 pm

**Conclusions/action items:**

This was a good team meeting that allowed for the team to get to know one another and created a good starting point for the year. We will all focus on researching the topic in general, and next week we will discuss more on the project and how we will tackle solving the problem the client introduced to us.

An action item for Yousef will be to email the client, Dr. McCoy, in order to set up a client meeting.



## 2023/11/10 Tong Lecture

---

YOUSEF GADALLA - Nov 10, 2023, 12:41 PM CST

**Title: Tong Lecture: One Engineer's Journey - Where preparation meets opportunity**

**Date:** 11/10/2023

**Content by:** Yousef

**Goals:**

**Content:**

1. Education

1. When to University of Pit for Undergrad

1. originally had wanted to be a doctor, but came across BME and enjoyed it a lot more

2. UW-Madison for grad school

1. learned how to concisely and properly present scientific data/info

2. wanted to focus on utilizing technology to improve quality of life

3. Should focus on your story

3. Med school

1. Didn't match for residency

1. Learn to pivot

2. Main points

1. Find your people

2. Do things that scare you

3. laugh until you cry

**Conclusions/action items:**



## 2023/09/10\_mortalityRiskFactorsInSJS

---

Richard YANG - Sep 10, 2023, 2:06 PM CDT

**Title:** mortality risk factors in primary sjogren syndrome

**Date:** 09/10

**Content by:** Richard

**Present:** NA

**Goals:** go through select articles to gain more knowledge on the subject

**Content:**

[article link](#)

The article focuses on baseline predictors that are involved in mortality in patients with primary SJS syndrome.

ESSDAI score: Euler Sjogren's syndrome disease activity index \_

Eight all-cause-of-death related variables are identified, ocular and oral tests, salivary biopsy, ESSDAI, ANA, anti-Ro, anti-La, and cryoglobulins

Five SS-death related variables are identified, oral tests, clinESSDAI, DAS-ESSDAI, ANA and cryoglobulins

Key mortality factors at the time of SS diagnosis are positive cryoglobulins and a high systemic activity scored using the ESSDAI

**Conclusions/action items:**

continue research



## 2023/09/10\_transferLearningLatentFactors

---

Richard YANG - Sep 10, 2023, 2:22 PM CDT

**Title:** Using transfer learning-based causality extraction to mine latent factors for Sjogren syndrome from biomedical literature

**Date:** 09/10

**Content by:** Richard

**Present:** NA

**Goals:** read up on literature

**Content:**

[article link](#)

using transfer learning techniques, a model trained on several datasets, an ELECTRA-based sentence-level relation extraction model can be used to extract causal relations and build a causal network with high precision and recall values.

**Conclusions/action items:**

Although this is not directly related to this project, this article presents a highly practical tool for future research in all medical databases and can significantly improve the quality-of-life, speeding up research process

continue research







**Title:** OMERACT Grading system

**Date:** 09/12

**Content by:** Richard

**Present:** NA

**Goals:** find out the current standard for physicians to diagnose SjS

**Content:**

[article link](#)

Score 0: normal glandular parenchyma in the absence of alterations;

Score 1: presence of fine echogenicity in the absence of clear alterations, or slight, diffuse glandular hypoechogenicity, mild glandular alteration;

Score 2: presence of focal hypochoic areas, but partial conservation of normal glandular parenchyma, moderate glandular alteration;

Score 3: diffuse presence of hypochoic areas in the absence of normal glandular parenchyma, or the presence of glandular fibrosis, severe glandular alteration.

echogenicity: the brightness of an image caused by the reflection of soundwaves and is influenced y sound beam characteristics and tissue density

**Conclusions/action items:**

the scoring system is highly subjective, reflecting the need for a more objective scoring system.



Title: Application of OMERACT in Sjs

Date: 9/15

Content by: Richard

Present: NA

Goals: Understand the practical application of OMERACT

Content:

Greyscale ultrasound of the parotid and submandibular gland showing good 0.91 specificity, though sensitivity of 0.72 can be improved.

Conclusions/action items:

This result can be used as a baseline for the performance of our model

**Application of the OMERACT Grey-scale Ultrasound Scoring System for salivary glands in a single-centre cohort of patients with suspected Sjögren's syndrome**

Victoria Faria, Uffe M Dohl, Simon Krabbe, Lena Terslev

**ABSTRACT**  
 Aim: To evaluate salivary gland involvement in suspected Sjögren's syndrome (SS) using the OMERACT Ultrasound Scoring System for SS. First, interobserver ultrasound cut-off, to assess the performance of the scoring system for diagnosis and then if OMERACT for all conclusions.

**Methods:** All patients referred to our department with a suspicion of SS in 12 months period were included. All underwent grey-scale ultrasound of the parotid and submandibular gland (PM) by 10 trained physicians (including subspecialty analysis). Label biopsy was performed according to physician judgement. Images of the four glands were scored 0-3 according to the scoring system and a consensus score was obtained using a three-physician panel.

**Results:** Of the 134 patients included in the analysis, 43 were diagnosed as primary SS according to the 2016 American College of Rheumatology (ACR)/European Scleroderma Study Group (ESG) criteria. Compared with the SS patients, the SS patients had a higher prevalence of SS (13% vs 2.0%, p=0.01). In patients with SS, the prevalence of SS was significantly higher than in patients without SS. Schirmer's test and positive label biopsy can predict SS with a sensitivity of 0.72 and a specificity of 0.91.

**Conclusion:** The OMERACT Ultrasound Scoring System showed good specificity (0.72) and excellent specificity (0.91) for a label biopsy. However, the sensitivity (0.72) was not as high as expected. Our data supports the use of ultrasound for the primary SS and suggests the need of ultrasound in the classification criteria.

**Key messages**  
 What is already known about this subject? Ultrasound is a non-invasive tool for evaluating organ-specific changes of the large salivary glands for Sjögren's syndrome.  
 What does this study add? The selected consensus-based OMERACT Scoring System has an excellent specificity for diagnosing Sjögren's syndrome in suspected patients.  
 How might this impact on clinical practice or future developments? The OMERACT Scoring System and other easy-to-use tools might be used to evaluate salivary glands in future care and support the use of ultrasound for diagnosing Sjögren's syndrome.

