From Pixels to Prognosis: Deep Learning for Uterine Cancer Detection

Krushna Sanjay Sharma
sharma.krus@northeastern.edu
Northeastern University
Seattle, WA, USA

Nandini Bommireddy
bommireddy.n@northeastern.edu
Northeastern University
Seattle, WA, USA

Zhizhou Gu
gu.zhiz@northeastern.edu
Northeastern University
Seattle, WA, USA

ABSTRACT

Histopathological analysis remains the cornerstone of cancer diagnosis, particularly in uterine malignancies. Manual inspection of slides is time-consuming and subject to variability. This report introduces an automated deep learning pipeline using four model architectures—ViT, ResNet34, EfficientNet-B0, and DenseNet121—to classify histology images as tumor or normal. We detail preprocessing, augmentation, and comprehensive evaluations, highlighting architectural strengths and limitations through detailed comparisons.

1 INTRODUCTION

Uterine cancer, particularly endometrial carcinoma, significantly impacts women's health globally. Histopathological examination, involving manual scrutiny of tissue slides, remains the gold standard despite inherent variability and labor intensity. Recent advancements in deep learning, notably convolutional neural networks (CNNs) [1] and transformer-based architectures [2], offer promising avenues to automate and improve this diagnostic process.

This project employs CNN and transformer models to classify histology patches extracted from whole-slide images, focusing on robust data preprocessing and balanced training to ensure fair model comparison.

2 DATASET AND PREPROCESSING

Data were sourced from The Cancer Imaging Archive (TCIA) with JSON labels categorizing slides into tumor and normal. Rigorous preprocessing was essential to transform gigapixel whole slide images into manageable, informative patches suitable for deep learning models.

2.1 Whole-Slide Image Representation

Histopathology slides are digitized at extremely high resolutions, capturing intricate tissue structures necessary for accurate cancer diagnosis. As shown in Figure 1, these whole-slide images are typically stored in a multi-scale pyramid format, where each level corresponds to a different magnification (e.g., 5×, 10×, 20×, and 40×). The left panel of the figure illustrates this pyramid structure, with each layer representing the same tissue section at increasing levels of detail and pixel dimensions. The right panel displays a representative hematoxylin and eosin (H&E) stained uterine tissue section at high resolution.

This multi-scale organization enables both pathologists and automated algorithms to efficiently navigate and analyze slides, from low-magnification overviews for region selection to high-magnification views for cellular and subcellular feature assessment. In our pipeline, we leverage this structure to extract informative patches at appropriate magnifications, ensuring both contextual and morphological features are captured for robust model training.

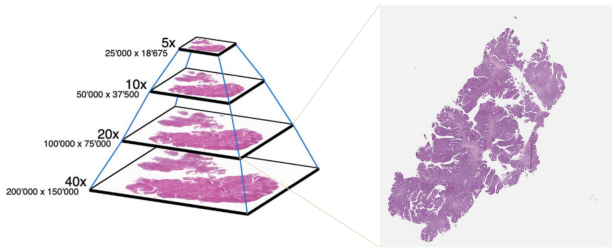

Figure 1: Multi-scale pyramid representation of a histopathology slide showing increasing resolution and detail from 5× (25,000 × 18,675 pixels) to 40× (200,000 × 150,000 pixels)

2.2 Data Preprocessing Pipeline

Our preprocessing pipeline, illustrated in Figure 2, transforms gigapixel whole slide images into manageable, informative patches through a systematic approach:

2.2.1 Optimal Resolution Selection. The first step involves selecting the appropriate magnification level from the slide pyramid. We determine the optimal resolution that balances detail preservation with computational efficiency, typically corresponding to 20× magnification.

2.2.2 Patch Extraction. From the selected resolution level, we extract random 224×224 pixel patches. This dimension aligns with the input requirements of standard deep learning architectures while capturing sufficient contextual information for accurate classification.

2.2.3 Quality Filtering. To ensure only diagnostically relevant tissue sections are included, we implement a multi-stage filtering process:

- **White Area Threshold:** Patches containing more than 30% white pixels (background) are excluded.
- **Grey Area Threshold:** Patches with more than 40% grey pixels (often indicating poorly stained or out-of-focus regions) are removed.
- **Tissue Content Requirement:** Only patches with more than 2% tissue content (determined through texture and color saturation analysis) are retained.

This filtering strategy significantly improves the signal-to-noise ratio in our dataset by ensuring models train on diagnostically relevant tissue regions rather than background or artifacts.

