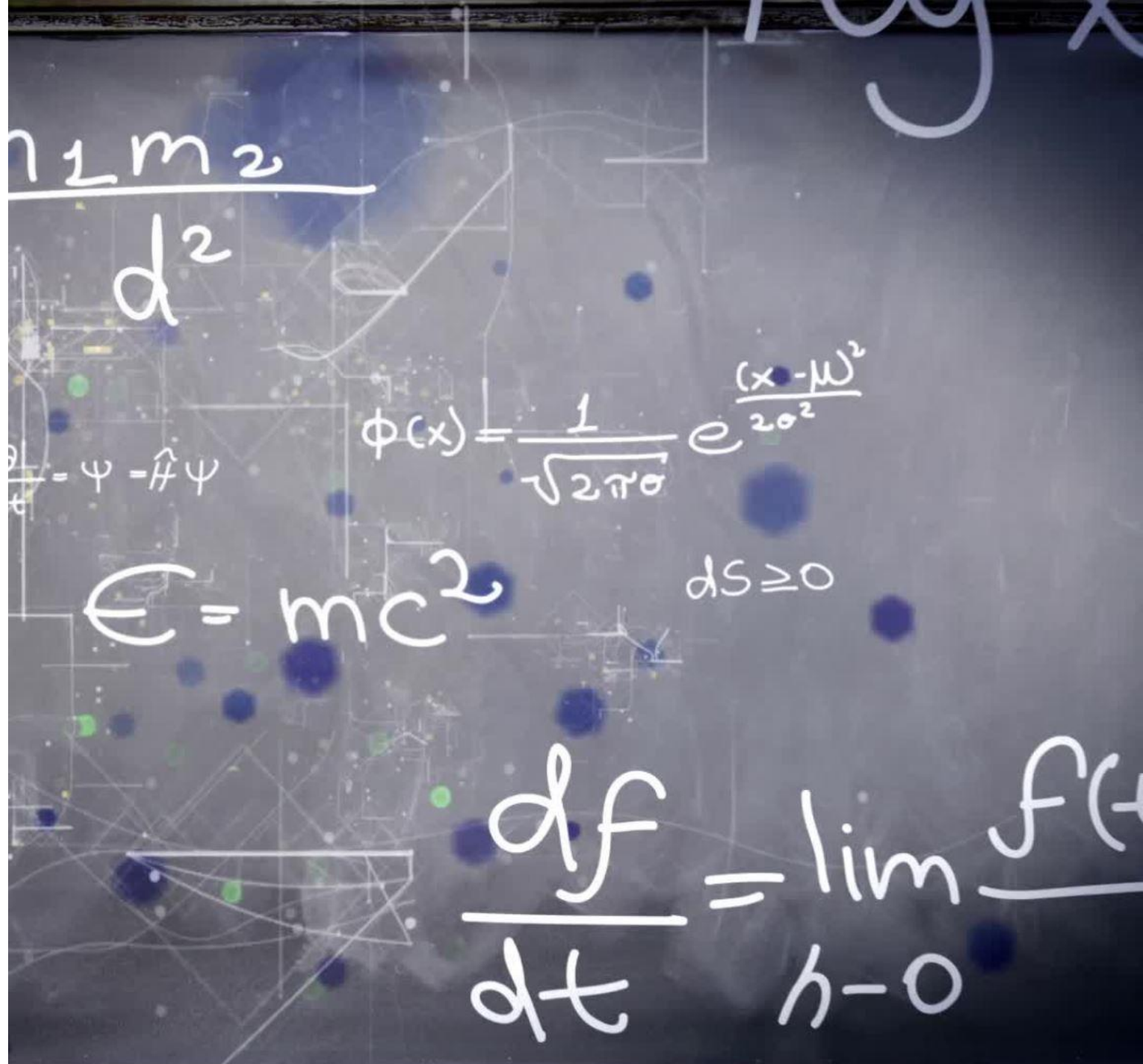
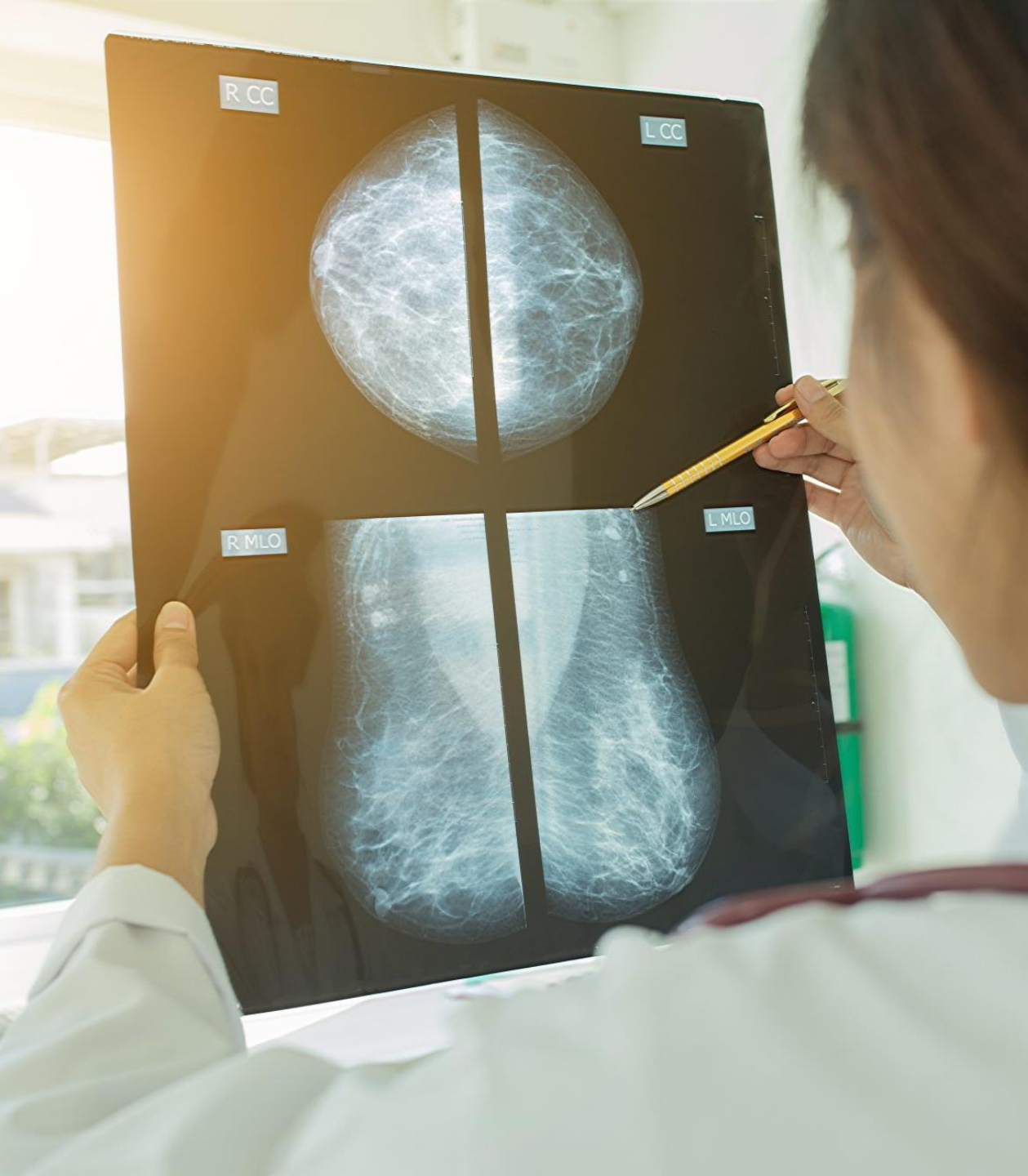