SS is related to primary SS (pSS), which is not associated with other connective tissue diseases (CTDs), and secondary SS, which is associated with other CTDs. It can be challenging to establish the diagnosis, as it is not based on a single component but on a combination of symptoms, decreased function of exocrine glands, autoantibodies and demonstration of lymphocytic infiltration on label salivary gland biopsy. The biopsy findings may be insensitive or inconclusive.

In cross-sectional studies, classification criteria were published in 2010 in a collaboration between American College of Rheumatology (ACR) and European Scleroderma Study Group (ESG). These criteria take into account several features with different weights: label biopsy, autoantibodies, ocular staining test, Schirmer's test and unstimulated salivary flow rate. Label biopsy and autoantibodies (anti-SSA, anti-SSB) and anti-SSB (anti-Ro) and anti-SSA (anti-La) have the highest weight. Label biopsy is an invasive procedure with potentially severe side effects such

Download

e001516.full.pdf (802 kB)





## 2023/09/21\_dataAcquisition

---

Richard YANG - Sep 21, 2023, 4:34 PM CDT

**Title:** acquisition of additional human data

**Date:** 09/21

**Content by:** Richard

**Present:** NA

**Goals:** find the CFR standard for the acquisition of human data

**Content:**

though most of the data is provided to us, it is nevertheless important to know the rules concerning the acquisition of these data, which may be helpful in the future.

The data acquired must fall under 505(i) or 520(g) for prior submission, or the results of which are intended to be submitted to the FDA at a later date for inspection. FDA defines a human subject as an individual who is or becomes a participant in this project and is either subject to the test article or as a control.

furthermore, informed consent must be obtained from participants, which is outlined [here](#) by the FDA guidelines. It must follow the five steps:

1. adequate information to allow for an informed decision about participation in the clinical investigation.
2. facilitating the potential participant's understanding of the information.
3. an appropriate amount of time to ask questions and to discuss with family and friends the research protocol and whether you should participate.
4. obtaining the potential participant's voluntary agreement to participate.
5. continuing to provide information as the clinical investigation progresses or as the subject or situation requires.

**Conclusions/action items:**

This would be a relevant addition to the PDS, as it defines where our data comes from and the limitations of our data.

