

Article

An Explainable Classification Method Based on Complex Scaling in Histopathology Images for Lung and Colon Cancer

Sudhakar Tummala ^{1,*}, Seifedine Kadry ^{2,3,4}, Ahmed Nadeem ⁵, Hafiz Tayyab Rauf ^{6,*} and Nadia Gul ⁷

¹ Department of Electronics and Communication Engineering, School of Engineering and Sciences, SRM University-AP, Amaravati 522240, Andhra Pradesh, India

² Department of Applied Data Science, Noroff University College, 4612 Kristiansand, Norway

³ Artificial Intelligence Research Center (AIRC), Ajman University, Ajman 346, United Arab Emirates

⁴ Department of Electrical and Computer Engineering, Lebanese American University, Byblos P.O. Box 36, Lebanon

⁵ Department of Pharmacology & Toxicology, College of Pharmacy, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

⁶ Centre for Smart Systems, AI and Cybersecurity, Staffordshire University, Stoke-on-Trent ST4 2DE, UK

⁷ Wah Medical College affiliated with POF Hospital, Wah Cantt 47040, Pakistan; mrsnadiagul@gmail.com

* Correspondence: sudhakar.t@srmmap.edu.in (S.T.); hafiztayyabrauf093@gmail.com (H.T.R.)

Abstract: Lung and colon cancers are among the leading causes of human mortality and morbidity. Early diagnostic work up of these diseases include radiography, ultrasound, magnetic resonance imaging, and computed tomography. Certain blood tumor markers for carcinoma lung and colon also aid in the diagnosis. Despite the lab and diagnostic imaging, histopathology remains the gold standard, which provides cell-level images of tissue under examination. To read these images, a histopathologist spends a large amount of time. Furthermore, using conventional diagnostic methods involve high-end equipment as well. This leads to limited number of patients getting final diagnosis and early treatment. In addition, there are chances of inter-observer errors. In recent years, deep learning has shown promising results in the medical field. This has helped in early diagnosis and treatment according to severity of disease. With the help of *EfficientNetV2* models that have been cross-validated and tested fivefold, we propose an automated method for detecting lung (lung adenocarcinoma, lung benign, and lung squamous cell carcinoma) and colon (colon adenocarcinoma and colon benign) cancer subtypes from LC25000 histopathology images. A state-of-the-art deep learning architecture based on the principles of compound scaling and progressive learning, *EfficientNetV2* large, medium, and small models. An accuracy of 99.97%, AUC of 99.99%, F1-score of 99.97%, balanced accuracy of 99.97%, and Matthew's correlation coefficient of 99.96% were obtained on the test set using the *EfficientNetV2-L* model for the 5-class classification of lung and colon cancers, outperforming the existing methods. Using gradCAM, we created visual saliency maps to precisely locate the vital regions in the histopathology images from the test set where the models put more attention during cancer subtype predictions. This visual saliency maps may potentially assist pathologists to design better treatment strategies. Therefore, it is possible to use the proposed pipeline in clinical settings for fully automated lung and colon cancer detection from histopathology images with explainability.

Keywords: lung cancer; colon cancer; *EfficientNetV2*; explainability; histopathology



Citation: Tummala, S.; Kadry, S.; Nadeem, A.; Rauf, H.T.; Gul, N. An Explainable Classification Method Based on Complex Scaling in Histopathology Images for Lung and Colon Cancer. *Diagnostics* **2023**, *13*, 1594. <https://doi.org/10.3390/diagnostics13091594>

Academic Editor: Malek Makki

Received: 11 March 2023

Revised: 15 April 2023

Accepted: 28 April 2023

Published: 29 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the World Health Organization, cancer is the leading cause of mortality worldwide, and by 2040, the global cancer burden is expected to be 28.4 million cases, a 47% increase from 2020 [1,2]. Lung and colorectal (both colon and rectum) cancers are the more prevalent types worldwide, after breast cancer, with incidence rates of 11.4% and 10%, respectively, in 2020 [2]. Although low, there is a chance of synchronous occurrence

between lung and colon cancers [3]. In addition, lung and colorectal cancers exhibit the top two mortality rates of 18% and 9.4%, respectively, among all cancers [2]. Therefore, a more accurate diagnosis of these cancer subtypes is necessary to explore the treatment options in the initial stages of the disease. The non-invasive methods for diagnosis include radiography and computed tomography (CT) imaging for lung cancer and flexible sigmoidoscopy, and CT colonoscopy for colon cancer. However, reliable subtyping of these cancers may not be possible using non-invasive means always, and minimally invasive procedures, such as histopathology, would be required for precise disease identification and improved quality of treatment. In addition, the manual grading of histopathological images may be tiresome to pathologists. Moreover, accurate grading of the lung and colon cancer subtypes requires trained pathologists, and manual grading could be error prone. Hence, automated image processing methods for cancer lung and colon cancer subtype screening are warranted to reduce the burden on pathologists.

Deep learning (DL) is a branch of machine learning (ML) that eliminates the need for manual feature engineering, and convolutional neural network (CNN) based DL models provide hierarchical feature maps for better representation of input images. In recent years, various state-of-the-art CNN-based DL frameworks, including the AlexNet [4], VGG Nets [5], GoogLeNet [6], Residual Nets [7], DenseNets [8], EfficientNets [9,10], and, lately, multi-head self-attention based vision transformer (ViT) [11,12] architectures were invented for various vision tasks, including classification. Although massive data would be required to train these large DL models from scratch, transfer learning (TL) helps to adapt the large pre-trained models for downstream tasks. Thus, TL reduces the need for massive data for training, which is scarce in specific fields, such as medicine. DL and TL have been performing a vital role in healthcare in building automated diagnostic systems using medical images that include histopathological images, retina images, radiographs, computed tomography images, magnetic resonance images, etc. These automated systems are primarily used for classification tasks and also assist clinical practitioners in situations of rapid data acquisition and automated quality checking [13–17]. *EfficientNetV2* is a recent DL architecture that was developed based on progressive learning with a combination of compound scaling and neural architecture search (NAS) to improve both the training speed and parameter efficiency [10], and it outperformed several existing state-of-the-art DL models, including ViT variants in image classification task on the ImageNet and other datasets.

