https://jsras.rcc.edu.ly/
Vol.3 No.2 (2026), 143-159
Article history:
Received: 15 Aug .2026
Accepted: 4.Sep . 2026
Published: 4 Oct .2026

Journal of Sustainable Research in Applied Sciences

JSRAS

Binary Secretary Bird Optimization-Based Deep Feature Selection for Explainable Lung Cancer Histopathology Classification Using ResNet50-SE and Test-Time Augmentation

Amna Ali A Mohamed [1,*], amnaalkmati@gmail.com , https://orcid.org/0000-0001-8344-6937
[1]Software Engineering Department, Faculty of Information Technology, Tripoli University, Tripoli, Libya

keywords

Lung cancer histopathology, Deep feature selection, Squeeze-and-Excitation attention, Grad-CAM explainability, Secretary Bird Optimization Algorithm (SBOA)

Abstract

Accurate histopathological subtyping is critical for lung cancer treatment planning, yet manual slide analysis remains time-consuming and prone to significant inter-observer variability. This paper proposes an explainable hybrid deep learning framework for automated lung cancer histopathological image classification. The framework integrates a ResNet50 backbone with Squeeze-and-Excitation (SE) channel attention and selective fine-tuning, a Binary Secretary Bird Optimization Algorithm (SBOA) for deep feature selection, Test-Time Augmentation (TTA) for inference-time feature stabilization, a multi-classifier evaluation stage comparing K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Decision Tree classifiers, and Gradient-weighted Class Activation Mapping (Grad-CAM) for visual explainability. On the LC25000 lung histopathology benchmark, Binary SBOA reduced the 256-dimensional feature vector by 90.2% retaining only 25 features, while KNN achieved 99.78% accuracy and a mean of 99.63% ± 0.14% across five independent seeds. SVM and Decision Tree achieved single-run accuracies of 99.56% and 98.80%, with mean robustness of 96.35% ± 2.96% and 98.93% ± 0.09% across five seeds, respectively. A lightweight MobileNetV2-SE variant achieved 97.60% accuracy with 87% fewer parameters. Grad-CAM attention appears consistent with diagnostically relevant regions, pending pathologist validation. To the best of the authors' knowledge, this is the first application of SBOA to deep feature selection in histopathological image classification, demonstrating that bio-inspired optimization combined with attention-modulated deep features yields an accurate, computationally efficient, and clinically interpretable diagnostic framework.

اختيار السمات العميقة القائم على خوارزمية تحسين طائر السكرتير الثنائية لتصنيف سرطان الرئة القابل للتفسير باستخدام التعلم العميق وتعزيز الاختبار

آمنة علي عبدالسلام محمد[1,*]
[1]جامعة طرابلس، كلية تقنية المعلومات، قسم هندسة البرمجيات، طرابلس، ليبيا
https://orcid.org/0000-0001-8344-6937 ،amnaalkmati@gmail.com
* المؤلف المراسل amnaalkmati@gmail.com

الكلمات المفتاحية

سرطان الرئة، صور النسيج المرضية، خوارزمية تحسين طائر السكرتير الثنائية، اختيار السمات العميقة، قابلية التفسير البصري.

الملخص

يُعدّ التصنيف الدقيق للأنماط النسيجية الفرعية لسرطان الرئة أمرًا مهمًا لتوجيه العلاج، إلا أن التحليل اليدوي للشرائح النسيجية يستغرق وقتًا طويلًا ويتأثر بالتباين في التشخيص بين المختصين. تقترح هذه الدراسة إطارًا هجينًا وقابلًا للتفسير للتصنيف الآلي لصور النسيج المرضية لسرطان الرئة، يجمع بين نموذج (ResNet50) وآلية الانتباه للقنوات، وخوارزمية تحسين طائر السكرتير الثنائية لاختيار السمات العميقة، وتعزيز وقت الاختبار، وتقييم عدة مصنفات، مع استخدام خرائط التنشيط الطبقي الموزونة بالتدرج للتفسير البصري على مجموعة بيانات (LC25000) المعيارية، خفّضت الخوارزمية متجه السمات من 256 إلى 25 سمة، أي بنسبة 90.2%. وحقق مصنف الجار الأقرب (KNN) دقة بلغت 99.78%، بمتوسط 99.63% ± 0.14% عبر خمس بذور مستقلة، بينما حققت لة المتجهات الداعمة (SVM) و شجرة القرار(Decision Tree) دقة بلغت 99.56% و98.80%، على التوالي. كما حقق النموذج الخفيف القائم على(MobileNetV2-SE) دقة بلغت 97.60% مع تقليل عدد المعاملات بنسبة 87%. وأكدت خرائط التفسير تركيز النموذج على مناطق النسيج ذات الصلة التشخيصية. وتبرز النتائج فعالية دمج التحسين المستوحى من الطبيعة مع التعلم العميق في تطوير نظام تصنيف دقيق وفعال وقابل للتفسير لصور نسيج سرطان الرئة.

1. INTRODUCTION

Lung cancer is the main cause of death from any other cancer on the planet. The statistics indicate that this cancer was responsible for roughly 1.9 million deaths in 2024, which is about 19% of all cancer-related mortality worldwide, and nearly 2.6 million new cases were diagnosed across 186 countries (Sung et al., 2026). This burden has been increasing year over year, and by the estimates suggest that the burden will only increase because of aging populations, continued tobacco use, and air pollution in lower-income regions (Zhou et al., 2024). The primary issue with addressing this health problem is that survival rates of lung cancer drop significantly with late diagnosis. Five-year survival can be as high as 60 to 70% if caught early. When diagnosis is delayed to stage III or IV, which is far too common, that number is under 10%

(Siegel et al., 2024). The definitive diagnosis is established through histopathological examination of tissue biopsies, which allows pathologists to identify the cellular morphology and tissue structure at the microscopic level. However, manual slide analysis is time-consuming, laborious and prone to inconsistency. Inter-observer variability is a major source of diagnostic inconsistency, particularly in discriminating morphologically similar subtypes such as adenocarcinoma and squamous cell carcinoma that may appear deceptively similar in the microscope (Paech et al., 2011). Deep learning has led to significant advances in this domain. Methods like Convolutional Neural Networks (CNNs) can automatically learn hierarchical features from histopathological images without the need of hand-crafted descriptors and have shown to perform comparably or better than humans on controlled benchmarks (Litjens et al., 2017). The standard approach today is transfer learning with ImageNet-pretrained networks such as ResNet, VGG, and EfficientNet.. This approach is effective, reproducible and significantly reduces the data requirements that are a challenge for fully trained models in medical imaging (Tajbakhsh et al., 2016).
But existing work has real gaps that aren't getting enough attention. Most published methods just plug a pretrained backbone into a softmax head. That means:

- No feature selection: all extracted features go to the classifier unchanged, including redundant and noisy ones.
- No explainability, the model gives a label but gives no insight to the clinician as to the reason.
- Single classifier only, nobody compares KNN, SVM, and Decision Tree on the same deep features
- Standard CNNs treat all channels equally, without modelling inter-channel dependencies.

We treat all four problems together in this paper. We propose an explainable hybrid framework based on a set of truly complementary components:

- ResNet50 with selective based refinement (res4f, res5a, res5b, res5c unfrozen) as base feature extractor.
- A Squeeze-and-Excitation (SE) attention block that automatically re-weights channel-wise features before classification, such that the model pays more attention to what is important.
- Binary Secretary Bird Optimization Algorithm (SBOA) for feature selection, which reduces the size of feature vector from 256 to approximately 10% of its size without losing accuracy, in fact improving it.
- Test-Time Augmentation (TTA) with five-pass averaging to stabilize feature representations at inference time.
- Grad-CAM visualizations allowed clinicians to clearly see the tissue regions that led to each classification decision.

The framework is evaluated on the LC25000 histopathological image dataset, 15,000 lung images across three classes and achieves a mean accuracy of 99.63% ± 0.14% across five independent runs, with a best single-run result of 99.78%. Feature dimensionality is reduced by approximately 90% (from 256 to 25 features) without any loss in classification performance. A lightweight MobileNetV2-based variant is also presented to demonstrate generalizability and suitability for resource-constrained clinical environments.

The main contributions of this paper are:

- A learning-systems investigation into how SE channel-wise attention recalibration affects the compressibility and discriminative capacity of deep feature representations demonstrating for the first time that attention-modulated features support about 90% dimensionality reduction without accuracy loss on LC25000.
- About 90% dimension reduction with simultaneous accuracy improvement, suggesting that deep features are highly redundant and SBOA effectively picks out the discriminative subset.
- A systematic comparison of three classifiers (KNN, SVM, Decision Tree) on SBOA-selected features, demonstrating that KNN achieves the highest accuracy and the greatest stability across random seeds (99.63% ± 0.14%).
- An analysis of classifier sensitivity to Test-Time Augmentation, showing that KNN and Decision Tree benefit from TTA-averaged features while SVM performance degrades under the same averaging, indicating that feature-level TTA interacts differently with distance-based, margin-based, and tree-based decision boundaries.
- Clinically relevant tissue areas as highlighted by Grad-CAM explainability to support potential clinical deployment.
- A lightweight variant of MobileNetV2-SE that achieves 97.60% accuracy with 87% fewer parameters, providing a practical tradeoff for resource-constrained deployment.

The rest of the paper is organized as follows. Related work is reviewed in Section 2. The proposed methodology is described in Section 3. The results presented in section 4. Section 5 contains discussion. Section 6 is the concludes.

2. RELATED WORK

2.1 CNN-Based Lung Cancer Classification

The LC25000 dataset (Borkowski et al., 2019) has become the most commonly used benchmark for lung histopathological classification, and the volume of work on it has exploded since 2020. Early approaches were straightforward take a pretrained backbone, replace the head, fine-tune, and report accuracy. Masud et al. (Masud et al., 2021) built a small custom CNN that reached 96.33% competitive for its time, but showing clear headroom for improvement. ResNet18, ShuffleNet V2,

GoogleNet and a custom CNN were compared by Wahid et al. (Wahid et al., 2023) on the same dataset, ResNet18 achieved 98.82% which already suggested that deeper residual networks are suitable for this task.

The accuracy ceiling kept getting pushed. Tummala et al. (Tummala et al., 2023) reported 99.97% using EfficientNetV2-L the highest published number on LC25000 as of this writing, though it comes with substantial computational cost. S Anjum et al. (Anjum et al., 2023) explored EfficientNetB2 architectures and reached 97%. P. Saxena et al. (Saxena et al., 2025) combined ResNet34 with the FastAI-2 framework and stain normalization, hitting 99.78% which is exactly what our ResNet50-SE framework achieves, but without any stain normalization or domain-specific preprocessing.