21 CFR 56.102 up to date as of 9/19/2023	21 CFR 56.102 (Sep. 19, 2023)
<b>Definitions.</b>	
This content from the eCFR is not authoritative but unofficial.	
<b>Title 21 -- Food and Drugs</b>	
<b>Chapter I -- Food and Drug Administration, Department of Health and Human Services</b>	
<b>Subchapter A -- General</b>	
<b>Part 56 -- Institutional Review Boards</b>	
<b>Subpart A -- General Provisions</b>	
Authority: 21 U.S.C. 321, 343, 346, 349, 348, 350a, 350b, 351, 352, 353, 355, 360, 360b, 360c, 360d, 360e, 360h, 360i, 360j, 360k, 360l, 360m, 360n, 360o, 360p, 360q, 360r, 360s, 360t, 360u, 360v, 360w, 360x, 360y, 360z, 360aa, 360ab, 360ac, 360ad, 360ae, 360af, 360ag, 360ah, 360ai, 360aj, 360ak, 360al, 360am, 360an, 360ao, 360ap, 360aq, 360ar, 360as, 360at, 360au, 360av, 360aw, 360ax, 360ay, 360az, 360ba, 360bb, 360bc, 360bd, 360be, 360bf, 360bg, 360bh, 360bi, 360bj, 360bk, 360bl, 360bm, 360bn, 360bo, 360bp, 360bq, 360br, 360bs, 360bt, 360bu, 360bv, 360bw, 360bx, 360by, 360bz, 360ca, 360cb, 360cc, 360cd, 360ce, 360cf, 360cg, 360ch, 360ci, 360cj, 360ck, 360cl, 360cm, 360cn, 360co, 360cp, 360cq, 360cr, 360cs, 360ct, 360cu, 360cv, 360cw, 360cx, 360cy, 360cz, 360da, 360db, 360dc, 360dd, 360de, 360df, 360dg, 360dh, 360di, 360dj, 360dk, 360dl, 360dm, 360dn, 360do, 360dp, 360dq, 360dr, 360ds, 360dt, 360du, 360dv, 360dw, 360dx, 360dy, 360dz, 360ea, 360eb, 360ec, 360ed, 360ee, 360ef, 360eg, 360eh, 360ei, 360ej, 360ek, 360el, 360em, 360en, 360eo, 360ep, 360eq, 360er, 360es, 360et, 360eu, 360ev, 360ew, 360ex, 360ey, 360ez, 360fa, 360fb, 360fc, 360fd, 360fe, 360ff, 360fg, 360fh, 360fi, 360fj, 360fk, 360fl, 360fm, 360fn, 360fo, 360fp, 360fq, 360fr, 360fs, 360ft, 360fu, 360fv, 360fw, 360fx, 360fy, 360fz, 360ga, 360gb, 360gc, 360gd, 360ge, 360gf, 360gg, 360gh, 360gi, 360gj, 360gk, 360gl, 360gm, 360gn, 360go, 360gp, 360gq, 360gr, 360gs, 360gt, 360gu, 360gv, 360gw, 360gx, 360gy, 360gz, 360ha, 360hb, 360hc, 360hd, 360he, 360hf, 360hg, 360hh, 360hi, 360hj, 360hk, 360hl, 360hm, 360hn, 360ho, 360hp, 360hq, 360hr, 360hs, 360ht, 360hu, 360hv, 360hw, 360hx, 360hy, 360hz, 360ia, 360ib, 360ic, 360id, 360ie, 360if, 360ig, 360ih, 360ii, 360ij, 360ik, 360il, 360im, 360in, 360io, 360ip, 360iq, 360ir, 360is, 360it, 360iu, 360iv, 360iw, 360ix, 360iy, 360iz, 360ja, 360jb, 360jc, 360jd, 360je, 360jf, 360jg, 360jh, 360ji, 360jj, 360jk, 360jl, 360jm, 360jn, 360jo, 360jp, 360jq, 360jr, 360js, 360jt, 360ju, 360jv, 360jw, 360jx, 360jy, 360jz, 360ka, 360kb, 360kc, 360kd, 360ke, 360kf, 360kg, 360kh, 360ki, 360kj, 360kk, 360kl, 360km, 360kn, 360ko, 360kp, 360kq, 360kr, 360ks, 360kt, 360ku, 360kv, 360kw, 360kx, 360ky, 360kz, 360la, 360lb, 360lc, 360ld, 360le, 360lf, 360lg, 360lh, 360li, 360lj, 360lk, 360ll, 360lm, 360ln, 360lo, 360lp, 360lq, 360lr, 360ls, 360lt, 360lu, 360lv, 360lw, 360lx, 360ly, 360lz, 360ma, 360mb, 360mc, 360md, 360me, 360mf, 360mg, 360mh, 360mi, 360mj, 360mk, 360ml, 360mm, 360mn, 360mo, 360mp, 360mq, 360mr, 360ms, 360mt, 360mu, 360mv, 360mw, 360mx, 360my, 360mz, 360na, 360nb, 360nc, 360nd, 360ne, 360nf, 360ng, 360nh, 360ni, 360nj, 360nk, 360nl, 360nm, 360nn, 360no, 360np, 360nq, 360nr, 360ns, 360nt, 360nu, 360nv, 360nw, 360nx, 360ny, 360nz, 360oa, 360ob, 360oc, 360od, 360oe, 360of, 360og, 360oh, 360oi, 360oj, 360ok, 360ol, 360om, 360on, 360oo, 360op, 360oq, 360or, 360os, 360ot, 360ou, 360ov, 360ow, 360ox, 360oy, 360oz, 360pa, 360pb, 360pc, 360pd, 360pe, 360pf, 360pg, 360ph, 360pi, 360pj, 360pk, 360pl, 360pm, 360pn, 360po, 360pp, 360pq, 360pr, 360ps, 360pt, 360pu, 360pv, 360pw, 360px, 360py, 360pz, 360qa, 360qb, 360qc, 360qd, 360qe, 360qf, 360qg, 360qh, 360qi, 360qj, 360qk, 360ql, 360qm, 360qn, 360qo, 360qp, 360qq, 360qr, 360qs, 360qt, 360qu, 360qv, 360qw, 360qx, 360qy, 360qz, 360ra, 360rb, 360rc, 360rd, 360re, 360rf, 360rg, 360rh, 360ri, 360rj, 360rk, 360rl, 360rm, 360rn, 360ro, 360rp, 360rq, 360rr, 360rs, 360rt, 360ru, 360rv, 360rw, 360rx, 360ry, 360rz, 360sa, 360sb, 360sc, 360sd, 360se, 360sf, 360sg, 360sh, 360si, 360sj, 360sk, 360sl, 360sm, 360sn, 360so, 360sp, 360sq, 360sr, 360ss, 360st, 360su, 360sv, 360sw, 360sx, 360sy, 360sz, 360ta, 360tb, 360tc, 360td, 360te, 360tf, 360tg, 360th, 360ti, 360tj, 360tk, 360tl, 360tm, 360tn, 360to, 360tp, 360tq, 360tr, 360ts, 360tt, 360tu, 360tv, 360tw, 360tx, 360ty, 360tz, 360ua, 360ub, 360uc, 360ud, 360ue, 360uf, 360ug, 360uh, 360ui, 360uj, 360uk, 360ul, 360um, 360un, 360uo, 360up, 360uq, 360ur, 360us, 360ut, 360uu, 360uv, 360uw, 360ux, 360uy, 360uz, 360va, 360vb, 360vc, 360vd, 360ve, 360vf, 360vg, 360vh, 360vi, 360vj, 360vk, 360vl, 360vm, 360vn, 360vo, 360vp, 360vq, 360vr, 360vs, 360vt, 360vu, 360vv, 360vw, 360vx, 360vy, 360vz, 360wa, 360wb, 360wc, 360wd, 360we, 360wf, 360wg, 360wh, 360wi, 360wj, 360wk, 360wl, 360wm, 360wn, 360wo, 360wp, 360wq, 360wr, 360ws, 360wt, 360wu, 360wv, 360ww, 360wx, 360wy, 360wz, 360xa, 360xb, 360xc, 360xd, 360xe, 360xf, 360fg, 360fh, 360fi, 360fj, 360fk, 360fl, 360fm, 360fn, 360fo, 360fp, 360fq, 360fr, 360fs, 360ft, 360fu, 360fv, 360fw, 360fx, 360fy, 360fz, 360ga, 360gb, 360gc, 360gd, 360ge, 360gf, 360gg, 360gh, 360gi, 360gj, 360gk, 360gl, 360gm, 360gn, 360go, 360gp, 360gq, 360gr, 360gs, 360gt, 360gu, 360gv, 360gw, 360gx, 360gy, 360gz, 360ha, 360hb, 360hc, 360hd, 360he, 360hf, 360hg, 360hh, 360hi, 360hj, 360hk, 360hl, 360hm, 360hn, 360ho, 360hp, 360hq, 360hr, 360hs, 360ht, 360hu, 360hv, 360hw, 360hx, 360hy, 360hz, 360ia, 360ib, 360ic, 360id, 360ie, 360if, 360ig, 360ih, 360ii, 360ij, 360ik, 360il, 360im, 360in, 360io, 360ip, 360iq, 360ir, 360is, 360it, 360iu, 360iv, 360iw, 360ix, 360iy, 360iz, 360ja, 360jb, 360jc, 360jd, 360je, 360jf, 360jg, 360jh, 360ji, 360jj, 360jk, 360jl, 360jm, 360jn, 360jo, 360jp, 360jq, 360jr, 360js, 360jt, 360ju, 360jv, 360jw, 360jx, 360jy, 360jz, 360ka, 360kb, 360kc, 360kd, 360ke, 360kf, 360kg, 360kh, 360ki, 360kj, 360kk, 360kl, 360km, 360kn, 360ko, 360kp, 360kq, 360kr, 360ks, 360kt, 360ku, 360kv, 360kw, 360kx, 360ky, 360kz, 360la, 360lb, 360lc, 360ld, 360le, 360lf, 360lg, 360lh, 360li, 360lj, 360lk, 360ll, 360lm, 360ln, 360lo, 360lp, 360lq, 360lr, 360ls, 360lt, 360lu, 360lv, 360lw, 360lx, 360ly, 360lz, 360ma, 360mb, 360mc, 360md, 360me, 360mf, 360mg, 360mh, 360mi, 360mj, 360mk, 360ml, 360mm, 360mn, 360mo, 360mp, 360mq, 360mr, 360ms, 360mt, 360mu, 360mv, 360mw, 360mx, 360my, 360mz, 360na, 360nb, 360nc, 360nd, 360ne, 360nf, 360ng, 360nh, 360ni, 360nj, 360nk, 360nl, 360nm, 360nn, 360no, 360np, 360nq, 360nr, 360ns, 360nt, 360nu, 360nv, 360nw, 360nx, 360ny, 360nz, 360oa, 360ob, 360oc, 360od, 360oe, 360of, 360og, 360oh, 360oi, 360oj, 360ok, 360ol, 360om, 360on, 360oo, 360op, 360oq, 360or, 360os, 360ot, 360ou, 360ov, 360ow, 360ox, 360oy, 360oz, 360pa, 360pb, 360pc, 360pd, 360pe, 360pf, 360pg, 360ph, 360pi, 360pj, 360pk, 360pl, 360pm, 360pn, 360po, 360pp, 360pq, 360pr, 360ps, 360pt, 360pu, 360pv, 360pw, 360px, 360py, 360pz, 360qa, 360qb, 360qc, 360qd, 360qe, 360qf, 360qg, 360qh, 360qi, 360qj, 360qk, 360ql, 360qm, 360qn, 360qo, 360qp, 360qq, 360qr, 360qs, 360qt, 360qu, 360qv, 360qw, 360qx, 360qy, 360qz, 360ra, 360rb, 360rc, 360rd, 360re, 360rf, 360rg, 360rh, 360ri, 360rj, 360rk, 360rl, 360rm, 360rn, 360ro, 360rp, 360rq, 360rr, 360rs, 360rt, 360ru, 360rv, 360rw, 360rx, 360ry, 360rz, 360sa, 360sb, 360sc, 360sd, 360se, 360sf, 360sg, 360sh, 360si, 360sj, 360sk, 360sl, 360sm, 360sn, 360so, 360sp, 360sq, 360sr, 360ss, 360st, 360su, 360sv, 360sw, 360sx, 360sy, 360sz, 360ta, 360tb, 360tc, 360td, 360te, 360tf, 360tg, 360th, 360ti, 360tj, 360tk, 360tl, 360tm, 360tn, 360to, 360tp, 360tq, 360tr, 360ts, 360tt, 360tu, 360tv, 360tw, 360tx, 360ty, 360tz, 360ua, 360ub, 360uc, 360ud, 360ue, 360uf, 360ug, 360uh, 360ui, 360uj, 360uk, 360ul, 360um, 360un, 360uo, 360up, 360uq, 360ur, 360us, 360ut, 360uu, 360uv, 360uw, 360ux, 360uy, 360uz, 360va, 360vb, 360vc, 360vd, 360ve, 360vf, 360vg, 360vh, 360vi, 360vj, 360vk, 360vl, 360vm, 360vn, 360vo, 360vp, 360vq, 360vr, 360vs, 360vt, 360vu, 360vv, 360vw, 360vx, 360vy, 360vz, 360wa, 360wb, 360wc, 360wd, 360we, 360wf, 360wg, 360wh, 360wi, 360wj, 360wk, 360wl, 360wm, 360wn, 360wo, 360wp, 360wq, 360wr, 360ws, 360wt, 360wu, 360wv, 360ww, 360wx, 360wy, 360wz, 360xa, 360xb, 360xc, 360xd, 360xe, 360xf, 360xg, 360xh, 360xi, 360xj, 360xk, 360xl, 360xm, 360xn, 360xo, 360xp, 360xq, 360xr, 360xs, 360xt, 360xu, 360xv, 360xw, 360xx, 360xy, 360xz, 360ya, 360yb, 360yc, 360yd, 360ye, 360yf, 360yg, 360yh, 360yi, 360yj, 360yk, 360yl, 360ym, 360yn, 360yo, 360yp, 360yq, 360yr, 360ys, 360yt, 360yu, 360yv, 360yw, 360yx, 360yy, 360yz, 360za, 360zb, 360zc, 360zd, 360ze, 360zf, 360zg, 360zh, 360zi, 360zj, 360zk, 360zl, 360zm, 360zn, 360zo, 360zp, 360zq, 360zr, 360zs, 360zt, 360zu, 360zv, 360zw, 360zx, 360zy, 360zz	
21 CFR 56.102 up to date as of 9/19/2023	page 1 of 8

