Engineering Applications of Artificial Intelligence 126 (2023) 106902

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images

Rehan Raza [a,b], Fatima Zulfiqar [c,d], Muhammad Owais Khan [b,c], Muhammad Arif [e], Atif Alvi [b], Muhammad Aksam Iftikhar [c], Tanvir Alam [e,*]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Department of Computer Science, School of Systems and Technology, University of Management and Technology, Lahore, Pakistan
[c] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Lahore, Pakistan
[d] Department of Computer Science, Bahria University Lahore Campus, Lahore, Pakistan
[e] College of Science and Engineering, Hamad Bin Khalifa University, Doha 34110, Qatar

ARTICLE INFO

Keywords:
Lung cancer
Deep learning
EfficientNetB1
Transfer learning
Medical imaging

ABSTRACT

Lung cancer (LC) remains a leading cause of death worldwide. Early diagnosis is critical to protect innocent human lives. Computed tomography (CT) scans are one of the primary imaging modalities for lung cancer diagnosis. However, manual CT scan analysis is time-consuming and prone to errors/not accurate. Considering these shortcomings, computational methods especially machine learning and deep learning algorithms are leveraged as an alternative to accelerate the accurate detection of CT scans as cancerous, and non-cancerous. In the present article, we proposed a novel transfer learning-based predictor called, Lung-EffNet for lung cancer classification. Lung-EffNet is built based on the architecture of EfficientNet and further modified by adding top layers in the classification head of the model. Lung-EffNet is evaluated by utilizing five variants of EfficientNet i. e., B0–B4. The experiments are conducted on the benchmark dataset "IQ-OTH/NCCD" for lung cancer patients grouped as benign, malignant, or normal based on the presence or absence of lung cancer. The class imbalance issue was handled through multiple data augmentation methods to overcome the biases. The developed model Lung-EffNet attained 99.10% of accuracy and a score of 0.97 to 0.99 of ROC on the test set. We compared the efficacy of the proposed fine-tuned pre-trained EfficientNet with other pre-trained CNN architectures. The predicted outcomes demonstrate that EfficientNetB1 based Lung-EffNet outperforms other CNNs in terms of both accuracy and efficiency. Moreover, it is faster and requires fewer parameters to train than other CNN based models, making it a good choice for large-scale deployment in clinical settings and a promising tool for automated lung cancer diagnosis from CT scan images.

1. Introduction

Cancer has become a major concern to public health and a silent killer of millions of people around the world. According to the survey report of the World Health Organization (WHO), cancer was considered the second highest mortality with an estimated 10 million deaths in 2020 (American Cancer Society, 2023). Cancer can occur in any part of the body and can affect people of all ages. There are many types of cancer i.e., with more than 100 different types, each with its own set of characteristics and behaviors (U.S.D.o.H.a.H. Services et al., 2023). The most prevalent forms of cancer are lung, breast, colorectal, liver, and stomach cancer (WHO, 2023). LC originates inside the lungs. The lungs are a pair of organs located in the chest that perform respiration i.e., exchanging oxygen and carbon dioxide in the body. In 2021, it was reported that approximately 1.80 million people have died due to lung cancer (Hiremath et al., 2022). Lung cancer can be categorized into two types from a broad spectrum: small cell lung cancer (SCLC) and non-SCLC (NSCLC) (Fontana et al., 1991).

The classification of LC provides in-depth knowledge to understand from a patient management perspective, because different types of lung cancer may require different treatment approaches. Conventional Wet-lab methods of lung cancer classification involve the visual interpretation of imaging studies applying certain criteria for example tumor size, shape, and location to determine the type and stage of cancer (Tekade

* Corresponding author.
E-mail address: talam@hbku.edu.qa (T. Alam).

https://doi.org/10.1016/j.engappai.2023.106902
Received 17 March 2023; Received in revised form 18 July 2023; Accepted 31 July 2023
Available online 3 September 2023

and Rajeswari, 2018). However, these methods are subjective and may result in inter-observer variability. To address this challenge, researchers have developed computational methods for automating the analysis of medical images for lung cancer classification (Asuntha et al., 2020). Recently, advancement in artificial intelligence has significantly contributed to designing automated approaches for the accurate analysis of medical imaging data, including the classification of lung cancer. These methods can provide more objective and precise results, and improve the overall performance of lung cancer classification (Ge et al., 2021).

Medical imaging, such as magnetic resonance imaging (MRI) and CT are traditional methods for lung cancer classification and diagnosis (Coudray et al., 2018). Images from the CT scans and MRI are the widely used image acquisition techniques in medical imaging, but each has its characteristics and purpose. CT scans method captures the internal body images using X-rays. They are particularly useful for detecting and identifying abnormalities in the chest, such as lung cancer or pneumonia (Gordienko et al., 2017). In contrast, MRI captures internal body images using radio waves and a strong magnetic field. MRI has been proven a significant tool due to capturing high-resolution images without using ionizing radiation for imaging problems likewise joints, the spine, and the brain (Lustberg et al., 2018).

The medical image analysis for lung cancer classification involves two types of approaches: machine learning algorithms, which can learn to classify cancer based on features extracted from the images. Another approach for image analysis is designing deep learning models (Kadir and Gleeson, 2018). There are many variants of DL models, however, Convolutional Neural Networks (CNNs) are the most widely used model for image and video classification tasks. CNN is particularly ideal due to the hierarchical learning representation of data, which allows them to capture complex information from raw data. The CNN is capable of automatically learning and extracting the prominent patterns from the input image without interference from manual feature encoding. This allows the CNN to effectively classify images even if they are significantly different from the training examples, as long as they contain the same types of features. CNNs require a large amount of data to achieve good performance. This can be a challenge in cases where it is difficult to obtain a large dataset (Ge et al., 2022).

Transfer learning (TL) is another widely used approach to tackle the shortcomings of the CNN model. For TL, the pre-trained CNN models are often used as a baseline and the model is subsequently improved using a new dataset. This can be done by unfreezing some of the layers in the pre-trained model and training them on the new data, or by adding additional layers on top of the pre-trained model and training the entire model end-to-end fashion. These approaches automatically extract the computationally better features. Medical image analysis using deep learning techniques has the potential to enhance the overall robustness, performance, and efficiency of LC classification (Sun et al., 2016). Transfer learning has emerged as a valuable technique in the field of lung cancer detection and classification, as evidenced by several studies. The study by Wang et al. (2020) focuses on the classification of different pathological types of lung cancer using CT images. They employ a transfer learning strategy with deep residual neural networks (ResNets) to achieve accurate classification results. By leveraging pre-trained models and fine-tuning them on a specific lung cancer dataset, the authors demonstrate the effectiveness of transfer learning in improving classification performance. In another study by Sajja et al. (2019) deep transfer learning is employed for lung cancer detection based on CT scan images. By utilizing pre-trained models and adapting them to the task of lung cancer detection, the researchers achieve promising results. This highlights the importance of transfer learning in leveraging pre-existing knowledge from related tasks to improve lung cancer detection accuracy. Katsamenis et al. (2020) explore the application of transfer learning in COVID-19 pneumonia detection and classification using chest X-ray images. Transfer learning is also utilized for the differential diagnosis of lung cancer using conventional CT and FDG PET/CT images. The study showed that transfer learning, combined with metadata, can enhance the performance of deep learning systems in distinguishing between benign and malignant lung lesions (Chen et al., 2021a). This highlights the importance of transfer learning in leveraging existing knowledge to improve the accuracy of a lung cancer diagnosis. Furthermore, Protonotarios et al. (2022) propose a few-shot U-Net deep learning model for lung cancer lesion segmentation via PET/CT imaging. They employ transfer learning to leverage pre-trained models and adapt them to the task of segmenting lung cancer lesions. This demonstrates the significance of transfer learning in enhancing the efficiency and accuracy of lung cancer lesion segmentation, a critical step in diagnosis and treatment planning. Collectively, the aforementioned studies emphasize the importance of transfer learning in the domain of lung cancer detection, classification, and segmentation. By leveraging pre-trained models and fine-tuning them on specific lung cancer datasets, transfer learning enables the utilization of existing knowledge and significantly improves the accuracy and performance of deep learning models. This approach proves valuable in enhancing the understanding and diagnosis of lung cancer, ultimately leading to more effective treatment strategies and improved patient outcomes (Hu et al., 2023).

The primary objective of this study is to ascertain the presence of cancerous cells in lung nodule CT-scan images. To achieve this, the research focuses on classifying lung cancer into three distinct categories: Benign, Malignant, and Normal, based on CT-scan slices of lung nodules. To accomplish this classification task, transfer learning is employed using five variants of pre-trained EfficientNet models, specifically EfficientNet B0 to B4. Each variant of the EfficientNet model is fine-tuned explicitly on CT-scan slices of lung cancer. Fine-tuning involves adapting the pre-trained models to the specific characteristics and features of the lung cancer dataset, allowing them to capture relevant patterns and information. During the fine-tuning process, the feature maps generated by the EfficientNet models are extracted. These feature maps represent high-level representations of the input CT-scan slices, capturing important spatial and semantic information. These extracted feature maps are then passed through fully connected layers, which are responsible for performing the classification task. The fully connected layers analyze the extracted features and make predictions regarding the presence of Benign, Malignant, or Normal lung cancer. By employing this transfer learning approach and leveraging the capabilities of the EfficientNet models, the study aims to achieve accurate and reliable classification results for lung cancer based on CT-scan slices of lung nodules (Lyu, 2021).

Transfer learning with EfficientNets can offer several advantages that may help address common limitations in the field of medical image analysis. EfficientNets are known for their ability to extract meaningful and representative features from images. By leveraging transfer learning, the pre-trained EfficientNet models can capture relevant features from large-scale datasets (e.g., ImageNet) and transfer this knowledge to the task of lung cancer segmentation. This can potentially improve the accuracy and robustness of the segmentation process (Humayun et al., 2022a). Additionally, EfficientNets have demonstrated their effectiveness in various computer vision tasks, including object recognition and segmentation. By fine-tuning these models on medical image datasets, such as lung cancer images, the network can learn to generalize well to the specific characteristics and variations present in medical images, contributing to improved segmentation performance. Moreover, EfficientNets are designed to achieve a good balance between model size and performance. They have shown superior efficiency in terms of computational resources and memory requirements compared to other deep learning architectures. This can be particularly advantageous in the medical field, where limited computational resources and time constraints are often a concern (Al-Yasriy et al., 2020a).

The present article primarily contributes to the following:

- We developed a novel transfer learning framework Lung-EffNet using EfficientNetB1 for lung cancer classification.

- We solved the severe imbalance issue by using the augmentation method to overcome the skewness of data.
- To evaluate the impact of data augmentation on the performance of the proposed model, extensive experiments were carried out on the benchmark dataset, both with and without the implementation of data augmentation.
- We compared the novel-designed model with other advanced methods in terms of execution time and computational complexity to demonstrate the performance of EfficientNet over other classification models.
- Our proposed model Lung-EffNet demonstrates superior performance in comparison to existing methods and targets lung cancer from CT scan images.

The content of the paper is structured as follows: Section 2 reviews existing techniques for predicting lung cancer. Section 3 elaborates on the proposed method, which utilizes fine-tuning and transfer learning of the EfficientNet model. The experiments and comparisons with recent techniques on IQ-OTH/NCCD datasets are explained in Section 4. An ablation study has been conducted in Section 5 to assess the generalizability and robustness of the proposed Lung-Effnet encompassing several key parameters. Finally, the conclusion and potential future research directions are discussed in Section 6.

2. Research background

Lung cancer diagnosis and treatment at an early stage is extremely important to prevent the loss of human lives. LC is considered a lethal disease causes due to abnormal growth of malignant cells either within one or both lungs that might spread further to the other human organs if not treated early. Therefore, an efficient computer-aided-diagnosis (CAD) based framework is required which is capable of automatically detecting and classifying the presence of lung cancer with higher accuracy. Detection, segmentation, and classification of cancerous lung nodules can be carried out from several medical imaging such as CT scans (Sun et al., 2016), X-rays (Gordienko et al., 2017), Histopathology images (Narin and Onur, 2022), and Sputum Smear Microscopy Images (Lyu, 2021).