MAMMOGRAM IMAGE CLASSIFICATION THROUGH MACHINE LEARNING

Aarya Gupta, Julia Kwiek,
Sourya Potti





1. WHAT IS A MAMMOGRAM

- A **mammogram** is a specialized low-dose X-ray of the breast.
 - It is used to **screen for and detect breast cancer**, often before symptoms appear.
 - Helps identify **tumors, cysts, and calcium deposits (calcifications)**.
 - Can detect cancer **early**, when it is most treatable.
 - Recommended as part of **routine breast cancer screening**, especially for women over 40 or those at higher risk.
-

WHAT CHALLENGES DO DOCTORS FACE WITH MAMMOGRAMS

Masses and calcifications are harder to detect in certain cases.

Radiologists may miss subtle signs due to image complexity.

Mammograms are well-suited for machine learning analysis.

An accurate algorithm could flag missed abnormalities and help save lives.

WHY WE CHOSE THIS TOPIC

01

Breast cancer causes high mortality.

02

Applying technology to save lives is meaningful to us.

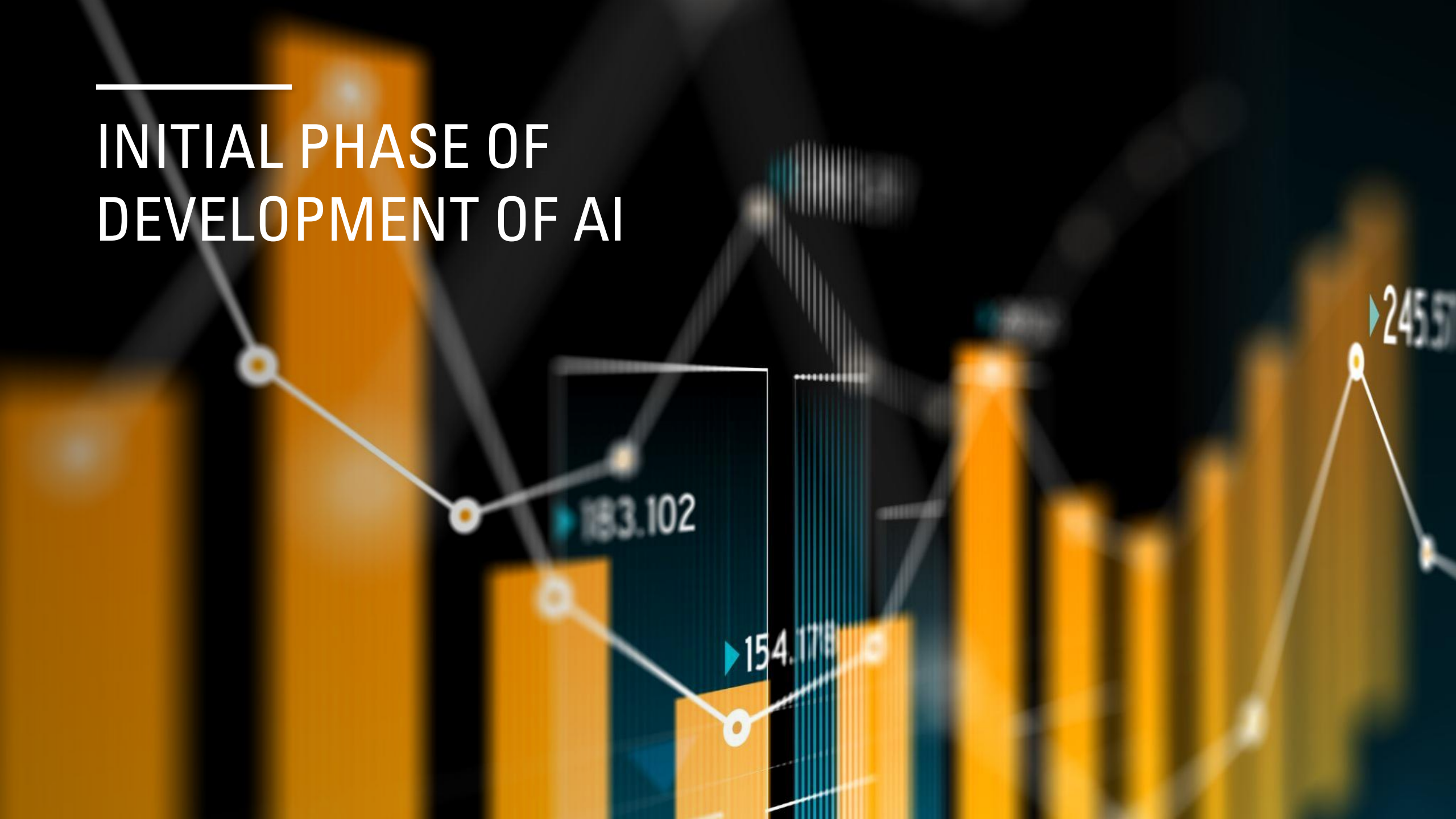
03

Machine learning can boost accuracy and reduce missed diagnoses.

04

This project blends social impact with technical innovation.

INITIAL PHASE OF DEVELOPMENT OF AI



DEVELOPMENT & REFINEMENT

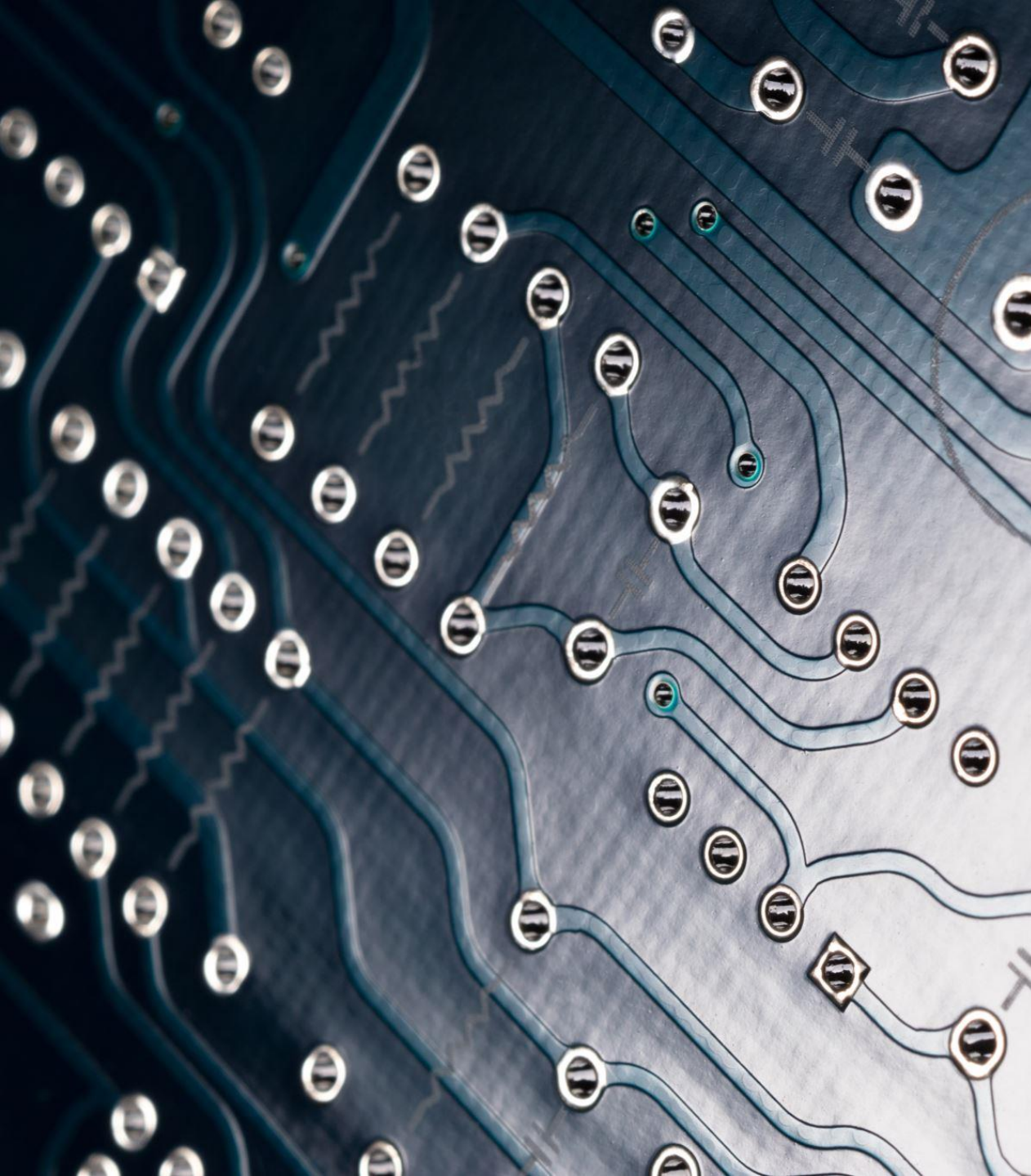
Began with a binary classifier: **no cancer** vs. **high likelihood of cancer**

