OWEN KUANG

Applied Science Project

Logbook

2023 – 2024

Schedule:

March

SUN	MON	TUE	WED	THU	FRI	SAT
25	26 Meeting w/ Dr. Pieper / Oral Pres	27	28 Create Graphs / Design Poster	THU 29 ⊙ 1 pending task	Mar 1 Decorate / Create Poster	2
3	4 Meeting w/ Dr. Pieper / Science Fair	5	6	7 Record Video Presentation	8	9
10	11 Meeting w/ Dr. Pieper / Work on results section of the conclusion	12	13 Continue working on the results section	14	15 Complete results and Significance + Future Directions	16
17	18 Meeting w/ Dr. Pieper	19 Send Conclusion to Dr. Garcia for edits	20	21 Complete Edits + Submit	22	23
24	25 Meeting w/ Dr. Pieper / Spring Break	26	27	28	29	30

February

SUN	MON	TUE	WED	THU	FRI	SAT
28	29 Meeting W/ Dr. Pieper	30 Finish CNN Methods	31	Feb 1 Finish Citations / Finish Training and / Validation of Method / ⊘ Submit for edits	2	3
4	5 Finish Intro of Oral Pres / Meeting w/ Dr. Pieper	6	7 Work on methods	8	9 Finish oral pres methods	10
11	12 Meeting w/ Dr. Pieper	13 Work on Results	14	15 Finish Results / Make final procedure Edi / No class	16	17
18	19 Meeting w/ Dr. Pieper	20 Finish Significance Pres	21	22 Oral rehearsal for Dr. Gar	23	24
25	26 Meeting w/ Dr. Pieper / Oral Pres	27	28 Create Graphs / Design Poster	29	Mar 1 Decorate / Create Poster	2

January

SUN	MON	TUE	WED	THU	FRI	SAT
31	Jan 1	2 Meeting W/ Dr. Pieper	3	4	5 Evaluate Model Accuracy	6
7	8	9	10	11	12 Midterms (NO ASP)	13
14	15	16	17	18	19 Midterms (NO ASP)	20
21	22 Midterm (NO ASP)	23	24 Finish Datasets Methods Meeting W/ Dr. Garcia	25	26 Finish Preprocessing Me	27
28	29 Meeting W/ Dr. Pieper	30 Finish CNN Methods	31	Feb 1	2	3

December 2023

Organize your life here! Place all your test dates, assignment due dates, extracurricular events (practices, games, concerts, tournaments), volunteering, and family and social events on this monthly calendar. Colour code it if you desire further organization! And don't forget to schedule in a day off now and then — your brain needs a rest too!

"Alone we can do so little, together we can do so much." —Helen Keller

MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY/SUNDAY
November 27 D3 — Robotics, ASP - Convert spectogram to greyscale	November 28 D4	November 29 D5 — Robotics, ASP - Obtain pixel values for spectogram	November 30 D6 — Meeting w/ Dr. Pieper	1 D1 — ASP - Review literature for RNN input - size down input — TERM 1 REPORT CARDS ISSUED	Robotics tournament 2/3
4 D2	5 D3 — ASP - Review literature for CNN input & convolutional layer	6 D4	7 D5 — ASP - Develop filters for conv. layer	8 D6	Robotics Build Day 9/10
11 D1 — Bonding unit test, Bio unit Quiz, Physics unit Quiz - ASP - continue developing filter, make edits for lit Review	12 D2	13 D3 — ASP - make edits + fix changes suggested by Dr. Garcia after re-review	14 D4 — Unit Quiz	15 D5 — Introduction section Due (ASP), Bio unit test, Physics unit test — ASP - Create max pooling layer + read literature	Robotics Tournament 16/17
18 D6 — ASP - continue max pooling, and reading literature to follow along.	19 D1 — ASP	20 D2 — Math test — LAST DAY OF CLASSES BEFORE WINTER BREAK	21	22	23/24
25	26	27	28 — WINTER BREAK	29 — WINTER BREAK	30/31 — WINTER BREAK
			WINTER BREAK BEGINS	WINTER BREAK	WINTER BREAK

November 2023

Organize your life here! Place all your test dates, assignment due dates, extracurricular events (practices, games, concerts, tournaments), volunteering, and family and social events on this monthly calendar. Colour code it if you desire further organization! And don't forget to schedule in a day off now and then — your brain needs a rest too!

"Take the attitude of a student, never be too big to ask questions, never know too much to learn something new." —Augustine Og Mandino

MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY/SUNDAY
October 30 D2 — Meeting w/ Dr. Pieper, take meeting notes (No meeting)	October 31 D3 — ASP - listn to presentations - Complete all edits on research proposal	1 D4	2 D5 — ASP - work on oral prep - Research Q - Methodology - Objective - Significance	3 D6	4/5 — ASP Practice oral prep
6 D1 — ASP *Oral Presentation due	7 D2	8 D3 — ASP - learn fourier transform implementation for Pytorch	9 D4	10 D5 — ASP - obtain input - transformed time-frequency plot	11/12
13 — SCHOOL CLOSED IN LIEU OF REMEMBRANCE DAY	14 D6	15 D1 — ASP - Develop ML model - find way to convolve the data	16 D2 — ASP - find way to convolve data	17 D3	18/19
20 D4 — chem AP quiz — ASP - find segments + patterns w/ filters	21 D5	22 D6	23 D1 — ASP - find segments + patterns w/ filters	24 D2 — TERM 1 ENDS	25/26
27 D3 — ASP - Attempt to create second layer	28 D4	29 D5 — ASP - Attempt to execute second layer	30 D6 — ASP - Apply neurons for second layer, find how to use auto weights — TERM 1 REPORT CARDS ISSUED	December 1 D1	December 2/3

October 2023

"Education is the most powerful weapon which you can use to change the world."
— Nelson Mandela

MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY/SUNDAY
2 D2 ASP -Meet w/ Dr.Pieper (cancelled b/c no mtg) - Complete chart of different waves & causes research -Ask what dataset.	**3** D3 ASP -come up w/ Research proposal outline - learn IEEE style	**4** ASP	-Programming club **5** D4 - Reading comp.	**6** D5 ASP -Complete working title, half of intro + extra research (if needed)	**7/8** ASP -Send half of intro to Dr. Pieper (deaths caused by AeS, traditional ECG methods, ECG wave information)
9 ASP Meeting w/ Dr. Pieper - Ask dataset - Ask for revisions abt proposal THANKSGIVING SCHOOL IS CLOSED	**10** D6 ASP INDIVIDUAL STUDENT PHOTO RETAKES	**11** D1 ASP -Complete Research Proposal Introduction (Abnormality, research attempts, different ml models) -complete revisions	**12** D2 - ELA Notes due - GH Quiz	**13** D3 ASP -Complete objectives (short, long term)	**14/15** ASP -Send revised proposal + objectives + completed intro.
16 ASP -meet Dr. Pieper - take meeting notes	**17** D4 ASP complete hypothesis+Q	**18** D5 D6 PARENT-TEACHER INTERVIEWS	**19** D1 ASP Complete Methodology +Significance + references -Research Prop. Done PARENT-TEACHER INTERVIEWS	**20** D2 ASP PARENT-TEACHER INTERVIEWS NOON DISMISSAL	**21/22** ASP -Send revised proposal + hypothesis and references
23 D3 ASP -meet Dr. Pieper -take meeting notes	**24** D4 ASP -familiarize w/ implementation of ml algo used for project	**25** D5	**26** D6 ASP	**27** D1 ASP -Familiarize w/ dataset → look for useful values	**28/29**
30 D2 ASP -Meet Dr. Pieper -take meeting notes	**31** D3 ASP -start tutorial on ml algo being used for study	November 1 D4	November 2 D5 ASP	November 3 D6	November 4/5

August/September 2023

"Don't let yesterday take up too much of today."
— Will Rogers

MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY/SUNDAY
August 28 D1 FIRST DAY OF SCHOOL	August 29 D2	August 30 D3	**31** D4	**1** D5	**2/3**
4 D6 LABOUR DAY SCHOOL CLOSED	**5** D6 Legacy meeting Sign up for prog. Club	**6** D1 ASP -Email Dr. Pieper -Download paperpile INDIVIDUAL STUDENT PHOTOS	**7** D2 ASP -Meet w/ Dr. Garcia -Research intro to ml source -Email Dr.Pieper	**8** D3	**9/10**
11 D4 ASP -Meet w/ Dr. Pieper - Hilton - Pytorch	**12** D5 ASP SENIOR SCHOOL CURRICULUM NIGHT	**13** D6	**14** D1 ASP - Research intro to ml - Research how to implement Pytorch w/ jupyter notebook ASP -Hinton's sources	**15** D2 -math unit quiz	**16/17** -Robotics meeting ASP -Pytorch
18 D3 ASP -Meet w/ Dr. Pieper - 1 Page Proposal -ECG -Chem 1-2 Quiz -Bio unit quiz	**19** D4 -GH Quiz -Math Unit test	**20** D5 ASP -Meet w/ Dr. Garcia -Small sources, just summary.	**21** D6 ASP -Finish researching general heart disease/subjects - get an understanding of ml field & where it is at WELCOME BACK BBQ	**22** D1 ASP Bio unit test	**23/24**
25 D2 -Comp arc. quiz -Meet w/ Dr. Pieper -Complete ECG Prop. paper for meet.	**26** D3 ASP -research other methods/ ways of prediction done by other people	**27** D4	**28** D5 ASP -Finish all research - ML models used - complete research on diff. ECG readings/waves -Chem quiz	**29** D6 Programming test. ASP LOGBOOK DUE TERRY FOX RUN	**30/October 1** TERRY FOX FUNDRAISER CONCERT

Checklist:

☑ Research Proposal Introduction (October 11)

☑ Research Proposal Objective (October 13)

☑ Research Proposal Questions (October 17)

☑ Research Proposal Methodology (October 19)

☑ Research Proposal References (October 23)

☑ Research Proposal Significance (October 23)

☑ Completed Research Proposal (October 25) Completed October 31

☑ Introduction to Oral Presentation (October 29)

☑ Research question and goals (October 31)

☑ Methodology to Oral Presentation (October 31)

☑ Significance to Oral Presentation (November 2)

☑ Completed Oral Presentation (November 2)

☑ Oral Presentation (November 6)

☑ Convert ECG Readings into 10 second segments (November 16)

☑ Convert ECG Readings to time frequency plots with STFT (November 16)

☑ Find CNN sources to read on and use as a guide (November 16)

☑ Log into CYSF & complete basic project info (November 27) Completed November 17

☑ Submit CYSF Ethics form (November 27) Completed November 17

☑ Complete the machine learning model (December 29)

    ☑ Create the input layer for the machine learning model (December 13)

- ☑ Obtain the dataset (November 12)
- ☑ Take the MLII portion of the dataset and values and export to csv (November 16)
- ☑ Apply fourier transform to get frequency domain (November 16)
- ☑ Convert spectrogram to grayscale (November 28)
- ☑ Get the pixel value of the greyscale image (December 5)
- ☑ Group the data/lower spatial size and find a suitable resolution (x pixels by x pixels) (December 11)
- ☑ Load pixel values into tensors (December 13)