In general, DL methods are similar to black box architectures. Therefore, it is often required to ensure that these DL models focus on the most relevant regions in the input image during target class prediction. Several methods exist in the literature to visualize most activated areas when a DL model predicts the class of a specific image to add explainability to the model. A few of these methods include class activation mapping (CAM) [18], gradCAM [19], and gradCAM++ [20]. In this study, we considered gradCAM for creating visual saliency maps for *EfficientNetV2* predictions.

Hence, the contributions of the present work are:

- i. A fully automated framework for the five-class diagnosis of most occurring lung and colon cancer subtypes is proposed using *EfficientNetV2*-large (*L*), medium (*M*), and small (*S*) models based on histopathology images.
- ii. These existing pretrained models are finetuned and tested using a large, openly available lung and colon cancer histopathology image dataset called LC25000.
- iii. Visual saliency maps are provided using the gradCAM method to understand the model decisions during testing better.

Related Work

Several works employing ML and DL techniques have been present in the literature during recent years for the classification of colon and lung cancer subtypes from histopathological images from private and public (LC25000) datasets. These works are stratified into 3-class classification of lung cancer subtypes (adenocarcinoma, squamous cell carcinoma,

and benign), 2-class classification of colon cancer subtypes (adenocarcinoma and benign), and 5-class classification of both lung and colon cancer subtypes, which are given in Table 1. In [21], a custom CNN model with heavy data augmentation from 298 microscopy images was developed and achieved an overall accuracy of 71.1% for subtyping lung cancer into adenocarcinoma, squamous cell carcinoma, and small cell carcinoma. In another recent study using LC25000 dataset [22], lung cancer subtyping is performed using a custom-made CNN, obtaining an accuracy of 97.2%. Furthermore, in [23], colon cancer subtyping was only implemented using a CNN and principal component analysis (PCA) from LC25000, and the framework has a classification accuracy of 99.8%. Few studies exist using feature extraction from the histopathology images and different ML classifiers, including random forest (RF) and XGBoost, for the lung and colon cancer subtyping and achieved accuracies above 96.3% [24,25].

A multi-input dual-stream capsule neural network was proposed [26] using LC25000 images that employed several pre-processing strategies, including color balancing, gamma correction, image sharpening and multi-scale fusion, to obtain an accuracy of 99.6%. Similarly, Ref. [27] employs histogram equalization as the pre-processing step followed by pretrained AlexNet to improve the colon cancer classification. In other recent studies, pretrained DarkNet-19 and support vector machine classifier [28], DenseNet-121, and RF classifier [29] were developed and demonstrated 99.7% and 98.6% accuracy, respectively. Integration of deep feature extraction and ensemble learning with high-performance filtering was found to be effective in a recent work [30] with an accuracy of 99.3% using LC25000 data. Lastly, a custom CNN model from the same dataset followed by several dimensionality reduction methods, such as PCA, discrete Fourier transform, and fast Walsh-Hadamard transform, was employed to obtain 99.6% accuracy for the five-class classification [31].

Although some previous studies obtained accuracies above 99.5%, they lacked explainability and incorporated extensive pre-processing steps. Therefore, the present study aimed at using compound scaling-inspired *EfficientNetV2* models for the five-class classification with added interpretability using the gradCAM method. Eventually, our framework outperformed all the existing methods based on LC25000 dataset with an overall test accuracy of 99.98%.

Table 1. Previous works on classifying lung and colon cancer subtypes using different machine learning and deep neural network methods based on LC25000 dataset and a private dataset. CNN: convolution neural network. ML: machine learning, PCA: principal component analysis, DWT: discrete wavelet transforms, SVM: support vector machine, RF: random forest, BA: balanced accuracy, AUC: area under the receiver operating characteristic curve, MCC: Matthew’s correlation coefficient, FWHT: fast Walsh-Hadamard transform.

Study	Year	Method	Dataset	Interpretability	Performance (%)
Chehade A. H. et al. [25]	2022	ML classifiers	LC25000	No	Accuracy: 99.0 F1-score: 98.80
Masud M. et al. [24]	2021	ML classifiers	LC25000	No	Accuracy: 96.33
Ali M. et al. [26]	2021	Multi-input capsule neural network	LC25000	No	Accuracy: 99.58
Togacar M. [28]	2021	DarkNet-19 and SVM	LC25000	No	Accuracy: 99.69
Mehmood S. et al. [27]	2022	Image enhancement and AlexNet	LC25000	No	Accuracy: 98.40
Teramoto A. et al. [21]	2017	Custom CNN model	Private dataset (298 microscopic images)	No	Accuracy: 71.10 (Only lung cancer)
Attallah O. et al. [31]	2022	Custom CNN + PCA, FWHT, DWT	LC25000	No	Accuracy: 99.60
Hatuwal B. K. et al. [22]	2020	Custom CNN	LC25000	No	Accuracy: 97.20 (Only lung cancer)

Table 1. Cont.

Study	Year	Method	Dataset	Interpretability	Performance (%)
Mangal S. et al. [32]	2020	Custom CNN	LC25000	No	Accuracy: 96.50
Talukder Md. A. et al. [30]	2022	Hybrid ensemble learning	LC25000	No	Accuracy: 99.30
Kumar N. et al. [29]	2022	DenseNet121 and RF	LC25000	No	Accuracy: 98.60 F1-score: 98.50
Hasan Md. I. et al. [23]	2022	Custom CNN and PCA	LC25000	No	Accuracy: 99.80 (Only colon cancer)
Present study	2023	<i>EfficientNetV2</i>	LC25000	Yes	Accuracy: 99.97 F1-score: 99.97 BA: 99.97 AUC: 99.99 MCC: 99.96

2. Methods

2.1. Dataset

For this study, we considered a publicly available dataset LC25000 [33]. Initially, 250 color images for each lung and colon cancer subtype were acquired using Leica microscope LM190 HD camera connected to an Olympus BX41 microscope, constituting 1250 images before data augmentation. Then, the 250 images for each cancer subtype were increased to 5000 by using augmentation methods, including right and left rotations and vertical and horizontal flips. Thus, after data augmentation, the dataset consists of 25,000 regular histopathology images. The original spatial resolution of the images was 1024×768 , but they were cropped to 768×768 before applying the data augmentation. Eventually, for the current study, the spatial resolution of the images was changed to 224×224 by resizing.