Converted the model framework from **TensorFlow** to **PyTorch**

Initially built a **custom CNN** from scratch

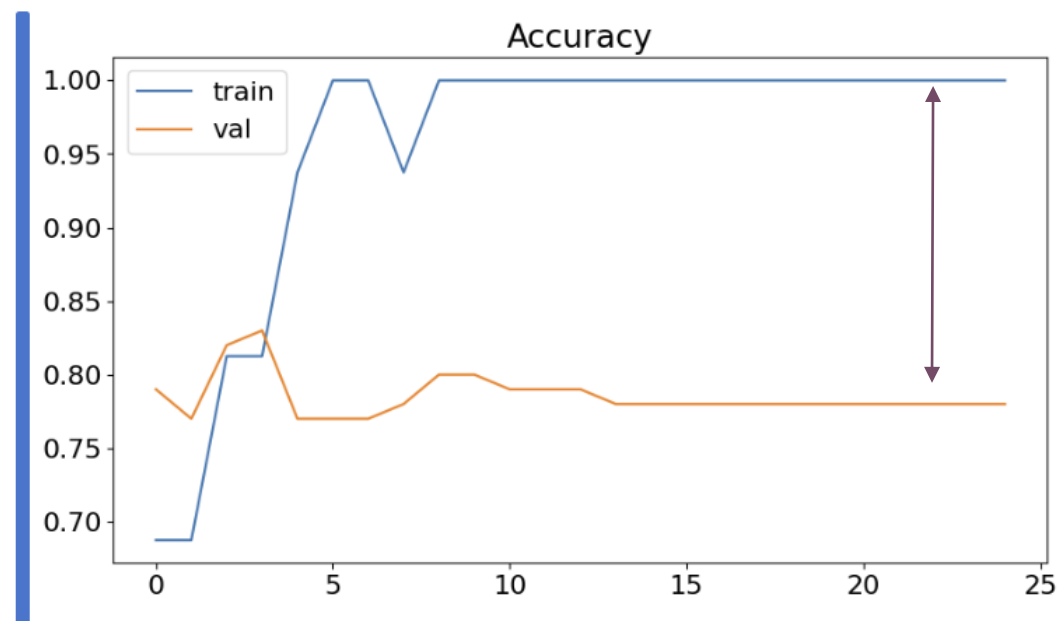
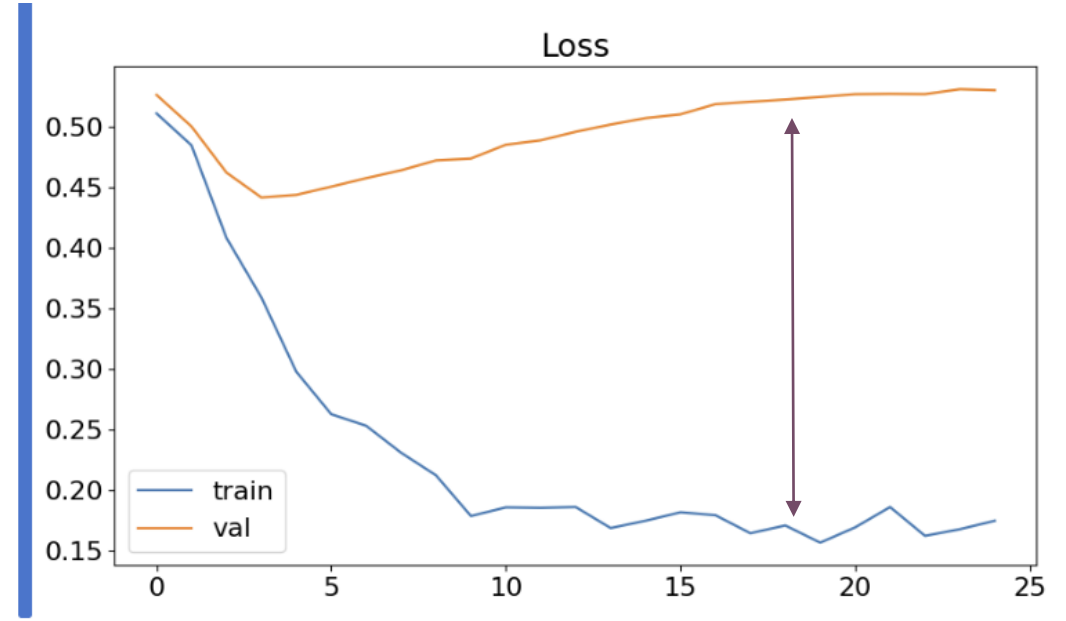
Later switched to **transfer learning** using **ResNet** pretrained on **ImageNet1k-V2**

Implemented changes to hyperparameters, such as implementing learning rate decay (decreasing learning rate overtime to continue learning)

