StarGPT (v2.0)

Lung Cancer Classification from Histopathological Images using Deep Transfer Learning

Youssef Negm, Omar Hossam

The Problem: Diagnosing Lung Cancer

- Leading cause of cancer deaths globally.
- Manual analysis of H&E slides: Time-consuming, prone to human error.
- Critical need for automated, rapid, accurate diagnostic support.

Early and accurate diagnosis significantly improves patient outcomes.

The Dataset: Histopathological Images

Source: Kaggle Lung Cancer Histopathological Images

Benign

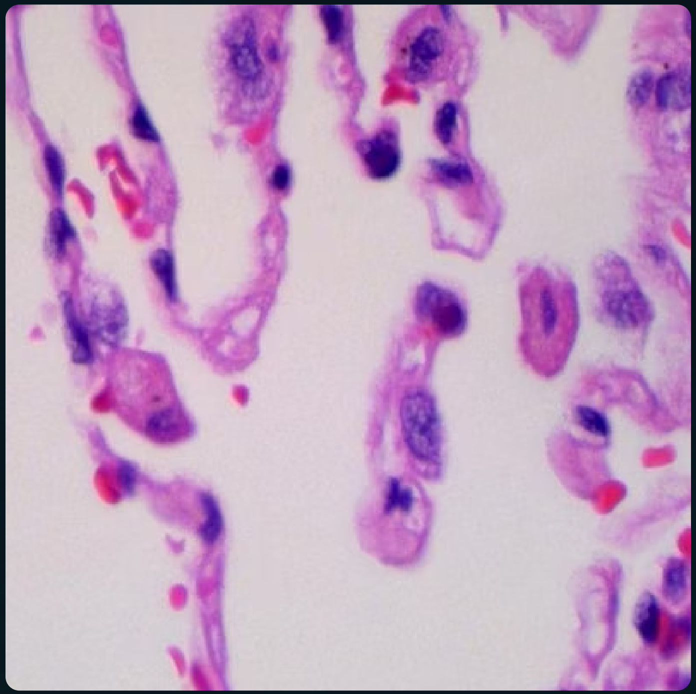
Healthy / Non-cancerous tissue.

