



FourthBrain

# Automating Explainability in Healthcare

## Supporting Org Information

Company	Samsung SDS Research America
Location (city/time zone)	San Jose, California

<b>Supporting Org Information</b>	<b>1</b>
<b>AI Product/Capstone Project Description</b>	<b>2</b>
Problem	2
Why	2
Success	3
Audience	3
What	4
Final Deliverables	4
Anything Else?	5



# AI Product/Capstone Project Description

## Problem

[Write a succinct statement of the problem that you're trying to solve (<50 words)]

Cancer-related diagnoses from tissue imaging are hard to explain, and high-resolution datasets are limited.

## Why

Write about why this is a problem worth solving. What is the business value hypothesis that connects to what success looks like and for whom? (~50-250 words)

Data collection technologies like [whole-slide imaging](#), 3D MRI, and X-Ray are all able to provide high-resolution data on biological tissue. Some of this data may be in three dimensions as a collection of two-dimensional slices.

From tissue images, doctors are currently able to leverage heat maps, lists of similar images from training sets, and highlighting of specific high-level image features. However, more detailed descriptions (both visual and written) of exactly why a categorical diagnosis of “cancer” is made is needed.

At the same time, breakthroughs in Computer Vision, Natural Language Processing, and training of large transformer models have paved the way for new types of healthcare-specific AI modeling opportunities.



## Success

Write about what success looks like. What is the Key Performance Indicator (or couple of KPIs)? How might they connect to a relevant ML model accuracy metric? (<50 words)

This is a *research*-focused project focused on *computer vision* and *explainable AI classification* techniques.

Successful projects will produce a generic Python code and ML web application capable of assessing whole-slide images of potentially cancerous tissue and producing a diagnosis and corresponding auto-generated explanation report.

As a modeling-focused project, *the key performance indicator for this work is a set of ML model accuracy metrics chosen for analyzing each model's performance.*

## Audience

Specify exactly which users/customers this AI/ML product is being built for. What is the customer's pain or need that connects back to the problem? (<50 words)

A functional AI product, once developed, would be used by a Doctor who is able to work with the product to help explain "this is a cancer cell because..."



## What

Now describe what the ML looks like. This includes a discussion of data and sources, potential/likely models, a choice of an accuracy metric to optimize for and a defense of your choice. How does your accuracy metric connect back to the KPI(s) named above?

### Data

- The fully annotated dataset of breast cancer whole slide images is [available as open-source](#). The original whole slide images have been cut into small tiles due to the high-resolution. Each tile has been expertly analyzed and annotated. A computer vision model that provides the correct output with high-probability is described in the documentation.

### Modeling

- Implement AI/ML models that ...
  - Automatically generate heat maps
  - Find similar images within the data set
  - Highlight specific high-level explainable features that correspond to cancer diagnoses
  - Provide an auto-generated written language summary of the explainability findings
- You may use pre-trained models, perform transfer learning, train from scratch, or use any methods that appear useful for the purpose of providing an actionable explanation to a non-AI-expert such as a physician or patient.
- Comparison of explainability approaches (including [SHAP](#) and [LIME](#)) on utility for doctors
- Survey of pre-trained [image-to-text models](#) and recent image-to-text model research (e.g., [Wang. et al.](#))

You may also find it helpful to fill out an [MLOps Stack Canvas](#) or [MLOps Stack Template](#). These tools really help to clarify tech stack requirements for our students.

## Final Deliverables

At the conclusion of a cohort, students are expected to deliver:

- Deployed Demo
- 10-minute Presentation
- GitHub Repo ( description in README + code )



For more detailed information on student capstone projects, you can check out the guide that we provide MLE students [here](#).

## Anything Else?

Please provide any additional information on key activities, technologies, datasets, expected learning outcomes, potential mentorship or employment opportunities, or anything else not listed above!. And thanks for supporting our students!

### Tool Recommendations:

- As a general framework, we recommend PyTorch.
- Specifically for this challenge, we recommend using any one of the many available source codes that accompany academic publications in the field of AI explainability.

### Recommended Course of Action

- Perform a wide literature research in the field of explainable AI for computer vision tasks. Most established computer vision explanations result in an image. Explanations resulting in text are preferable and more challenging. Your search will result in a collection of source codes that you can test/examine for usefulness.
- In order to examine these explanations, you will need a reference model. Use the data and annotations to train such a reference model. As the focus is on explanations here, the accuracy and general goodness of the reference model is not important so do not spend much effort in optimizing this model.

### Key research questions and technological constraints that the project will answer:

- How can we construct an understandable and trustworthy explanation for the AI diagnosis of cancer versus non-cancer given the obvious implications for the affected person? Trust, by doctors and the general population, is one of the central challenges in AI today. Explanations are the key instrument in overcoming it.

### Expected learning outcomes

- General understanding of explainable AI and its closely related research topic of ethical AI.
- Technical knowledge and experience in applying various explainability approaches to a computer vision use case. These might range from the



analysis of model weights (resulting in heatmaps), multi-model AI (resulting in text from images), efficient search (resulting in example images), and others.

- In depth understanding on how to measure the effectiveness of an explanation.

### **Background information**

- When a patient is diagnosed with breast cancer, the patient wants assurance that whatever or whoever said so is absolutely sure. They are anxious and afraid. They are looking for trust and certainty. The doctor, who has to approve or reject the AI findings, wants assurance that the AI has made this determination on the basis of evidence that the doctor agrees with. A mere black-box output is unacceptable. Currently over 5 billion medical images are taken every year, the number of pathologists is decreasing, costs are increasing, the market for second opinions is already a several billion dollar industry. AI is perfectly placed to relieve the pressure on this market if we can figure out how to provide consumable outputs. This is still a research area without an established result - there is no right answer.

### **About FourthBrain**

FourthBrain trains aspiring Machine Learning engineers in the technical and practical skills necessary to contribute immediately to an AI team. Our remote, online program is designed to be flexible and accessible for anyone with software experience. We infuse values of collaboration, communication, empathy, and equity throughout the program.

We are part of the AI Fund, founded by Andrew Ng.