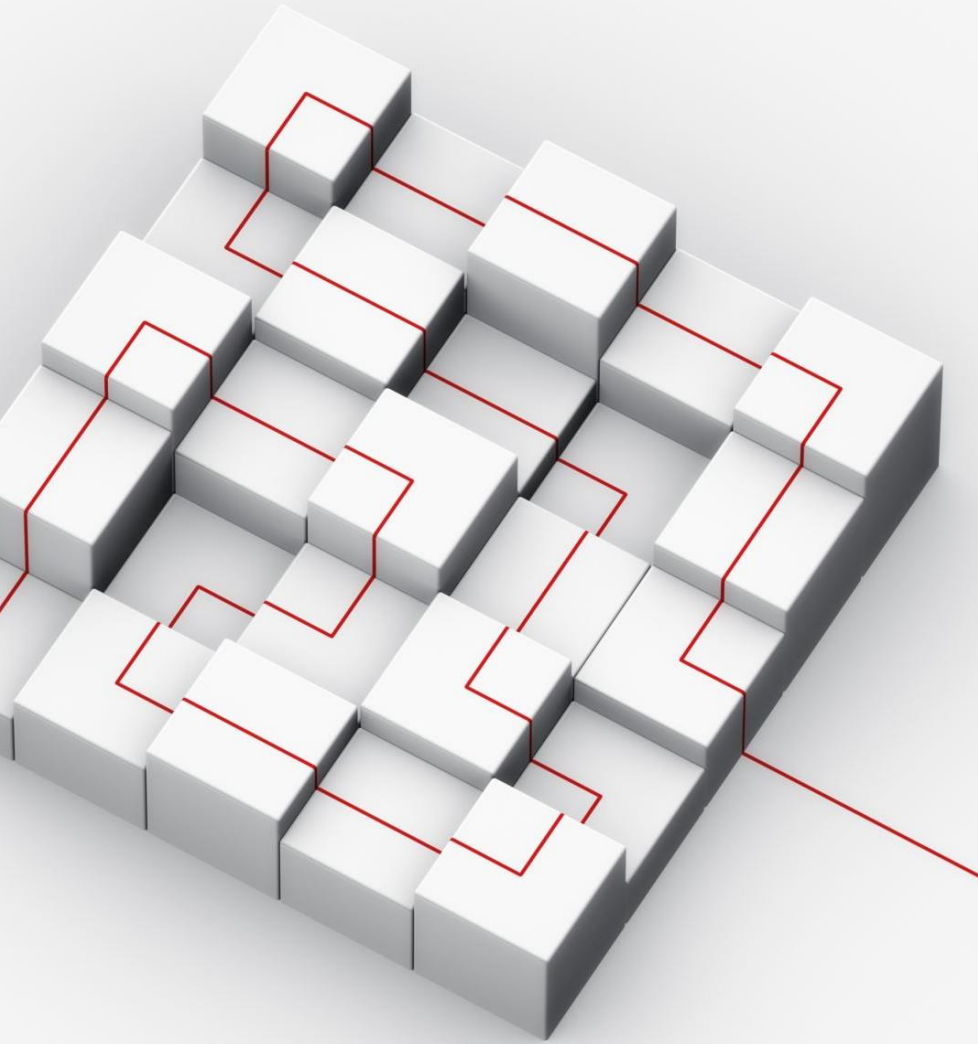


BINARY CLASSIFIER

- Built from scratch to analyze grayscale mammogram images
 - Uses **convolutional layers** to detect patterns in small image patches
 - **Pooling layers** reduce image size while keeping key features
 - Applies **batch normalization** and **dropout** to improve learning and prevent overfitting
 - Flattens the image and uses **fully connected layers** for final prediction
 - **Output:** A number between 0 and 1, where 1 = likely cancer, 0 = likely no cancer
-

