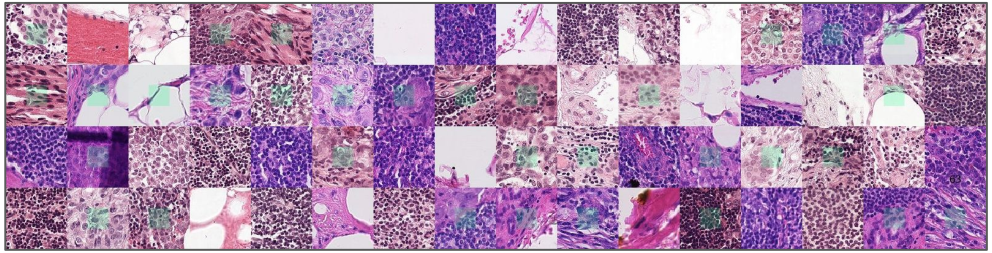
Deep Learning for Automated Metastatic Cancer Detection

Aayush, Bradon, Jason, Kadin, Om

Introduction and Background

- Metastatic cancer detection from pathology slides is crucial for early diagnosis and treatment
- Goal: Automate detection, improving speed and diagnostic accuracy for pathologists
- PatchCamelyon (PCam): Contains image patches from whole-slide images (WSIs) for binary classification aimed at detecting tumor tissue within a 32x32 pixel region



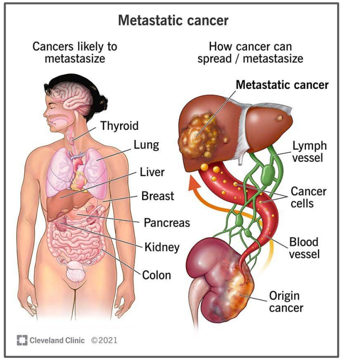
Example images of from PCam, green boxes indicate tumor tissue in center region

Dataset Details:

- Training samples: 262,144
- Validation & Test samples: 32,768 each
- Image size: 96x96 pixels
- Balanced sampling for high data quality

Problem Definition

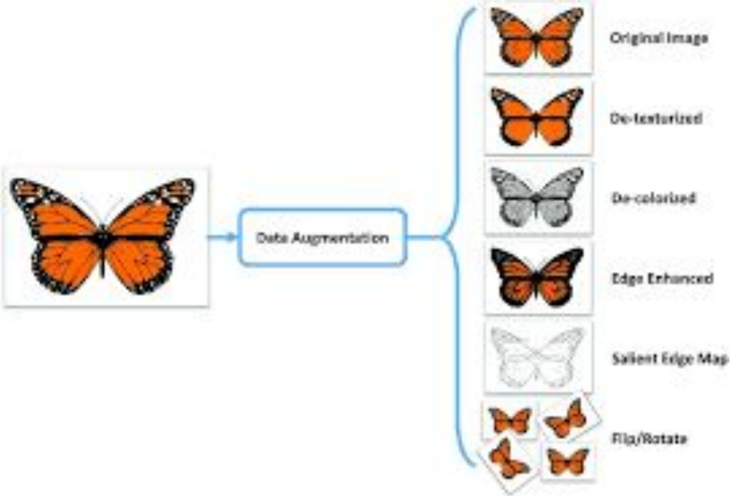
- Challenge: Manual identification of metastatic cancer in tissue samples is time-consuming and prone to errors, risking inaccurate diagnoses
- Solution: Automating cancer detection with machine learning can enhance diagnostic speed and accuracy, reducing human error
- Impact: The PCam dataset enables the application of ML techniques to improve real-world diagnostics and align with healthcare innovation trends



