Deep Learning Based Multi-class Classification Model for Lung and Colon Cancer Detection

# Abstract

Lung and colon cancers are recognized as one of the lethal cancer types accountings for high mortality rates globally. Identifying such cancers has become a challenging task due to their complex symptoms, insidious nature of cancer growth and variability in clinical manifestation. Machine learning models can act as effective tools in diagnosing lung and colon cancers due to the advanced histopathological analysis of imagery to discover cancerous patterns. Thus, this study aims to develop an effective deep learning model utilizing advanced preprocessing and optimization strategies that can accurately recognize lung and colon cancer using the LC25000 dataset as a multi-class classification five types of cancers. A methodical approach of data preprocessing techniques, included CLAHE, histogram equalization and noise removal followed by segmentation methods combined with Xception specific augmentations. Several deep learning models such as Xception, CNN, VGG16, RestNet50, MobileNetV2, EfficientNetB0, InceptionV3 and Swin Transformer were tested with Bayesian optimization to improve model performance. The Xception model achieved the highest accuracy with a value of 1.00 and ROC curve of 1.00. RestNet50 was the second-best model, attaining an accuracy 0.997 and ROC curve of 0.998. The InceptionV3 was fastest in training with 700.41 seconds yet, with a low accuracy of 0.65. The study reveals that Xception model is ideal for lung and colon cancer detection as a reliable and accurate model with easy to training parameters when used with systematic image preprocessing techniques.

# Keywords: *Colon cancer, Deep learning, Hyperparameter optimization, Image Processing, Lung cancer, Xception model*

# Introduction

Both lung and colon cancers are a leading causes of cancer deaths across the world and lung cancer has 11.6% of total cancer cases and is responsible for 18.4% of global cancer deaths, while colorectal cancer has 9.2% of all cancer deaths (Hadiyoso et al., 2023). It tends to develop unnoticeably and can quickly transfer to other body organs before the development of signs, making the early diagnosis very essential but difficult. Conventional diagnostic tools involve chest X-rays, CT scans, biopsies and sputum cytology but such strategies are invasive, expensive or they are not very sensitive. (Hamida et al., 2021). Images made by examining tissue under a microscope are used to diagnose cancer and help precisely group different types of cancer such as lung adenocarcinoma, squamous cell carcinoma or colorectal adenocarcinoma (Mangal et al., 2020).

Machine learning (ML) and deep learning (DL) propose effective solutions, as they allow detecting the problem at an early stage, accurately and non-destructively by analyzing medical images, genomic information, and a patient record. The limitations of the current studies, on lung and colon cancer classification using DL often suffer from limited dataset diversity and lack of validation and reduced model variability. Additionally, complex models and preprocessing dependencies pose challenges for clinical deployment. Thus, our study aims to come up with an effective deep learning with strong preprocessing and optimization strategies that can properly analyze lung and colon cancer in the LC25000 dataset based on its histopathological image to help in early diagnosis and prevention.

Several researchers have used the LC25000 dataset to detect colon and lung cancers using models like (Hadiyoso et al., 2023) Convolution Neural Network (CNN) with VGG16 architecture on the colon and lung cancer dataset with transfer learning, 80:20 train-test split. Results showed an accuracy of 98.96%, a specificity of 99.74% and a sensitivity of 98.96% when used with CLAHE. The accuracy of features consistently increased when CLAHE was used as compared to standard processing (Hadiyoso et al., 2023). A CNN architecture and the RMSprop way of updating weights are used (Khan et al., 2024; Mangal et al., 2020) resulting in training accuracies of 97.92% (lung cancer) and 96.95% (colon) and validation accuracies of 97.90% (lung cancer) and 96.61% (colon). The proposed model uses deep learning, focusing on the Vision Transformer (ViT), Swin Transformer, a modified Swin Transformer and ResNet-101 with best performance by the modified Swin Transformer at 99.80 (Al-Jabbar et al., 2023). Three hybrid deep learning methods are proposed for the classification of this dataset. The first method uses an Artificial Neural Network (ANN) and takes GoogLeNet’s and VGG-19’s results, each with 95.5% and 95.92% accuracy, after PCA. The secondary approach achieved 98.66% accuracy (after PCA) and 98.54% (before PCA) using ANN and fusing GoogLeNet and VGG-19 features. By combining CNN features and three handcrafted features (DWT, LBP, FCH, GLCM) and applying ANN using VGG-19, the study reached an accuracy of 99.64% (sensitivity 99.85%, precision 100%, specificity 100%, AUC 99.86%) which is better than anything seen previously (Tummala et al., 2023).

This study focuses on developing more accurate deep learning models for better detection and classification of lung and colon cancer in histopathological images from the LC25000 dataset. To increase the accuracy of our model, we applied a range of preprocessing steps by investigating the existing issues in the dataset such as lacks diversity in patient demographics and imaging variations. In addition, we added data augmentation to address the data imbalances in classes. By combining these advanced preprocessing methods, modern deep learning networks and optimization techniques, this research aimed to build a reliable, efficient and accurate system for discovering lung and colon cancer early. The novelty of this work approaches deep learning for lung and colon cancer classification by integrating advanced preprocessing techniques such as CLAHE and Xception-specific augmentations, with Bayesian optimization to enhance model performance on the LC25000 dataset.

# MATERIALS AND Methods

The following Figure 1 shows that methodology used in the study. Preprocessing techniques were applied to the acquired dataset in order to make it suitable for model training. Then the pre-processed dataset was split into train, validation, test sets. The models were trained and validated before it was evaluated using test set.