- ☑ Create the convolutional layer for the machine learning model
  - ☑ Create the filters (December 13)
- ☑ Create the max pooling layer for the machine learning model (December 20)
- ☑ Create the fully connected layer for the machine learning model (December 29)
- ☑ Check for model accuracy (January 5)
  - ☑ Test the model (January 1)
  - ☑ Benchmark accuracy (January 1)
  - ☑ Make iterations and adjust weights (January 2 - 5)
  - ☑ Final Model test + Benchmark (January 5)
- ☑ Introduction to final paper (December 15)
- ☑ Literature review (January)
- ☑ Rough Draft Procedure (February)
  - ☑ Datasets (January 24)

- ☑ ~~Data Preprocessing (January 26)~~
- ☑ ~~Convolutional Neural Network (January 30)~~
- ☑ ~~Training and validation (February 1)~~
- ☑ ~~Citations / proofread (February 1)~~
- ☑ ~~Final Draft Procedure (February 13)~~
  - ☑ ~~Send to Dr. Garcia for review (February 1)~~
  - ☑ ~~Edits and final changes (February 13)~~
- ☑ ~~Oral Presentation (February 26)~~
  - ☑ ~~Introduction (February 5)~~
  - ☑ ~~Punchline (February 5)~~
  - ☑ ~~Methods (February 9)~~
  - ☑ ~~Results (February 15)~~
  - ☑ ~~Significance (February 20)~~
  - ☑ ~~Rehearsal (February 22)~~
  - ☑ ~~Presentation (February 26)~~
- ☑ ~~Science Fair (March 4)~~
  - ☑ ~~Design Poster (February 28)~~
  - ☑ ~~Make graphs (February 28)~~
  - ☑ ~~Decorate & make poster (March 1)~~
- ☐ CYSF (April 11 - 13)
  - ☐ Complete portal (March 15)
    - ☐ Record video (March __)
- ☐ Final Paper (May tbd)
  - ☐ Results
  - ☐ Discussion
  - ☐ Acknowledge

☐ References

Daily Notes

2023

September Notes

September 1, 2023

1. Schedule
   - Task each class (at least 90 mins)
2. Organization
3. Content
- Citing: Cite the original source but you do not need to cite the source
- Sources: IEEE, Research Gate, Google Scholar, Arxiv

September 3, 2023

- Sent an email to Dr. Pieper
- Wanted to change to a topic more machine learning based
- Asked to be provided with some more resources / a general direction on what I should try to research
- Asked to see if I could do something similar to his current research (machine learning the amount of insulin to inject for diabetes)
- Gave them my schedule
- Asked for a further meeting date to discuss more if needed
- Waiting on response (articles and Dr. Pieper's schedule)

September 6, 2023

- Biweekly checks /15 (10%)

- Mentor eval /15 (10%)
- Sign into Paperpile - Reference Manager
- Citations immediately after using source
- Download annotated pdf
- ASP Meetings with Dr. Garcia on September 8, September 20
- Sent a follow up email, asking to see if Dr. Pieper saw my email (still waiting on response)

September 8, 2023

- Meeting with Dr. Garcia
- Sent another follow up email to Dr. Pieper (responded, set weekly meetings on monday at 7:00)
- Introduction to Machine Learning, Neural Networks, and Deep Learning (check background research)
- Talk to Dr. Pieper about math and programming needed
- See what matlab experience or other experience
- Decide what project
- What courses / requirements
- Set weekly meetings
- See if there are other people in the lab if it is too busy
- Research proposal
- Innovation project (goals)
- Ask what available times, tell what available times, set up weekly meetings, talk what I want to happen

September 11, 2023

- Meeting with Dr. Pieper at 7:00
- Things to ask Dr. Pieper

- ○ What are some project more related to med and machine / more deeply machine learning wise should I attempt and maybe do
- ○ What method of machine learning should I use
- ○ What courses / Requirements should or do I need to take
- ○ If I should use matlab or pytorch or …
- ○ What tasks should I complete for the next week
- ○ What sources should I look at
- ○ What citing method should I use (IEEE)?
- Dr. Pieper meeting notes
  - ○ Weekly meetings set with Dr. Pieper on Mondays at 7:00
  - ○ Pattern recognition
  - ○ Have the machine learn the characteristics of a person (glucose intake and outake)
  - ○ Have a look at https://www.cs.toronto.edu/~hinton/
  - ○ recurrent neural networks
  - ○ Have a look at pytorch
- Tasks for next week
  - ○ Have a look at pytorch and explain a little
  - ○ Read some of these sources https://www.cs.toronto.edu/~hinton/

September 12, 2023

- Young persons accessing UofC labs (not needed for me)
- Meet mentors during class time (Email attend at Webber)
- CC / Email Dr. Garcia
- Read introduction to neural networks background research (refer to background research)

September 14, 2023

- Innovation project (most likely, goals)
- Study (data already given)
- Experiment (choose your own variables and methods)
- Come up with an outline
- Finish introduction to neural networks background research (refer to background research)
- Read section one of the https://www.cs.toronto.edu/~hinton/ sources (refer to background research)
- Bring an outline to Dr. Pieper

September 16, 2023

- Research pytorch and takes notes on pytorch
- Downloaded (Pytorch, Jupyter Notebook, and Anaconda)
- Created small ml code project (detecting iris flower types 28/30)
- https://github.com/OwenKuang/Iris-Classification-ML (Code)

September 18, 2023

- Research Proposal
- Working title (enough specificity so that people know exactly what you are doing)
- Introduction
    - Need for your study
    - Short term and long term
- Are there variables in the proposal?
- Long term goals (Goals the field is trying to achieve or in the future that could be achieved)
- Short term goals (what you can achieve with your project now

- Question
- Hypothesis
- No word count
- Several Paged introduction
- 12 Point font
- Double spaced
- Figures (need legend)
- Hinton
- Dr Pieper Meeting notes
    - Could look at x rays or electrocardiograms
    - Look for tumors
    - Can check if they have a heart attack or predict a heart attack
    - 5 electrodes (5 neurons)
    - Output (5 more per neuron)
    - Yes no decision (predict heart attack)
    - Look for healthy ECG and unhealthy ECG
    - Time series data (measure every half a second or every 10th of a second)
- Task for next week
    - If ECG is not good, look for other
    - Next week written proposal about an objective (identify the probability of a heart attack of a victim)
    - One sentence statement
- Task for coming weeks
    - Simpler problem
    - 20 - 30 data points
    - 5 -6 data points
    - Divide into quadrants
    - Try to identify different characters

September 20, 2023

- Meeting with Dr. Garcia

- Research proposal

- Talk about statistics of cardiovascular diseases

- How is it diagnosed or predicted

- Any prevention methods

- Machine learning and prediction

- Goal to contribute to a new emerging predict and treat

- Could be to improve on the accuracy

- Other methods to look

- Can repeat something that has been done

- Why was it more or less accurate

- Confirm existing data and prediction results

- Compare methods for efficiency

- Narrow (try a different method and compare)

- Research ECG based machine learning algorithms (check background research)

- Goal, method, databases, in the draft for Dr. Pieper

- What algo has been or will be used?

September 22, 2023

- Logbook September 29th

- Weekly meetings (what is done, ask questions, send items)

- Need to include october and september schedule

- Presentation 11th - 17th

- Research general heart diseases and attacks

- Get understanding of the field
- Refer to background research
- Logbook (Organization (content), schedule, communication (evidence)

September 25, 2023

Proposal Written for Dr. Pieper

📄 ECG and Machine Learning Proposal

Dr. Pieper Meeting Notes

- 5 Signals
- 5 Pictures
- Identify the combination
- Compounding factors
- Percentage / Likelihood
- Different classifications
- Ultimate Goal (Probability) not yes or no
- Being more explicit on the PQRST inputs
- Other information that relates to issues
- Complete by next week search the
- PQRST
- Other information
- Table form

September 26th

Research Proposal Planning

- Working Title
- Abstract
- Introduction (some ML methods, PQRST waves, additional ECG information, how they could relate)

- Objective (to quickly identify percentage or probability of certain heart conditions and diseases)
- Variables (Not Needed)
- Questions/Hypothesis (how can machine learning be used to increase ECG accuracy)
- Methodology (first take data, use this model, etc)
- Significance (Allows for doctors to use the algorithm to quickly detect if there are any issues with the ECG reading with less error especially during rushed situations, also allows for people who need an ecg reading in a situation where it is inconvenient for them to go to the doctor for an ECG reading)
- References ()

September 28, 2023

- Sources (20% of canadians affected by disease A) need to give source
- Need a peer review source to back up the data
- Reputable sources (pubmed, NHS, Google Scholar, Government of Canada or Gov of USA, or other government websites)
- Descriptions (pubmed, textbook)
- Use IEEE (numbered)
- Can't have source 3 before source 1 etc.
- Due Tomorrow: Logbook
- Research 3 sources sent by Dr. Garcia
- Start on research proposal

October Notes

October 1, 2023

- Researched ECG waves and what heart diseases could lead to which deformities (Refer to October background research)
- Worked on table with different heart diseases and ECG readings that relate to them 📄 Owen ASP Table

October 2, 2023

- Calendar must include what needs to be accomplished every class
- Could include meeting w/ mentor w/ Dr. Garcia
- Could include what was researched
- Write proposal (which part of proposal)
- Background information, add enough information (paste work done digitally or summarize)
- Tasks have to be small enough so that it can be completed during class time
- Work on proposal outline

Meeting with Dr. Garcia

- Writing the sections
- Discuss draft with Dr. Pieper
- Complete draft, latest Oct 13, 2023
- Discuss outline
- Overschedule, make deadlines but may be done before
- Dr. Pieper contacted cardiologist, however they needed more time to prepare (MEETING CANCELED)

October 4, 2023

- Absent
- Unable to research IEEE style, learn about IEEE style next class

October 6, 2023

- Take 15 minutes to write tasks
- (Write entries into your logbook class)
- Have a plan of the deadline however, can talk about delays and how you are behind
- September notebook excused
- Take picture of October logbook and resubmit logbook
- Complete IEEE style - refer to background research
- Work on proposal introduction (deaths, traditional methodology, and ECG wave information) - refer to Links to Written Documents
- Write the proposal like how Dr. Garcia outlined but references and citations in IEEE style

October 11, 2023

- Turn it in into google classroom and turn it in
- Oral presentations
  - Punchline (elevator pitch,this is goal, this is briefly on what I am trying to do, no details)
  - 10 minutes
  - 10-12 slides
  - Background
  - Research question goals
  - Methodology (graphics)
  - Little text good graphics
  - Big font text
  - Flow diagram
  - significance
- Reard Arnav and Michael's research proposal

- Worked on research proposal introduction

October 12, 2023

- Complete for next meeting
  - Check forwarded email and read on the two sources provided by cardiologist
- Meeting with Dr. Pieper notes
  - Contacted, will get a masters student that will be able to help us
  - Sign up for website that allows you to see ECG data
  - Focus in on a couple of things in the table (instead of all of the things listed)
  - Masters student can help with libraries needed for python
  - Sign up for access to dataset
    https://www.physionet.org/about/database/
  - Automate process of ECG interpretation (replicate what a person can do) for goal
  - Continue working on proposal

October 13, 2023

Meeting with dr. Garcia

- October 25 deadline for research proposal (changed from october 19)
- Feedback from cardiologist and Dr. Pieper before send to Dr. Garcia
- Include a task checker
- Oral presentation deadline November 6
- Webber Academy Science Fair (March 4)
- March 14 Online portal for CYSF
- Join science fair google classroom
- Submit Basic Project / Ethics (need to submit)