For a fair differentiation with existing literature, the percentage of data used in training, validation, and testing from LC25000 is considered to match with existing studies, i.e., 80% of the data was used for cross-validation and the remaining 20% for testing. The images in the dataset were labeled as follows: 0 for lung-adenocarcinoma, 1 for lung-benign, 2 for lung-squamous cell carcinoma, 3 for colon-adenocarcinoma, and 4 for colon-benign by experienced pathologists. Example histopathological images with lung and colon cancer subtypes are shown in Figure 1. Furthermore, the dataset stratified with respect to lung and colon cancer subtypes are given in Table 2 for the train, validation, and test sets.

Table 2. Number of lung and colon cancer histopathology images in the training, validation, and testing. aca: adenocarcinoma, n: benign, scc: squamous cell carcinoma.

	Lung-aca	Lung-n	Lung-scc	Colon-aca	Colon-n
Training	3600	3600	3600	3600	3600
Validation	400	400	400	400	400
Testing	1000	1000	1000	1000	1000

2.2. Physiological Mechanisms of Lung and Colon Cancers

In this subsection, we briefly described the pathophysiological mechanisms about the lung and colon cancer subtypes dealt in the present study. Lung adenocarcinoma and squamous cell carcinoma falls under the category of non-small cell lung cancers where squamous carcinoma frequently occurs as a central endobronchial lesion, and adenocarcinoma has a tendency to start in the lung periphery and invade the pleura [34]. Lung benign is non-cancerous and will not spread to the surrounding tissues. Most occurring lung benign include hamartomas that usually occurs in outer portion of lung connective tissue and bronchial adenomas that grow in the bronchi and in the ducts or mucus glands of the windpipe. Colon adenocarcinoma and benign occur in a pedunculated polyp, sessile

polyp, or stricture. Polyp is an abnormal chunk of cells that also grow inside the colon. Small polyps rarely contain cancer [35].

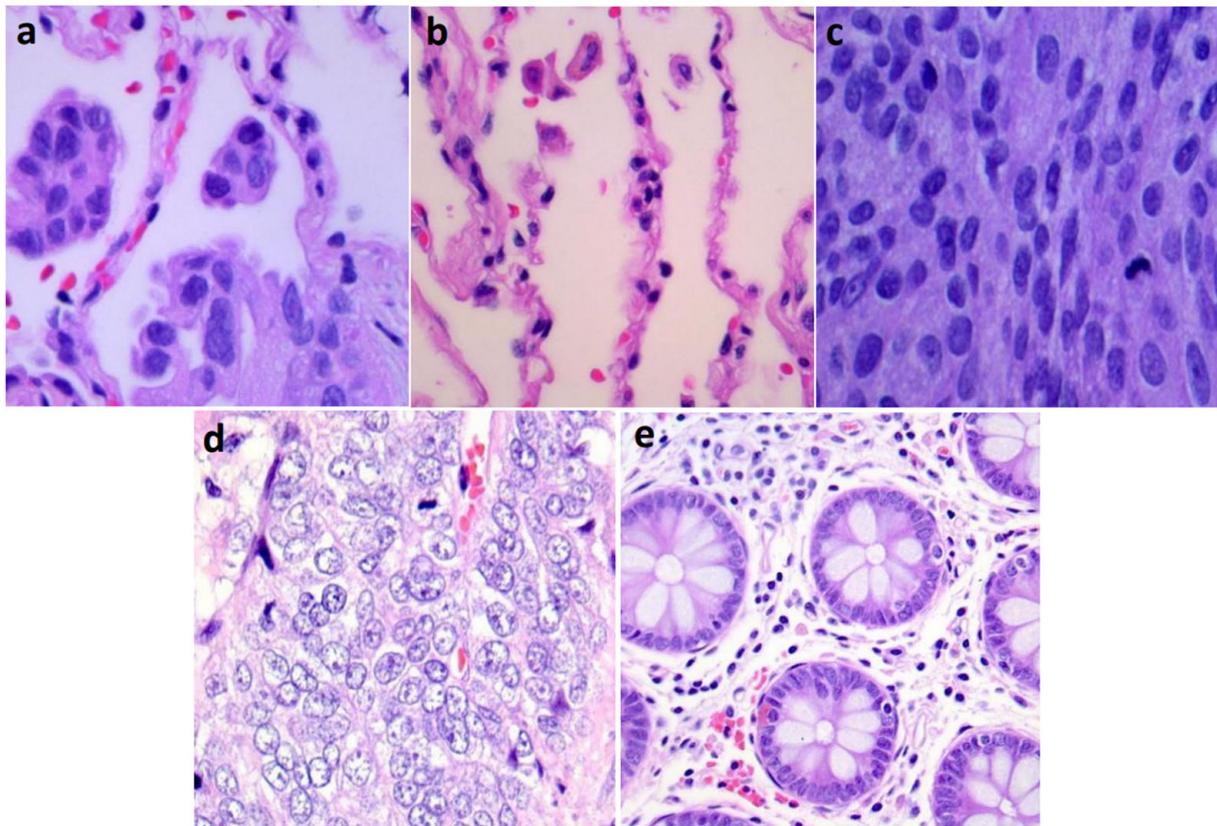


Figure 1. Sample lung and colon cancer histopathological images from LC25000 dataset. (a) lung-adenocarcinoma, (b) lung-benign, (c) lung-squamous cell carcinoma, (d) colon-adenocarcinoma, (e) colon-benign.

2.3. *EfficientNetV2* and Compound Scaling

EfficientNetV2 [10], the next version of *EfficientNetV1* [9], is a novel family of deep CNNs focusing on two significant aspects: enhancing the training speed and parameter efficiency. To accomplish this task, a combination of training-aware NAS, and compound scaling were used. The faster training was achieved by employing both MBConv and Fused-MBConv layers. Here, MBConv layers are the basic building blocks of MobileNetV2 [36] constructed from the inverted residual blocks. To obtain the Fused-MBConv layer, the first two blocks (depth-wise 3×3 convolution and the expansion 1×1 convolution block) of MBConv were replaced by a regular 3×3 convolution block, as shown in Figure 2. Afterward, a squeeze and excitation block in both MBConv and Fused-MBConv layers was used to weigh different channels adaptively. Finally, a 1×1 squeeze layer was inserted to reduce the number of channels equal to the channels present in the input of either MBConv or Fused-MBConv layers.