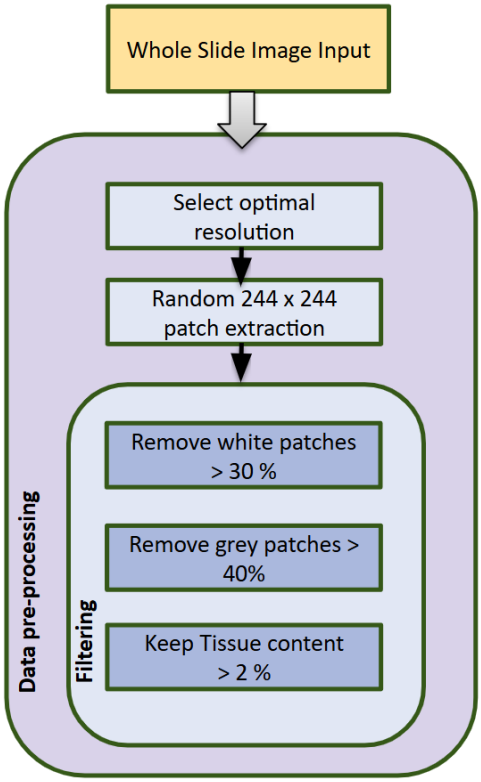

Figure 2: Data preprocessing workflow showing the filtering criteria for patch selection

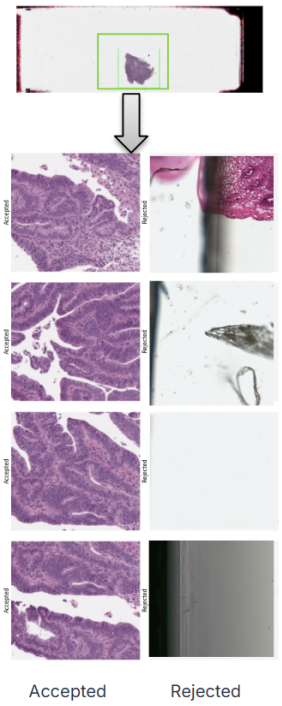

Figure 3: Representative examples of extracted patches after preprocessing and filtering. Left column: accepted patches with sufficient tissue content; right column: rejected patches due to excessive background, artifacts, or insufficient tissue.

2.3 Resulting Patch Dataset

The extracted and filtered patches, as exemplified in Figure 3, form our training and validation dataset. The left column shows representative tumor patches, characterized by irregular cellular architecture, nuclear atypia, and disrupted tissue patterns. The right column displays normal endometrial tissue patches with regular glandular structures and uniform cellular distribution.

While our patch extraction methodology substantially increased the available sample size for training, enhancing the model's ability to learn diverse tissue morphologies, it simultaneously amplified the inherent class imbalance present in the original slide-level data. This occurred because tumor slides typically yield more viable tissue patches than normal slides, due to their higher cellular density and architectural complexity.

This preprocessing approach resulted in a curated dataset with significant class imbalance: approximately 90,000 tumor patches versus 35,000 normal patches, reflecting the natural prevalence in our collected slides. We address this imbalance during model training through appropriate weighting and augmentation strategies.

3 TRAINING PIPELINE AND AUGMENTATION

Our training pipeline employs a systematic sequence of steps to ensure robust model performance and balanced learning across the uterine cancer histopathology dataset. The overall workflow is summarized in Figure 4, which details each stage from patch validation to final model checkpointing.

3.1 Class Balancing

Class imbalance is a significant challenge in medical imaging tasks, and in our dataset, the number of tumor patches (90,000) substantially exceeds the number of normal patches (35,000). While our patch extraction strategy increased the total sample size and improved the diversity of training data, it also amplified the inherent class imbalance present at the slide level.

To address this, we apply class weighting during model training. The weight for each class is calculated using the inverse frequency method:

$$w_j = \frac{N}{2 \cdot n_j} \tag{1}$$

where w_j is the weight for class j, N is the total number of samples, and n_j is the number of samples in class j. This ensures that the minority class contributes proportionally more to the loss, counteracting the bias towards the majority class.