[Download](#)

21\_CFR\_56.102\_up\_to\_date\_as\_of\_9-19-2023\_.pdf (49.9 kB)



## 2023/09/21\_de-identification

---

Richard YANG - Sep 21, 2023, 4:41 PM CDT

**Title:** de-identification of the data

**Date:** 09/21

**Content by:** Richard

**Present:** NA

**Goals:** determine the step to de-identify data and how to work with them

**Content:**

since the data as-is is identifiable, it must be processed before it can be used for this project. HIPAA privacy rule defines de-identified data as no reasonable basis to believe that the information can be used to identify an individual under 45 CFR 164.514.

After the data is de-identified, HIPAA does not restrict the use or disclosure of de-identified health information, as it is no longer considered protected health information.

HIPAA provides two de-identification methods per 45 CFR 164.514(b): 1) Expert determination 2) Safe harbor. The former requires “a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable” while the latter requires the removal of 18 types of identifiers, including but not limited to name, address, and phone number

**Conclusions/action items:**

these steps should be taken either by the client or us before the data can be used. This is also helpful information for the PDS.

45 CFR 164.514 (up to date as of 9/19/2023)  
Other requirements relating to uses and disclosures of protected health information.

The content in this CFR is a summary but not official.

Title 45 — Public Welfare  
Subtitle A — Department of Health and Human Services  
Subchapter C — Administrative Data Standards and Related Requirements  
Part 164 — Security and Privacy

Subpart E — Privacy of Individually Identifiable Health Information

**Authority:** 42 U.S.C. 1320a-2, 1320a-3, and 1320a-5; sec. 264 of Pub. L. 94-142, 116 Stat. 2003-2004 (42 U.S.C. 1320a-2) (1998); (codified at 45 CFR 164.514); Pub. L. 115-3, 323 Stat. 228-279.  
**Authority:** 42 U.S.C. 1320a-2; 42 U.S.C. 1320a-3; sec. 264, Pub. L. 94-142, 116 Stat. 2003-2004 (42 U.S.C. 1320a-2) (1998); (codified at 45 CFR 164.514); Pub. L. 115-3, 323 Stat. 228-279.  
**Authority:** 42 U.S.C. 1320a-2; 42 U.S.C. 1320a-3; sec. 264, Pub. L. 94-142, 116 Stat. 2003-2004 (42 U.S.C. 1320a-2) (1998); (codified at 45 CFR 164.514); Pub. L. 115-3, 323 Stat. 228-279.

