RESEARCH

Open Access

Deep learning ensemble approach with explainable AI for lung and colon cancer classification using advanced hyperparameter tuning

K. Vanitha[1], Mahesh T. R[2], S. Sathea Sree[3] and Suresh Guluwadi[4*]

Abstract

Lung and colon cancers are leading contributors to cancer-related fatalities globally, distinguished by unique histo-pathological traits discernible through medical imaging. Effective classification of these cancers is critical for accurate diagnosis and treatment. This study addresses critical challenges in the diagnostic imaging of lung and colon cancers, which are among the leading causes of cancer-related deaths worldwide. Recognizing the limitations of existing diagnostic methods, which often suffer from overfitting and poor generalizability, our research introduces a novel deep learning framework that synergistically combines the Xception and MobileNet architectures. This innovative ensemble model aims to enhance feature extraction, improve model robustness, and reduce overfitting.

Our methodology involves training the hybrid model on a comprehensive dataset of histopathological images, followed by validation against a balanced test set. The results demonstrate an impressive classification accuracy of 99.44%, with perfect precision and recall in identifying certain cancerous and non-cancerous tissues, marking a significant improvement over traditional approach.

The practical implications of these findings are profound. By integrating Gradient-weighted Class Activation Mapping (Grad-CAM), the model offers enhanced interpretability, allowing clinicians to visualize the diagnostic reasoning process. This transparency is vital for clinical acceptance and enables more personalized, accurate treatment planning. Our study not only pushes the boundaries of medical imaging technology but also sets the stage for future research aimed at expanding these techniques to other types of cancer diagnostics.

Keywords Enhanced deep learning, Histopathology, Lung cancer diagnosis, Colon cancer detection, Ensemble learning models, Xception and MobileNet integration, Gradient-weighted Class Activation Mapping (Grad-CAM), Medical image analysis, Automated cancer classification, Clinical AI applications

*Correspondence:
Suresh Guluwadi
suresh.guluwadi@astu.edu.et
Full list of author information is available at the end of the article

Introduction

Lung and colon cancers represent two of the most significant challenges in the field of oncology due to their high incidence and mortality rates globally. Effective and early diagnosis of these cancers is paramount for improving patient outcomes and survival rates. Histopathological examination of tissue samples, performed via microscopy, remains the gold standard for diagnosing these malignancies. However, the interpretation of histopathological images is highly dependent on the expertise of the pathologist and can be subject to variability. This underscores the need for more robust, accurate, and reproducible diagnostic methods that can augment human expertise.

In recent years, the field of medical imaging has been revolutionized by the advent of deep learning technologies, which have shown promise in enhancing the analysis of medical images, including histopathological images of cancerous tissues. Deep learning, a subset of machine learning, leverages neural networks with multiple layers to analyze various levels of abstraction in data. In medical applications, these technologies have been employed to identify, classify, and predict disease patterns with remarkable success.

Traditionally, the diagnosis of lung and colon cancers involves the histological examination of tissue sections under a microscope [1]. Pathologists review these sections to identify morphological features indicative of malignant transformations. The accuracy of these diagnoses heavily relies on the individual pathologist's skill and experience, which can lead to significant inter-observer variability. Moreover, the process can be time-consuming and is often limited by the volume of cases that a pathologist can examine within a given timeframe. Figure 1 depicts the histopathological images of lung and colon cancer.

While automated systems have been developed to assist pathologists by quantifying histopathological features, these systems often rely on conventional machine learning techniques that require manual feature extraction. This process can be labor-intensive and may not capture the nuanced features of the cancerous tissues effectively. Moreover, many existing automated systems do not generalize well across different datasets or medical centers due to variations in staining techniques, image capture conditions, and intrinsic biological heterogeneity.

Deep learning models, particularly convolutional neural networks (CNNs), have emerged as a powerful tool for medical image analysis, capable of automating feature extraction and providing robust generalizations across diverse datasets [2]. These models learn
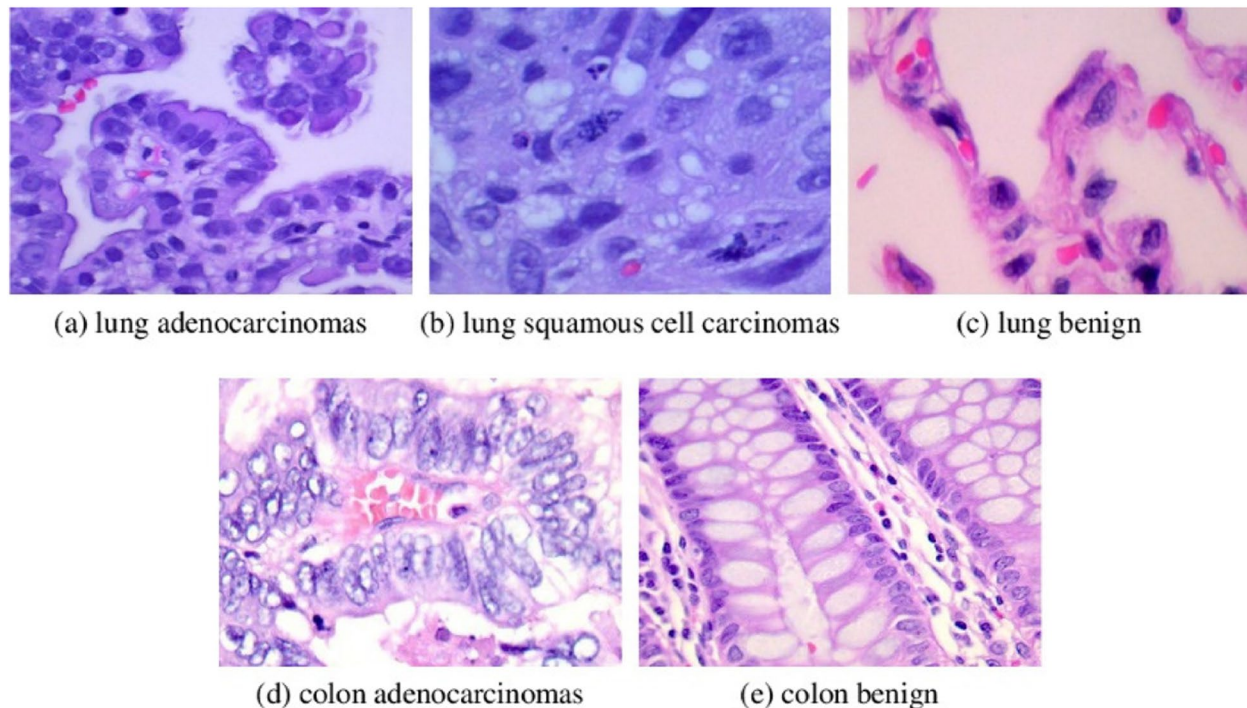

(a) lung adenocarcinomas   (b) lung squamous cell carcinomas   (c) lung benign

(d) colon adenocarcinomas   (e) colon benign

Fig. 1 Histopathological images of lung colon cancer

Vanitha *et al. BMC Medical Informatics and Decision Making* (2024) 24:222

Page 3 of 18

to identify disease markers directly from the images, reducing the reliance on manual feature labeling and potentially decreasing the variability introduced by human interpretation.

The Xception architecture, which extends the principles of Inception by replacing standard convolutions with depthwise separable convolutions, offers a high degree of model adaptability and learning capacity. On the other hand, MobileNet, designed for mobile and resource-constrained environments, utilizes depthwise separable convolutions to create lightweight, efficient models. By combining these two architectures, the approach harnesses both the depth and efficiency of these models, allowing for a sophisticated analysis of histopathological images that is both accurate and computationally feasible.

The proposed system integrates these two powerful architectures in a novel ensemble approach where both networks are first pre-trained on the ImageNet dataset to learn a wide range of image features. These pre-trained models are then fine-tuned on a curated dataset of lung and colon histopathological images, ensuring that the model specializes in features that are most predictive of cancerous conditions. The outputs of these networks are concatenated and fed into a series of dense layers, culminating in a classification layer that distinguishes between the different cancerous and non-cancerous tissue types.

An integral component of the methodology is the use of Gradient-weighted Class Activation Mapping (Grad-CAM), which provides visual explanations for the decisions made by the model. This technique generates heatmaps that highlight the areas within the images most influential to the model's predictions, thereby offering insights into what the model is "seeing" when it makes a diagnosis. This transparency is crucial for clinical acceptance of AI tools, as it provides pathologists with understandable and interpretable evidence of the model's diagnostic pathways.

Motivation

The motivation for this research stems from the critical need to enhance the accuracy and consistency of cancer diagnostics. Despite advances in imaging and computational tools, the potential of deep learning in medical imaging, particularly in improving the analysis and interpretation of histopathological images, has not been fully exploited. Existing automated systems often struggle with overfitting, limited generalizability across different datasets, and the inability to capture subtle yet crucial features in images. There is a pressing need for a robust model that not only improves diagnostic accuracy but also offers insights into its decision-making process, thereby aiding pathologists and enhancing trust in automated systems.