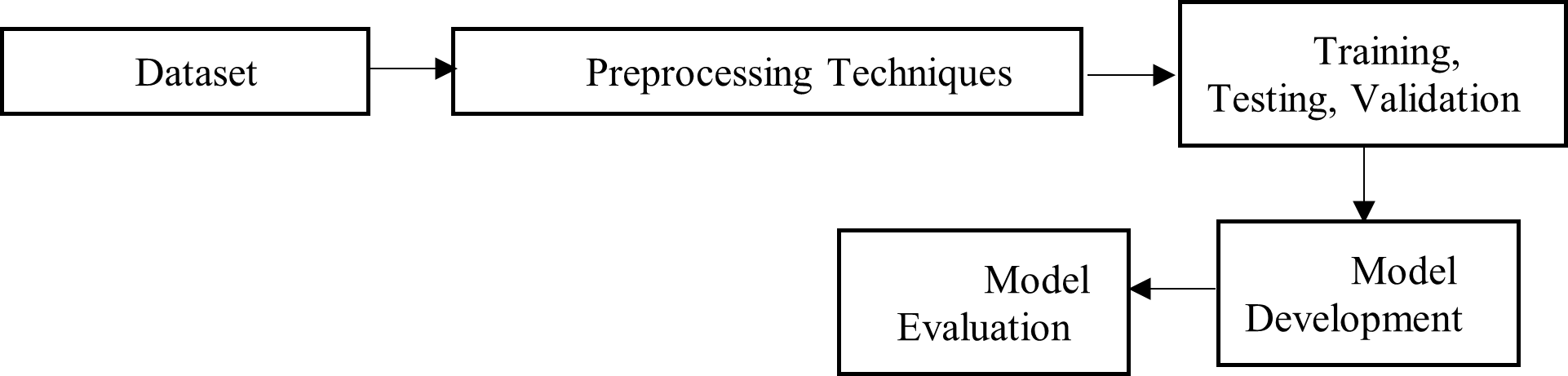


Figure 1: High level Architecture

## Dataset

The study follows lung and colon historical dataset which is 25000 images and have 5 classes. The lung cancers types such as adenocarcinoma, Squamous cell carcinoma, Benign Lung tissue and colon images are Colon adenocarcinoma, Benign colonic tissue. The dataset is publicly available at in Kaggle(Andrew A. Borkowski, 2020). Figure 2 shows the dataset architecture.

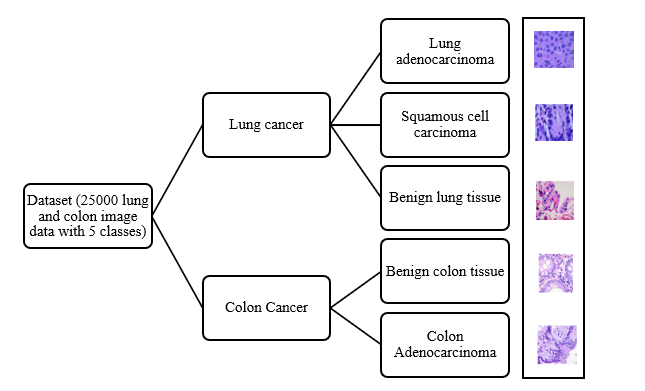


Figure 2: LC25000 Dataset with types of cancer

## Preprocessing techniques

Advanced methods of preprocessing were used on the lungs and colon images within the dataset was applied to enhance image quality, remove noise, standardize features, and highlight critical structures ultimately improving the performance of machine learning and deep learning models in accurate cancer detection and classification. Techniques used were resizing, making grayscale copies, normalization, CLAHE (Contrast Limited Adaptive Histogram Equalization), histogram equalization, contrast boosting and using filters called Gaussian blur, median blur, Laplacian filter, unsharp masking and bilateral filtering. A further set of steps included binarization and removal of background using segmentation methods such as thresholding, edge detection, region growing and the watershed algorithm. We added Xception specific steps and data augmentation by performing rotation, flipping, zooming, brightness changes, cropping and color mapping (Alotaibi et al., 2024). Examples of both raw lung and colon images and their preprocessed forms are shown in Figure 3.

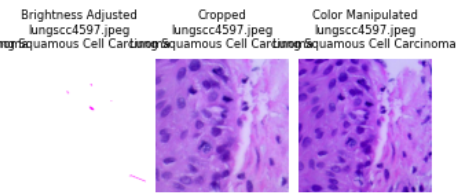
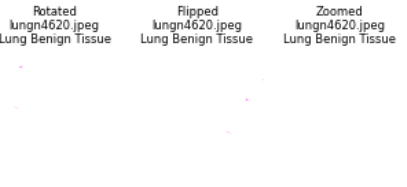
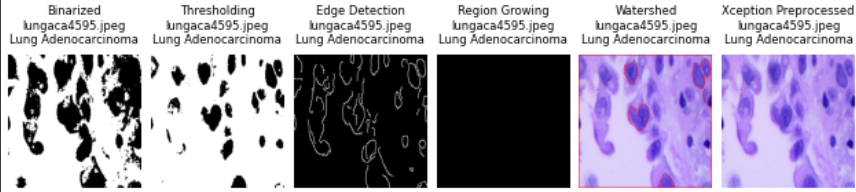
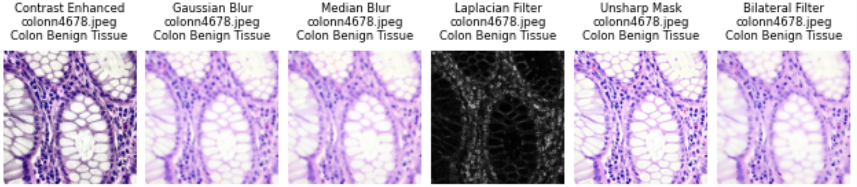
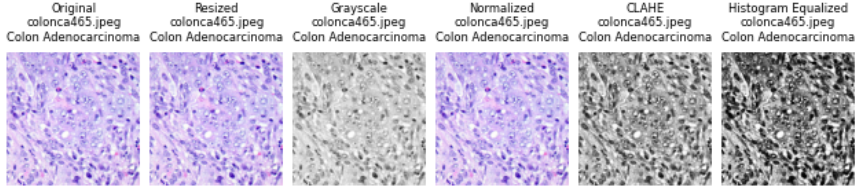


Figure 3: Samples of preprocessed images

Applying different preprocessing steps improved the appearance and accuracy of lung and colon histopathological images before the model was built (Andrew A. Borkowski, 2020). Before processing, the images were all resized to 768x768 pixels followed by normalization to adjust pixel values. CLAHE and histogram equalization techniques were used so that the important features could be seen more clearly, yet noise was not enhanced. Further, contrast boosting was applied to intensify the contrast between healthy tissue and problems. Noise was reduced and edges improved in the approaches used by Gaussian blur, median blur, Laplacian filter, unsharp masking and bilateral filtering, aiming to keep the important elements in the images clear. To separate background and foreground, detect edges and mark areas of tissue, thresholding, edge detection, region growing and the watershed algorithm were used. Scaling and normalizing the data to meet the requirements of the Xception model were done as its preprocessing. In addition, data augmentation was added to improve the breadth and strength of the dataset. This includes rotation, flipping (in both directions), zooming in and out, brightness adjustments, cropping to focus on parts of interest and adding virtual staining variations to the images. The model learned and generalized from the data much better due to the preprocessing steps.