Before the advancement of deep learning, manual feature extraction using several image processing techniques such as top-hat transformation, noise removal using median, adaptive bilateral filtering, segmentation using marker watershed algorithm, particle swarm optimization, fuzzy C-means clustering, and morphological operations have been used. The extracted feature vectors were then passed to the multiple ML classifiers for segmentation and classification (Kavitha and Prabakaran, 2019). Kareem et al. (2021) also proposed a CAD-based system for the classification of lung cancer into either cancerous/malignant or non-cancerous/non-malignant. Here, images were first preprocessed by applying three image-processing techniques of image enhancement, segmentation, and feature extraction using Gabor and GLCM filters which were then classified using an SVM classifier. Their proposed method obtained 89.88% accuracy using a polynomial kernel. Machine learning-based methods work well with fewer datasets however, they require manual feature extraction and feature selection processes which increase the computational time and complexity. The aforementioned limitations have been overcome using different variants of DL models especially CNN for image classification, detection, and segmentation tasks.

Several studies have used deep CNN (DCNN) for LC classification. Chen et al. (2021b) proposed a chatbot using CNN and natural language processing (NLP) for the classification of malignant nodules from non-malignant nodules of lung cancer from CT-scan images. Their proposed custom-built CNN architecture with four convolutional and pooling layers obtained a validation accuracy of 88%. The chatbot was developed using a Bi-directional long short-term memory (BLSTM) and attention model termed as Hierarchical Bi-Long Short-Term Memory Attention Model (HBAM). The primary objective of their study was to facilitate the doctor in the diagnosis and treatment planning of the patient by communicating with the patient who is suffering from lung cancer and determining their condition. Tekade and Rajeswari (2018) also proposed a 12-layer deep custom-built deep CNN architecture for the classification of lung cancer into two classes i.e., benign, and malignant. They also compared their custom-built DCNN model with the pre-trained VGG16 and Inception-V3, which showed that VGG-16 achieved better results.

Deep learning architectures are stochastic in nature which gives different results for each prediction, and the outcome from the deep learning model involves randomness and uncertainty. In addition, there is a possibility of a model overfitting the training data due to high variance in the dataset. To overcome the aforementioned issues, ensemble learning has been used by many studies in different research domains including medical imaging (Raza et al., 2022). Many studies used ensemble learning for the classification of lung cancer from CT-scan Images. Ensemble learning combines the prediction of different models that are trained and evaluated on the same dataset. Dass et al. Raza et al. (2022) proposed a deep ensemble convolutional neural network (DECNN) for the both classification and segmentation of lung and liver cancer into three classes i.e. lung nodule, tumor, and normal. The segmentation of CT-scan images of lung cancer and the liver tumor is first carried out using morphological segmentation and blob feature extractor that is then passed to the DCNN for classification. Their proposed DECNN with blob-Net is validated on static and dynamic test data by achieving approximate accuracy of 75% on dynamic data and 99.8% accuracy on static data. Lyu (2021) present another study that proposed an ensemble of four individual DCNN architectures. Their proposed ensemble of pre-trained AlexNet, VGG-16, DCNN, and DenseNet is carried out by using both voting and stacking classifiers. DenseNet outperformed other pre-trained architectures if compared individually by achieving 97.8% test accuracy. Their proposed stacking ensemble of the top three independently trained pre-trained architectures showed better results with 99% test accuracy.

The limitation of methods based on DCNN architectures is that it requires large annotated datasets to train the model efficiently and avoid overfitting. This is relatively extremely difficult while dealing with medical imaging-based problems. Due to the privacy concerns of the patients, many hospitals and clinics do not share patient records publicly. Therefore, manual data collection is a huge problem. Another limitation includes the availability of computational resources such as graphical processing unit (GPU), and random access memory (RAM).

In view of the above-mentioned limitations, many researchers in various research domains have utilized transfer learning. Transfer learning that is also known as domain adaptive training utilizes the weight of the pre-trained model that is originally trained on a large dataset utilizing the weight-sharing process. Transfer learning has gained huge success that allowed the researchers to efficiently either train the pre-trained model by employing layers fine-tuning or feature extraction. Mamoona et al. Humayun et al. (2022b) proposed a transfer learning-based approach for the multi-class classification of lung cancer from CT-scan images into three classes: benign, malignant, and normal. They utilized a data augmentation technique to automatically expand the dataset samples. Three pre-trained architectures that include VGG16, VGG19, and Xception were retrained on the publicly available IQ-OTH/NCCD lung cancer dataset. Their proposed transfer learning of pre-trained VGG-16, VGG-19, and Xception networks obtained an overall accuracy of 98.83%, 98.05%, and 97.4%, respectively. Similarly, the transfer learning of AlexNet is carried out in Al-Yasriy et al. (2020b) the researchers utilized the architecture of AlexNet which is trained on the CT-scan images in the IQ-OTH/NCCD lung cancer dataset. The dataset is divided into two classes of malignant, and non-malignant instead of the original three classes. Their proposed methodology obtained an overall test accuracy, sensitivity, and specificity of 93.54%, 95.71%, and 95% respectively. Another study Narin and Onur (2022)

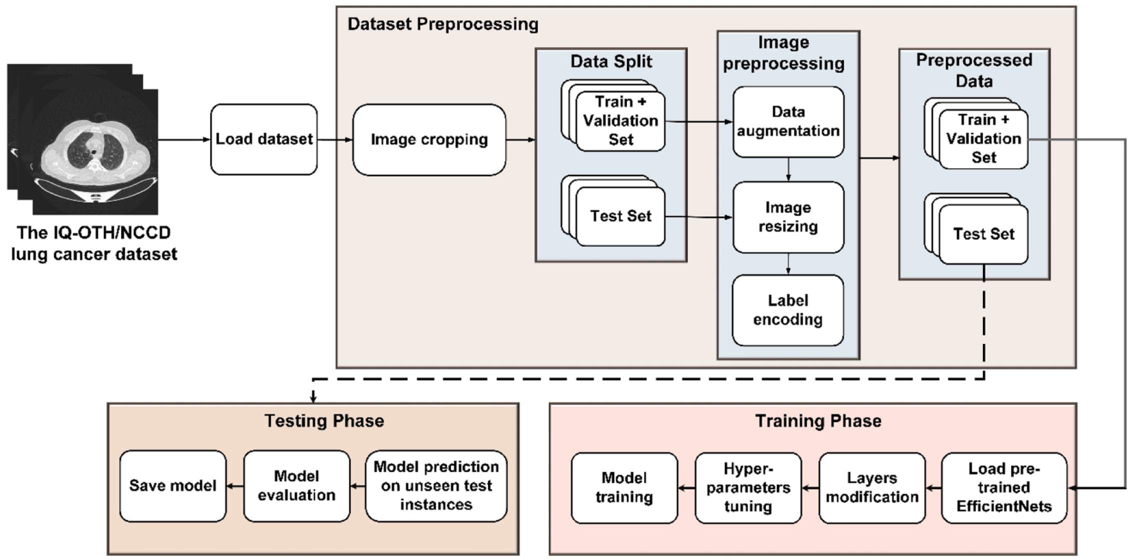

Fig. 1. The general workflow of the proposed method.

studied the effect of hyper-parameters tuning on the classification result for the lung cancer classification problem. They utilized AlexNet and ResNet50 architecture in their study. Both pre-trained architectures were used as feature extractors which were then passed to the SVM classifier for the classification of lung cancer. After explicitly hyper-parameters tuning, the proposed AlexNet model obtained higher accuracy of 98.58% when it was trained for 200 epochs with a batch size of 64.

Abida et al. (Sultana et al., 2021) compared the performance of their proposed hybrid 2D-CNN + SVM with four transfer learning-based models i.e. ResNet50, Inception-V3, InceptionResNetV2, and VGG19 for the classification of three types of lung cancer i.e., lung benign tissue, lung adenocarcinoma, and lung squamous cell carcinoma. The experiments were carried out on the histopathological images that are available on Kaggle. Their experiments showed that the transfer learning of Inception-V3 significantly outperformed other methods achieving 99.13% validation accuracy. Transfer learning of GoogleNet for lung cancer detection is proposed. The experiments were carried out on the dataset with two classes i.e., malignant, and non-malignant, and obtained a test accuracy of 94.38% (AL-Huseiny and Sajit, 2021).

Along with classification, some studies have focused on the detection and segmentation of lung tumors from the healthy surrounding tissues using either CT-scan or X-ray images of the lung by employing various transfer-learning-based techniques. Khan et al. (2021) proposed a novel approach for the simultaneous segmentation and classification of lung nodules in CT images using the VGG19 network. The authors begin by highlighting the importance of accurate detection and classification of lung nodules in the early diagnosis of lung cancer. The proposed approach utilizes the VGG19 network, a deep convolutional neural network (CNN), to jointly perform the segmentation and classification of lung nodules. The network is pre-trained on a large dataset and fine-tuned using a combination of nodule segmentation and classification loss functions. The segmentation module utilizes a combination of thresholding and morphological operations to delineate lung nodules accurately. Experimental results on a publicly available dataset demonstrate the effectiveness of the proposed approach. Jennifer and Sharmila (2023) proposed a novel approach for the automatic detection of lung infections using chest X-ray images based on neutrosophic sets. The author highlights the limitations of traditional methods in handling uncertainty and imprecision in medical data, emphasizing the importance of accurate and timely diagnosis of lung infections. The proposed approach leverages neutrosophic sets, which can represent indeterminacy, ambiguity, and inconsistency, to model the uncertainty in chest X-ray images and improve infection detection. Jaszcz et al. (2022) proposed a novel approach for lung X-ray image segmentation using the heuristic Red Fox Optimization (RFO) algorithm. The authors emphasize the importance of accurate lung image segmentation for various medical applications and highlight the challenges posed by the complex nature of lung structures and the presence of noise in X-ray images. The proposed approach utilizes the RFO algorithm, inspired by the foraging behavior of red foxes, to optimize the segmentation process. The algorithm aims to find the optimal thresholds for segmenting lung regions effectively. Experimental results demonstrate the effectiveness of the proposed approach in accurately segmenting lung regions in X-ray images. The approach achieves high segmentation accuracy and outperforms other existing methods in terms of performance metrics such as the Dice similarity coefficient and sensitivity. The researchers in Poap et al. (2018) proposed a method for segmenting chest radiographs to aid in the detection of lung diseases using a nature-inspired algorithm. The authors emphasize the significance of accurate segmentation in chest radiographs for the early detection and diagnosis of lung diseases. They highlight the challenges posed by image complexity and the need for robust segmentation techniques. The proposed approach utilizes a nature-inspired algorithm, which is inspired by natural phenomena or processes, to perform the segmentation of chest radiographs. Experimental results are presented to demonstrate the effectiveness of the proposed approach in segmenting chest radiographs for lung disease detection. The segmentation performance is evaluated using metrics such as accuracy, sensitivity, and specificity. Voulodimos et al. (2021) presented a deep-learning model for segmenting COVID-19-infected areas in CT images. The proposed model is based on the U-Net architecture and is designed to achieve accurate segmentation with limited training data (few-shot learning). The study addresses the challenge of COVID-19 diagnosis and monitoring by providing an automated method for identifying and quantifying affected regions in CT scans. The model's performance was evaluated using various metrics and compared favorably against other state-of-the-art approaches, demonstrating its potential for aiding medical professionals in COVID-19 analysis. Liu et al. (2021) introduced a two-stage cross-domain transfer learning framework for segmenting lung infections in COVID-19 patients. The proposed method leverage pre-trained models from a different domain and fine-tunes them using limited annotated data specific to COVID-19. The framework comprises an initial coarse segmentation stage followed by a refinement stage for improved accuracy. Experimental results demonstrate that the proposed method achieves robust performance in segmenting lung infections, outperforming other state-of-the-art

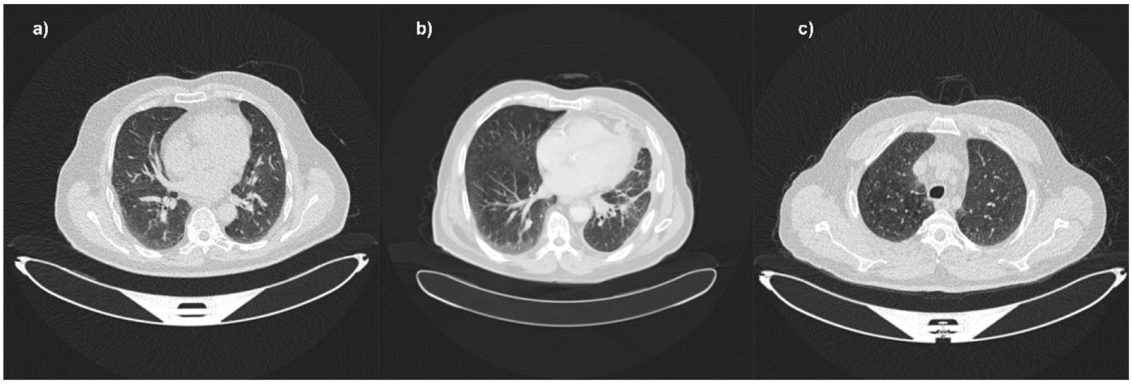

Fig. 2. Sample images in IQ-OTH/NCCD lung cancer dataset for each class. Here, (a) Benign, (b) Malignant, and (c) Normal.

approaches. The study highlights the potential of transfer learning techniques for enhancing COVID-19 image analysis and aiding in clinical decision-making.