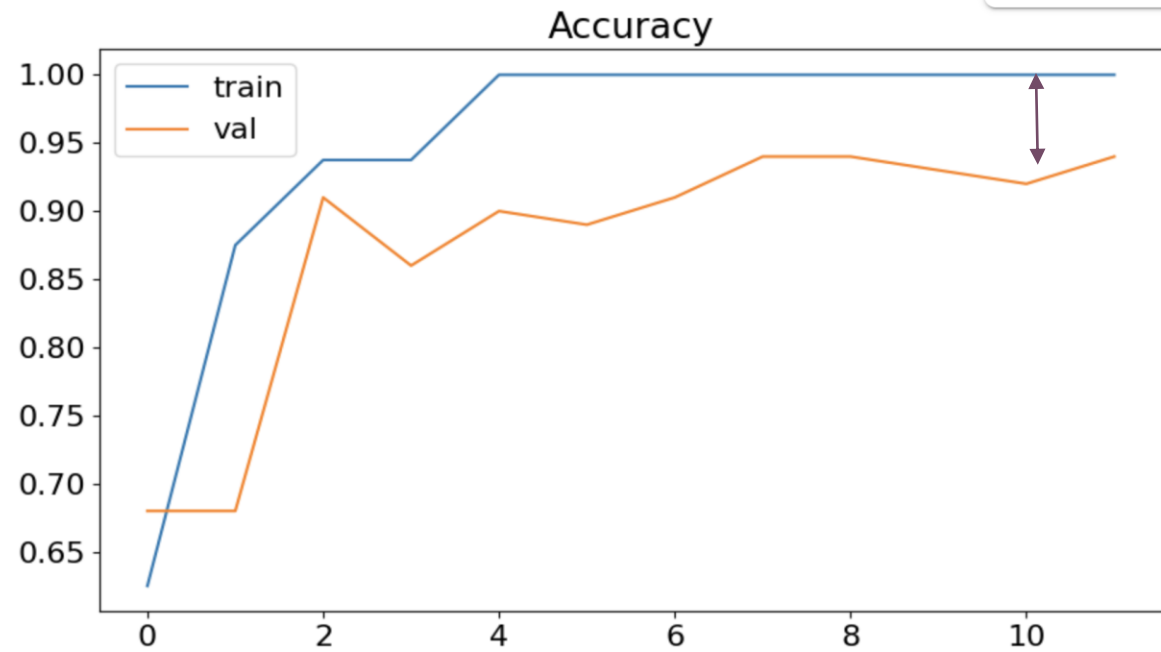
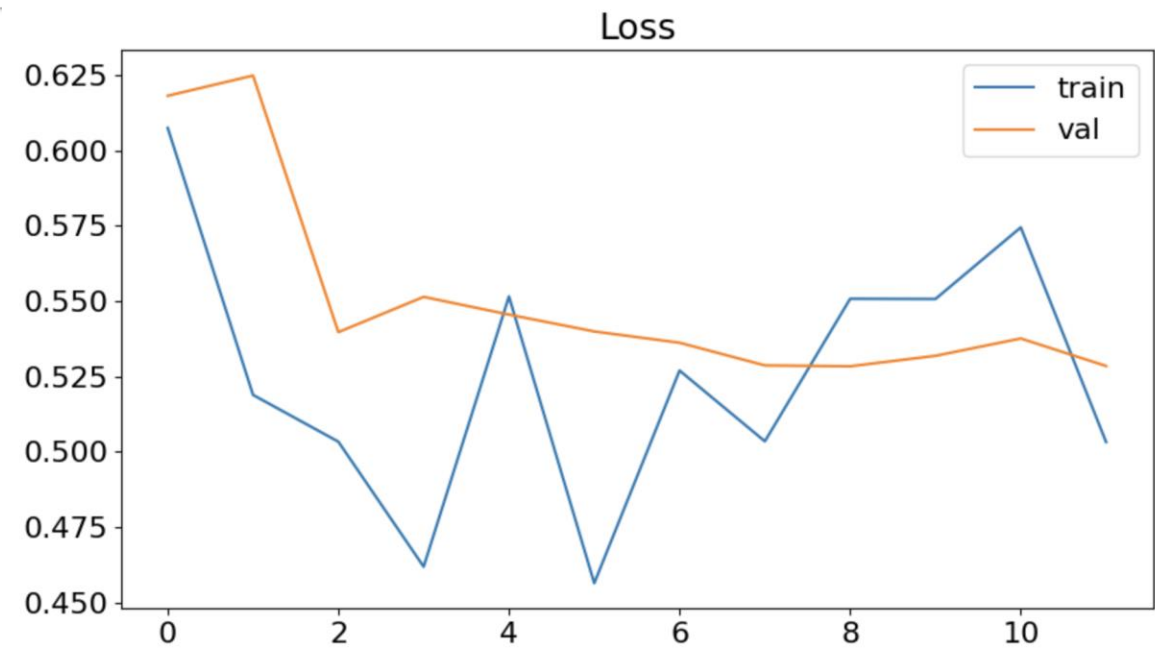


BINARY CLASSIFIER RESULTS



BINARY CLASSIFIER FINAL MODEL

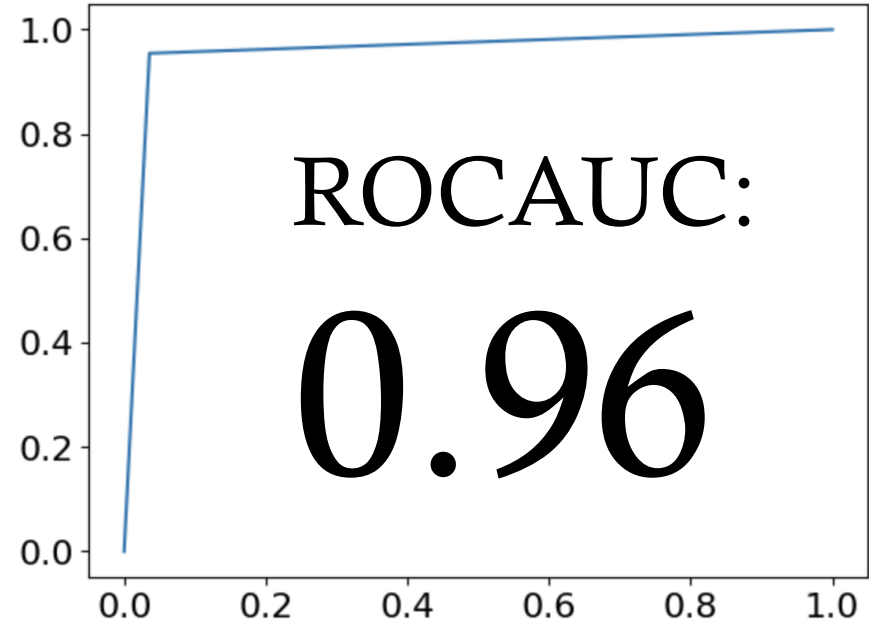
- Starts with a **pretrained base model** trained on millions of general images (recognizes basic features like **edges, textures, and shapes**)
- We **add custom layers** to help the model learn cancer-specific patterns
- **Resized all images** to reduce overfitting and computational cost
- Included a **dropout layer** to prevent overfitting
- **Manually tuned hyperparameters** (e.g., learning rate, batch size) for better performance
- **Performs well** even on **unseen** validation data, meaning that it **generalizes reliably** to new samples



