# Lung and colon cancer detection with InceptionResNetV2: a transfer learning approach

# Detección de cáncer de pulmón y colon con InceptionResNetV2: un enfoque de aprendizaje por transferencia

Ochoa-Ornelas, Raquel*[a], Gudiño-Ochoa, Alberto[b], García-Rodríguez, Julio Alberto[c] and Uribe-Toscano, Sofia[d]

[a] ROR Tecnológico Nacional de México - Instituto Tecnológico de Ciudad Guzmán • S-4687-2018 • 0000-0003-1824-5789 • 668976
[b] ROR Tecnológico Nacional de México - Instituto Tecnológico de Ciudad Guzmán • HDN-0235-2022 • 000-0002-2366-7452 • 1135216
[c] ROR Universidad de Guadalajara - Centro Universitario del Sur • LBI-3198-2024 • 0000-0003-0339-0545 • 367137
[d] ROR Universidad de Guadalajara - Centro Universitario del Sur • JRY-7988-2023 • 0009-0009-8325-8237

## Abstract

Lung and colon cancers are among the most common and lethal types of cancer. It is possible for individuals to develop both cancers simultaneously, as risk factors such as smoking, which is a leading cause of lung cancer, can also contribute to poor dietary habits, thereby increasing the risk of colon cancer. Traditionally, the diagnosis of these cancers relies on biopsies and subsequent laboratory analysis. This study proposes an Inception-ResNetV2-based model for accurately classifying lung and colon cancer using histopathological images. The dataset analyzed consists of 25,000 images categorized into five cancer types: Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, and Lung Squamous Cell Carcinoma. The developed approach achieved an accuracy of 99.15%. The model's performance is further enhanced through the use of Local Binary Patterns (LBP), improving both accuracy and computational efficiency. Additionally, the explainable AI method SHAP is employed to demonstrate the contribution of each feature to the predictive analysis

## Resumen

El cáncer de pulmón y colon son dos de los tipos de cáncer más comunes y letales. Es posible que una persona desarrolle ambos cánceres simultáneamente, ya que factores de riesgo como el tabaquismo, principal causa del cáncer de pulmón, también pueden contribuir a hábitos dietéticos inadecuados, lo que incrementa el riesgo de cáncer de colon. El diagnóstico de estos cánceres generalmente se basa en biopsias y su análisis en laboratorio. Este estudio propone un modelo basado en Inception-ResNetV2 para la clasificación precisa de cáncer de pulmón y colon en imágenes histopatológicas. El conjunto de datos analizado contiene 25,000 imágenes clasificadas en cinco tipos: Adenocarcinoma de colon, Tejido benigno de colon, Adenocarcinoma de pulmón, Tejido benigno de pulmón y Carcinoma de células escamosas de pulmón. El enfoque desarrollado alcanzó una precisión del 99.15%. La eficiencia del modelo se optimiza mediante patrones binarios locales (LBP), mejorando tanto la precisión como la eficiencia computacional. Además, el método de IA explicable SHAP se utiliza para mostrar la contribución de cada característica en el análisis predictivo.
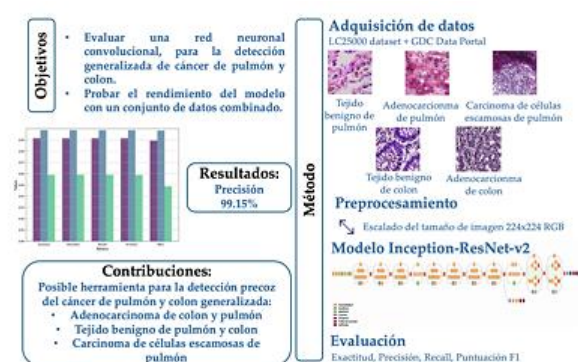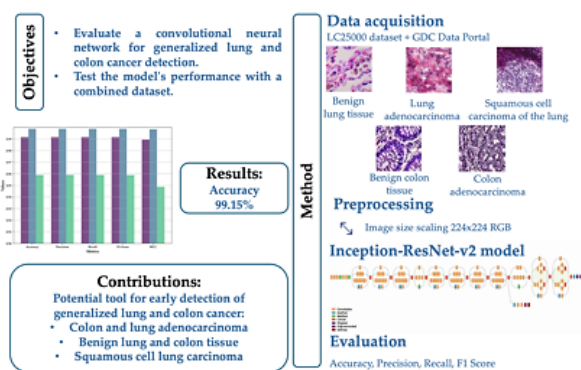
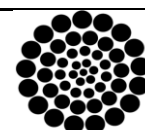**Lung cancer, Histopathological, InceptionResNet, Colon cancer**

**Cáncer de pulmón, Histopatólogicas, InceptionResNet, Cáncer de colon**

Article

## Introduction

Cancer stands as the second leading cause of death worldwide, with over 19 million new cases and approximately 10 million deaths reported in 2020 (Sung *et al.*, 2020). Characterized by genetic and metabolic abnormalities, cancer poses a substantial global health challenge, particularly with lung and colon cancers being among the most lethal. Cancerous cells demonstrate unchecked growth, genomic instability, and a high likelihood of metastasis, making these cancers especially difficult to treat in advanced stages (Kiri *et al.*, 2024).

Malignant tumors arise when damaged cells are not replaced by healthy ones, leading to abnormal proliferation that invades surrounding tissues. Consequently, lung and colon cancers are prevalent, contributing to 4.19 million cases and 2.7 million deaths in 2020 alone (Li *et al.*, 2021). Lung cancer alone accounts for 18.4% of cancer-related deaths, while colorectal cancer represents 9.2% (Cao *et al.*, 2020). Notably, risk factors such as smoking, obesity, alcohol consumption, and radiation exposure play significant roles in developing these cancers. Colorectal cancer originates in the colon or rectum—both parts of the large intestine—while lung cancer arises in the lung tissue, characterized by unregulated cell growth (Ranasinghe *et al.*, 2022). Although early detection improves outcomes, identifying these cancers at an initial stage often proves challenging without specialized medical screenings. Thus, regular screenings are crucial for individuals at increased risk (Crosby *et al.*, 2023).

The timely detection of these cancers is vital for enhancing survival rates, as metastasis between the lungs and colon can occur if diagnosis is delayed. Diagnostic approaches typically involve imaging techniques, such as X-rays, CT scans, and biopsies. Among these, histopathological analysis remains the gold standard for confirming cancer diagnoses and reducing mortality rates. However, histopathological inspection is labor-intensive, time-consuming, and dependent on pathologists' subjective expertise, which can result in diagnostic inconsistencies (Panayides *et al.*, 2023).