- Poster (cannot be bigger than the trifold, provided with trifold)
- Digital books have to be physical
- Complete research proposal introduction
- Research two sources given by cardiologist (refer to october research)

October 17, 2023

- March 4th school science fair
- Complete ethics for science fair portal
- Research two sources given by cardiologist (refer to october research)

October 19, 2023

- Research proposal
- Completed Introduction (behind was due on October 17th)
- Completed Objectives
- Compled Questions
- Emailed Dr. Pieper for meeting on Thursday or Friday (No response)
- Questions to ask Dr. Pieper For Meeting
    - Ask a little bit more about time frequency matrix (resolved)
    - What exactly is the input for time frequency matrix (resolved)
    - Is the pooling layer similar to that of the convolutional layer in the sense that it cuts the data down? (Resolved)

October 22, 2023

- Sent follow up email to Dr. Pieper to confirm meeting on Monday (meeting confirmed for 9:00 on Monday)
- Completed methodology
- Completed significance
- Completed References

- Draft of research proposal completed (ahead, was due on October 23)

October 23, 2023

- November 6 Oral Pres.
- Talk with Dr. Garcia about reviewing draft of research proposal (will take a look and give feedback)
- Marks in report card
    - October Logbook
    - Mentor Evaluation
    - Written Research Proposal
    - Oral Research Proposal
    - Biweekly checks
- Completed Oral Presentation Title slide
- Oral Presentation
    - Punchline (tells people what you are talking about)
    - Clarify terms, make it simple
    - Background
    - Methods
    - First slide (flow diagram of methodology)
    - Significance
    - Engagement with the audience
    - Large text, little text
- Meeting with Dr. Pieper at 9:00

    Questions:
    - Should I focus on just one arrhythmia or heart disease so far?
    - Missing information about CNN?

    Notes:
    - Checked over research proposal (refer to research proposal)
    - Extract / download database

- Finish proposal
- Figure out a sample rate (100 times a second?)

October 25, 2023

- Dr. Garcia gave feedback on ASP project proposal
- Worked on revising research proposal with feedback from Dr. Garcia
- Title / Procedure
- RQ (goals)
- Flow Diagram (how is my methodology allowing me to answer the research question)
- Significance
- Presentation dates
    - October 31:
        - Jessica
        - Cooper
        - Brynn
    - November 6:
        - Tosin and Zainab
        - Natalie
        - Owen
    - November 8:
        - Mariska
        - Elliot
        - Emily
    - November 15:
        - Vincent
        - Joel and Ashank
        - Tiffany
    - Paragraphs to add, Do not limit scope to machine learning,

October 27, 2023

- Completed Research Proposal Introduction edits
- Completed Research Proposal Conclusion edits
- Completed Research Proposal Objective edits

October 29, 2023

- Downloaded MIT arrhythmia database
- Converted files into CSV format
- Plotted values onto graph
- Found sampling frequency (360/second)
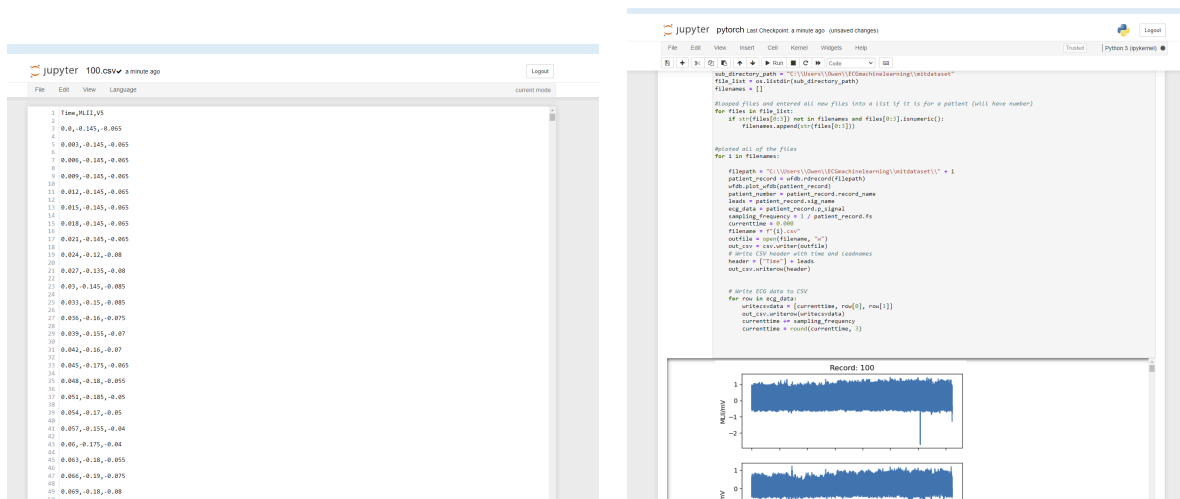
October 30, 2023

- Meeting with Dr. Pieper Notes (No meeting)

October 31, 2023

- Presentation Notes

Jessica (Design and construction of CRISPR / Cas9-Mediated Knockout Plasmids for Autism-Risk Genes)

- Strong genetic base for autism
- Crispr Cas 9 novel method
- Introduction

- 1 in 44
- 400 - 1000 autism risk genes
- Bacteria plasmid can replicate independently
- 2 gRNA containing gene will knock out if one fails

Objectives
- See how to develop 2 gRNA genes

Variables

- Molar vector to insert ratio (independent)
- Bacteria cultures and reagents (controlled
- Transformation efficiency of plasmids into competent cells (dependent)
- Quality of DNA oligos, reagents and competent bacteria cells (compounding)

Methodology
- Snapgene
- CRISPOR

Cooper (changes in complement protein deposition and its role in microglia mediated synaptic pruning)
- Brain pruning, extra dendrites are removed
- Concussions may lead to degenerative brain diseases

Previous studies
- Microglia loss -> synaptic pruning impairment -> behavior deficits

Research Question
- How do repeated mild traumatic brain injuries affect c3 and c1q complement protein expression in motor cortex of mice

Variables

Methodology

Significance
- Completed Research Proposal

November Notes

November 1, 2023

- Literature review December 15th
- Final Paper (Intro → Methods → Results → Analysis)
- Background information (if no correction copy and paste) otherwise fix and then use as literature review
- Methods February
- Results March
- Analysis April
- Completed Paper May

November 5, 2023

- Completed oral presentation slides
- Refer to written documents (google slides)

November 6, 2023

- Practice presentation with Dr. Garcia
    - Explain more p q r s t waves (what the represent)
    - Simpler explanation of convolutional layer
    - Explain that the CNN showcased is a simpler version for detecting numbers
    - Show how bradyarrhythmia and tachyarrhythmia can be detected with ecg
    - Enlarge all of the text and images, fill up the space
    - Explain fourier transforms a little bit more in depth and how that can be used as the input for the cnn
    - Explain what database will be used and what the database comprises of https://physionet.org/content/ptb-xl/1.0.3/

November 7, 2023

- Made all edits previously mentioned by Dr. Garcia
- Dr. Pieper was unable to make Monday meetings, switched to Thursday
- Did not reply to confirm time, send reminder email to confirm time tomorrow

November 8, 2023

- Completed oral presentation
- Sent an email to Dr. Pieper, asking for confirmation of meeting time as well as current progress.
- Mariska Sugar on muscle health
    - Cytokines (positive impact)
    - Macrophage Immune cells
    - Only males but need to look at females as well as they could gain different diseases and reaction
- Natalie compression garments on POTS
    - Responding (Heart rate stroke heart rate variability)
    - Independent (Compression garments)
    - Elliot (gut microbiota)
        - Dysbiosis
        - Autism (lack of social communication skills)
        - Gut and brain neural development

November 9, 2023

Meeting with Dr. Pieper Notes:

- References for CNN + Fourier transforms
- See how code fits with the problem
- Write some code that segments the data
- 1 - 2 minutes in the saw tooth pattern that indicates trouble or not trouble
- What patterns the CNN should be looking for

November 15, 2023

Tosin and ZZ (effects of ketamine and marijuana on neural function)

- Ketamine blocks percepter (hallucination)
- How do ketamine and CBD oil exposure affect neuronal cell viability
- What effect does the exposure of ketamine and CBD oil have on neural growth
- How does the synaptic connectivity of neurons exposed to ketamine and CBD oil compare to those that are not exposed
- Research for specific features on Atrial Fibrillation
- Completed STFT on ECG readings and displayed it on a time frequency plot (split into 10 second segments)
- Coding CNN references
  https://www.analyticsvidhya.com/blog/2019/10/building-image-classification-models-cnn-pytorch/
- https://pytorch.org/tutorials/beginner/blitz/cifar10_tutorial.html

```
# Set the parameters for the spectrogram (you may need to adjust these)
fs = 360  # Set your sampling frequency
noverlap = 200

# Create a time-frequency plot using specgram (STFT)
plt.figure(figsize=(10, 6))
plt.specgram(your_data, Fs=fs, noverlap=noverlap, cmap='viridis')

# Add labels and title
plt.xlabel('Time (s)')
plt.ylabel('Frequency (Hz)')
plt.title('Time-Frequency Plot using STFT')

# Add a colorbar for reference
plt.colorbar(label='Intensity')

# Show the plot
plt.show()
```


November 16, 2023

Meeting with Dr. Pieper Notes

(no response / meeting)

- Sent an email to let Dr.Pieper know what I did in the past week and what I will be working on, also asked to see if he found some sources and also asked to set another meeting time

November 17, 2023

Read

https://www.analyticsvidhya.com/blog/2019/10/building-image-classification-models-cnn-pytorch/ and created code (refer to code below)

Completed ethics in CYSF and basic background information

Meeting with Dr. Garcia notes:

- Email him one day before the link for reminder
- Deeper understanding of the fourier transform / time analysis
- Send him progress (once or twice a week, mix of progress or asking for help)
- Add the robotics schedule into the calendar

CODE:

Importing libraries needed

```
from skimage.io import imread
import matplotlib.pyplot as plt
%matplotlib inline

from sklearn.model_selection import train_test_split


from sklearn.metrics import accuracy_score
from tqdm import tqdm

import torch
from torch.autograd import Variable
from torch.nn import Linear, ReLU, CrossEntropyLoss, Sequential, Conv2d, MaxPool2d, Module, Softmax, BatchNorm2d, Dropout
from torch.optim import Adam, SGD
```

Loaded the dataset

```
directory_path = "C:\\Users\\Owen\\ECGmachinelearning"
filecsv = []
file_list = os.listdir(directory_path)
```

```
# Iterate through the CSVs
for files in file_list:
    if files[0:3].isdigit() and files[0:3] not in filecsv:
        filecsv.append(files[0:3])

# Iterate through the CSVs
for i in filecsv:
    df = pd.read_csv(os.path.join(directory_path, i + ".csv"))

train.head()
```

November 21, 2023

Add goals section to the introduction of the final literature review (at the end)

Encountered problem with normalization (unwanted noise from electrocardiogram readings were detected and therefore needed to be filtered out otherwise the normalization would be affected, refer to background research)

Continued to follow along with source:

https://www.analyticsvidhya.com/blog/2019/10/building-image-classification-models-cnn-pytorch/

