Accurate Cancer Prediction Using AI

Design Document

Team 47

Potential Collaboration with Digistain

Advisor: Ashraf Gaffar

Jack Sebahar, Isaiah Mundy, Helen Lau, Mason Wichman, Lal Siama, Nicholas Otto

sdmay24-47@iastate.edu

https://sdmay24-47.sd.ece.iastate.edu

Revised: 9/7/23 Version 1
9/24/23 Version 2
10/8/23 Version 3
10/22/23 Version 4
11/12/23 Version 5

Executive Summary

Development Standards & Practices Used

HIPAA - This standard ensures the protection of patients' medical data. Since the project deals with medical information, we will need to ensure patient privacy and security [1].

IEEE 1058: It provides guidelines for the preparation of software project management plans [2].

IEEE 3051 - This standard revolves around the trustworthiness of Artificial Intelligence and Autonomous Systems, and provides guidelines for ensuring the trustworthiness and reliability of AI systems. This standard is particularly relevant to projects in critical domains like healthcare [2].

IEEE 3123 - This standard provides clear definitions for terminology utilized in artificial intelligence and machine learning [2].

IEEE 12207 - This standard defines the processes involved in the software development life cycle. It provides a framework for managing software projects and ensuring the quality and reliability of software systems [2].

ISO 13485:2016 - This standard discusses quality management systems regarding medical devices. Since our AI is intended for medical purposes, there is a high level of quality that is expected [3].

Summary of Requirements

- **Functional**
    1. AI Model Functionality: Develop a basic AI model for cancer prediction with at least 70% accuracy.
    2. Training Process: Train the AI model using medical data.
    3. Iterative Improvement: Continuously improve the AI model with additional data.
    4. Student Training: AI training with the use of Java/Python and Tensorflow, Keras.
    5. Data Integration: Establish a workflow for acquiring and integrating medical data into the AI model.
    6. Accuracy Evaluation: Evaluate AI model's prediction accuracy.
    7. User Interface: Design a user-friendly interface for data input.
- **Resource**
    1. Access to a computational infrastructure capable of training and running the AI model

    2. UI

    1. Design a user-friendly and intuitive interface for interacting with the AI model
- **Performance**
    1. Achieve a minimum level of accuracy and reliability in cancer prediction.

    3. Legal

    1. Comply with data privacy and security regulations when handling medical data

2. Ensure only patients and medical experts ever see data

- **Maintainability**
  1. Develop a system for updating and maintaining the AI model as new data becomes available
  2. continuously update model to maintain high accuracy
- **Testing**
  1. Implement rigorous testing procedures to validate the accuracy and reliability of the AI model
  2. Implement tests to ensure confidentiality of patient data

Applicable Courses from Iowa State University Curriculum

- SE/CprE 185 - Problem solving in SE/CprE
- Com S 227, 228 - Object Oriented Programming
- SE 329 - Software Project Management
- Com S 309 - Software Development Practices

New Skills/Knowledge acquired that was not taught in courses

- Machine learning model development
  - During this project, we have done extensive learning in machine learning platforms, strategies, and neural networks
- Cloud computing platforms
  - This project requires us to learn about various cloud computing platforms such as GCP and AWS. It is important to know how these platforms function, what services they provide, and how these services can be useful for our project
- Cancer Pathology
  - During this project, we had to familiarize ourselves with pathology data and cell culture images. This new knowledge is important for us in developing our model

Table of Contents

List of figures/tables/symbols/definitions (This should be the similar to the project plan)

Figures

Tables

Definitions

- Keras - an open-source library that provides a Python interface for artificial neural networks [4].
- Neural Network - a computer system modeled on the human brain and nervous system
- Pathology - branch of medical science that is focused on the study and diagnosis of disease
- Tensorflow - a free and open-source software library for machine learning and artificial intelligence [5].
- Tkinter - a Python binding to the Tk GUI toolkit [6].

1 Team

1.1 Team Members

- Jack Sebahar
- Isaiah Mundy
- Helen Lau
- Lal Siama
- Nicholas Otto
- Mason Wichman

1.2 Required Skill Sets for Your Project

- Software Development
- Software Project Management
- Software Testing
- Version Control
- Code Review

1.3 Skill Sets covered by the Team

Jack - Code Review, Version Control, Software Project Management, Software Development

Mason - Software Development, Software Testing, Version Control, Code Review

Helen - Code Review, Version Control, Software Development

Siama - Code Review, Version Control, Software Testing

Isaiah - Software Development, Software Project Management, Software Testing, Version Control

Nick - Code Review, Version Control, Software Development, Software Testing

1.4 Project Management Style Adopted by the team

Our team will follow a typical waterfall project management style with a project manager, development team, testing team, and documentation team. Some members of the team will need to be on multiple teams, and we may need to adjust the roles throughout the semester if necessary.

1.5 Initial Project Management Roles

- Jack - Project Manager
- Siama - QA Engineer, Tester
- Nick - Client Liaison
- Isaiah - Team Coordinator
- Mason - GUI Developer
- Helen - Software Developer

1.6 Problem Statement

The current challenge in cancer diagnosis and prediction lies in the limited ability to accurately foresee its occurrence and recurrence. Traditional diagnostic methods may not be as precise as desired, and human doctors, despite their expertise, face challenges in providing consistently accurate predictions. Recent research suggests that artificial intelligence (AI) has the potential to significantly improve predictive accuracy, surpassing the capabilities of human doctors. The goal of this project is to develop a simple AI model, leveraging tools such as Tensorflow and Keras, to enhance cancer prediction, providing a valuable asset in the realm of medical diagnosis.

