Lung Cancer Histopathological Classification: CNN vs ScatNet Approaches

Lorenzo Mioso

April 2025

Abstract

This report compares Convolutional Neural Networks (CNNs) and Scattering Networks (ScatNets) for lung cancer histopathological image classification. Evaluating both models on adenocarcinoma and benign tissue samples, I found that CNNs achieve superior performance (99.26% accuracy vs ScatNet's 92.99%), despite ScatNet's theoretical advantages of translation and rotation invariance. Grayscale features proved sufficient for accurate classification, indicating tissue morphology rather than staining drives model decisions. Analysis through guided backpropagation revealed CNNs produce more interpretable and pathologically relevant attribution maps compared to ScatNet's diffuse visualizations.

1 Introduction

Automated classification of histopathological images plays a crucial role in computer-aided diagnosis systems for cancer detection. This study compares two fundamentally different approaches for classifying lung tissue samples into adenocarcinoma (malignant) and benign categories: Convolutional Neural Networks (CNNs) with learned filters and Scattering Networks (ScatNets) with predefined wavelet filters.

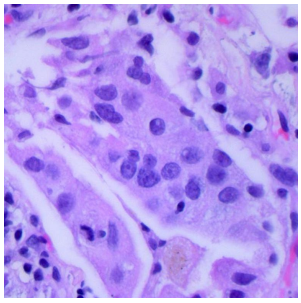
While CNNs have demonstrated remarkable success in image classification tasks, their lack of built-in invariance to transformations and black-box nature remain concerns. ScatNets offer mathematical guarantees of invariance to translations and rotations, potentially beneficial for histopathological image analysis where orientation is not diagnostically significant.

This work compares these approaches not only in terms of classification accuracy but also interpretability, using guided backpropagation to visualize features influencing model decisions and provide insights into their reasoning processes.

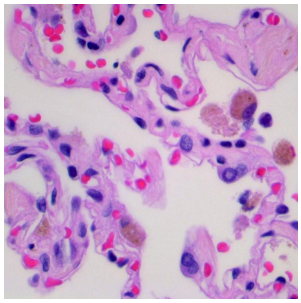
2 Dataset and Preprocessing

2.1 Dataset Description

The dataset consists of lung cancer histopathological images categorized into three classes: adenocarcinoma, squamous cell carcinoma, and benign tissue. For this study, I focus on binary classification between adenocarcinoma and benign tissue. The images are high-resolution (768×768 pixels) microscopic views of lung tissue samples.



(a) Adenocarcinoma sample



(b) Benign tissue sample

Figure 1: Representative images from the dataset showing distinct histopathological patterns

2.2 Preprocessing Pipeline

Initial experiments revealed that models achieved near-perfect accuracy by learning color distributions rather than morphological features (Figure 2). Since staining variations are not diagnostically relevant, I converted images to grayscale to focus the models on tissue structure.

My preprocessing pipeline included:

- Grayscale conversion
- Per-fold normalization and standardization
- Data augmentation: random rotations, flips, color jittering, cropping, and Gaussian noise

Figure 3 demonstrates the effect of augmentation.

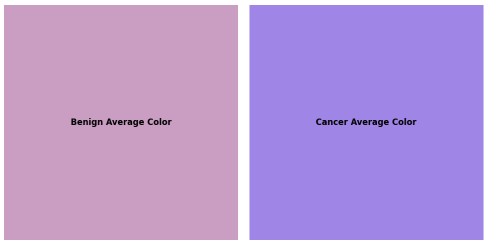
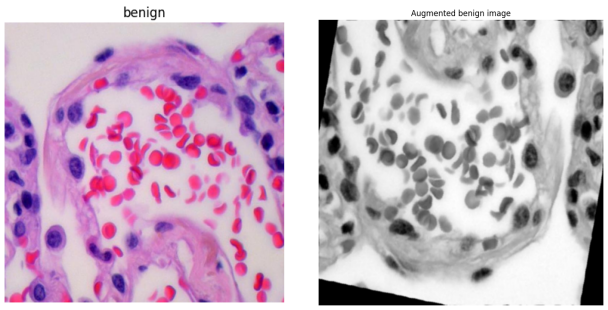


Figure 2: Average color distribution across classes, showing distinct color profiles that could lead to artificial classification cues



(a) Original image (b) Augmented image

Figure 3: Example of data augmentation applied to a histopathological image

3 Model Architectures

3.1 CNN Architecture

I designed an efficient CNN architecture inspired by ResNet but without skip connections. The model consists of three progressive convolutional blocks with increasing filter complexity (16→16→24 filters), as illustrated in Figure 4.

