Predicting Breast Cancer Using Computer Vision

By Deanna Hedges

The problem

We all know someone who has been impacted by breast cancer

The problem

We all know someone who has been impacted by breast cancer

We all want to **cure cancer**

The problem

We all know someone who has been impacted by breast cancer

We all want to **cure cancer**

There is a **market for technology** that assists in this goal

The solution

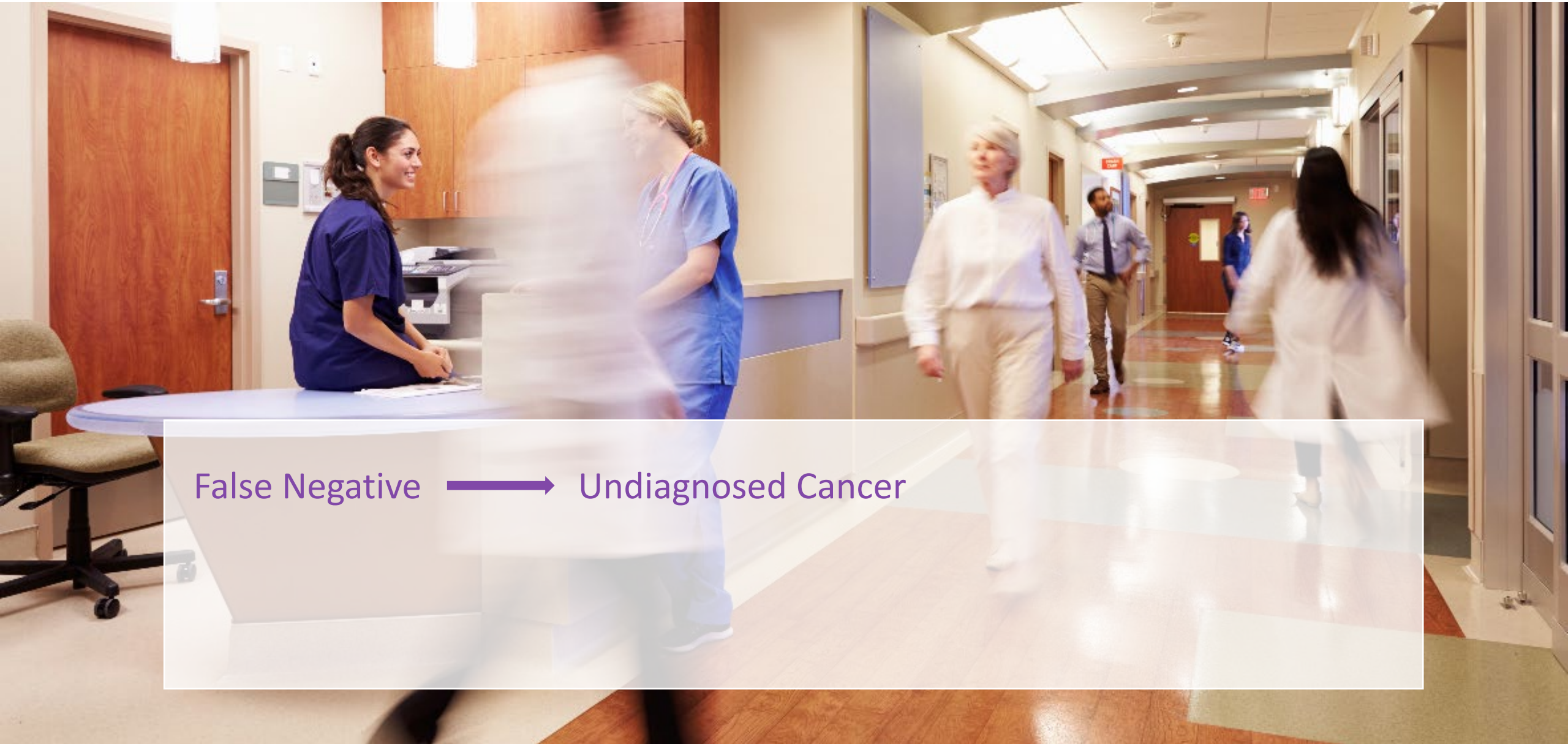
Saving a life starts with a diagnosis

The solution

Saving a life starts with a diagnosis

The solution

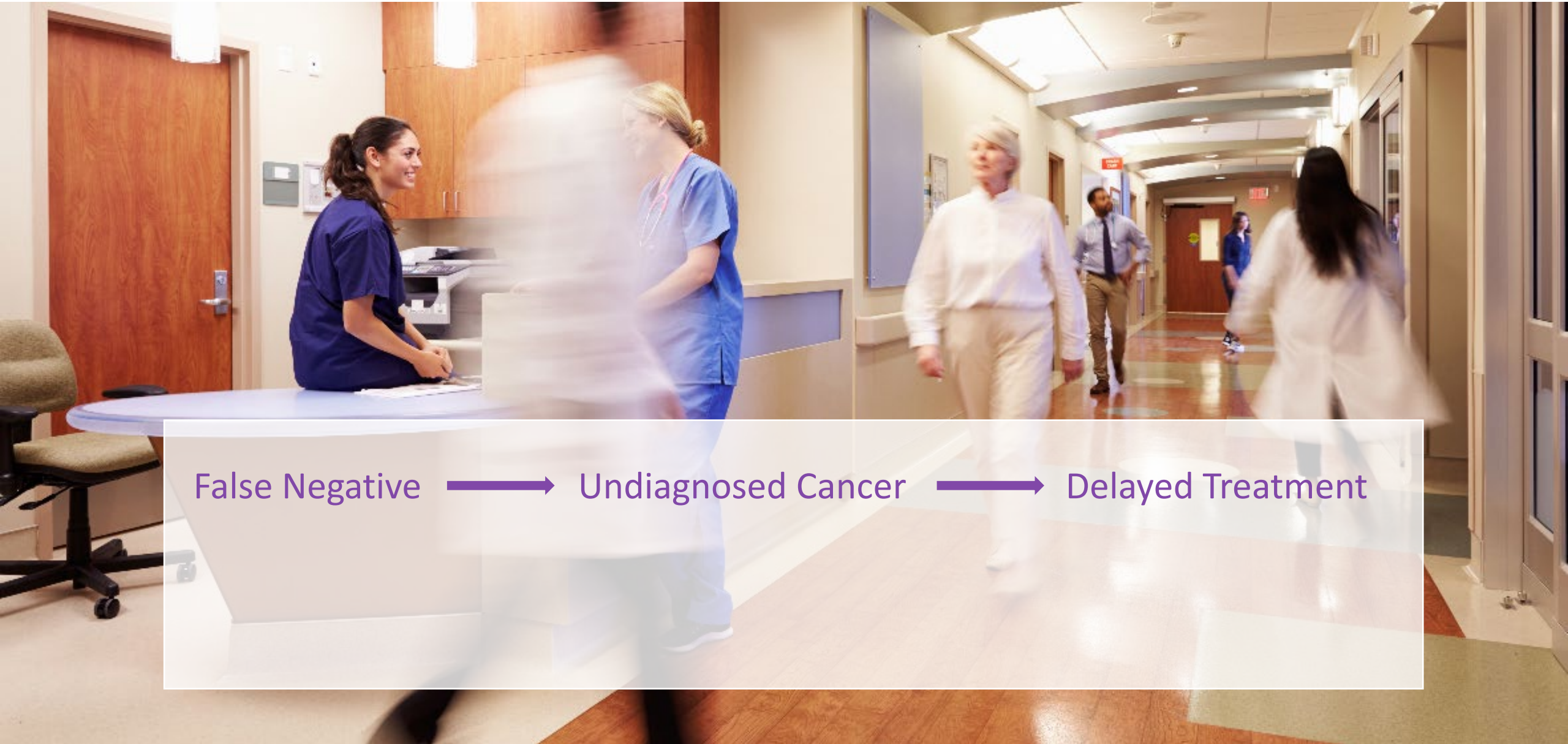
Saving a life starts with a diagnosis



False Negative → Undiagnosed Cancer

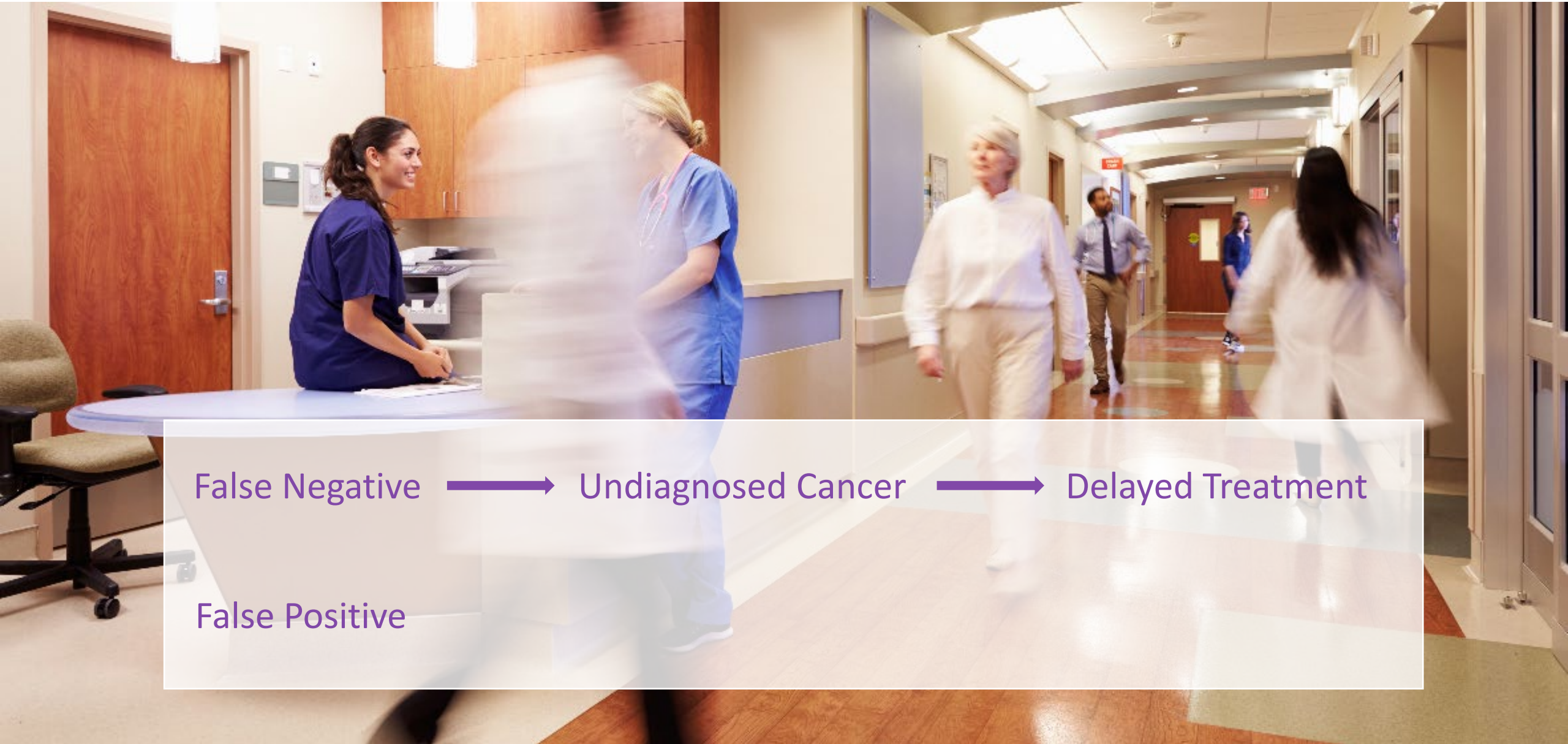
The solution

Saving a life starts with a diagnosis



The solution

Saving a life starts with a diagnosis