1.7 Requirements and constraints

- Functional
    - AI Model Functionality: Develop a basic AI model for cancer prediction with at least 70% accuracy.
    - Training Process: Train the AI model using medical data.
    - Iterative Improvement: Continuously improve the AI model with additional data.
    - Student Training: AI training with the use of Java/Python and Tensorflow, Keras.
    - Data Integration: Establish a workflow for acquiring and integrating medical data into the AI model.
    - Accuracy Evaluation: Evaluate AI model's prediction accuracy.
    - User Interface: Design a user-friendly interface for data input.
- Resource
    - Access to a computational infrastructure capable of training and running the AI model
- UI
    - Design a user-friendly and intuitive interface for interacting with the AI model
- Performance
    - Achieve a minimum level of accuracy and reliability in cancer prediction.
- Legal
    - Comply with data privacy and security regulations when handling medical data
    - Ensure only patients and medical experts ever see data
- Maintainability
    - Develop a system for updating and maintaining the AI model as new data becomes available
    - continuously update model to maintain high accuracy
- Testing
    - Implement rigorous testing procedures to validate the accuracy and reliability of the AI model
    - Implement tests to ensure confidentiality of patient data

1.8 Engineering Standards

HIPAA - This standard ensures the protection of patients' medical data [1]. Since the project deals with medical information, we will need to ensure patient privacy and security

IEEE 1058: It provides guidelines for the preparation of software project management plans [2].

IEEE 3051 - This standard revolves around the trustworthiness of Artificial Intelligence and Autonomous Systems, and provides guidelines for ensuring the trustworthiness and reliability of AI systems [2]. This standard is particularly relevant to projects in critical domains like healthcare.

IEEE 3123 - This standard provides clear definitions for terminology utilized in artificial intelligence and machine learning [2].

IEEE 12207 - This standard defines the processes involved in the software development life cycle. It provides a framework for managing software projects and ensuring the quality and reliability of software systems [2].

ISO 13485:2016 - This standard discusses quality management systems regarding medical devices. Since our AI is intended for medical purposes, there is a high level of quality that is expected (ISO - International Organization) [3].

1.9 Intended Users and Uses

Intended Users:

- Medical professionals such as oncologists and healthcare providers seeking enhanced predictive tools for cancer outcomes.
- Users with varying levels of AI expertise, from novices to experienced practitioners.

Intended Uses:

- Assisting medical professionals in making more accurate and timely cancer-related predictions.
- Serving as a supplementary tool for informed decision-making in cancer treatment.

2 Project Plan

2.1 TASK DECOMPOSITION

1. Understand the domain and the problem
    a. Research and gather information on cancer and pathology
    b. Analyze the requirements for diagnosing conditions based on patient information and images
2. Investigate AI as a solution
    a. Explore the limitations of traditional computer science methods in solving the problem
    b. Investigate the advantages of AI, especially in medical image processing
3. Model selection and training
    a. Set up the necessary infrastructure and tools for AI development
    b. Explore models in Keras.io for diagnosing cancer conditions
    c. Train and familiarize with one or two simple AI models

   d. Retrain the selected model with collected data and, if necessary, consider switching to another model
4. User interface and prognosis
   a. Develop a user interface to upload patient information and images
   b. Implement the functionality to provide a prognosis based on user inputs and the AI model
5. Performance and comparison
   a. Collect metrics related to the AI environment and document model parameters
   b. Record details about the size and quality of the dataset
   c. Replicate the model training process on both AWS and Google Cloud and provide a comparison report

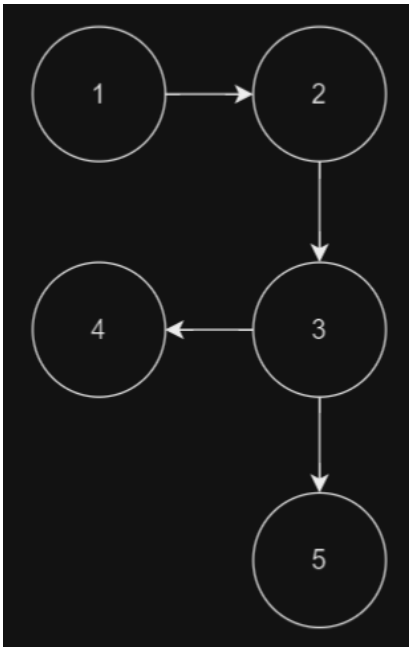

Figure 1. Process Flow

2.2 PROJECT MANAGEMENT/TRACKING PROCEDURES

We will be utilizing the waterfall+agile project management style. Waterfall makes sense for our project since our group is smaller and our project has clear requirements. Each phase of our project will most likely need to be completed before the other. Some aspects of agile might be utilized, such as when making adjustments to project performance, in which case we would need to revisit certain phases and meet with the client to get a clear deliverable. We'll be utilizing Github and Discord to keep track of progress.