Displaying all the different cancers which can metastasize throughout the body

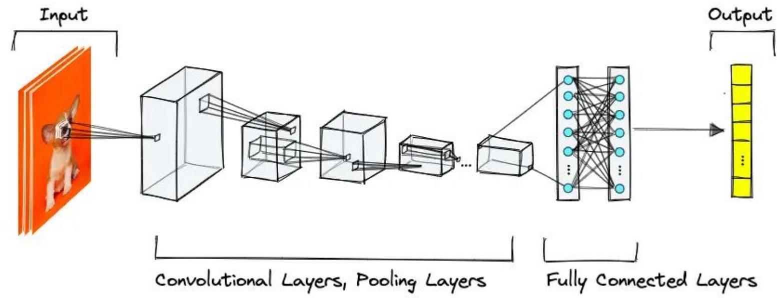
Methods (Data Preprocessing)

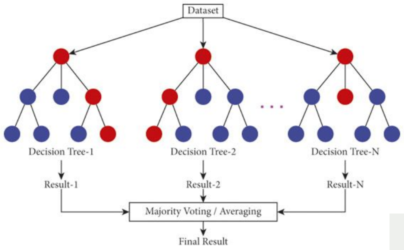
1. Image Resizing
    a. Concept: Resizing all images to a consistent size of 96x96 pixels
    b. Why: Uniform size simplifies the model and reduces computations
2. Normalization
    a. Concept: Standardize pixel values across all colors between 0 and 1
    b. Why: Causes model to treat all pixel values equally, speeding up convergence and improving accuracy
3. Data Augmentation
    a. Concept: Apply transformations (flips/rotations) to training images
    b. Why: Increases the variety of the dataset without collecting more data, adding to robustness and reducing overfitting

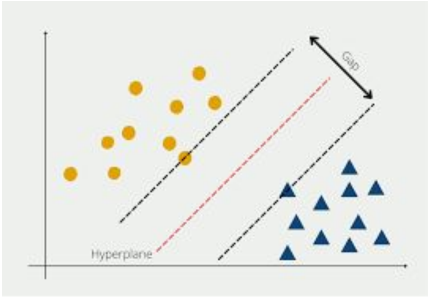


Methods (Supervised ML Algorithms)

1. Convolutional Neural Networks (CNNs)
   a. Concept: Deep learning architectures like ResNet and VGG16
   b. Why: CNNs are best for detecting patterns in images, ideal for tumor detection in the PCam dataset (Esteva)
2. Random Forest
   a. Concept: builds multiple decision trees to determine outcomes
   b. Why: A baseline for feature importance and model interpretability, but expected to perform less effectively than CNNs on images
3. Support Vector Machines (SVM)
   a. Concept: Designed for binary classification that creates a hyperplane to separate data points
   b. Why: Useful for comparison, though likely less effective in high-dimensional images compared to CNNs (Esteva)







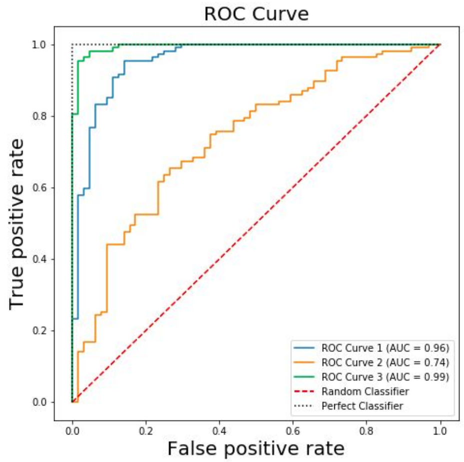
Results

Metrics:

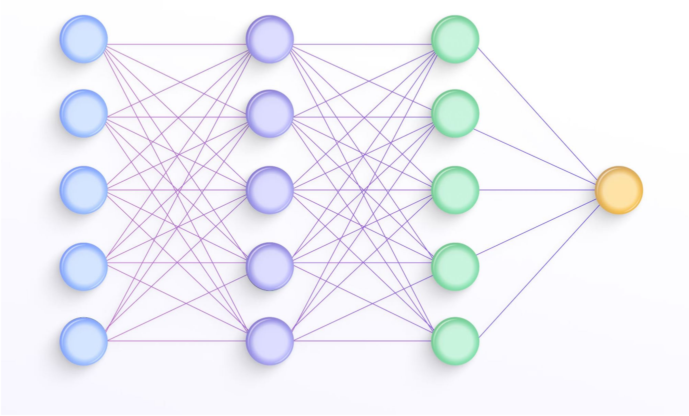
1. Accuracy - the percent of images classified correctly
2. Precision and Recall- ability of the model to identify true positives and avoid false negatives
3. AUC-RUC - visualize and measure the model's ability to differentiate between positive and negative classes. (Stokes)

Expected Results:

We expect CNNs to outperform traditional methods because of their effectiveness in image classification. Data augmentation will improve model generalization leading to improved quantitative metrics.