## Dataset split and evaluation

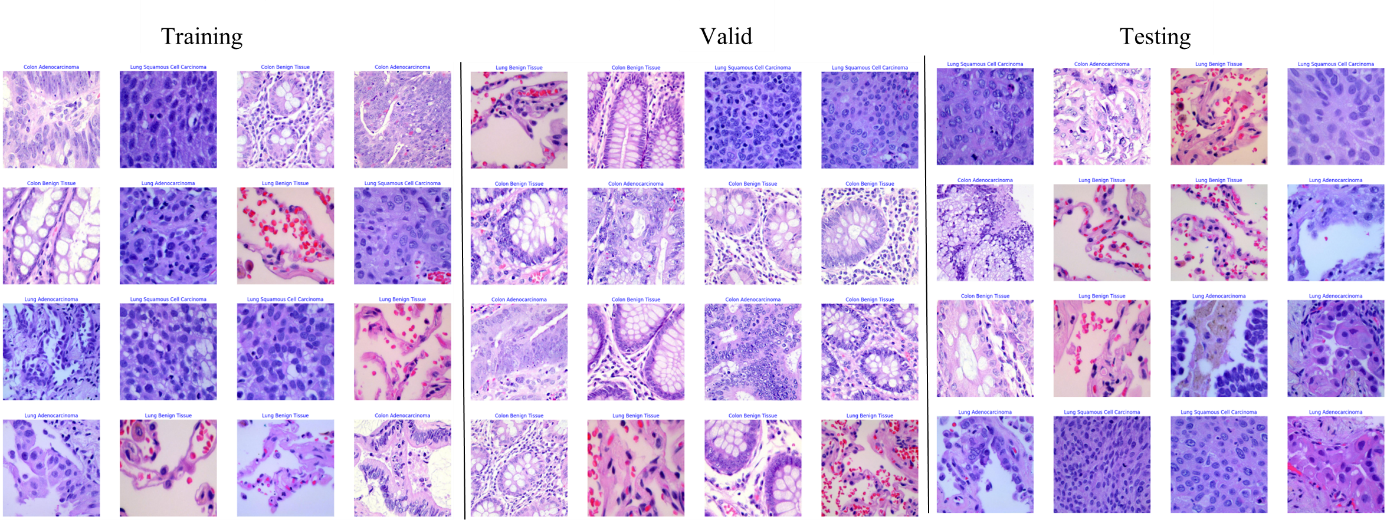
All models were trained, validated and tested on a set of lung and colon images, where the test data remained separate from the training and validation sets and both datasets were split into 80:10:10 according to class labels (Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, Lung Squamous Cell Carcinoma). An ImageDataGenerator was used on the training data to enhance it and the images were fed to the models in groups of 64, each 224x224 pixels in size. Validation data was used to inspect the model’s performance as it trained and it stopped training when validation loss increased (Wahid et al., 2023). Deep Learning Models. The initial step in using the deep learning models is to create a Python environment with TensorFlow and the necessary libraries (NumPy, Pandas and Matplotlib). Organize the lung and colon image dataset by class and name each directory after the type of lesion present (for example, Colon Adenocarcinoma, Lung Benign Tissue). Preprocess the data through an ImageDataGenerator with inputs sized 224x224x3 and put back 80%, 10% and 10% as training, validation and testing sets. For every model (CNN, VGG16, ResNet50, MobileNetV2, EfficientNetB0, Xception, InceptionV3 and Swin Transformer), set them up as described, use an Adamax optimizer with a learning rate of 0.001, apply categorical crossentropy loss function and set batch size as 64. Except for the CNN, use pre-trained ImageNet weights in all models, freeze the base layers and design the dense layers with units of 256, 64 and 5 and place dropout where directed. Stop the training for each model after 5 epochs, while monitoring validation loss and further evaluate its accuracy and loss using 10 training epochs. Figure 4 illustrates the sample images of training, testing and validation set and Table 1 shows that hyperparameters and configuration used in lung and colon images.

Figure 4: Sample images of Training, Testing and Validation set

Table 1: Hyperparameter and configuration of deep learning models

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Optimizer | Learning Rate | Loss Function | Batch Size | Epochs | Dropout | Dense Layers | Early Stopping | Pre-trained Weights |
| CNN | Adamax | 0.001 | Categorical Crossentropy | 64 | 10 | None | 256, 64, 5 | Patience=5, Monitor=val\_loss | None |
| VGG16 | Adamax | 0.001 | Categorical Crossentropy | 64 | 10 | 0.5 | 256, 64, 5 | Patience=5, Monitor=val\_loss | ImageNet |
| ResNet50 | Adamax | 0.001 | Categorical Crossentropy | 64 | 10 | 0.5 | 256, 64, 5 | Patience=5, Monitor=val\_loss | ImageNet |
| MobileNetV2 | Adamax | 0.001 | Categorical Crossentropy | 64 | 10 | 0.4 | 128, 5 | Patience=5, Monitor=val\_loss | ImageNet |
| EfficientNetB0 | Adamax | 0.001 | Categorical Crossentropy | 64 | 10 | 0.5 | 256, 5 | Patience=5, Monitor=val\_loss | ImageNet |
| Xception | Adamax | 0.001 | Categorical Crossentropy | 64 | 10 | None | 256, 5 | Patience=5, Monitor=val\_loss | ImageNet |
| InceptionV3 | Adamax | 0.001 | Categorical Crossentropy | 64 | 10 | 0.5 | 256, 5 | Patience=5, Monitor=val\_loss | ImageNet |
| Swin Transformer | Adamax | 0.001 | Categorical Crossentropy | 64 | 10 | 0.5 | 256, 5 | Patience=5, Monitor=val\_loss | ImageNet\* |