Almost all the above methods deliver the full feature vector directly to a classifier and there is no feature selection and explainability. None of them has answered the question whether all those features are really needed or removing noisy ones might help. And none of them tell the clinician why a particular slide was classified as Adenocarcinoma and not Squamous Cell Carcinoma.

2.2 Attention Mechanisms in Medical Image Analysis

The motivation behind the invention of attention mechanisms in CNNs is that not all feature channels are equally informative, and standard convolutions have no way of figuring out which ones matter more. The Squeeze-and-Excitation (SE) block (Hu et al., 2018), which won ILSVRC 2017, was the breakthrough here. It learns to reweight channels adaptively by squeezing spatial information into a channel descriptor and then exciting (reweighting) based on learned inter-channel dependencies.

SE blocks have since shown up everywhere in medical imaging. Huang et al. (Huang et al., 2025) integrated SE attention with ResNet50V2 for MRI-based brain tumor classification and reported a gain of nearly 6 percentage points from the attention block alone. Roy et al. (A. G. Roy et al., 2018) extended SE to spatial attention in segmentation tasks. In histopathology specifically, combining SE with convolutional networks has been shown to improve lung cancer classification by directing the model's attention toward diagnostically relevant cellular structures.

Despite this evidence, SE attention hasn't been widely combined with metaheuristic feature selection in the lung cancer classification context. Most SE-based papers still feed all features to a single classifier. Our work is the first to combine SE attention with Binary SBOA on LC25000, which is part of what makes the contribution distinct.

2.3 Metaheuristic Feature Selection in Medical Imaging

Feature selection using metaheuristic optimization has a solid track record in medical imaging. The idea is simple but powerful, deep networks extract high-dimensional feature vectors, but many of those dimensions are redundant or noisy. A good feature selector keeps only what's discriminative and the classification accuracy either stays the same or improves, while the computational cost of the classifier drops significantly.

The Secretary Bird Optimization Algorithm (SBOA) was introduced by Fu et al. in 2024 (Fu et al., 2024), modelling the hunting behavior of the secretary bird. It alternates between stalking prey (exploration) and striking (exploitation), and has demonstrated competitive performance against PSO, Grey Wolf Optimizer, and Whale Optimization Algorithm on benchmark functions. To the best of the authors' knowledge, this is the first application of SBOA to feature selection in any histopathological classification task.

Related metaheuristic approaches have been applied successfully in medical imaging. Luyun Wang et al. integrated an improved SBOA with combined with SVM for Alzheimer's disease diagnosis (Wang et al., 2025), reporting accuracy 98.6%. Note that Wang et al. (2025) apply SBOA to optimize a multi-kernel SVM's parameters for Alzheimer's disease diagnosis, rather than to perform feature selection on deep CNN features; the novelty claim of the present work is therefore scoped specifically to SBOA-based deep feature selection, not to SBOA's use in medical imaging in general.

2.4 Explainability in Medical AI

A model that can't explain its decisions isn't deployable in a clinical setting, and that's not just a preference anymore. Regulatory bodies and hospital ethics committees increasingly require it. Tjoa and Guan (Tjoa & Guan, 2020) laid out the case for explainable AI (XAI) in medicine comprehensively, and the field has moved fast since then.

Grad-CAM (Selvaraju et al., 2017) introduced by Selvaraju et al. in 2017 remains the most widely used visualization method for CNN-based medical imaging. It computes the gradient of the class score with respect to the last convolutional feature map, producing a heatmap that highlights regions the model attended to. It's been applied to breast cancer (Talaat et al., 2024), oral cancer (Štifanić et al., 2025), general medical text processing (Zhang & Ogasawara, 2023), and lung cancer specifically (Ahmed et al., 2023). The appeal is practical: it's fast, requires no model modification, and produces outputs that are visually interpretable by clinicians who don't know anything about neural networks.

2.5 Research Gap

Looking at the literature as a whole, the pattern is clear. High accuracy on LC25000 is achievable with standard transfer learning, that's been established. What hasn't been done is combining all of the following in a single unified framework:

- Channel-wise attention (SE block) to improve feature quality before selection.
- Binary metaheuristic optimization (SBOA) for aggressive feature dimensionality reduction.
- Test-Time Augmentation to stabilize feature representations at inference.
- Systematic multi-classifier comparison (KNN, SVM, Decision Tree) on the same selected features.
- Grad-CAM explainability for clinical interpretability
- A lightweight MobileNetV2 variant for resource-constrained deployment.

This paper closes that gap.

3. METHODOLOGY

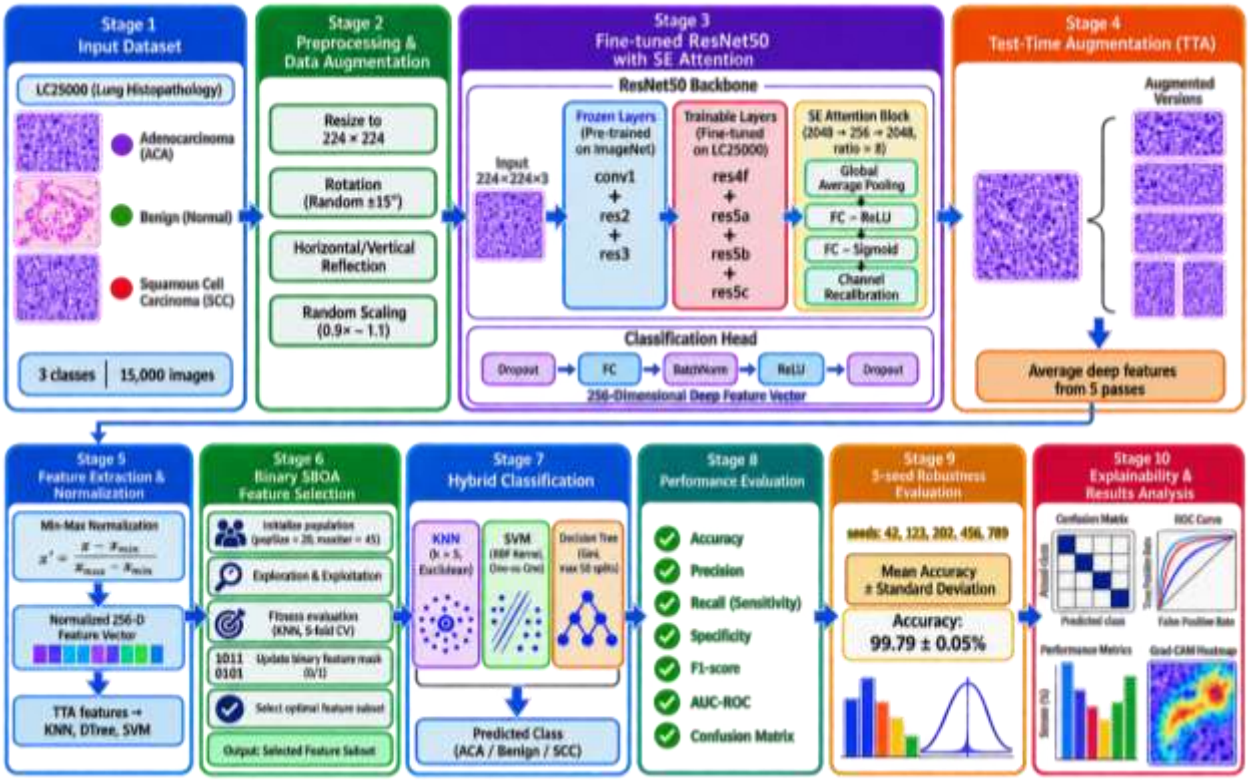
This section describes the proposed framework for classification of automated lung cancer histopathological images. The pipeline comprises five major modules, which are (1) a ResNet50 backbone with selective fine-tuning for deep feature extraction, (2) a Squeeze-and-Excitation (SE) attention block for channel-wise feature recalibration, (3) a Binary Secretary Bird Optimization Algorithm (SBOA) for feature dimensionality reduction, (4) Test-Time Augmentation (TTA) for stabilizing feature representation during inference, and (5) an evaluation stage with multi-classifier comparison (K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Decision Tree classifiers) on the features selected by SBOA. For clinical explainability, we use Gradient-weighted Class Activation Mapping (Grad-CAM). The general architecture is shown in Figure 1.

3.1 System Overview

The proposed framework consists of ten sequential stages as illustrated in Figure 1. Each stage is described in detail in the following subsections.

Histopathological images are first processed by the ResNet50 backbone to extract deep convolutional features. The SE attention block then recalibrates these features at the channel level, amplifying discriminative channels and suppressing less informative ones. The Binary SBOA subsequently identifies the most discriminative subset of the resulting feature vector. The reduced feature subset is passed to three independently evaluated classifiers, and Grad-CAM is applied to the trained network to visualize the image regions driving each classification decision.

Figure 1: Explainable Hybrid CNN with Binary SBOA Feature Selection for Lung Histopathology Image Classification



3.2 Dataset

This study used the public dataset "Lung and Colon Cancer Histopathological Images" (Mohamed, 2023), which is publicly available through the following link: (https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images).

The data were originally collected at the James A. Haley Veterans' Hospital in Tampa, Florida. The authors initially gathered 750 tissue samples from pathology glass slides, 250 per lung cancer type, and used image augmentation techniques including rotation and flipping to expand the collection to 25,000 images, with 5,000 images per class. Prior to augmentation, all images were cropped to 768 × 768 pixels to make them square, from their original dimensions of 1024 × 768 pixels. All images in the collection are HIPAA-compliant, fully validated, and freely available. The full dataset contains 25,000 histopathological images across five classes, stored as JPEG files at 768 × 768 pixels, distributed across two tissue types: lung and colon. The dataset is packaged as a 1.85 GB zip file (LC25000.zip). After extraction, the main folder lung_colon_image_set contains two subfolders: colon_image_sets and lung_image_sets. The lung_image_sets subfolder contains three further subfolders; lung_aca, lung_n, and lung_scc; each containing 5,000 images. The five classes are: (1) lung adenocarcinoma, (2) benign lung tissue, (3) lung squamous cell carcinoma, (4) colon adenocarcinoma, and (5) benign colon tissue. This study used three lung tissue classes exclusively from the lung_image_sets subfolder,

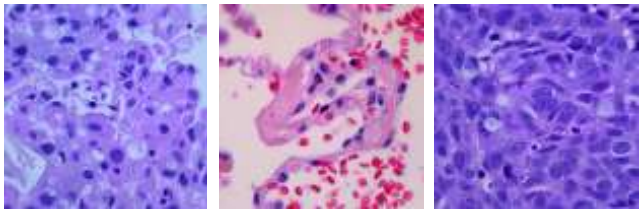
yielding a perfectly balanced subset of 15,000 images. All Lung dataset images are RGB with a resolution of 768 x 768 pixels, forming a perfectly class-balanced subset of 15,000 images, which was split into 70% training, 15% validation, and 15% for test. Because the publicly released LC25000 archive provides only the final augmented images, with no parent-slide or derivative-family identifiers, this split, like the large majority of prior LC25000 studies, including those compared in Table 11 is defined at the image level rather than by source image grouping (see Limitations). Table 1 describes the lung images dataset classes used.