2.3 PROJECT PROPOSED MILESTONES, METRICS, AND EVALUATION CRITERIA

1. Understand the domain medical problem
   a. Conduct at least 3 hours of research on technical documents regarding the domain
   b. Be able to confidently identify cancerous conditions

2. Investigate AI as a solution
    a. Determine the best AI environment for development
    b. Be able to confidently navigate the AI
3. Model selection and training
    a. Comprise data set of at least 100 data points per category
    b. Successfully train an AI model
4. User Interface
    a. Develop a working user interface
    b. Allows patients to upload info and provides prognosis
5. Performance
    a. Ensure same model can be trained on other cloud platforms
    b. Refine model until accuracy of at least 70% on a platform

2.4 PROJECT TIMELINE/SCHEDULE

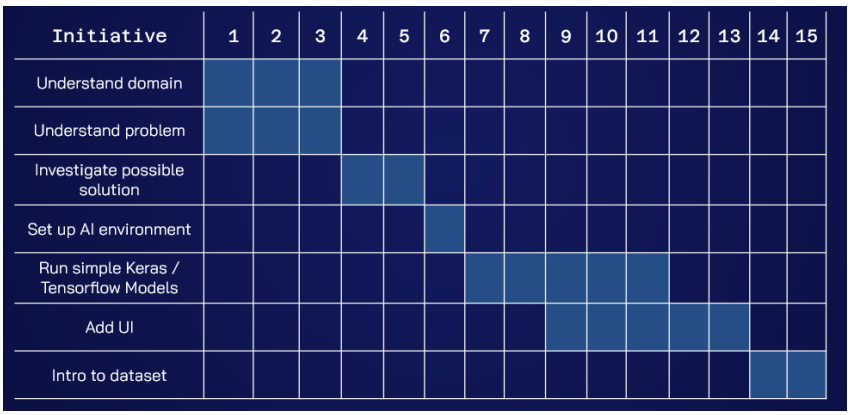

Figure 2. Task Timeline (491)

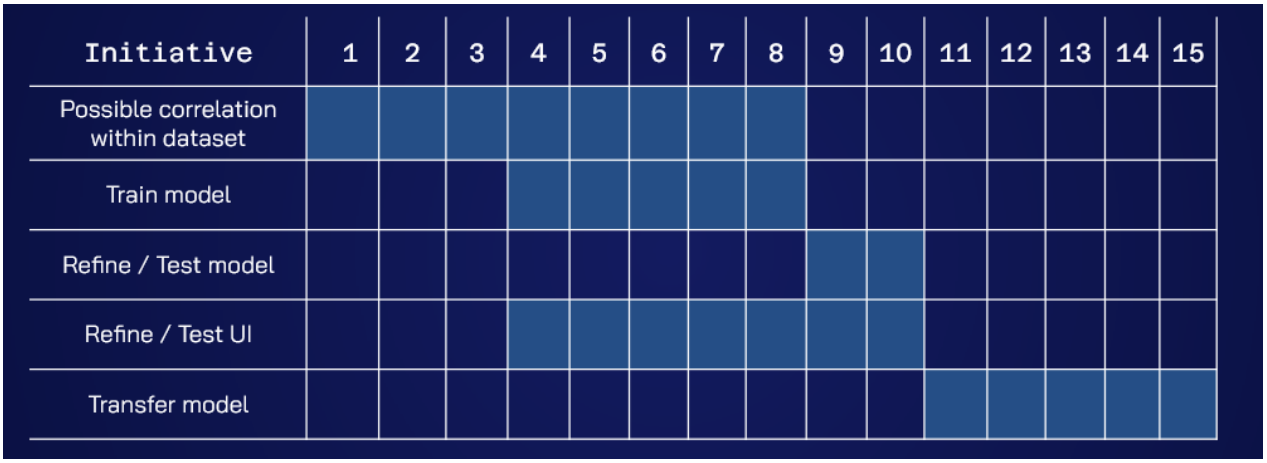

Figure 3. Task Timeline (492)

2.5 RISKS AND RISK MANAGEMENT/MITIGATION

Agile project can associate risks and risk mitigation with each sprint.

- Difficulty in understanding AI and cloud concepts - 0.1
- Difficulty in setting up on-premise environment - 0.2
- Model may not achieve desired accuracy - 0.6
    - Mitigation: Experiment with other pre-built models, try changing parameters values within the current model, attempt to extract other information from dataset
- Insufficient training data - 0.25
- Difficulties Moving Model to other Cloud (data inconsistencies, resource differences, cost variability) - 0.6
    - Mitigations: Change data preprocessing methods, use similar hardware configurations, setting budget requirements
- Data leaks - 0.8
    - Mitigations: Perform security tests, implement encryption

2.6 PERSONNEL EFFORT REQUIREMENTS

Task	Time Requirements per week
Understand the domain and problem	3 hours - only a high level understanding of the problem is required
Investigate AI as a solution	3 hours - high level understanding of AI techniques in healthcare
Model selection and training	6 hours - decision between AWS or GCP, this is where the development of the model starts, will take more effort
Improve model accuracy	8 hours - will take some more time, as it also involves developing strategies to improve model
UI development	5 hours - developing UI should be less complex than designing the model, but may take some time
Porting Model on cloud and training	3 hours - once the initial model is trained, it should not be too difficult to train on different cloud platform

Table 1. Personnel Effort Requirements

2.7 OTHER RESOURCE REQUIREMENTS

Resources needed:

- AI training data
    - The pathology dataset that acts as the input for the model
- Tensorflow
- Keras
- Cloud computing environment
    - GCP or AWS
- Tkinter for UI Design

4 Design

4.1 Design Content

The design content of our project involves creating an AI-based medical diagnosis tool for predicting the likely prognosis of cancer. This tool will consist of multiple components and subsystems, including data preprocessing, AI model selection and training, user interface development, and performance evaluation.