CODE (tried normalize frequency, set cut offs however it was still not correct and lots of intensity levels at 0) :

```
directory_path = "C:\\Users\\Owen\\ECGmachinelearning"
filecsv = []
file_list = os.listdir(directory_path)

# Iterate through the CSVs
for files in file_list:
   if files[0:3].isdigit() and files[0:3] not in filecsv:
      filecsv.append(files[0:3])

# Iterate through the CSVs
for i in filecsv:
   df = pd.read_csv(os.path.join(directory_path, i + ".csv"))
```

```python
# Check if 'MLII' column is present in the DataFrame
if 'MLII' not in df.columns:
    print(f"Skipping {i} as it does not contain the 'MLII' column.")
    continue

# Create an empty list to store DataFrames for each CSV file
file_rows_list = []

# Iterate through rows and check the condition
total_time = 0
for index, row in df.iterrows():
    total_time = row['Time']

    if total_time >= 10:
        break  # Stop when the condition is met

    # Convert the row to a DataFrame and append it to the list
    file_rows_list.append(pd.DataFrame(row).T)

# Concatenate the list of DataFrames into a single DataFrame for each CSV file
selected_rows_df = pd.concat(file_rows_list, ignore_index=True)

time = selected_rows_df['Time']
raw_data = selected_rows_df['MLII']

# Set the parameters for the spectrogram
fs = 360  #sampling frequency
noverlap = 200

# Create a time-frequency plot using specgram (STFT)
plt.figure(figsize=(10, 6))
spec, freqs, t, im = plt.specgram(raw_data, Fs=fs, noverlap=noverlap,
cmap='viridis')

# Close the plot to prevent it from being displayed
plt.close()
# Display the image
plt.imshow(image, cmap='viridis', aspect='auto')
plt.xlabel('Time (s)')
plt.ylabel('Frequency (Hz)')
plt.title('Spectrogram Image')
plt.colorbar(label='Intensity')
```

```
plt.show()
```

November 23, 2023

Amy: neurobiological abnormalities of the brain risk factors associated with disproportionate incarceration for serious violent crime of males compared to females and young compared to older individuals?

Ashank and Joel: Mimotopes that cause type one diabetes

Tiffany: Disparities between public available and internally provided information on psychological interviewers

Inform mentors about winter break, and how the break aligns with theirs

Read source about normalization and how to remove unwanted noise

Meeting with Dr. Pieper notes:

- Work on the filtering
- How many neurons
- Generate an array of numbers instead of the picture of the spectogram
- Average of the numbers between intensity levels (0-5hz)]
- Leave the code open with variables

November 27, 2023

- Continue reading papers
- Logbook due 30th
- Leave Februrary for preparation for posters and presentations
- School Science fair March 4th
- Questions for mentor:
    - What are your plans for winter break going to be like?
    - Are you available to meet during the winter break or not?
    - Tell them about midterm plans and that I probably will not be working on ASP during that date

- Tell them when winter break starts (December 21, and ask about their winter break)
- Main focus
  - Data collection
  - Deep understanding of data collection
  - How does your data link to your hypothesis
  - What do these numbers mean, what am I actually concluding, did I answer my research question or not?
  - What would I need to do or take into account for the future and to improve my research?
  - Be critical of your analysis of data and come up with new ways to test what you are trying to test
- Worked on a deeper understanding of time frequency domain and how to apply filter and short time fourier transform (refer to November research)

November 28,2023

Converted to grayscale and obtained pixel values

```python
directory_path = "C:\\Users\\Owen\\ECGmachinelearning"
output_csv_directory = "C:\\Users\\Owen\\output_spectrogram_data"  # Set desired
output directory
filecsv = []
file_list = os.listdir(directory_path)

# Create the output directory if it doesn't exist
os.makedirs(output_csv_directory, exist_ok=True)

# iterate through the csvs
for record_number, file_name in enumerate(file_list, start=1):
    if str(file_name[0:3]).isnumeric():
        filecsv.append(str(file_name))

# iterate through the csvs
for record_number, file_name in enumerate(filecsv, start=1):
    df = pd.read_csv(os.path.join(directory_path, file_name))
```

```python
# Check if 'MLII' column is present in the DataFrame
if 'MLII' not in df.columns:
    print(f"Skipping {file_name} as it does not contain the 'MLII' column.")
    continue

# Create an empty list to store DataFrames for each CSV file
file_rows_list = []

# Iterate through rows and check the condition
total_time = 0
for index, row in df.iterrows():
    total_time = row['Time']

    if total_time >= 10:
        break  # Stop when the condition is met

    # Convert the row to a DataFrame and append it to the list
    file_rows_list.append(pd.DataFrame(row).T)

# Concatenate the list of DataFrames into a single DataFrame for each CSV file
selected_rows_df = pd.concat(file_rows_list, ignore_index=True)

time = selected_rows_df['Time']
raw_data = selected_rows_df['MLII']

# Set the parameters for the spectrogram (you may need to adjust these)
fs = 360
noverlap = 200

# Create a time-frequency plot using specgram (STFT) with grayscale colormap
plt.figure(figsize=(10, 6))
specgram_values, freqs, times, _ = plt.specgram(raw_data, Fs=fs, noverlap=noverlap,
cmap='gray')

# Round pixel values to the nearest 10 decimal points
rounded_specgram_values = np.round(specgram_values, 10)

# Create a CSV file for each record
filename = f"{record_number}.csv"
output_csv_path = os.path.join(output_csv_directory, filename)
```

```python
    with open(output_csv_path, "w", newline='') as outfile:
        out_csv = csv.writer(outfile)

        # Write CSV header with time and frequency bins
        header = ["Time"] + [f"Freq_{int(freq)}Hz" for freq in freqs]
        out_csv.writerow(header)

        # Write spectrogram data to CSV
        for idx, time_value in enumerate(times):
            writecsvdata = [round(time_value, 3)] + list(rounded_specgram_values[:,
idx])
            out_csv.writerow(writecsvdata)

    # Display the plot
    plt.xlabel('Time (s)')
    plt.ylabel('Frequency (Hz)')
    plt.title(f'Grayscale Time-Frequency Plot using STFT - Record {record_number}')
    plt.show()

# Output a confirmation message
print(f"Data exported to {output_csv_directory}")
```

November 29,2023

Februrary 26 Oral Presentation 15 minutes

Know the poster very well (fancy poster != advantage)

GIVE PUNCHLINE, LESS THAN ONE MINUTE EXPLANATION OF PROJECT

Research about number of neurons in dataset (refer to background research)

Work averaging and reducing spatial size of the input

Spacial size for input layer: 36 * 36

Concrete plan for when to finish

November 30,2023

Meeting with Dr. Pieper Notes: (he did not join the meet)

December Notes

December 1, 2023

- Worked on segmenting the frequency by increments of 5 hz instead of 1hz (highlighted lines are the added lines)

```python
directory_path = "C:\\Users\\Owen\\ECGmachinelearning"
output_csv_directory = "C:\\Users\\Owen\\output_spectrogram_data"  # Set desired
output directory

filecsv = []
file_list = os.listdir(directory_path)

# Create the output directory if it doesn't exist
os.makedirs(output_csv_directory, exist_ok=True)

# Iterate through the csvs
for record_number, file_name in enumerate(file_list, start=1):
    if str(file_name[0:3]).isnumeric():
        filecsv.append(str(file_name))

# Iterate through the csvs
for record_number, file_name in enumerate(filecsv, start=1):
    df = pd.read_csv(os.path.join(directory_path, file_name))

    # Check if 'MLII' column is present in the DataFrame
    if 'MLII' not in df.columns:
        print(f"Skipping {file_name} as it does not contain the 'MLII' column.")
        continue

    # Create an empty list to store DataFrames for each CSV file
    file_rows_list = []

    # Iterate through rows and check the condition
    total_time = 0
    for index, row in df.iterrows():
        total_time = row['Time']

        if total_time >= 10:
            break  # Stop when the condition is met
```

```python
        # Convert the row to a DataFrame and append it to the list
        file_rows_list.append(pd.DataFrame(row).T)

    # Concatenate the list of DataFrames into a single DataFrame for each CSV file
    selected_rows_df = pd.concat(file_rows_list, ignore_index=True)


    time = selected_rows_df['Time']
    raw_data = selected_rows_df['MLII']


    # Set the parameters for the spectrogram
    fs = 360
    noverlap = 200
    freq_increment = 5  # Set the frequency increment


    # Create a time-frequency plot using specgram (STFT) with grayscale colormap
    plt.figure(figsize=(10, 6))
    specgram_values, freqs, times, _ = plt.specgram(
        raw_data, Fs=fs, noverlap=noverlap, cmap='gray')


    # Increment frequency values by freq_increment
    freqs = np.arange(0, fs / 2 + freq_increment, freq_increment)
    specgram_values = specgram_values[:len(freqs), :]


    # Calculate the average value of the pixels
    avg_specgram_values = np.mean(specgram_values, axis=1)


    # Round pixel values to the nearest 10 decimal points
    rounded_avg_specgram_values = np.round(avg_specgram_values, 10)


    # Create a CSV file for each record
    filename = f"{record_number}.csv"
    output_csv_path = os.path.join(output_csv_directory, filename)


    with open(output_csv_path, "w", newline='') as outfile:
        out_csv = csv.writer(outfile)


        # Write CSV header with time and frequency bins
        header = ["Time"] + [f"Freq_{int(freq)}Hz" for freq in freqs]
        out_csv.writerow(header)


        # Write averaged spectrogram data to CSV
        writecsvdata = [round(total_time, 3)] + list(rounded_avg_specgram_values)
```

```
    out_csv.writerow(writecsvdata)


# Display the plot
plt.xlabel('Time (s)')
plt.ylabel('Frequency (Hz)')
plt.title(f'Grayscale Time-Frequency Plot using STFT - Record {record_number}')
plt.show()


# Output a confirmation message
print(f"Data exported to {output_csv_directory}")
```

- Followed this source to start creating the convolutional layer of the cnn
  https://machinelearningmastery.com/building-a-convolutional-neural-network-in-pytorch/ (refer to background research and code)

Dr. Garcia notes:

- Schedule a meeting for Monday with Dr. Pieper
- General time frame for how long each task takes

December 4, 2023

Dr. Garcia notes:

- Needs to be detailed
- Work on research proposal is not detailed enough, link to research proposal
- Task dates

Mentor Notes:

- (Did not see Dr. Pieper's email notification confirming for Monday as I had robotics when he sent the email)
- Dear Dr. Pieper,
  Sorry for the late reply, unfortunately I did not see the email you sent until now. Would you be able to meet today at 7:30 or do you want to move the meeting to the usual Thursday date?

Thanks, Owen

- Dr. Pieper confirmed time for Tuesday

    Meeting Notes:

    - Questions to ask / answers:

        - When is your winter break, and what that schedule looks like for you (start meeting evening on the 21st, email him schedule)

        - Midterm week January 12 - 19 Most likely will not be working on the project

        - Main focus will be to complete it during winter break, which is from December 21 - January 5th (will you be able to meet anytime during then or not?)