The literature showed that the transfer learning-based methods proved to be more robust in comparison to custom-built DCNN methods. However, for the aforementioned pre-trained architectures, the accuracy increases with increasing architecture depth, which also increases the network complexity and parameters. As a result of which the efficiency of the model decreases in both training and testing. Moreover, while training the deep learning model also faces the issue of gradient vanishing and exploding problems as while going deeper the gradient of the features becomes zero which tends to be overfitting of the model. Also increasing the hidden layers in the deep model results in the model complexity and take a lot of time in its execution. The interpretation of these images can be challenging for radiologists due to the large amount of information they contain. Furthermore, the inter and intra-visual similarities between the different classes make it harder to classify the cancerous images from the benign and normal images. The primary objective of this study is to propose a transfer-learning-based method for the classification of lung cancer from the CT-scan slices using a more efficient CNN architecture that generalizes well with fewer network parameters and less system overload.

3. Material and methods

This section discusses the dataset used and the methodology adopted for the training of the proposed model for the multi-class lung cancer classification from the CT scans. The general workflow of the proposed methodology is illustrated in Fig. 1. First, the CT scan slices in the dataset are loaded then several pre-processing steps are applied to the loaded images. Since the problem is based on medical imaging, therefore, it is difficult to collect a large number of annotated data for model training. As a result, data augmentation is utilized to synthetically expand the number of training examples. The proposed methodology is based on the transfer-learning approach using different variants of EfficientNet that are fine-tuned for the multi-class classification of lung cancer into three categories i.e., benign, malignant, and normal. The details regarding the pre-processing steps, data augmentation, and the proposed model are described in the subsequent sections.

3.1. Dataset

The experiments are carried out on the "Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD)" lung cancer dataset (Alyasriy and Muayed, 2020). In 2019, the dataset was collected from the "Iraq-Oncology Teaching Hospital and National Center for Cancer Diseases" for more than three months. The dataset consists of CT scans of patients who are either healthy or diagnosed with different stages of lung cancer. The dataset was annotated by multiple oncologists and radiologists. The dataset consists of a total of 1097 images of a human chest representing CT scan slices of 110 cases varying in age, gender, education status, residence, and living status. The cases in this study were divided into three categories: benign, malignant, and normal, as depicted in Fig. 2. A total of 110 cases were analyzed, with 40 being identified as malignant, 15 as benign, and 55 as normal. The CT scans used in the study were initially obtained in DICOM format, with a resolution of 512 × 512, but were later converted to JPEG for ease of access. The IQ-OTH/NCCD dataset is publicly available in the Kaggle repository (Anon, 2023b). The breakdown of the dataset by class-wise category is presented in Table 1.

Table 1
The class-wise distribution of the "IQ-OTH/NCCD lung cancer dataset".

Class	Patients	No. of samples
Benign	15	120
Malignant	40	561
Normal	55	416
Total	110	1097

3.2. Data preprocessing

This section discusses the pre-processing steps applied to the dataset before model training and testing. All images of each class are first shuffled for the non-bias training and then split into the train–test ratio of 80:20 in such a way that 80% of the total images are sampled into the training set for model training whereas 20% of the total images are sampled in the testing set for model evaluation on unseen test instances. The class-wise distribution of the dataset after the train–test-split is summarized in Table 2. Since the original CT scan images contain unwanted regions i.e., background and noise that might cause noisy training, therefore it was removed by cropping out the extreme points of the biggest lung contour from the rest as shown in Fig. 3. After cropping,

Table 2
The class-wise distribution of the dataset after the train–test split of 80:20 ratio.

Class	Split	No. of sample	Total
Benign	Train	96	876
Malignant		448	
Normal		332	
Benign	Test	24	221
Malignant		113	
Normal		84	

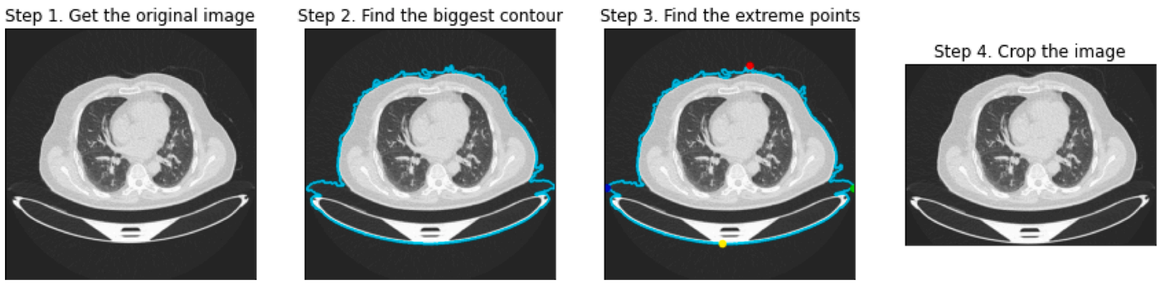

Fig. 3. The sequence of steps applied to crop unwanted regions from CT scan images of lung cancer.

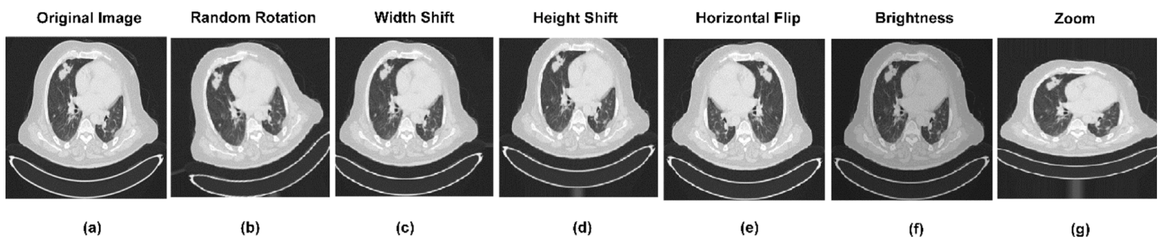

Fig. 4. Sample of original and augmented CT scan slices of lung cancer after different data augmentation technique is applied.

Table 3
The breakdown of the dataset by class-wise category after data augmentation is applied.

Class	Split	Before augmentation	After augmentation	% per class	Total
Benign	Train	96	1344	33.47%	4016
Malignant		448	1344	33.44%	
Normal		332	1328	33.07%	
Benign	Test	24	–	–	221
Malignant		113	–	–	
Normal		84	–	–	

data augmentation is applied to the cropped images. The details regarding data augmentation are discussed in the next section. Images in both the train and test set are then resized to the fixed resolution of 240 × 240 × 3 to match the shape of the input tensor with those of the input shape required by the pre-trained EfficientNet models. Image resizing also helps to reduce computational overload during model training by keeping contextual information and features in an image intact. Finally, class labels in both the train and test set are label-encoded into 0, 1, and 2 for benign, malignant, and normal respectively.

3.3. Data augmentation

The original CT scan slices of 1190 samples are not enough to train deep CNN architecture effectively; therefore, data augmentation is applied to the original samples to increase dataset samples for each class. Data augmentation also helps to increase the robustness of the model and to prevent the model from overfitting. It is a technique that is widely used to increase the number of instances of datasets automatically employing different image transformation functions such as translation, rotation, shearing, mirroring, cropping, horizontal flip, and vertical flip, etc. In this study, six data augmentation techniques i.e. horizontal flip, rotation, brightness, zooming, width shift, and height shift are applied over the dataset in such a way that images in benign, malignant, and normal classes are augmented by the factor of 13, 2, and 3 respectively to partially balance dataset samples for each class. Because the benign class contains the least number of images, therefore, it has been augmented the most. Table 3 summarizes the number of CT scan slices for each class before and after data augmentation. Here, data augmentation is applied on CT scan slices in a train set only. Fig. 4 shows some samples of augmented CT scan slices of lung cancer after the aforementioned data augmentation is applied.

3.4. Proposed architecture

This section discusses the transfer learning paradigm with the proposed fine-tuned architecture of the EfficientNetB1 in detail.

3.4.1. Transfer learning

In contrast to conventional machine learning methods, CNN enables the automatic extraction of both low-level and high-level feature maps from the model's convolutional base, pooling, and batch-normalization layers. The one-dimensional feature vector created from these extracted feature maps is then sent to a set of single or multiple fully connected layers for classification. Despite its enormous success, one of CNN's drawbacks is that it needs a lot of data samples to train the model effectively and avoid high-bias (underfitting) and high-variance (overfitting) issues. However, it is not practical to gather a significant amount of annotated data for various research challenges, particularly in the field of medical imaging. Additionally, most of the data are not even freely available. To overcome the aforementioned issue, the transfer-learning technique can be applied. Transfer-learning transfers the knowledge taken from architectures that were originally trained on a bigger benchmark dataset such as ImageNet (He et al., 2015) to problems that are either similar to or different from their original context, such as the classification of lung cancer from CT-scan slices with fewer data points. Fig. 5 illustrates the broad concept of transfer learning. None of the pre-trained CNN designs can be utilized directly for inference and expect sufficient generalizability on unseen test instances due to the difference in the domains of the source and target dataset, i.e., CT scans. Instead, to adapt to the images in the target domain, the layers of the pre-trained models are refined empirically. Instead, to adapt to the images in the target domain, the layers of the pre-trained models are fine empirically. The technique of fine-tuning involves retraining the weights taken from a few top layers of a deep CNN architecture for different specific problems. These weights were initially trained on a very large dataset. By unfreezing all or some of the layers in convolutional base layers (He et al., 2015), or by employing pre-trained architectures as

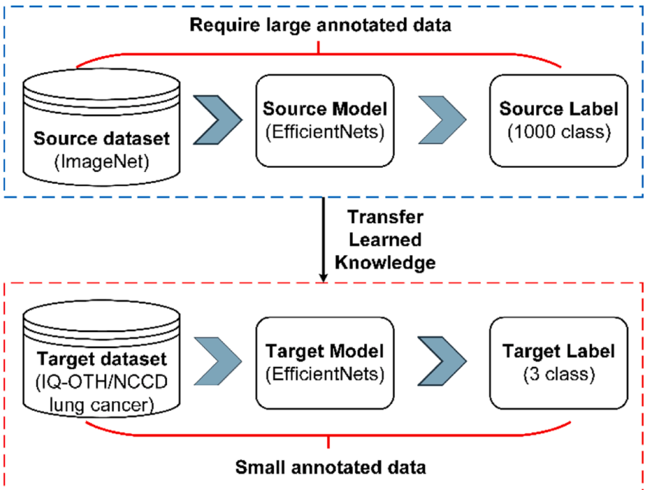

Fig. 5. The general concept of transfer learning.

fixed feature extractors that are then fed to other classifiers like SVM for classification, pre-trained architectures can be fine-tuned.

In this study, transfer learning of five variants of pre-trained EfficientNet i.e., EfficientNet B0–B4 is performed where each variant of the EfficientNet model is fine-tuned explicitly on CT-scan slices of lung cancer. The feature maps from the EfficientNet are extracted that are then passed to the fully connected layers for classification. The following section goes over the details for optimizing the classification layers of the pre-trained EfficientNet architecture.

3.4.2. *Classification using fine-tuned EfficientNet*

The deep CNN-based architectures are usually overparameterized, due to the increasing number of convolutional layers, network depth, and width as a result of which, the efficiency of the network is compromised and makes an architecture computationally expensive. There is a tradeoff between network efficiency and accuracy. Deep networks might generalize well on test data but their efficiency in terms of inference speed, floating-point operations per second (flops), network parameters, and model size increases. In 2019, the Google AI research team proposed a family of EfficientNet series, namely EfficientNetB0–EfficienNetB7 (Tan and Le, 2019), as a backbone architecture that has outperformed many state-of-the-art deep CNN-based architectures such as Inception-V3, ResNet50, Inception-ResNetV2, DenseNet for image classification from ImageNet, segmentation, and other transfer learning based problems. In contrast to conventional scaling methods utilized by previous studies in their proposed architecture, which include arbitrarily increasing network width, depth, and resolution to increase the generalizability of the network. The CNN architecture is structurally scaled up by the EfficientNet utilizing fixed sets of scaling coefficients using a uniform compound scaling approach. Compound scaling is based on the idea of balancing the dimension of depth d, width w, and resolution r of a network by scaling it with a constant ratio. Mathematically equation is given in (1).

$$\mathrm{d} = \alpha\varnothing, \mathrm{w} = \beta\varnothing, \mathrm{r} = \gamma\varnothing \tag{1}$$

Such that $\alpha.\beta 2.\gamma 2 \approx 2$ where, $\alpha \geq 1$, $\beta \geq 1$, $\gamma \geq 1$. The values of α, β, and γ are determined by a grid search algorithm. $\varnothing$ is a user-defined parameter that determines the increase in the computational resource to the network. Flops of the convolutional operation of the network are directly proportional to d, w2, and r2 in such a way that flops will double if the depth of the network is doubled. On the other hand, if the network width and resolution are doubled, flops will increase to four times. However, scaling the network with respect to Eq. (1), the increase in flops is according to the relation $(\alpha.\beta 2.\gamma 2)\varnothing$ in such a way that for any new value $\varnothing$, the total flops are increased by $2\varnothing$.