4.2 Design Complexity

Our project will be of sufficient technical complexity as it will consist of multiple components and subsystems, including data preprocessing, AI model training, user interface development, and cloud integration. Each of these components involves various scientific, mathematical, and engineering principles such as understanding the biological and medical aspects of cancer, mathematical principles in AI model training, and software engineering principles in user interface development.

4.3 Modern Engineering Tools

The modern engineering tools used in our design include:

- Tensorflow: A deep learning framework used for developing/training AI models [5]
- GCP/AWS: Cloud computing platforms that will be used for model deployment and scalability [7]
- Keras: Open source library that provides an interface for Neural Networks [4]
- UI: Tkinter for UI development platform [6]

4.4 Design Context

Our design problem deals with the current accuracy of cancer diagnosis tools. Our goal is to create a tool to fill the gap and create an accurate model of cancer recognition. This tool would help serve communities affected and as well as help contribute to the community of research behind cancer diagnosis.

Area	**Description**	**Examples**
Public health, safety, and welfare	Our product directly impacts public health by providing a tool that can aid in the diagnosis of cancer, which can decrease cancer related deaths and complications	Our product directly impacts public health by providing a tool that can aid in the diagnosis of cancer.

Global, cultural, and social	Our project aims to benefit 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.	Our project aims to benefit 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	We don't expect our project to have much of a direct or indirect environmental impact.	We don't expect our project to have much of a direct or indirect environmental impact.
Economic	As our project could potentially lead to improved patient outcomes in the future, this could result in reduced healthcare costs as cancer diagnoses will become more efficient.	As our project could potentially lead to improved patient outcomes in the future, this could result in reduced healthcare costs.

Table 2. Design Considerations

4.5 Prior Work/Solutions

- IBM Watson for Oncology: IBM Watson for Oncology is an AI-System developed by IBM to assist oncologists in making treatment decisions [8]. It has access to a large amount of medical research and clinical trial data and offers treatment recommendations based on a patient's medical records. Its advantage lies with its extensive knowledge base but it falls short in that it relies heavily on structured data.
- BreastIHC: BreastIHC is an AI System that can detect and quantify breast cancer cells for primary diagnosis [9]. The software is able to distinguish between tumorous and non-tumorous structures on a cellular level. Along with its capabilities, an added advantage is that it is a plug and play solution that doesn't require complex setup or calibration, but a disadvantage is that it is only for breast cancer.

4.6 Design Decisions

List key design decisions (at least three) that you have made or will need to make in relation to your proposed solution. These can include, but are not limited to, materials, subsystems, physical components, sensors/chips/devices, physical layout, features, etc.

1. Model will need to be able to run with our designed UI application
2. Model will need to have an accuracy of at least 70%
3. Model will be developed using GCP and Tensorflow
4. Model will be transferable to other AI platforms

4.7 Proposed Design

We have started designing basic models on GCP to gain some familiarity with AI and the process of developing ML models. We can monitor what parameters lead to an accurate model by training basic models.

4.7.1 Design 0 (Initial Design)

Design Visual and Description

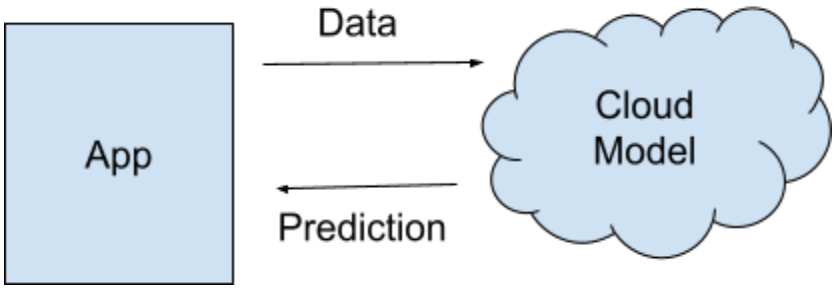

Figure 4. Initial Design

Our design will send data to our model on the cloud, and the model will return the diagnosis prediction.

Functionality

Our design is intended to allow a user to input an image, and our GCP model will identify whether the image is cancerous or not with 70% or more accuracy. The front-end should be visually appealing and easy to navigate for the user.

Our current design satisfies the requirements of allowing a user to put an image through our model, but doesn't accurately identify the image with a high enough accuracy. It also doesn't have a UI yet so it can't satisfy the requirement of being easy to use or looking nice.

4.7.2 Design 1 (Design Iteration)

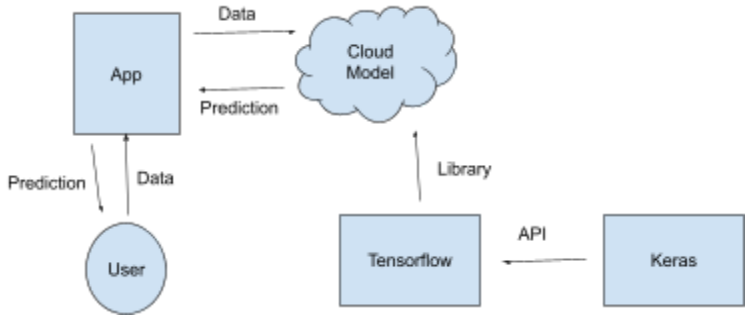

Figure 5. Design 1

Design Visual and Description

This design shows the user interacting with the front-end UI. The model predictions are retrieved the same way as design 0, except now it considers the two frameworks running on top of it: Tensorflow and Keras.