BINARY CLASSIFIER RESULTS

BINARY CLASSIFIER FINAL RESULTS

Confusion Matrix Score: 87, ROCAUC: 0.959
[<matplotlib.lines.Line2D at 0x7de27e5a2e90>]



	Predicted Positive	Predicted Negative
True Positive	28	<u>0</u>
True Negative	2	20

Binary Confusion Matrix Scoring

Pred. Truth	0 (Neg)	1 (Pos)
0 (Neg)	+2	-3
1 (Pos)	-6	+2

score:
87

NECESSITY OF A MULTICLASS MODEL



A **binary model** only detects cancer vs. no cancer



But **not all abnormalities are malignant** — some are benign



Multiclass classification
allows us to distinguish
between:

No abnormality (0)
Benign Calcification (1)
Benign Mass (2)
Malignant Calcification (3)
Malignant Mass (4)



Helps provide **more detailed diagnostic support**



Enables doctors to **prioritize cases** and **plan appropriate care**



Reflects the **real complexity of mammogram interpretation**

MULTICLASS MODEL



This model builds on the **Binary Model**, expanding from **binary** to **multiclass classification** by identifying one of **five possible conditions**.



It uses a **pretrained base model** to extract general image features.

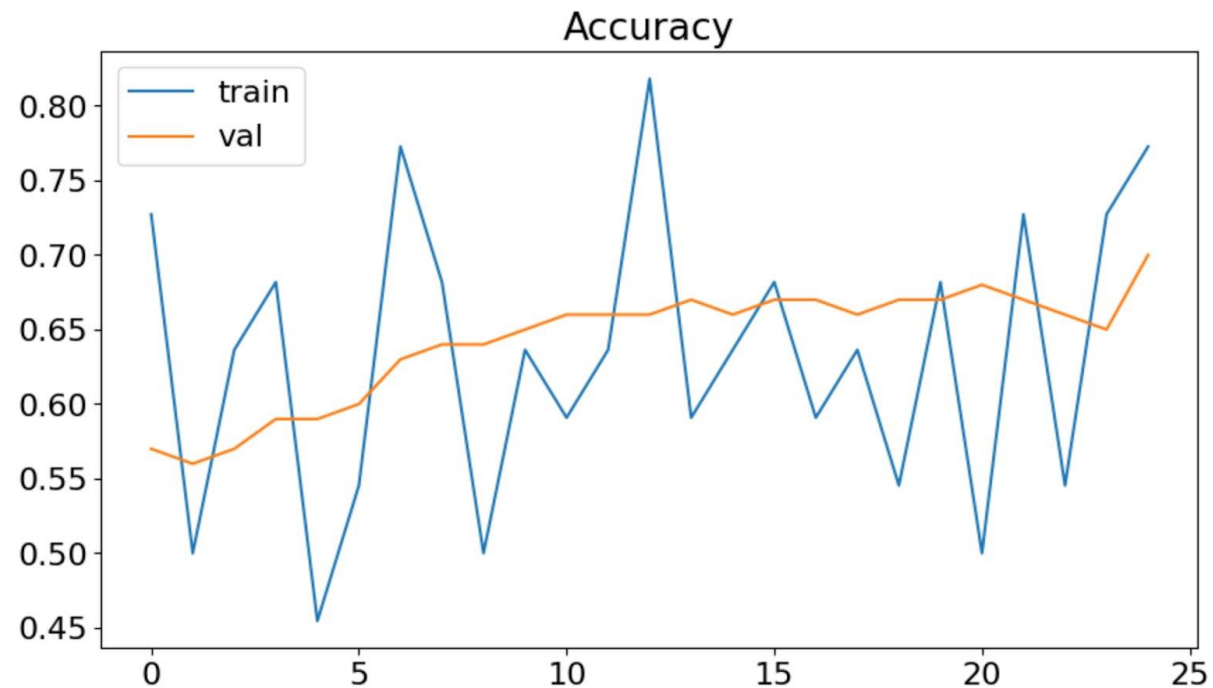


Custom layers are added to help the model learn distinctions between the five disease categories.



Output: Five **prediction scores** (one per condition); the **highest score** indicates the model's predicted diagnosis.

MULTICLASS RESULTS



Training Accuracy: 0.61	Validation Accuracy: 0.73
Training Accuracy: 0.74	Validation Accuracy: 0.79
Training Accuracy: 0.80	Validation Accuracy: 0.74
Training Accuracy: 0.88	Validation Accuracy: 0.73
Training Accuracy: 0.92	Validation Accuracy: 0.77
Training Accuracy: 0.95	Validation Accuracy: 0.76
Training Accuracy: 0.97	Validation Accuracy: 0.82
Training Accuracy: 0.98	Validation Accuracy: 0.78
Training Accuracy: 0.99	Validation Accuracy: 0.78
Training Accuracy: 0.98	Validation Accuracy: 0.75
Training Accuracy: 0.99	Validation Accuracy: 0.80
Training Accuracy: 0.99	Validation Accuracy: 0.76
Training Accuracy: 0.99	Validation Accuracy: 0.77
Training Accuracy: 0.99	Validation Accuracy: 0.77
Training Accuracy: 0.99	Validation Accuracy: 0.77
Training Accuracy: 0.99	Validation Accuracy: 0.77
Training Accuracy: 0.99	Validation Accuracy: 0.76
Training Accuracy: 1.00	Validation Accuracy: 0.78
Training Accuracy: 1.00	Validation Accuracy: 0.74
Training Accuracy: 0.99	Validation Accuracy: 0.76
Training Accuracy: 1.00	Validation Accuracy: 0.77
Training Accuracy: 1.00	Validation Accuracy: 0.75
Training Accuracy: 1.00	Validation Accuracy: 0.74
Training Accuracy: 1.00	Validation Accuracy: 0.79
Training Accuracy: 1.00	Validation Accuracy: 0.79
Training Accuracy: 1.00	Validation Accuracy: 0.79