The architecture of all EfficientNet consists of a stem block, followed by 7 blocks, and a final layer as shown in Figs. 6, and 7.

As seen in Fig. 8, each block in EfficientNet has a variable number of modules, and as one proceeds from EfficientNetB0 to EfficientNetB7, the number of modules increases. Each variant of EfficientNet has variable depth and parameters. The simplest version of EfficientNet i.e. EfficientNetB0, has 237 layers and 5.3M parameters, while EfficientNetB7 has 813 layers and 66M parameters. EfficientNet architecture makes use of mobile inverted bottleneck convolution (MBConv) layers, just like MobileNetV2 and MnasNet. Since the normalization layer is already present in the stem layer of EfficientNet therefore, no additional image normalization is required as a pre-processing step hence, it takes an input image with pixel intensity values in the range of 0–255.

In this research, five variants of pre-trained EfficientNet i.e., EfficientNet B0–B4 are utilized as a backbone for the classification of lung cancer from the CT-scan slices. The criteria for selecting the best EfficientNet variations depend on several variables, including the dataset size, the resources available for model training and evaluation, the depth of the model, the network parameters, and the batch size. EfficientNetB5 through EfficientNetB7 are larger variations of EffeicientNet that have a deeper network and more parameters. Because of this, the model may overfit the training set, and larger computational resources (GPU + RAM) are needed for the model training. EfficientNetB0–EfficientNetB4 is exclusively used as a backbone for the classification of lung cancer for the aforementioned reasons.

This research involves the transfer learning of five pre-trained EfficientNet variations, EfficientNetB0 through EfficientNetB4, which were initially trained on the ImageNet benchmark dataset. The CT scan slices of lung cancer are specifically used to fine-tune these models. Fig. 9 depicts the network architecture of the modified EfficientNetB1. By first initializing the base model with the ImageNet weights as a backbone, the pre-trained EfficientNet is fine-tuned. To reduce dimensionality, the Global Average Pooling (GAP) layer is inserted on top of the EfficientNet backbone while keeping the weights in the convolutional base of each block fixed. Without affecting the model's performance in terms of accuracy, the GAP layer also aids in simplifying the network in terms of a number of parameters. A dropout layer with a 0.5 probability is added to the network after the GAP layer. A regularizer called dropout aids in keeping the model from overfitting. Since the dataset has three class labels, the initial output layer with 1000 units is replaced by an output layer with 3 units with a Softmax activation layer. The entire

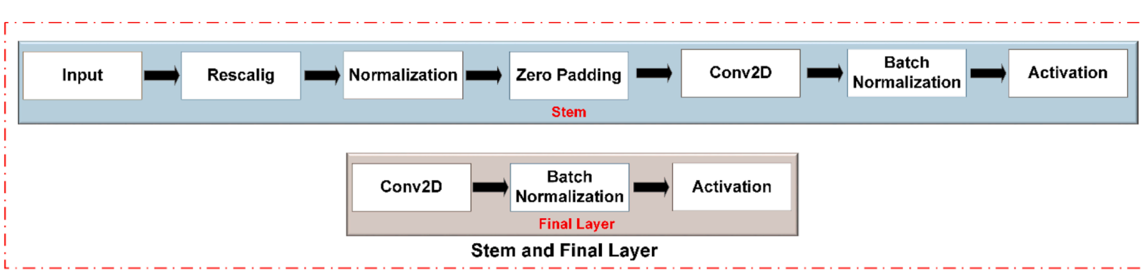

Fig. 6. Stem and Final layer in EfficientNet.

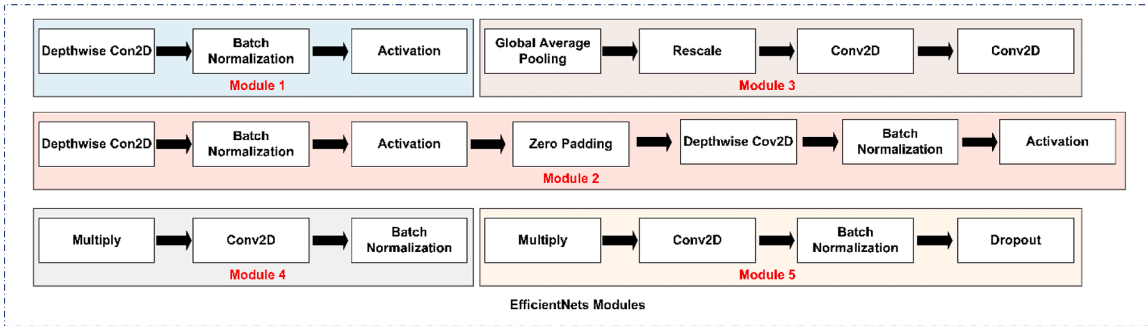

Fig. 7. Five modules in EfficientNet. Each module has a different type of layer based on the nature of the required feature maps.

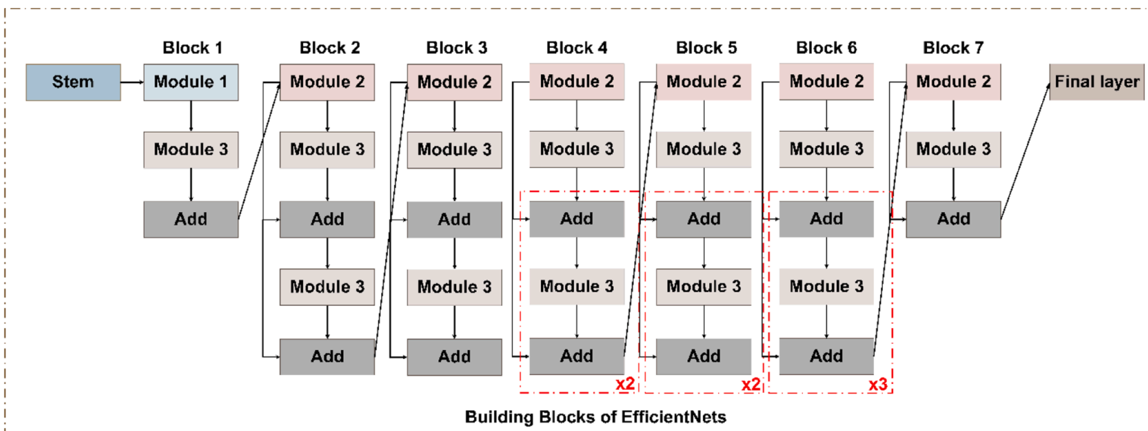

Fig. 8. The general architecture of EfficientNetB1. Each EfficientNet has seven blocks with a variable number of modules in each block.

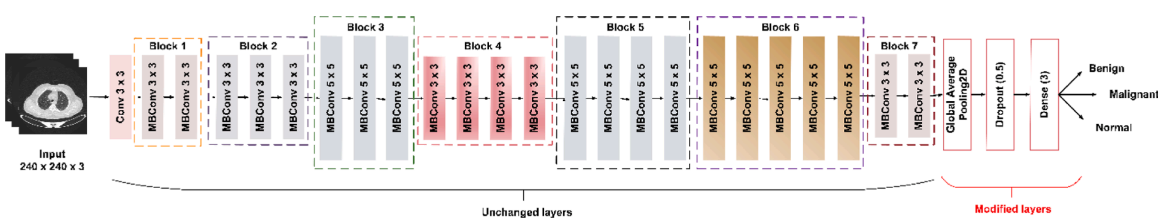

Fig. 9. The proposed method of fine-tuning pre-trained EfficientNetB1. The classification layer of the pre-trained EfficientNet is modified by the addition of GAP, dropout, and output layers. The layers in each EfficientNet block are kept unchanged.

architecture is re-trained on the IQ-OTH/NCCD lung cancer dataset.

4. Results and discussion

This section provides a comprehensive discussion of the evaluation measures employed to assess the performance of the proposed method. Additionally, it covers the system and software requirements necessary for model training and evaluation. Detailed information regarding the various hyper-parameters and their corresponding values is also provided. Furthermore, a thorough analysis of the results obtained through the proposed method is presented in this section.

4.1. Evaluation measures

Evaluation measures are quantitative metrics that are used to evaluate the performance of a deep learning model. They are used to compare the performance of different models or algorithms on a particular task, assess the effectiveness of a model or algorithm in solving a particular problem, and identify areas for improvement. The evaluation measures that are being used in this research work are accuracy, F1-Score, precision, recall/sensitivity, ROC curve, and confusion matrix (Ge et al., 2022, 2023).

Accuracy: The accuracy of a diagnostic test or classifier is calculated as the number of true positive and true negative results divided by the total number of results. It can be expressed by Eq. (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where, TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative respectively.

Precision: Precision is the proportion of correct positive results, and is calculated by Eq. (3).

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Recall: Recall is the proportion of actual positive cases that are correctly identified, and is by Eq. (4).

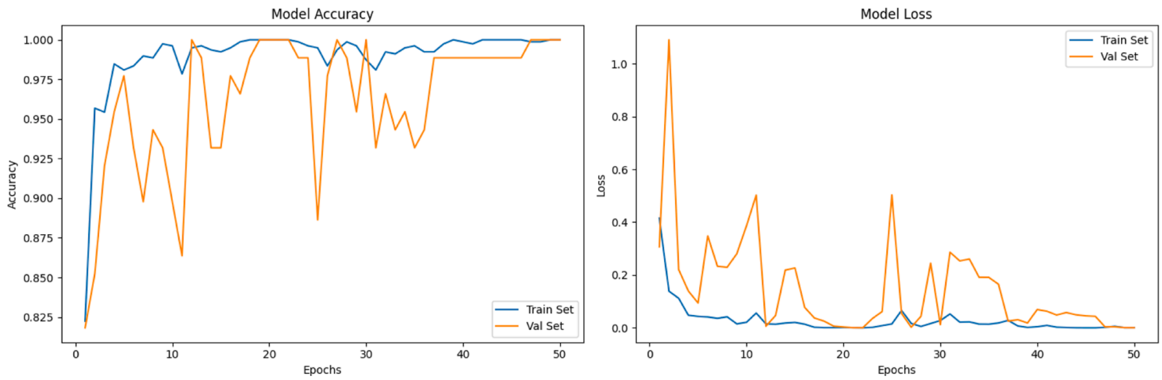

Fig. 10. Model training and validation accuracy and loss curves without data augmentation.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

F1 score: The F1 score is a balanced harmonic mean of precision and recall, and is calculated by Eq. (5).

$$F1 - Score = \frac{2{*}Precision{*}Recall}{Precision + Recall} \tag{5}$$

where precision is the proportion of correct positive results, and recall is the proportion of actual positive cases that are correctly identified.

Receiver Operating Characteristic (ROC) curve: a graph that represents performance at each threshold of a classification model. ROC represents two parameters. i.e., True Positive rate (TPR) and False Positive rate (FPR) that is given by Eqs. (6) and (7).

$$TPR = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

$$FPR = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{7}$$

Confusion matrix: A confusion matrix is a summarized form of predicted results that helps to measure different evaluation measures like accuracy, recall, precision, etc. It summarizes the overall performance of the model by breaking down the number of correct and incorrect predictions for each class.

4.2. Experimental setup

The proposed fine-tuned EfficientNetB1 model is implemented on the Google Colab Pro framework which allows for faster training and evaluation of the model, which could be useful for developing and testing the proposed model. The experimental setup used in this research work is as follows: This research employs the Python programming language, the Keras library, and TensorFlow as a backend to train the model. The experiments are performed using a Tesla T4 GPU on Google Colab Pro, which has 25 GB of RAM.

4.3. Hyper-parameters settings

In order to achieve optimal performance in model training and obtain the desired results for lung cancer classification, various hyper-parameters were fine-tuned through empirical experimentation. These hyper-parameters include batch size, optimizers, learning rate, epochs, and loss function.

