Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

DenseNet201-Based Deep Transfer Learning Framework for Accurate Multi-Class Lung Cancer Classification Using Computed Tomography Images

*D. Priyanka[1], Dr E. Aravind[2]
Research Scholar Assistant Professor Department of CSE Chaitanya – Deemed to be University
Kupriyanka1223@gmail.com
Department of CSE Chaitanya – Deemed to be University
aravind@chaitanya.edu.in

ARTICLEINFO

Received: 03 Nov 2024

Revised: 17 Dec 2024

Accepted: 27 Dec 2024

ABSTRACT

Timely and precise diagnosis of lung cancer is vital to better survival and effective clinical decision making. With the advent of deep learning, automated medical image analysis has shown great promise; however, simultaneously attaining high classification accuracy, robustness, and interpretability of the model is difficult. In this study, a Deep Transfer Learning (DTL) approach with DenseNet201 network is proposed for classification of lung cancer if it is multi-class in nature, based on the computed tomography (CT) images.In this study, a DenseNet201 based deep transfer learning (DTL) framework is proposed for multi-class lung cancer classification using computed tomography (CT) images from publicly available IQ-OTHNCCD dataset. The proposed architecture consists of a DenseNet201 backbone pre-trained with ImageNet and a custom classification head (GAP, BN, FC, Dropout, Softmax). A comprehensive data augmentation, label smoothing, class-weight balancing, early stopping, adaptive learning rate scheduling, and two-stage fine-tuning were used in the training process for improving generalization and preventing overfitting. The CT images of the dataset were divided into benign, malignant and normal classes, with 767, 164 and 166 images in the training, validation, and testing sets respectively. Experimental results showed that DenseNet201 framework attained the highest accuracy of 96.39%, precision of 96.49%, recall of 96.39%, and F1-score of 96.30%, in the test phase. The proposed model has been compared with ResNet50 (87.95%) and EfficientNetB0 (72.89%) under the same experimental conditions, and it has been confirmed that the proposed model has a stronger feature extraction ability for lung CT image classification. In addition, the classification accuracy was validated by a confusion matrix analysis, receiver operating characteristic (ROC) curves, area under the curve (AUC), and class-wise sensitivity and specificity analysis were performed, and the model interpretability was increased by using the Gradient-weighted Class Activation Mapping (Grad-CAM). The experimental results show that the proposed DenseNet201 framework is able to classify lung cancers accurately, robustly and explainably into multi-class and has great potential for use in computer-aided diagnosis (CAD) system for early detection of lung cancer by radiologists.

Keywords DenseNet201; Lung Cancer Classification; Deep Transfer Learning; Computed Tomography (CT); Medical Image Analysis; Computer-Aided Diagnosis (CAD); Explainable Artificial Intelligence (XAI); Grad-CAM; Deep Learning; Multi-Class Classification.

1. Introduction

While significant advances have been made in the field of diagnostic imaging, therapeutic interventions and precision medicine, lung cancer still stands as one of the most severe public health problems in the world, and it continues to be a primary cause of cancer-related mortality. This disease involves abnormal growth of cells in the lung tissue, and its symptoms are not usually apparent until the disease is well developed. As a result, many lung cancer cases are identified at late stages, making the treatment

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

method and the likelihood of survival much less effective. Many clinical trials have shown that early diagnosis can make a dramatic difference in treatment results, allowing for intervention and treatment planning to be tailored to the individual patient. Thus, a reliable and accurate computer-aided diagnostic system to assist a doctor in early diagnosis of lung cancer has been an important research goal in medical image analysis [1–4].