The objective of the paper is to:

- Concatenate the outputs of Xception and MobileNet and feed them into dense layers, leading to a classification layer that distinguishes between cancerous and non-cancerous tissue types, enhancing diagnostic accuracy.
- Utilize Gradient-weighted Class Activation Mapping (Grad-CAM) to generate heatmaps, offering visual explanations for the model's decisions and providing transparent insights into its diagnostic pathways, crucial for clinical acceptance.
- Contribute to advancing medical diagnostics by providing a scalable solution applicable to various cancers and complex medical imaging tasks, fostering more automated, precise, and reliable diagnostic processes to meet the demands of modern healthcare systems.

The paper is structured as follows: Related Work reviews deep learning in cancer diagnosis. Methodology details the model, dataset, training, and Grad-CAM. Results and Discussion presents performance metrics and Grad-CAM findings. Conclusion and Future Work summarizes findings, applications, and future research.

Related work

The application of artificial intelligence in medical imaging, particularly using deep learning techniques, has marked a significant progression in the diagnosis and characterization of various cancers. The utilization of convolutional neural networks (CNNs) has been extensively studied due to their ability to automatically detect intricate patterns in imaging data that are often indiscernible to human eyes.

Before the widespread adoption of deep learning, several traditional machine learning methods were employed in medical image analysis. Techniques such as Support Vector Machines (SVM), Random Forests, and k-Nearest Neighbors (kNN) were commonly used. These models often required meticulous feature engineering and were generally limited by the handcrafted nature of feature extraction which impacted their efficacy in complex image classification tasks.

The shift from traditional models to CNNs introduced a significant breakthrough in medical diagnostics. CNNs, such as AlexNet, VGG, and Inception, have been applied to classify and predict various forms of cancers with promising results. These models eliminate the need for manual feature extraction by learning image representations directly from the data, leading to improved accuracy and efficiency in medical image analysis.

For lung and colon cancer, several studies have highlighted the use of specific architectures like MobileNet and Xception. MobileNet is a model designed for efficient performance in mobile and embedded vision applications, which has been adapted for medical imaging to handle resource constraints without compromising diagnostic accuracy. Xception model, which employs depthwise separable convolutions and has demonstrated superior performance on tasks requiring high-level feature extraction, such as distinguishing between different types of lung and colon tissue samples [3].

Recent studies have explored the integration of multiple deep learning models to leverage the strengths of various architectures. For instance, the combination of MobileNet and Xception has been examined for its potential to enhance feature representation and reduce overfitting, providing a more robust analysis of complex medical images. This approach aligns with the ongoing trend of ensemble models in deep learning, where multiple networks are used to improve predictive performance and reliability.

The importance of model interpretability in medical AI has also been increasingly recognized. Techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) allow clinicians to visualize which areas of an image influence the network's predictions. This transparency is crucial for clinical acceptance, as it aids in the verification of AI-driven diagnostics by highlighting the decision-making process of the model.

Table 1 Literature review

Authors	Dataset	Remarks
Bruntha, P. Malin, et al. (2021) [4]	LIDCIRRI dataset	Developed ConvLung, a two-layered CNN. Xception and Inception-ResNet50v2 showed superior performance
Ibrahim, Dina M., Nada M. Elshennawy, and Amany M. Sarhan. (2021) [5]	Digital chest x-ray and CT datasets with four classes	Multi-classification model diagnosing COVID-19, pneumonia, and lung cancer, using VGG19-CNN and ResNet152V2 variants
Chouat, Ines, et al. (2022) [6]	Chest X-ray (CXR) images	Introduced two CNN models using transfer learning to detect pneumonia
Malik, Hassaan, and Tayyaba Anees (2022) [7]	Eight COVID-19 datasets	BDCNet integrates Vgg-19 with CNN for multi-classification, outperforming other models
Aitazaz, Tayyab, et al. (2022) [8]	Lung and colon histopathology image dataset	Transfer Learning model for detection of histopathology images showcases ViT Performance
Kaya, Yasin, and Ercan Gürsoy **(2023)** [9]	IQ-OTH/NCCD and LIDC-IDRI	Improved COVID-19 classification using novel fine-tuning in deep transfer learning
Karaddi, Sahebgoud Hanamantray, and Lakhan Dev Sharma (2023) [10]	CXR-image data	Utilized eight pre-trained CNNs for lung disease classification, highlighting Densenet-201's performance
Sharma, Neelam, et al. (2023) [11]	Lungs Disease Dataset	Explored various models including ResNet, DenseNet, VGG-16, EfficientNet, Xception, and MobileNet for lung disease diagnosis
Sethy, Prabira Kumar, et al. (2023) [12]	Lung and colon histopathology image dataset	Developed a hybrid model for the detection of lung colon cancer
Chhagan Patil, Nandkishor, and Nitin Jagannath Patil (2024) [13]	Chest X-ray Dataset	Hybrid model combining VGG19 and Capsule Network for lung cancer recognition, improving orientation challenges
Seth, Amit, and Vandana Dixit Kaushik (2024) [14]	Lung and colon histopathology image dataset	Used Swin Transformer, Cascade Classifier and Adaptive Tasmanian Devil Algorithm
Chutia, Upasana, et al. (2024) [15]	NIH Chest X-rays	Enhanced X-ray images using CLAHE; employed DenseNet201 with hybrid pooling and channel attention mechanisms
Chiet, Chai Chee, et al. (2024) [16]	CT Scan Images	Used pre-trained CNNs for lung cancer detection, indicating potential for further optimization

Vanitha *et al. BMC Medical Informatics and Decision Making*   (2024) 24:222

Page 5 of 18

The integration of deep learning technologies in medical imaging, especially for lung and colon cancer diagnosis, continues to evolve. This paper builds on these developments by proposing an innovative model that combines the strengths of MobileNet and Xception, enhanced by interpretative mechanisms such as Grad-CAM, to set new standards in accuracy and reliability in histopathological image analysis.

To provide a structured comparison of these studies, the Table 1 summarizes their contributions, methodologies, and key findings.

This table illustrates the diversity and progression in methodologies from pure image classification models to complex architectures integrating features like attention mechanisms and hybrid model configurations. Each study contributes uniquely towards refining the diagnostic processes for lung and colon cancers, demonstrating the ongoing advancements and potential of deep learning in medical imaging. This literature review forms the backdrop against which the current research is positioned, aiming to leverage and expand upon these recent innovations.

Methodology

This research investigates the application of convolutional neural networks (CNNs) in the classification of histopathological images of lung and colon cancer. The primary objective is to compare the performance of two well-known architectures, MobileNet and Xception, integrated into a single model framework to enhance classification accuracy. This section details the dataset, preprocessing steps, model architecture, training process, and evaluation metrics employed in the study. Figure 2 showcases the workflow of the proposed model.

Dataset description

The Lung and Colon Cancer Histopathological Image Dataset (LC25000), developed by a team of researchers, provides a robust resource for machine learning research in medical diagnostics. This dataset contains 25,000 high-quality, de-identified, and HIPAA-compliant histopathological images divided into five classes: Colon Adenocarcinoma, Benign Colonic Tissue, Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, and Benign Lung Tissue with each class contributing 5,000 images. These images are integral for developing algorithms capable of distinguishing between malignant and benign tissue samples in both colon and lung tissues [17]. Validated by expert pathologists and structured for machine learning applications, each image adheres to standardized formatting and resolution specifications to ensure consistency and reliability in training and testing AI models, thereby advancing the field of medical image analysis and the early detection of cancer. The Table 2 provides an overview of the dataset distribution across different classes. It helps illustrate the balance or imbalance within the dataset used for training the model. Figure 3 depicts the sample input images from lung and colon cancer histopathological image dataset.