In recent years, advancements in artificial intelligence (AI) and machine learning (ML) have transformed cancer diagnostics, enabling the automated analysis of medical images. Deep learning, a subset of AI, has been particularly effective in identifying subtle patterns within histopathological images that might elude human observation (Ahmed *et al.*, 2022). Moreover, the use of convolutional neural networks (CNNs), especially transfer learning models like Inception-ResNetV2, has shown remarkable promise in improving cancer detection accuracy (Alabdulqader *et al.*, 2023). These models can automatically learn and extract hierarchical features from raw medical images, consequently reducing the need for manual feature engineering and enhancing classification accuracy.

Nevertheless, challenges remain in applying AI-driven diagnostic tools in clinical practice. A primary challenge is the generalizability of AI models across diverse datasets and imaging modalities, as most models are trained on a limited set of data that may not fully represent the variability seen in real-world scenarios (Tummala *et al.*, 2023). Additionally, interpretability is a critical concern. Although AI models often achieve high accuracy, the "black box" nature of many deep learning algorithms makes it difficult for clinicians to understand the decision-making process—an essential factor for building trust and adoption in clinical settings. To address this, explainable AI (XAI) techniques, such as SHAP (Shapley Additive Explanations), have been developed to provide insights into model predictions by highlighting the impact of individual features in the diagnostic process (Hasan *et al.*, 2024).

Moreover, the cost and accessibility of advanced diagnostic tools remain significant barriers, particularly in low-income regions, where up to 70% of cancer-related deaths occur. Supporting these regions with AI-based diagnostic systems and adequately trained personnel could help mitigate disparities in cancer care and enhance early detection efforts worldwide. Recent studies have examined hybrid approaches that combine AI with traditional diagnostic techniques to improve sensitivity and specificity, further demonstrating the potential of AI to revolutionize cancer diagnostics (Al-Jabbar *et al.*, 2023).

This research centers on the application of the Inception-ResNetV2 model—a deep learning architecture that combines the strengths of Inception and ResNet networks—to classify lung and colon cancer from histopathological images. By leveraging transfer learning and explainable AI techniques, this study aims to enhance diagnostic accuracy and interpretability, offering a promising approach for early cancer detection.

**Literature Review**

Over the years, numerous studies have investigated the use of neural networks to diagnose lung and colon cancers through histopathological image analysis. Consequently, research in this area has provided a strong foundation for developing reliable, automated cancer detection systems.

Yahia and Talaat (2022), for example, developed an intelligent classification system that combines CLAHE-based image enhancement with modified neural networks. They used deep learning algorithms to automate the detection of five types of colon and lung tissues (three malignant and two benign), achieving a remarkable 99.5% accuracy. Moreover, their system demonstrated potential in assisting clinicians by providing a reliable tool for identifying various tumor types.

In a similar vein, Raju and Rao (2022) employed visualization techniques such as GradCam and SmoothGrad alongside MobileNetV2 and InceptionResNetV2 models to analyze histopathological images of colon and lung tissues. With an impressive accuracy of 99.95%, their work underscored the importance of visualization methods for enhancing model interpretability, making AI systems more transparent and dependable for clinical applications.

Additionally, Mengash et al. (2023) integrated the Marine Predators Algorithm (MPA) with deep learning to develop the MPADL-LC3 model, which achieved a classification accuracy of 99.27%. Similarly, Alqahtani et al. (2023) introduced an improved Water Strider algorithm combined with a Convolutional Autoencoder, achieving 99.41% accuracy alongside notable computational efficiency.

Meanwhile, Obayya et al. (2023) implemented the BICLCD-TSADL model, which utilizes the Tuna Swarm Algorithm with deep learning, achieving high precision in colon and lung cancer detection, surpassing 99.33% accuracy.

More recently, Singh et al. (2024) explored the application of state-of-the-art convolutional neural networks to differentiate between squamous cell carcinoma and adenocarcinoma of the lung and colon using histopathological images. They tested five pretrained architectures—VGG-19, InceptionResNetV2, DenseNet201, EfficientNetB6, and MobileNetV2—on the LC25000 dataset, achieving accuracies between 93.12% and 99.32%, with MobileNetV2 demonstrating the highest accuracy.

In addition, Farhadipour (2024) proposed a preprocessing algorithm based on CLAHE to enhance histopathological images for cancer diagnosis using CNNs. Evaluating models such as VGG19, SqueezeNet, and EfficientNet-b0 on the LC25000 dataset, they found VGG19 reached 99.84% accuracy, while SqueezeNet offered an optimal balance between accuracy (99.58%) and computational efficiency, requiring only 0.56 ms per image.

Furthermore, Moin et al. (2024) observed that recent advancements in deep learning have continued to propel the use of CNNs for analyzing histopathological images of lung and colon cancers. They adapted pretrained CNN models, including Xception, DenseNet, and ResNet, to enhance classification performance, achieving accuracies between 97% and 99%. Additionally, they applied visualization techniques like GradCAM to improve interpretability in classifying malignant and benign images.

This study builds upon prior research by introducing an Inception-ResNetV2 model for classifying colon and lung cancers, utilizing Local Binary Patterns (LBP) to improve precision and computational efficiency. Moreover, by integrating explainable AI techniques, such as SHAP, this study provides detailed insights into the contributions of individual features in the predictive analysis, thereby enhancing the model's transparency and clinical applicability (Hasan et al., 2024).

Ochoa-Ornelas, Raquel, Gudiño-Ochoa, Alberto, García-Rodríguez, Julio Alberto and Uribe-Toscano, Sofia. [2024]. Lung and colon cancer detection with InceptionResNetV2: a transfer learning approach. Journal of Research and Development. 10[25]-1-13: e11025113. https://doi.org/10.35429/JRD.2024.10.25.1.13

### Dataset and Application of CLAHE

This study utilized histopathological images from the LC25000 dataset, which includes five distinct classes: Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, and Lung Squamous Cell Carcinoma, with 5000 images per class (Borkowski et al., 2019). To enhance model generalization and variability, we replaced 1000 images within the malignant classes (adenocarcinoma and squamous cell carcinoma) with newly acquired ones. Consequently, this adjustment improves interpretability across different cases. The new images were captured using an Amperio Image Scope at zoom levels of 20x and 40x. Figure 1 provides a representative sample from each class. Moreover, each image underwent contrast enhancement using CLAHE to ensure the model effectively captures essential features during the classification process.