Key architectural elements include:

- Initial 11×11 convolutional kernel to capture broad tissue patterns
- Strategic max pooling for dimensionality reduction
- Batch normalization after each convolutional layer
- ReLU activations for modeling non-linear relationships

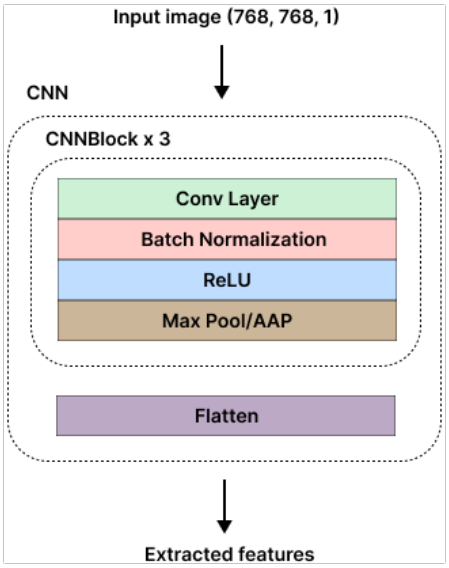


Figure 4: CNN architecture with three convolutional blocks followed by a classifier

3.2 Scattering Transform

The scattering transform is a mathematical operation that creates representations of signals invariant to translations and stable to deformations. It operates by cascading wavelet transforms with non-linear modulus operators, followed by local averaging.

The process works in multiple stages:

- First-order coefficients capture basic edge and texture information
- Second-order coefficients extract more complex patterns resistant to deformation
- Low-pass filtering provides translation invariance while preserving spatial information

Figure 5 illustrates the cascaded structure of the scattering transform, where each level provides increasing invariance to geometrical transformations while preserving discriminative information.

Unlike CNNs, which learn filters from data, the scattering transform uses predefined wavelet filters with mathematical guarantees. This approach offers several advantages for image analysis, including stability to small deformations, preservation of high-frequency information, and reduced sensitivity to rotation and scale changes—qualities potentially beneficial for analyzing biological structures in medical imaging.

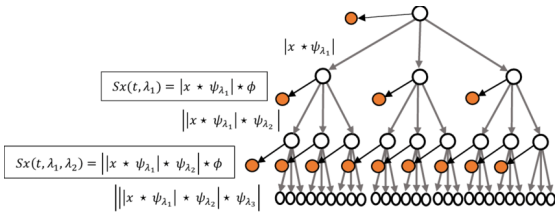


Figure 5: Diagram of the scattering transform showing wavelet decomposition and feature extraction process

3.3 ScatNet Architecture

The ScatNet architecture employs wavelet scattering transforms rather than learned filters. My implementation uses:

- J=3 scale parameter for wavelet decomposition
- L=8 orientations for directional sensitivity
- M=2 scattering order to capture higher-order interactions
- 4×4 global average pooling to reduce dimensionality

The ScatNet processes images through a series of wavelet transformations, producing 3,472 translation, rotation, and scaling invariant features that are then fed to the classifier, as shown in Figure 6.