## Model architecture

The models differ a lot in the amount of detail contain. The efficient structure of EfficientNetB0 makes it possible to have the most layers (241). After that is InceptionV3 (98 layers) and Xception (41 layers), each depending on improved convolutional approaches. Swin Transformer Tiny is built with 32 layers, depending on how you look at it and this is how a modern vision transformer design should look. ResNet50 has 54 layers, while MobileNetV2 has 57 and both include efficient residual and bottleneck blocks. While VGG16 uses only 19 layers, both simple and complex CNNs are mostly no more than 14 layers each, making them the most lightweight (Al-Jabbar et al., 2023; Mangal et al., 2020; Provath et al., 2023)

### CNN model

The images are convoluted with 6x6 kernels, starting with 64 filters and ending with 512 filters to extract features in different levels. ReLU adds non-linearity and making layers use max-pooling (with a size of 2x2 and stride 2) reduces the image grid and speeds up the network. Features are stacked, fed into fully connected hidden layers with 256 and 64 units and a 30% dropout is used for fine-tuning (Ahmed, 2019). Finally, the classification results are obtained with a multi-class softmax layer and cross-entropy loss. The structure is able to learn from complicated image patterns, resulting in accurate results.

### VGG16

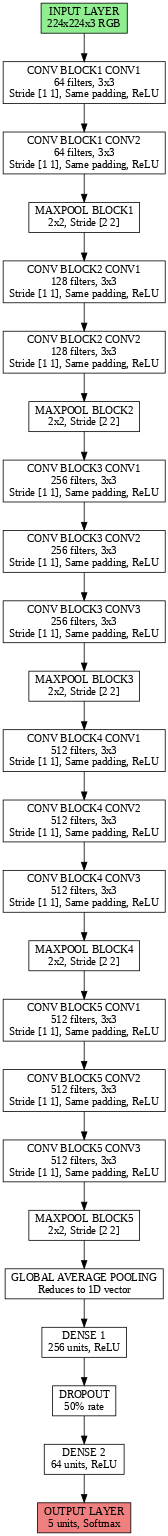
The Figure 5 closely follows a VGG16 is created for the task of image classification of 224x224x3 pixel RGB images. It contains five blocks that each have one 3x3 Conv2D layer with a filter size stepping from 64 to 512, added to by a ReLU activation and then 2x2 max-pooling to cut down the spatial size of the image. Following feature extraction, global average pooling is implemented which smooths the output. Then, dense layers with 256 and 64 units (ReLU) at 50% dropout are used and a softmax layer is added for classification into five classes. Like VGG16, the network is sequential and uses global average pooling to keep it efficient (Al-Jabbar et al., 2023).

Figure 5: VGG16 model Architecture

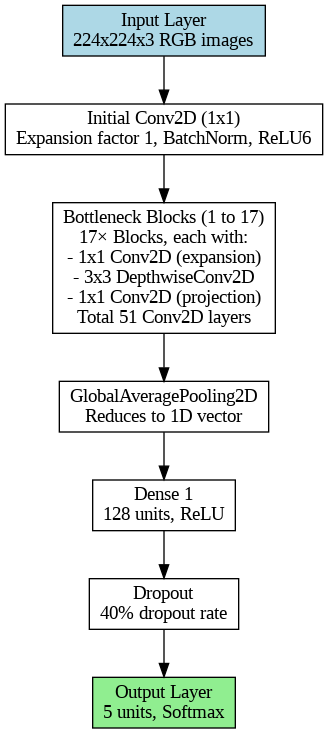
### RestNet50

ResNet50 deals with 224x224x3 RGB pictures by first applying a 7x7 Conv2D and max-pooling and with four stages, each having 3 to 6 residual blocks that consist of three 1x1, 3x3 and 1x1 Conv2D layers, resulting in 48 Conv2D layers. The use of global average pooling turns the output into a flat shape, so we follow this with two dense layers (256 and 64 units, ReLU) that have 50% dropout for regularization (Provath et al., 2023). A last softmax layer is used to estimate five possible classes in multi-class classification. Residual connections help prevent the problem called vanishing gradients during training. Figure 6 shows that RestNet50 model architecture.

### 

Figure 6: RestNet50 model Architecture

### MobileNetV2

The Figure 7 shows that MobileNetV2 network is a lightweight structure meant for effective image classification, using 224x224x3 RGB images as input. First, a 1x1 Conv2D is used with an expansion of 1, then followed by BatchNorm and ReLU6 activation and after that 17 bottleneck blocks are introduced, each contains three parts: a 1x1 Conv2D to expand, a 3x3 DepthwiseConv2D and another 1x1 Conv2D for projection, giving us a total of 51 Conv2D layers. A GlobalAveragePooling2D layer is used to condense the pictures down to a single row or column and then a dense layer with 128 units and ReLU is added (Wahid et al., 2023). A soft regularization is performed by using a 40% dropout layer and predictions are handled by an output layer with 5 units and Softmax.