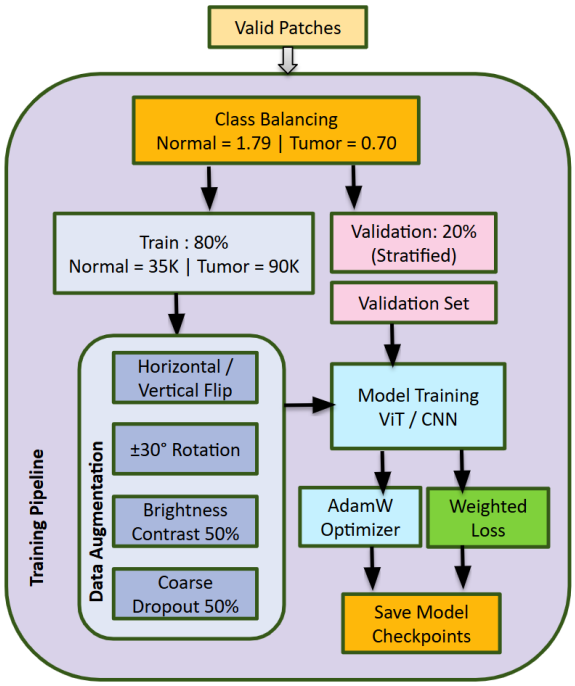

Figure 4: Training pipeline

For our dataset, this results in the following weights:

$$w_{\text{normal}} = \frac{125,000}{2 \times 35,000} \approx 1.79 \tag{2}$$

$$w_{\text{tumor}} = \frac{125,000}{2 \times 90,000} \approx 0.70 \tag{3}$$

These weights are incorporated into the loss function during training, as shown in the training pipeline (Figure 4), ensuring that the model learns to accurately classify both normal and tumor patches despite the imbalance. This class balancing step is critical for achieving robust and fair performance across both classes.

3.2 Dataset Split

The dataset is split into training and validation sets using a stratified approach, ensuring that both sets contain the same proportion of normal and tumor patches. The training set consists of 80% of the data (35K normal patches and 90K tumor patches), while the validation set comprises 20% (stratified) of the dataset.

3.3 Data Augmentation

Data augmentation is employed to increase the diversity of the training set and reduce overfitting. The following transformations are applied to the images:

- **Horizontal/Vertical Flip**: Random flipping of images helps the model learn spatial invariances.
- **Rotation**: Random rotations of ±30° provide additional perspective variations.
- **Brightness/Contrast Adjustment (50%)**: Adjusting the brightness and contrast of images to simulate different lighting conditions in real-world images.
- **Coarse Dropout (50%)**: Random dropout of image patches to encourage the model to focus on different parts of the image, improving generalization.

3.4 Model Training

The model training involves training Vision Transformer (ViT) and CNN architectures with a combination of the following:

- **Optimizer**: The AdamW optimizer is used due to its ability to handle sparse gradients and apply weight decay more effectively than traditional Adam. AdamW separates the L2 regularization from the gradient update step, ensuring proper regularization regardless of gradient magnitude. This results in better generalization, particularly important for histopathology datasets where small, discriminative features can be easily overlooked.
- **Weighted Loss Function**: We integrate the class weights calculated during the class balancing step directly into the cross-entropy loss function:

$$\mathcal{L}_{\text{weighted}} = -\sum_{i=1}^{n} w_{y_i} \log\left(\frac{e^{z_{y_i}}}{\sum_{j=1}^{C} e^{z_j}}\right) \tag{4}$$

Where w_{y_i} is the weight for the true class of sample i, ensuring that errors on normal patches contribute 1.79 times more to the loss than they would in standard cross-entropy, while errors on tumor patches contribute 0.70 times as much.

The standard cross-entropy loss is formulated as:

$$\mathcal{L} = -\log\left(\frac{e^{z_y}}{\sum_{j=1}^{C} e^{z_j}}\right) = -z_y + \log\left(\sum_{j=1}^{C} e^{z_j}\right) \tag{5}$$

3.5 Model Checkpoints

To prevent overfitting and save the best-performing models, checkpoints are saved during training. This allows the best model to be retrieved based on validation performance and ensures that the training process can be resumed if interrupted.

3.6 Final Model Evaluation

After training, the model is evaluated on the validation set to assess its generalization performance. Performance metrics such as accuracy, precision, recall, F1-score, AUROC, and log loss are used to summarize the model's effectiveness. Additionally, confusion matrices and ROC curves are generated to provide further insights into the model's performance on both classes (tumor and normal).

4 MODEL ARCHITECTURES AND COMPARATIVE ANALYSIS

Building upon our preprocessing pipeline and class-balanced training approach, we evaluated four distinct deep learning architectures, each representing different design philosophies in computer vision. The selection includes both convolutional neural networks (CNNs) and transformer-based models to comprehensively assess their applicability to histopathological analysis of uterine cancer.

4.1 Vision Transformer (ViT)

Vision Transformers represent a paradigm shift from convolutional approaches by dividing input images into fixed-size patches (typically 16×16), which are linearly embedded and treated as tokens in transformer encoders [2]. Each patch undergoes positional encoding to preserve spatial information, followed by self-attention mechanisms that capture global dependencies and multilayer perceptron (MLP) processing. This architecture excels at modeling long-range dependencies between image regions, which theoretically could benefit analysis of tissue architecture in histopathology.