Table 2 Dataset distribution

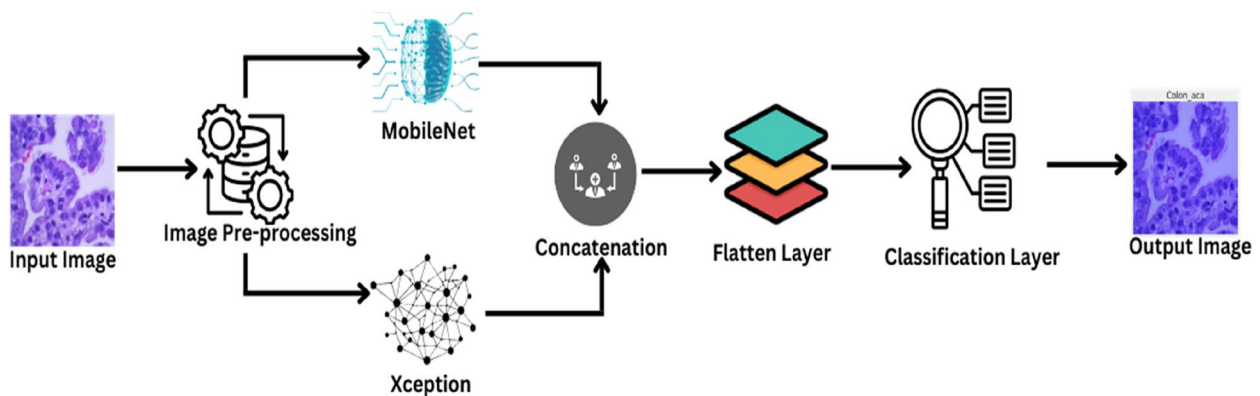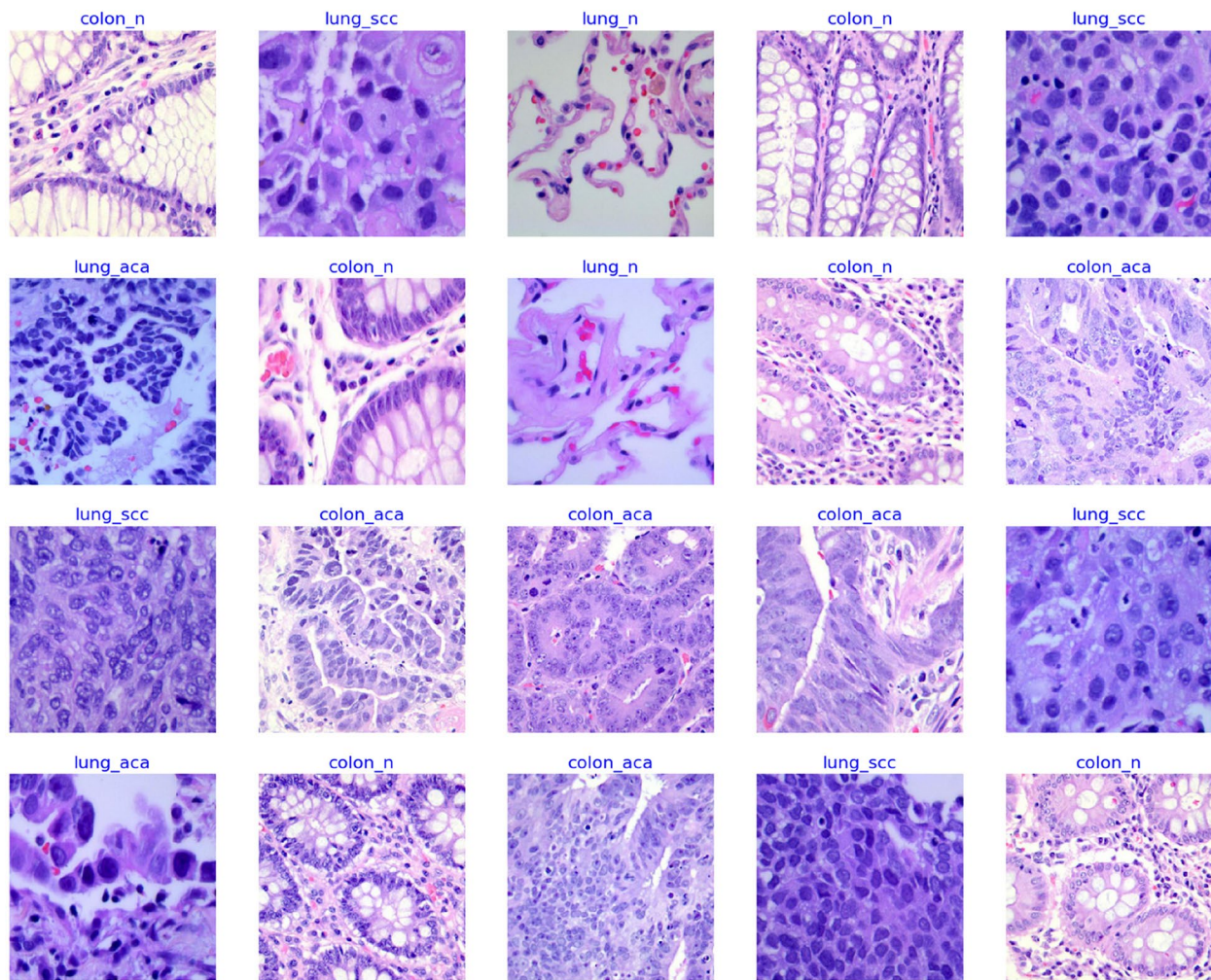| Class	Number of images
Colon Adenocarcinoma	5000
Benign Colonic Tissue	5000
Lung Adenocarcinoma	5000
Lung Squamous Cell Carcinoma	5000
Benign Lung Tissue	5000
Total	**25000**

Fig. 2 Workflow of the proposed model

Fig. 3 Sample input images from dataset

Data preprocessing

Data preprocessing is a crucial step in preparing raw images for effective model training. The first step involves resizing all images to 224×224 pixels to match the input size requirements of the pre-trained MobileNet and Xception models. This standardization ensures consistency across the dataset. Next, pixel values are normalized to the range [-1, 1], which helps in facilitating faster convergence during the training process by stabilizing the learning rate and improving numerical stability. To further enhance the robustness of the model against overfitting and to ensure generalization across different imaging conditions, several image augmentation techniques were implemented. Specifically:

- Random rotations up to 20 degrees to simulate variations in sample positioning.
- Width and height shifts up to 10% to mimic variations in the scanning field.
- Horizontal flipping to represent mirror variations in image orientations.
- Zoom augmentation up to 10% to simulate variations in the zoom level of imaging devices.

These augmentations help in making the model robust to variations that might be encountered in real-world diagnostic settings. Table 3 summarizes the image augmentation techniques used to enhance the training dataset:

Table 3 Summary of image augmentation techniques

Augmentation Technique	Description
Random Rotations	Up to 20 degrees to simulate sample tilting
Width and Height Shifts	Up to 10% shift to mimic scanning variations
Horizontal Flipping	To reflect horizontal asymmetry in images
Zoom Augmentation	Up to 10% to simulate closer or farther views

Equation 1 provides the mathematical representation of resizing an image I to width w and height h, ensuring uniform input size for the model and Eq. 2 describes the normalization process where pixel values (P) in the image are scaled to the range [-1, 1].

$$\text{Resized Image} = \text{resize}(I, (w, h)) \quad (1)$$

$$P' = \frac{P - \min(P)}{\max(P) - \min(P)} \times 2 - 1 \quad (2)$$

Finally, the dataset is divided into training (80%), validation (10%), and test (10%) sets. This split ensures that the model is trained on a substantial portion of the data, validated on a separate subset to tune hyperparameters and monitor performance, and tested on unseen data to evaluate its generalization ability. This comprehensive preprocessing pipeline is essential for building a robust and effective image classification model. Table 4 details about how the dataset was divided into training, validation, and testing sets. Equation 3 splits dataset (D) into training and testing sets based on ratio (r).

$$\text{Train Set, Test Set} = \text{split}(D, r) \quad (3)$$

Model architecture

The model architecture described leverages the strengths of two powerful convolutional neural networks (CNNs), MobileNet and Xception, which are widely used in image classification tasks. By combining these models, the approach aims to enhance the feature extraction capabilities, thus improving classification performance for histopathological image analysis.

Table 4 Data splitting

Set	Percentage of dataset	Number of images
Training	80%	20,000
Validation	10%	2,500
Testing	10%	2,500
Total	**100%**	**25,000**

Depthwise separable convolutions are a refined type of convolution that significantly reduces the computational cost and number of parameters compared to traditional convolutions. This convolution technique decomposes the standard convolution operation into two layers: depthwise and pointwise convolutions. In depthwise convolution, a single filter is applied per input channel, mapping each input channel to an output channel independently, thereby not mixing inputs across channels. This is followed by pointwise convolution, which uses 1×1 convolutions to combine the outputs of the depthwise convolution across channels. This method enhances the model's efficiency by reducing the computational load and the model size while maintaining effective feature extraction. Equation 4 explains the mathematical operation behind depthwise separable convolutions used in MobileNet and Xception, highlighting their computational efficiency.

$$Y_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot X_{k+i,l+j,m} \quad (4)$$

MobileNet leverages depthwise separable convolutions to offer an architecture optimized for mobile and embedded applications. By simplifying the convolution process into depthwise and pointwise steps, MobileNet dramatically reduces computational requirements, making it suitable for devices with limited processing power. It introduces tunable parameters such as width multiplier and resolution multiplier, which allow for flexible architecture scaling to balance between speed and accuracy. MobileNet's efficiency makes it particularly useful for real-time applications on mobile devices, including face recognition and augmented reality. Xception, or "Extreme Inception," advances the design of the Inception models by employing depthwise separable convolutions across its architecture, replacing traditional convolutions. This adjustment not only reduces the model's parameter count but also enhances its capability to process high-resolution images effectively. Xception's architecture facilitates the independent learning of spatial features and channel-wise features by separating the convolution operations, which improves the model's ability to discern more complex patterns in the data. Due to its efficient handling of model parameters and depthwise learning approach, Xception is well-suited for tasks that involve large-scale learning from extensive, high-dimensional data sets, such as detailed image classification tasks.