**Box 1**



Lung Adenocarcinoma    Lung Benign Tissue    Lung Squamous Cell Carcinoma

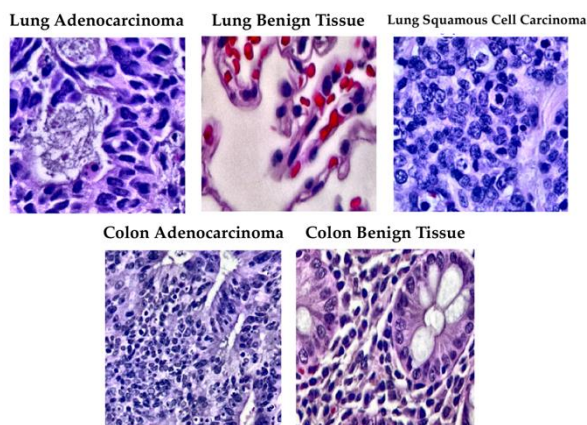Colon Adenocarcinoma    Colon Benign Tissue

**Figure 1**

Representative histopathological images of lung and colon tissues, including benign and malignant samples from adenocarcinoma and squamous cell carcinoma

*Source: Own elaboration*

The contrast of the images was improved using CLAHE, a sophisticated technique designed to enhance the visibility of details in regions where contrast is naturally low, without amplifying noise or unwanted artifacts. CLAHE, an advanced version of Adaptive Histogram Equalization (AHE), redistributes intensity levels within small, localized regions (sub-images or tiles) of the image rather than adjusting the entire image globally.

Consequently, this approach allows for the enhancement of local details without affecting regions where contrast is already sufficient (Al-Jabbar et al., 2023).

The CLAHE process involves dividing the image into small subregions or tiles. Each tile undergoes local contrast adjustment through adaptive histogram equalization, which enhances specific details within each subregion. The gray-level histogram of each tile is adjusted to redistribute pixel intensities more evenly, thus improving contrast in areas with low visibility. To prevent over-amplification of noise or artifacts in these low-contrast areas, CLAHE introduces a "clipping" mechanism in the histogram (Mengash et al., 2023). This mechanism sets a maximum threshold for contrast redistribution, thereby limiting excessive enhancement in homogeneous regions where noise could otherwise increase if left uncontrolled. By applying CLAHE, the images fed into the model contained enhanced local contrast, enabling the model to extract and utilize finer details during the classification of colon and lung cancers. This preprocessing step is crucial for improving the model's accuracy, ensuring it can effectively differentiate between benign and malignant tissues.

### InceptionResNetV2 Model

The InceptionResNetV2 architecture was chosen for this study due to its state-of-the-art performance in image classification tasks. It combines the strengths of both the Inception and ResNet architectures by integrating multi-scale feature extraction through inception modules with residual connections, which help mitigate the vanishing gradient problem in deep networks (Neshat et al., 2017). Consequently, this architecture is particularly effective for complex tasks like histopathological cancer detection, where hierarchical feature extraction across various spatial scales is essential for accurate classification.

For this study, the dataset was divided into three subsets: training (80%), validation (12%), and testing (8%). This division ensures that the model has sufficient data to learn patterns while also providing a robust framework for evaluating its performance on unseen validation and test sets.

Ochoa-Ornelas, Raquel, Gudiño-Ochoa, Alberto, García-Rodríguez, Julio Alberto and Uribe-Toscano, Sofia. [2024]. Lung and colon cancer detection with InceptionResNetV2: a transfer learning approach. Journal of Research and Development. 10[25]-1-13: e11025113. https://doi.org/10.35429/JRD.2024.10.25.1.13

Moreover, this split aligns with approaches commonly used in studies involving the LC25000 dataset, reducing the likelihood of overfitting. The training set was used to adjust the model's weights, while the validation set allowed for hyperparameter tuning and performance monitoring during training. The testing set, comprising 8% of the total data, was reserved for assessing the model's final performance on unseen data, thereby ensuring that the results reflect real-world generalization capabilities.

The InceptionResNetV2 model is designed to facilitate residual learning, where shortcut connections enable gradients to bypass certain layers, effectively reducing the likelihood of vanishing gradients during backpropagation. Mathematically, this can be expressed as:

$$y = \mathcal{F}(x, \{W_i\}) + x \qquad [1]$$

where $y$ is the output, $x$ is the input, and $\mathcal{F}$ represents the residual function learned by the layers, with $W_i$ denoting the layer weights. The addition of the input $x$ to the output of $\mathcal{F}$ allows the network to learn identity mappings more easily, stabilizing the training process. For this study, the InceptionResNetV2 model pre-trained on ImageNet was used as a feature extractor (Mehmood *et al*., 2022). The top layers of the model were excluded, enabling the fine-tuning of the architecture on the specific task of classifying histopathological images. The input dimensions were set to *224x224x3*, suitable for the image resolution of the dataset.

After initializing the pre-trained InceptionResNetV2, additional fully connected layers were added to customize the model for multi-class classification. To stabilize activations and accelerate training convergence, a Batch Normalization layer was introduced. This was followed by a fully connected layer with 256 units and L2 regularization to prevent overfitting. The output layer, configured with a softmax activation function, was designed to predict class probabilities across the five classes of the dataset: Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, and Lung Squamous Cell Carcinoma. The softmax function for the output layer is defined as:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \text{for } i = 1, \dots, K \qquad [2]$$

where $K$ is the number of classes, $z_i$ is the input to the softmax function for class $i$, and $\sigma(z)_i$ represents the predicted probability for class $i$.

The model was trained using the Adamax optimizer, which is a variant of Adam that is particularly effective for models with sparse gradients. The learning rate was set at 0.001 to balance training speed and model stability. For the loss function, categorical crossentropy was selected to measure the discrepancy between predicted class probabilities and the true class labels. This function is expressed as:

$$L(y, \hat{y}) = -\sum_{i=1}^{K} y_i \log(\hat{y_i}) \qquad [3]$$

where $y_i$ is the true label for class $i$ and $\hat{y_i}$ is the predicted probability for class $i$. The model was trained for 20 epochs, and early stopping was implemented to prevent overfitting. Specifically, EarlyStopping was used to halt the training if the validation loss did not improve for three consecutive epochs, ensuring that the best version of the model was preserved. Additionally, ModelCheckpoint was employed to save the model whenever an improvement in validation loss was detected.