However, lacking the spatial inductive biases intrinsic to CNNs, ViTs typically require extensive datasets (14M+ images) to achieve high performance, potentially limiting their effectiveness on specialized medical datasets like ours [10]. Recent studies have shown that while ViTs can effectively localize cellular structures without explicit supervision, their performance significantly degrades when training data is limited [6].

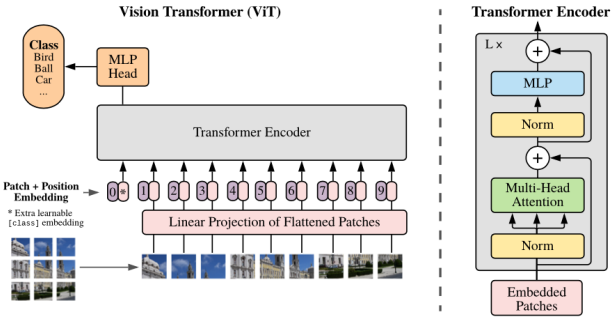

Figure 5: Vision Transformer Architecture: Input images are divided into fixed-size patches, which are flattened, linearly embedded with position encodings, and processed through transformer encoder blocks utilizing multi-head self-attention mechanisms.

4.2 ResNet34

ResNet34 introduces a powerful solution to the vanishing gradient problem through 34 convolutional layers with skip (residual) connections that facilitate gradient flow during backpropagation [1]. These residual connections allow the network to learn identity mappings, effectively mitigating vanishing gradients and enabling much deeper architectures than previously possible. Its balanced depth-to-performance ratio makes it particularly suited for histopathological analysis, where fine-grained feature extraction is crucial.

Studies specifically examining ResNet architectures for cancer histopathology have demonstrated that the 34-layer variant often achieves optimal performance balance, with an AUC of 0.992 on large histopathology datasets—outperforming both shallower and deeper variants [9]. The architecture's capacity to capture hierarchical features maps well to the multi-scale nature of histopathological patterns, from cellular details to tissue-level organization.

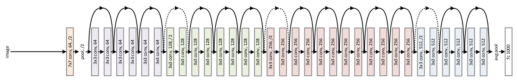

Figure 6: ResNet-34 Architecture: The network features 34 layers organized into residual blocks that enable direct gradient flow through skip connections, effectively addressing the vanishing gradient problem common in deep networks.

4.3 EfficientNet-B0

EfficientNet-B0 revolutionizes neural architecture design by employing compound scaling that uniformly scales network depth, width, and resolution dimensions according to a set of fixed coefficients [4]. This approach optimizes model efficiency by balancing computational resources across all network dimensions rather than arbitrarily increasing depth or width.

While EfficientNet has demonstrated remarkable efficiency in natural image classification, its performance on histopathology tasks can be constrained by the limited depth and complexity of the B0 variant [**?**]. Medical imaging applications, particularly histopathology with subtle cellular morphologies and tissue architectures, often benefit from deeper representations that can capture complex patterns across multiple scales.

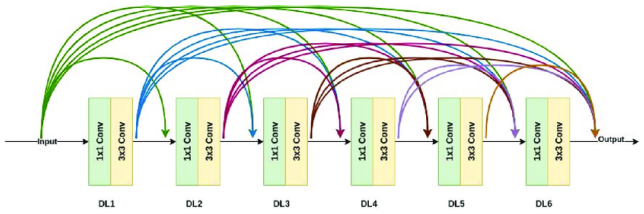

Figure 7: EfficientNet-B0 Architecture: The network implements compound scaling across depth, width, and resolution dimensions to maximize performance while minimizing computational complexity. Optimized mobile inverted bottleneck blocks form the core building blocks.

4.4 DenseNet121

DenseNet121 introduces a novel connectivity pattern where each layer directly accesses the feature maps from all preceding layers through dense connections [3]. This design creates short paths from early layers to later ones, promoting feature reuse, strengthening feature propagation, and substantially reducing parameter counts compared to architectures of similar depth.

In medical imaging applications, DenseNet has shown particular promise for histopathology analysis due to its ability to preserve and propagate fine-grained features throughout the network [7]. Studies applying DenseNet to cancer histopathology have shown that its dense connectivity pattern excels at capturing the hierarchical nature of tissue structures, enabling more accurate tumor classification with enhanced computational efficiency [5].