§ 164.514 Other requirements relating to uses and disclosures of protected health information.

- (a) **Standard: De-identification of protected health information.** Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information.
- (b) **Implementation specifications: Requirements for de-identification of protected health information.** A covered entity may determine that health information is not individually identifiable health information only if:
  - (1) A person with appropriate knowledge and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:
    - (i) Applying each principle and method, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
    - (ii) Documents the methods and results of the analysis that justify such determinations; and
  - (2) The following definitions of the individual or relatives, employees, or household members of the individual, are met:
    - (A) Names;
    - (B) All geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly available data from the Bureau of the Census:
      - (1) The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and
      - (2) The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000.
    - (C) All elements of dates (except year) for dates directly related to an individual (including birth date, admission date, discharge date, date of death) and all ages over 89 and all elements of dates (including year) indicating such ages except that such ages and elements may be aggregated into a single category of age 90 or older.

45 CFR 164.514 (up to date as of 9/19/2023) (enhanced display) page 1 of 8

[Download](#)

45\_CFR\_164.514\_up\_to\_date\_as\_of\_9-19-2023\_.pdf (84.9 kB)



## 2023/09/22\_data structure

---

Richard YANG - Sep 22, 2023, 11:38 AM CDT

**Title:** data structure idea

**Date:** 09/22

**Content by:** Richard

**Present:** NA

**Goals:** determine feasible data structures to implement

**Content:**

there are two main scenarios where data structure for the input data is necessary: 1. slow processing and fast input rate, 2. processing from a list/database

in any case, a FIFO should be the optimal solution as it ensures that no task would be skipped due to slow processing rate, which is also used in data transmission

Alternatively, a priority cue can also be used if there are any reasons some patients might be prioritize.

Most likely, a data structure would not be needed, since the processing time is likely to be in the frames of a few seconds.

**Conclusions/action items:**

determine baseline algorithm





**Title:** IVM

**Date:** 09 26

**Content by:** Richard

**Present:** NA

**Goals:** assess IVM for baseline model

**Content:**

what is a kernel:

a kernel trick refers to instead of calculating the corresponding coordinate of the data after non-linear (or linear) transformation, one can simply calculate the difference of the transformation said data point and other transformed data points by taking the inner product. This becomes much more helpful with used with transformation functions that have an infinite feature dimension.

IVM is built to be a better version of SVM specifically for multi-class classification

it is computationally less expensive than SVM

it uses a logistic regression kernel

**Conclusions/action items:**

literature shows that IVM can be just as good as SVM at two-class classification and less computationally expensive at the same time. Although currently multi-class capability is not needed, it is worth considering as a candidate for the baseline model



### Kernel Logistic Regression and the Import Vector Machine

Ji Zhu & Trevor Hastie

To cite this article: Ji Zhu & Trevor Hastie (2005) Kernel Logistic Regression and the Import Vector Machine, Journal of Computational and Graphical Statistics, 14(1), 185-205, DOI: 10.1198/153860305000005619

To link to this article: <https://doi.org/10.1198/153860305000005619>

Published online: 23 Jun 2012

Submit your article to this journal

Article views: 892

View related articles

Cite articles: 30 View citing articles

Full terms & conditions of access and use can be found at <https://www.int.wiley.com/doc/etop/journals/information/jocgs/jocgs.html>

[Download](#)

**Kernel\_Logistic\_Regression\_and\_the\_Import\_Vector\_Machine.pdf (745 kB)**



**Title:** KNN for image classification

**Date:** 09 27

**Content by:** Richard

**Present:** NA

**Goals:** the teams chose KNN for our baseline model, the goal of this entry is to find out information and areas of KNN that we can exploit specifically for image classification

**Content:**

[post](#)

with KNN (as with image classification generally) light intensity variations, rotations, etc. are important to consider, and slight variations can change the result dramatically. This post converts an image to a vector and then generates a color histogram, but data augmentation should also be considered.

Also, a few other important aspects to consider:

distance metrics/similarity functions: simple Euclidean, city block

number of neighbors: this parameter is very important for KNN, in the past, I have exhaustively searched through a reasonable set of  $k$ , but there might exist a more heuristic approach

**Conclusions/action items:**

Find a heuristic approach to determine the best number of neighbors for KNN. Since the dataset might be slow to process, a heuristic approach might be necessary if the brute force approach takes unreasonably long to execute.



**Title:** plans on working on the algorithm

**Date:** 09 29

**Content by:** Richard

**Present:** Team

**Goals:** determine how the team will be working on the algorithm

**Content:**

The team has decided that GUI is less important than the algorithm and should be worked on last

The team will split up into three teams, KNN, VGG-19, and image to tensor conversion

Basic algorithm to process numerical data set:

KNN: Dhruv, Brandon

VGG-19 : Yousef

Algorithm to import image data : Aran

**Conclusions/action items:**

This plan will start on oct 2nd Monday, in the mean time the team will prepare for the preliminary presentation



**Title:** VGG 19

**Date:** 10/02

**Content by:** Richard

**Present:** NA

**Goals:** find out how to implement VGG19 with PyTorch

**Content:**

[link](#)

model is created with "torchvision.models.vgg19"

there are two optional parameters:

1. weights, can be set to any pre-trained weights, "VGG19\_weights" can also be used, default is NULL

weights='DEFAULT' is equivalent to weights='IMAGENET1K\_V1'

2. progress, a bool, when set to TRUE a progress bar is displayed, default is true

**Conclusions/action items:**

document any other important methods or classes associated with the model

import the model in VScode and make sure that an arbitrary dataset can be processed using the model



**2023/10/06VGG19Continued**

---

**Title:** VGG19

**Date:** 10 06

**Content by:** Richard

**Present:** NA

**Goals:** document further important methods etc

**Content:**

[vgg19](#)