Functionality

This design satisfies the requirement of allowing the user to input data and retrieve a prediction based on the data. This design also includes more depth about what is powering the cloud model. It's important to consider the libraries and APIs of the model so that we can better understand its parameters and methodology. This will ensure our model is most optimized based on the platforms we are using, thus increasing accuracy.

4.8 Technology Considerations

As of now, we intend on using computers compatible with AWS/GCP, and open-sourced platforms such as Keras and TensorFlow. It is also important to ensure the computer can run the UI development platform. Those are the only major technology considerations we need since GCP runs model training on the cloud.

4.9 Design Analysis

The design, as visually depicted above, should work because it meets all the requirements to crosstalk between user and the cloud model. We will continue to refine our model as needed until desired accuracy is achieved. The design also considers the APIs that work behind the scenes. It is important to consider this in a design because we can ensure our model works well with the APIs and platforms powering it.

5 Testing

5.1 Unit Testing

Units being tested: The units to be tested in our project include individual components such as data preprocessing, AI model training, user interface, and cloud integration.

How: Unit testing will be performed during development using testing frameworks like JUnit for java or PyTest for Python. Each unit will be tested in isolation to verify its correctness.

Tools: JUnit, PyTest, or other testing frameworks relevant to the programming languages used in this project.

5.2 Interface Testing

Interfaces in the Design: The key interfaces in our design include interaction between the user interface and the AI model, as well as connection between the cloud computing environment and the model.

How: Interface testing will ensure that data is transmitted correctly between these components. It involves testing how different units interact and pass data between each other.

Tools: Custom testing scripts designed specifically to test the interaction between the user interface and the AI model. These scripts will simulate various user interactions with the UI, including data input and retrieval.

5.3 Integration Testing

1. **Data Preprocessing and Validation:**
    - **Criticality:** Proper preprocessing of the input data is crucial for the model's performance. Validation is essential to ensure that the input data meets the required format and quality standards.
    - **Testing:** Implement unit tests to check data preprocessing functions. Use sample datasets to validate that the model accepts and processes data correctly.
2. **User Interface (UI) Design and User Input Handling:**
    - **Criticality:** The UI should be user-friendly, allowing users to easily upload cancer data. Proper input handling is crucial for collecting the required information for prediction.
    - **Testing:** Perform usability testing to ensure the UI is intuitive. Conduct tests to verify that the UI sends the correct input to the model and handles errors gracefully.
3. **Security and Privacy:**
    - **Criticality:** Medical data is sensitive, and ensuring the security and privacy of patient information is paramount.
    - **Testing:** Perform security audits to identify vulnerabilities. Potentially implement encryption for data transmission and storage.
4. **Model Monitoring and Updating:**
    - **Criticality:** Regularly monitoring the model's performance and updating it with new data is crucial for maintaining accuracy over time.
    - **Testing:** Implement tests to simulate model updates and verify that the system seamlessly integrates new versions. Monitor the model's performance over time to ensure it meets accuracy requirements.

5.4 System Testing

- **Objective:** Verify the entire system against specified requirements, ensuring it functions as a unified whole.
    - Requirements: data maintains confidentiality, prediction has accuracy of at least 70%, UI is easy to navigate
- **Scope:** Comprehensive end-to-end testing, encompassing user interactions, data processing, and external system integrations. Need to ensure that the user can upload data, data maintains confidentiality, and the model can return a prediction to a user
- **Potential Tools:**
    - Selenium with JUnit/TestNG: For end-to-end testing of web applications [10].
    - Postman: For end-to-end testing of APIs [11].
- **Considerations:** Validate that the system meets all specified requirements, and simulate real-world scenarios to ensure robustness
    - Compile several test data, run several round trips to ensure product behaves as expected

5.5 Regression Testing

To prevent new additions from breaking existing functionality, we'll employ a robust regression testing strategy. This involves maintaining comprehensive tests, utilizing version control for tracking changes, incorporating automated testing tools, and managing test data. To ensure that critical features do not break, we will focus on those directly linked to the accurate prediction of cancer risk, such as data input processing, model training, and prediction outputs. This approach is driven by the project's core requirements, emphasizing the stability of cancer prediction functionality.

5.6 Acceptance Testing

For acceptance testing, we'll create test scenarios based on project requirements, involving the client in defining acceptance criteria. Also, demo sessions will be organized to showcase the AI model's performance, ensuring that it aligns with specified requirements and meets client expectations. This collaborative and iterative approach guarantees that the AI model is validated against user needs and meets both functional and non-functional design requirements.

5.7 Security Testing (if applicable)

Since the system deals with sensitive medical data, security testing is essential. It will assess the system's ability to protect data from unauthorized access, breaches, and vulnerabilities.

Some security testing measures we could take to ensure patient security would be implementing authentication and authorization systems, data encryption, and access control.
What this looks like is ensuring that only authorized individuals have access to the model via a login.