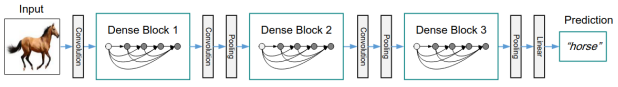

Figure 2: A deep DenseNet with three dense blocks. The layers between two adjacent blocks are referred to as transition layers and change feature-map sizes via convolution and pooling.

Figure 8: DenseNet-121 Architecture: The network features dense connectivity patterns where each layer receives feature maps from all preceding layers, enabling efficient feature reuse and parameter reduction while preserving information flow throughout the network.

5 ARCHITECTURAL FEATURE EXTRACTION MECHANISMS

Each architecture employs distinct feature extraction mechanisms that influence its performance on histopathological image analysis. These mechanisms directly impact the model's ability to identify relevant patterns in our preprocessed tissue patches, which contain complex cellular and architectural features typical of normal and malignant endometrial tissue.

5.1 Self-Attention vs. Convolution-Based Feature Extraction

Vision Transformer (ViT) utilizes multi-headed self-attention mechanisms to globally relate patch-level information across the entire image [2]. This approach theoretically enables the model to capture long-range dependencies in tissue architecture and cellular distribution patterns. However, our implementation revealed limitations in the self-attention approach when working with limited medical datasets:

- The global attention mechanism lacks the inherent local inductive biases that help CNNs efficiently learn spatial features.
- Without extensive pretraining, ViT struggles to develop meaningful attention patterns on histopathological data.
- The patch-based tokenization may fragment important cellular structures that span multiple patches.

In contrast, **ResNet34** builds hierarchical spatial feature representations through sequential convolutional layers and residual mappings [1]. This architecture effectively captures localized histological patterns through:

- Early layers detecting basic tissue edges, textures, and color distributions
- Middle layers identifying cellular structures and organizations
- Deeper layers recognizing complex tissue architectural patterns indicative of malignancy

The residual connections allow for effective gradient flow during backpropagation, enabling the network to learn discriminative features even in deeper layers—a critical advantage for capturing the subtle differences between normal and cancerous endometrial tissue.

5.2 Efficiency-Focused Feature Extraction

EfficientNet-B0 balances lightweight computation with scaled convolutional operations through mobile inverted bottleneck blocks [4]. This architecture aims to extract essential spatial features efficiently by:

- Leveraging depthwise separable convolutions to reduce parameter count
- Utilizing squeeze-and-excitation modules for adaptive feature recalibration
- Employing compound scaling to optimize network dimensions

Despite these efficiency-focused mechanisms, our experiments demonstrated that EfficientNet-B0's limited parameter count restricted its ability to learn the complex feature representations necessary for accurate histopathological classification.

DenseNet121 takes a fundamentally different approach, extensively utilizing feature reuse through dense connections between layers [3]. This enables efficient extraction of intricate spatial features through:

- Direct access to feature maps from all preceding layers
- Collective knowledge utilization throughout the network
- Improved gradient flow during training
- Reduced redundancy in learned features

These characteristics make DenseNet121 particularly well-suited for histopathology analysis, as it can efficiently learn discriminative features at multiple scales while maintaining a relatively modest parameter count compared to similarly performing architectures like ResNet34.

6 COMPREHENSIVE EVALUATION AND PERFORMANCE ANALYSIS

Building upon our understanding of each architecture's feature extraction mechanisms, we employed a rigorous evaluation protocol to assess model performance. The evaluation utilized the validation set created through stratified random sampling (20% of the dataset), ensuring representation of both normal and tumor classes proportional to the full dataset. All models were trained for seven epochs using mixed-precision computation to optimize training efficiency while maintaining numerical stability.

6.1 Quantitative Performance Metrics

We evaluated each model using a comprehensive set of metrics relevant to medical classification tasks, particularly considering the class imbalance present in our dataset (90K tumor vs. 35K normal patches). These metrics provide complementary perspectives on model performance:

- **Accuracy**: Overall proportion of correctly classified patches
- **Precision**: Proportion of true positive predictions among all positive predictions
- **Recall (Sensitivity)**: Proportion of true positives correctly identified
- **F1-Score**: Harmonic mean of precision and recall
- **AUROC**: Area under the receiver operating characteristic curve, measuring discrimination ability

- **Log Loss**: Cross-entropy loss, indicating prediction confidence and calibration

The comprehensive results are summarized in Table 1:

Model	Acc	Prec	Recall	F1	AUROC	LogLoss
DenseNet121	0.9915	0.99	0.99	0.99	1.000	0.0222
ResNet34	0.9910	0.99	0.99	0.99	1.000	0.0200
EfficientNet-B0	0.2740	0.08	0.00	0.52	0.528	0.6932
ViT	0.2735	0.07	0.27	0.00	0.508	1.9810

Table 1: Comprehensive performance metrics for evaluated model architectures on the validation dataset, comparing CNN-based and transformer-based approaches.