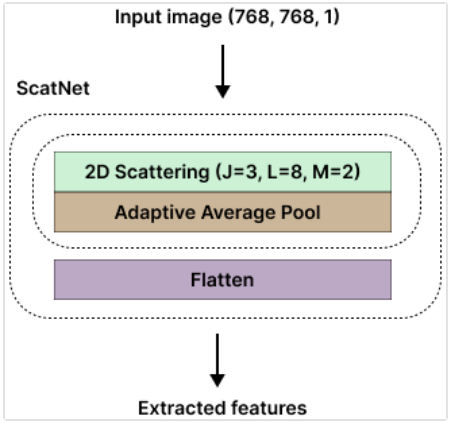


Figure 6: ScatNet architecture with wavelet transforms and classifier

3.4 Feature Classifier

Both the CNN and ScatNet feature extractors feed into the same classifier architecture:

- Single hidden layer with 16 neurons
- Batch normalization for training stability
- ReLU activation for non-linear patterns
- Dropout (0.5) for regularization
- Output layer with 2 neurons and softmax activation

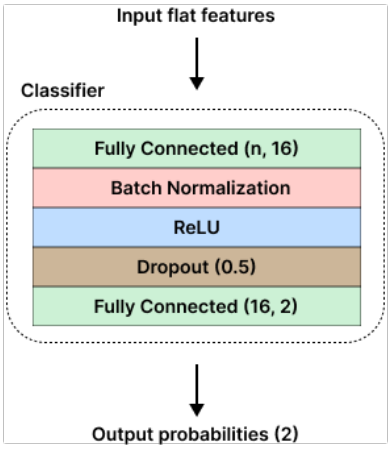


Figure 7: Feature classifier architecture used with both CNN and ScatNet

4 Training and Evaluation

4.1 Training Methodology

Both models were trained using 10-fold cross-validation with Adam optimizer, but required different learning rate approaches. CNNs benefited from higher learning rates (1e-3 to 5e-3) to develop effective filters capturing subtle tissue patterns, while ScatNet used more conservative rates. The CNN demonstrated rapid convergence (typically within 15 epochs) with minimal overfitting and consistent performance across folds. In contrast, ScatNet required fewer training epochs but showed slower per-epoch convergence and greater cross-fold variance. Early stopping was implemented to prevent overfitting, with model checkpointing to save the best-performing weights.

5 Results and Analysis

5.1 Performance Metrics

Both models were evaluated using accuracy and F1 score across all 10 folds. Table 1 summarizes the comparative performance.

Table 1: Performance comparison between CNN and ScatNet

Metric	CNN	ScatNet
Mean Accuracy	99.26% ± 0.72%	92.99% ± 1.59%
Mean F1 Score	99.27% ± 0.72%	92.83% ± 1.70%
Best Fold	99.90% (Fold 9)	95.10% (Fold 6)

The CNN significantly outperformed ScatNet by 6.27% in mean accuracy. Furthermore, the CNN demonstrated more consistent performance across folds, with a standard deviation of 0.72% compared to ScatNet's 1.59%.

5.2 Performance Comparison

Both models showed strong performance, with CNN significantly outperforming ScatNet:

- **CNN**: Achieved exceptional accuracy (98.00%-99.90%) and F1 scores (98.01%-99.90%) across all folds, with the best performance in Fold 9 (99.90% accuracy).
- **ScatNet**: Demonstrated good but less consistent performance with accuracy ranging from 89.10% to 95.10% and F1 scores from 88.68% to 95.06%. The best performing fold was Fold 6 (95.10%), while Fold 1 performed worst (89.10%).

CNN not only achieved higher peak performance but also showed greater consistency across folds, with a standard deviation of 0.72% compared to ScatNet's 1.59%.

6 Feature Analysis

6.1 CNN Filter Analysis

Visual inspection of the first-layer CNN filters revealed structured patterns relevant to histopathological analysis, as shown in Figure 8.

The filters exhibited several distinct patterns:

- Diagonal and vertical strips for detecting edges and transitions between tissue textures

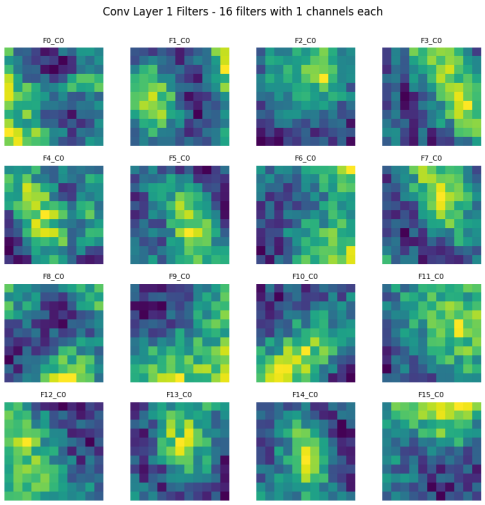


Figure 8: Learned filters from the first convolutional layer of the CNN

- Circular patterns for identifying cell nuclei and localized features
- Two-part filters capturing gradients and contrast changes