False Negative → Undiagnosed Cancer → Delayed Treatment

False Positive

The solution

Saving a life starts with a diagnosis

False Negative → Undiagnosed Cancer → Delayed Treatment

False Positive → Wrong Diagnosis

The solution

Saving a life starts with a diagnosis

False Negative → Undiagnosed Cancer → Delayed Treatment

False Positive → Wrong Diagnosis → Superfluous Treatment

The data

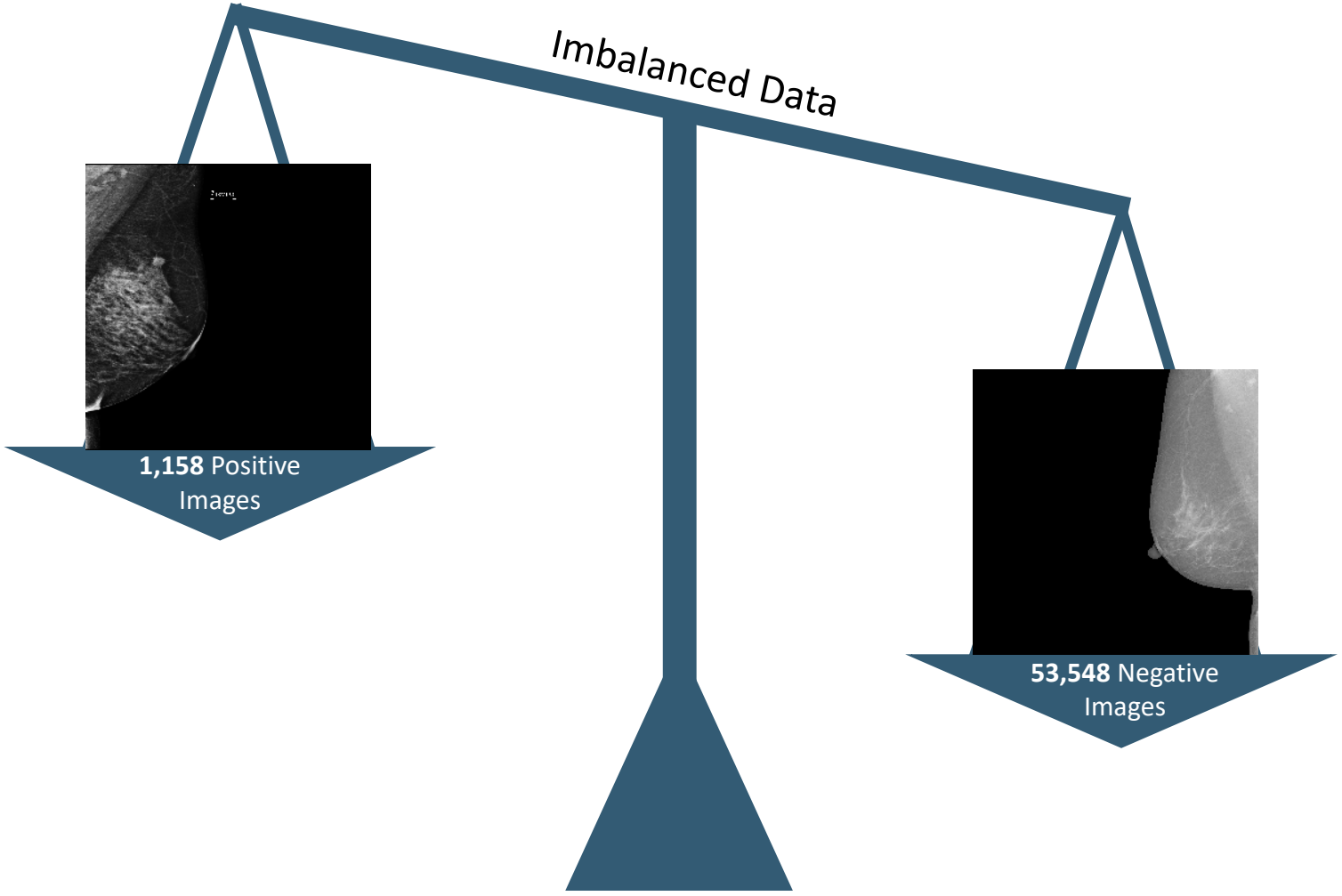
Mammography scan images provided by the Radiological Society of North America (RSNA)

About the source

- The RSNA is a non-profit organization
- 31 radiologic subspecialties are represented
- 145 countries are represented
- Accessed through Kaggle

About the data

- 54,708 scans from 11,913 patients
- DICOM file format
- Scans of individual breasts, some from multiple angles
- Varying placement, exposure, and negative space