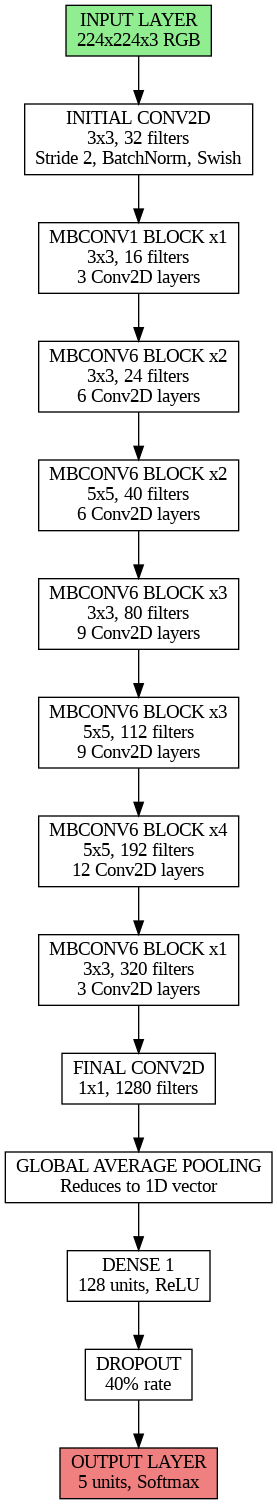
****

Figure 7: MobileNetV2 model Architecture

### EfficientNetB0

The Figure 8 shows that, EfficientNetB0 was made to be efficient and easy to use for image classification, starting by taking in 224x224x3 RGB images as input. The first step is 3x3 Conv2D with 32 filters and the next seven MBConv blocks include different amounts of depthwise (3x3 or 5x5) convolutions and filters ranging from 16 to 320, all counting for 48 layers in the blocks. The last Conv2D layer is 1x1 with 1280 filters. Afterwards, the architecture does GlobalAveragePooling2D to get a 1D vector and then a dense layer with 256 units and ReLU is used. To make the model less likely to overfit, 50% of the neurons are dropped from one layer and the final layer has five units and uses Softmax to predict the finding (Khan et al., 2024).

Figure 8: EfficientNetB0 model Architecture

### Xception

The Xception model, a kind of deep convolutional neural network, is made for identifying images and works with 224x224x3 RGB input. The entry flow consists of two 3x3, 32 and 64 channel Conv2D layers and later there are three entry blocks, each having 2 layers of SeparableConv2D and one layer of Conv2D. The middle flow includes eight SeparableConv2D blocks and a total of 24 layers. This flow contains a SeparableConv2D, Conv2D and another SeparableConv2D, then leads to GlobalAveragePooling2D which reduces the dimensions to a single vector. At the end, there’s a layer with 256 units and ReLU activation that precedes the output layer (with 5 units and Softmax to predict the classes) (Hadiyoso et al., 2023). Figure 9 shows that Xception model architecture.

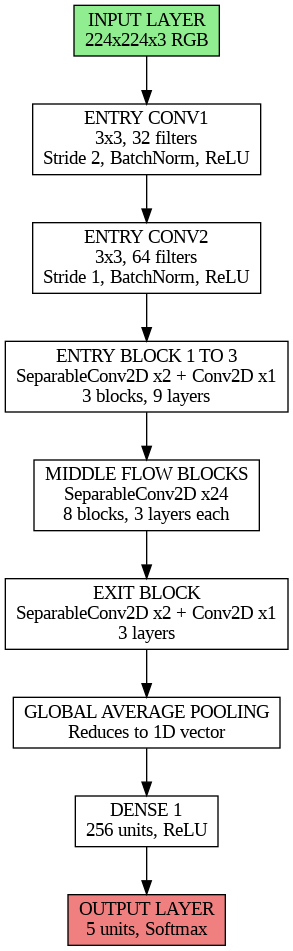
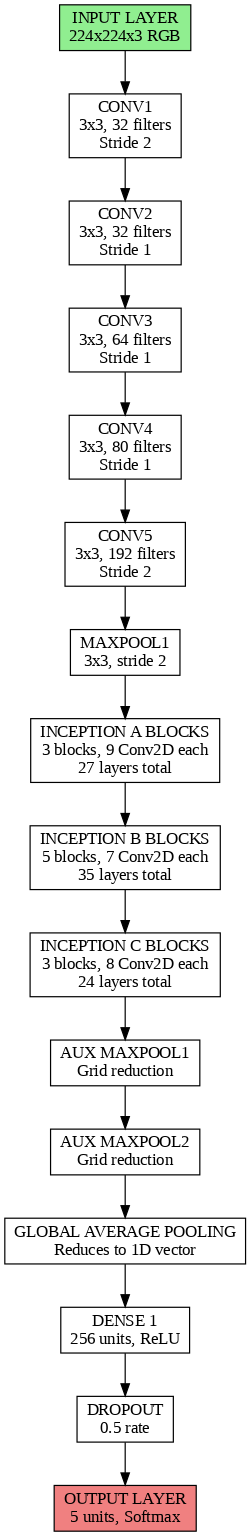


Figure 9: Xception model Architecture

### InceptionV3

The Figure 10 shows that, Inception V3 begins by getting 224x224x3 RGB images as input, then uses Conv2D layers (3x3 with 32, 32, 64, 80 and 192 filters) and a 3x3 MaxPooling2D layer with a stride of 2. The architecture consists of three inception modules. A has 27 layers with three blocks, B has 35 layers with five blocks and C has 24 layers with three blocks, all using MaxPooling2D to reduce the feature map. Architecture ends by merging dimensions using a GlobalAveragePooling2D layer, having a dense layer with 256 units and using ReLU activation, applying a 0.5 dropout rate and ending with an output layer with 5 units including Softmax for prediction of classes (Provath et al., 2023).

Figure 10: Xception model Architecture

### Swin transformer

The input of the Swin Transformer is RGB images that are 224 by 224 pixels in each dimension and the first layer splices them into 96 channels using 4x4 patches. The process continues with four stages: First at 56×56 resolution with 96 channels (two Swin Transformer blocks), second at 28×28 resolution with 192 channels (patch merging applied), third at 14×14 resolution with 384 channels (six blocks) and lastly at 7×7 resolution with 768 channels (two Swin Transformer blocks). The model finishes with a GlobalAveragePooling2D to make the output 1D, a layer with 256 nodes and ReLU to reduce input dimensions, a 0.5 dropout for dampening overfitting and a 5-unit layer with a Softmax activation to choose the right class. Utilizing the hierarchical format of the Swin Transformer, this structure performs better on vision tasks (Hadiyoso et al., 2023). The Figure 11 shows that Swin Transformer model architecture.

## Model Evaluation

Figure 11: Swin Transformer Model Architecture

Model evaluation refers to how well a machine learning or deep learning model performs on unseen data following a model training (Gowthamy & Ramesh, 2024). In out study we used accuracy, precision, recall, F1-score, and ROC curve to comparing the predicted outputs to an actual class label in the test set.