**Explainable AI with SHAP**

To enhance the interpretability of the model's predictions, SHAP was applied. SHAP assigns importance scores to each feature in the model, indicating its contribution to the final prediction. This is particularly important in medical applications, where understanding the decision-making process of the model is crucial for clinical validation (Hasan et al., 2024). SHAP values are computed as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \qquad [4]$$

where $N$ is the set of all features, $S$ is a subset of features excluding feature $i$, and $f(S)$ represents the model's prediction based on the subset $S$. This method evaluates the marginal contribution of each feature, offering a robust explanation of how the model makes its predictions.

Ochoa-Ornelas, Raquel, Gudiño-Ochoa, Alberto, García-Rodríguez, Julio Alberto and Uribe-Toscano, Sofia. [2024]. Lung and colon cancer detection with InceptionResNetV2: a transfer learning approach. Journal of Research and Development. 10[25]-1-13: e11025113. https://doi.org/10.35429/JRD.2024.10.25.1.13

Article

## Evaluation and performance metrics

The model's performance was evaluated by tracking the loss and accuracy during training and validation over multiple epochs. As shown in Figure 2, the training loss decreased significantly during the initial epochs and stabilized around epoch 10, where the best validation loss was observed. To prevent overfitting, early stopping was applied, halting the training process after the 10th epoch. Similarly, Figure 3 illustrates the trends in training and validation accuracy. The validation accuracy peaked at epoch 6, reaching 96%, after which it fluctuated, likely indicating that the model was approaching its generalization limit. These observations underscore the importance of balancing training and validation to avoid overfitting.
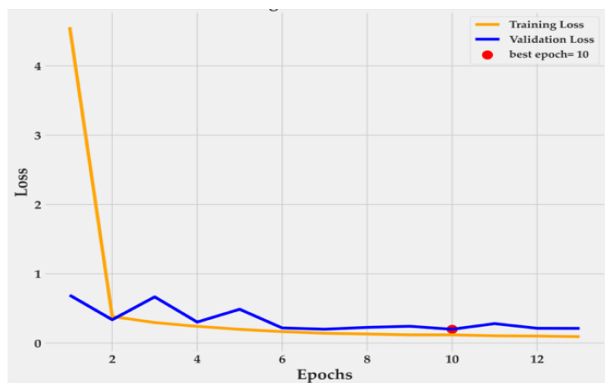
### Box 2



**Figure 2**

Training and validation loss over epochs, with the best validation loss at epoch 10 (red point)

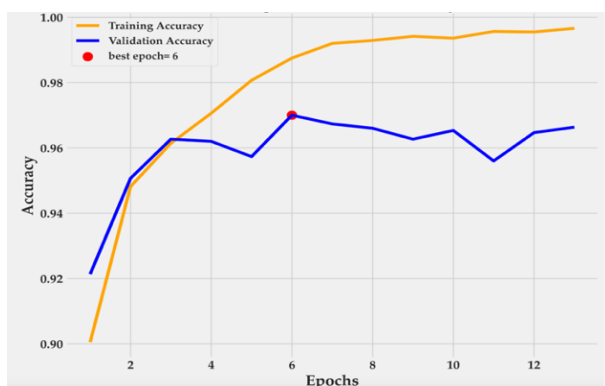*Source: Own elaboration*

### Box 3



**Figure 3**

Training and validation accuracy over epochs, with the highest validation accuracy at epoch 6 (red point)

*Source: Own elaboration*

The confusion matrix, shown in Figure 4, demonstrates the model's performance across the entire dataset by comparing the true labels with the predicted labels for each of the five cancer classes: Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, and Lung Squamous Cell Carcinoma.
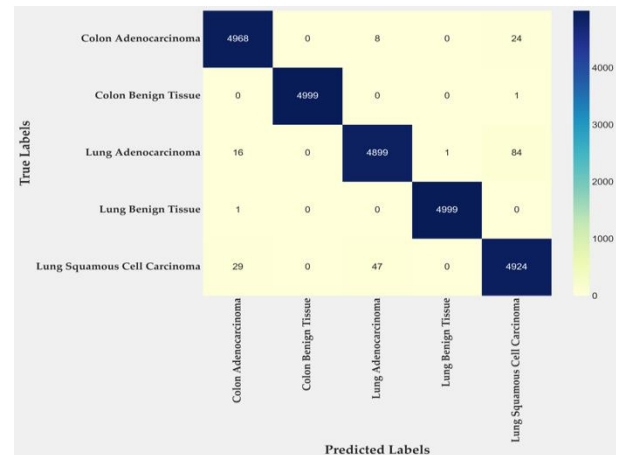
### Box 4



**Figure 4**

Confusion matrix for the classification of the entire dataset.

*Source: Own elaboration*

The matrix provides insights into the model's classification accuracy for each class. Notably, the model achieved near-perfect accuracy for benign tissue classes, with minimal misclassifications. For instance, Colon Benign Tissue and Lung Benign Tissue both achieved an accuracy of 99.98%, with only one misclassification out of 5000 images in each case.

However, some degree of misclassification was observed in the malignant classes. Specifically, Lung Squamous Cell Carcinoma had the highest misclassification rate, with 47 instances incor classified as Lung Adenocarcinoma. Similarly, Lung Adenocarcinoma showed 84 instances misclassified as Lung Squamous Cell Carcinoma. These errors may be attributed to the histopathological similarities between these two cancer subtypes, as noted in previous studies. Table 1 summarizes the classification performance across all classes, with benign tissues showing over 99% accuracy, while malignant tissues range from 97% to 99%. The average accuracy across all classes is 99.15%.