In the present work, we considered *EfficientNetV2-L*, *-M*, and *-S* models that employed Fused-MBConv blocks in the initial layers. The *EfficientNetV2-S* model architecture begins with a standard 3×3 convolution layer followed by three Fused-MBConv and three MBConv blocks. The eventual layers contain a 1×1 convolution, pooling, and concluded by a fully connected layer. Furthermore, the *EfficientNetV2-S* model was scaled up using a compound scaling strategy to get *EfficientNetV2-M* and *-L* models. The idea behind compound scaling is to balance the dimensions of depth (d), width (w), and input image

resolution (r) by scaling them to a constant ratio. Mathematically, it was formulated as given below.

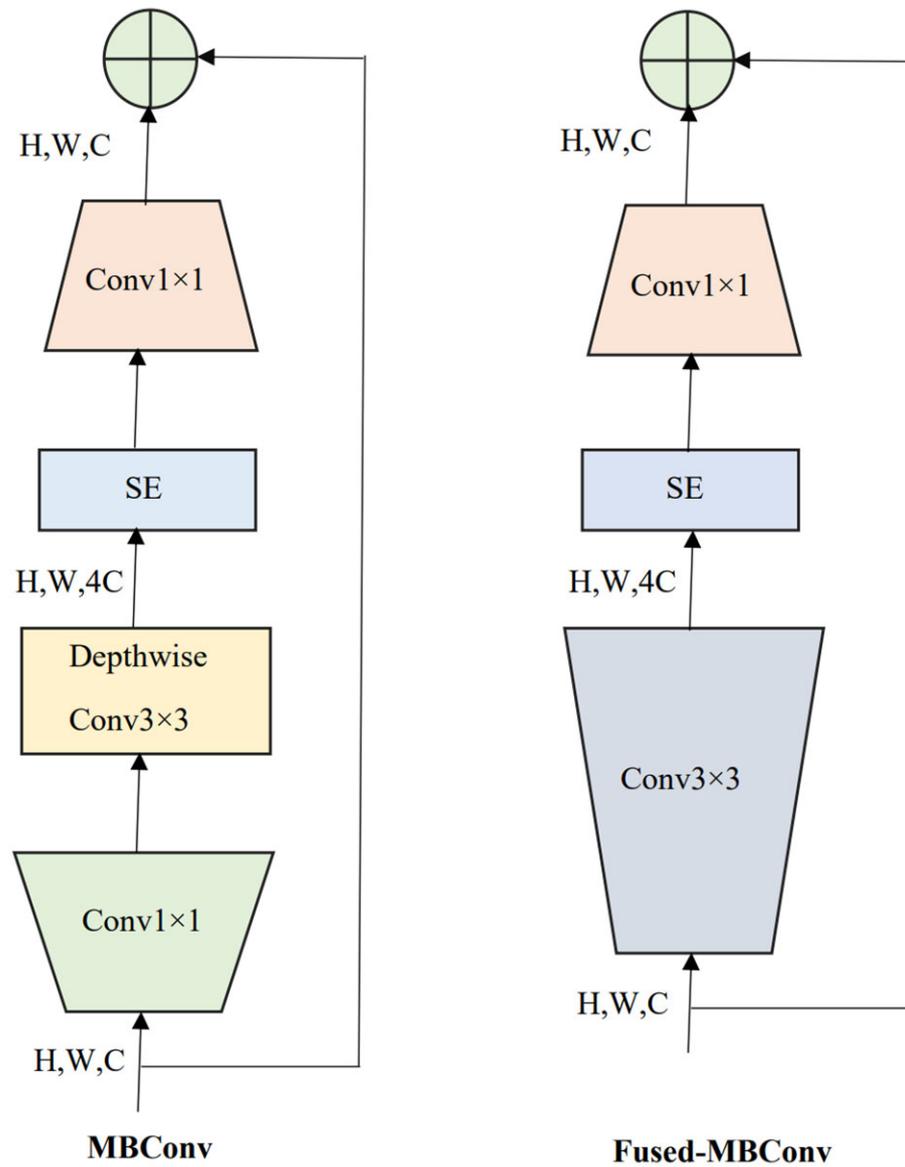


Figure 2. MBConv and Fused-MBConv layers are used as building blocks of *EfficientNetV2* models. SE: squeeze and excitation block. H, W, C : image height, width, and the number of channels.

$$d = \alpha^\varphi, w = \beta^\varphi, r = \gamma^\varphi$$

$$\text{such that } \alpha \cdot \beta^2 \cdot \gamma^2 = 2 \tag{1}$$

The values of α, β , and γ are always greater than or equal to one and could be determined by grid search. Intuitively, φ determines the extra computational resources required for model scaling, which is user defined. In practice, the convolution operations dominate the computational cost in CNNs. Hence, scaling a CNN using Equation (1) would roughly increase the floating-point operations per second (FLOPS) by $(\alpha \cdot \beta^2 \cdot \gamma^2)^\varphi$. However, based on the constraint set in Equation (1), for any new φ , the FLOPS in total will approximately increase by 2^φ . By progressively increasing the size of the image during training, the training speed was further improved. However, this gradual increase in the image size during training often leads to a drop in performance which was handled by adaptive regularization schemes, such

as data augmentation and dropout. That means for smaller image sizes, weak augmentation was used and vice versa. Furthermore, for complete details, refer to [9].

2.4. Model Training and Validation

The final softmax layer of the pre-trained *EfficientNetV2-S*, *-M*, and *-L* models were discarded, and a new softmax layer is added to classify lung and colon cancer subtypes. The model cross-validation and testing were conducted under the Google Colab Pro cloud environment using TensorFlow 2.0 with high-level *Keras* API at the backend. Furthermore, all the hyperparameters for all the models were empirically selected, and hence the validation set was used to ensure that the individual models were not over-fitting during training. For training, the *Adadelta* optimizer was used at 0.1 learning rate, 32 batch size, and 5 epochs. Since it is a five-class classification problem, sparse categorical cross-entropy (SCCE), as given in Equation (2), was used as the loss function.

$$SCCE_{loss} = -\frac{1}{N} \sum_{i=0}^N \sum_{j=1}^5 y_j \log(\hat{y}_j) \quad (2)$$

Above, N is the total number of images during training/validation, \hat{y}_j is the label of predicted class, and y_j is the label of the true class. For all three models, the parameters of the last 50 percent of layers were fine-tuned during training, and the parameters of the first half of the network remain unaltered. We have used two repetitions for splitting the data into training, and testing and the average performance metric values are reported.