Given that lung cancer classification involves distinguishing between normal, benign, and malignant cases, categorical cross-entropy was selected as the appropriate loss function. The Adam optimizer (Kingma and Ba, 2014) with an initial learning rate of 0.001 (1e−2), was utilized for all five EfficientNets models. To optimize the learning rate further, a decay factor of 0.3 was applied after evaluating the model's validation accuracy at every 5 iterations. To introduce additional regularization during fine-tuning without affecting the ImageNet weights, a drop connect rate of 0.2 was set. This rate serves as a trigger for regularization techniques. During training, the training set images were loaded with a mini-batch size of 32. The fine-tuned EfficientNets underwent training for a total of 50 epochs. For each epoch, 10% of the images from the training set were randomly separated to form a validation set. This validation set was used to assess the performance of the trained model and determine if any overfitting had occurred. All selected variants of EfficientNets (B0–B4) were trained and evaluated under consistent experimental and hyper-parameter settings to ensure a fair comparison. The optimized values of the hyper-parameters used throughout the experiments are summarized in Table 4. These combinations of the hyperparameter are finalized after fine-tuning and conducting several experiments.

Table 4
List of hype-parameters and their corresponding values.

Hyper-parameters	Values
Input shape	(240, 240, 3)
Drop connect rate	0.2
Output layer activation function	Softmax
Epoch	50
Batch size	32
Optimizer	Adam
Initial learning rate	0.001
Learning rate decay factor	0.3
Patience	5
Validation split	0.1
Loss function	Categorical cross-entropy

4.4. Analysis of results

In this study, a deep learning approach is proposed for the classification of lung cancer using pre-trained EfficientNet on CT scan images. The dataset without data augmentation consists of 1097 CT scan images taken from 110 patients of which 120 were benign, 561 were malignant

Table 5
Results of fine-tuned EfficientNetB1 without data augmentation.

	Precision	Recall/Sensitivity	F1-Score
Benign	95.65	91.67	93.62
Malignant	100.00	100.00	100.00
Normal	97.65	98.81	98.22
Accuracy			98.64
macro avg	97.77	96.83	97.28
weighted avg	98.63	98.64	98.63

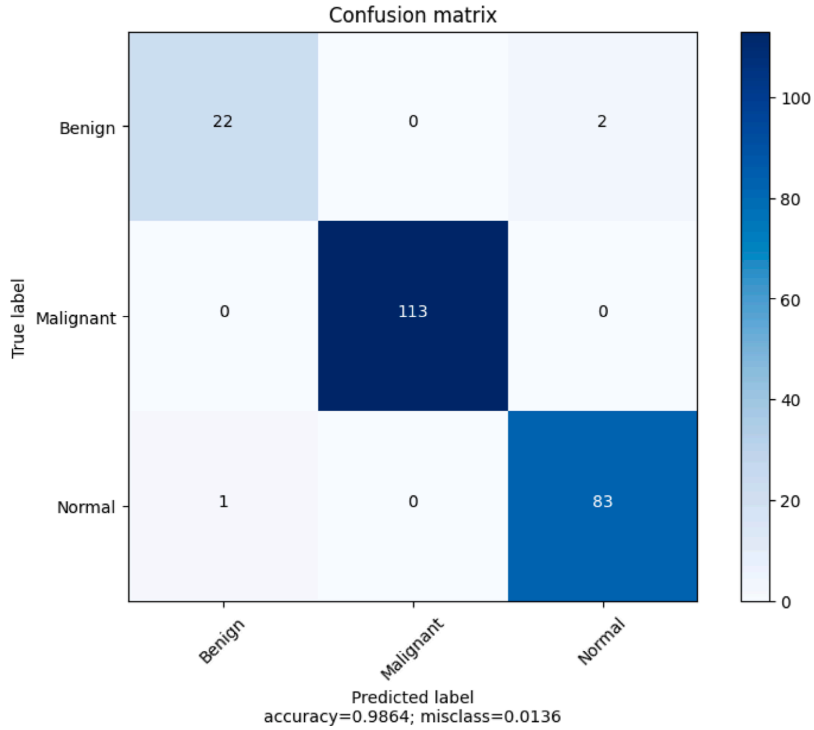

Fig. 11. Confusion matrix without data augmentation dataset.

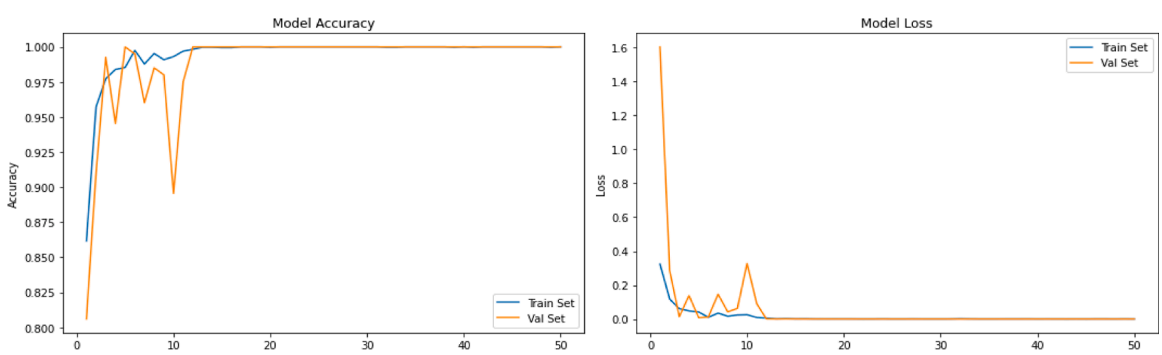

Fig. 12. Model training and validation accuracy and loss curves with Data Augmentation.

and the remaining 416 were normal classes. The experiments were conducted on both without and with data augmentation dataset and the split ratio were kept the same in both cases. All the CT scan images were pre-processed before passing through training purposes. The proposed model obtained 98.64% test accuracy with a precision of 98.63%. The detailed results obtained through the Lung-EffNet model without data augmentation are given in Table 5.

The curves of the model accuracy and the loss function without data augmentation can be seen in Fig. 10.

A confusion matrix is used to assess how well a classification model is generalizing on unseen test data. For each class, the confusion matrix highlights the number of the predicted labels in the horizontal x-axis with its true label in the vertical y-axis. This is done by first comparing labels obtained from the model prediction with the true labels and then counting the number of times the correct combination of predicted and true value occurs. The performance of the proposed technique in terms of the confusion matrix without data augmentation is shown in Fig. 11.

The dataset with data augmentation used in this study consisted of 4236 images of lung CT scans taken from the "IQ-OTH/NCCD dataset", of which 1368 were benign, 1456 were malignant and 1412 were normal. The images were pre-processed before being fed into the pre-trained EfficientNet model. The model was fine-tuned using a training dataset and tested on a separate test dataset. The results obtained from the model showed an overall test accuracy of 99.10% with a precision of 100% and a ROC score of 0.97 to 0.99. The EfficientNet models were developed using a compound scaling method, which involves scaling the network's depth, width, and resolution in a way that maintains a consistent level of accuracy while reducing the number of parameters and computational complexity. The model used 6.5 million parameters and perform well on the test dataset. The detailed result of the model with data augmentation is given in Table 6.

Table 6
Results of fine-tuned EfficientNetB1 with data augmentation.

	Precision	Recall/Sensitivity	F1-Score
Benign	100.00	91.67	95.65
Malignant	100.00	100.00	100.00
Normal	97.67	100.00	98.82
Accuracy			99.10
macro avg	99.22	97.22	98.16
weighted avg	99.12	99.10	99.08

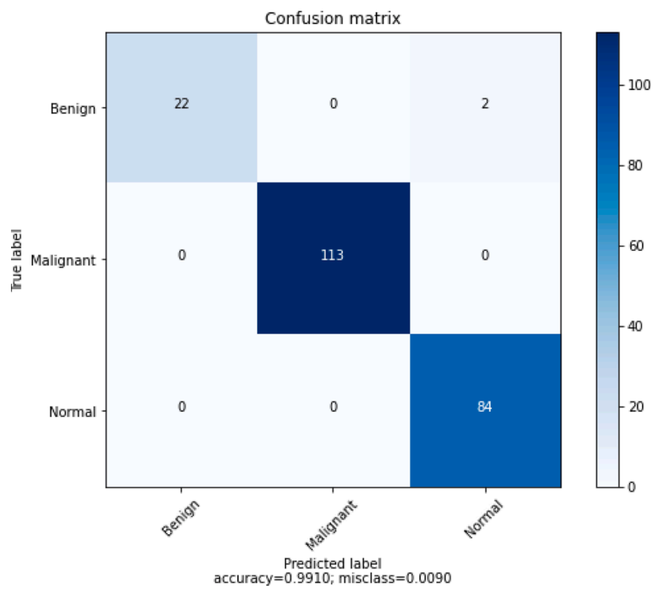

Fig. 13. Confusion matrix with data augmentation dataset.

The curves of the model accuracy and the loss function with data augmentation can be seen in Fig. 12.

The performance of the proposed technique in terms of the confusion matrix with data augmentation is shown in Fig. 13.

A ROC curve plots the TPR and FPR at various classification thresholds, allowing us to see how the classifier's performance changes as the threshold is varied. A curve that is closer to the top-left corner represents a classifier with better performance. The ROC curve can be seen in Fig. 14.

Gradient-weighted Class Activation Mapping (GradCAM) visualization technique is also used to indicate or visualize the regions in an image that are most important for a CNN to classify a given image. The heatmap is generated by taking the gradient of the class score with respect to the feature map of a convolutional layer and weighing the channels in the feature map by the gradient magnitude. The resulting heatmap is then up-sampled and overlaid on the input image to highlight the regions that are most important for the CNN classification of the image. The GradCAM visualization for each class is shown in Fig. 15.

4.5. Comparative analysis of Lung-EffNet with previous state-of-the-art approaches

EfficientNetB1 is one of the models in this family and has been demonstrated to produce outstanding results on image classification benchmark datasets. The outcomes attained by this study indicate that using pre-trained EfficientNet is a promising approach for predicting lung cancer on CT scan images. The model achieved impressive results on image classification problems because of the pre-processing steps employed on the lung images and fine-tuning of the EfficientNetB1 model. The dataset consists of three classes i.e., Benign, Malignant, and Normal. Due to a small amount of dataset data augmentation and transfer learning is being used in this research work. The overall accuracy of 99.10% and the precision of 100% are particularly noteworthy, as they suggest that the model is able to accurately detect cancerous lung images from benign and normal. Additionally, the specificity of 100% implies that the model is able to correctly identify normal lung images. The ROC curve of 0.97 to 0.99 further supports the validity of the proposed approach, indicating that the model can differentiate between cancerous and normal images with a high degree of accuracy. It is mandatory to mention that the proposed approach is not only limited to lung cancer but can also be applied to another medical imaging-based diagnoses.

By comparing with other existing models, EfficientNet-B1 has been shown to achieve superior performance in lung cancer classification tasks. One of the reasons for EfficientNetB1's superior performance is its ability to extract features from the input images. EfficientNetB1 uses a combination of convolutional and pooling layers to extract features from the images, with each layer building on the features extracted by the previous layer. This allows the model to learn complex representations of the images, which is crucial for accurate classification. The proposed architecture has been demonstrated to be highly efficient and generalized than conventional ML algorithms such as SVM and RF, and other DL-based CNN architectures such as VGG16, ResNet50, and DenseNet121. The proposed method is compared with state-of-the-art models in terms of accuracy and it was making sure that the train–test split ratio and the dataset were the same for the fair comparison. But due to the limited work on the same dataset and some of the studies used private datasets, we conduct a comparison with them by making sure that the train–test split ratio is the same i.e., 80:20. In Table 7 we presented a comprehensive analysis of the Lung-EffNet protocol with state-of-the-art approaches.

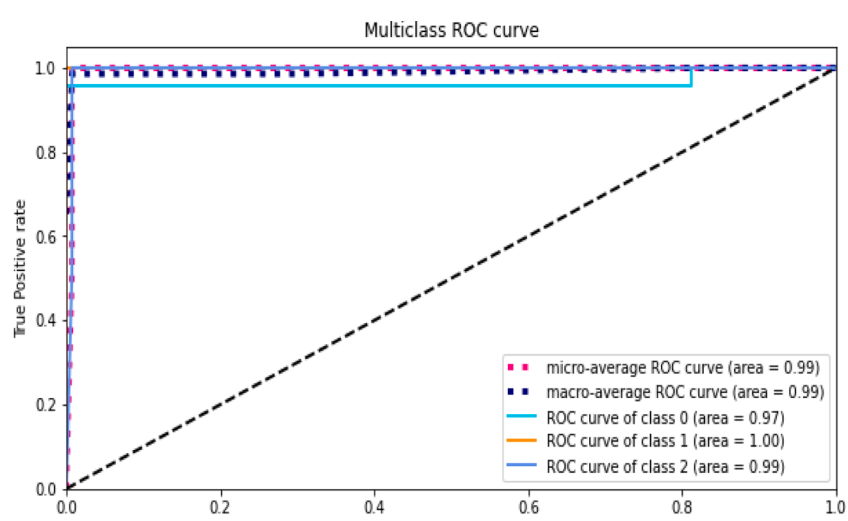

Fig. 14. ROC curve on test data.