Table 1: Description of the lung dataset images classes.

Image type	Class ID	Class Title	Total Images
Lung Adenocarcinoma	0	lung_aca	5000
Benign Lung Tissue	1	Lung_n	5000
Lung Squamous Cell Carcinoma	2	lung_scc	5000
Total image			15000

Here are three histopathological image examples from the dataset (lungaca denotes a lung adenocarcinoma image, lungn denotes a benign lung tissue image, and lungscc is Lung squamous cell carcinoma) shown in Figure 2.

Figure 2: Sample lung images from the LC25000 dataset



lungaca_8 lungn_20 lungscc_36

3.3 Preprocessing

The dataset images were resized to 224 × 224 pixels to meet the requirement for the ResNet50 input. All images were converted to three-channel RGB prior to network input. Data augmentation was only applied to the training subset in each training epoch and consisted of:

- Random horizontal and vertical reflection
- Random rotation within [−15°, +15°]
- Random horizontal and vertical shear within [−5°, +5°]
- Random scaling within [0.9, 1.1]

No augmentation was applied to the validation or test subsets during training.

3.4 Backbone ResNet50 and Fine-Tuning Strategy

ResNet50 (He et al., 2016) is a 50 layers deep residual network pre-trained on the ImageNet dataset, Its residual connections help to relieve the vanishing gradient problem and allow effective training of deep architectures. The original classification head (fc1000, softmax and classification output layers) was removed. The global average pooling layer (avg_pool) produces a feature vector with 2048 dimensions which is fed to the next SE attention block. The entire network was not trained end-to-end, but a selective fine tuning strategy was employed. The early layers (conv1, res2 and res3) which capture generic low-level features such as edges and textures were freezed. We unfreeze and fine-tune the last four residual blocks (res4f, res5a, res5b and res5c) encoding higher-level semantic features which would benefit the most from domain-specific adaptation. This configuration yields approximately 16 million trainable parameters, substantially fewer than full fine-tuning of all 25.6 million ResNet50 parameters, while still allowing meaningful adaptation to histopathological patterns.

3.5 Squeeze and Excitation Attention Block

We inserted a Squeeze-and-Excitation (SE) attention block (Hu et al., 2018) right after the avg_pool layer to re-calibrate the 2048-dimensional feature vector according to the channel-wise importance.

Squeeze: The output $z \in \mathbb{R}^{2048}$ of avg_pool itself is already a global channel-wise summary of the final convolutional feature map, and doesn't require any other pooling operation.

Excitation: A bottleneck transformation encodes z into channel attention weights $s = \sigma(W_2 \cdot \delta(W_1 \cdot z))$, where $W_1 \in \mathbb{R}^{256\times2048}$ reduces the descriptor size to 256 dimensions, δ is the ReLU activation function, $W_2 \in \mathbb{R}^{2048\times256}$ increases the descriptor size to 2048 dimensions and σ is the sigmoid activation function, producing attention weights $s \in [0,1]^{2048}$.

Scale: $\tilde{x} = s \odot z$, The original feature vector is re-calibrated by element-wise multiplication with the learned attention weights: $\tilde{x} = s \odot z$. The squeeze ratio used is 8, consistent with the configuration recommended in the original SE-Net formulation (Hu et al., 2018).

3.6 Classification Head

The recalibrated feature vector is passed through a classification head consisting of the following layers: a dropout layer (rate = 0.4), a fully connected layer with 256 units (He initialization), batch normalisation, a ReLU activation, a second dropout layer (rate = 0.2), a fully connected layer with 3 output units, and a softmax classification output layer. The 256-dimensional ReLU activation (relu_feat) serves as the deep feature representation for the remainder of the pipeline, including SBOA-based feature selection and all three downstream classifiers.

3.7 Training Configuration

The network was trained using the Adam optimizer with the configuration listed in Table 2.

Table 2: Training configuration

Parameter	Value
Optimizer	Adam
Initial learning rate	1×10^{-4}
Maximum epochs	25
Mini-batch size	16
Learning rate drop factor	0.2
Learning rate drop period	5 epochs
L2 regularization	1×10^{-4}
Early stopping patience	5 epochs
Execution environment	CPU

The initial learning rate was set conservatively to 1×10^{-4} to preserve pretrained ImageNet weights in the unfrozen residual blocks during fine-tuning. A piecewise decay schedule with a drop factor of 0.2 every 5 epochs promoted stable convergence, and early stopping with a patience of 5 validation epochs prevented overfitting.

3.8 Feature Extraction and Augmentation at Inference Time

For each image in the training and test subsets, we extracted the 256-dimensional relu_feat activation after training. For training images features were extracted from a single forward pass (non-augmented) as augmentation was already performed in a training loop itself. Test-Time Augmentation (TTA) was performed on the test images. Each test image was passed through the network 5 times: once as it was, unaugmented, and 4 further times with random augmentation (horizontal reflection, rotation within [−10°, +10°], and scaling within [0.95, 1.05]). This resulted in five 256-dimensional feature vectors which were averaged to obtain a single, more stable feature representation for each test image. TTA-averaged features were used uniformly across all three classifiers. Relative to single-pass (non-averaged) features, TTA improved KNN accuracy by 0.76 percentage points and left Decision Tree accuracy unchanged (0.00 percentage points), while SVM achieved 99.56% accuracy on TTA-averaged features. All three classifiers were therefore evaluated on TTA-averaged test features throughout this study, while training used the single-pass (non-augmented) training features described above. TTA is applied only at inference time, consistent with its intended role as a test-time stabilization technique rather than a training-time procedure. All features were normalized using min-max scaling with statistics computed exclusively on the training set, and the same statistics were applied to the test set to prevent data leakage.

3.9 Binary Secretary Bird Optimization Algorithm for Feature Selection

3.9.1 Overview of SBOA

The Secretary Bird Optimization Algorithm (SBOA) is a nature-inspired metaheuristic proposed by Fu et al. (2024) (Fu et al., 2024), which mimics the hunting behavior of the secretary bird, a large terrestrial raptor known for its unique prey-stalking and striking behavior. In its original formulation, SBOA structures its search around three iteration-dependent stages: an early searching-for-prey stage, a mid-run consuming-prey stage, and a late-stage escaping strategy, with the active stage determined by the proportion of the total iteration budget elapsed. SBOA has exhibited competitive performance compared to well-known optimizers like Particle Swarm Optimization (PSO) (Mohamed et al., 2025), Grey Wolf Optimizer (GWO) (Mirjalili et al., 2014) and Whale Optimization Algorithm (WOA) (Mirjalili & Lewis, 2016) on standard benchmark functions (Fu et al., 2024). In this work, we adopt a simplified binary adaptation of SBOA, in which the algorithm's core exploration and exploitation behaviors are consolidated into two phases applied at every iteration rather than partitioned across the three iteration-dependent stages of the original formulation. This design choice reduces implementation complexity and computational overhead for the high-dimensional binary feature selection problem addressed here, while preserving the two fundamental search mechanisms, Lévy-flight-driven exploration and best-guided exploitation, that characterize the original algorithm. To the best of our knowledge, this is the first application of SBOA, in either its original or simplified form, to feature selection in histopathological classification tasks.

3.9.2 Binary Encoding and Transfer Function

Feature selection is inherently a binary optimization problem, each of the 256 feature dimensions is either retained or discarded. The continuous-valued SBOA formulation was adapted to the binary domain using the V-shaped transfer function proposed by Mirjalili and Lewis (Mirjalili & Lewis, 2013).

$$T(x) = \left|\frac{x}{\sqrt{1+x^2}}\right| \qquad (1)$$

A feature dimension j is selected if a uniformly sampled random value r ~ U(0,1) satisfies r < T(x_j). The V-shaped function is a deliberate design choice it produces a bit-flip probability that increases symmetrically with the magnitude of the continuous position value. Agents with strong positive or negative positions make confident selection decisions, while values near zero preserve diversity in the population. This is actually more suitable than the S-shaped transfer function for feature selection tasks, since it avoids the saturation problem at extreme values.

3.9.3 Fitness Function

For each candidate binary mask, a weighted fitness function is used to evaluate the balance between two competing objectives: classification accuracy and feature subset size.

$$F(mask) = \alpha \cdot \varepsilon(mask) + \beta\left(\frac{|mask|}{D}\right) \qquad (2)$$

where *ε(mask)* is the 5-fold cross-validated KNN classification error on the training set with only features selected by the mask, |mask| is the number of selected features, D=256 is the total feature dimensionality and α=0.99 and β=0.01 are the weighting coefficients. The 99:1 weighting ratio is intentional – accuracy is far more important than the number of features, but the small β penalty consistently pushes the algorithm towards more compact subsets when two candidate masks have comparable accuracy. This is what drives the about 90% feature reduction we observed in the results.

3.9.4 Search Phases

Each SBOA iteration consists of two sequential phases:
Phase 1, Exploration (Lévy-flight stalking): Each search agent updates its position using a Lévy-flight-scaled step directed toward a randomly selected peer agent:

$$X_i(t+1) = X_i(t) + 0.01 \cdot L(\beta) \odot \left(X_r(t) - X_i(t)\right) \quad (3)$$

where X_r is a randomly selected agent from the population, $\odot$ denotes element-wise multiplication, and *L(β)* is a

Lévy-distributed step generated via Mantegna's algorithm with stability index $\beta = 1.5$. The scale factor 0.01 controls step magnitude, preventing overshooting while maintaining meaningful displacement. The heavy-tailed nature of the Lévy distribution enables occasional large jumps that help escape local optima, something a standard Gaussian random walk cannot reliably do. The resulting continuous position is binarized using the V-shaped transfer function described in Section 3.9.2.

Phase 2, Exploitation (striking toward the global best): Each agent refines its position by moving toward the current global best solution found so far:

$$X_i(t+1) = X_i(t) + (r_1 + r_2) \odot \left(X_{best} - X_i(t)\right) \quad (4)$$

where r_1 and r_2 are independently sampled random vectors drawn from $U(0,1)^D$. Using two independent random vectors rather than a single combined vector introduces additional stochastic diversity in the step magnitude, consistent with the original SBOA formulation (Fu et al., 2024), effectively allowing each feature dimension to be attracted toward the global best with a different random weight. The result is binarized using the V-shaped transfer function. If the new position achieves a lower fitness value than the current position, the agent updates; otherwise it stays.