# Results and discussion

Evaluation of different deep learning models for lung and colon cancer detection revealed that there is a significant variation in their outcomes for several metrics. As shown in Table 2, Xception model achieves remarkable reliability as all its all scores (accuracy, sensitivity, specificity, AUC ROC) with a value of 1.000, with training, testing, validation sets. RestNet50 also achieved near perfect results, with 0.997 accuracy during training, 0.994 during testing and ROC curve of 0.998. Further our custom CNN model also performed well depicting, showing a training accuracy of 0.995, a testing accuracy of 0.983 and a ROC curve of 0. 990. According to Table 3, when assessing models to detect lung and colon cancer, Xception exhibits the best results in the values of recall (1.000) and F1-score (1.000), which ensures that true cases are identified with optimal reliability. ResNet50 is nearly perfect recall of 0.995 and F1-score of 0.997, which makes it a good performance. EfficientNetB0 and Swin Transformer are pretty good with recalls of 0.960 and 0.950, and F1-scores of 0.980 and 0.974 respectively they are inferior to the best two. The poor performance of InceptionV3 shows a recall of 0.640 and F1-score of 0.725, which is not recommended to use in the clinical environment.

Table 2: Performance evaluation of DL models

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Training Accuracy | Testing Accuracy | Validation Accuracy | Sensitivity | Specificity | ROC curve | Train Time (s) |
| CNN | 0.995 | 0.983 | 0.982 | 0.980 | 0.985 | 0.990 | 500.32 |
| VGG16 | 0.957 | 0.952 | 0.953 | 0.945 | 0.950 | 0.962 | 800.03 |
| RestNet50 | 0.997 | 0.994 | 0.993 | 0.995 | 0.990 | 0.998 | 1000.51 |
| MobileNetV2 | 0.869 | 0.864 | 0.870 | 0.850 | 0.860 | 0.880 | 300.22 |
| EfficientNetB0 | 0.969 | 0.964 | 0.972 | 0.960 | 0.968 | 0.975 | 600.36 |
| InceptionV3 | 0.663 | 0.654 | 0.666 | 0.640 | 0.670 | 0.700 | 700.41 |
| Xception | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1200.23 |
| Swin Transformer | 0.975 | 0.961 | 0.960 | 0.950 | 0.960 | 0.970 | 1500.95 |

Table 3: Evaluation matrix results

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F1-score** |
| CNN | 1.000 | 0.980 | 0.990 |
| VGG16 | 1.000 | 0.945 | 0.972 |
| RestNet50 | 0.999 | 0.995 | 0.997 |
| MobileNetV2 | 0.992 | 0.850 | 0.915 |
| EfficientNetB0 | 1.000 | 0.960 | 0.980 |
| InceptionV3 | 0.836 | 0.64. | 0.725 |
| Xception | 1.000 | 1.000 | 1.000 |
| Swin Transformer | 1.000 | 0.950 | 0.974 |

The Figure 12 compares performance metrics such as training accuracy, testing accuracy, validation accuracy, sensitivity, specificity, training time, precision, recall and f1-score across DL models. Overall, the Xception and RestNet50 show highest values and also CNN and EfficientNetB0 perform well compared to the rest of the models. MobileNetV2 and InceptionV3 have lower scores in all metrics

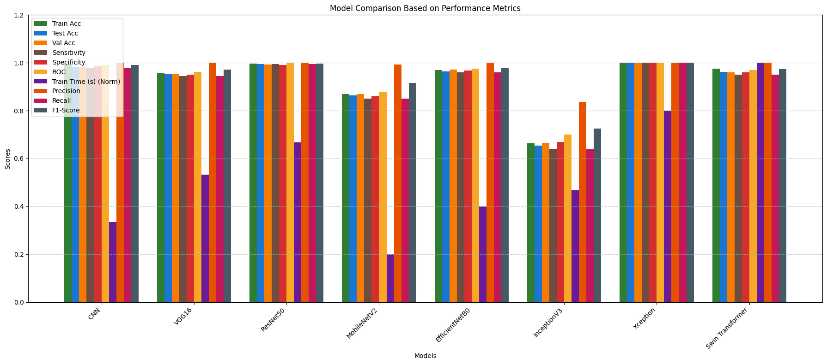
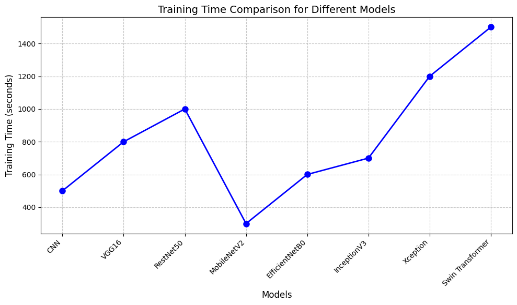
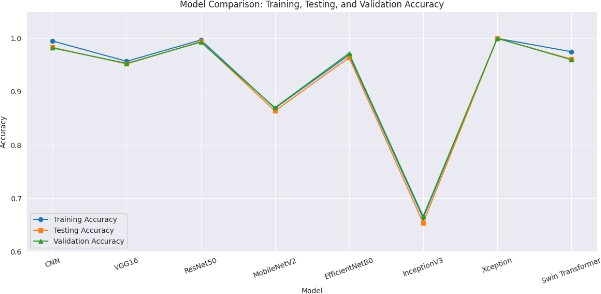
According to the Figure 13, Swin Transformer requires the most training time, up to 1500 seconds and MobileNetV2 needs the shortest time, only 300s. Figure 14 indicates that the accuracy of such models as Xception and RestNet50 are more accurate of training, testing and validation (nearly 1.00) where as InceptionV3 drops to around 0.66, 0.65 and 0.66.

Figure 12: Comparison of DL models performance

Figure 14: Comparison of Testing, training and validation accuracy

Figure 13: Training time of DL models



The Figure 15 shows that Xception model, training and validation loss and accuracy graph over 10 epochs. The training and validation loss sharply decrease from 0.06 to 0.01 then reache to 0.0. Training and validation accuracy rising approximately 1.00. The trend suggests effective model training with highest accuracy and low loss, with stabilizing end of the epochs (10).