The following is the import procedure, this is using .10pytorch, though .12or .13 can be used

```
import torch
```

```
model = torch.hub.load('pytorch/vision:v0.10.0', 'vgg11', pretrained=True)
```

```
# or any of these variants
```

```
# model = torch.hub.load('pytorch/vision:v0.10.0', 'vgg11_bn', pretrained=True)
```

```
# model = torch.hub.load('pytorch/vision:v0.10.0', 'vgg13', pretrained=True)
```

```
# model = torch.hub.load('pytorch/vision:v0.10.0', 'vgg13_bn', pretrained=True)
```

```
# model = torch.hub.load('pytorch/vision:v0.10.0', 'vgg16', pretrained=True)
```

```
# model = torch.hub.load('pytorch/vision:v0.10.0', 'vgg16_bn', pretrained=True)
```

```
# model = torch.hub.load('pytorch/vision:v0.10.0', 'vgg19', pretrained=True)
```

```
# model = torch.hub.load('pytorch/vision:v0.10.0', 'vgg19_bn', pretrained=True)
```

```
model.eval()
```

it also provided methods for execution:

```
from PIL import Image
```

```
from torchvision import transforms
```

```
input_image = Image.open(filename)
```

```
preprocess = transforms.Compose([
```

```
    transforms.Resize(256),
```

```
    transforms.CenterCrop(224),
```

```
transforms.ToTensor(),
transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]),
])
input_tensor = preprocess(input_image)
input_batch = input_tensor.unsqueeze(0) # create a mini-batch as expected by the model

# move the input and model to GPU for speed if available
if torch.cuda.is_available():
    input_batch = input_batch.to('cuda')
    model.to('cuda')

with torch.no_grad():
    output = model(input_batch)

# Tensor of shape 1000, with confidence scores over ImageNet's 1000 classes
print(output[0])

# The output has unnormalized scores. To get probabilities, you can run a softmax on it.
probabilities = torch.nn.functional.softmax(output[0], dim=0)
print(probabilities)
```

torvision is required for this execution, which I have not yet worked with.

### **Conclusions/action items:**

Find out what torvision is, how it can be downloaded and relevant documentation.





**Title:** ML for US

**Date:** 10 07

**Content by:** Richard

**Present:** NA

**Goals:** find out the current standard and state-of-art for machine learning specifically in applications of Ultrasound related tasks.

**Content:**

Table 1 of the paper provides a collection of papers on US-ML applications on different organs.

US presents numerous challenges: operator dependences, noise, artifacts, limited field of view, difficulty in imaging structures behind bone and air, and variability across different manufacturers' systems

These are important things to consider when interpreting our result, since most of the images are taken from the same device

a common problem among classification studies is their reliance on manual region-of-interest selection, the image set is also usually small

video clips might be more useful than single frames since they provided spatial and temporal information.

It might also be important in the future to provide real-time feedback to the sonographer during image acquisition and not only to interpret US post hoc.

**Conclusions/action items:**

put the above information in the discussion section of the report.





# 2023/10/11\_loadingLocalImages

---

**Title: load local images with pytorch****Date:** 11 11**content by:** Richard**Present:** NA**Goals:** find out how to load images from local storage to script**Content:**<https://towardsdatascience.com/beginners-guide-to-loading-image-data-with-pytorch-289c60b7afec>

numpy is required, specifically the np.load function

here is a sample code:

```
image_size = 64
```

```
DATA_DIR = './input/vaporarray/test.out.npy'
```

```
X_train = np.load(DATA_DIR)
```

```
print(f"Shape of training data: {X_train.shape}")
```

```
print(f"Data type: {type(X_train)}")
```

to find out array composition run this following line:

```
print(type(X_train[0][0][0][0]))
```

numpy.uint8 must be used for pytorch, if the data isn't already in the correct data type, use the following for conversion:

```
data = X_train.astype(np.float64)
```

```
data = 255 * data
```

```
X_train = data.astype(np.uint8)
```

use the following to visualize a random image in order to verify that the images are loaded correctly:

```
random_image = random.randint(0, len(X_train))
```

```
plt.imshow(X_train[random_image])
```

```
plt.title(f"Training example #{random_image}")
```

```
plt.axis('off')
```

plt.show()

**Conclusions/action items:**

the same article also documents data augmentation techniques and how to create a class from the above methods to increase abstraction. A data loader as a class would be beneficial, though not entirely important. The data augmentation techniques are, however, extremely useful to our application and should be explored.



## 2023/10/13\_loadingImages

---

Richard YANG - Oct 13, 2023, 12:50 PM CDT

**Title:** loading images

**Date:** 10 13

**Content by:** Richard

**Present:** NA

**Goals:** find out methods to convert image into tensors, since np.load did not work as intended

**Content:**

<https://www.pluralsight.com/guides/importing-image-data-into-numpy-arrays>

the above article illustrated several ways of loading image files, with matplotlib and pillow. The article only showed single image conversion and the project requires bulk conversion. furthermore, the compatibility with pytorch and torchvision has to be tested.

<https://koushik1102.medium.com/transfer-learning-with-vgg16-and-vgg19-the-simpler-way-ad4eec1e2997>

this above article uses glob for batch conversion

**Conclusions/action items:**

test the compatibility of the image conversion algorithm



**Title:** plans forward

**Date:** 10 15

**Content by:** Richard

**Present:** NA

**Goals:** device the plan forward after the second client meeting

**Content:**

1. two algorithms will be developed separately for the parotid and submandibular. When the patient's US is taken, 2(one of each gland) or 4 (one for each side of each gland) images will be input into the algorithm.
  1. this means that the dataset must be separated
  2. can train each of the two algorithms to output on the OMERACT scale and train a third algorithm to take two OMERACT values to produce 1/0 result.
2. an additional database has been provided
  1. <https://github.com/ArsoVukicevic/Assessment-of-pSS-from-SGUS-images/tree/master/3%20HarmonicSS%20benchmark%20dataset>
  2. the number of negative patients must be assessed, SMOTE might be necessary if there is an imbalance
3. highlighted box images should be ignored
4. images should be cropped to the appropriate window

**Conclusions/action items:**

discuss plan with team



## 2023/10/16\_thirdClassifier

---

Richard YANG - Oct 16, 2023, 3:26 PM CDT

**Title:** third classifier

**Date:** 10 16

**Content by:** Richard

**Present:** NA

**Goals:** brainstorm the structure of the third classifier

**Content:**

the structure of the data should be visualized first, i.e. the relationship between the two OMERACT scores and the Sjs status.