3.9.5 SBOA Configuration

SBOA was configured with a population of 20 agents and a maximum of 45 iterations. The iteration count was selected empirically based on convergence behavior observed during preliminary experiments; the fitness curve consistently stabilized before reaching iteration 45, confirming that additional iterations would not meaningfully improve the selected feature subset. The population was initialized with random binary vectors sampled from a Bernoulli distribution with p=0.5. In each iteration all agents perform the exploration phase followed by the exploitation phase. The global best solution is updated after both phases are completed. A safety constraint guarantees that at least one feature dimension remains selected at all times, if SBOA converges to an all-zero mask (which is rare but possible), the dimension with the lowest individual fitness is activated.

3.10 Classification stage

The three classifiers were trained and evaluated independently on the feature subset selected by SBOA, using the same TTA-averaged feature representation described in Section 3.8.

K-Nearest Neighbors (KNN): The value of k was selected from {1, 3, 5, 7, 9} via 5-fold cross-validation on the training set. KNN was additionally used as the fitness evaluator within SBOA, ensuring a consistent objective between the feature selection stage and the final classification stage.

Support Vector Machine (SVM): A multiclass SVM was implemented using error-correcting output codes with a one-vs-one coding scheme and a radial basis function (RBF) kernel. Hyperparameters (box constraint and kernel scale) were tuned via cross-validation grid search on the training set.

Decision Tree: The decision tree classifier was trained using Gini impurity as the split criterion, with a maximum of 50 splits and a minimum leaf size of 5. This classifier was included to assess whether the SBOA-selected features remain discriminative even under a simple, fully interpretable model, independent of the more complex decision boundaries learned by KNN and SVM.

3.11 Explanation with Grad-CAM

The trained ResNet50-SE network's classification decisions were visually explained through Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017). Grad-CAM identifies the gradient of the predicted class score in connection with the feature maps of a target layer (activation_49_relu), performs a global average pooling on these gradients to obtain importance weights per channel, and then combines the importance weights with the corresponding activation maps to produce a class-discriminative localization map: $L^c = \text{ReLU}(\Sigma_k\, \alpha_k{}^c \cdot A^k)$, where A^k is the k^{th} feature map and $\alpha_k{}^c$ is the importance weight for class c calculated from the pooled gradients. The resulting heatmap was upsampled to 224 × 224 pixels with bicubic interpolation and overlaid on the original image with 45% opacity.

3.12 MobileNetV2-SE

A MobileNetV2-SE as a lightweight version was also implemented to evaluate the suitability of the framework for resource-constrained clinical settings, using MobileNetV2 [29] as the backbone. MobileNetV2 has about 3.4 million parameters, 87% less than ResNet50, and all layers were kept frozen and not fine-tuned then SE attention block was applied on the 1,280-dimensional global average pooling output with a squeeze dimension of 128 (reduction ratio about 10). All other components of the pipeline were kept the same as in the full ResNet50-SE framework, including SBOA feature selection, the three classifiers, TTA and Grad-CAM.

3.13 Statistical Robustness Procedure

To evaluate the reproducibility of the proposed framework, the entire pipeline including SBOA feature selection and classifier training was independently repeated five times with different random seeds (42, 7, 99, 2024, 555). For each seed, we gather all evaluation metrics and compute the mean and standard deviation over the five runs for each classifier.

3.14 Metrics for Evaluation

For each class c, class values of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were evaluated the proposed framework using the metrics computed on the test set shown in Table 3:

Table 3: The performance metric

Measure	Formula
Accuracy	(TP + TN) / (TP + FP+ TN + FN)
Specificity	TN /(TN + FP)
Precision (per class)	TP / (TP + FP)
Recall / Sensitivity	TP / (TP + FN)

F1-Score	2 × TP / (2 × TP + FP + FN)

3.15 Summary of the Proposed Framework

Table 4 describes the complete configuration of the proposed framework, covering dataset composition, network architecture, training setup, feature selection settings, classifier configurations, and evaluation protocol, in a single reference that supports precise reproduction of the reported experiments.

Table 4: Summary of framework components and configurations.

Component	Method / Setting
Dataset	LC25000 (lung only, 3 classes, 15,000 images)
Split	70% train , 15% val , 15% test (manual stratified)
Backbone	ResNet50 (ImageNet pre-trained), selective fine-tuning
Trainable parameters	~16M (res4f, res5a, res5b, res5c unfrozen)
Attention	Squeeze-and-Excitation (reduction ratio = 8)
Feature vector	256-dim (relu_feat layer)
Training optimizer	Adam, LR = 1×10^{-4}, piecewise decay (0.2 / 5 epochs)
Test-Time Augmentation	5-pass averaging
Feature selection	Binary SBOA (pop = 20, iter = 45)
Transfer function	V-shaped (Mirjalili & Lewis, 2013)
Fitness evaluator	5-fold CV KNN (k = 5, Euclidean)
Classifiers	KNN, SVM (RBF), Decision Tree (Gini)
Explainability	Grad-CAM (activation_49_relu layer)
Lightweight variant	MobileNetV2-SE (3.4M params, fully frozen)
Robustness protocol	5 independent seeds (42, 7, 99, 2024, 555)
Implementation	MATLAB R2022a, CPU execution

4. RESULTS

This section presents the experimental results of the proposed framework on the LC25000 lung histopathology dataset. Results are reported for the primary ResNet50-SE configuration with Binary SBOA feature selection and Test-Time Augmentation (Section 4.1), followed by per-class performance analysis (Section 4.2), ROC Curve Analysis (Section 4.3), an ablation study isolating the contribution of each framework component (Section 4.4), a statistical robustness analysis across five random seeds (Section 4.5), a comparison against a lightweight MobileNetV2-SE backbone (Section 4.6), and a comparison against state-of-the-art methods on the same dataset (Section 4.7) and Explainability analysis (Grad-Cam) (Section 4.8).

4.1 Overall Classification Performance

Table 5 shows the classification results of all three classifier methods (KNN, SVM, Decision Tree) trained on the Binary SBOA selected feature subset, for the best performing run (seed = 42). The K-Nearest Neighbour achieved the highest accuracy of 99.78% with a precision, recall of 99.78%, Specificity of 99.89% and F1-score of 99.78%. In this run 99.56% accuracy for SVM and 98.80% accuracy for Decision Tree were obtained. For this seed, the original feature vector was reduced from 256 dimensions to 25 features using the binary SBOA, a reduction of 90.2%.

Table 5: Overall classification performance of ResNet50-SE + Binary SBOA + TTA (seed = 42).

Classifier	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)	Features Selected
KNN (k=5)	99.78	99.78	99.78	99.89	99.78	25/256
SVM (RBF)	99.56	99.56	99.56	99.78	99.56	25/256
Decision Tree	98.80	98.82	98.80	99.40	98.80	25/256

The visual comparison of three classifiers over six evaluation metrics is shown in Figure 3. KNN consistently obtained the highest scores for all metrics with the smallest performance gap in AUC, with all three classifiers exceeding 99.60%, indicating the high discriminative power of the feature subset selected by the SBOA regardless of the classifier. The most important difference between classifiers are in Accuracy, Precision and F1-Score, where KNN (99.78%) is ahead of SVM (99.56%) by 0.22 percentage points and Decision Tree (98.80%) by 0.98 percentage points. All classifiers have high specificity values, showing low false-positive rates for all three lung cancer subtypes.

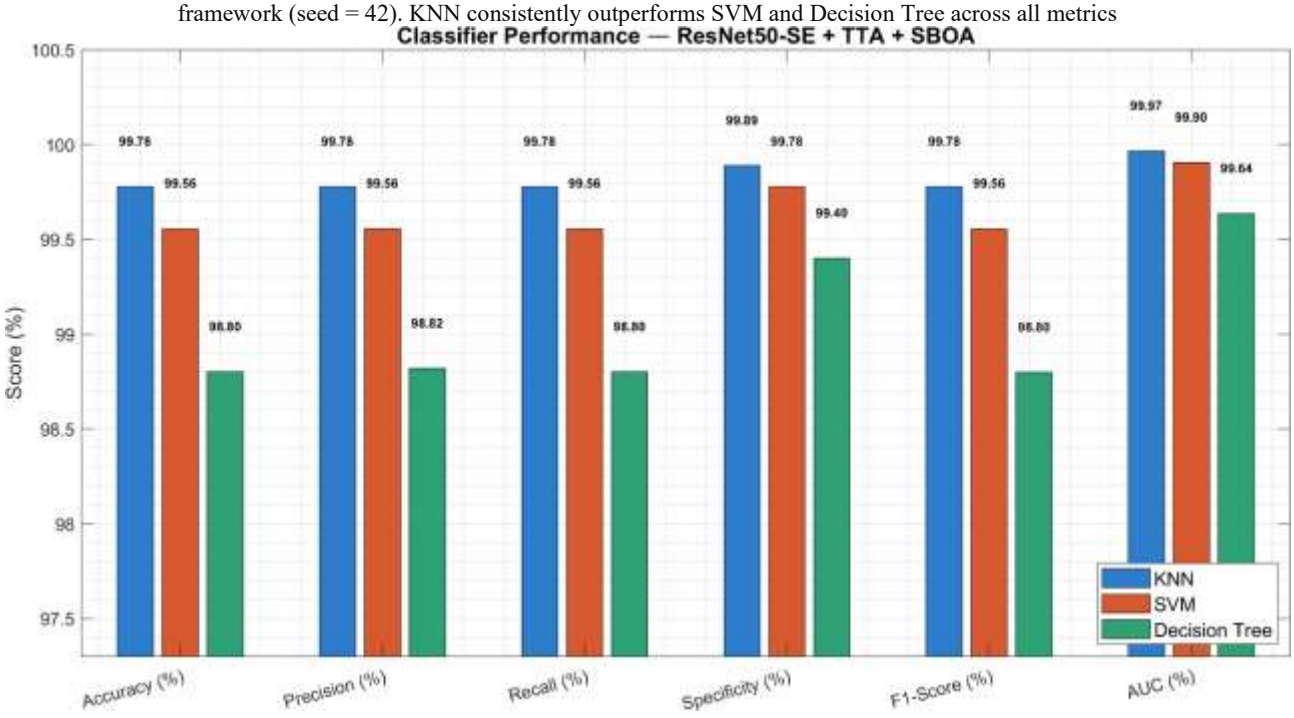
4.2 Per-Class Performance Analysis

Table 6 shows the per-class classification results obtained by the three classifiers. For the Benign class, the three models had a high result, with both KNN and SVM having 100% precision, recall and F1 score for this class. Adenocarcinoma and Squamous Carcinoma cells showed a slightly lower but still high performance on all the classifiers, consistent with the inter-observer diagnostic difficulty reported in the histopathology literature for the differentiation of these two subtypes. For KNN, recall for Adenocarcinoma was 99.60% and precision for Squamous Cell Carcinoma was 99.60% Decision Tree had the lowest per-class value observed among all three classifiers with 99.59% precision for Adenocarcinoma and 97.27% for Squamous Cell Carcinoma.