- **Objective:** Identify and address security vulnerabilities in the system.
- **Scope:** Test for vulnerabilities, including data encryption, authentication, and authorization mechanisms. Ensure data is not being compromised.
- **Potential Tools:**
    - **Burp Suite:** For manual security testing [12].
- **Considerations:** Ensure that the system complies with security standards and regulations.
- Methods: Upload data and perform tests to ensure data integrity. This includes testing every step of the application.

5.8 (Expected) Results

As we are developing the UI and model, we hope that all tests can pass successfully before moving on to the next phase of each feature. For the UI, ideally we would have checks in place to verify and authenticate users in order to protect patient data. As we develop each feature, unit tests should cover as many cases as possible and future proof our code. We'll utilize various testing tools and software to robustly test our models and code. To ensure compliance with the requirements we will work closely with the client and develop scenarios to ensure performances.

Predicted Regression: In the context of regression testing, it's crucial to acknowledge that AI models, including ones developed for predicting cancer risk, may encounter regression over time. Regression in AI models refers to a decline in its performance or accuracy compared to a previous state. We expect the AI

model to become exposed to new data and, with this, we predict the model's accuracy has a chance to decrease due to changes in underlying patterns of the data or the introduction of novel factors. To mitigate the impact of regression and uphold the model's accuracy, we will implement mechanisms for continuous monitoring of the AI model's performance, establish automated testing tools, and utilize version control to ensure previous versions can be reverted to.

6 Implementation

For next semester, we will begin working directly with the data. So while our progress this semester involved mostly research and practice, next semester we will:

1. Extract useful information from dataset
    a. This involves determining which features in the data will be most useful and valuable to teach a ML model
2. Modify preliminary models to work with extracted information
    a. We have worked with creating generic models. The next step is to modify these generic models to work with our data
3. Incorporate working model with UI
    a. Our model allows us to run ML models on the backend. Once the final model is created, we can incorporate it into the GUI
4. Port working model to other development platforms (AWS)
    a. The final step of our project is to compare model performance on different ML model platforms. Since our initial model will be designed on GCP, we need to ensure it can work on AWS and potentially more.
5. Compare results
    a. Compare the results of the models on both platforms to determine the best one

7 Professionalism

7.1 Areas of Responsibility

IEEE Ethic: I. To uphold the highest standards of integrity, responsible behavior, and ethical conduct in professional activities

Area of responsibility	**Definition**	**NSPE Canon**	**Ethic**
Work Competence	Perform work of high quality, integrity, timeliness, and professional competence	Perform services only in areas of their competence; Avoid deceptive acts	Working only in areas of competence ensures you are upholding highest standards of integrity
Financial Responsibility	Deliver products and services of realizable value and at reasonable costs.	Act for each employer or client as faithful agents or trustees.	Financial responsibility exhibits responsible behavior

Communication Honesty	Report work truthfully, without deception, and understandable to stakeholders.	Issue public statements only in an objective and truthful manner; Avoid deceptive acts	Having open and honest communication shows responsible behavior and integrity
Health, Safety, Well-Being	Minimize risks to safety, health, and well-being of stakeholders.	Hold paramount the safety, health, and welfare of the public.	By having high standards of responsible behavior, you will mitigate safety and health risks
Property Ownership	Respect property, ideas, and information of clients and others.	Act for each employer or client as faithful agents or trustees.	Responsible behavior includes respecting property and ideas. Aligns with NSPE Canon
Sustainability	Protect environment and natural resources locally and globally.		Respecting the environment exhibits a high standard or integrity and behavior
Social Responsibility	Produce products and services that benefit society and communities.	Conduct themselves honorably, responsibly, ethically, and lawfully so as to enhance the honor, reputation, and usefulness of the profession	This encapsulates ethic code 1, overall being responsible, honorable, and maintaining integrity in social situations

Table 3. Areas of Responsibility

7.2 Project Specific Professional Responsibility Areas

- **Work Competence** - We are all performing work in areas of competence since our project only accepts majors of study in a relevant area. We are all software/computer engineers developing software systems. Team performance = High
- **Financial Responsibility** - We are only using programs approved by the university and advisors. We are using as many cost effective programs as possible. Team performance = High
- **Communication Honesty** - We are communicating thoroughly with each other and our advisors. We are consulting them before any major decisions to prevent any confidentiality issues. Team performance = High
- **Health, Safety, Well-Being** - We are ensuring that confidential information remains confidential, which minimizes safety and well-being of patients. Team performance = High
- **Property Ownership** - We are respecting the confidential patient data by complying with NDA standards and ensuring vulnerabilities are minimized. Team performance = High
- **Sustainability** - We are not contributing to any environmental issues. Our project utilizes resources already available. Team performance = High

- **Social Responsibility** - Our project aims to minimize risks related to cancer and help maximize productivity for healthcare professionals, which benefits many people. Team performance = High

7.3 Most Applicable Professional Responsibility Area

The most applicable professional responsibility area related to our project is Health, Safety, and Well-Being since our project is directly concerned with improving the health and well-being of individuals. There are also numerous security concerns we must consider with our project that may affect the safety and well-being of patients.

8 Closing Material

8.1 Discussion

Our project meets the qualifications of an experiment-oriented project. Thus far, our results suggest that it is possible to design a model capable of determining the prognosis of a cancer diagnosis. However, the accuracy of such a model remains to be known.

