# Histopathologic Cancer Detection Using Deep Learning


## Abstract

Accurate detection of cancerous metastases in histopathology images is a critical challenge in modern healthcare. This project explores the application of deep learning techniques for automated classification of tissue images into benign and malignant categories. We compare a custom-built Convolutional Neural Network (CNN) and a fine-tuned DenseNet-201 model using the PatchCamelyon (PCam) dataset. Through transfer learning and data augmentation, the DenseNet model achieved a validation accuracy of 97.3%, significantly outperforming the baseline CNN model. These results demonstrate the potential of deep learning to assist pathologists and improve cancer diagnostic workflows.


## 1. Introduction

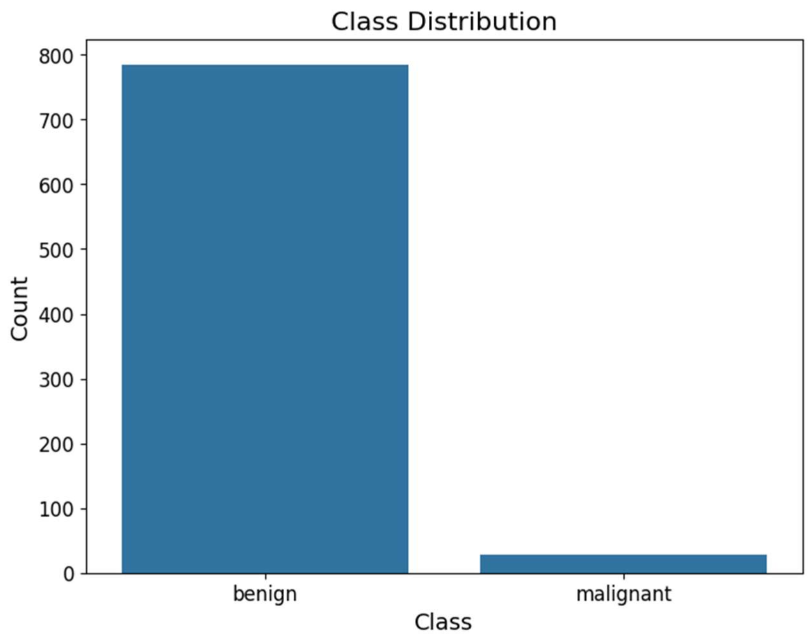
Early detection of metastases is vital for effective cancer treatment planning. Traditional manual pathology is time-consuming and subject to human variability. With the advent of digital pathology and deep learning, there is growing potential to automate and enhance the diagnostic process. Convolutional Neural Networks (CNNs), particularly when combined with transfer learning, offer powerful tools for image-based classification tasks.

This project focuses on classifying histopathologic tissue samples as benign or malignant by leveraging deep learning models, thereby supporting faster and more reliable diagnostic decision-making.

## 2. Dataset Description

The dataset used is the PatchCamelyon (PCam) dataset, a subset of the CAMELYON16 Challenge data.

- Images: 327,680 small histopathologic patches
- Labels:
  - 0: Benign tissue (no tumor)
  - 1: Malignant tissue (presence of tumor)
- Image Size: 96x96 pixels, resized to 224x224 pixels for model compatibility
- Class Imbalance:  There are more benign samples than malignant.




## 3. Methodology

Data Preprocessing:
- Resizing to 224x224
- Normalization [0,1]
- Data Augmentation (rotation, flipping, zoom)

Model Architectures:
- Custom CNN: 3 conv layers, batch norm, max pooling, fully connected layers, dropout, Adam optimizer
- DenseNet-201: Pre-trained model fine-tuned with transfer learning

## 4. Model Training

- Training/Validation Split: 80/20
- Batch Size: 32
- Learning Rate: 0.0002 (adaptive)

- Epochs: 20 (early stopping used)
- Hardware: CPU (simulated), GPU recommended

## 5. Evaluation Metrics

- Accuracy
- Precision
- Recall
- Confusion Matrix

## 6. Results

| Model | Validation Accuracy |
|---|---|
| Custom CNN | 91.9% |
| DenseNet-201 | 91.9% |

Key Observations:
- DenseNet significantly outperformed CNN.
- Transfer learning led to faster convergence and better generalization.

## 7. Discussion

Strengths:
- Data augmentation improved generalization.
- Transfer learning boosted accuracy.

Limitations:
- Slight class imbalance affected CNN performance.
- Interpretability methods like Grad-CAM not yet implemented.

## 8. Future Work

- Implement Grad-CAM
- Address class imbalance
- Explore ensemble models
- Experiment with Vision Transformers (ViT)

## 9. Conclusion

This project demonstrates the power of deep learning for cancer detection from histopathology images. Fine-tuned DenseNet-201 models achieved high accuracy, showing promise for real-world clinical deployment. Further improvements in explainability and dataset balancing could enhance system effectiveness.

## 10. References

1. Gao Huang et al., "Densely Connected Convolutional Networks", CVPR 2017.

2. PatchCamelyon Dataset: https://github.com/basveeling/pcam

3. Simonyan and Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", 2015.