Ochoa-Ornelas, Raquel, Gudiño-Ochoa, Alberto, García-Rodríguez, Julio Alberto and Uribe-Toscano, Sofia. [2024]. Lung and colon cancer detection with InceptionResNetV2: a transfer learning approach. Journal of Research and Development. 10[25]-1-13: e11025113. https://doi.org/10.35429/JRD.2024.10.25.1.13

## Box 5

**Table 1**

Classification report for the entire dataset.

| Class | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|
| Colon Adenocarcinoma | 99.08 | 99.36 | 99.22 |
| Colon Benign Tissue | 100 | 99.98 | 99.99 |
| Lung Adenocarcinoma | 98.88 | 97.98 | 98.43 |
| Lung Benign Tissue | 99.98 | 99.98 | 99.98 |
| Lung Squamous Cell Carcinoma | 97.83 | 98.48 | 98.15 |

*Source: Own elaboration*

The confusion matrix in Figure 5 illustrates the model's classification performance on the testing dataset. Overall, the model demonstrated high accuracy, particularly in the benign tissue categories. For Colon Adenocarcinoma, the majority of samples were correctly classified, with only a small fraction misclassified. Similarly, for Colon Benign Tissue, the model achieved perfect accuracy, indicating robust performance in distinguishing benign from malignant colon tissues.

## Box 6
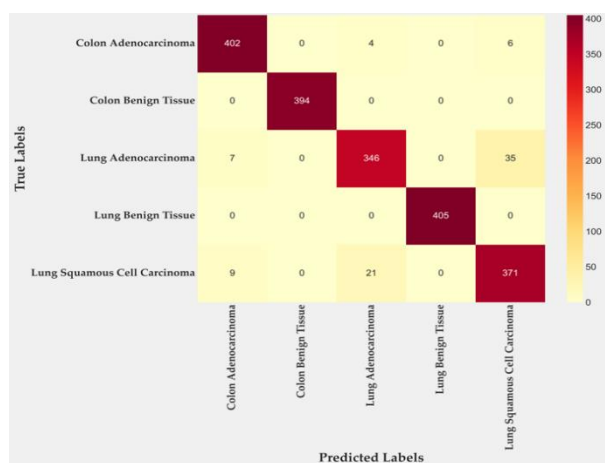


**Figure 5**

Confusion matrix illustrating the model's performance on the testing dataset

*Source: Own elaboration*

In the case of Lung Adenocarcinoma, the model exhibited strong classification accuracy but showed a tendency to confuse it with Lung Squamous Cell Carcinoma, likely due to overlapping histological features.

A similar pattern was observed for Lung Squamous Cell Carcinoma, where a higher misclassification rate occurred. Notably, the model performed exceptionally well in identifying benign lung tissue, achieving a perfect classification for Lung Benign Tissue. These results suggest that the model generalizes effectively to unseen data, although the confusion between lung adenocarcinoma and squamous cell carcinoma highlights potential areas for further optimization. Table 2 provides a classification report for the testing dataset.

The confusion matrix in Figure 6 illustrates the model's performance on the training dataset. For Colon Adenocarcinoma, out of 4013 samples, the model correctly classified 4012, with only one misclassified sample. Similarly, the Colon Benign Tissue class achieved near-perfect classification, with only one misclassification among 4001 samples.

## Box 7

**Table 2**

Classification report for the testing dataset.

| Class | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|
| Colon Adenocarcinoma | 96.17 | 97.57 | 96.86 |
| Colon Benign Tissue | 100 | 100 | 100 |
| Lung Adenocarcinoma | 93.26 | 89.17 | 91.17 |
| Lung Benign Tissue | 100 | 100 | 100 |
| Lung Squamous Cell Carcinoma | 90.04 | 95.90 | 95.89 |

*Source: Own elaboration*

The model also showed strong performance in the Lung Adenocarcinoma and Lung Benign Tissue classes, where it accurately classified 4004 and 3987 samples, respectively. A few misclassifications were noted, particularly between Lung Adenocarcinoma and Lung Squamous Cell Carcinoma, which is likely due to overlapping histopathological features between these malignancies. Nonetheless, the number of misclassified instances is negligible compared to the total number of samples. For Lung Squamous Cell Carcinoma, the model classified 3972 samples correctly, with only 6 instances misclassified. Table 3 shows the report classification of the training dataset.

Ochoa-Ornelas, Raquel, Gudiño-Ochoa, Alberto, García-Rodríguez, Julio Alberto and Uribe-Toscano, Sofia. [2024]. Lung and colon cancer detection with InceptionResNetV2: a transfer learning approach. Journal of Research and Development. 10[25]-1-13: e11025113. https://doi.org/10.35429/JRD.2024.10.25.1.13
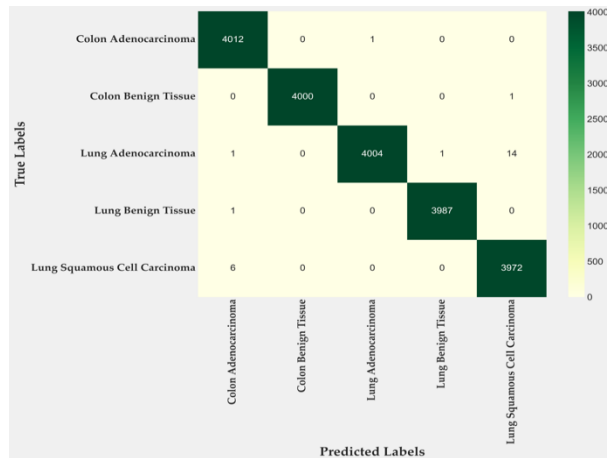
## Box 8



**Figure 6**

Confusion matrix illustrating the model's performance on the training dataset

*Source: Own elaboration*

## Box 9

**Table 3**

Report classification of the training dataset.

| Class | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|
| Colon Adenocarcinoma | 99.80 | 99.97 | 99.88 |
| Colon Benign Tissue | 100 | 99.97 | 99.98 |
| Lung Adenocarcinoma | 99.97 | 99.60 | 99.78 |
| Lung Benign Tissue | 99.97 | 99.97 | 99.97 |
| Lung Squamous Cell Carcinoma | 99.62 | 99.84 | 99.73 |