These patterns align with pathologically relevant features in histological images, such as cell boundaries, nuclei clustering, and tissue organization differences between normal and malignant samples.

6.2 ScatNet Filter Analysis

Unlike CNN, ScatNet uses predefined wavelet filters rather than learned ones. Figure 9 shows representative ScatNet filters.

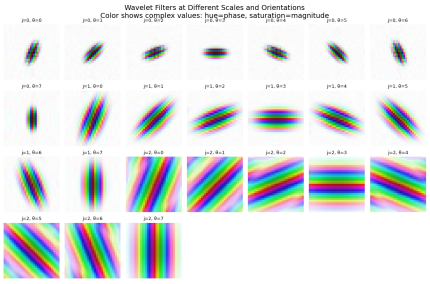


Figure 9: Predefined wavelet filters used in ScatNet

While these filters provide theoretical advantages:

- Scale and rotation invariance
- Mathematical guarantees on stability
- No learning required

Their fixed nature limits adaptability to the specific characteristics of histopathological images, potentially explaining ScatNet's lower performance despite its solid theoretical foundation.

7 Guided Backpropagation for Model Interpretation

7.1 Methodology

Guided backpropagation is an enhancement of standard backpropagation that helps visualize which image regions most influence a neural network's classification decision. This technique functions by:

1. Starting with a forward pass through the network
2. Computing gradients of the output with respect to the input image
3. Modifying the backpropagation of ReLU layers to only allow positive gradients to flow through positive activations

This approach produces clearer, less noisy visualizations by eliminating negative influences that typically create visual noise in standard backpropagation. The result is a saliency map that highlights image regions most strongly activating the target class.

Mathematically, guided backpropagation modifies the gradient flow through ReLU layers as follows:

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial f} \cdot \mathbb{1}_{x>0} \cdot \mathbb{1}_{\frac{\partial y}{\partial f}>0} \tag{1}$$

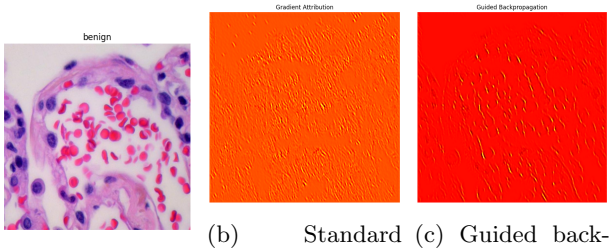
Where x is the input to the ReLU, f is the output, y is the target output, and $\mathbb{1}$ is the indicator function.

7.2 CNN Attribution Analysis

Guided backpropagation revealed distinctive patterns in the CNN's decision-making process, as shown in Figure 10.

Key observations:

- Guided backpropagation significantly reduced noise compared to regular backpropagation
- Higher resolution feature attribution with clearer cellular structure patterns
- Strong correlation with pathologically relevant regions
- Consistent visualization patterns across different implementation methods



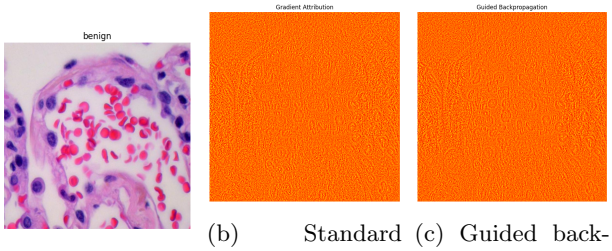
(a) Original image (b) Standard backprop (c) Guided backprop

Figure 10: CNN attribution analysis comparing standard and guided backpropagation

The guided backpropagation visualizations for CNN highlighted cellular structures and tissue organization patterns that pathologists would typically examine, suggesting that the model focuses on clinically relevant features.