If the relationship is approximately linear or not too high-dimension, an SVM or a simple MLP should be enough.

**Conclusions/action items:**

visualize the data with scatter plot









## 2023/10/20\_OMERACTRegressor

---

Richard YANG - Oct 20, 2023, 12:12 PM CDT

**Title:** regressor for OMERACT

**Date:** 10 20

**Content by:** Richard

**Present:** NA

**Goals:** build a preliminary regressor to take the OMERACT score from one gland and output the score of the other

**Content:**

after the data is scatter-plotted, the relationship is approximately linear, with some randomness. Both XGBoost and SVM performed with a coefficient of determination of 0.4.

after visualizing the data, it would be difficult to improve the model any further with the same input parameter since there is a large amount of randomness that cannot be explained by the current parameter.

Maybe it is possible to add age etc as additional parameters.

**Conclusions/action items:**

experiment with additional parameters



## 2023/10/21\_goalsForShowNTell

---

Richard YANG - Oct 21, 2023, 1:35 PM CDT

**Title:** goals for show and tell

**Date:** 10 21

**Content by:** Richard

**Present:** NA

**Goals:** determine the plan for progression for the next two weeks

**Content:**

the goal is to have a complete system working from raw images to training

this entails the following:

1. image preprocessing
  1. cropping
  2. filtering
  3. sorting
2. training VGG19 and KNN
  1. importing images
  2. importing class labels
  3. appropriate training functions
  4. confusion matrix and accuracy
3. training secondary classifier
  1. planA: use OMERACT
  2. planB: use probability output from VGG19
4. GUI
  1. this is a secondary objective

**Conclusions/action items:**

discuss with team



**Title:** difficulties of the final classifier

**Date:** 10 23

**Content by:** Richard

**Present:** NA

**Goals:** summarize the difficulties with the final classifier

**Content:**

the final classifier based on OMERACT numerical input is complete. the accuracy hangs around 70%. The difficulty seems to be differentiating patients' disease status with lower OMERACT scores of 0-1.5.

in order to improve the accuracy, there needs to be additional information other than the OMERACT score provided as input. As discussed before, the probability of classification is a potential option from the previous classifiers. Another regressed can also be trained to specifically output probability.

**Conclusions/action items:**

upload script, look into outputting probability in KNN and VGG



**Title:** ideas

**Date:** 2023

**Content by:** Richard

**Present:** NA

**Goals:** document ideas on the secondary classifier

**Content:**

since the classifier seems to only struggle with lower OMERACT scores, maybe another classifier can be trained to only look at lower OMERACT scores. The flow of which would be similar to the previous idea of getting additional output from the OMERACT grader; however, in this case only the OMERACT score  $\leq 2$  will be passed through this additional step to reduce processing time. Though the classifiers built so far seem to be really efficient at executing commands, terminating within seconds.

**Conclusions/action items:**

as before, look into additional output parameters of KNN and VGG



**Title:** show and tell

**Date:** 11 01

**Content by:** Richard

**Present:** NA

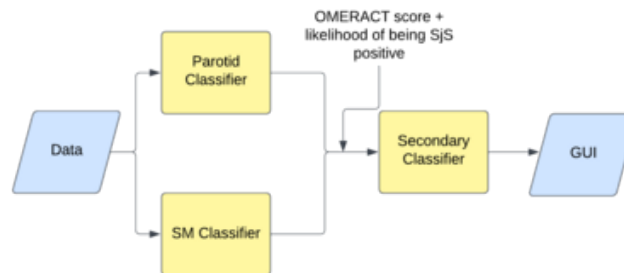
**Goals:** create graphic content for show and tell

**Content:**

Created a couple flowchart

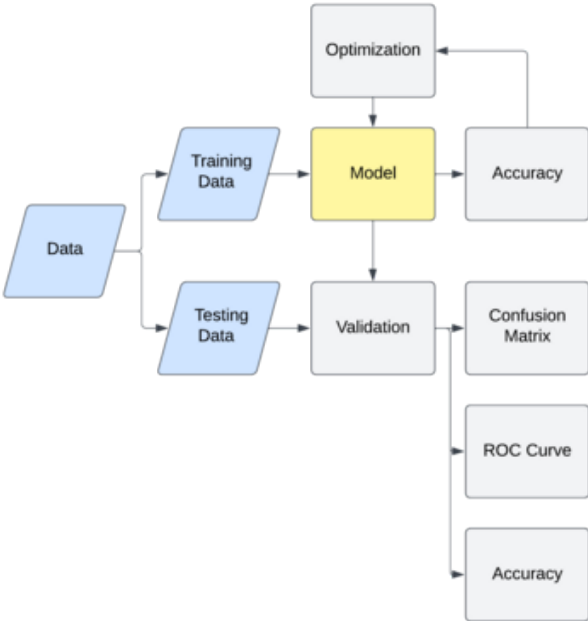
**Conclusions/action items:**

include some ultrasound samples and make code more readable, continue work on GUI



[Download](#)

clientReq\_-\_training\_1\_.png (51 kB)



[Download](#)

clientReq\_-\_Page\_4.png (72.4 kB)





**Title:** GUI progress

**Date:** 11 01

**Content by:** Richard

**Present:** NA

**Goals:** document current status on the GUI

**Content:**

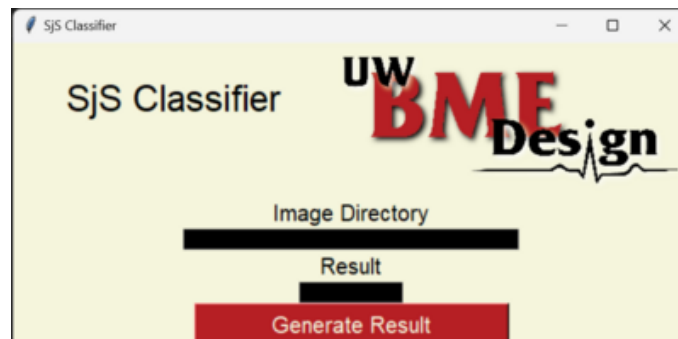
a basic interface has been built

I am moderately comfortable with tkinter at this point and have a good idea of how to implement the rest of the GUI

a screenshot of the GUI is included

**Conclusions/action items:**

next step would be integrating the classifier with the GUI. Since I have already created a GUIlibrary.py file, we will likely keep the model in there, and save the weights in a separate file. I think the best course of action would be to instantiate the model as an object and create relevant methods, so that the weights don't have to be reloaded every time



[Download](#)

Screenshot\_2023-11-01\_120642.png (91.2 kB)



## 2023/11/08\_plansForward

---

Richard YANG - Nov 08, 2023, 6:43 PM CST

**Title:** plans forward

**Date:** 11 08

**Content by:** Richard

**Present:** NA

**Goals:** devise the goals for the next month

**Content:**

As we have finished the preprocessing stage of the project, the next immediate objective should be finish the first two classifiers, which three team members are working on. In the meanwhile, we should explore methods of data augmentation.