*Source: Own elaboration*

This high level of performance across all classes underscores the model's effectiveness in learning from the training data, achieving high precision and accuracy in distinguishing between malignant and benign tissues. The near-perfect classification performance on the training dataset indicates that the model is well-optimized for this task. Nevertheless, caution is necessary to ensure that this high accuracy translates to generalization on unseen data, as overfitting remains a potential concern. Figure 7 provides a comparative overview of the model's performance across the training dataset, testing dataset, and entire dataset using key evaluation metrics: Accuracy, Precision, Recall, F1-Score, and Matthews Correlation Coefficient (MCC). The results highlight the model's robust performance and generalization capabilities.
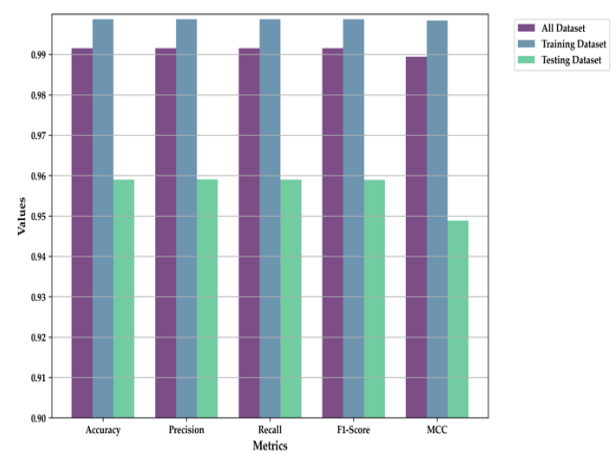
## Box 10



**Figure 7**

Comparison of average Accuracy, Precision, Recall, F1-Score, and MCC across the training, testing, and full dataset

*Source: Own elaboration*

Accuracy measures the overall correctness of the model's predictions and is defined as the ratio of correct predictions (true positives TP and true negatives TN) to the total number of predictions:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad [5]$$

Precision quantifies the proportion of positive predictions that are correct, calculated as:

$$\text{Precision} = \frac{TP}{TP+FP} \quad [6]$$

Recall, also known as sensitivity, evaluates the model's ability to correctly identify positive instances:

$$\text{Recall} = \frac{TP}{TP+FN} \quad [7]$$

The F1-Score provides a balance between Precision and Recall, particularly useful in cases of imbalanced datasets. It is the harmonic mean of Precision and Recall:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad [8]$$

Lastly, the MCC offers a robust evaluation metric that considers all elements of the confusion matrix (true positives, true negatives, false positives, and false negatives), making it particularly valuable for imbalanced datasets. It is calculated as:

Ochoa-Ornelas, Raquel, Gudiño-Ochoa, Alberto, García-Rodríguez, Julio Alberto and Uribe-Toscano, Sofia. [2024]. Lung and colon cancer detection with InceptionResNetV2: a transfer learning approach. Journal of Research and Development. 10[25]-1-13: e11025113. https://doi.org/10.35429/JRD.2024.10.25.1.13

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad [9]$$

The training dataset demonstrated near-perfect performance, with all metrics—Accuracy, Precision, Recall, and F1-Score—scoring approximately 99.87%, and an MCC of 99.84%. These values indicate that the model fits the training data exceptionally well. On the testing dataset, the model maintained strong generalization, achieving Accuracy, Precision, and Recall around 95.90%, and an MCC of 94.88%, reflecting its effectiveness on unseen data with minimal overfitting. For the entire dataset, the aggregated metrics confirm consistent performance, with Accuracy, Precision, and Recall reaching approximately 99.15%, and an MCC of 98.94%, verifying the model's reliability in classifying histopathological images across both training and testing phases.

Figure 8 presents the Receiver Operating Characteristic (ROC) curves for each of the five histopathological classes in the testing dataset: Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, and Lung Squamous Cell Carcinoma. The ROC curve evaluates the model's ability to distinguish between true positives and false positives for each class, with the Area Under the Curve (AUC) serving as a key performance indicator, where TPR is the True Positive Rate or Recall:

$$AUC = \int_0^1 TPR(x) \ dx \quad [10]$$

The results indicate near-perfect performance across all classes, with Area Under the Curve (AUC) values ranging from 0.9939 to 1.0000, demonstrating excellent discriminatory power. Both benign tissue classes—Colon Benign Tissue and Lung Benign Tissue—achieved an AUC of 1.0000, reflecting perfect classification with no false positives. For the malignant classes, Colon Adenocarcinoma and Lung Adenocarcinoma also exhibited outstanding performance, with AUCs of 0.9977 and 0.9928, respectively. Although these values are slightly lower than those of the benign classes, they still represent exceptional accuracy in detecting cancerous tissues.

The Lung Squamous Cell Carcinoma class, with an AUC of 0.9939, also demonstrated strong classification; however, the marginally lower AUC compared to other classes suggests minor misclassifications, consistent with the findings in the confusion matrix. Overall, the ROC curves confirm the model's robustness in handling both benign and malignant histopathological images with minimal error rates.
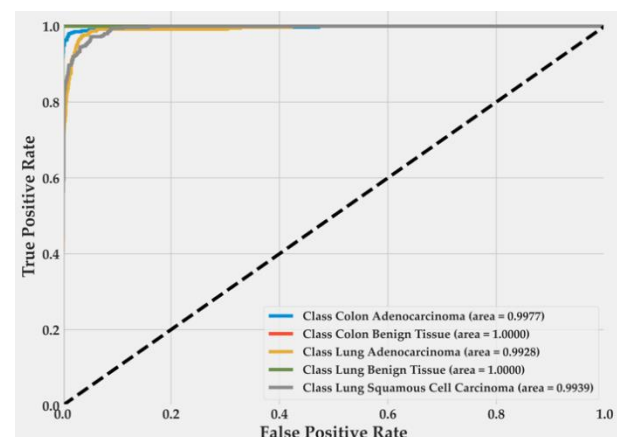
**Box 11**



**Figure 8**
ROC curves for the testing dataset

*Source: Own elaboration*

Figure 9 presents the Precision-Recall (PR) curves for the five histopathological classes in the testing dataset. PR curves are particularly useful for evaluating performance in datasets with class imbalances, as they emphasize the model's ability to accurately predict positive instances.
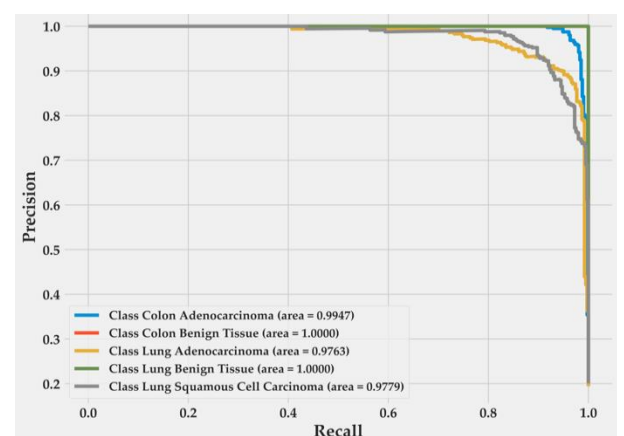
**Box 12**