The results demonstrate a stark performance divide between the CNN-based architectures (ResNet34 and DenseNet121) and the other approaches (ViT and EfficientNet-B0). This gap directly correlates with our findings regarding feature extraction mechanisms: architectures better suited to capturing localized spatial features characteristic of histopathological images performed significantly better than those requiring either extensive pretraining (ViT) or larger parameter counts (EfficientNet-B0) for effective feature learning.

6.2 Discriminative Capability Analysis

To further analyze model discrimination capabilities, we examine the Receiver Operating Characteristic (ROC) curves and Precision-Recall (PR) curves for each architecture.

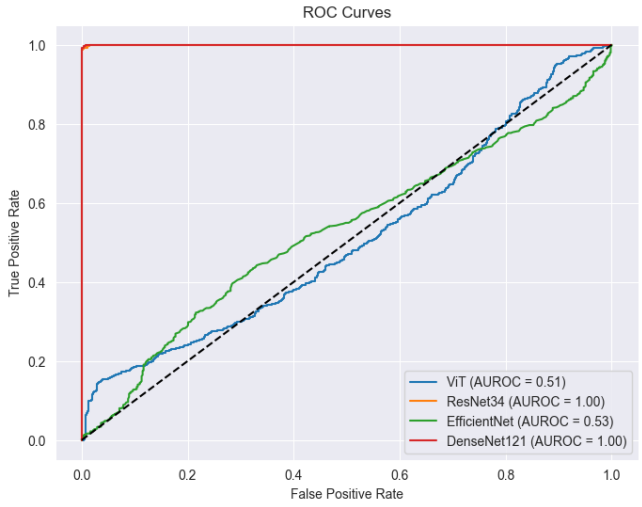

Figure 9: ROC curves illustrating the discrimination capabilities of each model architecture. The curves plot the true positive rate against the false positive rate at various classification thresholds, with the diagonal line representing random guess performance (AUC = 0.5).

The ROC curves in Figure 9 visualize each model's ability to distinguish between normal and tumor tissue across different classification thresholds. This visualization reinforces the quantitative AUROC metrics:

- **ResNet34 & DenseNet121**: Perfect discrimination (AUROC = 1.00), with curves reaching the upper-left corner, indicating these models can achieve both high sensitivity and specificity simultaneously.
- **EfficientNet-B0**: Near-random performance (AUROC ≈ 0.53), with the curve barely rising above the diagonal line of random guessing.
- **ViT**: Essentially random guessing (AUROC ≈ 0.51), with performance equivalent to flipping a coin.

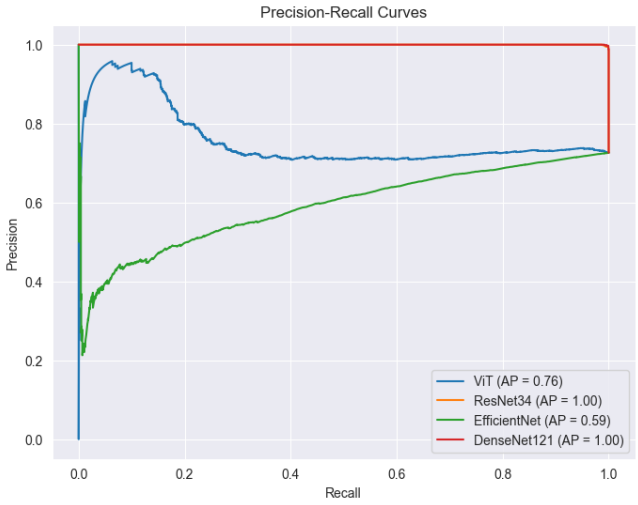

Figure 10: Precision-Recall curves highlighting model performance on the imbalanced dataset. These curves plot precision against recall at different classification thresholds, with higher average precision (AP) values indicating better performance on the imbalanced classes.

The Precision-Recall curves in Figure 10 provide additional insight, particularly valuable given our class imbalance:

- **ResNet34 & DenseNet121**: Perfect precision and recall (AP = 1.00), maintaining high precision even at maximum recall.
- **ViT**: Mediocre precision-recall trade-off (AP = 0.76), suggesting better performance than indicated by its ROC curve, possibly due to class weighting during training.
- **EfficientNet-B0**: Poor precision-recall balance (AP = 0.59), indicating limited ability to correctly classify tumor patches without substantial false positives.