The data

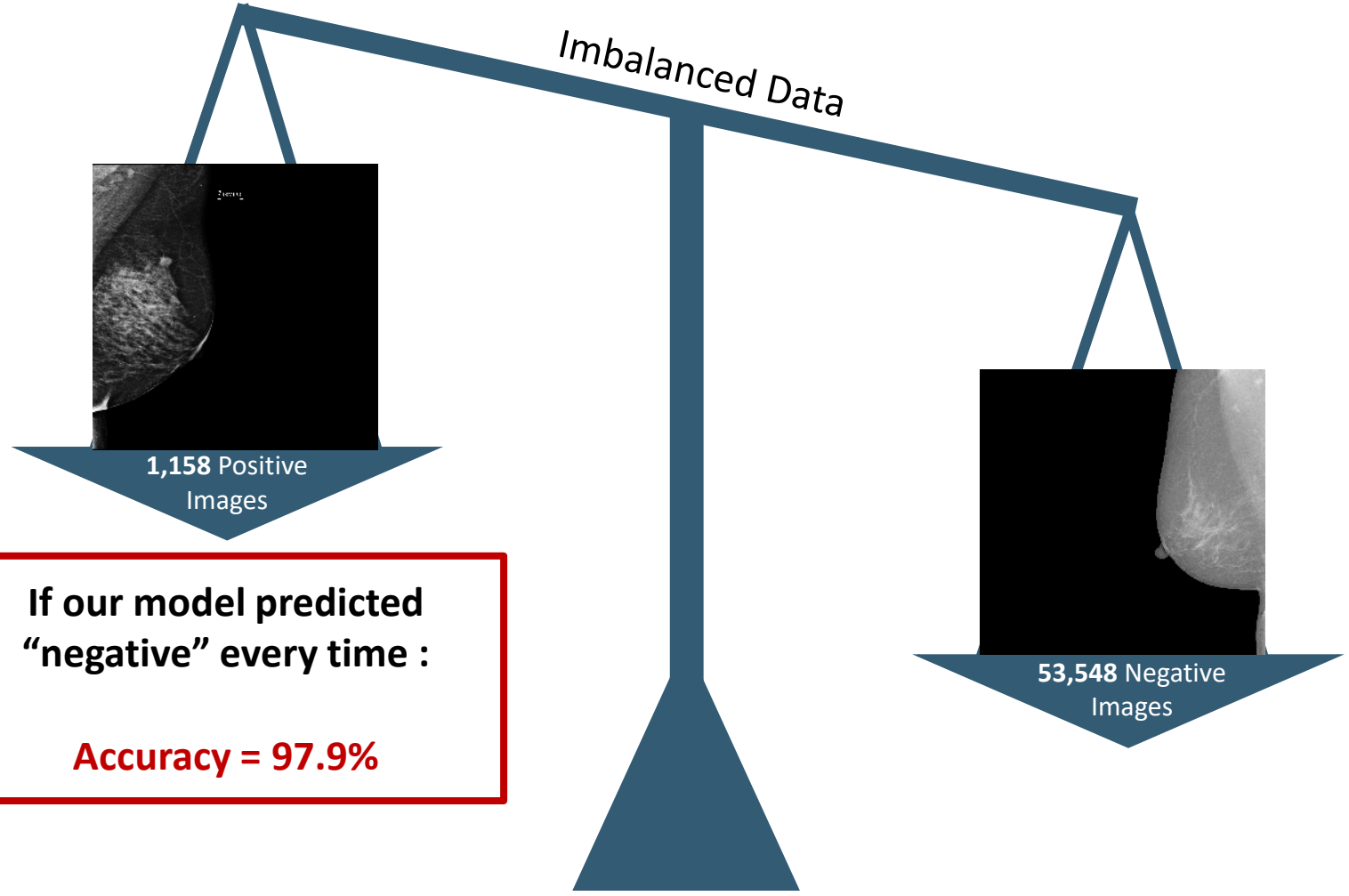
Mammography scan images provided by the Radiological Society of North America (RSNA)

About the source

- The RSNA is a non-profit organization
- 31 radiologic subspecialties are represented
- 145 countries are represented
- Accessed through Kaggle

About the data

- 54,708 scans from 11,913 patients
- DICOM file format
- Scans of individual breasts, some from multiple angles
- Varying placement, exposure, and negative space



If our model predicted "negative" every time :