**Figure 9**
Precision-Recall curves for the testing dataset

*Source: Own elaboration*

Ochoa-Ornelas, Raquel, Gudiño-Ochoa, Alberto, García-Rodríguez, Julio Alberto and Uribe-Toscano, Sofia. [2024]. Lung and colon cancer detection with InceptionResNetV2: a transfer learning approach. Journal of Research and Development. 10[25]-1-13: e11025113. https://doi.org/10.35429/JRD.2024.10.25.1.13

**Journal of Research and Development**

Article

The PR curve results further demonstrate the model's strong performance, especially across the benign classes, with Colon Benign Tissue and Lung Benign Tissue achieving an AUC of 1.0000. This outcome suggests that the model can classify benign samples without any false positives or false negatives, a highly desirable feature in clinical applications. For the malignant classes, the model's performance remains very strong. Colon Adenocarcinoma achieves an AUC of 0.9947, indicating high precision and recall. Meanwhile, Lung Adenocarcinoma and Lung Squamous Cell Carcinoma show slightly lower AUCs of 0.9763 and 0.9779, respectively, reflecting minor misclassifications between these subtypes, as previously noted in the testing confusion matrix.

Figure 10 presents the SHAP value analysis across a range of histopathological images for each of the five cancer classes. SHAP values allow for interpreting the contribution of individual features to the model's predictions, providing valuable insights into the decision-making process of the deep learning model. In this analysis, red areas indicate features with a positive contribution to a specific class, while blue areas represent features that reduce the likelihood of classifying an image into a particular class. The intensity of each color correlates with the magnitude of the SHAP value, highlighting important regions that the model uses to differentiate between classes.

For instance, in the Lung Adenocarcinoma and Lung Squamous Cell Carcinoma classes, the SHAP values highlight distinct regions in the tissue where cellular abnormalities significantly contribute to classification. These regions correspond to areas of abnormal cellular structure, such as increased nucleus density or irregular tissue patterns. Similarly, in benign tissue classes, blue regions representing negative SHAP values align with areas that reinforce non-cancerous classification by the absence of malignant features. This SHAP analysis enhances model transparency by emphasizing the importance of specific features, thereby confirming the model's capability to accurately distinguish between lung and colon cancer subtypes based on critical image regions. The consistent interpretability across cancer classes underscores the robustness of the model, ensuring that its predictions are based on biologically meaningful features rather than noise.
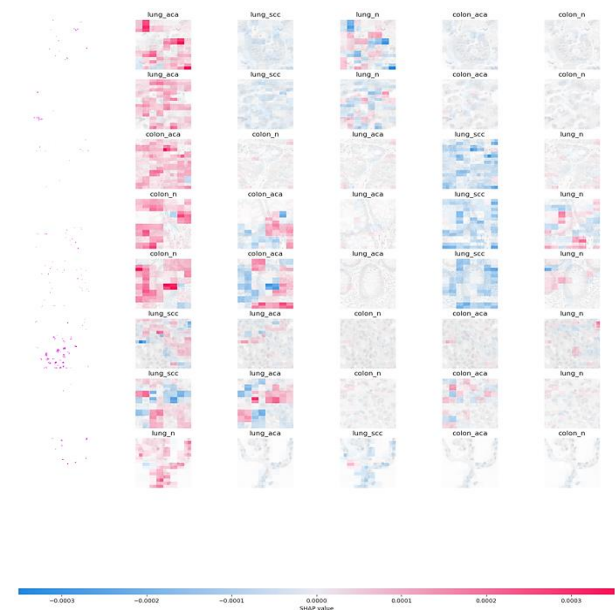
**Box 13**



**Figure 10**

SHAP value analysis across multiple histopathological images

*Source: Own elaboration*

## Conclusions

This study effectively leveraged the Inception-ResNetV2 architecture and transfer learning to classify histopathological images for detecting lung and colon cancers. The model demonstrated strong performance, with the testing dataset achieving an accuracy of 95.90%, a precision of 95.91%, a recall of 95.90%, and an F1-Score of 95.89%, along with a Matthews Correlation Coefficient (MCC) of 94.88%. The training dataset results were near-perfect, with all metrics exceeding 99.87%, confirming the robustness of the model's training.

Despite these promising results, there is room for further improvement. Future work could focus on expanding the dataset with more diverse samples to enhance generalization. Additionally, optimizing the feature extraction and classification processes—perhaps through the integration of more advanced preprocessing techniques or by combining multiple models—could further improve the model's ability to differentiate between similar cancer subtypes (Tummala et al., 2023; Al-Jabbar et al., 2023). Real-world validation in clinical settings is also necessary to evaluate the model's reliability in practical applications.

Ochoa-Ornelas, Raquel, Gudiño-Ochoa, Alberto, García-Rodríguez, Julio Alberto and Uribe-Toscano, Sofia. [2024]. Lung and colon cancer detection with InceptionResNetV2: a transfer learning approach. Journal of Research and Development. 10[25]-1-13: e11025113. https://doi.org/10.35429/JRD.2024.10.25.1.13

In conclusion, the proposed method offers both high accuracy and interpretability, making it a promising tool for automated cancer detection. Continuous enhancements to the model will further refine its clinical applicability, potentially improving early detection and diagnosis in oncological settings.

## Declarations

### Conflict of interest

The authors declare no conflicts of interest.

### Authors' contribution

*Ochoa-Ornelas, Raquel:* Contributed to the project idea, conceptualization, methodology, software, validation, resources, data curation, writing—original draft preparation, supervision, project administration and funding acquisition.

*Gudiño-Ochoa, Alberto:* Contributed to the project idea, conceptualization, methodology, software, investigation, data curation, writing—original draft preparation, visualization and supervision.

*García-Rodríguez, Julio Alberto*: Methodology, validation, resources, writing—original draft preparation, visualization, project administration and funding acquisition.

*Uribe-Toscano, Sofia*: Conceptualization, validation, investigation, data curation and visualization.

### Availability of data and materials

The datasets used in this study are publicly available. The enhanced LC25000 dataset, including additional images and CLAHE preprocessing, can be accessed via the following links:
https://github.com/AlbertoGudinoOchoa/Enhanced-LC25000-CLAHE-Dataset-Cancer-Classification and https://drive.google.com/drive/folders/1aQNez61naAiuveaQlSzJ2VBsMI5_KUYm?usp=sharing.