MULTICLASS FINAL MODEL



Adjusted hyperparameters by implementing a variable learning rate schedule that reduces the learning rate whenever the loss function plateaus, helping the model converge more effectively.

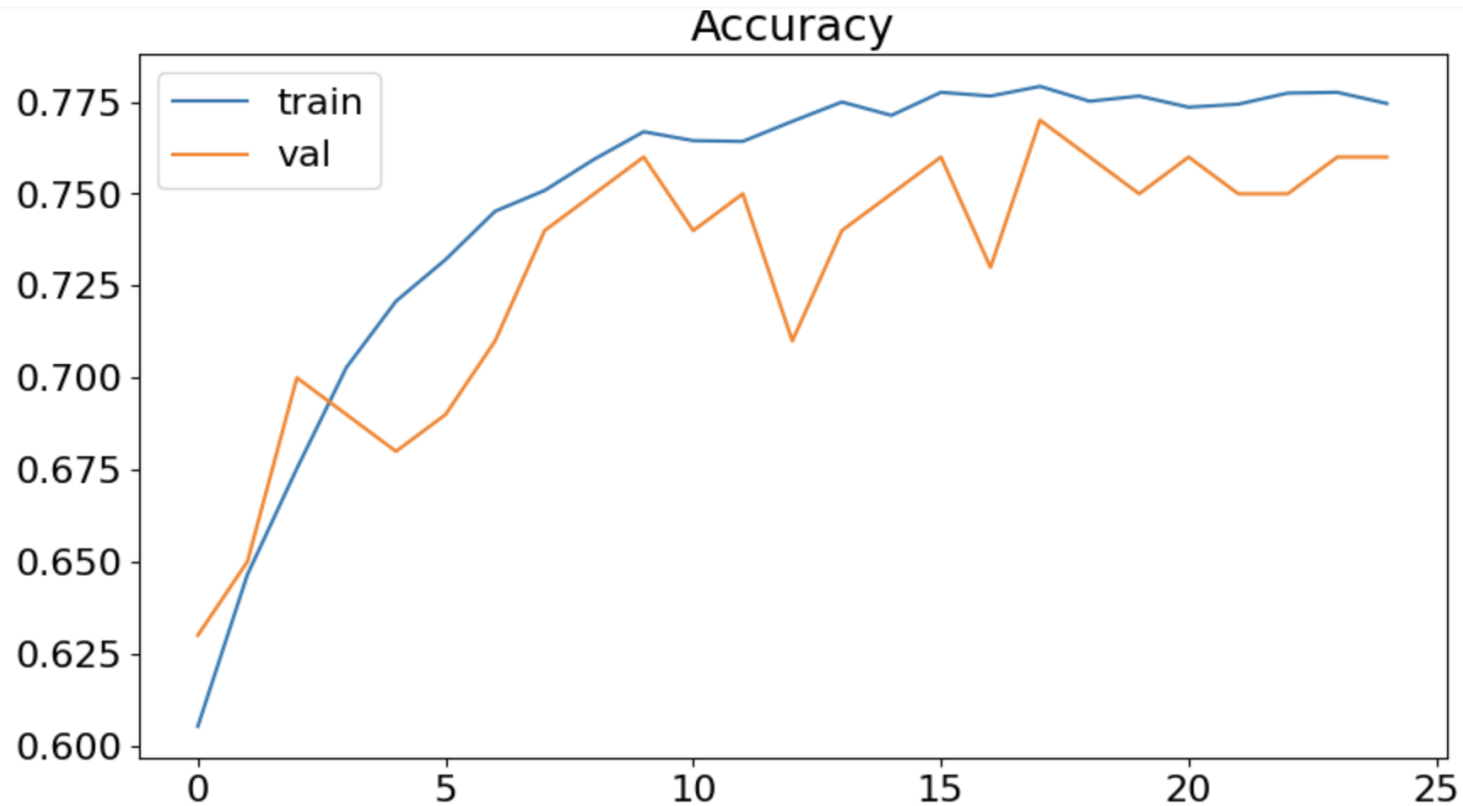


Resized images to 180×180 , choosing a higher resolution than the binary classification model (150×150) to capture finer details necessary for distinguishing between multiple classes.



Applied random horizontal flip as a data augmentation technique to introduce more variability into the training data, which helps reduce overfitting and improves the model's ability to generalize.

MULTICLASS FINAL RESULTS



CHALLENGES AND FUTURE IMPROVEMENTS

- Rented high-memory GPUs (L4, H100, A100) before transitioning to Student Colab Pro to speed up training.
- Memory constraints required downscaling images from $299 \times 299 \times 1$ to $150 \times 150 \times 1$ (binary) and 180×180 (multiclass).
- Multiclass model showed overfitting on validation data.
- Future work should explore larger datasets and more advanced data augmentation to improve generalization.

THANK YOU
