

# EE / CprE / SE 491 – sdmay24-47

## Accurate Cancer Prediction Using AI

### Bi-weekly Report 3

Feb 11th – Feb 24th

Client: Ashraf Gaffar

Faculty Advisor: Ashraf Gaffar

#### Team Members:

Jack Sebahar — *Project Manager / Client Liason*

Nicholas Otto — *Model Design Lead*

Mason Wichman — *UI Design Lead*

Lal Siama - QA Engineer, Tester

Helen Lau - UI Design

Isaiah Mundy - UI Design Test

#### Weekly Summary

Following the agile style of our project, we are continuing to make iterative changes to our code to improve model accuracy. This includes dealing with data normalization and reducing overall model error. When expanding our model input data from being very small to encompassing almost every piece of data we have, we actually noticed a decrease in model accuracy compared to our initial model. This indicates to us that the small sections of data we used in the early versions of our model actually had a better indication of outcome than when including all the data in the new version of our model. This makes sense to us, since we were told that the peaks in the data points provide a better estimate of the outcome than the entire data as a whole. As a team, we have been discovering new and effective ways to extract only the key data from the dataset, such as the peaks, so that we can discover which pieces of data provide the strongest correlation to the outcome. Simultaneously, we need a good way to test our model in a real-world environment beyond the basic testing metrics provided by Keras. Although these are a good indication of your model's overall performance, they don't show you real results with your data. We have been implementing different ways to do this via built-in Tensorflow and Keras methods. Research, improvement, and testing are the name of the game right now, and each week we make incremental changes in our model with the ideal goal of improving its accuracy and performance each time.

#### Past Week Accomplishments

- Model Error Reduction
  - Wrote functions to normalize our data to a common scale to attempt to reduce model error. We found no significant difference in the model error, but in the creation of these functions, we discovered an already existing error in the way the model was being presented with training data. In rectifying this, we found that the model produced better results when it was given a smaller representation of each data sample to the point where a model trained on one data point per sample produced a significantly lower error than a model trained on every data point for each sample.

- Furthered investigating AWS approach
  - An important portion of our project is to generate models on different platforms. The purpose of this is to determine if there are differences between the accuracies of models on different platforms. Our current development environment consists of GCP and Colab. We began initializing our AWS environment. This has not been as trivial as GCP since nobody on our team has prior experience with it.
  
- Model performance evaluation
  - Keras provides some built-in metrics to gauge a model's performance such as the "evaluate()" function which provides the mean squared error of the data. Using this knowledge, we can get a relative idea of our model's performance, however, this doesn't help us really see how the model is actually performing in a real-world scenario. As such, we want the ability to run some unknown data through the model and see what it returns to gauge the practical accuracy of the model. There are plenty of ways to do this such as using the "load\_model()" function. In the past two weeks, we have been testing some valuable ways to do this.

### Pending Issues

- Determine collaborative environment for AWS
  - GCP uses Colab for collaborative development, we need to set up our AWS environment in a similar fashion.
  - AWS uses Sagemaker notebook snapshots, but this means changes made won't affect other notebooks
  - We need to figure out how to create a similar development environment as in GCP
- Determine billing options for AWS
  - Some features of AWS may require a billing account. We have been approved to use funds for our project, but this will have to be consulted with our advisor

### Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Jack Sebahar	<ul style="list-style-type: none"> <li>● Model performance evaluation               <ul style="list-style-type: none"> <li>○ Finding methods to sample our model with unknown data</li> </ul> </li> </ul>	5	28
Nicholas Otto	<ul style="list-style-type: none"> <li>● Data normalization</li> <li>● Model error reduction</li> </ul>	6	35
Mason Wichman	<ul style="list-style-type: none"> <li>● Data preprocessing               <ul style="list-style-type: none"> <li>○ Data peak extraction script</li> </ul> </li> <li>● AWS initialization</li> </ul>	5	30
Lal Siama	<ul style="list-style-type: none"> <li>● Data preprocessing</li> </ul>	3	24