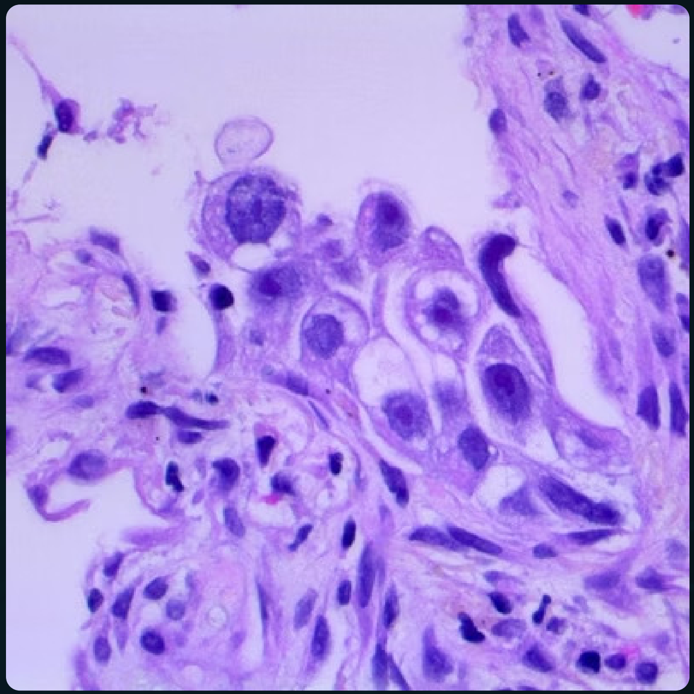
Adenocarcinoma

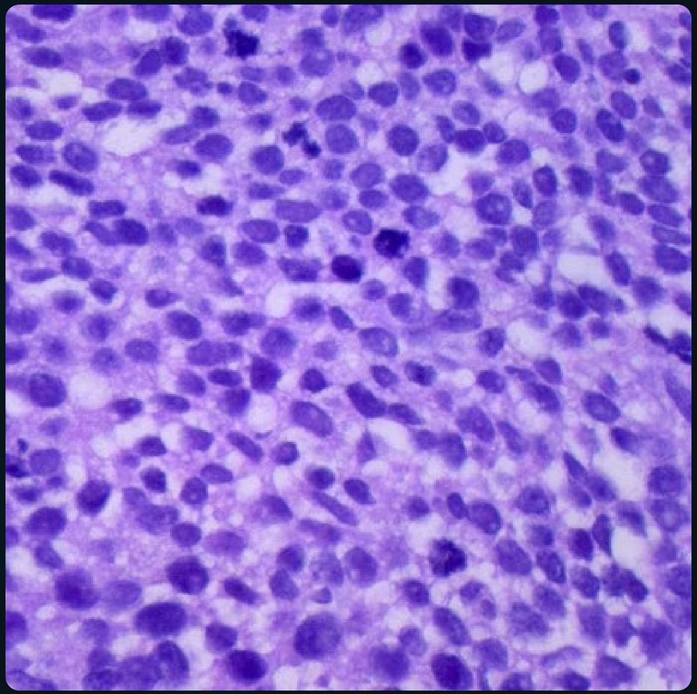
Malignant lung cancer type.

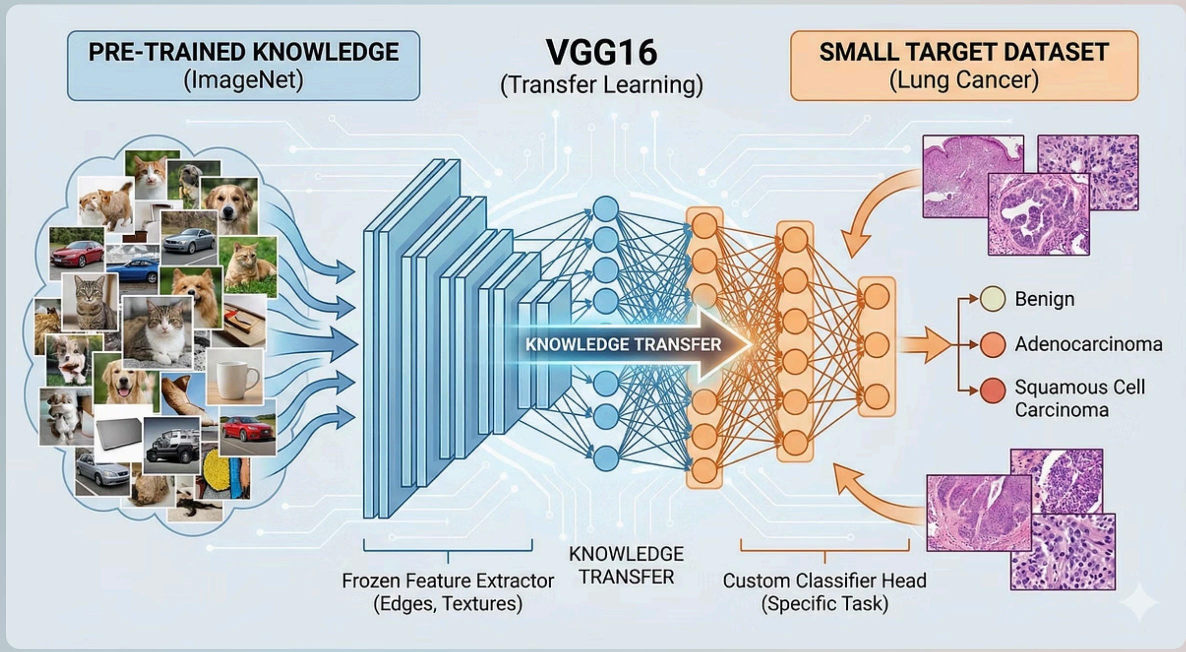
Squamous Cell Carcinoma

Another malignant lung cancer type.