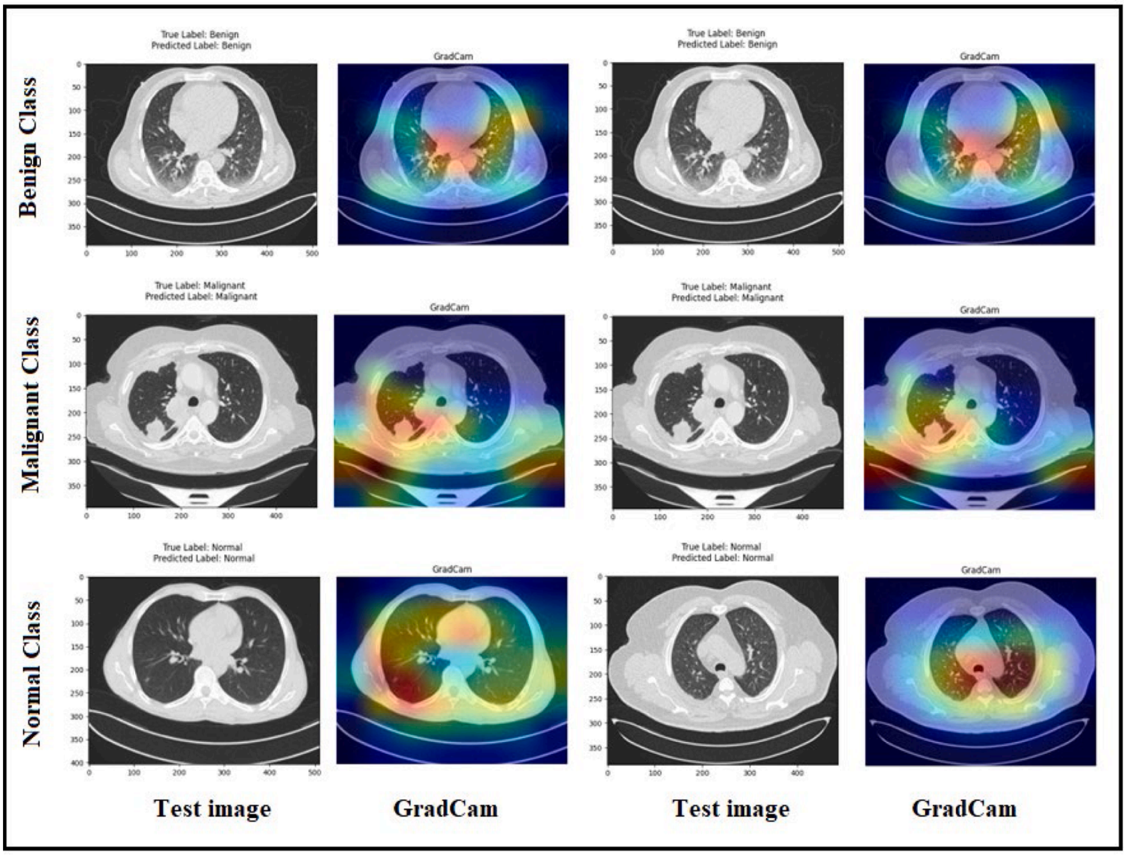

Fig. 15. GradCam visualization of class-wise results.

Table 7
Performance comparison of the Lung-EffNet with existing methods over IQ-OTH/NCCD dataset.

Reference	Methodology	Dataset	Accuracy
Hamdalla et al. (Al-Yasriy et al., 2020a)	AlexNet with CNN	CT scan dataset	93.54%
Chen et al. (Chen et al., 2021a)	CNN with embedded NLP	CT scan images	88.00%
AL-Huseiny et al. (Kareem et al., 2021)	GoogLeNet DNN	IQ-OTH/NCCD	94.38%
Narin and Onur, 2022	AlexNet and ResNet	IQ-OTH/NCCD	98.58% accuracy on AlexNet
Lei Lyu et al. (Lyu, 2021)	Ensemble of DenseNet. DCNN, VGG, and AlexNet	CT scan images	99.00%
Abida et al. (Kavitha and Prabakaran, 2019)	Transfer learning based VGG-16 and Inception V3	CT scan images	99.8% accuracy on static data 75% accuracy on dynamic data
Mamoona et al. (Humayun et al., 2022a)	Ensemble approach TSCNN	CT scan images taken from LUNA dataset	98.83% on VGG16, 98.05% on VGG19, and 97.4% on Xception
Proposed Model	**Transfer learning of Fine-tuned EfficientNetB1**	**IQ-OTH/NCCD**	**99.10%**

5. Ablation study

In order to assess the generalizability and robustness of the proposed Lung-Effnet for the classification of lung cancer, we conducted an ablation study encompassing several key parameters. This study aimed to investigate the impact of different factors on the performance of our model and gain insights into its behavior under various conditions. In this section, we present the findings of our ablation study, which include the analysis of computational complexity, the influence of data augmentation techniques, the effect of different data splits, the comparison of models with and without transfer learning, and the training of the proposed model for the classification of different lung cancer types. Through these sub-sections, we provide a comprehensive evaluation of the proposed model and its adaptability in diverse scenarios, shedding light on its strengths, limitations, and potential applications in the field of lung cancer diagnosis and classification.

5.1. Computational complexity

The comprehensive analysis of the computational complexity of the proposed fine-tuned EfficientNet models (B0 to B4) is determined. The evaluation considers various factors, including the number of network parameters, floating-point operations (FLOPS), network size, training time, inference time, and test accuracy. Table 8 provides a summary of the computational cost of the proposed fine-tuned EfficientNets and its variants along with other deep learning architectures that are trained and evaluated on the same dataset under the same hyper-parameter settings. Starting with EfficientNetB0, it is observed that this model has a relatively simpler structure, consisting of 4.01 million parameters. As we move towards higher-order models, such as EfficientNetB1, the number of parameters increases to 6.51 million. Further, EfficientNetB2, B3, and B4 exhibit even greater complexity, with 7.70 million, 10.70 million, and 17.55 million parameters, respectively. It is important to note that its depth, specifically the number of convolutional layers

Table 8
The computational complexity of the proposed fine-tuned EfficientNets.

Model	FLOPs (G)	No. of parameters	Training time (hh:mm:ss)	Testing time (hh:mm:ss)	Model size (MB)	Test accuracy (%)
EfficientNetB0	0.95	4,011,391	00:03:20	00:00:02	47.2	93.67
EfficientNetB1	1.42	6,517,027	00:46:40	00:00:01	76.4	**99.10**
EfficientNetB2	1.64	7,705,221	00:50:50	00:00:02	90.0	97.74
EfficientNetB3	2.38	10,700,843	00:65:20	00:00:02	124.6	98.19
EfficientNetB4	3.71	17,553,995	01:05:50	00:00:04	203.6	97.29
ResNet50	9.13	23,534,592	00:38:02	00:00:03	–	94.40
MobileNet	1.59	3,210,051	00:20:50	00:00:01	37.3	98.28
MobileNetV2	2.72	2,227,715	00:23:20	00:00:01	26.5	98.28
MobileNetV3Small	1.72	928,739	00:00:00	00:00:01	11.7	97.84
Attention-InceptionResNet-V2	15.0	47,476,851	01:10:50	00:00:07	546.7	97.41

Table 9
The influence of fine-tuned EfficientNets with and without data augmentation.

Model	Without Data Augmentation				With Data Augmentation			
	Test accuracy	Precision	Recall	F1- score	Test accuracy	Precision	Recall	F1-score
EfficientNet-B0	92.62	86.15	88.43	87.13	93.67	87.20	89.48	88.18
EfficientNet-B1	98.64	97.77	96.83	97.28	99.10	99.22	97.22	98.16
EfficientNet-B2	96.83	95.54	95.24	95.37	99.10	99.22	97.22	98.16
EfficientNet-B3	98.19	98.48	94.44	96.19	98.19	98.48	94.44	96.19
EfficientNet-B4	96.74	95.89	92.11	93.74	97.29	96.44	92.66	94.29

Table 10
Results of proposed fine-tuned EfficientNets with different data split.

Model	70:30 Split				80:20 Split				90:10 Split			
	Test accuracy	Precision	Recall	F1- score	Test accuracy	Precision	Recall	F1-score	Test accuracy	Precision	Recall	F1-score
EfficientNet-B0	**97.58**	96.37	93.91	95.03	93.67	87.20	89.48	88.18	94.59	95.83	85.53	88.71
EfficientNet-B1	98.48	97.95	96.03	96.93	**99.10**	99.22	97.22	98.16	**98.20**	98.48	94.44	96.19
EfficientNet-B2	96.83	95.54	95.24	95.37	**99.10**	99.22	97.22	98.16	94.59	91.91	94.09	92.87
EfficientNet-B3	96.67	95.43	91.13	92.93	**98.19**	98.48	94.44	96.19	98.10	98.10	98.10	98.10
EfficientNet-B4	96.06	95.73	88.62	91.24	**97.29**	96.44	92.66	94.29	96.4	97.1	88.9	91.82

incorporated, influences the number of parameters in a deep neural network architecture. In this case, as the model depth increases from B0 to B4, the number of parameters also increases, indicating a more complex architecture.

5.2. The influence of data augmentation techniques

In this section, the effectiveness of data augmentation techniques is evaluated by comparing the results of the fine-tuned EfficientNet B0–B4 using both augmented and non-augmented data. The respective results

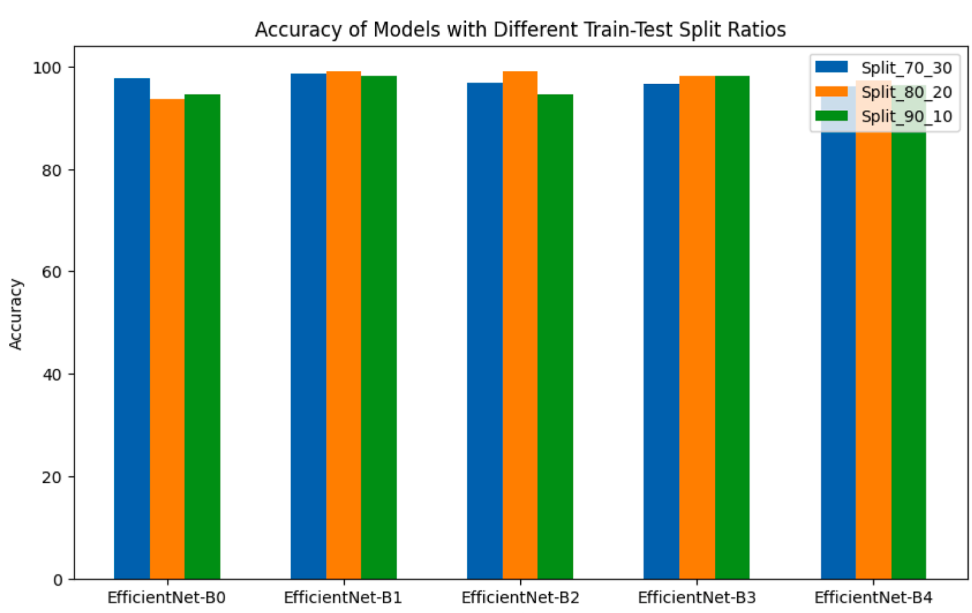

Fig. 16. The Performance comparison of the proposed fine-tuned EfficientNets with different data split.

Table 11
Ablation study results with and without transfer learning.

Model	Without transfer learning				With transfer learning			
	Test accuracy	Precision	Recall	F1- score	Test accuracy	Precision	Recall	F1-score
EfficientNet-B0	70.59	58.95	57.80	57.76	93.67	87.20	89.48	88.18
EfficientNet-B1	79.19	65.35	65.19	64.95	99.10	99.22	97.22	98.16
EfficientNet-B2	77.38	65.47	59.89	57.20	99.10	99.22	97.22	98.16
EfficientNet-B3	73.30	61.97	60.97	60.83	98.19	98.48	94.44	96.19
EfficientNet-B4	81.45	66.98	67.17	67.04	97.29	96.44	92.66	94.29

are provided in Table 9.