7.3 ScatNet Attribution Analysis

ScatNet's attribution maps showed markedly different characteristics, as illustrated in Figure 11.



(a) Original image (b) Standard backprop (c) Guided backprop

Figure 11: ScatNet attribution analysis comparing standard and guided backpropagation

Notable differences:

- Minimal differences between guided and regular backpropagation
- Less interpretable, diffuse attribution regions
- Weaker correlation with pathologically significant structures
- Limited impact of guided backpropagation due to ReLUs only in the classifier layers

The limited improvement seen with guided backpropagation in ScatNet can be attributed to its architecture, where non-linearities (ReLU) are only present in the classifier component, not in the feature extraction process. This architectural difference significantly affects how attribution methods perform.

8 Discussion

8.1 Performance Gap Analysis

The substantial performance gap between CNN and ScatNet (6.27% mean accuracy difference) can be attributed to several factors:

- **Adaptability**: CNN filters are learned specifically for the histopathology domain, while ScatNet uses fixed mathematical wavelets.
- **Feature specificity**: CNN develops task-specific feature detectors that identify diagnostically relevant patterns in lung tissue.
- **Hierarchical learning**: CNN's multi-layer structure progressively builds more complex feature representations.
- **Over-generalization**: ScatNet's invariance properties, while theoretically advantageous, may discard discriminative information relevant to cancer detection.

8.2 Interpretability Comparison

From an interpretability perspective:

- CNN attribution maps showed stronger correspondence with pathologically relevant structures
- ScatNet attribution maps were more diffuse and less aligned with cellular structures
- Guided backpropagation significantly enhanced CNN visualizations but had minimal impact on ScatNet visualizations

These findings suggest that while ScatNet has theoretical interpretability advantages due to its mathematical foundation, in practice, CNN produces more clinically interpretable explanations.

8.3 Color vs. Structure

My experiments confirmed that grayscale images provided sufficient information for accurate classification. This finding has important implications:

- Tissue morphology, not staining characteristics, drives model decisions
- Models are less likely to be affected by staining variations between laboratories
- Potential for application to unstained or differently stained samples

9 Limitations

Despite the strong results, this study has several limitations that should be acknowledged:

- **Dataset diversity**: The analysis is limited to a single institutional dataset, which may not capture the full spectrum of tissue appearance variations across different laboratories and staining protocols.
- **Binary classification focus**: The study addresses only binary classification (adenocarcinoma vs. benign), whereas clinical applications often require multi-class discrimination between different cancer subtypes.
- **Fixed architecture comparison**: The comparison evaluates specific implementations of CNN and ScatNet rather than the full range of possible architectures within each approach.
- **Limited interpretability methods**: While guided backpropagation provides valuable insights, additional explainability techniques could offer complementary perspectives on model decision processes.

10 Conclusion

This study compared CNN and ScatNet approaches for lung cancer histopathological image classification, evaluating both performance and interpretability aspects. My findings clearly demonstrate that CNN outperforms ScatNet in this domain despite ScatNet's theoretical advantages.

The superior performance of CNNs can be attributed to their ability to learn domain-specific features directly from the data, adapting to the unique characteristics of histopathological images. The guided backpropagation analysis further revealed that CNN models focus on pathologically relevant regions, providing more interpretable and clinically aligned visualizations than ScatNet.

Based on these findings, I recommend:

- **Clinical implementation**: Deploy CNN-based systems for computer-aided diagnosis in lung histopathology with confidence in their accuracy and interpretability.
- **Model validation**: Conduct multi-center validation studies to verify CNN performance across different laboratory and staining protocols.

- **Feature extraction**: Consider using the identified CNN filters as standardized feature extractors for other histopathology applications.
- **Interpretability tools**: Integrate guided backpropagation visualization into pathologist workflow to enhance model trust and adoption.

Future work should explore hybrid approaches combining the adaptability of CNNs with the mathematical guarantees of ScatNets to potentially achieve both high performance and robust interpretability.