As for the secondary classifier, we should add another set of input of probability.

As for the GUI, initial steps should be taken to integrate the ml algorithm.

**Conclusions/action items:**

discusses with team the update the progress



## 2023/11/10\_savingWeights

---

Richard YANG - Nov 10, 2023, 12:03 PM CST

**Title:** saving and loading model weights

**Date:** 11 10

**Content by:** Richard

**Present:** NA

**Goals:** discover codes that can be used for saving model weights and biases so that it does not have to be retrained.

**Content:**

[https://scikit-learn.org/stable/model\\_persistence.html](https://scikit-learn.org/stable/model_persistence.html)

pickle and joblib can be used to dump and load weights, as described above by sklearn

**Conclusions/action items:**

try with model. See if there is any compatibility issues.



## 2023/11/13\_modelResults

---

Richard YANG - Nov 13, 2023, 6:42 PM CST

**Title:** model results

**Date:** 11 13

**Content by:** Richard

**Present:** NA

**Goals:** decide what results need to be produced

**Content:**

1. OMERACT accuracy for each gland
2. Accuracy of disease status based on two glands
  1. classifier needs to output OMERACT and the probability of positive
  2. the secondary classifier then takes this information to produce result
3. Accuracy of disease status based on one gland

**Conclusions/action items:**

make sure the team members know what the expectations are



**Title:** data imbalance

**Date:** 11 15

**Content by:** Richard

**Present:** NA

**Goals:** address the problem of data imbalance

**Content:**

the class distributions of the 4 OMERACT scores are highly imbalanced, as a result, the neural network keeps predicting one of the four labels in order to minimize loss instead of learning the difference between the four classes. There are different techniques, but here I will implement class weight first.

my code takes inspiration from this [article](#)

```
from torch import optim from sklearn.utils import class_weight import torch.nn as nn
tmp = [y.detach().numpy() for X, y in train_dataloader]
train_labels = np.hstack(tmp)
class_weights = torch.from_numpy(class_weight.compute_class_weight(class_weight='balanced', classes=np.unique(train_labels), y=train_labels))
optimizer = optim.Adam(model.parameters(), lr = 0.00001)
criterion = nn.CrossEntropyLoss(weight=class_weights)
```

this essentially tell the classifier how much to penalize, or change how much of the weights dependent on the target class label.

this worked well on a simple CNN.

**Conclusions/action items:**

try this on VGG and implement of the disease status set.



## 2023/11/27\_batchSizeAndLearningRate

---

Richard YANG - Nov 27, 2023, 5:56 PM CST

**Title:** batch size and learning rate

**Date:** 11 27

**Content by:** Richard

**Present:** NA

**Goals:** guide to determine the best combination of batchsize and learning rate

**Content:**

[link](#)

to have the same training performance, one can double the learning rate and half the batch size

large batch sizes may lead to a generalization gap, though this is most likely caused by the low number of steps

**Conclusions/action items:**

training the models with batch size 256 to 1024



## 2023/11/29\_savingWeights

---

Richard YANG - Nov 29, 2023, 3:48 PM CST

**Title:** saving weights

**Date:** 11 29

**Content by:** Richard

**Present:** NA

**Goals:** official method of saving weights

**Content:**

pickle can be used to save weights, but pytorch has its own methods

to save :

```
torch.save(model, PATH)
```

to load:

```
# Model class must be defined somewhere  
model = torch.load(PATH)  
model.eval()
```

**Conclusions/action items:**

save the weights of the models that has been trained



**Title:** future work

**Date:** 12 12

**Content by:** Richard

**Present:** NA

**Goals:** document future work

**Content:**

<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0276-2>

**This project was similar to ours but with xrays and a much larger and balanced dataset**

**it implemented data augmentation, which we can take inspiration from**

**we can also reference their parameters of input resolution and batch size**

**Conclusions/action items:**

implement this next semester





## 2023/09/21\_citiTraining

---

Richard YANG - Sep 21, 2023, 4:44 PM CDT

**Title:** CITI training

**Date:** 09/21

**Content by:** Richard

**Present:** NA

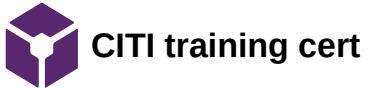
**Goals:** work on the CITI training

**Content:**

the team is undergoing CITI training per the request of the client. as of this entry, 4 of the 10 modules are complete. Some of the information is definitely helpful for the PDS and a great source for standards regarding human subjects as well.

**Conclusions/action items:**

continue CITI training, after which, additional necessary training should be completed.



Richard YANG - Sep 22, 2023, 12:34 PM CDT



[Download](#)

**citiCompletionCertificate\_12660889\_58550507.pdf (77 kB)**



## 2014/11/03-Entry guidelines

---

John Puccinelli - Sep 05, 2016, 1:18 PM CDT

Use this as a guide for every entry

- Every text entry of your notebook should have the **bold titles** below.
- Every page/entry should be **named starting with the date** of the entry's first creation/activity. subsequent material from future dates can be added later.

You can create a copy of the blank template by first opening the desired folder, clicking on "New", selecting "Copy Existing Page...", and then select "2014/11/03-Template")

**Title:** Descriptive title (i.e. Client Meeting)

**Date:** 9/5/2016

**Content by:** The one person who wrote the content

**Present:** Names of those present if more than just you (not necessary for individual work)

**Goals:** Establish clear goals for all text entries (meetings, individual work, etc.).

**Content:**

Contains clear and organized notes (also includes any references used)

**Conclusions/action items:**

Recap only the most significant findings and/or action items resulting from the entry.



**Title:**

**Date:**

**Content by:**

**Present:**

**Goals:**

**Content:**

**Conclusions/action items:**