The use of computed tomography (CT) imaging has become the imaging modality of choice for lung cancer screening because it allows high-resolution cross sectional imaging of the pulmonary structures and the ability to detect very small nodules that are not often seen on conventional chest radiography. CT images deliver better anatomical information than other imaging modalities, enabling the size, shape, texture, density, and location of lesions be assessed. These imaging characteristics are crucial to distinguishing malignant tumours from benign abnormalities, and normal pulmonary tissue. But, with the surge of the CT examination in the current health care setting, radiologists' workload has also risen substantially. Reading 100s of slices, one by one, from a patient's CT scan is time-consuming and can lead to varying interpretations due to fatigue. In addition, there can be a lot of variation in the appearance of benign and malignant nodules which can make it difficult to accurately diagnose them even for experienced clinicians. These restrictions have led to the creation of intelligent CAD systems which help radiologists with objective and reproducible diagnostic decisions [2–4].

In recent years, the field of automated medical image analysis, especially deep learning, has seen remarkable advancements with the potential of computer systems to learn very discriminative representations directly from raw imaging data. Deep learning models learn a hierarchically structured feature representation by using several convolutional layers, without the need for handcrafted feature extraction based on intensity, texture, shape and statistical descriptors, as is the case with traditional machine learning methods. This ability can greatly enhance the accuracy of classification with minimal manual feature engineering. Convolutional Neural Networks (CNNs) are thus one of the most successful deep learning architectures used for disease detection, medical image segmentation, lesion localization and image classification. It has been consistently reported that CNN-based models outperform traditional machine learning models in lung cancer diagnosis due to their ability to learn the complexity of the lung CT image structure and pathological characteristics [2–4].

Another area where the practical use of healthcare data has been hindered by the scarcity of annotated clinical images has been addressed by the use of transfer learning, which has facilitated the application of deep learning to the medical imaging domain. Transfer learning reduces the need to train deep neural networks from scratch from the largest natural image collections, such as ImageNet, and then fine-tune the network to the domain-specific medical imaging task. This approach can be very cost-effective, fast to train and has better accuracy for classification and requires less reliance on annotated big data sets. Previous work has shown that transfer learning greatly increases the robustness and generalisation power of lung cancer classification systems by transferring generic visual representations to the CT image analysis domain [1,8,9]. Therefore, transfer learning has emerged as one of the most popular approaches in the development of computer-aided diagnosis (CAD) systems for reliable deep learning.

While automatic diagnosis of lung cancer has made remarkable advances, the following problems are still to be solved. The multi-class classification is challenging due to the high variation in lesion morphology, size, texture, anatomic location and imaging contrast seen in lung CT images. Although benign pulmonary nodules may have imaging appearances similar to normal lung, early malignant lesions may have only subtle imaging appearances that differ from those of benign nodules. These similarities make the prediction of these conditions more likely to be overestimated (false-positive) or underestimated (false-negative) which impacts the accuracy of the diagnosis. Moreover, the datasets used for medical imaging are frequently imbalanced, samples are sparse, and there are variations caused by different imaging protocols and hardware. All those things hinders the generalization of the model and makes it more likely to overfit, stressing the importance of developing stronger feature extraction architectures that are able to learn more highly discriminative image representations for a wider range of clinical conditions [3,4,8].

Several deep convolutional neural network models have been suggested to enhance classification of lung cancer. In the field of medical imaging, several pretrained models such as VGGNet, ResNet, DenseNet, InceptionNet, MobileNet, EfficientNet, and hybrid CNN models have shown promising results in different medical image classification applications. In these architectures, Dense Convolutional Networks (DenseNets) have gained massive popularity due to their dense connectivity pattern, in which each convolutional layer not only receives features from all previous layers but also passes its learned features to all subsequent layers. This architecture is efficient to reuse features, boosts gradient propagation, eliminates parameter redundancy and maintains low-level texture information and high-level semantic features across the network. Therefore, DenseNet architectures display a good learning capacity with relatively efficient parameter usage, which makes them particularly suitable for the medical image analysis domain, where the pathological variation is subtle.