Both models are initialized with weights from training on the ImageNet dataset, a large dataset featuring over a million images across 1000 categories [18]. Using pre-trained weights allows the model to start with a robust understanding of visual features that are

applicable across a wide range of image classification tasks.

In the initial training phase, the convolutional bases of both MobileNet and Xception are frozen. This means that the weights in these layers do not change during the first few training epochs. Freezing is critical because it prevents the well-learned features from being distorted while the latter parts of the network are being trained. This is especially beneficial when the new dataset (histopathological images in this case) has relatively fewer examples compared to the original ImageNet dataset. Equation 5 is the modified convolution equation incorporating strides (*s*) to reduce the dimensionality of the feature maps.

$$F_{xy} = \sum_{a=1}^{A} \sum_{b=1}^{B} I_{sx+a,sy+b} \cdot K_{a,b} \tag{5}$$

After feature extraction through both the MobileNet and Xception networks, the outputs are concatenated. MobileNet captures more granular details due to its lightweight architecture, whereas Xception, leveraging a deeper and more complex structure, might extract higher-level features. Concatenating these features ensures that the final feature set is both comprehensive and robust, capturing a broad spectrum of image characteristics at different levels of abstraction. Equation 6 mathematically represents the concatenation of features from MobileNet and Xception.

$$C = \text{Concat}(A, B) \tag{6}$$

Concatenation layer involves merging the output tensors of both networks. If both feature extractors output feature maps of the same dimensions, these can be concatenated along the depth axis to form a single, unified feature map. This concatenated feature map then serves as the input to the subsequent layers of the network. Equation 7 Shows how concatenated features are linearly combined to form the input to the next layer.

$$O_i = \sum_{j \in S} W_{i,j} \cdot C_j + b_i \tag{7}$$

The classification head is the part of the model that makes the final prediction. The first step in the classification head is to flatten the 3D output of the concatenated feature maps into a 1D vector. This transformation prepares the data for processing in the fully connected dense layers. Following flattening, the feature vector is passed through a series of dense layers. Each dense layer consists of a set number of neurons, each with ReLU (Rectified Linear Unit) activation. ReLU is chosen for its ability to introduce non-linearity into the network, helping to learn complex

patterns in the data [19]. The number of neurons and layers can be tuned based on the complexity of the task and the computational resources available. Equation 8 Clarifies the non-linear activation function used in the neural networks, essential for learning complex patterns.

$$f(x) = \max(0, x) \tag{8}$$

In the architecture of our proposed model, the flatten layer plays a pivotal role by transforming the complex, multi-dimensional feature maps produced by the convolutional layers into a flat, one-dimensional array. This transformation is essential as it bridges the convolutional base of the network, which is adept at spatial feature extraction, with the dense layers that perform classification. The flatten layer ensures that the spatial relationships within the feature maps are linearized, allowing the dense layers to effectively learn from the entirety of the extracted features for accurate cancer classification.

Following the concatenation of outputs from MobileNet and Xception, our model includes several dense layers designed to effectively harness and interpret the rich feature sets provided by these two powerful networks. The sequence begins with a Flatten layer, transitioning the multidimensional feature maps into a one-dimensional array, which is then processed through three dense layers: a first dense layer with 1024 neurons utilizing ReLU activation to facilitate the learning of complex, high-level features from the concatenated inputs, a subsequent second dense layer with 512 neurons also equipped with ReLU activation to further refine these features, and a third dense layer with 256 neurons continuing the pattern of ReLU activation to prepare the refined features for final classification. The culmination of this dense layer sequence is an output layer that employs softmax activation to produce a probabilistic distribution across the defined classes, enabling precise cancer classification based on histopathological images. This structured approach in our dense layers ensures robust learning and contributes significantly to the high accuracy and reliability of our model in medical diagnostics. Equation 9 describes the output layer's activation function, which normalizes the output to a probability distribution over predicted output class.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \text{ for } j = 1, \dots, K \tag{9}$$

Algorithm I expresses the steps included in the proposed model.

Algorithm I: Proposed model Algorithm

Input: Image Data(A directory containing histopathological images categorized by cancer types)

Output: Training/validation loss and accuracy graphs, Grad-CAM visualizations, and prediction results.

Steps:

1. **Load and Pre-process Data**:
   - Traverse directories to gather image file paths and labels.
   - Balance the dataset by sampling.
   - Split the dataset into training, validation, and testing sets.
   - Normalize and augment images for training robustness.

2. **Build and Setup Model**:
   - Utilize pre-trained MobileNet and Xception models for feature extraction.
   - Combine features from both models.
   - Append dense layers to form the final output classification layers.
   - Compile the model with loss functions and optimizers appropriate for classification tasks.

3. **Train the Model**:
   - Employ image generators for efficient data handling during training.
   - Use callbacks for early stopping and dynamic learning rate adjustments based on performance metrics.

4. **Evaluate and Visualize**:
   - Assess model performance using the test dataset with metrics such as accuracy and AUC.
   - Generate confusion matrices and classification reports for detailed performance analysis.
   - Produce visualizations like training curves and Grad-CAM heatmaps for interpretative insights.

5. **Save and Manage Outputs**:
   - Store the trained model and weights.
   - Save generated plots and visualizations for further analysis or reporting.

The architecture is designed for deep feature extraction followed by classification, leveraging the strengths of both MobileNet and Xception models for a potentially robust performance in image classification tasks.

Grad-Cam implementation

Gradient-weighted Class Activation Mapping (Grad-CAM) is a technique designed to enhance the interpretability of convolutional neural networks (CNNs), especially those applied in vision-related tasks. Grad-CAM aids in visualizing which regions of an input image are significant for the model's predictions, making it particularly valuable in fields like medical image analysis where understanding the model's decision-making process is crucial. The process involves several steps: a forward pass where the image is processed through the CNN resulting in feature maps; the selection of a target convolutional layer to focus on, which provides a spatial map highlighting crucial image areas for the class prediction; the computation of gradients of the class score with respect to these feature maps, followed by a global-average-pooling to derive importance weights for each channel; and finally, a weighted combination of these maps to produce a coarse heatmap. A ReLU function is then

applied to this heatmap to focus only on the features that positively influence the class, enhancing the interpretative utility of the CNN. Figure 4 showcases the images after applying Grad-CAM.

This architecture, combining two powerful pre-trained models and a robust classification head, is designed to effectively classify complex medical images, providing both high accuracy and efficient computation.

Training procedure

Training a deep learning model effectively requires careful consideration of various components, including the choice of optimizer, loss function, learning rate adjustments, and techniques to prevent overfitting [20, 21]. Our model uses the Adamax optimizer, known for handling sparse gradients effectively. The initial learning rate is set at 0.001, with dynamic adjustments made using a custom Learning Rate Adjustment (LRA) callback based on training performance. We use a batch size of 32 for training to balance computational efficiency and performance, with test batch size adjusted to fit memory constraints. The loss function is categorical crossentropy, ideal for our multi-class classification task. An early stopping mechanism monitors validation loss with a patience of 10 epochs to prevent overfitting. These settings ensure robust performance across different data distributions, facilitating efficient model convergence and high accuracy, which is crucial for medical image analysis. Equation 10 details the update rule for Adamax, emphasizing its role in adjusting the model weights.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \widehat{m}_t \tag{10}$$

Categorical crossentropy is selected as the loss function, which is appropriate for multi-class classification tasks in Eq. 11:

$$\mathcal{L} = -\sum_{i=1}^{C} t_i \log(p_i) \tag{11}$$

This loss function penalizes incorrect classifications more heavily, ensuring that the model learns to output probabilities close to the actual class labels.

To enhance the training process, a custom Learning Rate Adjustment (LRA) callback has been developed. This callback utilizes a dynamic mechanism to adjust the learning rate based on ongoing training performance, as outlined by the Eq. 12:

$$\eta_{\text{new}} = \eta \cdot \text{decay\_factor} \tag{12}$$

This method ensures that the learning rate is fine-tuned in response to the model's progress, promoting more effective learning and convergence.

The LRA callback actively monitors both training accuracy and validation loss during the training phase. If the training accuracy fails to meet a set threshold or if the validation loss does not show improvement over a defined number of epochs specified by the patience parameter the learning rate is adjusted downward. This patience mechanism is crucial as it prevents premature adjustments of the learning rate, allowing sufficient time for the model to explore the solution space thoroughly and avoid local minima.

When the conditions for adjustment are met, the learning rate is decreased by a predetermined factor, such as 0.5. This reduction helps in subtly fine-tuning the model's weights, which is essential for achieving optimal performance without causing significant disruptions that could negatively impact the training dynamics.

Additionally, the callback includes an optional 'dwell' parameter, which offers a safety net by reverting the model to the best weights observed before the plateau or decline in performance.