Figure 4 shows the confusion matrices for the three classifiers as heatmaps. The matrices show the raw counts of predictions as well as the row-normalized percentages for each class. The image shows the distribution of misclassifications and complements the precision, recall and F1-score values reported in Table 6. For KNN, all errors are between Adenocarcinoma and Squamous Cell

Carcinoma (3 Adenocarcinoma samples misclassified as Squamous Cell Carcinoma, and 2 Squamous Cell Carcinoma samples misclassified as Adenocarcinoma). No Benign samples are misclassified in either direction.
SVM shows the same error pattern with a higher count (7 and 3 samples, respectively), and likewise no Benign-related errors. Decision Tree exhibits a different and more varied error pattern: in addition to confusion between the two malignant subtypes (21 Adenocarcinoma samples misclassified as Squamous Cell Carcinoma, and 3 Squamous Cell Carcinoma samples misclassified as Adenocarcinoma), 3 Adenocarcinoma samples were misclassified as Benign, the only instance across all three classifiers in which a malignant sample was assigned to the benign class. This suggests that while KNN and SVM restrict all residual error to distinguishing between the two malignant subtypes, Decision Tree's simpler decision boundary occasionally fails to separate malignant from benign tissue entirely, a more clinically significant category of error than subtype confusion alone.

Figure 3: Classifier performance comparison across six evaluation metrics for the proposed ResNet50-SE + Binary SBOA + TTA framework (seed = 42). KNN consistently outperforms SVM and Decision Tree across all metrics



.

Table 6: Per-class results ResNet50-SE + Binary SBOA + TTA (seed = 42, best run).

Classifier	Class	TP	FP	FN	TN	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
KNN (k=5)	Adenocarcinoma	747	2	3	1498	99.60	99.73	99.60	99.87	99.67
KNN (k=5)	Benign	750	0	0	1500	100.00	100.00	100.00	100.00	100.00
KNN (k=5)	Squamous Cell Carcinoma.	748	3	2	1497	99.73	99.60	99.73	99.80	99.67
SVM (RBF)	Adenocarcinoma	743	3	7	1497	99.07	99.60	99.07	99.80	99.33
SVM (RBF)	Benign	750	0	0	1500	100.00	100.00	100.00	100.00	100.00
SVM (RBF)	Squamous Cell Carcinoma.	747	7	3	1493	99.60	99.07	99.60	99.53	99.34
Decision Tree	Adenocarcinoma	726	3	24	1497	96.80	99.59	96.80	99.80	98.17
Decision Tree	Benign	750	3	0	1497	100.00	99.60	100.00	99.80	99.80
Decision Tree	Squamous Cell Carcinoma.	747	21	3	1479	99.60	97.27	99.60	98.60	98.42

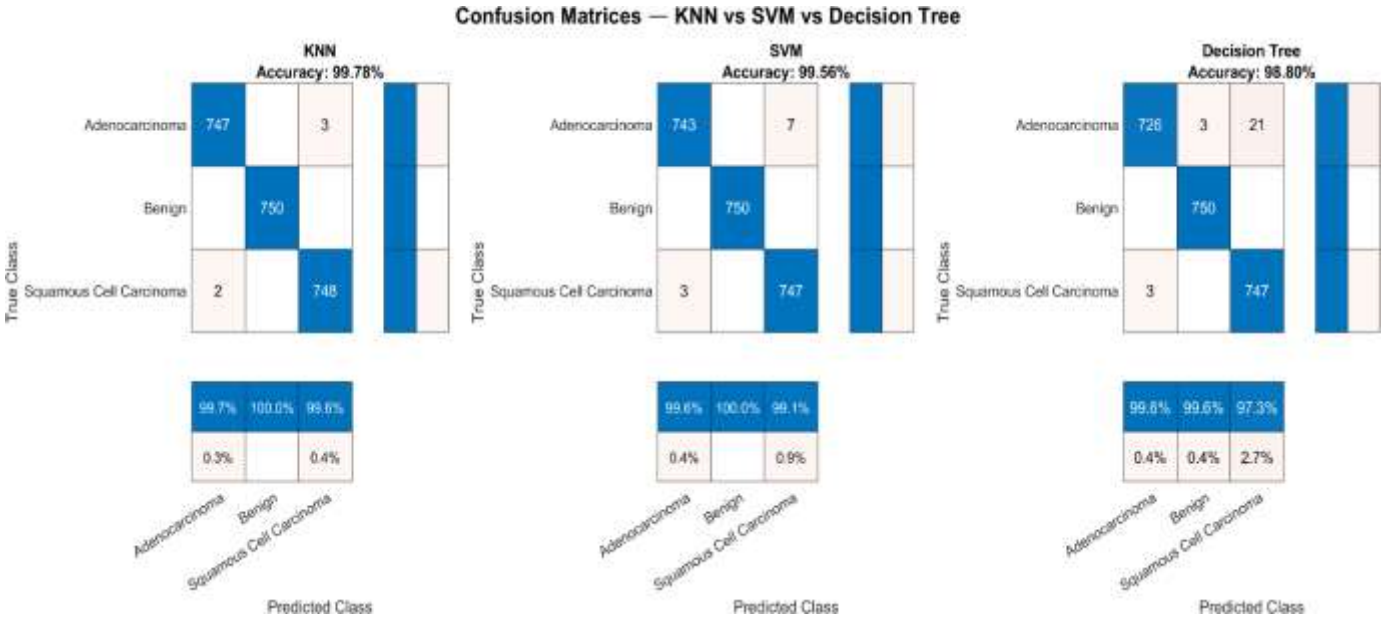


Figure4 4: Confusion matrices for KNN, SVM, and Decision Tree on the Binary SBOA-selected feature subset (seed = 42), showing raw counts (top) and row-normalized percentages (bottom) for each class.

4.3 ROC Curve Analysis

Figure 5 presents receiver operating characteristic (ROC) curves for all three classifiers on the three classification categories, and enlarged views of the top-left corner to distinguish between near-ceiling performance. KNN achieved the best mean AUC (99.97%) with per class AUC of 99.97% (Adenocarcinoma), 100.00% (Benign), 99.93% (Squamous Cell Carcinoma). The Decision Tree gave the lowest average AUC of 99.64% (99.19%, 99.93% and 99.79% respectively) while SVM gave a similar average AUC of 99.90% (99.96, 100.00%, and 99.75% respectively).

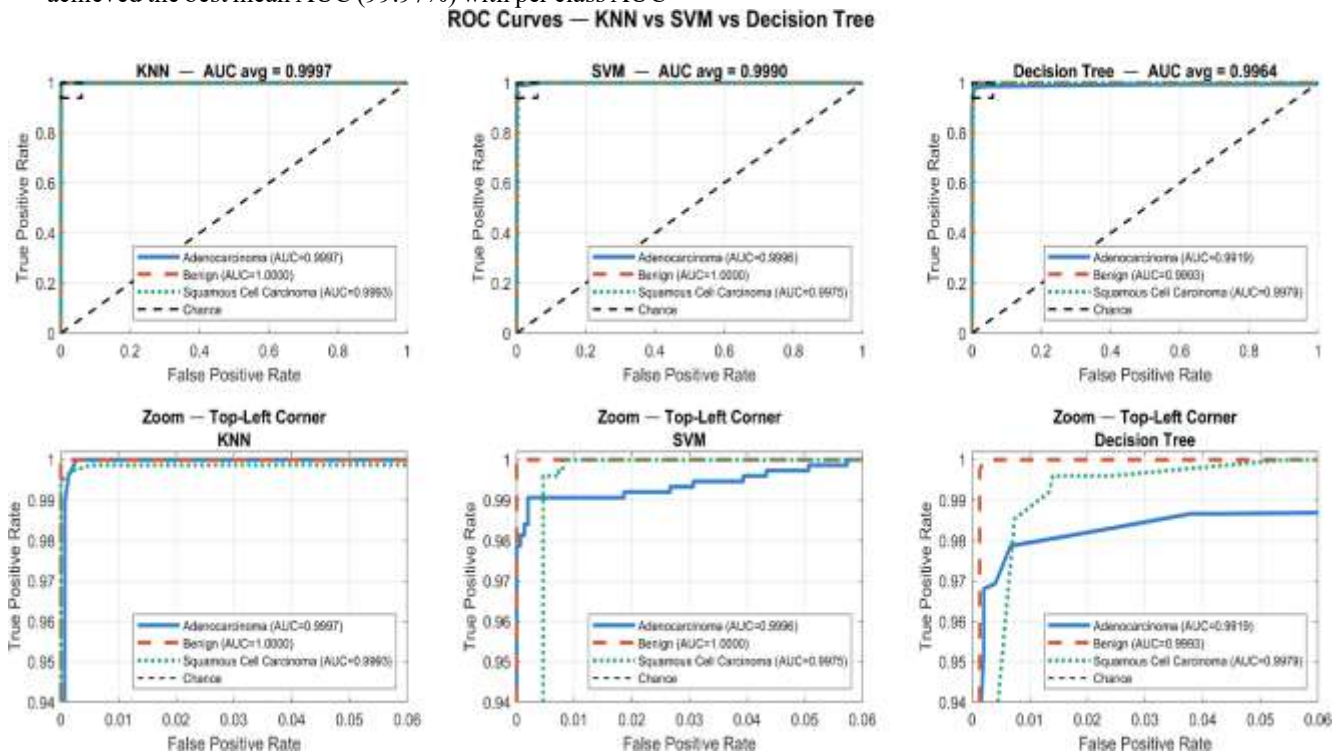


Figure 5: ROC curves for KNN, SVM, and Decision Tree on the Binary SBOA-selected feature subset (seed = 42), with zoomed insets of the top-left region.

The zoomed top-left views shows a qualitative difference in the smoothness of the curves between classifiers. KNN and SVM produce visually smooth ROC curves that sharply rise toward a true positive rate near 1.0 at very low false positive rates. The Decision Tree curve shows a visibly step-wise, jagged course trajectory in the same

region. This reflects the discrete, threshold-based decision mechanism of a single decision tree, which produces a comparatively small number of distinct score values, in contrast to the continuous distance and margin scores produced by KNN and SVM respectively. Across all three classifiers, the Benign class consistently achieved the highest or near-highest AUC, consistent with the perfect or near-perfect per-class precision and recall reported for this class in Table 7.