        - Compress frequencies

        - Research built in pytorch functions

Task Notes:

- Read source about developing convolutional layer of neural network (Refer to background research)

    - Worked on loading the dataset to a tensor (Refer to code)

    - Worked on converting the dataframe to a 2D list (refer to code)

    - Worked on understanding number of parameters (Refer to background research)

Code:

Converting dataset to tensor

```
list_of_tensors = [torch.tensor(np.array(df)) for df in data]
torch.stack(list_of_tensors)
```

Machine learning model

4 learned filters that are 3 * 3→ 4 learned filters that are 3 * 3

```python
class Net(Module):
    def __init__(self):
        super(Net, self).__init__()

        self.cnn_layers = Sequential(
            # first convolutional layer, 4 filters, 3x3, stride of 1 and
an external added padding of 1
            Conv2d(1, 4, kernel_size=3, stride=1, padding=1),
            # normalization of a 4D object (first conv layer had 4
filters, making the 1D object become 4D)
            BatchNorm2d(4),
            ReLU(inplace=True),
            #  first pooling layer
            MaxPool2d(kernel_size=2, stride=2),
            #  second convolutional layer, kernel size 4 filters, but stay
at a dimension of 4, with kernel size 3x3, stride 1, and padding 1
            Conv2d(4, 4, kernel_size=3, stride=1, padding=1),
            #  normalization
            BatchNorm2d(4),
            ReLU(inplace=True),
            #  second pooling layer
            MaxPool2d(kernel_size=2, stride=2),
        )

        self.linear_layers = Sequential(
        #  fully connected layer, goes from 216 tensor size to 41 tensor
size
            Linear(216, 41),
            ReLU(inplace=True),
            # final classification, 41 neurons to 5 (different classes)
            Linear(41, 5)
        )

        # Set the data type of parameters to torch.float32
        self.apply(lambda m: m.to(torch.float32) if 'weight' in
m._parameters else None)
```

```python
        self.apply(lambda m: m.to(torch.float32) if 'bias' in
m._parameters else None)

    def forward(self, x):


        x = self.cnn_layers(x)

        x = x.view(x.size(0), -1)

        x = self.linear_layers(x)


        return x
```

December 7, 2023

Dr. Garcia Notes:

- Produce document
- Plan for Science Fair (Rest of December, January, and February until presentation in class)
- March fourth for science fair goal
- Each week what are your tasks
- Big tasks identified with deadlines when completed
    - Smaller tasks
- Poster - Design/create
- Plan into each week between now (what are the tasks to complete each project)

Task Notes:

- Read on functions from pytroch and understand (For mentor and refer to background research)

December 11, 2023

Task Notes:

- Studied for Bio

Dr. Garcia Notes:

- Add time for how long each task will take
- Final literature review is not due in Februrary (only methodology)
- 15 Minutes for oral presentation and Science fair presentation
- Slideshow for oral pres
- Trifold for Science Fair

December 12, 2023

Dr. Pieper meeting notes:

- Reformat data (cut high frequency after 100hz)
- Approx 2 seconds
- Test the neural network see if it makes any sense

December 13, 2023

Task Notes:

- Worked on making edits to introduction of lit review

    Dr. Garcia comment: Title too vague

    - Edit: **Use of Convolutional Neural Network to identify and predict heart attacks from ECG signals converted into frequency domain**

    Dr. Garcia comment: (small errors like capitalization and world choice)

    - Edit: Lowercased arrhythmia, used the word "fortunately" instead of luckily, capitalized "ST", and also added space between the citation.

Dr. Garcia Comment: ECG does not give wavelengths, and also asked about what other Fourier transforms:

- Edit: The Short-Time Fourier Transform (STFT), or another Fourier transform like the Fast Fourier Transform is needed as it will be able to convert the electrical signals from time domain, which is in 2 dimensions, to frequency domain, which is 3 dimensions. This conversion will be in the form of a spectogram and the luminosity levels of each pixel will be the data used as the input to the CNN [8].

December 15, 2023

Task Notes: Worked on Making edits to introduction for lit review

Dr. Garcia Comment: Need citation when making claim

- Edit: Machine learning however, is not limited to the detection of arrhythmia and has also been implemented in the different ACS detections [10]. (Found a source to cite and also added footnote)

Dr. Garcia Comment: Do not need to say it is a type of neural network

- Edit: The model utilizes a CNN and was trained to a confidence level of 95% by analyzing the slopes and amplitudes of the ST segments [11].

Dr. Garcia comment: Change word choice:

- Edi: With the new implementation of CNN and fourier transforms in ECG reading for ACS and arrhythmias, the goals of this study are as follows.

December 19, 2023

Task notes:

- Downsized to 2 seconds
- Cut off frequency at 100hz
- Implemented CNN

Code:

```
    #
            for freq in freqs:
                if count > 72:
                    break
                count += 1
                if count % 2 ==0:
                    ap = str(freq) + "hz"
                    freqlist.append(ap)

            header = ["Time"] + freqlist
            out_csv.writerow(header)
            #create list to load onto a tensor

            tenslist = []
            for idx, time_value in enumerate(times):
                freqcount = 0
                avg_pixel_values = []
                 # if time is greater than 2, break
                if time_value > 2:
                    break
#get a cut of at 100 hz which is a dimension of 72 / 4 (18)
                while freqcount != 72:
                    totalpixel = 0
                    for i in range(2):
                        totalpixel +=
rounded_specgram_values[freqcount][idx]
                        freqcount += 1
                    avg_pixel_values.append(totalpixel / 2)
                tenslist.append(avg_pixel_values)
                writecsvdata =  [round(time_value, 7)] + avg_pixel_values
                out_csv.writerow(writecsvdata)
```

December 22, 2023

Meeting with Dr. Pieper:

- He was unable to meet as he was on a last minute christmas vacation
  (resecheduled to tomorrow)

December 23, 2023

Meeting with Dr. Pieper:

- Model was not working because of type error

Type error:

Cell **In[58], line 111**

109 inputs, labels = batch['data'].float(), batch['label']

110 outputs = model(inputs)

--> **111** loss = criterion(outputs, labels)

112 loss.backward()

113 optimizer.step()

File **~\anaconda3\lib\site-packages\torch\nn\modules\module.py:1501**, in

Module._call_impl**(self, *args, **kwargs)**

1496 # If we don't have any hooks, we want to skip the rest of the logic in

1497 # this function, and just call forward.

1498 **if not** (self._backward_hooks **or** self._backward_pre_hooks **or** self._forward_hooks **or**

self._forward_pre_hooks

1499   **or** _global_backward_pre_hooks **or** _global_backward_hooks

1500   **or** _global_forward_hooks **or** _global_forward_pre_hooks):

-> **1501**   **return** forward_call(*args, **kwargs)

1502 # Do not call functions when jit is used

1503 full_backward_hooks, non_full_backward_hooks = [], []

File **~\anaconda3\lib\site-packages\torch\nn\modules\loss.py:1174**, in

CrossEntropyLoss.forward**(self, input, target)**

1173 **def** forward(self, input: Tensor, target: Tensor) -> Tensor:

-> **1174**   **return** F.cross_entropy(input, target, weight=self.weight,

1175   ignore_index=self.ignore_index, reduction=self.reduction,

1176   label_smoothing=self.label_smoothing)

File **~\anaconda3\lib\site-packages\torch\nn\functional.py:3029**, in cross_entropy**(input, target,**

weight, size_average, ignore_index, reduce, reduction, label_smoothing)

3027 **if** size_average **is not None or** reduce **is not None**:

```
3028    reduction = _Reduction.legacy_get_string(size_average, reduce)
-> 3029 return torch._C._nn.cross_entropy_loss(input, target, weight,
_Reduction.get_enum(reduction), ignore_index, label_smoothing)
```

RuntimeError: expected scalar type Long but found Int

2024

January Notes

January 2

Task notes:

- Found out that the model turned the tensors into a different data type after every output solved by manually converting to float32 (which is equivalent to native python long)

Code (highlighted were the changes):

```python
def forward(self, x):
    x = x.float()


    x = self.cnn_layers(x)


    x = x.view(x.size(0), -1)

    # Explicitly cast the tensor to float32 before going through
linear_layers
    x = x.to(torch.float32)


    x = self.linear_layers(x)


    return x
```

January 4

Dr. Pieper Meeting Notes:

- report on overfitting
- Expand Time, double the time
- Send him a picture over email of current image data

Task Notes:

- Made a report on overfitting
- 📄 **Overfitting** not available for me to review. Please make sure you provide access.

Code:

Doubled time, changed from 2 seconds to 4 seconds

Pooled the frequency less

```
#
        for freq in freqs:
            if count > 72:
                break
            count += 1
            if count % 2 ==0:
                ap = str(freq) + "hz"
                freqlist.append(ap)

        header = ["Time"] + freqlist
        out_csv.writerow(header)
        #create list to load onto a tensor

        tenslist = []
        for idx, time_value in enumerate(times):
            freqcount = 0
            avg_pixel_values = []
# switched time value from 2 to 4
            if time_value > 4:
                break
```

```
              #get a cut of at 100 hz which is a dimension of 72 / 4
(18)
            while freqcount != 72:
                totalpixel = 0
                for i in range(2):
                    totalpixel +=
rounded_specgram_values[freqcount][idx]
                    freqcount += 1
                avg_pixel_values.append(totalpixel / 2)
            tenslist.append(avg_pixel_values)
            writecsvdata =  [round(time_value, 7)] + avg_pixel_values
            out_csv.writerow(writecsvdata)
```

January 5

Dr. Pieper Meeting Notes:

- Report frequency bucket / partitioning (wide frequency ranges vs smaller ranges) does not make much difference
- Increasing amount of time also does not really make much difference
- System is insensitive to more time / cycles
- Changed parameters (doubled the amount 1,
- How the neural network was created
- How it trains / how does it adjust the weights / gradient method / steepest descent
- Test out the dataset
- Instead of 5 classes, just give 2 classes (only healthy or unhealthy)

Code:

```
    self.linear_layers = Sequential(
    #  fully connected layer, goes from 216 tensor size to 41 tensor
size
        Linear(216, 41),
        ReLU(inplace=True),
        # final classification, 41 neurons to 2 (different classes)
        Linear(41, 2)
```

)

January 24

Class Notes:

- January 31st (december / january logbooks)
- Feb 13th Procedure Section Deadline
    - What was done
    - How was done
    - Past tense
    - NOT WHY
    - Order does not matter
    - Good subsections / titles
    - Citations (datasets from other sources need to be cited)
    - Python citations / libraries
    - Only proper sources get cited in reference list
    - Pytorch citation is needed in the reference list
- March 1st Marks lock (term 2)

Dr. Garcia Meeting Notes:

- Think about places to get images
- Visualisations
- Future stances
- Can explain a tiny bit but does not need to be super detailed

Task Notes:

- Prepared and sent email to Dr. Pieper about meeting again and midterms finished

    I have finished midterms and am going to resume my work on my Applied Science Project. Let me know if you want to continue to have regular meetings and if you would still like to meet on Mondays at around 7:00 for these regular meetings.
- Started working on procedure section

- (wrote the datasets section of the procedure)
- Refer to 📄 Procedure section titled "datasets" (will add citations + specific amounts of data used afterwards) Here again your doc is not available!