Early stopping is employed to prevent overfitting and ensure the model generalizes well to unseen data. Early stopping monitors the validation loss. If the validation loss does not improve for a set number of consecutive epochs (patience), training is halted. Equation 13
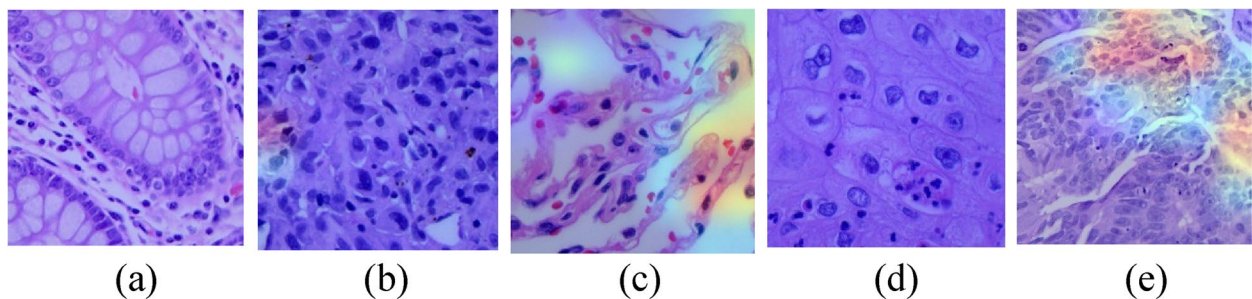

Fig. 4 Grad-CAM images

provides the criterion for stopping the training process early to prevent overfitting.

$$\text{If } \Delta\text{loss} < \text{threshold for } n \text{ epochs, stop training} \tag{13}$$

The choice of the number of epochs and batch size is critical to ensure efficient training without overfitting. The model is trained for a maximum of 10 epochs. This is a preliminary setting to evaluate the model's convergence. A batch size of 32 is chosen for training. This batch size provides a balance between computational efficiency and model performance. It allows the optimizer to update weights more frequently within each epoch, which can lead to faster convergence. The test batch size is chosen such that the entire test set can be divided into batches evenly, optimizing memory usage and processing time [22].

The decision to train our model for up to 10 epochs was strategically made based on the model's performance during preliminary trials and the incorporation of an early stopping mechanism. This mechanism is designed to halt training once the validation loss ceases to decrease over a series of epochs, effectively preventing overtraining and ensuring model robustness. Our experimental results demonstrated that further increases in the number of training epochs did not substantially enhance model performance, indicating that the optimal learning capacity was achieved within the first 10 epochs. Additionally, this approach aligns with our goal to maintain computational efficiency and reduce the environmental impact of extensive training sessions, without sacrificing the model's performance and accuracy in diagnosing lung and colon cancers.

Statistical methods
Evaluating the performance of a classification model, especially in medical imaging, requires a comprehensive set of metrics to ensure robustness, reliability, and generalizability. This study employs several evaluation metrics to assess the performance of the model in classifying histopathological images. Each metric provides unique insights into the model's strengths and weaknesses.

Accuracy is the proportion of correctly classified instances among the total instances evaluated. It is one of the most straightforward metrics for evaluating classification models. Accuracy gives a general sense of how well the model performs across all classes and is solved using the formula in the Eq. 14. However, it may not be sufficient alone, especially in cases of class imbalance where the number of instances per class varies significantly.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{14}$$

Precision, also known as positive predictive value, measures the accuracy of the positive predictions made by the classifier. High precision indicates that the model has a low false positive rate. In medical imaging, this is crucial to minimize the misdiagnosis of non-cancerous images as cancerous. Equation 15 showcases formula to calculate precision.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

Recall, also known as sensitivity or true positive rate, measures the ability of the classifier to identify all relevant instances (true positives) and is calculated using the formula in Eq. 16. High recall indicates that the model successfully identifies most of the positive instances. In medical applications, high recall is essential to ensure that most cancerous images are correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both. It is particularly useful when the class distribution is imbalanced. The F1-score ranges from 0 to 1, with 1 being the best possible score. It offers a more comprehensive evaluation than accuracy alone by considering both false positives and false negatives. F1 score can be calculated by using the formula in Eq. 17.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is a performance measurement for classification problems at various threshold settings and can be calculated using formula in Eq. 18. The ROC curve is a plot of the true positive rate (recall) against the false positive rate. The AUC value ranges from 0 to 1. A higher AUC indicates better model performance. An AUC of 0.5 suggests no discriminative ability, whereas an AUC close to 1 indicates excellent model performance. The AUC-ROC curve provides a visual representation of the trade-off between true positive and false positive rates, helping to assess the model's discriminatory power.

$$AUC = \int_0^1 TPR(t) \, dFPR(t) \tag{18}$$

A confusion matrix is a table used to describe the performance of a classification model. It shows the number

Vanitha *et al. BMC Medical Informatics and Decision Making* (2024) 24:222

Page 12 of 18

of true positive, true negative, false positive, and false negative predictions. The confusion matrix provides a detailed breakdown of the model's performance across all classes, highlighting potential biases towards particular classes. It allows the identification of specific classes where the model may struggle, enabling targeted improvements.

The use of multiple evaluation metrics ensures a comprehensive assessment of the model's performance. Accuracy provides a general overview, while precision and recall give insights into the handling of false positives and false negatives. The F1-score offers a balanced measure, the AUC-ROC curve evaluates discriminative ability, and the confusion matrix provides a detailed breakdown of performance across all classes. This multifaceted evaluation approach is crucial in medical imaging applications, where the implications of misclassification are significant.

This methodology section provides a detailed description of the processes involved in training a deep learning model for the classification of histopathological images. By employing advanced CNN architectures and a robust training procedure, the study aims to achieve high accuracy and reliability, which are crucial for medical image analysis.

Experiment and analysis

The experimental evaluation was structured to assess the effectiveness of the integrated MobileNet and Xception model in classifying histopathological images of lung and colon cancer. The LC25000 dataset was utilized, involving an extensive series of experiments to validate the model's performance under various conditions.

Training details

The model was trained for up to 10 epochs, with training potentially stopping earlier based on the early stopping criteria, which monitored validation loss. The early stopping mechanism ensured that the model did not continue to train once performance ceased to improve, thus preventing overfitting. The batch size for training was set to 32, which is a common choice that balances computational efficiency and model performance. For the test set, an appropriate batch size was calculated to ensure that the entire test dataset was processed without leaving out any samples, optimizing memory usage and processing time. Figure 5 depicts the graphical representation of training and validation loss and accuracy.

The optimizer used was Adamax, a variant of the Adam optimizer known for its effectiveness in handling sparse gradients, which are common in high-dimensional data such as images. The initial learning rate was set to 0.002 and was dynamically adjusted according to a custom Learning Rate Adjustment (LRA) callback. This callback monitored both training accuracy and validation loss, making necessary adjustments to the learning rate to optimize the training process.

The loss function employed was categorical crossentropy, which is appropriate for multi-class classification tasks. This function computes the loss by comparing the predicted probabilities to the actual class labels, penalizing incorrect predictions more heavily. This ensures
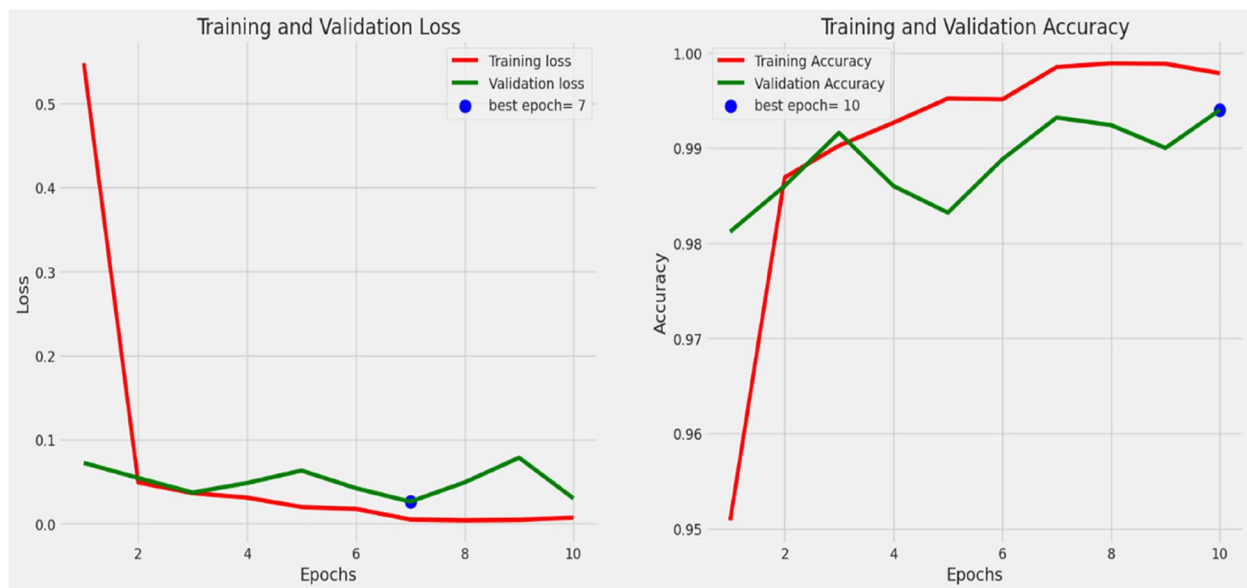

Fig. 5 Graphical representation of training and validation

Vanitha *et al. BMC Medical Informatics and Decision Making* (2024) 24:222

Page 13 of 18

that the model learns to output probabilities close to the actual class labels, thus improving classification accuracy. The Fig. 6 showcases the learning curve of the model.