It can be seen that the results of fine-tuned EfficientNet B0–B4 are relatively better with data augmentation. By introducing variations in instances in the train set, the model becomes more resilient to changes in image orientation, position, and scale, making it better equipped to handle different instances of lung cancer images that may exhibit variations in shape, size, or orientation. Data augmentation also helps to alleviate the issue of limited training data by artificially expanding the dataset, effectively increasing the number of available samples for training. This helps to mitigate overfitting, where the model becomes too specialized to the training data and fails to generalize well to unseen data. Overall, data augmentation serves as a regularization technique that improves the generalizability of the deep CNN by exposing it to a wider range of variations and ensuring that it can effectively classify lung cancer images under different conditions and variations commonly encountered in real-world scenarios.

5.3. The effect of different data split

Several experiments have been conducted to explore the impact of different data splits on the performance of the proposed fine-tuned EfficientNets. We evaluated the model's performance with varying proportions of training and testing sets. We have carefully analyzed and compared the results obtained from different data splits, including the originally proposed split of 80:20, as well as splits of 70:30 and 90:10 as shown in Table 10. The histogram, that highlight the performance comparison of the proposed fine-tuned EfficientNet B0–B4 with varying data split is shown in Fig. 16. It has been observed that the model's performance was influenced by the data split; indicating that the training set size plays a crucial role in the effectiveness of the transfer learning approach.

Based on our experiments, we found that the model's performance is comparatively better than the originally proposed split of 80:20 compared to the other splits. This observation suggests that using a smaller training set, as facilitated by transfer learning, can indeed yield favorable results in terms of classification accuracy and overall performance. It has also been observed that by allocating a significant portion (80%) of the dataset to the training set, the model is provided with an ample amount of data for learning and generalization. This approach helps prevent potential overfitting, where the model might memorize the training instances instead of learning meaningful patterns.

5.4. The comparison of models with and without transfer learning

The proposed methodology is based on a transfer learning approach that used different pre-trained models using medical imaging datasets. While conducting the experiments the pre-trained model was trained using ImageNet weights. In the ablation study a number of experiments were also conducted to analyze the impact of using the transfer learning on the specific type of dataset. So, for this purpose, the experiments were carried out with or without using (i.e., ImageNet weight was set to None while training of pre-trained models that is mean use only the architecture of pre-trained models) the transfer learning. The results achieved through this ablation study are summarized in Table 11.

Without transfer learning, the following observations have been made:

- The model tends to overfit during training and testing.
- The model is biased towards the Malignant class as it contains more images in the test set.
- The model is unable to differentiate between the Benign and Normal classes giving poor generalizability on the Benign class.
- As the model gets deeper, the generalizability of the EfficientNet gets better. Test accuracy, precision, recall, and F1-score of EffcientNetB4 are comparatively better than the previous versions.

However, the pre-trained EfficientNets, with their deep architectures, have demonstrated strong feature extraction capabilities. They are trained on large-scale natural image datasets, such as ImageNet, which enables them to learn general visual representations. These learned features can be highly useful for tasks like lung cancer classification, as they capture relevant patterns and discriminative information.

- Transfer learning with EfficientNets has been shown to enhance the generalization performance of models. By leveraging the learned representations from natural images, the models can capture robust and transferable features that are applicable to medical imaging tasks. This can lead to better performance and improved accuracy in lung cancer classification, especially when working with limited medical imaging data.
- Training deep neural networks from scratch on large medical imaging datasets can be computationally expensive and time-consuming. By using transfer learning with EfficientNets, we can significantly reduce the training time and computational resources required, as we are building upon pre-trained models. This enables researchers and practitioners to iterate and experiment with different architectures and hyperparameters more efficiently.

Overall, transfer learning with EfficientNets offers a powerful approach for leveraging pre-trained models and exploiting their strengths in lung cancer classification. It allows us to capitalize on the knowledge acquired from natural image datasets and apply it effectively to medical imaging tasks, ultimately improving the accuracy and efficiency of our model.

5.5. Classification of different lung cancer types

To further analyze and highlight the robustness of the proposed fine-tuned EfficientNetB1, additional experiments are also carried out which aim to classify different types of lung cancer from CT scan images. The experiments are carried out on the "Chest CT-Scan images Dataset" taken from Kaggle (Anon, 2023a). The dataset is a collection of CT scan images of the chest. The dataset aims to facilitate research and development in the field of medical imaging analysis, particularly in the context of chest-related disease. The dataset was compiled and publically shared by Mohammad Hany, an AI enthusiast, and researcher. The dataset consists of 1000 CT scan images of three cancer types i.e. adenocarcinoma, large cell carcinoma, and squamous cell carcinoma in

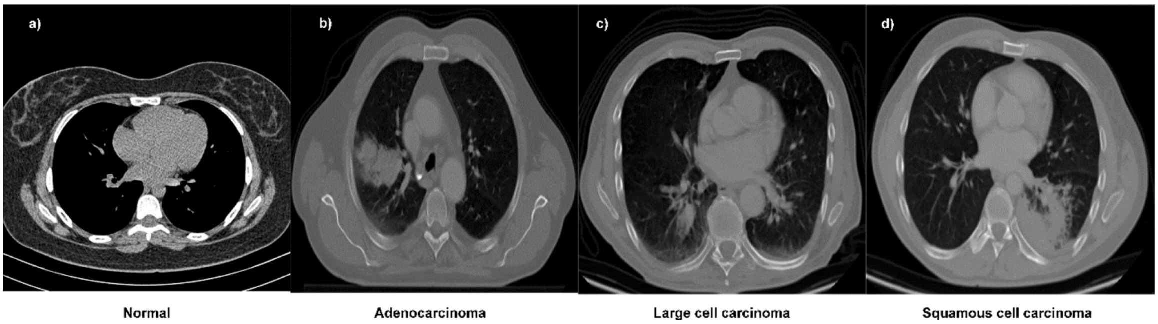

Fig. 17. Sample CT-scan images in the "Chest CT-scan image dataset" of each class. Here, (a) Normal, (b) Adenocarcinoma (c) Large cell carcinoma, and (d) squamous cell carcinoma.

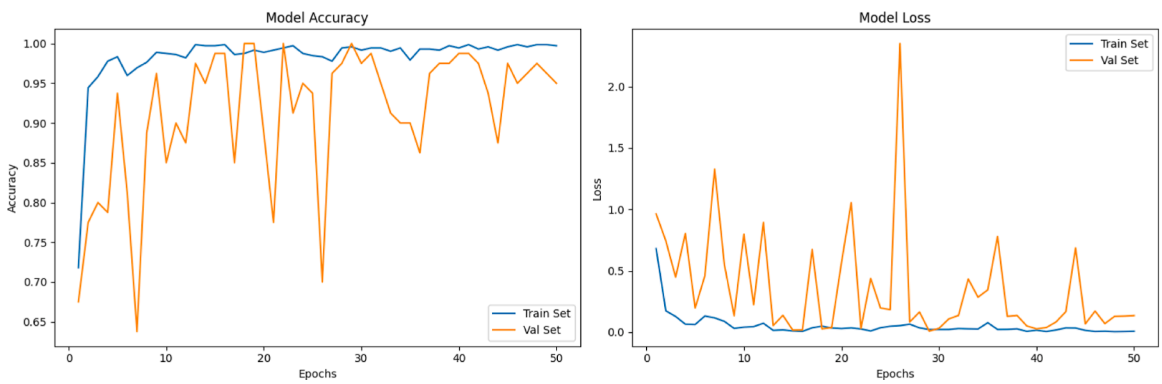

Fig. 18. Training and validation accuracy-loss curve of the proposed model.

Table 12
The class-wise distribution of "Chest CT-scan image dataset".

Classes	Split	#Sample	Total
Normal	Train (70%)	148	613
Adenocarcinoma		195	
Large cell carcinoma		115	
Squamous cell carcinoma		155	
Normal	Validation (10%)	13	72
Adenocarcinoma		23	
Large cell carcinoma		21	
Squamous cell carcinoma		15	
Normal	Test (20%)	54	315
Adenocarcinoma		120	
Large cell carcinoma		51	
Squamous cell carcinoma		90	

Portable Network Graphics (.png) format as shown in Fig. 17. CT-scan images of the normal chest are also included with the CT scans of the affected chest to accurately classify and differentiate normal lung nodules from infected ones. The dataset is divided into three sets i.e. train, test, and validation. The class-wise distribution of the dataset in the train, validation, and test set is summarized in Table 12.

The dataset undergoes several pre-processing steps before model training and testing. CT-scan slices in the train, validation, and test set are first resized to the fixed resolution of 240 × 240 × 3 to match the shape of the input tensor with those of the input shape required by the pre-trained EfficientNet models. Data augmentation is not applied in this case. Finally, class labels are label-encoded into 0, 1, 2, and 3 for normal, adenocarcinoma, large cell carcinoma, and squamous cell carcinoma respectively.

In this case, the proposed fine-tuned EfficientNetB1 is re-trained on four classes, therefore the units in the output layer are set to 4 with a *softmax activation* function. The model is trained for 50 epochs using the

Table 13
The class-wise evaluation of the proposed fine-tuned EfficientNetB1 on test data for the classification of lung cancer types.

Classes	Precision	Recall/ Sensitivity	F1-score	Test accuracy
Normal	100.00	98.25	99.12	98.24
Adenocarcinoma	86.84	98.51	92.31	98.50
Large cell carcinoma	93.10	79.41	85.71	79.41
Squamous cell carcinoma	100.00	92.86	96.30	92.85

Adam optimizer with an initial learning rate of 0.001 (1e−2). The initial learning rate is decayed with a decay factor of 0.3 after visualizing the model's validation accuracy after every 5 iterations. The drop connects rate is set to 0.2 which acts as a trigger for additional regularization during fine-tuning without affecting ImageNet weights. All images in the train set are loaded with a mini-batch size of 32.

The proposed fine-tuned EfficientNetB1 obtained an overall accuracy of 99.00%, and 94.00% on the train set and test set respectively. Table 13 summarizes the prediction of the proposed fine-tuned EfficientNetB1 (Lung-EffNet) on unseen test data for each category in terms of precision, recall/sensitivity, F1-score, and test accuracy. These results showed that the proposed methodology of fine-tuning the pre-trained EfficientNetB1 showed good performance for multi-class classification of lung cancer types and significant performance in terms of all evaluation metrics. The proposed model obtained relatively lower recall and test accuracy on large cell carcinoma due to its complex nature.

Fig. 18 depicts the curves illustrating the training and validation accuracy as well as the training and validation loss for the proposed model. As the number of epochs increased, there was a noticeable improvement in both training and validation accuracy. Simultaneously, the training and validation loss exhibited a decreasing trend.

The performance evaluation of the proposed fine-tuned

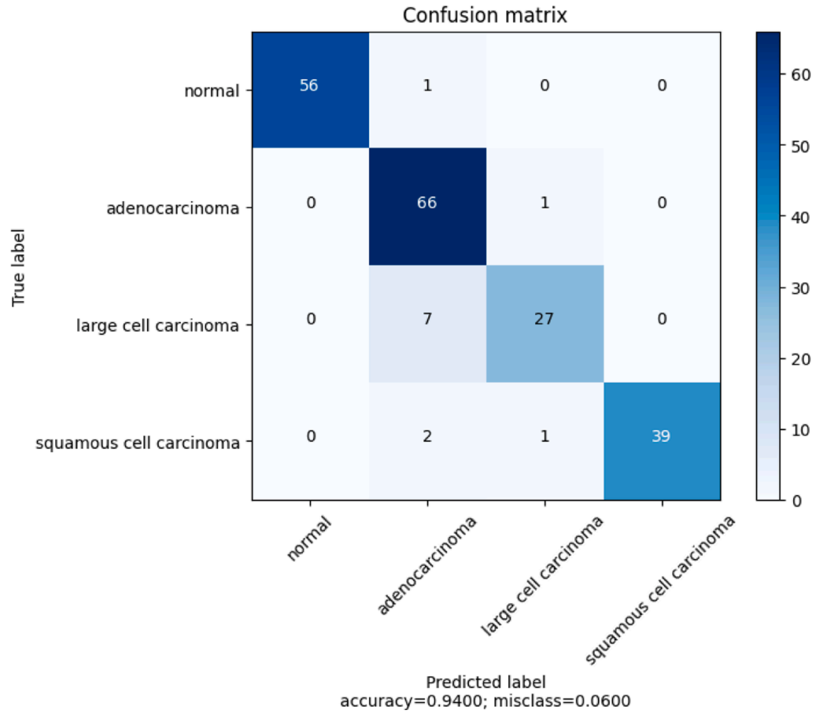

Fig. 19. Confusion matrix of the proposed EfficientNetB1 for multi-class classification of lung cancer types into four classes i.e. normal, adenocarcinoma, large cell carcinoma, squamous cell carcinoma.