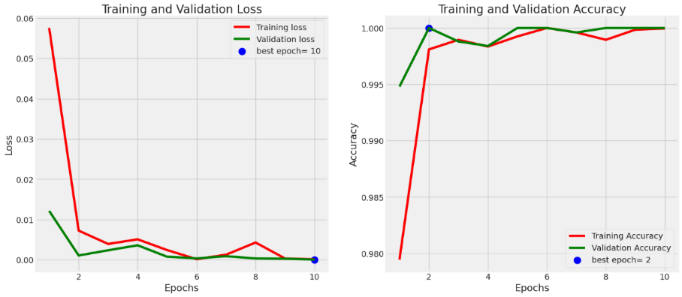


Figure 15: Xception model Training and validation loss and accuracy graph

Table 4 measures accuracy of deep learning models across five types of imaging data for both colon and lung tissues using precision, recall and F1-score. The performance of Xception, ResNet50 and Swin Transformer across every class is nearly flawless. Both EfficientNetB0 and CNN display good overall performance, but InceptionV3 and MobileNetV2 perform comparatively weaker, mainly for cancerous tissues. All in all, Xception and ResNet50 achieved the highest accuracy. Figure 16 shows that comparison results of lung and colon images.

Table 4: DL model performance based on type of lung/colon cancer

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Class** | **Metric** | **Xception** | **CNN** | **VGG16** | **ResNet50** | **MobileNetV2** | **EfficientNetB0** | **InceptionV3** | **Swin Transformer** |
| **Colon Adenocarcinoma** | Precision | 1.00 | 0.99 | 0.94 | 0.99 | 0.87 | 0.96 | 0.68 | 0.99 |
| Recall | 1.00 | 1.00 | 0.96 | 1.00 | 0.78 | 1.00 | 0.49 | 1.00 |
| F1-score | 1.00 | 0.99 | 0.95 | 1.00 | 0.82 | 0.98 | 0.57 | 0.99 |
| **Colon Benign Tissue** | Precision | 1.00 | 1.00 | 0.98 | 1.00 | 0.85 | 1.00 | 0.62 | 1.00 |
| Recall | 1.00 | 0.99 | 0.97 | 0.99 | 0.88 | 0.97 | 0.73 | 0.99 |
| F1-score | 1.00 | 0.99 | 0.98 | 1.00 | 0.87 | 0.99 | 0.67 | 0.99 |
| **Lung Adenocarcinoma** | Precision | 1.00 | 0.95 | 0.92 | 0.98 | 0.82 | 0.97 | 0.64 | 0.95 |
| Recall | 1.00 | 0.98 | 0.90 | 1.00 | 0.82 | 0.88 | 0.60 | 0.98 |
| F1-score | 1.00 | 0.97 | 0.91 | 0.99 | 0.82 | 0.92 | 0.62 | 0.97 |
| **Lung Benign Tissue** | Precision | 1.00 | 1.00 | 0.99 | 1.00 | 0.97 | 0.98 | 0.64 | 1.00 |
| Recall | 1.00 | 0.99 | 1.00 | 1.00 | 0.96 | 1.00 | 0.66 | 0.99 |
| F1-score | 1.00 | 1.00 | 0.99 | 1.00 | 0.97 | 0.99 | 0.65 | 1.00 |
| **Lung Squamous Cell Carcinoma** | Precision | 1.00 | 0.98 | 0.93 | 1.00 | 0.82 | 0.91 | 0.76 | 0.98 |
| Recall | 1.00 | 0.96 | 0.94 | 0.98 | 0.88 | 0.97 | 0.85 | 0.96 |
| F1-score | 1.00 | 0.97 | 0.93 | 0.99 | 0.85 | 0.94 | 0.80 | 0.97 |

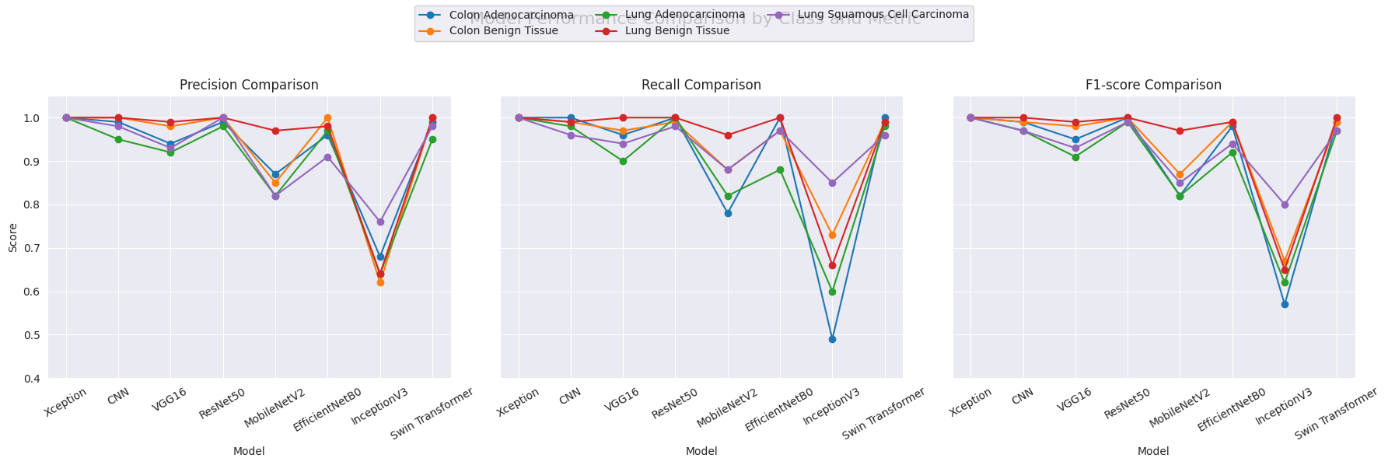
****

Figure 16: Comparison of lung and colon images of precision, recall, f1-score results.