The model achieved an overall accuracy of 99.44% on the test set, demonstrating exceptional performance across various classes. Precision ranged from 0.9559 for Lung Squamous Cell Carcinoma (SCC) to 1.0000 for both Colon Adenocarcinoma (ACA) and benign Lung tissue (Lung N). Recall values were also high, with the lowest being 0.9508 for Lung Adenocarcinoma (ACA), indicating the model's strong ability to correctly identify true positive cases. The F1-scores were consistently high across all classes, reflecting a balanced performance between precision and recall. These results underscore the model's robustness and effectiveness in accurately classifying histopathological images of lung and colon cancers. Figure 7 is the heatmap of classification report of the proposed model.

To gain a deeper understanding of the model's performance, a detailed error analysis was conducted, scrutinizing misclassifications to identify patterns and potential causes. Key findings from this analysis include inter-class confusion, where most misclassifications occurred between visually similar classes such as Lung Squamous Cell Carcinoma (SCC) and Lung Adenocarcinoma (ACA), suggesting a need for the model to learn more discriminative features specific to each subtype. Additionally, errors were more frequent in images with lower resolution or poor staining quality, indicating the importance of consistent image quality to enhance model reliability. Despite the dataset being balanced, intrinsic variations in texture and pattern complexity within classes might have contributed to biased error rates towards more complex patterns, commonly seen in benign versus malignant distinctions. Furthermore, Grad-CAM visualizations of misclassified cases revealed that incorrect predictions often focused on non-relevant regions of the images, indicating potential improvements needed in the model's region of interest detection capabilities. Figure 8 is the image showcasing truly classified instances whereas Fig. 9 representing misclassified instances.

The AUC-ROC curves for each class demonstrated excellent discriminative performance, with values close to 1. This indicates that the model has a high true positive rate while maintaining a low false positive rate across all classes. Figure 10 is the graph representing ROC curve of the model.

The confusion matrix provided a detailed breakdown of performance, highlighting few confusions between Lung SCC and Lung ACA, underscoring areas for potential improvement in future model iterations. Figure 11 depicts the confusion matrix of the proposed model.

To contextualize the performance of the integrated MobileNet and Xception model, compared it with existing methodologies in histopathological image classification. Table 5 shows comparison of proposed model with existing model and their techniques.

The integrated MobileNet and Xception model demonstrates a high degree of accuracy and robustness in classifying histopathological images of lung and colon cancers. The architecture effectively captures and utilizes complex features from the images, significantly outperforming traditional models. Despite the advancements, the proposed model, like many deep learning approaches, faces potential limitations that could impact its broad applicability in clinical settings. A significant limitation is the model's reliance on the specific dataset used for training.
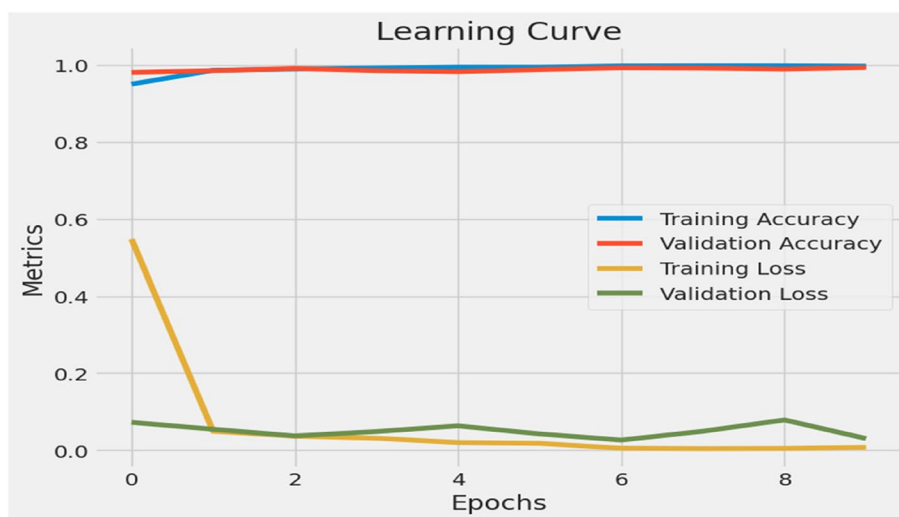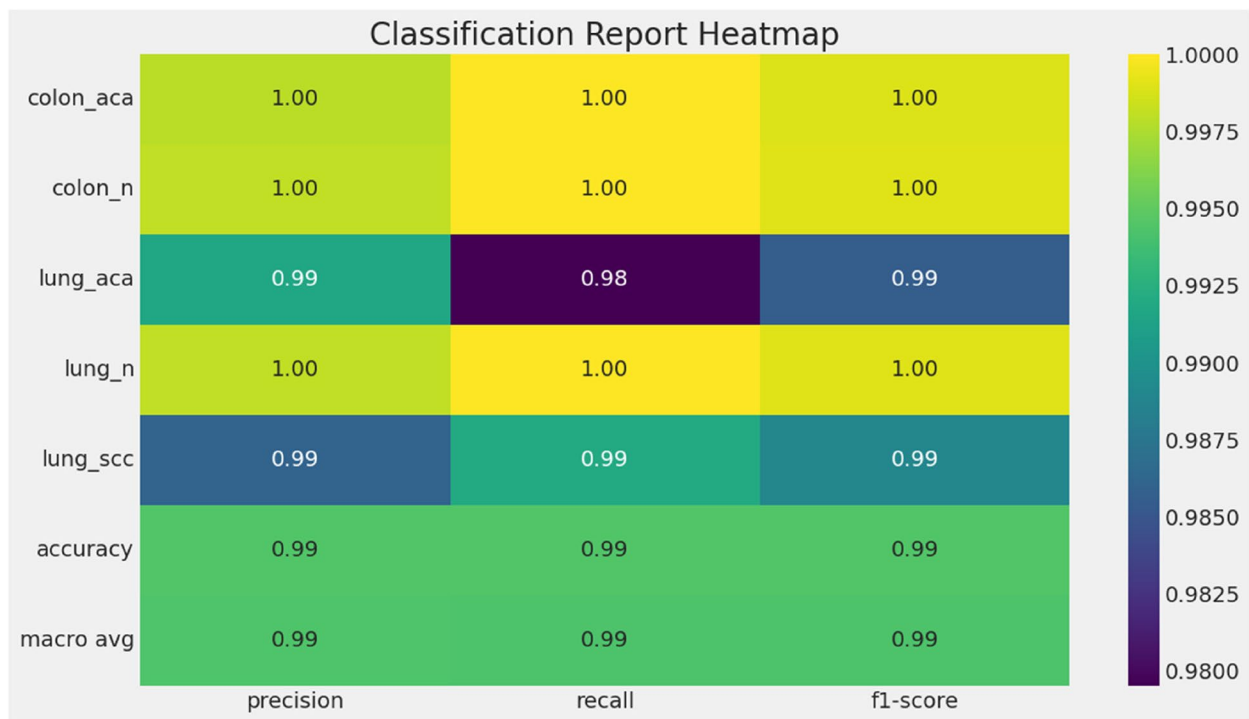

Fig. 6 Learning curve

Fig. 7 Classification report

While the LC25000 dataset provides a robust platform for developing and testing the model, its unique characteristics, such as image acquisition techniques, staining protocols, and the distribution of cancer types, may not be representative of broader clinical environments. This could potentially limit the model's performance when applied to data from different hospitals or labs with varying imaging standards.

Furthermore, there is a risk of inherent biases within the dataset, such as overrepresentation or underrepresentation of certain demographic groups or cancer stages, which could skew the model's predictions. Such biases might result in reduced accuracy when the model is deployed in diverse real-world scenarios, where the distribution of cases might differ significantly from the training dataset.