EfficientNetsB1 model on the test data is presented through a confusion matrix, depicted in Fig. 19. Analyzing the confusion matrix provides insights into the accuracy of predictions for different lung cancer types and normal images. Notably, it is observed that among the various lung cancer types and normal images, the class representing large cell carcinoma exhibits the lowest accuracy in terms of correct predictions.

6. Conclusion and future directions

This research paper presents a transfer learning-based fine-tuning approach using EfficientNetB1 for the classification of lung cancer from CT scan images. The proposed method demonstrates several strengths that contribute to its effectiveness and potential for practical application. One of the key strengths is the utilization of pre-trained deep neural networks and transfer learning. By leveraging the knowledge learned from large-scale natural image datasets, EfficientNetB1 is fine-tuned to effectively extract relevant features from CT scan images of lung cancer. This approach significantly reduces computational resources and time required for training while maintaining high accuracy levels. The experimental results showcase the strengths of the proposed method. EfficientNetB1 consistently outperforms other CNN architectures in terms of accuracy and efficiency, achieving an impressive accuracy of 99.10% and demonstrating robust ROC scores ranging from 0.97 to 0.99 on test sets. These results highlight the potential of EfficientNetB1 as a powerful tool for automated diagnosis of lung cancer classification. However, it is important to acknowledge certain limitations of the current method. The performance of the proposed model could be further enhanced by utilizing a larger dataset and employing synthetic data generation techniques such as Generative Adversarial Networks (GANs). Incorporating additional diverse data samples would improve the model's ability to generalize and handle variations in lung cancer cases. Additionally, the implications of our findings extend to future research in the field of lung cancer diagnosis and classification, particularly focusing on the potential of transfer learning with EfficientNets. This approach opens up exciting avenues for exploration and improvement in medical image analysis. One potential future direction is the investigation of alternative deep learning architectures in combination with transfer learning. While EfficientNetB1 has shown remarkable performance in our study, exploring other state-of-the-art architectures could lead to further enhancements in accuracy and efficiency. Models such as ResNet, DenseNet, or InceptionNet could be evaluated to assess their effectiveness in the context of lung cancer classification. Furthermore, the incorporation of additional clinical data holds promise for future research. Expanding the dataset to include a wider range of cases, demographic information, and medical history could provide valuable insights into the interplay between different factors and improve the robustness of the classification model. Lastly, future studies could explore the application of transfer learning with EfficientNets on larger datasets. Scaling up the dataset size would further validate the effectiveness of the proposed approach in real-world scenarios and increase its generalizability. Additionally, investigating the potential benefits of data augmentation techniques and synthetic data generation, such as GANs, could help overcome limitations posed by limited data availability and improve the model's performance in rare or challenging cases.

CRediT authorship contribution statement

Rehan Raza: Experiments and initial draft. **Fatima Zulfiqar:** Experiments and initial draft. **Muhammad Owais Khan:** Writing manuscript. **Muhammad Arif:** Experiments and initial draft. **Atif Alvi:** Wrting manuscript. **Muhammad Aksam Iftikhar:** Wrting manuscript. **Tanvir Alam:** Initial draft, Conceptualization, Conceiving and designing experiments, Writing manuscript.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Data availability

We used publicly available dataset

Acknowledgment

The open-access publication of this article was supported by the Qatar National Library (QNL), Doha, Qatar. This work was supported by Qatar National Reseach Fund under RRC02-0805-210019 to Tanvir Alam.

References

AL-Huseiny, M.S., Sajit, A.S., 2021. Transfer learning with GoogLeNet for detection of lung cancer. Indonesian J. Electr. Eng. Comput. Sci. 22, 1078–1086.
Al-Yasriy, Hamdalla F., AL-Husieny, Muayed S., Mohsen, Furat Y., Khalil, Enam A., Hassan, Zainab S., 2020a. Diagnosis of lung cancer based on CT scans using CNN. IOP Conf. Ser.: Mater. Sci. Eng. 928 (2), 022035.
Al-Yasriy, H.F., AL-Husieny, M.S., Mohsen, F.Y., Khalil, E.A., Hassan, Z.S., 2020b. Diagnosis of lung cancer based on CT scans using CNN. IOP Conf. Ser.: Mater. Sci. Eng., 022035
Alyasriy, H., Muayed, A., 2020. The IQ-OTHNCCD lung cancer dataset. Mendeley Data 1, 1–13 (Accessed 19 January 2023).
American Cancer Society, I., 2023. What is lung cancer. Available: http://www.cancer.org/cancer/what-is-cancer.html.
Anon, 2023a. Chest CT-Scan images Dataset. Available: Chest CT-scan images dataset. https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images. (Accessed 07 June 2023).
Anon, 2023b. The IQ-OTH/NCCD lung cancer dataset — Kaggle. https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset. (Accessed 19 January 2023).
Asuntha, A., Srinivasan, A.J.M.T., Applications, 2020. Deep learning for lung cancer detection and classification, 79, pp. 7731–7762.
Chen, Jiahao, Ma, Qianli, Wang, Weixin, 2021a. A lung cancer detection system based on convolutional neural networks and natural language processing. In: 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology. AINIT. IEEE.
Chen, J., Ma, Q., Wang, W., 2021b. A Lung Cancer Detection System Based on Convolutional Neural Networks and Natural Language Processing. In: 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology. AINIT, pp. 354–359.
Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., et al., 2018. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning, 24, pp. 1559–1567.
Fontana, R.S., Sanderson, D.R., Woolner, L.B., Taylor, W.F., Eugene Miller, W., Muhm, J. R., et al., 1991. Screening for lung cancer. A critique of the Mayo Lung Project 67, 1155–1164.
Ge, F., Li, C., Iqbal, S., Muhammad, A., Li, F., Thafar, M.A., et al., 2023. VPatho: A deep learning-based two-stage approach for accurate prediction of gain-of-function and loss-of-function variants. Brief. Bioinform. 24, bbac535.
Ge, F., Zhang, Y., Xu, J., Muhammad, A., Song, J., Yu, D.-J., 2022. Prediction of disease-associated nssnps by integrating multi-scale ResNet models with deep feature fusion. Brief. Bioinform. 23, bbab530.
Ge, F., Zhu, Y.-H., Xu, J., Muhammad, A., Song, J., Yu, D.-J.J.C., et al., 2021. MutTMPredictor: Robust and accurate cascade XGBoost classifier for prediction of mutations in transmembrane proteins, 19, pp. 6400–6416.
Gordienko, Y., Gang, P., Hui, J., Zeng, W., Kochura, Yu., Alienin, O., Rokovyi, O., Stirenko, S., 2017. Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition arXiv preprint arXiv:151203385.
Hiremath, I.S., Goel, A., Warrier, S., Kumar, A.P., Sethi, G., o. C. P. Garg, M.J.J., 2022. The multidimensional role of the WNT/β-catenin signaling pathway in human malignancies, 237, pp. 199–238.
Hu, J., Zeng, W.-W., Jia, N.-X., Arif, M., Yu, D.-J., Zhang, G.-J., 2023. Improving DNA-binding protein prediction using three-part sequence-order feature extraction and a deep neural network algorithm. J. Chem. Inf. Model. 63, 1044–1057.
Humayun, Mamoona, Sujatha, R., Almuayqil, Saleh Naif, Jhanjhi, N.Z., 2022a. A transfer learning approach with a convolutional neural network for the classification of lung carcinomaHealthcare, 10 (6), 1058. MDPI.
Humayun, M., Sujatha, R., Almuayqil, S.N., Jhanjhi, N., 2022b. A transfer learning approach with a convolutional neural network for the classification of lung carcinoma. Healthcare 1058.
Jaszcz, A., Połap, D., Damaševičius, R., 2022. Lung X-Ray image segmentatioň using heuristic red fox optimization algorithm. Sci. Program. 2022, 1–8.
Jennifer, J.S., Sharmila, T.S., 2023. A neutrosophic set approach on chest X-rays for automatic lung infection detection. Inf. Technol. Control 52, 37–52.
Kadir, T., Gleeson, F., 2018. Lung cancer prediction using machine learning and advanced imaging techniques. Transl. Lung Cancer Res. 7, 304.
Kareem, H.F., AL-Husieny, M.S., Mohsen, F.Y., Khalil, E.A., Hassan, Z.S., 2021. Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset. Indonesian J. Electr. Eng. Comput. Sci. 21, 1731.
Katsamenis, I., Protopapadakis, E., Voulodimos, A., Doulamis, A., Doulamis, N., 2020. Transfer learning for COVID-19 pneumonia detection and classification in chest X-ray images. In: 24th Pan-Hellenic Conference on Informatics. pp. 170–174.
Kavitha, P., Prabakaran, S., 2019. A novel hybrid segmentation method with particle swarm optimization and fuzzy c-mean based on partitioning the image for detecting lung cancer.
Khan, M.A., Rajinikanth, V., Satapathy, S.C., Taniar, D., Mohanty, J.R., Tariq, U., et al., 2021. VGG19 network assisted joint segmentation and classification of lung nodules in CT images. Diagnostics 11, 2208.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization arXiv preprint arXiv:1412.6980.
Liu, J., Dong, B., Wang, S., Cui, H., Fan, D.-P., Ma, J., et al., 2021. COVID-19 lung infection segmentation with a novel two-stage cross-domain transfer learning framework. Med. Image Anal. 74, 102205.
Lustberg, T., van Soest, J., Gooding, M., Peressutti, D., Aljabar, P., van der Stoep, J., et al., 2018. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiotherapy Oncol. 126, 312–317.
Lyu, Lei, 2021. Lung cancer diagnosis based on convolutional neural networks ensemble model. In: 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology. AINIT. IEEE.
Narin, D., Onur, T.Ö., 2022. The effect of hyper parameters on the classification of lung cancer images using deep learning methods. Erzincan Univ. J. Sci. Technol. 15, 258–268.
Poap, D., Wozniak, M., Damaševičius, R., Wei, W., 2018. Chest radiographs segmentation by the use of nature-inspired algorithm for lung disease detection. In: 2018 IEEE Symposium Series on Computational Intelligence. SSCI, pp. 2298–2303.
Protonotarios, N.E., Katsamenis, I., Sykiotis, S., Dikaios, N., Kastis, G.A., Chatziioannou, S.N., et al., 2022. A few-shot U-net deep learning model for lung cancer lesion segmentation via PET/CT imaging. Biomed. Phys. Eng. Express 8, 025019.
Raza, R., Zulfiqar, F., Tariq, S., Anwar, G.B., Sargano, A.B., Habib, Z., 2022. Melanoma classification from dermoscopy images using ensemble of convolutional neural networks. Mathematics 10, 26.
Sajja, T., Devarapalli, R., Kalluri, H., 2019. Lung cancer detection based on CT scan images by using deep transfer learning. Traitement du Signal 36, 339–344.
Sultana, A., Khan, T.T., Hossain, T., 2021. Comparison of Four Transfer Learning and Hybrid CNN Models on Three Types of Lung Cancer. In: 2021 5th International Conference on Electrical Information and Communication Technology. EICT, pp. 1–6.
Sun, W., Zheng, B., Qian, W., 2016. Computer aided lung cancer diagnosis with deep learning algorithms. In: Medical imaging 2016: computer-aided diagnosis. pp. 241–248.
Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114.
Tekade, R., Rajeswari, K., 2018. Lung cancer detection and classification using deep learning. In: 2018 fourth international conference on computing communication control and automation. ICCUBEA, pp. 1–5.
U.S.D.o.H.a.H. Services, N.I.o. Health, N.C. Institute, 2023. Cancer treatment. Available: http://www.cancer.gov/about-cancer/treatment.
Voulodimos, A., Protopapadakis, E., Katsamenis, I., Doulamis, A., Doulamis, N., 2021. Deep learning models for COVID-19 infected area segmentation in CT images. In: The 14th Pervasive Technologies Related to Assistive Environments Conference. pp. 404–411.
Wang, S., Dong, L., Wang, X., Wang, X.J.O.M., 2020. Classification of pathological types of lung cancer from CT images by deep residual neural networks with transfer learning strategy, 15, pp. 190–197.
WHO, 2023. Cancer. Available: http://www.who.int/health-topics/cancer#tab=tab_1.