Table 7: Per-class AUC by classifier (seed = 42; source: ROC curve analysis).

Classifier	Adenocarcinoma (%)	Benign (%)	Squamous Cell Carcinoma (%)	Average AUC (%)
KNN	99.97	100.00	99.93	99.97
SVM	99.96	100.00	99.75	99.90
Decision Tree	99.19	99.93	99.79	99.64

4.4 Ablation Study

To isolate the contribution of each component of the proposed framework, three configurations were compared: The unoptimized ResNet50 baseline reflects a single training run, the SE+SBOA and SE+SBOA+TTA rows reflect the seed=42 run reported in Table 5, with five-seed robustness reported separately in Table 9. Results are summarized in Table 8.

The base ResNet50 backbone was enhanced with SE channel attention combined with Binary SBOA feature selection, resulting in an increase in KNN accuracy from 95.00% (single-run baseline) to 99.02%. Because SE attention and Binary SBOA feature selection are only ever evaluated jointly in this ablation design, this improvement cannot be decomposed into the individual contribution of each component. Isolating them requires two additional configurations (SE-only and SBOA-only), which we identify as future work (see Limitations).

TTA on top of this configuration gave a further improvement (albeit smaller): KNN improved by 0.76 percentage points (99.02% to 99.78%), SVM by 0.36 percentage points (99.20% to 99.56%), and Decision Tree remained unchanged, 98.80%.

4.5 Statistical Robustness Considering Random Seeds

To evaluate reproducibility, we repeated the entire pipeline five times with different random seeds (42, 7, 99, 2024 and 555) and results are summarised in Table 9. KNN was the most accurate and stable classifier across all five seeds with mean accuracy 99.63 ± 0.14%. The number of features selected by SBOA ranged from 21 to 32 (mean 27.6 ± 4.51) which corresponds to a mean feature reduction of 89.2%%. SVM had much higher variance (96.35 ± 2.96%) than KNN or Decision Tree (98.93 ± 0.09%), which means it is more sensitive to the particular feature subset chosen by SBOA in a run.

Table 8: Ablation study: the contribution of SE attention, Binary SBOA, and TTA.

Configuration	SE	SBOA	TTA	Classifier	Accuracy (%)	F1-Score (%)	AUC (%)	TTA Gain
ResNet50 (unoptimized baseline)	×	×	×	KNN	95.00	95.00	97.00	-
ResNet50 (unoptimized baseline)	×	×	×	SVM	96.00	96.00	98.00	-
ResNet50 (unoptimized baseline)	×	×	×	Decision Tree	93.00	93.00	95.00	-
ResNet50-SE + SBOA (no TTA)	✓	✓	×	KNN	99.02	99.01	99.50	-
ResNet50-SE + SBOA (no TTA)	✓	✓	×	SVM	99.20	99.19	99.60	-
ResNet50-SE + SBOA (no TTA)	✓	✓	×	Decision Tree	98.80	98.79	99.20	-
ResNet50-SE + SBOA + TTA (proposed)	✓	✓	✓	KNN	99.78	99.78	99.97	+0.76
ResNet50-SE + SBOA + TTA (proposed)	✓	✓	✓	SVM	99.56	99.56	99.90	+0.36
ResNet50-SE + SBOA + TTA (proposed)	✓	✓	✓	Decision Tree	98.80	98.80	99.64	+0.00

Table 9: Statistical robustness across five random seeds: ResNet50-SE + Binary SBOA + TTA.

Seed	Classifier	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)	AUC (%)	Features
42	KNN	99.78	99.78	99.78	99.89	99.78	99.97	25
42	SVM	99.56	99.56	97.56	98.78	97.56	99.90	25
42	Decision Tree	98.80	98.82	98.80	99.40	98.80	99.64	25
7	KNN	99.60	99.60	99.60	99.80	99.60	99.96	30
7	SVM	99.56	99.56	99.56	99.78	99.56	99.89	30
7	Decision Tree	99.02	99.03	99.02	99.51	99.02	99.80	30
99	KNN	99.42	99.42	99.42	99.71	99.42	99.96	30
99	SVM	93.47	94.47	99.47	99.73	99.47	99.88	30
99	Decision Tree	99.02	99.02	99.51	99.51	99.02	99.63	30
2024	KNN	99.73	99.73	99.73	99.87	99.73	99.90	21
2024	SVM	94.71	94.71	94.71	97.36	94.71	98.43	21
2024	Decision Tree	98.89	98.90	98.89	99.44	98.89	99.81	21
555	KNN	99.64	99.64	99.64	99.82	99.64	100.00	32
555	SVM	94.47	94.47	99.47	99.73	94.47	99.96	32
555	Decision Tree	98.93	98.94	98.93	99.47	98.93	99.58	32
Mean ± SD	KNN	99.63 ± 0.14	99.63 ± 0.14	99.63 ±0.14	99.82 ±0.07	99.63 ±0.14	99.96 ±0.04	27.6 ±4.51
Mean ± SD	SVM	96.35 ±2.96	96.55 ±2.75	98.15 ±2.10	99.08 ±1.05	97.15 ±2.47	99.61 ±0.66	27.6 ±4.51
Mean ± SD	Decision Tree	98.93 ± 0.09	98.94 ±0.09	99.03 ±0.28	99.47 ±0.05	98.93 ±0.09	99.69 ±0.11	27.6 ±4.51

4.6 Backbone Comparison: ResNet50-SE versus MobileNetV2-SE

To evaluate the framework's applicability under resource-constrained deployment conditions, a lightweight variant using a fully frozen MobileNetV2 backbone (3.4 million parameters, 87% fewer than ResNet50's 25.6 million) was evaluated under the same SBOA and TTA pipeline. Table 10 compares both backbones across all three classifiers. ResNet50-SE outperformed MobileNetV2-SE in every metric and classifier, with accuracy increases from 2.18 percentage points (KNN) to 2.94 percentage points (SVM). Despite the lower accuracy, MobileNetV2-SE reached a similar feature reduction ratio (89.1% vs 90.2%), with good absolute performance (KNN accuracy = 97.60%, AUC= 99.41%), indicating that the framework could still be viable in situations where the computational resources do not allow the use of the larger ResNet50 backbone.

4.7 Comparison with State-of-the-Art Methods

Table 11 situates the proposed framework relative to previously published methods evaluated on the LC25000 lung subset. The methods compared in Table 11 were evaluated under differing, and in most cases unreported, train/test split ratios, preprocessing pipelines, and augmentation strategies. Because this information is not consistently available in the source publications, the comparison should be read as contextual positioning within the literature rather than a controlled head-to-head benchmark. While several prior methods report marginally higher single-metric accuracy (e.g., EfficientNetV2-L at 99.97%), the proposed ResNet50-SE framework is, to the authors' knowledge, the only method in this comparison that simultaneously incorporates metaheuristic feature selection, multi-classifier evaluation, and Grad-CAM-based explainability, in addition to competitive accuracy (99.63% ± 0.14% mean across five seeds; 99.78% best single run).

Table 10: Backbone comparison: MobileNetV2-SE versus ResNet50-SE, both with Binary SBOA and TTA (seed = 42).

Metric	KNN MobileNetV2	KNN ResNet50	Δ	SVM MobileNetV2	SVM ResNet50	Δ	DT MobileNetV2	DT ResNet50
Accuracy (%)	97.60	99.78	+2.18	96.62	99.56	+2.94	95.29	98.80
Precision (%)	97.61	99.78	+2.17	96.73	99.56	+2.83	95.27	98.82
Recall (%)	97.60	99.78	+2.18	96.62	99.56	+2.94	95.29	98.80
Specificity (%)	98.80	99.89	+1.09	98.31	99.78	+1.47	97.64	99.40
F1-Score (%)	97.60	99.78	+2.18	96.61	99.56	+2.95	95.28	98.80

AUC (%)	99.41	99.97	+0.56	99.26	99.90	+0.64	99.27	99.64

Table 11: Comparison with state-of-the-art methods on the LC25000 lung dataset.

Method	Year	Backbone	Accuracy (%)	Feature Sel.	Explain.	Multi-Clf.
Custom CNN (Hatuwal & Thapa, 2020)	2020	Custom CNN	97.20	-	-	-
Lightweight CNN (Masud et al., 2021)	2021	Custom CNN	96.33	-	-	-
ResNet18 (Wahid et al., 2023)	2021	ResNet18	98.82	-	-	-
EfficientNet-B0 (Mishra & Agarwal, 2022)	2022	EfficientNet-B0	98.67	-	-	-
EfficientNetV2-L (Tummala et al., 2023)	2023	EfficientNetV2-L	99.97	-	-	-
Self-ONN (Said et al., 2024)	2024	Self-ONN	99.74	-	-	-
ResNet50 Transfer Learning (Adekunle et al., 2025)	2025	ResNet50	98.80	-	-	-
DenseNet201 + FSRCNN + SVM (Patharia et al., 2025)	2025	DenseNet201	98.00	-	-	-
ResNet34 + FastAI-2 (Saxena et al., 2025)	2025	ResNet34	99.78	-	-	-
DPCSE-Net (Dual-path + SE) (AlShehri, 2025)	2025	Dual CNN+SE	99.88%	-	Grad-CAM	-
DenseNet121 + GA + KNN (A. Roy et al., 2025)	2025	DenseNet121	99.75	GA	-	-
Proposed (MobileNetV2-SE+SBOA+TTA)	2026	MobileNetV2-SE	97.60	89.1%	Grad-CAM	✓
Proposed (ResNet50-SE+SBOA+TTA)	2026	ResNet50-SE	99.78	90.2%	Grad-CAM	✓