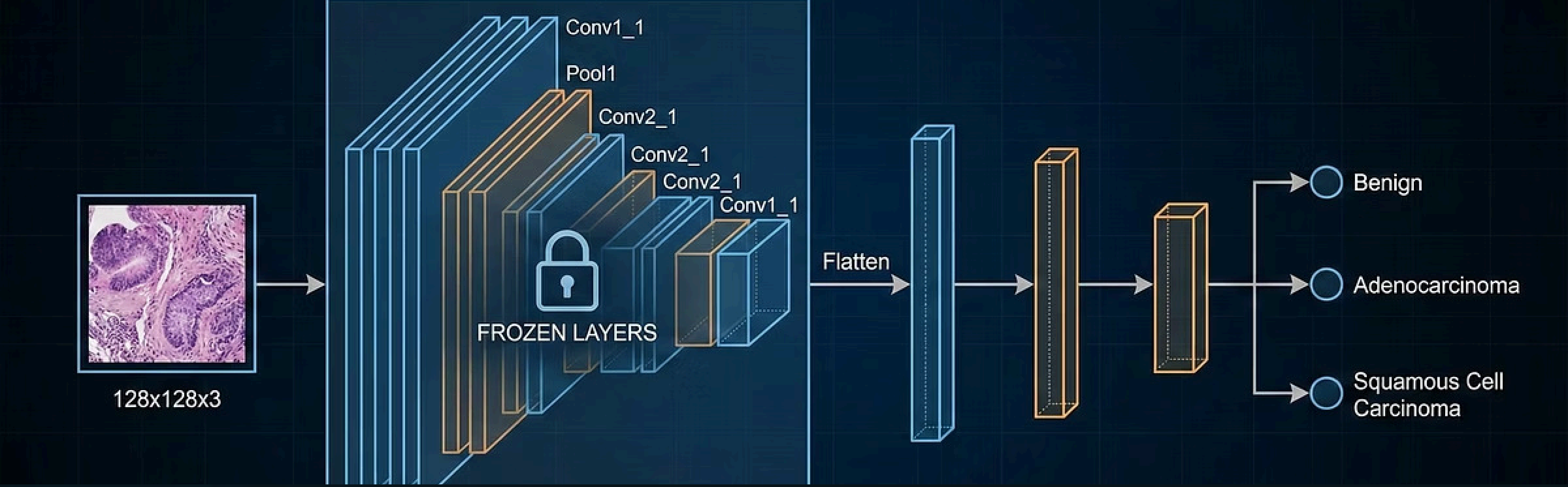
Methodology: Deep Transfer Learning

Why Transfer Learning?

- Small dataset limitations.
- Leverages pre-trained knowledge from vast datasets.
- Accelerates training, improves generalization.

Model Utilized: VGG16

- Visual Geometry Group (VGG) architecture.
- Pre-trained on ImageNet for complex feature extraction.
- Identifies edges, textures, and object parts.



VGG16 Model Architecture

- **Input:** 128x128 pixel RGB images.
- **Base:** Frozen VGG16 Convolutional Layers (Feature Extractor).
- **Head (Custom Classifier):**
  - Flatten Layer
  - Dense Layer (512 neurons, ReLU activation)
  - Dropout (0.5 for regularization)
  - Output Layer (3 neurons, Softmax activation)

Preprocessing & Training Strategy

Data Augmentation

Increased diversity and robustness:

- Zoom, Shear
- Width & Height Shift
- Horizontal Flip

Normalization

Standardized pixel values:

- Scaled to 1/255
- Ensures consistent input range

Optimization

Efficient model training:

- Adam Optimizer
- Categorical Crossentropy Loss

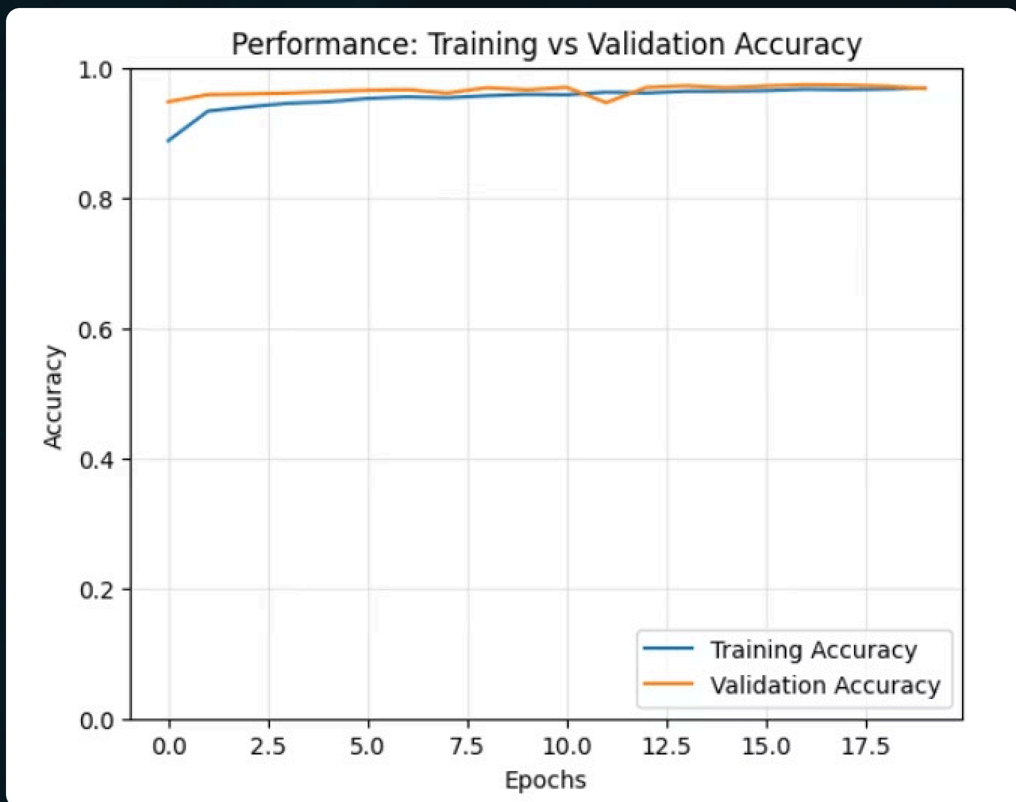
Early Stopping

Prevented overfitting:

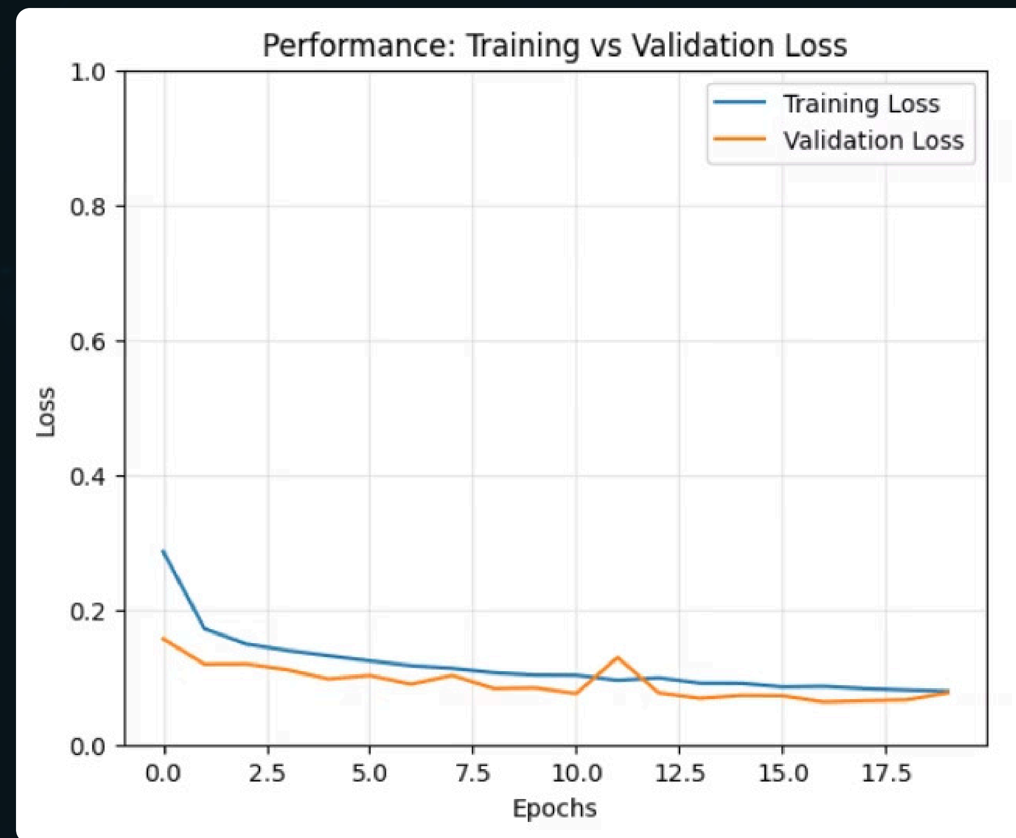
- Monitored validation loss
- Stopped training when performance plateaued