The MIT-BIH Arrhythmia Database provided by PhysioNet was utilized. The dataset contained 58 records of patients in which 46 of these records were used. The dataset was sampled at a frequency of 360hz and each reading was independently annotated by two or more cardiologists with discrepancies then resolved afterwards. (https://physionet.org/content/mitdb/1.0.0/) This dataset contained (number of normal readings, etc) The second dataset utilized was the PTB-XL dataset from PhysioNet and was sampled at a frequency of 500hz. This dataset contains 9514 normal electrocardiogram readings, 5469 readings of Mycardio Infarction (MI) and 12782 unused samples of other abnormalities. (https://physionet.org/content/ptb-xl/1.0.3/)

January 26
Task notes:

- Continued to work on the procedure
- Refer to 📄 Procedure section titled "data preprocessing" (will add citations + specific amounts of data used afterwards)

From the MIT-BIH Arrythmia Database, 46 of the 58 patients' records were utlized as they all contained an MLII column. The amplitude data from these patients was then transformed into a frequency domain with the use of the matplotlib library. The parameters for the matplotlib library included a sampling frequency of 360, mirroring that of the dataset. Additionally the spectogram was given the no overlap parameter, and was graphed on a greyscale colourmap. Prior to feeding the spectogram values into the

CNN, the spectogram pixel values were averaged into frequency buckets of every 5hz and 2 seconds to reduce the spatial size of the input. The frequency buckets resulted in a final input size of 24 by 36 for the CNN.

January 29

Meeting with Dr. Pieper

- Learn Understand Adam optimizer / Stochastic Gradient Descent
- Make the model print out weights
- Describe back propagation
- View the weights converging (not changing)
- Try different numbers for the FC layer (bigger and smaller)

January 30

Class notes:

- Explain figure before the figure is introduced
- Fig 1. Title + more concise description at the bottom
- Explain table before the figure is introduced
- Table 1. Title + more concise description at the top

Task Notes:

- Worked on 📄 Procedure section titled "convolutional neural network" (will add citations + specific amounts of data used afterwards)

  The convolutional neural took in a tensor of size 1 x 24 x 36. This network included two convolutional layers as well as two fully connected layers. The convolutional layers both featured a 3 x 3 kernel. The first convolutional layer included a kernel that added 4 feature maps, but also reduced the original image size by 2, making the final image 4 x 22 x 34.

The second convolutional layer again decreased the image size by 2 but did not add any extra depth. This resulted in a final 4 x 20 x 32 image size.

February Notes

February 1

Class notes:

- Monday March 4
- 8:15 - 8:45 set up
- 9:00 am - Judging
- 11:30 am done
- Email teachers about missing class

Task notes:

- Took screenshot of CYSF portal for Ms. Grelowski
- Finished ☰ Procedure section titled "convolutional neural network"

  After each convolutional layer, the tensors were normalized with the BatchNorm2d function provided by Pytorch. After normalization, the tensors were placed inside the pooling layer. This pooling layer consists of a 2 x 2 kernel with a stride of 2. The CNN consisted of 2 fully connected layers which flattened the tensors into 216 individual inputs. These inputs were connected to 41 neurons in the first fully connected layer. The second fully connected layer starts with 41 neurons which are connected to 5 final classification neurons.

February 5

Class notes:

- Be specific, what part of the procedure

- Portals will be previewed
- Participate in 3 - 4 presentations
- Presentation
    - Background (time) depends on what you have
    - Research question
    - Goals
    - Hypothesis
    - Methods
    - Results
    - Conclusion and analysis
    - Answer to my question
    - Not clear / negative → Future directions
    - Poster funding
    - Significance

Task notes:

- Worked on introduction for oral presentation
- Refer to: 🟨 ASP Pres new slides 1 - 3
- Sent procedure to Dr. Garcia to check over

Meeting notes with Dr. Garcia

- Oral presentation rehearsal 23 of February 11:30 during lunch
- Can do both but believe one value is inflated
- If I don't get any more data, future direction is more data
- Don't be too vague

Meeting with Dr. Pieper

- Take each element , subtract, and square, and add up the 9 each
- Graph, shows that it has learned
- Z axis, accuracy
- Y axis, number of neurons

- X axis, learning rate
- Show convergence vs epoch (could also be number of neurons in hidden layer as)
- 2 week show sketch

February 7

Class notes:

- March 15 city science fair finish the CYSF portal

Task notes:

- Worked on methodology for oral presentation
- Refer to: 🟨 ASP Pres new slides 7 - 8

February 9

Task notes:

- Completed methodology for oral presentation
- Refer to: 🟨 ASP Pres new 9 - 11

February 13

Task notes:

- Completed Results → Future directions
- Refer to : 🟨 ASP Pres new 12 - 15

Meeting with Dr. Garcia:

- reinforce cs aspect

- format for poster

- Why neural networks in particular
- Objective problem   results   conclusion
- Background   future directions
- Methodology   references

February 14

Meeting with Dr. Pieper

- Report on accuracy

- Create a graph and see if the loss converge

- Make a table on accuracy relative to learning rate

- Make a conclusion which learning rate is the best

- Number of neurons of the layers

- Fuzziness of the ecg might be hindering the model performance

- Find which ones the model is having difficulty finding

February 19

Meeting with Dr. Pieper

- Have an explanation (the noise is messing it up)

- Future direction Filter the noise

- For graphs: look for the highest accuracy

- Future directions could be amplifying the signal

- For next meeting: Finish graphs + draft of poster

February 20

Task Notes:

- Worked on creating draft of the poster:
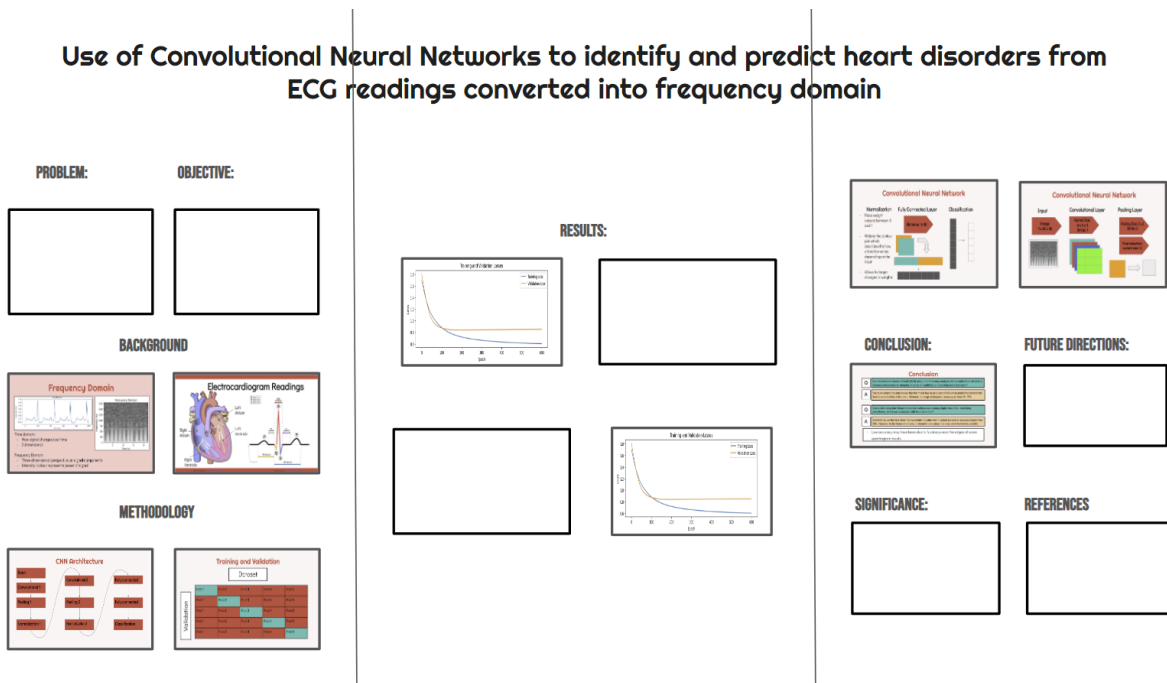
February 23

Meeting with Dr. Garcia (Presentation edits)

- Remove double space for separation
- Have units
- Focus on explanation
- What does data look like
- At some point what does those pixels mean in terms of if it higher or lower
- Put in conclusion training vs validation
- Be clear in the ecg
- Other projects not somebody else
- Scale was not very clear
- Explain where the dataset was from
- Validation be clear
- Explain lack of data

February 26

Class Notes:

- Took notes on other presentations

Brynn:
- Environmental stress leads to endangered species
- gene difference in endangered and non endangered species

Jessica
- Gene knockout on finding the affect a certain gene has on it
- Ideal method to construct crispr knockout

Cooper
- Synaptic pruning destroys extra adolescence neurons

February 27

Meeting with Dr. Pieper

- The project was not only used to achieve an accuracy for the model rather:
    - Identified the behavior of the network
    - Identified Sensitivity learning rate
    - Add short term objective that states: Identify the behavior of the CNN

February 28

Class Notes:

- ## Took notes on other presentations

Vincent
- MDR bacteria
- When one become mutant, only ones with mutation are left
- Antibiotics will no longer work
- Asked question: When identifying turbid or clear, how did you tell? With an apparatus or inspected by eyes?

Tiffany
- First responder suffer from mental health issue
- Main issue not getting or actively seeking them (public or internal sigma)
- Asked question: Why do you think ems had more words than police

Mariska

- Cytokines
- Mcp 1 pro inflammatory
- Call macrophages
- Cd68+ researched mark information of what happening in the tissue

March Notes

March 1

Class Notes:

- ## Took notes on other presentations

ZZ + Tosin

- Effect on Ketamine and cannabidiol on hippocampal brain cell cultures
- How marijuana affects function of brain
- Neonatal ?
- Promise of antidepressant
- Hippocampus - learning for the brain
- Blocks NMDA receptors what do they do exactly
- Exposed (is it just in the breast milk or is it in the air as well)?
- Asked Question: Why do you break down to neurons and expose the rat itself?

Joel + Ashank

- Silico Approach to mining mimotopes from the gut microbiome for type 1 diabetes
- Humans cannot regulate sugar anymore - pancreatic beta cells
- Asked Question: What is functional enrichment analysis in future directions?

March 5

- ## Class notes
    - March 15 complete CYSF portal
    - April 12th or 13 is CYSF
    - April 11th poster setup
    - Start writing results

- Results section

- Discussion

- Acknowledgements

- References

- March and april together

- Final paper (may 30%)

March 7

- ## Class notes

  - Paper that is full of figures and tables

  - Look at owl purdue link about how to cite figures and tables

  - Explain above the figure all there is to do with the figure

  - Explain what the figure is about

  - What we see from the figure

  - Description

  - Comparison of ____ and ___ groups

- ## Task notes

  - Made tasks for march → may

  - ### Reviewed science fair comments

    - Overall good project

    - More in depth about the methods

    - More in depth about on how fourier transform helps use filters (quick example)

March 11

- ## Class notes

  - Reflection - judges questions

  - (implement)

- ## Task notes

  - Worked on

  - Results are a description of results but not explanation

- ## Meeting with Dr. Pieper

- Visualized data

- Convergence vs number of neurons

- Maybe apply filters

- Sensitivity based on the design parameters

- Demonstrated convergence of the network

- Meet with cardiologist and see how they would like to use it

March 13, 2024

- Task Notes

  - Completed CYSF portal

    - Problem

    - Method

    - Analysis

    - Conclusion

    - Citations

Links to Written Documents

📄 ECG and Machine Learning Proposal

https://github.com/OwenKuang/Iris-Classification-ML

📄 Owen ASP Table

📄 Owen ASP Proposal

🟨 ASP Oral Presentation ECG

📄 Overfitting

🟨 ASP Pres new

Background Research

September Research

September 8, 2023

Source: https://pubmed.ncbi.nlm.nih.gov/32704420/

- Artificial Intelligence (AI) - automating intellectual tasks
- Machine learning (ML) and deep learning (DL) - Specific methods to achieve tasks
Data Science → Artificial Intelligence → Machine learning → Deep learning → Artificial Neural Networks (NN) (INSERT IMAGE)
- Four commonly used learning methods: Supervised, unsupervised, semi supervised, and reinforcement learning