Our initial experiments involve generating basic ML models using Tensorflow and Keras. We have been able to follow Tensorflow documentation and reproduce the flower recognition model, capable of recognizing flower types. This exhibits our ability to successfully generate ML models using Tensorflow. We have also discovered plenty of Keras models that take in several data types such as vectors and text. This is relevant because, at this point, we are unsure of exactly what data our model will need to take as input. This question will be answered next semester when we start working directly with the data.

We know it is possible to develop a UI capable of accepting model metadata and providing results to the user. As such, our experimentation with Tensorflow, Keras, and Tkinter suggests it is possible to create a model capable of providing a cancer prognosis, but the accuracy of such a model is still in question

Our work with the cancer pathology dataset will provide more clarity on some of our questions.

8.2 Conclusion

Much of our project thus far has included research and preliminary work so that we can start working with the dataset in 492. Up to this point we have:

1. Researched the medical problems. This includes understanding pathology data and some traditional cancer detection methods. One of our goals was to fully understand the medical problem so this step helped achieve that goal and create an efficient app as possible.
2. Understood the dataset. This includes reviewing the data we have access to and planning what useful data can be extracted and used as input to a neural network. This aligns with our previous goal of understanding the medical problem and traditional methodologies. There is still more to learn with the dataset, but as we work closely with it next semester, this goal will be met.
3. Researched feasible model development platforms. We have begun experimentation with Google Colab using Tensorflow. One of our goals was to determine the best cloud development platform.

We used our research to determine that Google Cloud would be the best course of action to design our model.

4. We have begun designing basic ML models using Tensorflow Core and its supplemental documentation. One model we have generated so far is the flower image classification model. This aligns with our goal of creating a model of at least 70% accuracy. If we can understand the development process as much as possible, this will lead to better results in the final product. Working with the dataset and understanding it as much as possible will help ensure our goal of high performance is met.
5. Designed a prototype UI. The basic UI allows users to input their data, and runs the data through our model on the backend. The prediction is then returned to the user on the front end. One of our goals was to develop a user interface that allows users to insert their own patient data. This prototype UI is a good starting point for this goal. Continuing to refine and research efficient UI design solutions will help us reach this goal.

8.3 References

[1] O. for C. Rights (OCR), "Health Information Privacy," *HHS.gov*, Jun. 09, 2021. http://www.hhs.gov/hipaa/index.html

[2] "IEEE Code of Ethics," *IEEE*, 2019. http://www.ieee.org/about/corporate/governance/p7-8.html (accessed Sep. 20, 2023).

[3] "ISO Code of Ethics and Conduct," *ISO*, Jan. 01, 1970. http://www.iso.org/publication/PUB100011.html (accessed Sep. 20, 2023).

[4] "Keras: Deep Learning for humans," *Keras.io*, 2018. http://keras.io (accessed Sep. 20, 2023).

[5] "API Documentation | TensorFlow v2.14.0," *TensorFlow*. http://www.tensorflow.org/api_docs (accessed Sep. 15, 2023).

[6] "Graphical User Interfaces with Tk," *Python Documentation*. http://docs.python.org/3/library/tk.html (accessed Nov. 10, 2023).

[7] "Google Cloud Platform Overview," *Google Cloud*. http://cloud.google.com/docs/overview (accessed Sep. 20, 2023).

[8] "IBM 5725-W51 IBM Watson for Oncology.," *IBM Documentation*. http://www.ibm.com/docs/en/announcements/watson-oncology?region=CAN (accessed Nov. 02, 2023).

[9] "Mindpeak News | BreastIHC CE-IVD certified," *mindpeak.ai*. http://www.mindpeak.ai/post/breastihc-ce-ivd-certified (accessed Oct. 15, 2023).

[10] "SeleniumHQ Browser Automation," *www.selenium.dev*, 2023. https://www.selenium.dev/ (accessed Nov. 08, 2023).

[11] "Postman | The Collaboration Platform for API Development," *Postman*, 2021. https://www.postman.com/ (accessed Nov. 08, 2023).

[12] "Burp Suite Scanner," *Portswigger*, 2019. https://portswigger.net/burp (accessed Nov. 09, 2023).

8.4 Appendices

- Tensorflow Documentation:
  - https://developers.google.com/machine-learning/crash-course/first-steps-with-tensorflow/toolkit
  - https://www.tensorflow.org/api_docs
- Google Cloud Platform Documentation
  - https://cloud.google.com/docs/overview
- Google Colab Documentation:
  - https://colab.research.google.com/?utm_source=scs-index
- Keras.io Documentation:
  - https://keras.io/
- Tkinter Documentation:
  - https://docs.python.org/3/library/tk.html

8.4.1 Team Contract

Team Members:

1) Jack Sebahar 2) Nicholas Otto

3) Mason Wichman 4) Isaiah Mundy

5) Helen Lau 6) Lal Siama

Team Procedures

1. Day, time, and location (face-to-face or virtual) for regular team meetings:

- Every Thursday after the senior design lecture at about 4:00. End meetings when satisfied with progress.
- Face-to-face in Howe lecture hall

2. Preferred method of communication updates, reminders, issues, and scheduling (e.g., e-mail, phone, app, face-to-face):

- Communication will take place using Discord
- Allows messages to be directed towards certain individuals or the entire group

3. Decision-making policy (e.g., consensus, majority vote):

- Decision-making will be done based on group consensus

4. Procedures for record keeping (i.e., who will keep meeting minutes, how will minutes be

shared/archived):

- Meeting minutes will be tracked via Discord, as it can automatically track when the meeting starts and when it ends based on messages sent. All messages sent are saved on Discord and can be viewed any time.
-

Participation Expectations

1. Expected individual attendance, punctuality, and participation at all team meetings:

- Team members are expected to attend and be punctual for scheduled meetings. In the case of conflicts, prior notification is expected.