Accuracy = 97.9%

The data

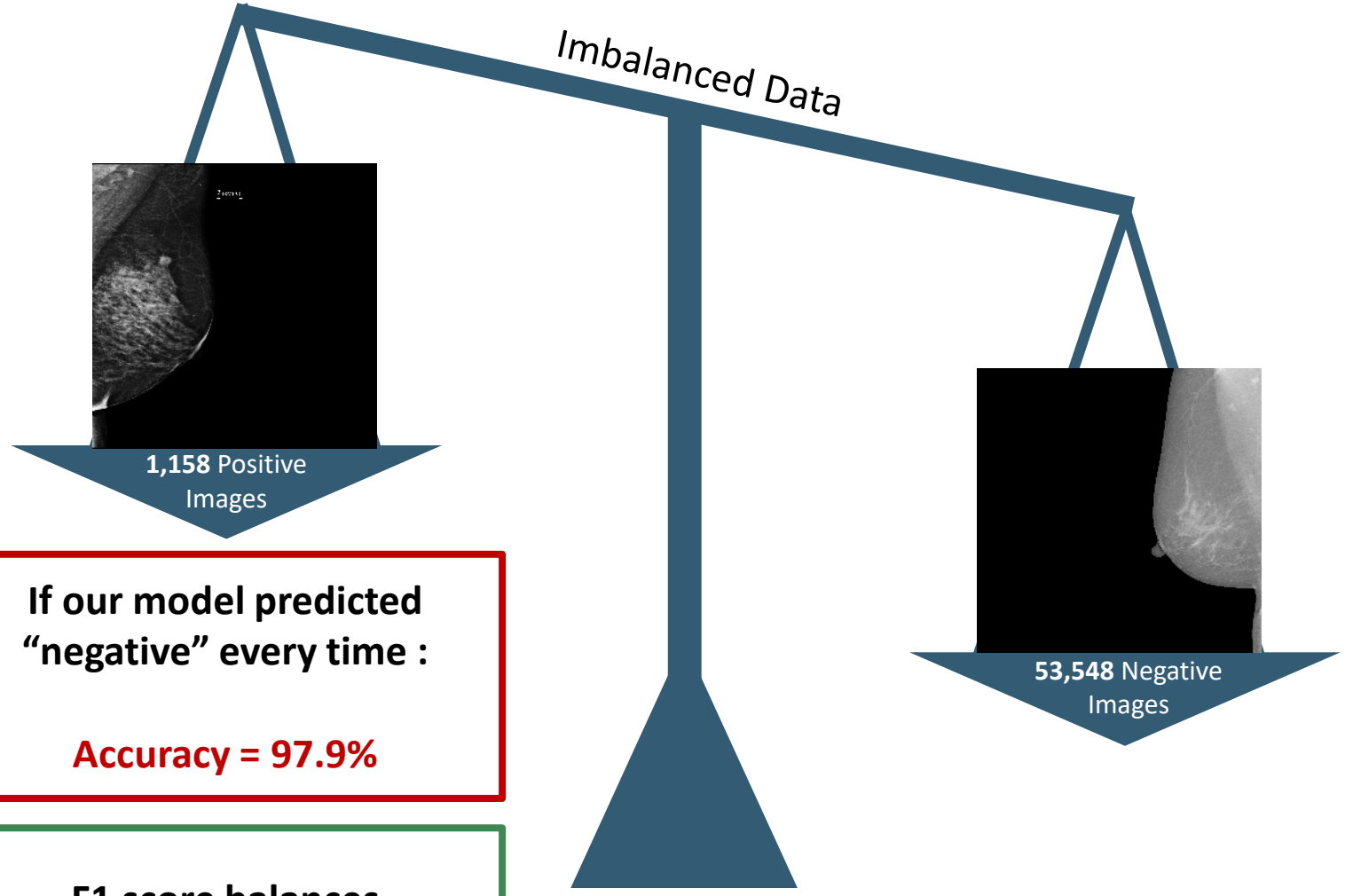
Mammography scan images provided by the Radiological Society of North America (RSNA)

About the source

- The RSNA is a non-profit organization
- 31 radiologic subspecialties are represented
- 145 countries are represented
- Accessed through Kaggle

About the data

- 54,708 scans from 11,913 patients
- DICOM file format
- Scans of individual breasts, some from multiple angles
- Varying placement, exposure, and negative space



If our model predicted "negative" every time :