Supervised Learning:
- Given a dataset
- Split into training, validation, and testing datasets
- Uses patterns in training dataset
- Infers an algorithm (algo) from feature-target pairs
- Know if prediction is correct or wrong
  Basic Steps
    - Acquire dataset
    - Split into training and validation
    - Use training and validation datasets to inform model

- Evaluate model with test dataset
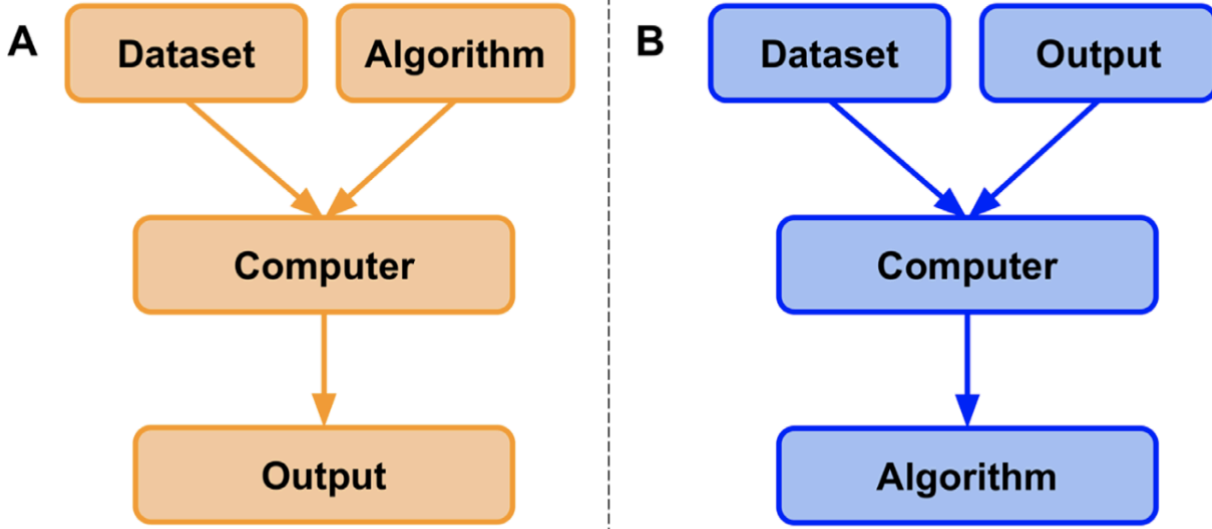
A) Classical Programming (computer is given a dataset and algo)

B) Machine learning (Given a dataset and associated output and computer develops the algo)

Common Supervised learning tasks

Regression

- Predicting numeric data

Classification

- Category of an example

September 12, 2023

Unsupervised learning

- Left to determine patterns by itself (not given a specific set of patterns)
- Clustering algorithm

Semi Supervised Learning

- Useful for datasets that contain labeled and unlabeled data
- Often used for medical images

- In theory, outperforms unsupervised models

Reinforcement Learning
- No single correct answer
- Overall goal or outcome
- Similar to human learning experience
- Trial and error rather than just data
- Applications in medicine are limited

Evaluating Performance
- Trainly dataset split into a smaller dataset and a validation dataset
- If validation on y% and x% are increasing and converging after every training iteration, model is learning
- If converge but do not increase, not learning
- Overfit when learning model has learned specific features to the training data and not a generalizable dataset
- Validation dataset can be used to tune the algorithm
- Supervised learning evaluated based on prediction accuracy
- Underfit when learning model has not learned or memorized dataset
- Ideally fit when model performs strongly on both datasets
- Average mean squared error (MSE) indicates performance on a regression model (how close a predicted value is to the intended value)

Linear Regression
- Specify a relationship between one or more numeric features and a single numeric target
- Analysis technique
- Straight line to explain dataset
- Residual: vertical distance between predicted y and actual y

Logistic Regression
- Probability of a particular outcome
- Sigmoid function (s shaped curve) and converts discrete or continuous (x) in a y value between 0 and 1

Decision Trees
- Supervised learning
- Root node
- Single defining feature to split the data into a specific class
- Every decision node has another split

September 14, 2023

Neural Networks and Deep Learning
- Inspired by biological neural networks
- Connections between nodes are weighted
- Weighting depends on outcome

Feedforward Neural Networks
- Attempts to find a line, plane, or hyperplane that separates the classes
- Transformed using sigmoid function
- Perceptron: Machine learning algo
- Feedforward: information from each previous node used in the next layer

Challenges
- Black box problem
- Not sure what features were identified
- Simple ml algos are easier to read but harder to interpret

Annotated PDF: 📄 IntrotoML.pdf

September 16, 2023

Code: https://github.com/OwenKuang/Iris-Classification-ML

Source (tensors): ▶ What the HECK is a Tensor?!?

- Tensors: Mathematical object that has n indices and m^n components that obeys certain transformation rules
- Tensors have different ranks
- Rank X = needs x pieces of information

September 17, 2023

Source: ▶ The Next Generation of Neural Networks

First generation of neural networks:
- Perceptrons
- Hand coded features

Second generation neural networks:
- Back propagations
- Find the derivative
- Change the weights after feeding it backwards
- Takes too long to learn lots of layers

Support Vector machine (clever perceptron)
- Not handed coded
- Feature computes the similarity between test and training example

- Selects the best subset and decides on weight

Restricted Boltzmann Machine

- One layer of hidden units
- No connections

Alternating Gibbs sampling

- Compare the fantasies of the computers with that of the given pixels
- Adjust the weights until the fantasies of the computer is similar/the same to that of the pixels
- Model has stopped learning when it is the same or similar
- Slow

Greedy Learning

- Find something simpler than the original distribution
- Use the simpler distribution to solve the problem
- Good for scaling

September 20, 2023

Source: https://www.nature.com/articles/s41598-021-97118-5

ECG: Electrical activities of the heart

- Shape and duration of wavelength can be used to determine heart diseases
- Two - event related moving averages (TERMA)
- Fractional fourier transform (frFT)

September 22, 2023

Source: [Diagnosis and Management of Myocarditis in Children: A Scientific Statement From the American Heart Association](#)

Myocarditis: inflamed heart tissue, affects electrical system and muscle cells, leads to irregular heartbeat and heart function

Clinical: Possible micardis

CMR: confirmed micardis

Biopsy: Proven myocarditis

Source: [How useful is the smartwatch ECG? - ScienceDirect](#)

Utilize Photoplethysmography

Relatively reliable

Source: [Prediction of heart disease and classifiers' sensitivity analysis](#)

Numerous ML algorithms have been tested for heart disease (Naive Bayes , SGD, SVM, KNN, and more)

- 82% detection rate
- UCI Heart Disease Data Set

ANN

- 95% prediction reate

Source: [ECG / EKG Components: AmperorDirect](#)

5 Different Waves given by an ECG (P, Q, R, S, T)

Different phases of heart activities

P - Upper Heart Chambers

QRS complex - Lower heart chambers

T - Recovery of ventricles

September 26, 2023

Source: [ECG Diagnosis and Classification of Acute Coronary Syndromes - PMC](#)

ECG can have important information about acute coronary syndrome

Can be typical or subtle

Essential Initial evaluation for myocardial ischemia

Source: [A vector-free ECG interpretation with P, QRS & T waves as unbalanced transitions between stable configurations of the heart electric field during P-R, S-T & T-P segments](#)

Depolarize produces a weaker negative electric field

Repolarize produces positive stable electric field

September 28, 2023

Drafting Research Proposal: ▤ Owen ASP Proposal

Source: [Machine Learning Based Heart Disease Diagnoses](#)

- Cardiovascular Diseases (CVD) one of leading cause of death
- Coronary Heart Disease

- Cerebrovascular disease
- Peripheral arterial disease
- Rheumatic heart disease
- Congenital heart disease
- ECG used to diagnose but visual identification requires time
- Drawbacks of ML (Hidden layers, therefore no explanation for final prediction)

Source: [Deep Learning Electrocardiographic Analysis for Detection of Left-Sided Valvular Heart Disease](#)

- Valvular heart disease
- ECG for aortic stenosis

October Research

October 1, 2023

Source: [ECG diagnosis and classification of acute coronary syndromes](#)

Acute coronary syndromes (ACS)

ST Segment Elevation (STE)

ST Segment Depression (STD)

Acute myocardial ischemia - effects all components of heart waves

Transmural myocardial ischemia:

ST segment changes

- ST segment elevation facing ischemic zone
- ST segment depression facing anatomically opposite

T and U wave Changes

- T wave amplitude increase followed rapidly by STE
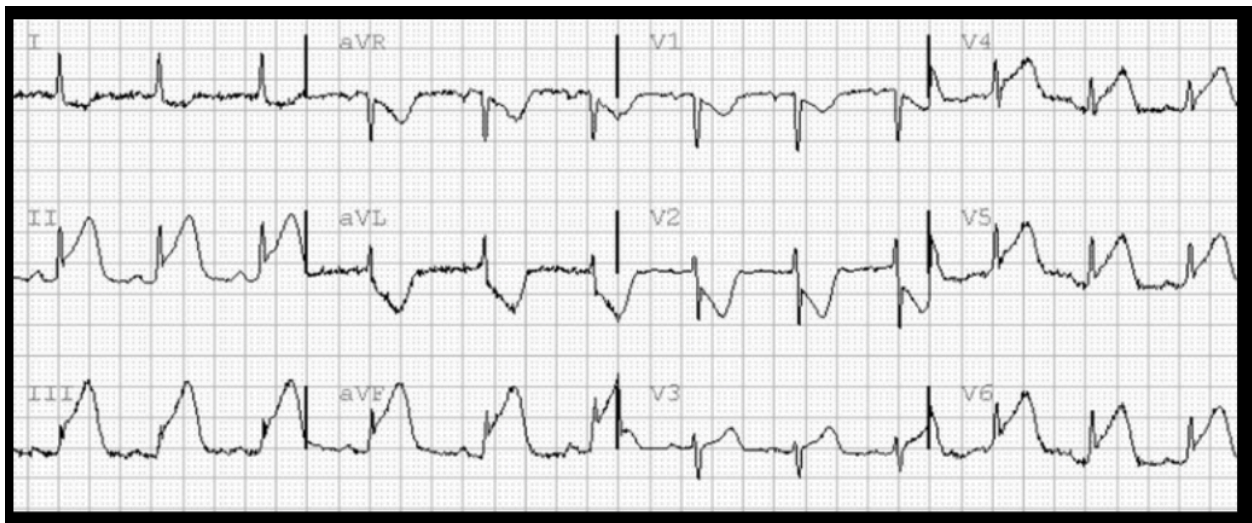
Chronic Phase of Myocardial infarction

- Q waves sign of irreversible necrosis

Non-ST elevation Acute Coronary Syndrome

- T wave inversion with isoelectric ST segments or minor ST deviation
- Up sloping STD w/ tall T waves (common in tachycardia which is when bpm >= 100)
- STD seen across multiple leads (diffuse) in inferior and anterolateral leads with ST elevation in lead AVR
- Wellens' sign = inverted T waves in precordial leads

Source: Advanced ECG in 2016: is there more than just a tracing?