2.5. Visual Saliency Maps

To better understand the model's decisions on where it is keeping more attention on the histopathology image during prediction, the visual saliency maps are created for each *EfficientNetV2* model using gradCAM [19] for all lung and colon cancer subtypes. To obtain the gradCAM map $L_{gradCAM}^c \in \mathbb{R}^{u \times v}$ of width u and height v for class c , indicating the most representative regions, we initially compute the first order derivative of the score for class c denoted as y^c (before the softmax), with respect to the feature maps A^k of the last convolutional layer. Furthermore, these first order derivatives propagated back are global mean pooled over the width and height of A^k (indexed by i and j , respectively) to get the neuron significance weights α_k^c . Mathematically, it is described as given below in Equation (3). Here, Z is the product of the width and height of the feature map A^k . The importance weights α_k^c captures the 'importance' of feature map A^k for a class of interest c . Furthermore, to get $L_{gradCAM}^c$, a weighted sum of final convolution layer output maps followed by ReLU (rectified linear unit) is performed as shown in Equation (4). Furthermore, ReLU is given in Equation (5). A ReLU is applied eventually to extract the 'positive' features that influence the class of interest.

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (3)$$

$$L_{gradCAM}^c = ReLU \left(\sum_k \alpha_k^c A^k \right) \quad (4)$$

$$ReLU(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (5)$$

2.6. Evaluation Metrics

To conduct the performance evaluation of the proposed models, the Python-based scikit-learn toolbox was used. The metrics include accuracy, F1-score, balanced accuracy (BA), area under the receiver operating characteristic curve (AUC), and Matthew's correlation coefficient (MCC), as described in the below equations. Here, F1-score is calculated

from the harmonic mean of precision and sensitivity, whereas *BA* is computed as the average of recall and specificity. Since it was a five-class classification study, the performance scores are obtained from the corresponding confusion matrix (*CM*) by employing one vs. rest approach. Given a specific class, the correctly classified images are categorized as true positives (*TP*). The false positives (*FP*) are the misclassifications above the half-diagonal of *CM*. The number of correctly classified present in the diagonal of *CM* other than the specific class are called true negatives (*TN*). Eventually, the misclassifications below the half diagonal are considered as false negatives (*FN*).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$F1 - score = \frac{2 * precision * recall}{precision + recall} \quad (7)$$

$$BA = \frac{sensitivity + specificity}{2} \quad (8)$$

$$sensitivity (recall) = \frac{TP}{TP + FN} \quad (9)$$

$$specificity = \frac{TN}{TN + FP} \quad (10)$$

$$precision = \frac{TP}{TP + FP} \quad (11)$$

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (12)$$

3. Results

All the models converged within five iterations during the cross-validation. Therefore, the evaluation scores during validation are very close to the performance scores during testing. Table 3 presents the complete evaluation details of the proposed *EfficientNetV2-S*, *-M*, *-L* models on the test set. The *EfficientNetV2-L* model performed better among all three, with an accuracy of 99.97% and an AUC of 99.99%. However, the other two models (small and medium) abilities are also very close to the performance metrics of the large model. For example, from Figure 3, we can see that the large model achieved almost 100% accuracy for the three-class stratification of lung cancer. Similarly, the medium model has obtained 100% accuracy for the two-class classification of colon cancer.

Table 3. Evaluation metrics on the test set for classifying lung and colon cancer subtypes using *EfficientNetV2-S/M/L* models given in percentages. BA: balanced accuracy, AUC: area under the curve, MCC: Matthew’s correlation coefficient.

	<i>EfficientNetV2-S</i>	<i>EfficientNetV2-M</i>	<i>EfficientNetV2-L</i>
Accuracy	99.90	99.96	99.97
AUC	99.99	99.99	99.99
F1-Score	99.90	99.96	99.97
BA	99.90	99.97	99.97
MCC	99.87	99.94	99.96

Furthermore, Figure 4 shows the visual saliency maps for a sample image for all lung and colon cancer subtypes using gradCAM. For comparison, the maps were generated for all three employed models of *EfficientNetV2*. In general, the highlighted regions in the histopathology image are similar among the different models. However, some notable dif-

ferences are present. For instance, the most activated regions during colon adenocarcinoma prediction using medium and large models are slightly different. In addition, the activated regions for the three-class lung cancer classification are wider for the small model compared to the medium and large models. Furthermore, we have given a color bar applicable to all the sub-saliency maps present in Figure 4 for quantitative estimate of attention. Here, red color indicates more attention (maximum value being one), and blue color indicates less attention (minimum value being zero) that the model put over the test histopathology image during the class prediction.