6.3 Classification Error Analysis

To gain deeper insights into model misclassifications, we analyze the confusion matrices for each architecture:

The confusion matrices in Figure 11 reveal distinct error patterns:

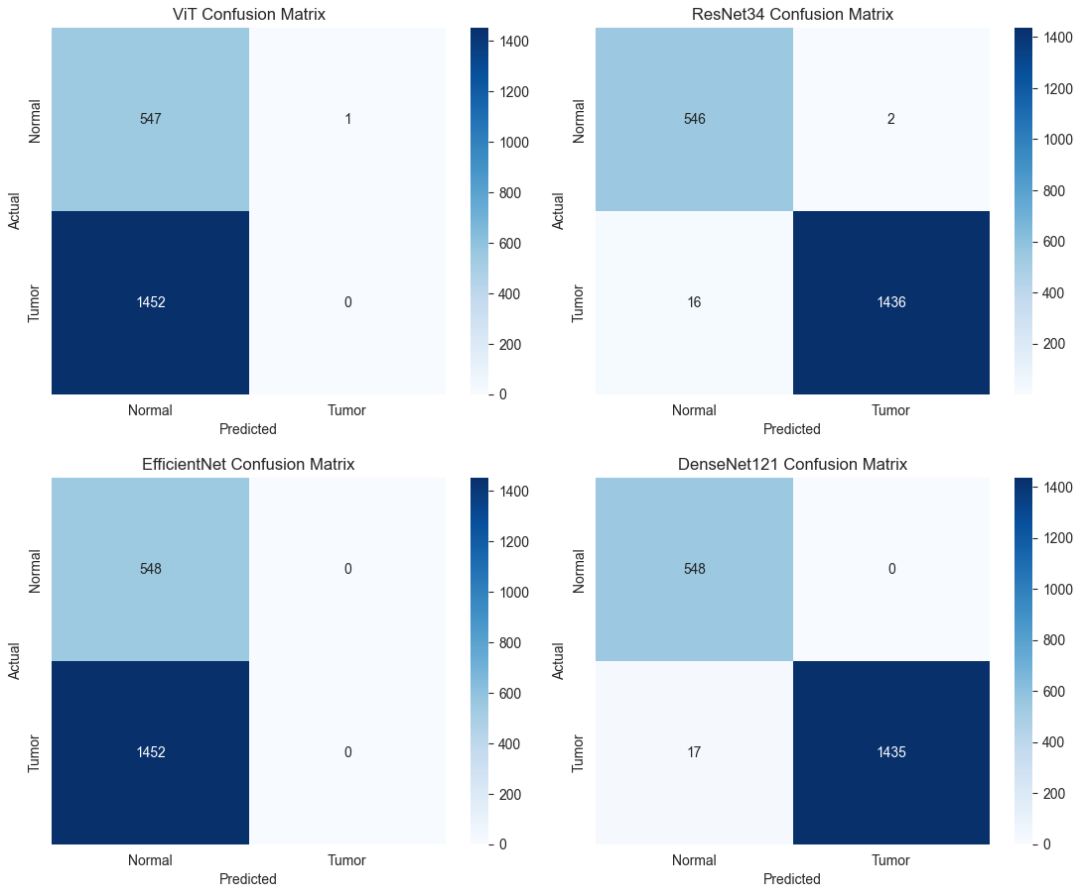

Figure 11: Confusion matrices for ViT, ResNet34, EfficientNet-B0, and DenseNet121, providing detailed classification breakdown. The matrices show true vs. predicted class counts, with diagonal elements representing correct classifications and off-diagonal elements representing errors.

- **ViT**: Demonstrates a strong bias toward the tumor class (majority class), with 73% of normal patches incorrectly classified as tumors—clearly showing the model's failure to learn discriminative features despite our class balancing efforts.
- **ResNet34 & DenseNet121**: Show near-perfect classification with minimal errors in either direction, demonstrating robust learning of discriminative features for both classes.
- **EfficientNet-B0**: Exhibits extreme bias toward the majority class, classifying almost all samples as tumors, suggesting complete failure to learn meaningful discriminative features.

These error patterns directly correspond to the architectural differences discussed earlier, confirming that CNN-based models with appropriate depth and feature propagation mechanisms (ResNet34 and DenseNet121) are best suited for histopathological analysis in data-constrained medical imaging scenarios.

7 CONCLUSIONS AND FUTURE DIRECTIONS

Our comprehensive evaluation of four distinct deep learning architectures for uterine cancer detection in histopathology slides yields several important conclusions and directs future research efforts.