Accuracy = 97.9%

F1 score balances importance of False Negatives and False Positives

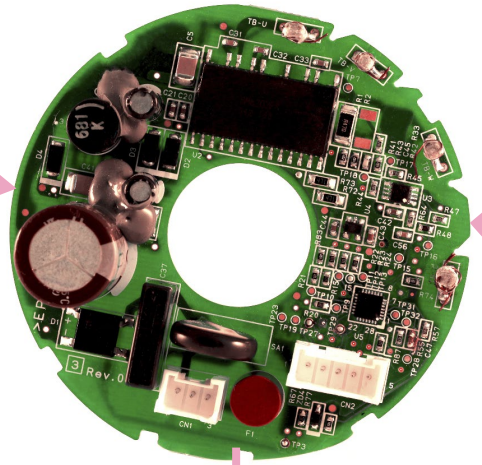
The method

Utilizing machine learning and computer vision to assist in diagnosis

Gather data and preliminary exploration

Convert images from DICOM to PNG format

Divide data into training and testing sets



Use transfer learning to train model on our data

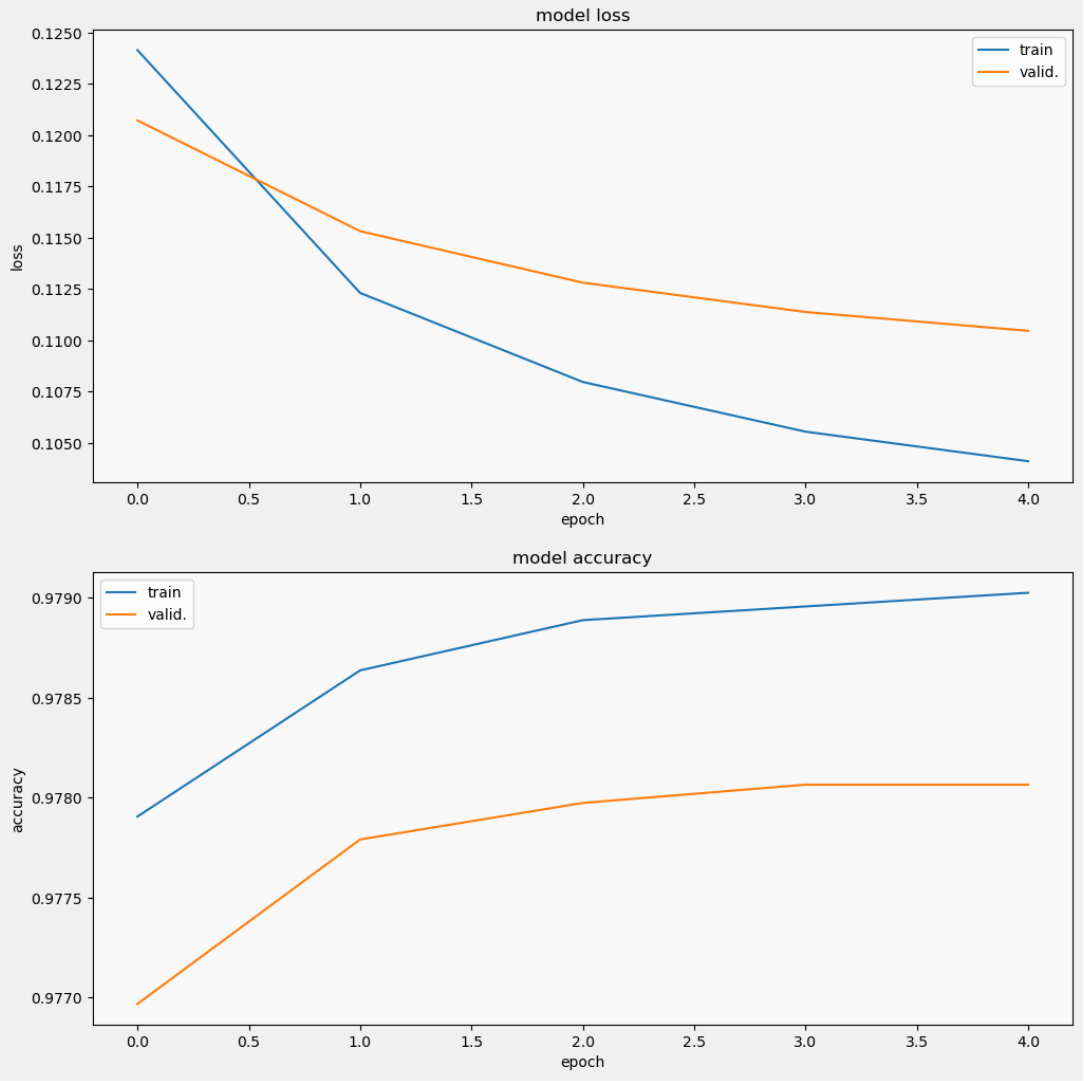
Use model to predict cancer in testing set images

Iterate and adjust hyperparameters to increase success metrics

Assess the model's potential value in the real-life diagnosis process

The results

How the model performed



About the model

- Convolutional Neural Network
- Transfer learning on VGG19 model
- Fairly shallow neural network
- Appropriately fit overall
- Overfit on the negative class

Next steps

- Over or under sampling
- Incorporating F1 score as a metric
- Data augmentation
- Adjusting hyperparameters
- Adding more layers for a deeper neural network

Questions?