Conclusion

The integration of Xception and MobileNet architectures, along with the application of Gradient-weighted Class Activation Mapping (Grad-CAM), has significantly advanced the classification of histopathological images for lung and colon cancers. This study has demonstrated the capability of these combined architectures to enhance feature extraction, improve generalizability, and reduce overfitting, achieving an accuracy rate of 99.44% on a
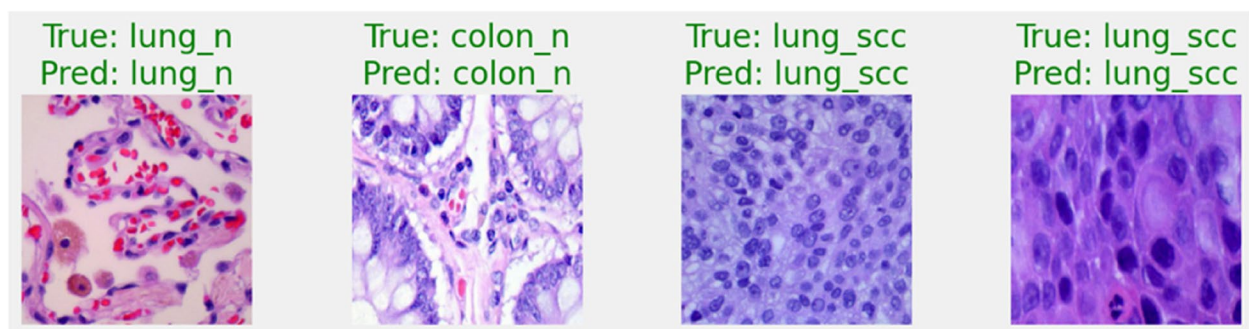

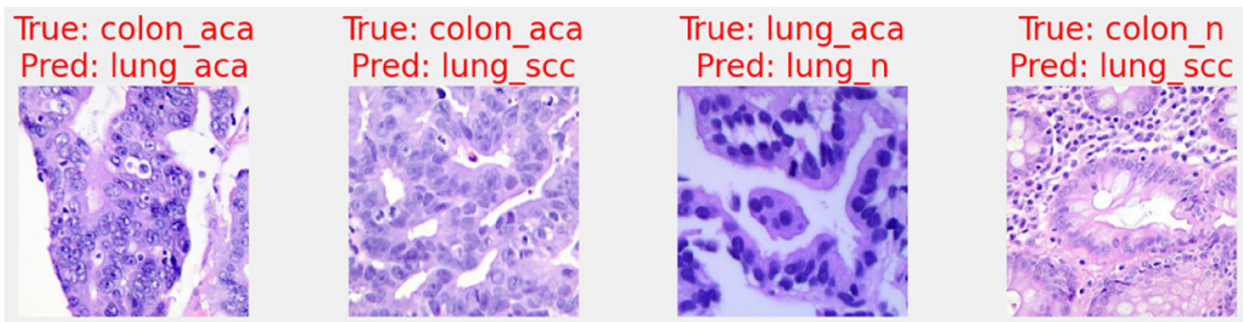**Fig. 8** Truly classified instances

Fig. 9 Misclassified instances

balanced test set. The precision and recall metrics indicate the model's exceptional performance in identifying specific cancerous and non-cancerous tissue types, with several categories achieving perfect scores.

The utilization of Grad-CAM has further augmented the interpretability of the model, offering a transformative approach to understanding deep learning decisions in medical diagnostics. By generating heatmaps that highlight influential regions used by the CNN for making predictions, Grad-CAM enables clinicians to visually verify these automated insights. This visualization facilitates a deeper understanding and trust in the model's functionality, crucial for integrating AI tools into clinical workflows. Moreover, such detailed visual explanations help in educational settings, where medical professionals can observe how advanced models discern subtle nuances in histopathological images

that may be overlooked in standard examinations. By enhancing the transparency of the model's decision-making process through Grad-CAM visualizations, it not only bolsters diagnostic confidence but also assists in advancing the discussion regarding AI's role in augmenting diagnostic accuracy and reliability in clinical settings. This step forward is pivotal for the adoption of AI in routine clinical practices, ensuring that AI-supported diagnostics are both interpretable and verifiable by expert clinicians.

The findings from this research suggest that the sophisticated blending of neural network architectures offers a promising pathway to refining diagnostic processes in pathology. The architecture of the combined MobileNet and Xception models, enhanced with Grad-CAM for interpretability, provides a robust foundation that can potentially be adapted for a wide range of diagnostic
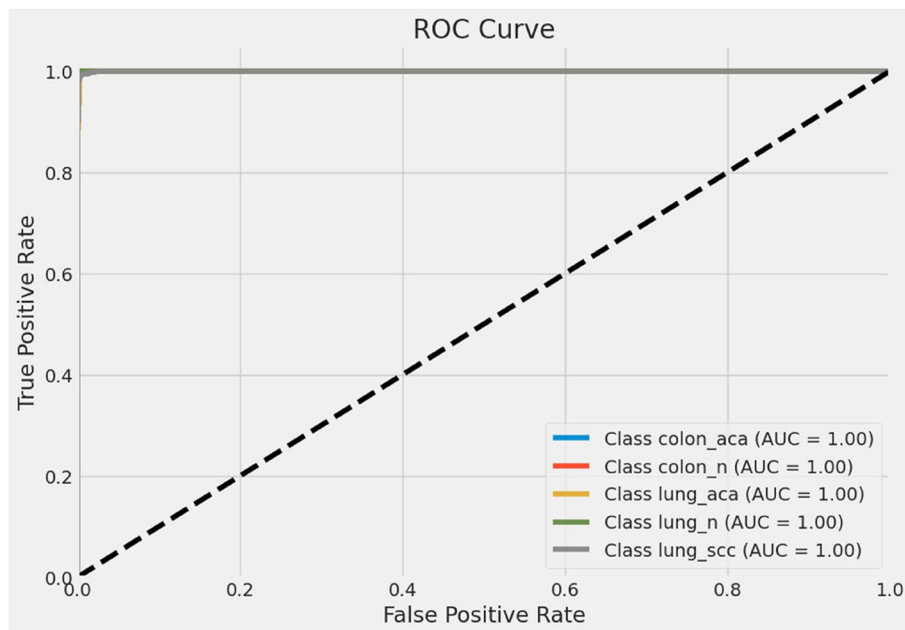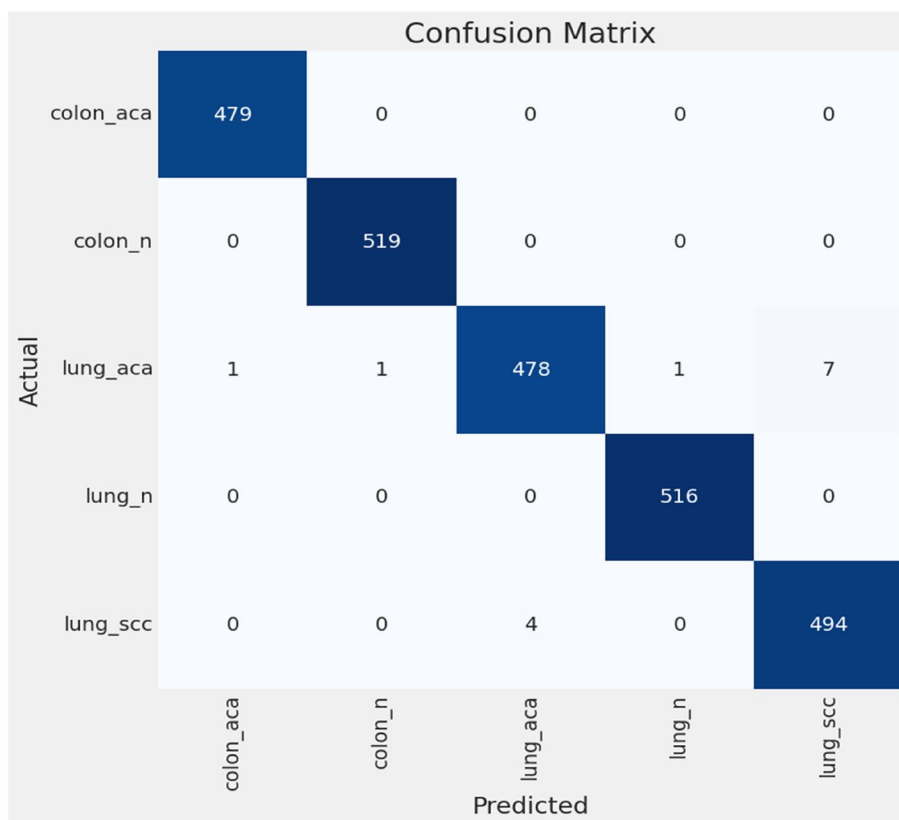

Fig. 10 ROC curve

Fig. 11 Confusion matrix

tasks beyond lung and colon cancers. This adaptability is crucial for extending the model's application to other types of cancer, such as breast, skin, or prostate cancers, where histopathological analysis plays a pivotal role in diagnosis.

Moreover, the modular nature of the proposed model framework allows for flexibility in tuning and retraining for different medical imaging tasks, such as MRI analysis, CT scans, or ultrasound image interpretation. By retraining the model on specific datasets corresponding to these tasks, or by integrating other specialized neural network layers tailored for different imaging modalities, the model can be made suitable for a broad spectrum of diagnostic applications.