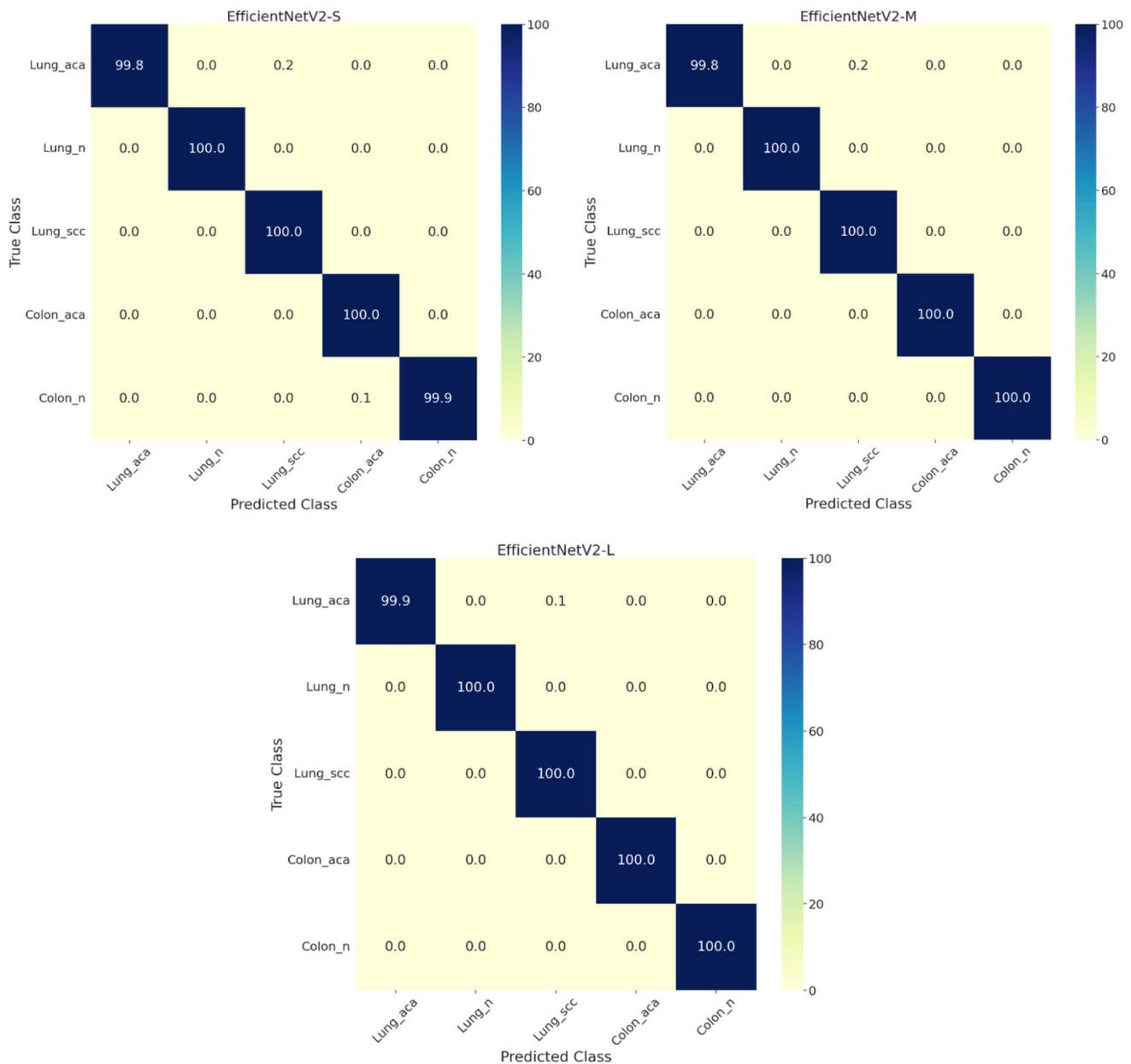


Figure 3. Multi-class confusion matrices for the test set for lung and colon cancer classification by employing *EfficientNetV2-S*, *-M*, and *-L* models. Lung_aca: lung adenocarcinoma, Lung_n: lung benign, Lung_scc: lung squamous cell carcinoma, Colon_aca: colon adenocarcinoma, Colon_n: colon benign.

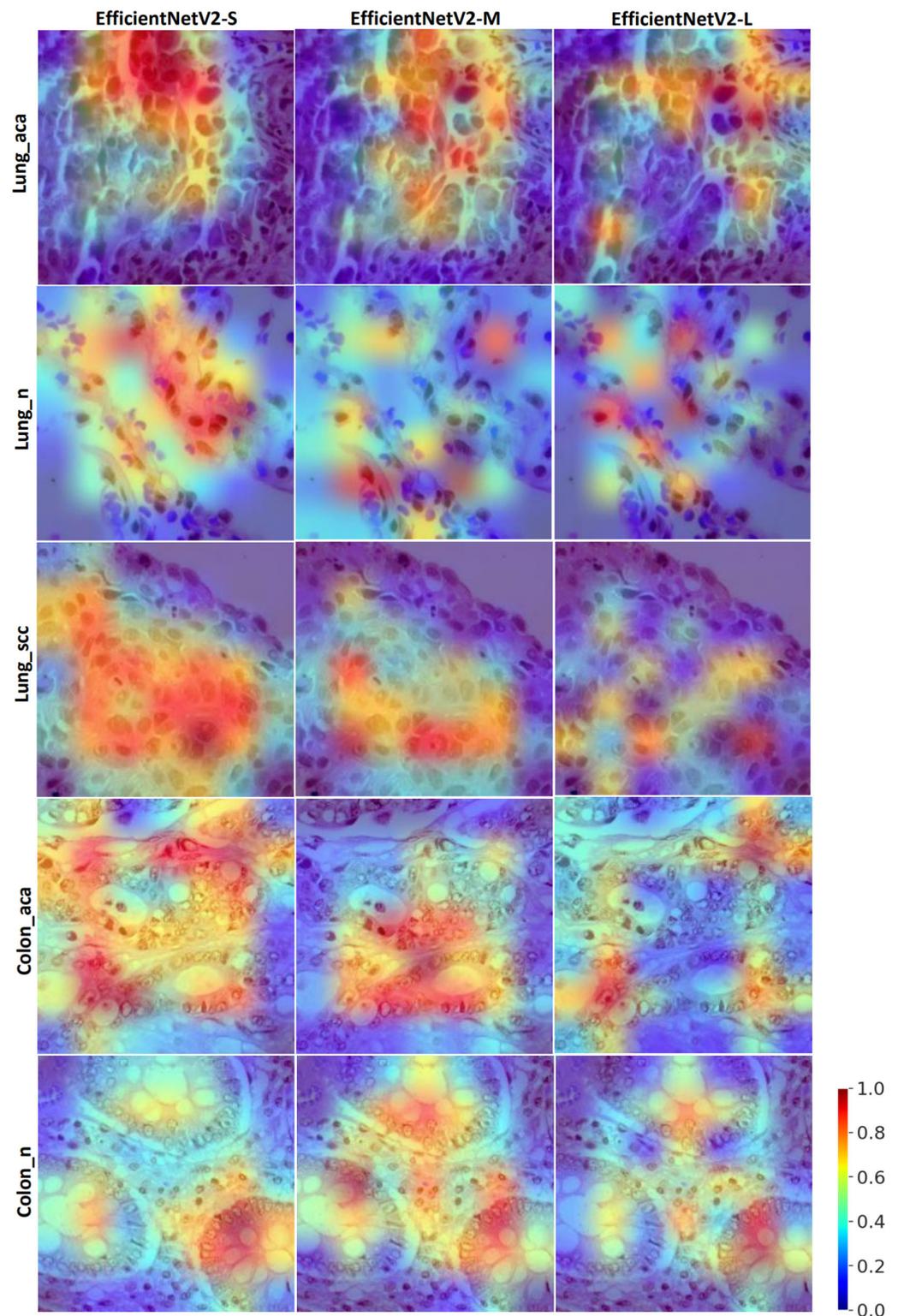


Figure 4. Visual saliency maps for explainability of the model's decisions during class prediction, created using gradCAM. For each class, one image is randomly picked from the test set. Lung_aca: lung adenocarcinoma, Lung_n: lung benign, Lung_scc: lung squamous cell carcinoma, Colon_aca: colon adenocarcinoma, Colon_n: colon benign. The red color in the maps indicates that more attention is given in those regions, and the blue color indicates that less attention is put to those regions during model prediction.