- ECG most important technology for diagnosis of patients with cardiac arrhythmias
- Atrial Fibrillation - Associated with stroke

Source: Arrhythmias - StatPearls - NCBI Bookshelf.

Less than 60 bpm: bradyarrhythmia

Over 100 bpm: tachyarrhythmia

Supraventricular Tachyarrhythmia (SVT)

- Atrial fibrillation

- Atrial flutter

- Atrial tachycardia

Ventricular Tachycardia (VT)

- Ventricular fibrillation

- Ventricular tachycardia

Narrow QRS complex tachycardia (QRS < 120 ms duration)

- Atrial tachycardia

- Atrial flutter

    - Regular tachycardia

    - Saw tooth appearance of P wave

- Atrial Fibrillation

    - Irregularly narrow complex

    - No discernable P-Waves

Wide QRS complex tachycardia (QRS >= 120 ms in duration)

- Ventricular tachycardia

    Non-sustained ventricular tachycardia

    - Monomorphic wide complex

    - More than three beats in a row with duration less than three seconds

    Sustained ventricular tachycardia

    - Monomorphic wide complex tachycardia

- Ventricular Fibrillation

    - Polymorphic fibrillatory waves

October 6, 2023

Source: [IEEE Overview - Purdue OWL](#)

General format:

- Title centered at the top of the page with 24 point font
- Byline centered below title with 10 point font
- After line break, include Author's name, author's affiliation, city & country location, email address (one line for each)
- Body written in 10 point font
- Columns in the last page = same length
- Begin with abstract
- First 8-10 characters in all caps
- Bracketed numbers for in text citation
- Refer to as "in [1]" not "in reference [1]"
- Refer to multiple citations, [1], [2]
- Refer to a list of 3 or more consecutive citations [first]-[last]
- First cited reference is always the first, etc
- References
    - Separate page titled "References"
    - Listed in number of citation, not alphabetically
    - Authors: surname and initials, suffixes included
    - <7 authors, list in order of byline
    - >7 authors, et al.
    - Months shortened to 3-4 letters
    - [#]   Author, "Title," *Journal*, volume, number, page range, month, year, DOI.
    - [7]   K. M. Bivens and K. Cole, "The grotesque protest in social media as embodied, political rhetoric," *Journal of Communication*

Inquiry, vol. 42, no. 1, pp. 5-25, Jan. 2018, doi: 10.1177/0196859917735650.

-

Source: [Heart Disease in Canada](#)

- Second leading cause of death
- 1/12 Canadian adults over 20 have diagnosed heart disease
- Every hour, 14 canadian with diagnosed heart disease die

October 8, 2023

Source: [The hidden waves in the ECG uncovered revealing a sound automated interpretation method - PMC](#).

October 13, 2023

Source: https://ieeexplore.ieee.org/document/8913462

- Method to detect multiple arrhythmias using time frequency analysis and convolutional neural network
- Divided into 12 classes
- Three kinds of time frequency was analyzed
    - Short time fourier transform
    - Wavelet transform
    - Pseudo wigner ville distribution
- STFT best accuracy AUC 0.9987
- Early studies: extract ECG features before arrhythmia detection (find P wave, T wave, QRS complex)

October 17, 2023

Source: https://ieeexplore.ieee.org/document/8913462
- Signal fit to 10 seconds
- Resampled to 125 hz
- Second pass with 0.5 hz to remove baseline
- Time frequency transform used to obtain time frequency distribution matrix was normalized and used input of design CNN
- Five fold cross validation method
- Good for public data however, in real life applications, there are lots of noise contamination making it useless
- CNN (Convolutional Neural Network)
    - Convolutional Layer
    - Pooling Layer
    - Fully connected layer
    - Dropout layer
- Using wavelet transform allowed CNN to find arrhythmias without specifically finding the P and S waves

Source: https://pubmed.ncbi.nlm.nih.gov/31450977/

Source: Convolutional Neural Networks, Explained | by Mayank Mishra | Towards Data Science

November Research

November 15, 2023

Source: https://litfl.com/atrial-flutter-ecg-library/

Atrial fibrillation Features

Saw tooth pattern

Loss of isoelectric baseline

Normal at 300 bpm (1:1)

2:1 block = 150 bpm

3:1 block = 100 bpm

4:1 block = 75 bpm

November 21, 2023

Source: https://jscholaronline.org/articles/JBER/Signal-Processing.pdf

Baseline wander: 0.5 - 0.6Hz cut off frequency

Powerline interference: 50 - 60Hz cut off frequency

EMG noise: 100Hz+

November 27, 2023

Source: https://www.youtube.com/watch?v=Nz4WB8-gNBg

- Time domain (all the components of the signals are together, cannot get the idea of the frequency information of a signal)
- Frequency domain (can get all the components and see the various different signals)
- Frequency domain gives a different view than time domain
- Fourier transforms are used to convert from time domain to frequency domain

Source: https://www.youtube.com/watch?v=FE6Ep82J9g8

- Spectograms is a 2D representation of a 3D waveform
- Y axis = frequency
- X axis = Time

- Z axis = intensity (darkness / lightness)

Source: https://www.youtube.com/watch?v=KTB_OFoAQcc

Source: https://www.youtube.com/watch?v=s2K1JfNR7Sc

November 28

Source: **https://www.youtube.com/watch?v=i76B1pBoN4Y**

- High pass and low pass filters
- Allows for certain waves to pass / smooth out the signal
- Lessons the impact of either higher frequency or lower frequency waves (depending on the filter)

November 29

Source: https://www.youtube.com/watch?v=YhatZb5SzE0

- For cnn use case generally one neuron per pixel value for the input layer
- For the output layer in the CNN, one neuron per classification type
- Better to deeper network than wider network (less size for input layer but rather more fully connected layers)

Source: https://www.youtube.com/watch?v=YSNLMNnlNw8

- No perfect amount of kernels
- Based on how detailed of features you are looking for
- More features can give more to learn

December Research

December 1

Source: Building a Convolutional Neural Network in PyTorch - MachineLearningMastery.com

- Neurons inside of the filters
- Need to convert pixel values to a pytorch tensor (read additional sources: https://pytorch.org/tutorials/beginner/introyt/tensors_deeper_tutorial.html How to load a pandas dataframe in tensorflow)

```
new_tf_dataset = tf.data.Dataset.from_tensor_slices((dataframe.values,
var_target.values))
```

- Format for tensors (channel, height, width)
- (1, 36, 36)

December 5

Source: PyTorch: Training your first Convolutional Neural Network (CNN) - PyImageSearch (Refer to code)

Source:

https://towardsdatascience.com/understanding-and-calculating-the-number-of-parameters-in-convolution-neural-networks-cnns-fc88790d530d

- Parameters are the number of learnable features / filter parameters
- Input layer, nothing to learn, therefore param = 0
- Conv layer ((width * height * filters in previous layer) + 1) filters in current layer)
- Pool, 0 to learn
- https://www.geeksforgeeks.org/introduction-convolution-neural-network/

December 7

Source:

https://pytorch.org/docs/stable/generated/torch.nn.Conv2d.html#torch.nn.Conv2d

- Nn.con2d → 2D convolution over input signal

```
Parameters: (in_channels, out_channels, kernel_size, stride=1, padding=0,
dilation=1, groups=1, bias=True, padding_mode='zeros', device=None,
dtype=None)
```

- torch.tensor → converts data into tensors

Parameters: Like normal lists, can directly convert list to tensor

- torch.nn.MaxPool2d(*kernel_size*, *stride=None*, *padding=0*, *dilation=1*, *return_indices=False*, *ceil_mode=False*)
- torch.nn.Flatten(*start_dim=1*, *end_dim=-1*)
- torch.ReLu

Experimental Procedure

Data Preprocessing:

- Obtained the MIT arrhythmia database
- Applied a fourier transform using the built in Pytorch functions
- Took the pixel values of the transformed graph and used as input for CNN

Created CNN architecture:

- 1 Input layer
- 2 Convolutional Layers
- 2 Pooling Layers
- 2 Normalization Layers
- 2 Fully connected Layers
- 1 Classification Layer

5 Fold Cross Validation:

- Partitioned the Data into 5 different folds (80 / 20 split)
- Utilized 1 fold for validation and the other 4 folds for training
- Repeated process 5 times (each fold becomes used as the validation)
- Accuracy + Validation accuracy = Average of the 5 times

Training

- Ran the model for 600 epochs and obtained final accuracy + validation score

Data Collection
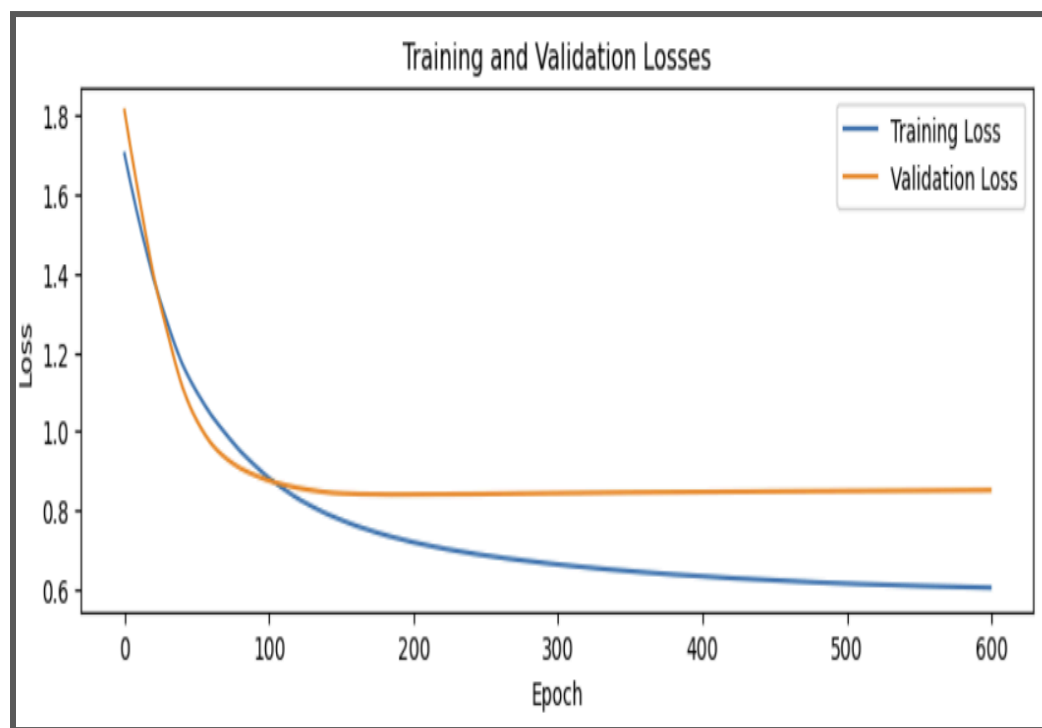
Training Accuracy: 95%

Validation Accuracy: 83%

Learning Rate	Training Accuracy	Validation Accuracy
1 * 10^-3	82%	76%
1 * 10^-4	95%	83%
1 * 10^-5	78%	80%

Discussion and Conclusion

The convolutional neural network was able to achieve an accuracy of 83 percent in the validation and 95% in the training. The most optimal learning rate is 1 * 10^-4. In the future, I hope to add filters to remove noise and allow the model to identify the patterns more easily.