Example graph of AUC-RUC Curve

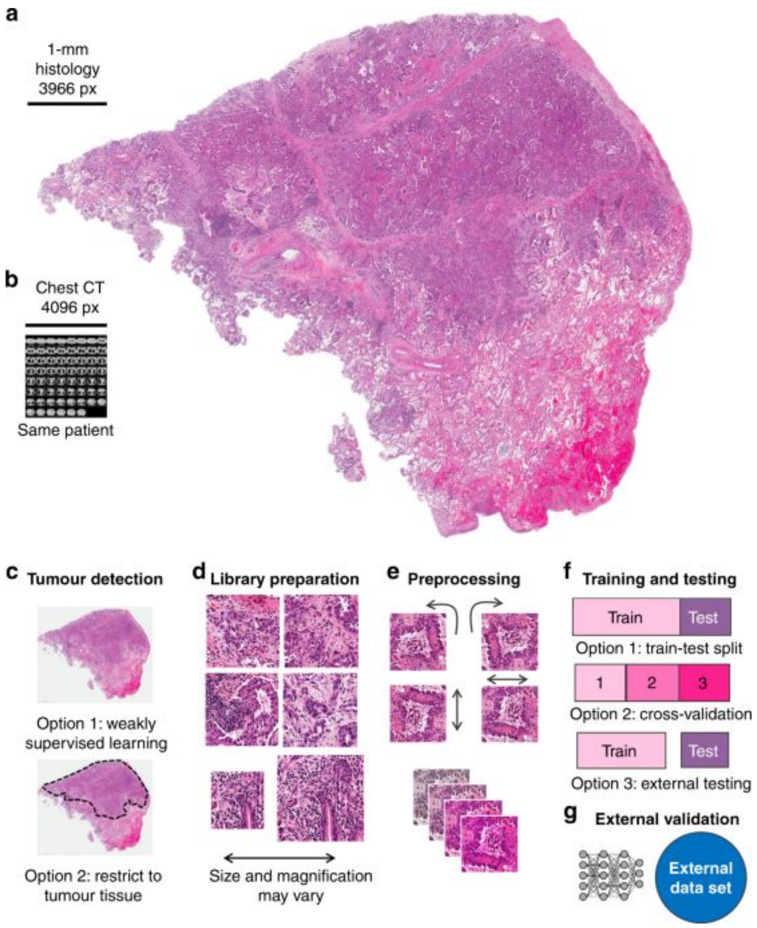


Example image of CNN

Discussion

Project Goals: We want to develop an efficient ML model that can accurately detect cancer tissues, outperforming traditional methods and reduce time and energy spent in manual interpretation

Ethical Consideration: We want to consider if the model is deployable in clinical settings without introducing implicit bias.



Example of goals, imaging, and process

References

1. H. C. Stokes, “Association of dietary cholesterol or egg consumption with incident cardiovascular disease and mortality,” JAMA, vol. 321, no. 11, pp. 1081-1095, Mar. 2019. (Online). Available: https://jamanetwork.com/journals/jama/fullarticle/2665774. (Accessed: Oct. 3, 2024).
2. A. Esteva, et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, pp. 115-118, Feb. 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28778026/. [Accessed: Oct. 3, 2024].
3. B. E. H. Bejnordi, et al., “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” Nature Medicine, vol. 24, pp. 1549-1558, Oct. 2018. [Online]. Available: https://www.nature.com/articles/s41591-018-0316-z. [Accessed: Oct. 3, 2024].