The dataset is distributed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license, which allows sharing, adaptation, and redistribution of the material, provided appropriate credit is given and any derivative works are shared under the same license. Full details of the license can be found at Creative Commons License.

### Abbrevations

| | |
|---|---|
| AHE | Adaptive Histogram Equalization |
| AI | Artificial Intelligence |
| AUC | Area Under the Curve |
| CAD | Computer-Aided Detection |
| CLAHE | Contrast Limited Adaptive Histogram Equalization |
| CNN | Convolutional Neural Networks |
| CT | Computed Tomography |
| DenseNet | Dense Convolutional Network |
| F1-Score | Harmonic mean of Precision and Recall |
| GradCAM | Gradient-weighted Class Activation Mapping |
| LBP | Local Binary Patterns |
| MCC | Matthews Correlation Coefficient |
| ML | Machine Learning |

| | |
|---|---|
| MPA | Marine Predators Algorithm |
| PR | Precision-Recall |
| ResNet | Residual Network |
| ROC | Receiver Operating Characteristic |
| SHAP | Shapley Additive Explanations |
| SqueezeNet | Lightweight Convolutional Neural Network |
| VGG | Visual Geometry Group |
| XAI | Explainable Artificial Intelligence |

**References**

*Background*

Kiri, S., & Ryba, T. (2024). Cancer, metastasis, and the epigenome. *Molecular Cancer*, *23*(1), 154.

Li, N., Zhai, Z., Zheng, Y., Lin, S., Deng, Y., Xiang, G., ... & Wang, M. (2021). Association of 13 occupational carcinogens in patients with cancer, individually and collectively, 1990-2017. JAMA Network Open, 4(2), e2037530-e2037530.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 71(3), 209-249.

*Basics*

Cao, M., Li, H., Sun, D., & Chen, W. (2020). Cancer burden of major cancers in China: a need for sustainable actions. Cancer Communications, 40(5), 205-210.

Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., ... & Balasubramanian, S. (2022). Early detection of cancer. *Science*, *375*(6586), eaay9040.

Ranasinghe, R., Mathai, M., & Zulli, A. (2022). A synopsis of modern-day colorectal cancer: Where we stand. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *1877*(2), 188699.

*Support*

Ahmed, A. A., Abouzid, M., & Kaczmarek, E. (2022). Deep learning approaches in histopathology. *Cancers*, *14*(21), 5264.

Alabdulqader, E. A., Umer, M., Alnowaiser, K., Wang, H., Alarfaj, A. A., & Ashraf, I. (2024). Image Processing-based Resource-Efficient Transfer Learning Approach for Cancer Detection Employing Local Binary Pattern Features. Mobile Networks and Applications, 1-17.

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). *arXiv preprint arXiv:1912.12142*.

Hasan, M. A., Haque, F., Sabuj, S. R., Sarker, H., Goni, M. O. F., Rahman, F., & Rashid, M. M. (2024). An End-to-End Lightweight Multi-Scale CNN for the Classification of Lung and Colon Cancer with XAI Integration. *Technologies*, *12*(4), 56.

Panayides, A. S., Amini, A., Filipovic, N. D., Sharma, A., Tsaftaris, S. A., Young, A., ... & Pattichis, C. S. (2020). AI in medical imaging informatics: current challenges and future directions. *IEEE journal of biomedical and health informatics*, *24*(7), 1837-1857.

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.

*Differences*

Al-Jabbar, M., Alshahrani, M., Senan, E. M., & Ahmed, I. A. (2023). Histopathological Analysis for Detecting Lung and Colon Cancer Malignancies Using Hybrid Systems with Fused Features. *Bioengineering*, *10*(3), 383.

Alqahtani, H., Alabdulkreem, E., Alotaibi, F., Alnfiai, M. M., Singla, C., & Salama, A. S. (2023). Improved Water Strider Algorithm With Convolutional Autoencoder for Lung and Colon Cancer Detection on Histopathological Images. *IEEE Access*.

Farhadipour, A. (2024). Lung and colon cancer detection with convolutional neural networks and adaptive histogram equalization. *Iran Journal of Computer Science*, *7*(2), 381-395.

Mengash, H. A., Alamgeer, M., Maashi, M., Othman, M., Hamza, M. A., Ibrahim, S. S., ... & Yaseen, I. (2023). Leveraging marine predators algorithm with deep learning for lung and colon cancer diagnosis. *Cancers*, *15*(5), 1591.

Moin, M. B., Faria, F. T. J., Saha, S., Rafa, B. K., & Alam, M. S. (2024). Exploring Explainable AI Techniques for Improved Interpretability in Lung and Colon Cancer Classification. *arXiv preprint arXiv:2405.04610*.

Obayya, M., Arasi, M. A., Alruwais, N., Alsini, R., Mohamed, A., & Yaseen, I. (2023). Biomedical image analysis for colon and lung cancer detection using tuna swarm algorithm with deep learning model. *IEEE Access*.

Raju, M. S. N., & Rao, B. S. (2022). Classification of Colon and Lung Cancer Through Analysis of Histopathology Images Using Deep Learning Models. Ingenierie des Systemes d'Information, *27*(6).

Singh, O., Kashyap, K. L., & Singh, K. K. (2024). Lung and Colon Cancer Classification of Histopathology Images Using Convolutional Neural Network. *SN Computer Science*, *5*(2), 223.

Yahia Ibrahim, N., & Talaat, A. S. (2022). An enhancement technique to diagnose colon and lung cancer by using double CLAHE and deep learning. *International Journal of Advanced Computer Science and Applications*, *13*(8).

*Discussions*

Mehmood, S., Ghazal, T. M., Khan, M. A., Zubair, M., Naseem, M. T., Faiz, T., & Ahmad, M. (2022). Malignancy detection in lung and colon histopathology images using transfer learning with class selective image processing. *IEEE Access*, *10*, 25657-25668.

Neshat, M., Ahmed, M., Askari, H., Thilakaratne, M., & Mirjalili, S. (2024). Hybrid Inception Architecture with Residual Connection: Fine-tuned Inception-ResNet Deep Learning Model for Lung Inflammation Diagnosis from Chest Radiographs. *Procedia Computer Science*, *235*, 1841-1850.