Performance Evaluation

Training vs. Validation Accuracy



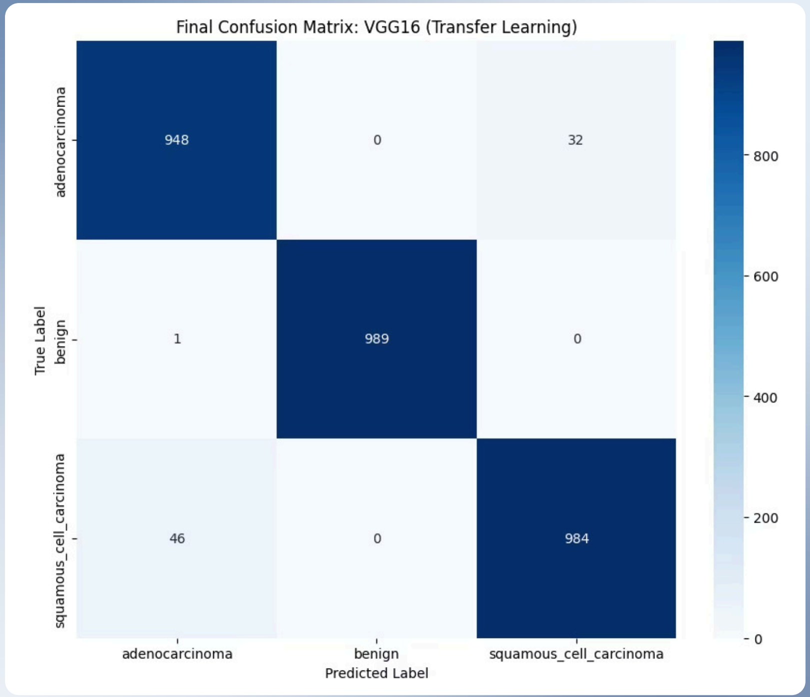
Training vs. Validation Loss



Final Testing Accuracy

(97.37%) --- VGG16 Classification Report ---

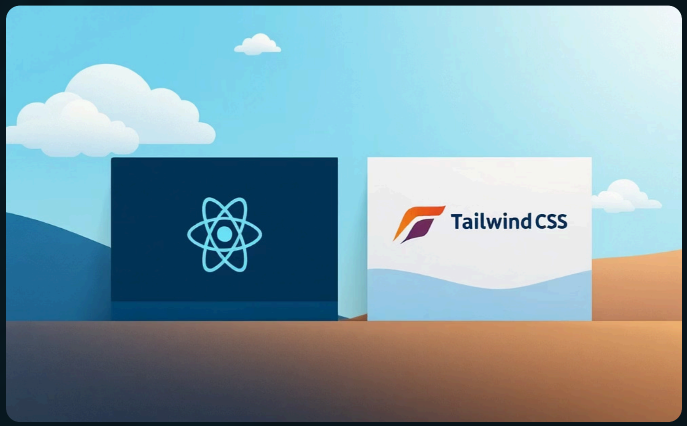
	precision	recall	f1-score	support
adenocarcinoma	0.95	0.97	0.96	980
benign	1.00	1.00	1.00	990
squamous_cell_carcinoma	0.97	0.96	0.96	1030
accuracy			0.97	3000
macro avg	0.97	0.97	0.97	3000
weighted avg	0.97	0.97	0.97	3000



Confusion Matrix Analysis

- Analysis of **False Positives** vs. **False Negatives**.
- Identified classes most prone to confusion.
- Typically, malignant types (Adenocarcinoma and Squamous Cell Carcinoma) can be challenging to differentiate.