4. Discussion

In the present study, we proposed a pipeline using pretrained *EfficientNetV2* models (*L*, *M*, and *S*) for the automated classification of lung and colon cancer subtypes from histopathology images of LC25000 dataset. These compound-driven architectures outperformed the existing works on the same dataset by achieving an accuracy up to 99.97%, including all five classes indicating the power of both compound scaling and TL. Hence, the models may essentially replace the pathologist and make the classification of lung and colon cancer fully automatic. Furthermore, our framework is end-to-end, requiring neither any pre-processing methods nor any dimensionality reduction strategies as employed in some previous studies to achieve accuracies above 99.5% [26,31]. For instance, the method in [27] used histogram equalization for colon cancer images to boost the overall accuracy from 89% to 98.4%, and we believe that extensive pre-processing may hamper the generalizable ability of the model to unseen data. The better overall performance of *EfficientNetV2-L* model could be due to the presence of a greater number of MBConv and Fused-MBConv layers that helped in learning the most relevant abstract features required for very accurate classification.

Looking at the visual saliency maps in Figure 4, we can understand the most activated regions during the target class prediction by the models. In general, the most activated areas in the image are widespread for small model relative to medium and large models since the small model has comparatively few parameters/layers, and to achieve better differentiability among subtypes, attention over large area of the image may be necessary. This trend was more apparent for lung cancer subtypes and colon adenocarcinoma. Interestingly, the medium model demonstrated wider activations for colon benign compared to the small and large models. Overall, we can observe from the saliency maps that all models' feature abstraction is from the appropriate areas of the histopathological images consistent across all lung and colon cancer subtypes. Furthermore, the color bar guides the pathologists to quantitatively measure (in the scale between zero and one) the amount of attention/importance the model put over the subregions of the test histopathology images during their class prediction. Furthermore, these visual saliency maps may assist pathologists in potentially designing individual treatment strategies.

Since the dataset was largely generated by augmenting the original dataset containing 250 histopathology images for each cancer subtype, data augmentation may not provide true data variability. Hence, future studies should involve testing the proposed models on larger datasets created without using any data augmentation. Although the hyperparameters during training were chosen empirically, a thorough grid search, including the selection of optimizer, could be conducted using cross validation. Nonetheless, the performance metrics are quite impressive on the test set across all three models, thus, strongly supporting the empirically chosen hyperparameters. In addition, it will be interesting to implement few-shot learning methods [37] that work based on small sample sizes as an alternative to increasing the dataset size using heavy data augmentation.

Deep learning with *EfficientNetV2* large, medium, and small models with high accuracies of 99.96% can perform an important role in diagnosis and treatment of carcinoma lung and colon. This algorithm can be employed to analyze the vast amounts of data generated for cancer diagnosis, including images of tissue samples viewed under a microscope, genetic data, and other clinical information. One of the key advantages of this deep learning model is its ability to analyze large datasets and identify patterns that may be difficult for human experts to discern. This can significantly improve the accuracy of cancer diagnosis, particularly in cases where subtle differences between healthy and cancerous tissue may be difficult to distinguish. In addition to improving diagnosis, it can also be used to develop personalized treatment plans for cancer patients according to disease severity. This can be done by analyzing data from large numbers of patients with similar genetic profiles. The presented algorithm can identify the most effective treatment options for individual patients based on their unique characteristics. Overall, the role of deep learning in CA lung and CA colon histopathology is significant, as it has the potential to improve the accuracy

of cancer diagnosis, reduces histopathologist's burden, identify new treatment options, and ultimately help save lives.

5. Conclusions

The *EfficientNetV2* based *-L*, *-M*, and *-S* models presented in this study have achieved accuracies above 99.96% AUCs of 99.99%, and MCC of up to 99.96% on the test dataset for the five-class classification that includes three lung cancer subtypes and two colon cancer subtypes from histopathology images. The performance is superior to the existing works using the LC25000 dataset, and, furthermore, we employed gradCAM to highlight the most important regions while target class prediction. The performance metrics of the classification are marginally superior for *-M* and *-L* models compared to *-S* model. Hence, the proposed framework may assist pathologists in fully automating the lung and colon cancer subtyping from histopathological images and interpretability. In the future, we would like to propose lightweight models for the same task that could be deployable on edge devices. The code of the proposed pipeline can be found here.

Author Contributions: Conceptualization and writing, S.T.; methodology, S.K.; software, A.N.; validation, H.T.R.; writing, N.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by Researchers Supporting Project Number (RSP2023R124), King Saud University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: This research study was conducted retrospectively using human subject data made available in open access by Kaggle. Ethical approval was not required, as confirmed by the license attached with the open access data.

Informed Consent Statement: This research study was conducted retrospectively using human subject data made available in open access by Kaggle. Hence, written informed consent is not required.

Data Availability Statement: The data used in the study is publicly available from Kaggle.

Acknowledgments: We thank the providers of LC25000 dataset. The authors acknowledge and extend their appreciation to the Researchers Supporting Project Number (RSP2023R124), King Saud University, Riyadh, Saudi Arabia for funding this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. de Martel, C.; Georges, D.; Bray, F.; Ferlay, J.; Clifford, G.M. Global Burden of Cancer Attributable to Infections in 2018: A Worldwide Incidence Analysis. *Lancet. Glob. Health* **2020**, *8*, e180–e190. [[CrossRef](#)] [[PubMed](#)]
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
3. Kurishima, K.; Miyazaki, K.; Watanabe, H.; Shiozawa, T.; Ishikawa, H.; Satoh, H.; Hizawa, N. Lung Cancer Patients with Synchronous Colon Cancer. *Mol. Clin. Oncol.* **2018**, *8*, 137–140. [[CrossRef](#)] [[PubMed](#)]
4. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [[CrossRef](#)]
5. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
6. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9. [[CrossRef](#)]
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. [[CrossRef](#)]
8. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [[CrossRef](#)]
9. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning ICML, Long Beach, CA, USA, 9–15 June 2019; pp. 10691–10700.
10. Tan, M.; Le, Q.V. *EfficientNetV2*: Smaller Models and Faster Training. *arXiv* **2021**, arXiv:2104.00298.
11. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:010.11929.

12. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, New Orleans, LA, USA, 18–24 June 2022; pp. 9992–10002. [\[CrossRef\]](#)
13. Tummala, S. Deep Learning Framework Using Siamese Neural Network for Diagnosis of Autism from Brain Magnetic Resonance Imaging. In Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2 April 2021; pp. 1–5.
14. Yousef, R.; Gupta, G.; Yousef, N.; Khari, M. A Holistic Overview of Deep Learning Approach in Medical Imaging. *Multimed. Syst.* **2022**, *28*, 881. [\[CrossRef\]](#)
15. Tummala, S.; Kim, J.; Kadry, S. BreaST-Net: Multi-Class Classification of Breast Cancer from Histopathological Images Using Ensemble of Swin Transformers. *Mathematics* **2022**, *10*, 4109. [\[CrossRef\]](#)
16. Galib, S.M.; Lee, H.K.; Guy, C.L.; Riblett, M.J.; Hugo, G.D. A Fast and Scalable Method for Quality Assurance of Deformable Image Registration on Lung CT Scans Using Convolutional Neural Networks. *Med. Phys.* **2020**, *47*, 99–109. [\[CrossRef\]](#)
17. Tummala, S.; Thadikemalla, V.S.G.; Kadry, S.; Sharaf, M.; Rauf, H.T. *EfficientNetV2* Based Ensemble Model for Quality Estimation of Diabetic Retinopathy Images from DeepDRiD. *Diagnostics* **2023**, *13*, 622. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 December 2016; pp. 2921–2929. [\[CrossRef\]](#)
19. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2016**, *128*, 336–359. [\[CrossRef\]](#)
20. Chattopadhyay, A.; Sarkar, A.; Howlader, P.; Balasubramanian, V.N. Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 839–847. [\[CrossRef\]](#)
21. Teramoto, A.; Tsukamoto, T.; Kiriya, Y.; Fujita, H. Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks. *Biomed Res. Int.* **2017**, *2017*, 4067832. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Hatuwal, B.K.; Thapa, H.C. Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images. *Int. J. Comput. Trends Technol.* **2020**, *68*, 21–24. [\[CrossRef\]](#)
23. Hasan, I.; Ali, S.; Rahman, H.; Islam, K. Automated Detection and Characterization of Colon Cancer with Deep Convolutional Neural Networks. *J. Healthc. Eng.* **2022**, *2022*, 5269913. [\[CrossRef\]](#)
24. Masud, M.; Sikder, N.; Nahid, A.A.; Bairagi, A.K.; Alzain, M.A. A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework. *Sensors* **2021**, *21*, 748. [\[CrossRef\]](#)
25. Hage Chehade, A.; Abdallah, N.; Marion, J.M.; Oueidat, M.; Chauvet, P. Lung and Colon Cancer Classification Using Medical Imaging: A Feature Engineering Approach. *Phys. Eng. Sci. Med.* **2022**, *45*, 729–746. [\[CrossRef\]](#)
26. Ali, M.; Ali, R. Multi-Input Dual-Stream Capsule Network for Improved Lung and Colon Cancer Classification. *Diagnostics* **2021**, *11*, 1485. [\[CrossRef\]](#)
27. Mehmood, S.; Ghazal, T.M.; Khan, M.A.; Zubair, M.; Naseem, M.T.; Faiz, T.; Ahmad, M. Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning with Class Selective Image Processing. *IEEE Access* **2022**, *10*, 25657–25668. [\[CrossRef\]](#)
28. Toğaçar, M. Disease Type Detection in Lung and Colon Cancer Images Using the Complement Approach of Inefficient Sets. *Comput. Biol. Med.* **2021**, *137*, 104827. [\[CrossRef\]](#)
29. Kumar, N.; Sharma, M.; Singh, V.P.; Madan, C.; Mehandia, S. An Empirical Study of Handcrafted and Dense Feature Extraction Techniques for Lung and Colon Cancer Classification from Histopathological Images. *Biomed. Signal Process. Control* **2022**, *75*, 103596. [\[CrossRef\]](#)
30. Talukder, M.A.; Islam, M.M.; Uddin, M.A.; Akhter, A.; Hasan, K.F.; Moni, M.A. Machine Learning-Based Lung and Colon Cancer Detection Using Deep Feature Extraction and Ensemble Learning. *Expert Syst. Appl.* **2022**, *205*, 117695. [\[CrossRef\]](#)
31. Attallah, O.; Aslan, M.F.; Sabanci, K. A Framework for Lung and Colon Cancer Diagnosis via Lightweight Deep Learning Models and Transformation Methods. *Diagnostics* **2022**, *12*, 2926. [\[CrossRef\]](#)
32. Mangal Engineerbabu, S.; Chaurasia Engineerbabu, A.; Khajanchi, A. Convolution Neural Networks for Diagnosing Colon and Lung Cancer Histopathological Images. *arXiv* **2020**, arXiv:2009.03878.
33. Borkowski, A.A.; Bui, M.M.; Thomas, L.B.; Wilson, C.P.; DeLand, L.A.; Mastorides, S.M. Lung and Colon Cancer Histopathological Image Dataset (LC25000). *arXiv* **2019**, arXiv:1912.12142.
34. Ihde, D.C.; Minna, J.D. Non-Small Cell Lung Cancer. Part I: Biology, Diagnosis, and Staging. *Curr. Probl. Cancer* **1991**, *15*, 65–104. [\[CrossRef\]](#)
35. Cappell, M.S. Pathophysiology, Clinical Presentation, and Management of Colon Cancer. *Gastroenterol. Clin. N. Am.* **2008**, *37*, 1–24. [\[CrossRef\]](#)
36. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [\[CrossRef\]](#)
37. Tummala, S.; Suresh, A.K. Few-Shot Learning Using Explainable Siamese Twin Network for the Automated Classification of Blood Cells. *Med. Biol. Eng. Comput.* **2023**, *1*, 1–15. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.