4.8 Explainability Analysis (Grad-CAM)

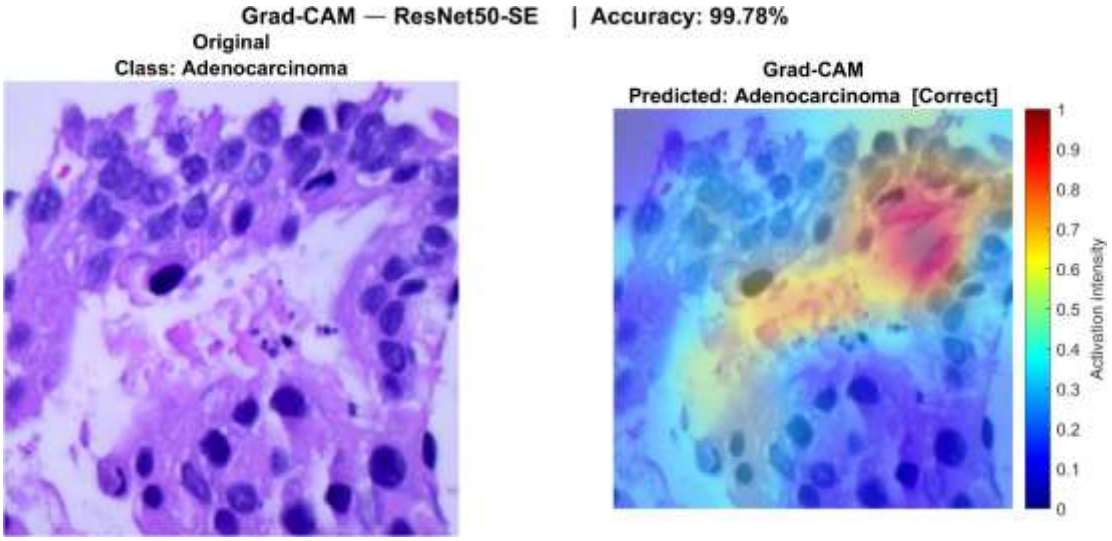
To assess whether the trained ResNet50-SE network bases its predictions on diagnostically meaningful tissue regions rather than incidental image artifacts, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to representative test samples from each of the three lung tissue classes, following the procedure described in Section 3.11. Figure 6 presents Grad-CAM heatmaps overlaid on original histopathological images for (a) Adenocarcinoma, (b) Benign tissue, and (c) Squamous Cell Carcinoma, alongside the network's corresponding predicted class for each sample.

Because Grad-CAM is computed on the ResNet50-SE network's own softmax classification head (Section 3.11), these visualizations explain the CNN's internal classification decision and the deep feature representation from which SBOA subsequently selects features. They do not directly explain the final decision of the downstream KNN, SVM, or Decision Tree classifiers reported in Section 4.1 onward.

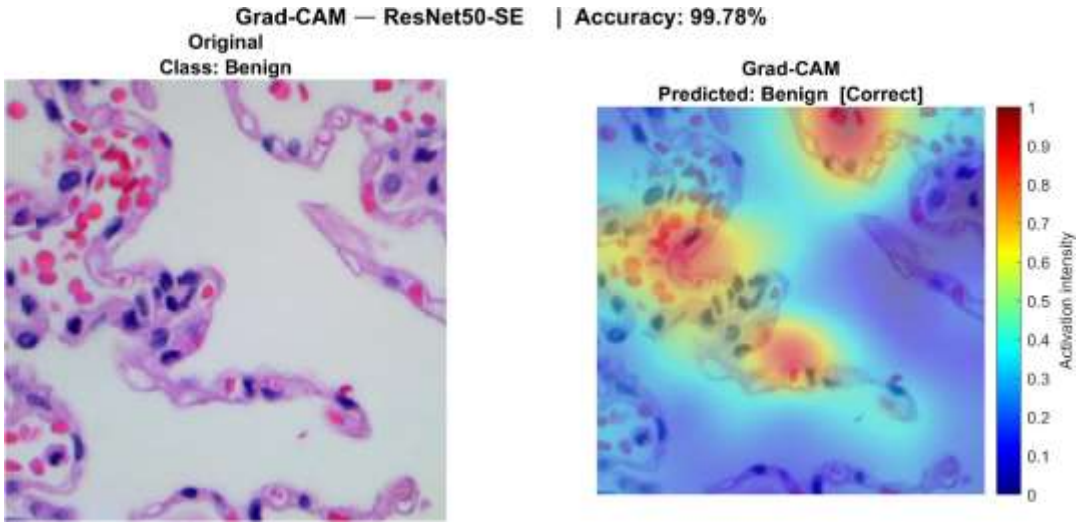
The following qualitative analysis is exploratory and has not been validated against expert pathologist annotation, conclusions about clinical relevance should be treated as preliminary.

For the Adenocarcinoma sample, activation is concentrated on a single dominant cellular cluster rather than distributed across the tissue field, suggesting the network identifies a specific region of abnormal cellular architecture as the primary basis for its prediction. For the Squamous Cell Carcinoma sample, two concentrated activation regions correspond to areas of dense nuclear clustering, consistent with the compact cellular sheets characteristic of this subtype. Notably, for the Benign sample, activation is not diffusely distributed across the tissue field, as might be expected for a class defined by the absence of malignant morphology; instead, it concentrates on two distinct regions coinciding with clusters of red blood cells and nuclei within the alveolar structure. This suggests the network's Benign classification may be driven in part by the presence of normal vascular and cellular content rather than solely by the absence of malignant features, an observation that warrants further investigation using region-level annotation from a pathologist.

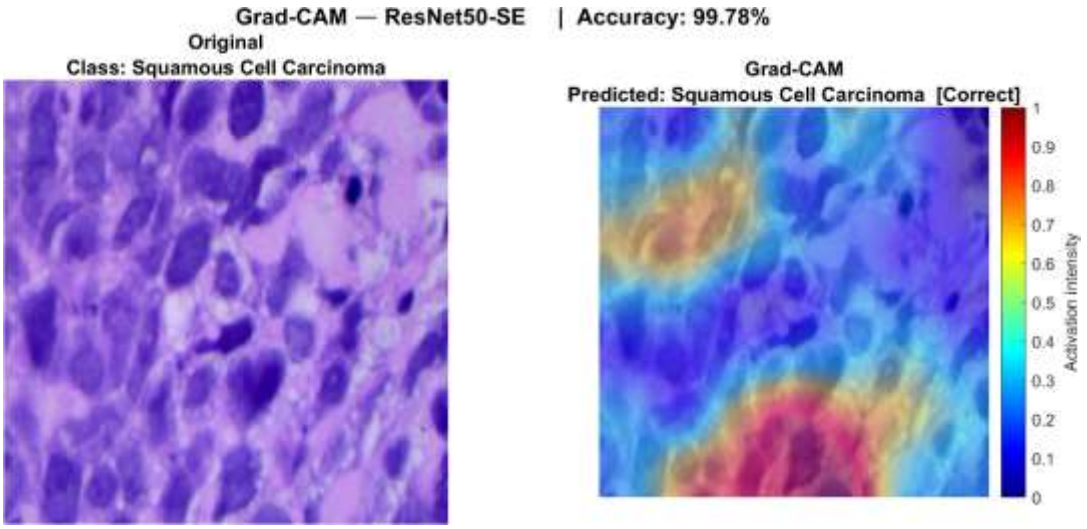
This qualitative analysis indicates that the network's attention is spatially concentrated and class-dependent rather than uniform or background-driven, providing preliminary support for the interpretability of its predictions. However, the finding that Benign-class activation coincides with vascular content rather than a more general tissue pattern highlights the importance of the pathologist-validation study identified as future work in Section 6, since visual plausibility alone does not confirm that the network's reasoning aligns with actual diagnostic criteria.



(a) Adenocarcinoma Grad-CAM



(b) Benign Grad-CAM



(c) Squamous Cell Carcinoma Grad-CAM

Figure 6: Grad-CAM visualizations for representative test samples of each class. Warmer colors (red/yellow) indicate parts of the image contributing most strongly to the network's classification decision, while cooler colors (blue) indicate parts with minimal contribution

5. DISCUSSION

5.1 Interpretation of Key Findings

The ablation results in Section 4.4 indicate that the combination of SE channel attention and Binary SBOA feature selection accounts for the majority of the improvement over the unoptimized baseline, raising KNN accuracy from 95.00% (single run) to 99.02 prior to the introduction of Test-Time Augmentation. This joint improvement is larger than the subsequent TTA gain; however, because SE and SBOA were not evaluated separately, we cannot attribute it to channel-wise recalibration or feature-subset reduction individually, this requires the two-factor ablation identified as future work.

TTA subsequently provided a smaller but consistent additional gain for KNN (+0.76 percentage points) and SVM (+0.36 percentage points), while leaving Decision Tree performance unchanged. This pattern is consistent with the underlying decision mechanisms of each classifier: KNN and SVM both rely on continuous distance or margin computations that are sensitive to small shifts in the feature vector, whereas a decision tree partitions the feature space using discrete, axis-aligned thresholds that are comparatively insensitive to the small perturbations introduced by feature averaging.

K-Nearest Neighbor was consistently the most accurate and stable classifier across all five random seeds (99.63% ± 0.14%) on the other hand, SVM was exhibited substantially higher variance (96.35% ± 2.96%). This difference is attributable to the interaction between the SBOA selected feature subset and each classifier's decision mechanism. Because SBOA selects a different feature subset in each run, depending on the random initialization of the search population, the RBF kernel's similarity computation which depends on the specific geometry of the selected feature space. it is more sensitive to these subset changes than KNN's direct distance computation in the same reduced space. This finding suggests that classifier choice should not be treated as independent of the feature selection method in future SBOA-based pipelines, and that KNN may be the more robust default pairing for Secretary Bird Optimization' selected features specifically.

5.2 Positioning Relative to Prior Work

As shown in Table 11, several prior methods evaluated on the LC25000 lung subset report marginally higher single-metric accuracy than the proposed framework, most notably EfficientNetV2-L at 99.97%. However, none of the compared methods include metaheuristic feature selection, multi-classifier evaluation and Grad-CAM based explainability, and the vast majority report a single accuracy number from a single backbone with no interpretability or feature-efficiency analysis. Instead, the proposed framework achieves competitive accuracy (99.63% ± 0.14% mean across five seeds, 99.78% best single run) while reducing the feature representation by approximately 90% and providing visual explanations of individual predictions. This is, to our knowledge, the first reported application of the Secretary Bird Optimization Algorithm to deep feature selection specifically in histopathological image classification (as distinct from its use for classifier hyperparameter/kernel optimization in Wang et al., 2025), which distinguishes this work from prior metaheuristic-based approaches on this dataset that rely on longer-established optimizers such as Genetic Algorithms.

5.3 Clinical and Practical Implications

The substantial feature reduction achieved by Binary SBOA (89–90% across both backbones) has practical implications beyond classification accuracy. A trained classifier operating on 21–40 features rather than the full 256-dimensional representation requires less storage and permits faster inference at deployment time, which may be relevant for integration into existing laboratory information systems where computational resources are shared across multiple diagnostic workflows. The MobileNetV2-SE variant further demonstrates that the proposed pipeline remains functional, though with a measurable accuracy trade-off of approximately 2 to 3 percentage points, when the backbone is constrained to a fraction of ResNet50's parameter count, indicating a viable deployment path for resource-limited clinical settings.