Table 5 Comparative study

Author	Techniques	Accuracy
Garg, Satvik, and Somya Garg (2020) [22]	Pre-trained CNN models	96 to 100%
Masud, Mehedi, et al. (2021) [23]	Modern Deep Learning (DL) and Digital Image Processing (DIP) techniques	96.33%
Singh, Onkar, and Koushlendra Kumar Singh. (2023) [24]	an ensemble classifier utilizing three distinct approaches random forest (RF), support vector machine (SVM), and logistic regression (LR) model	99.00%
Al-Mamun Provath, Md, Kaushik Deb, and Kang-Hyun Jo. (2023) [25]	Deep Learning Method	97%
Chilyatun Nisa, Nanik Suciati, and Anny Yuniarti [26]	Hybrid convolutional neural networks	98.88%
Hage Chehade, Aya, et al. (2022) [27]	Feature Engineering	99%
Mehmood, Shahid, et al. (2022) [28]	Pretrained Neural Network (AlexNet)	98.4%
Alamin Talukder, Md, et al. (2022) [29]	Ensemble Learning	99.3%
Wahid, Radical Rakhman, et al. (2023) [30]	Convolutional Neural Network	93.02%
Proposed model	Deep learning with combination of MobileNet and Xception model	99.44%

Vanitha *et al. BMC Medical Informatics and Decision Making*   (2024) 24:222

Page 17 of 18

While the current model demonstrates high accuracy and reliability in lung and colon cancer classification, its potential for generalization to other medical imaging tasks holds promise. With further development and validation, this approach could significantly contribute to the advancement of automated, precise, and reliable diagnostic processes across various domains of medical imaging. Future research could explore the application of this combined model framework to other types of cancer and more complex diagnostic tasks, potentially extending its utility to broader medical imaging applications. The pursuit of these advancements could pave the way for more automated, accurate, and reliable diagnostic processes, ultimately enhancing patient outcomes in oncology.

Authors' contributions
V.K took care of the review of literature and methodology. M.T.R has done the formal analysis, data collection and investigation. S.S.S has done the initial drafting and statistical analysis. S.G has supervised the overall project. All the authors of the article have read and approved the final article.

Availability of data and materials
The datasets used for the findings are publicly available at https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images.

Declarations

Ethics approval and consent to participate
Not Applicable as the research is done on the publicly available dataset.

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Computer Science and Engineering, Faculty of Engineering, Karpagam Academy of Higher Education (Deemed to Be University), Coimbatore, India. [2]Department of Computer Science and Engineering, JAIN (Deemed-to-Be University), Bengaluru 562112, India. [3]Department of Computer Science and Engineering, School of Engineering, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India. [4]Adama Science and Technology University, Adama 302120, Ethiopia.

References
1. Kumar A, Vishwakarma A, Bajaj V. Crccn-net: Automated framework for classification of colorectal tissue using histopathological images. Biomed Signal Process Control. 2023;79:104172.
2. Musthafa MM, Manimozhi I, Mahesh TR, et al. Optimizing double-layered convolutional neural networks for efficient lung cancer classification through hyperparameter optimization and advanced image pre-processing techniques. BMC Med Inform Decis Mak. 2024;24:142. https://doi.org/10.1186/s12911-024-02553-9.
3. Subashchandrabose U, John R, Anbazhagu UV, Venkatesan VK, Thyluru Ramakrishna M. Ensemble federated learning approach for diagnostics of multi-order lung cancer. Diagnostics. 2023;13:3053. https://doi.org/10.3390/diagnostics13193053.
4. A Sampling of Highlights from the Literature. Cancer Immunol Res. 2021;9(10):1109. https://doi.org/10.1158/2326-6066.cir-9-10-wwr.
5. Ibrahim DM, Elshennawy NM, Sarhan AM. Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. Comput Biol Med. 2021;132:104348.
6. Chouat I, Echtioui A, Khemakhem R, Zouch W, Ghorbel M, Hamida AB. Lung Disease Detection in Chest X-ray Images Using Transfer Learning. 2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). 2022. https://doi.org/10.1109/atsip55956.2022.9805892.
7. Malik H, Anees T. BDCNet: Multi-classification convolutional neural network model for classification of COVID-19, pneumonia, and lung cancer from chest radiographs. Multimedia Syst. 2022;28(3):815–29.
8. Aitazaz T, et al. Transfer learning for histopathology images: an empirical study. Neural Comp Appl. 2023;35(11):7963–74.
9. Kaya Y, Gürsoy E. A MobileNet-based CNN model with a novel fine-tuning mechanism for COVID-19 infection detection. Soft Comput. 2023;27(9):5521–35.
10. Karaddi SH, Sharma LD. Automated multi-class classification of lung diseases from CXR-images using pre-trained convolutional neural networks. Expert Syst Appl. 2023;211:118650.
11. Sharma N, Garg N, Garg B, Ranjan S, Goyal P. Image Classification of Various Lung Diseases Using Transfer Learning and Machine Learning Techniques. SSRN Electron J. 2023. https://doi.org/10.2139/ssrn.4628662.
12. Sethy, Prabira Kumar, et al. Lung cancer histopathological image classification using wavelets and AlexNet. J X-Ray Sci Technol. 2023;31(1):211–221.
13. Chhagan Patil N, Nitin Jagannath Patil. A hybrid VGG 19 and capsule network based deep learning model for lung cancer diagnosis using CT scan images. Int J Comp Digital Syst. 2024;16(1):1–12.
14. Seth A, Kaushik VD. Automatic lung and colon cancer detection using enhanced cascade convolution neural network. Multimed Tools Appl. 2024. https://doi.org/10.1007/s11042-024-18548-7.
15. Chutia U, Tewari AS, Singh JP, Raj VK. Classification of Lung Diseases Using an Attention-Based Modified DenseNet Model. J Imaging Inform Med. 2024;37(4):1625–41. https://doi.org/10.1007/s10278-024-01005-0.
16. Chiet CC, et al. A lung cancer detection with pre-trained CNN models. J Inform Web Eng. 2024;3(1):41–54.
17. Wang R, et al. Development of a novel combined nomogram model integrating deep learning-pathomics, radiomics and immunoscore to predict postoperative outcome of colorectal cancer lung metastasis patients. J Hematol Oncol. 2022;15(1):11.
18. Hatuwal BK, Thapa HC. Lung cancer detection using convolutional neural network on histopathological images. Int J Comput Trends Technol. 2020;68(10):21–4.
19. Wang K-S, et al. Accurate diagnosis of colorectal cancer based on histopathology images using artificial intelligence. BMC Med. 2021;19:1–12.
20. Thakur A, Mahesh TR, Khan SB, Palaiahnakote S, Vinoth Kumar V, Almusharraf A, Mashat A. An adaptive Xception model for classification of brain tumors. Int J Pattern Recognit Artif Intell. 2024. https://doi.org/10.1142/s0218001424560056.
21. Shaban M, et al. Context-aware convolutional neural network for grading of colorectal cancer histology images. IEEE Transact Med Imag. 2020;39(7):2395–405.
22. Garg S, Garg S. Prediction of lung and colon cancer through analysis of histopathological images by utilizing Pre-trained CNN models with visualization of class activation and saliency maps. 2020 3rd Artificial Intelligence and Cloud Computing Conference. 2020. https://doi.org/10.1145/3442536.3442543.
23. Masud M, et al. A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. Sensors. 2021;21(3):748.
24. Singh O, Singh KK. An approach to classify lung and colon cancer of histopathology images using deep feature extraction and an ensemble method. Int J Inform Technol. 2023;15(8):4149–60.

25. Al-Mamun Provath M, Deb K, Jo KH. Classification of lung and colon cancer using deep learning method. In: International Workshop on Frontiers of Computer Vision. Singapore: Springer Nature Singapore; 2023.
26. Mobilia SJ. Classification of Hyperspectral Colon Cancer Images Using Convolutional Neural Networks. https://doi.org/10.31979/etd.7hhn-2p3k.
27. Chehade A, et al. Lung and colon cancer classification using medical imaging: a feature engineering approach. Phys Eng Sci Med. 2022;45(3):729–46.
28. Mehmood S, et al. Malignancy detection in lung and colon histopathology images using transfer learning with class selective image processing. IEEE Access. 2022;10:25657–68.
29. Talukder MA. Machine Learning-based Lung and Colon Cancer Detection using Deep Feature Extraction and Ensemble Learning. 2022. https://doi.org/10.31224/2731.
30. Wahid RR, Nisa' C, Amaliyah RP, Puspaningrum EY. Lung and colon cancer detection with convolutional neural networks on histopathological images. THE PROCEEDINGS OF THE 5TH INTERNATIONAL CONFERENCE ON MARITIME EDUCATION AND TRAINING (The 5th ICMET). 2021, 2023. https://doi.org/10.1063/5.0114327.

Publisher's Note