	<ul style="list-style-type: none"> <li>○ Methods research</li> <li>● UI design improvements</li> </ul>		
Helen Lau	<ul style="list-style-type: none"> <li>● AWS initialization <ul style="list-style-type: none"> <li>○ Set up development environment similar to GCP</li> </ul> </li> </ul>	3	24
Isaiah Mundy	<ul style="list-style-type: none"> <li>● Data Preprocessing</li> <li>● Model Design <ul style="list-style-type: none"> <li>○ Normalizing and extracting peaks from data</li> </ul> </li> </ul>	3	24

### Plans for Coming Week

- Investigate more effective options for data preprocessing - Mason
  - Determine if there are more suitable options for preprocessing our data
  - Determine if any Keras models are designed to work with data like ours
- Reduce Model Error - Nick, Isaiah
  - Utilize functions to identify the “peaks” of each sample spectrum to create a more condensed representation of each sample to train the model on.
- Investigate more model testing methods - Jack
  - Built-in Keras functions?
  - Run via UI?
  - Is there a better way to automate this?
- Fully initialize the AWS environment - Helen
  - Current development is in GCP
  - Part of the goal is to compare results on different platforms
  - Using Sagemaker, similar to Colab
  - Determine if there are any billing requirements
- Improve UI - Mason, Siama
  - We started a basic UI last semester but haven’t done much with it this semester as most of our focus has been on model accuracy. However, we need to continue improving the UI so it can interact with our specific GCP or AWS model

### Broader Context

(1)

Area	Description	Examples
Public health, safety, and welfare	For many software projects, there is an expected level of quality since many of them deal with safety, health, and welfare, which can directly impact people	Our product directly impacts public health by providing a tool that can aid in the diagnosis of cancer.
Global,	It is important to consider factors such as	Our project aims to benefit

cultural, and social	accessibility, language barriers, and cultural sensitivities when designing and employing a piece of software	communities worldwide by offering a technology-driven solution for cancer diagnosis but does not necessarily reflect the values, practices or aims of any specific cultural group.
Environmental	Some software can have a direct impact on environmental processes	We don't expect our project to have much of a direct or indirect environmental impact.
Economic	Software projects can be expensive and have issues that can cost money. It is important to make a design as cost-efficient as possible. Some software can also be used to help decrease spending on other things.	As our project could potentially lead to improved patient outcomes in the future, this could result in reduced healthcare costs.
Ethical	There are certain ethical considerations to be made with our project since we are dealing with confidential data.	Since we are dealing with real patient medical data, we are legally bound to ensure this data remains confidential to only ourselves and medical professionals.
User Experience	There must be a quality user experience when the user is interacting with our product.	Since the users will be allowed to upload medical data via a front-end application, this process must be seamless and have no flaws or security leaks.
Continuous Improvement	Software must incorporate continuous improvements as new information surfaces and bugs are discovered in the design.	If new cancer diagnosis techniques get discovered or new medical data become available, we need to incorporate them to ensure our model maintains the highest accuracy possible.

(2)

- One way we can argue for the positive ethical effects is to show our process of handling sensitive medical data. As a team, we are only sharing the data via a physical connection, and have not uploaded the data to any cloud-based software. By demonstrating our process of using the data, we can show that we are complying with our ethical design considerations
- We are demonstrating evidence of positive effects in global, cultural, and social design considerations since our project is as objective as possible. Our product can be used cross-culturally, with essentially no language barriers. Furthermore, the idea of practicing good health exists in pretty much every culture and social setting, so our application can be used by virtually anyone

- We are demonstrating success with continuous improvement since every week we are discovering new ways to improve the performance of our model. We are constantly making improvements, especially when new information surfaces. This demonstrates our willingness to ensure our model is always performing as well as possible.
- We can demonstrate positive effects on public health, safety, and welfare by showing the capabilities and success of our model and how it can benefit people's medical experience.

(3)

- One potential negative effect we have in our design right now is with user experience. We have a base UI designed but haven't really improved on this design since last semester. This is mostly due to the fact that most of our time has been spent improving the model's accuracy, and it is hard to envision how our user interface should function without a finished product. We can address this by designating more time working with the user experience, especially now that our model is starting to come to fruition.