The Grad-CAM visualizations generated for representative test samples indicate that the network's attention aligns with regions of glandular formation in adenocarcinoma and keratinization patterns in squamous cell carcinoma, consistent with the morphological features pathologists use for manual subtyping. This alignment provides a preliminary, qualitative indication that the underlying deep feature representation is grounded in diagnostically plausible regions. It does not constitute a validated explanation of the final KNN/SVM/Decision Tree classification decision, nor a substitute for formal validation against expert pathologist annotation, both identified as priority directions for future work.

6. CONCLUSION

This paper presented an explainable hybrid deep learning framework for lung cancer histopathological image classification, integrating a ResNet50 backbone with Squeeze-and-Excitation attention, a Binary Secretary Bird Optimization Algorithm for feature selection, Test-Time Augmentation for inference-time stabilization, a multi-classifier evaluation stage, and Grad-CAM for visual explainability. On the LC25000 lung dataset, the proposed framework achieved a mean classification accuracy of 99.63% ± 0.14% across five random seeds using a K-Nearest Neighbor classifier, with a best single-run accuracy of 99.78%, while reducing the original 256 dimensional deep feature representation by approximately 90%. An ablation study confirmed that SE attention and Binary SBOA feature selection together account for the majority of the improvement over an unoptimized baseline, with Test-Time Augmentation. A lightweight MobileNetV2-SE variant demonstrated that the framework remains viable, with a measurable but modest accuracy trade-off, under substantially reduced computational budgets. This work presents the first

application of the Secretary Bird Optimization Algorithm to deep feature selection in histopathological image classification, and the first framework to combine metaheuristic feature selection, multi-classifier comparison, and gradient-based explainability within a single lung cancer histopathology classification pipeline.

Several limitations of this study should be acknowledged alongside these contributions. All experiments were conducted on a single dataset (LC25000), acquired under a consistent staining and imaging protocol, and generalization to histopathological images acquired under different staining protocols, scanners, or institutional practices was not evaluated. In addition, because the publicly distributed LC25000 archive does not include parent-image or derivative-family identifiers, the train/validation/test split used here is defined at the image level. We cannot rule out that augmented derivatives of the same original slide are distributed across partitions, which may inflate the reported accuracy relative to a strictly derivative-free evaluation; this constraint is shared with the large majority of prior LC25000 classification studies. The ablation study (Section 4.4) evaluates SE attention and Binary SBOA feature selection only in combination; a full factorial ablation isolating each component's individual contribution (ResNet50+SE only, ResNet50+SBOA only) was not performed and is identified as future work. Grad-CAM explains the CNN's own classification head rather than the downstream KNN/SVM/Decision Tree decision used for the final reported results. Model-agnostic feature-attribution methods (e.g., SHAP, LIME) applied to the SBOA-selected feature subset are identified as future work to directly address this gap.

The results support the broader premise that combining bio-inspired optimization with attention-modulated deep features and post-hoc explainability can yield diagnostic systems that are simultaneously accurate, computationally efficient, and clinically interpretable. The reported accuracy figures reflect within-dataset, same-protocol performance on LC25000 and should not be interpreted as evidence of cross-institutional or cross-protocol generalization until the external validation described below is completed. Future work will pursue three directions arising directly from the limitations above:

(1) external validation on independent histopathology datasets acquired under different staining and imaging protocols, such as LungHist700 or TCGA-LUAD, to assess cross-institutional generalization, (2) a formal validation study in which pathologists assess the clinical relevance of Grad-CAM attention regions against expert-annotated regions of interest, extending the preliminary explainability analysis presented here into a quantitatively validated measure of clinical trust, and (3) a full factorial ablation isolating the individual contributions of SE attention and Binary SBOA feature selection, evaluating ResNet50+SE-only and ResNet50+SBOA-only configurations separately, to determine each component's effect independently of their combination.

REFERENCES

Adekunle, O. O., Akinyemi, J. D., Ladoja, K. T., & Onifade, O. F. W. (2025). Lung cancer classification from ct images using resnet. *ArXiv Preprint ArXiv:2510.16310*.

Ahmed, M. S., Iqbal, K. N., & Alam, M. G. R. (2023). Interpretable lung cancer detection using explainable AI methods. *2023 International Conference for Advancement in Technology (ICONAT)*, 1–6.

AlShehri, H. (2025). Dual-Path Convolutional Neural Network with Squeeze-and-Excitation Attention for Lung and Colon Histopathology Classification. *Journal of Imaging*, *11*(12), 448.

Anjum, S., Ahmed, I., Asif, M., Aljuaid, H., Alturise, F., Ghadi, Y. Y., & Elhabob, R. (2023). Lung cancer classification in histopathology images using multiresolution efficient nets. *Computational Intelligence and Neuroscience*, *2023*(1), 7282944.

Borkowski, A. A., Bui, M. M., Thomas, L. B., Wilson, C. P., DeLand, L. A., & Mastorides, S. M. (2019). *Lung and Colon Cancer Histopathological Image Dataset (LC25000)*. http://arxiv.org/abs/1912.12142

Fu, Y., Liu, D., Chen, J., & He, L. (2024). Secretary bird optimization algorithm: a new metaheuristic for solving global optimization problems. *Artificial Intelligence Review*, *57*(5), 123.

Hatuwal, B. K., & Thapa, H. C. (2020). Lung cancer detection using convolutional neural network on histopathological images. *Int. J. Comput. Trends Technol*, *68*(10), 21–24.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141.

Huang, K. A., Alkadri, A., Prakash, N., & Prakash, N. S. (2025). Employing squeeze-and-excitation architecture in a fine-tuned convolutional neural network for magnetic resonance imaging tumor classification. *Cureus*, *17*(3).

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88.

Masud, M., Sikder, N., Nahid, A.-A., Bairagi, A. K., & AlZain, M. A. (2021). A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework. *Sensors*, *21*(3), 748. https://doi.org/10.3390/s21030748

Mirjalili, S., & Lewis, A. (2013). S-shaped versus V-shaped transfer functions for binary particle swarm optimization. *Swarm and Evolutionary Computation*, *9*, 1–14.

Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software*, *95*,

51–67.

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, *69*, 46–61.

Mishra, S., & Agarwal, U. (2022). Lung cancer detection (LCD) from histopathological images using fine-tuned deep neural network. *Proceedings of the International Conference on Intelligent Computing, Communication and Information Security*, 249–260.

Mohamed, A. A. A. (2023). *A Hybrid Machine Learning Techniques with Deep Neural Network Model for Colon Cancer Diagnosis*.

Mohamed, A. A. A., Rahebi, M., Hançerlioğulları, A., & Rahebi, J. (2025). An approach based on convolutional neural network and ACO-PSO for colon cancer disease diagnosis. *Politeknik Dergisi*, *28*(2), 649–659.

Paech, D. C., Weston, A. R., Pavlakis, N., Gill, A., Rajan, N., Barraclough, H., Fitzgerald, B., & Van Kooten, M. (2011). A systematic review of the interobserver variability for histology in the differentiation between squamous and nonsquamous non-small cell lung cancer. *Journal of Thoracic Oncology*, *6*(1), 55–63.

Patharia, P., Rao, B. S. P., Sethy, P. K., Ratha, A. K., & Nanthaamornphong, A. (2025). Enhanced Diagnosis of Lung and Colon Cancer Severity Through Deep Feature Analysis Using DenseNet201 and SVM With Histopathological Images: A Super-Resolution Approach. *Cancer Reports*, *8*(12), e70439.

Roy, A. G., Navab, N., & Wachinger, C. (2018). Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 421–429.

Roy, A., Saha, P., Gautam, N., Schwenker, F., & Sarkar, R. (2025). Adaptive genetic algorithm based deep feature selector for cancer detection in lung histopathological images: A. Roy et al. *Scientific Reports*, *15*(1), 4803.

Said, M. M. R., Islam, M. S. Bin, Sumon, M. S. I., Vranic, S., Al Saady, R. M., Alqahtani, A., Chowdhury, M. E. H., & Pedersen, S. (2024). Innovative deep learning architecture for the classification of lung and colon cancer from histopathology images. *Applied Computational Intelligence and Soft Computing*, *2024*(1), 5562890.

Saxena, P., Singh, S. K., Rashid, M., Alshamrani, S. S., & Alnfiai, M. M. (2025). Efficient deep learning model for classifying lung cancer images using normalized stain agnostic feature method and FastAI-2. *PeerJ Computer Science*, *11*, e2903.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, 618–626.

Siegel, R. L., Giaquinto, A. N., & Jemal, A. (2024). Cancer statistics, 2024. *CA: A Cancer Journal for Clinicians*, *74*(1), 12–49.

Štifanić, J., Štifanić, D., Anđelić, N., & Car, Z. (2025). Explainable AI for oral cancer diagnosis: Multiclass classification of histopathology images and Grad-CAM visualization. *Biology*, *14*(8), 909.

Sung, H., Filho, A. M., Laversanne, M., Ferlay, J., Siegel, R. L., Soerjomataram, I., Jemal, A., & Bray, F. (2026). Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. *CA: A Cancer Journal for Clinicians*, *76*(4), e70090. https://doi.org/https://doi.org/10.3322/caac.70090

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, *35*(5), 1299–1312.

Talaat, F. M., Gamel, S. A., El-Balka, R. M., Shehata, M., & ZainEldin, H. (2024). Grad-CAM enabled breast cancer classification with a 3D inception-ResNet V2: Empowering radiologists with explainable insights. *Cancers*, *16*(21), 3668.

Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(11), 4793–4813.

Tummala, S., Kadry, S., Nadeem, A., Rauf, H. T., & Gul, N. (2023). An explainable classification method based on complex scaling in histopathology images for lung and colon cancer. *Diagnostics*, *13*(9), 1594.

Wahid, R. R., Nisa', C., Amaliyah, R. P., & Puspaningrum, E. Y. (2023). Lung and colon cancer detection with convolutional neural networks on histopathological images. *AIP Conference Proceedings*, *2654*(1), 20020.

Wang, L., Sheng, J., Zhang, Q., Song, Y., Zhang, Q., Wang, B., & Zhang, R. (2025). Diagnosis of Alzheimer's disease using FusionNet with improved secretary bird optimization algorithm for optimal MK-SVM based on imaging genetic data. *Cerebral Cortex*, *35*(2), bhae498.

Zhang, H., & Ogasawara, K. (2023). Grad-CAM-based explainable artificial intelligence related to medical text processing. *Bioengineering*, *10*(9), 1070.

Zhou, J., Xu, Y., Liu, J., Feng, L., Yu, J., & Chen, D. (2024). Global burden of lung cancer in 2022 and projections to 2050: Incidence and mortality estimates from GLOBOCAN. *Cancer Epidemiology*, *93*, 102693.