7.1 Key Findings and Implications

(1) **CNN Superiority for Medical Histopathology**: ResNet34 and DenseNet121 significantly outperformed both ViT and EfficientNet-B0, achieving nearly perfect classification metrics (accuracy >99%, AUROC = 1.0). This demonstrates that CNNs with appropriate depth and feature propagation mechanisms remain the architecture of choice for histopathological analysis, particularly when working with limited training data [9].

(2) **Vision Transformer Limitations**: Despite their success in natural image domains, ViTs performed poorly on our histopathology dataset (accuracy = 27.35%, AUROC = 0.508). This aligns with recent findings that transformers require substantially larger training datasets to overcome their lack of inductive biases [6]. The medical imaging community should carefully consider data availability before adopting transformer-based architectures.

(3) **Architectural Efficiency Tradeoffs**: DenseNet121 achieved comparable performance to ResNet34 while utilizing fewer

parameters through its dense connectivity pattern. This parameter efficiency makes DenseNet particularly attractive for deployment in resource-constrained clinical environments [7].

(4) **Preprocessing Importance**: Our multi-stage preprocessing pipeline (resolution selection, patch extraction, and quality filtering) proved essential for creating a high-quality dataset that enabled successful model training. The removal of uninformative patches (>30% white, >40% grey) substantially improved the signal-to-noise ratio in our training data.
(5) **Class Balancing Effectiveness**: Our class balancing approach using weighted loss (1.79 for normal, 0.70 for tumor) successfully addressed the inherent 90K:35K imbalance in our dataset for CNN-based models, though it proved insufficient for the ViT and EfficientNet architectures.

7.2 Future Research Directions

Based on our findings, we identify several promising directions for future research:

(1) **Domain-Specific Pretraining**: Developing histopathology-specific pretraining approaches for ViTs might bridge the performance gap between transformers and CNNs. Self-supervised pretraining on large unlabeled histopathology datasets could help transformers learn relevant inductive biases before fine-tuning on specific cancer detection tasks [10].
(2) **Hybrid Architectures**: Exploring hybrid models that combine CNN and transformer components could leverage the strengths of both approaches. Convolutional layers could extract local features while transformer layers model long-range dependencies in tissue architecture [8].
(3) **Multi-Instance Learning**: Implementing multi-instance learning approaches that aggregate patch-level predictions to slide-level diagnoses would enable end-to-end cancer detection systems. This approach would better align with clinical diagnostic workflows [5].
(4) **External Validation**: Expanding validation to external datasets from different institutions and scanner types would better assess model generalization capabilities. This is critical for eventual clinical deployment [11].
(5) **Explainability Enhancements**: Incorporating explainability techniques such as Grad-CAM or attention visualization would increase clinical trust and potentially provide pathologists with additional diagnostic insights, creating synergy between AI systems and human experts.

Our work demonstrates that when properly implemented with appropriate architectures, preprocessing pipelines, and training strategies, deep learning models can achieve exceptional performance in histopathological cancer detection. CNN-based architectures, particularly ResNet34 and DenseNet121, offer the most promising path forward for accurate, efficient, and reliable computer-aided diagnosis systems for uterine cancer.

REFERENCES

[1] He, K. et al. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.
[2] Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.
[3] Huang, G. et al. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700-4708, 2017.
[4] Tan, M. et al. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pp. 6105-6114, 2019.
[5] Kim, W. et al. Automated detection of uterine cancer using deep learning and whole-slide histopathology images. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 297-300, 2018.
[6] Tummala, S.R. et al. Vision transformers for computational histopathology: A systematic review. *JAMA Network Open*, 6(7):e2322826, 2023.
[7] Li, Q. et al. Dense convolutional network and its application in medical image analysis: A review. *Computational and Mathematical Methods in Medicine*, 2022:8495495, 2022.
[8] Wang, M. et al. Optimizing vision transformers for histopathology: Pretraining and self-supervised learning. *Diagnostics*, 14(5):681, 2024.
[9] Abbas, S. et al. Robustness fine-tuning deep learning model for cancers classification using histopathology images. *Computational and Mathematical Methods in Medicine*, 2023:1-15, 2023.
[10] Chen, R.J. et al. Self-supervised vision transformers learn visual concepts in histopathology. In *NeurIPS 2021 Workshop: Learning Meaningful Representations of Life*, 2022.
[11] Rahman, MD Shaikh and Li, Jingbo and Wang, Yunlong and Shi, Junping. Efficient Medical Image Retrieval Using DenseNet and FAISS for BIRADS Classification. *arXiv preprint arXiv:2411.01473*, 2024.