System Implementation: The Web App

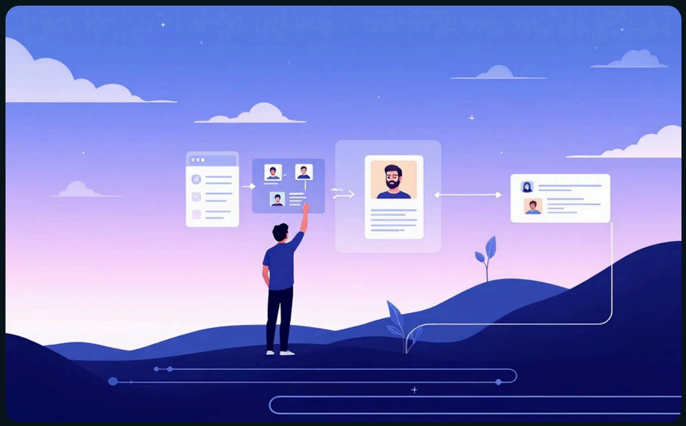


Frontend

- React.js + Vite
- Tailwind CSS
- Modern, Dark Mode UI

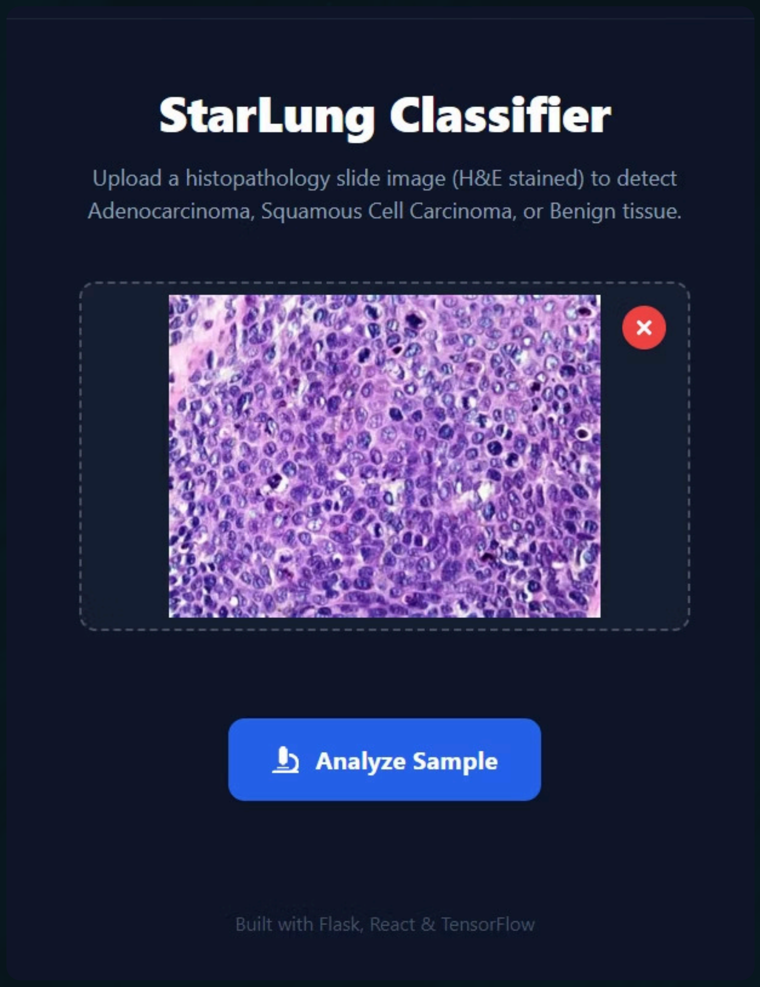
Backend

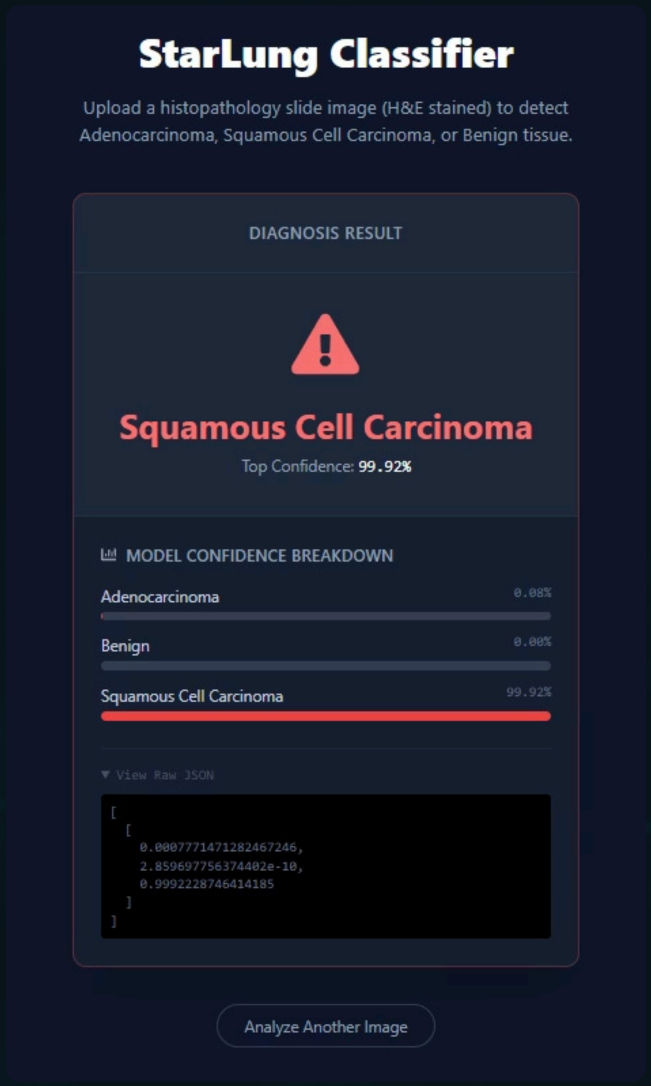
- Flask (Python)
- Serves VGG16 model via REST API



Workflow

- User uploads image.
- API processes image.
- Real-time prediction returned.





Conclusion & Future Work

Conclusion:

Successfully developed an end-to-end diagnostic tool for lung cancer classification with high accuracy, leveraging transfer learning and a robust web application.

Future Work:

- Integration of additional datasets for broader generalizability.
- Exploration of advanced transformer models for improved feature learning.
- Deployment of the tool in clinical settings for real-world validation.
- Development of explainable AI features to enhance trust and interpretability.