2. Expected level of responsibility for fulfilling team assignments, timelines, and deadlines:

- Each team member is responsible for fulfilling assigned tasks within agreed-upon timelines. Delays should be communicated promptly, and collaborative problem-solving is encouraged.

3. Expected level of communication with other team members:

- All team members are expected to actively engage in discussions, share progress updates, and promptly respond to messages or emails from other team members.

4. Expected level of commitment to team decisions and tasks:

- Once a team decision is made, all members are expected to commit to it. If there are individual concerns or disagreements, they should be addressed constructively within the team framework.

Leadership

1. Leadership roles for each team member (e.g., team organization, client interaction,

individual component design, testing, etc.):

- Jack Sebahar - Project Manager/Client interaction
- Nick Otto - Model Design lead
- Isaiah Mundy - Model Design Tester
- Mason Wichman - UI lead
- Helen Lau - UI Design
- Siama - UI Tester

2. Strategies for supporting and guiding the work of all team members:

- Regular check-ins and team meetings will provide an opportunity for discussing progress and potential challenges. Collaborative problem-solving will be emphasized, and team members are encouraged to seek help when needed.

3. Strategies for recognizing the contributions of all team members:

- Recognition will be given during team meetings and presentations, acknowledging individual achievements and contributions. Members of the team wil be able to present their progress, thus displaying their contributions.

Collaboration and Inclusion

1. Describe the skills, expertise, and unique perspectives each team member brings to the team.

- Jack - Computer engineering student with experience in machine learning and developing models using tensorflow and GCP.
- Nick - Software Engineering student with experience in a number of programming languages and frameworks, as well as some industry experience in cloud computing
- Mason - Software Engineering student with experience collaborating in many software related group projects. With a strong foundation in programming and teamwork.
- Isaiah - Software Engineering student with experience developing software for a variety of different team and individual projects.
- Helen - Software Engineering student with experience in developing software in a number of programming languages.
- Siama - Electrical Engineering student with experience in some programming languages, hardware related designs, and research experience in bioinformatics.

2. Strategies for encouraging and support contributions and ideas from all team members:

- Strategies involve creating an open and collaborative environment. This includes regular team meetings and check-ins where active engagement and the sharing of diverse perspectives are valued.

3. Procedures for identifying and resolving collaboration or inclusion issues (e.g., how will a team member inform the team that the team environment is obstructing their opportunity or ability to contribute?)

- Team members are encouraged to communicate openly about obstacles hindering their ability to contribute. Regular check-ins will provide opportunities to discuss team dynamics, and addressing any arising issues promptly is essential to maintaining a collaborative environment.

Goal-Setting, Planning, and Execution

1. Team goals for this semester:

Since our project is more experimental, these 3 things are important to prepare for next semester:

- Understand pathology data characteristics
    - View scholarly articles regarding cell culture
- Generate a simple AI model using Tensorflow
    - Results do not necessarily matter, purpose is to gain an understanding of the process
- Design a prototype UI
    - Capable of reading model data

2. Strategies for planning and assigning individual and team work:

- Individual work will be assigned based on volunteering, however, the team will ensure all aspects of the assignment are covered by someone. If certain tasks are open and nobody volunteers, the project manager will assign roles that are fair. We will assign teams based on each individual's skill set. If they have the skills to complete a certain task, then that task will be delegated to them. For example, the UI team will be comprised of people who have experience developing UI.

3. Strategies for keeping on task:

- The Project manager will ensure each member is on task during meetings. If a member of the team is not on task, the project manager and/or other team members will provide constructive feedback aimed toward getting that individual back on task. Meetings will not be held for more than several hours, as long work periods are not conducive to maintaining focus.

Consequences for Not Adhering to Team Contract

1. How will you handle infractions of any of the obligations of this team contract?

- Initial infractions will be addressed through open communication within the team. A member who fails to adhere to the contract will be approached in a constructive manner, and the reasons behind the infraction will be discussed. The team will collectively assess whether there are underlying issues that need resolution.

2. What will your team do if the infractions continue?

- If infractions continue despite initial discussions, a team meeting will be held to address the issue more formally. Potential consequences may include revisiting role assignments, redefining individual tasks, or, in extreme cases, seeking intervention from the project advisor.

****

a) I participated in formulating the standards, roles, and procedures as stated in this contract.

b) I understand that I am obligated to abide by these terms and conditions.

c) I understand that if I do not abide by these terms and conditions, I will suffer the

consequences as stated in this contract.

1) Jack Sebahar DATE 12/2/23

2) Nicholas Otto DATE 12/2/23

3) Mason Wichman DATE 12/2/23

4) Isaiah Mundy DATE 12/3/23

5) Helen Lau DATE 12/3/23

6) Lal Siama DATE 12/3/2023