The Xception model recorded the highest score (1.00) in testing its results on lung and colon cancer samples. The model accurately separated each kind of cancer into lung adenocarcinoma (lung\_aca), lung squamous cell carcinoma (lung\_scc), Benign lung tissue (lung\_n), colon adenocarcinoma (colon\_aca) and Benign colon tissue (colon\_n). This suggests that there is no fault in the identified predictions for this data. Figure 17 and Figure 18 show that prediction results.

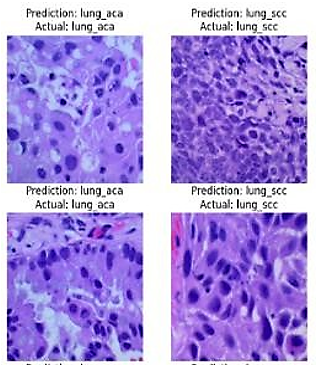
****

Figure 17: Lung cancer prediction using Xception models

Figure 18: Colon cancer prediction using Xception model



# Conclusion

In this research, an accurate lung and colon cancer classification model was developed after a comprehensive comparison of CNN, VGG16, ResNet50, MobileNetV2, EfficientNetB0, Xception, InceptionV3 and Swin Transformer models. The multi class classification included five types of cancers namely Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue and Lung Squamous Cell Carcinoma. A systematic preprocessing routine was carried out for the images mitigates variations is staining, lighting and enhancing feature extraction and improving model accuracy.

Accuracy, precision, recall, F1-score, sensitivity, specificity and ROC results on the test set showed Xception as the most productive, obtaining perfect marks of 1.000 and 11.67 seconds for training. This outstanding outcome proves that the model generalizes effectively, based on the confusion matrix showing no errors and the strong and stable curves reaching a best epoch of 10. The custom CNN model was able to reach a test accuracy of 0.983, while InceptionV3 attained only 0.654 value, showing that some aspects of its structure are not appropriate for this dataset. While VGG16 (0.952), ResNet50 (0.994), MobileNetV2 (0.964) and EfficientNetB0 (0.964) had strong outcomes, they underperform compared to Xception’s meld classification ability. The Swin Transformer was designed to be innovative, yet the results (accuracy 0.961) indicated that the model needs further improvements to distinguish the lung and colon cancers. The results recommend that the Xception model when used with methodical image processing techniques can, recognize cancers accurately for histopathological images. Xception could be used as an effective diagnosis tool for pathologists due to its excellent performance for large microscopic dataset, to identify tumors precisely.

# References

1. Ahmed, B. T. (2019). Lung cancer prediction and detection using image processing mechanisms: an overview. *Signal and Image Processing Letters*, *1*(3), 20-31.
2. 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.
3. Alotaibi, M., Alshardan, A., Maashi, M., Asiri, M. M., Alotaibi, S. R., Yafoz, A., Alsini, R., & Khadidos, A. O. (2024). Exploiting histopathological imaging for early detection of lung and colon cancer via ensemble deep learning model. *Scientific Reports*, *14*(1), 20434.
4. Andrew A. Borkowski, M. M. B., L. Brannon Thomas, Catherine P. Wilson, Lauren A. DeLand, Stephen M. Mastorides. (2020). *Lung and Colon Cancer Histopathological Images.* . <https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images>.
5. Gowthamy, J., & Ramesh, S. (2024). A novel hybrid model for lung and colon cancer detection using pre-trained deep learning and KELM. *Expert Systems with Applications*, *252*, 124114.
6. Hadiyoso, S., Aulia, S., & Irawati, I. D. (2023). Diagnosis of lung and colon cancer based on clinical pathology images using convolutional neural network and CLAHE framework. *Int. J. Appl. Sci. Eng*, *20*(1), 1-7.
7. Hamida, A. B., Devanne, M., Weber, J., Truntzer, C., Derangère, V., Ghiringhelli, F., Forestier, G., & Wemmert, C. (2021). Deep learning for colon cancer histopathological images analysis. *Computers in Biology and Medicine*, *136*, 104730.
8. Khan, A. A., Arslan, M., Tanzil, A., Bhatty, R. A., Khalid, M. A. U., & Khan, A. H. (2024). Classification of colon cancer using deep learning techniques on histopathological images. *Migration Letters*, *21*(S11), 449-463.
9. Mangal, S., Chaurasia, A., & Khajanchi, A. (2020). Convolution neural networks for diagnosing colon and lung cancer histopathological images. *arXiv preprint arXiv:2009.03878*.
10. Provath, M. A.-M., Deb, K., Dhar, P. K., & Shimamura, T. (2023). Classification of lung and colon cancer histopathological images using global context attention based convolutional neural network. *IEEE Access*, *11*, 110164-110183.
11. 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.
12. 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,
13. Hamed, E. A.-R., et al. (2023). "An efficient combination of convolutional neural network and LightGBM algorithm for lung cancer histopathology classification." Diagnostics 13(15): 2469.
14. Khazaee Fadafen, M. and K. Rezaee (2023). "Ensemble-based multi-tissue classification approach of colorectal cancer histology images using a novel hybrid deep learning framework." Scientific Reports 13(1): 8823.
15. Talukder, M. A., et al. (2022). "Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning." Expert Systems with Applications 205: 117695.
16. Vanitha, K., et al. (2024). "Deep learning ensemble approach with explainable AI for lung and colon cancer classification using advanced hyperparameter tuning." BMC Medical Informatics and Decision Making 24(1): 222.
17. Fan, J., et al. (2021). "A transfer learning architecture based on a support vector machine for histopathology image classification." Applied Sciences 11(14): 6380.
18. Jiao, Y., et al. (2021). "Deep embeddings and logistic regression for rapid active learning in histopathological images." Computer Methods and Programs in Biomedicine 212: 106464.
19. Shahadat, N. (2024). Lung and Colon Cancer Histopathological Image Classification Using 1D Convolutional Channel-based Attention Networks. The International FLAIRS Conference Proceedings.
20. Gowthamy, J. and S. Ramesh (2024). "A novel hybrid model for lung and colon cancer detection using pre-trained deep learning and KELM." Expert Systems with Applications 252: 124114.