Beyond its ability to provide high classification accuracy, interpretability of deep learning models has grown in significance for clinical deployment. Most deep neural networks are black box networks that are hard for the clinician to interpret. This lack of transparency can undermine the trust physicians have even in prediction models that perform well. Explainable Artificial Intelligence (XAI) techniques have thus become a crucial element in recent times of the medical image analysis field. Of these methods, Gradient-weighted Class Activation Mapping (Grad-CAM) has become one of the most popular visualization methods due to its ability to identify the most important regions of an image that lead the network to the final prediction. These visual explanations help to enhance transparency, support qualitative validation by clinicians, and boost the trust placed in AI-powered diagnostic systems.

In addition to lung cancer, deep learning has shown significant promise in other medical imaging tasks, highlighting its versatility and potential for automated disease diagnosis. Transfer learning, regularization, and data augmentation have been successfully used for brain tumor classification in recent studies, where the advanced CNN architectures are able to learn discriminative imaging features across modalities of medical imaging [14,19]. The findings in this study indicate the potential for transfer learning between different diagnostic tasks, and stimulate the investigation of advanced DenseNet architectures to assist with automated classification of lung cancer in other clinical settings.

While significant advancements have been made in the diagnosis of lung cancer using deep learning techniques, there are still some research gaps in the current literature. Most existing studies are limited to binary classification and/or use relatively small data sets for the evaluation of their frameworks without detailed comparative analysis. Moreover, some of the existing methods focus mostly on general classification performance and lack assessment of clinically meaningful parameters like sensitivity, specificity, receiver operating curve (ROC), area under the curve (AUC), and model explainability. Moreover, the majority of works do not provide a thorough comparison with multiple state-of-the-art backbone architectures in the same experimental setup, under the same experimental conditions and with the inclusion of strong transfer learning and complete optimization techniques, and the inclusion of explainable artificial intelligence. Such restrictions highlight the importance of having a comprehensive and easily understood deep learning system that will offer robust and reliable multi-class lung cancer classification with high generalization performance.

To meet these challenges, the present study introduces a deep transfer learning framework for automated multi-class lung cancer classification from computed tomography (CT) images by using the publicly available IQ-OTHNCCD dataset, which is based on DenseNet201. The proposed network design includes an ImageNet pre-trained DenseNet201 backbone and a customized classification head, which is composed of Global Average Pooling, Batch Normalization, fully connected layers, dropout regularization, and Softmax classification. To enhance convergence stability and mitigate overfitting, a two-stage transfer learning approach is used with image preprocessing, a large data augmentation set, label smoothing, adaptive learning-rate scheduling, class-weight balancing, and early stopping. Additionally, the proposed framework is tested extensively with various quantitative performance measurements such as accuracy, precision, recall, F1 score, confusion matrix, sensitivity, specificity, ROC analysis, AUC and explainability measures using Grad-CAM. To directly evaluate the effectiveness of the proposed DenseNet201 framework, comparative experiments with ResNet50 and EfficientNetB0 are also performed under the same experimental setup. The main features of this work include: (i) a

DenseNet201-based transfer learning framework for automated multi-class lung cancer classification; (ii) multiple optimization techniques were combined to achieve the two-stage fine-tuning strategy, thereby improving the robustness in learning; (iii) the proposed framework was comprehensively evaluated with the conventional classification metrics and explainable artificial intelligence; and (iv) the performance of the proposed architecture was compared with the existing transfer learning architectures and the effectiveness of the proposed framework was demonstrated. The overall goal of the proposed study is to develop an accurate, robust and interpretable computer-aided diagnosis system which can help radiologists make early diagnosis of lung cancer and contribute to intelligent clinical decision making in the future.

2. Related Work

The recent developments in AI have revolutionized the computer-aided diagnosis (CAD) system for lung cancer, allowing automatic analysis of computed tomography (CT) images with higher accuracy, effectiveness, and diagnostic consistency. In the last decade, researchers have suggested a wide range of machine learning and deep learning algorithms to detect, classify, analyse the histological features and predict the prognosis of lung cancer. Deep learning models can learn hierarchical feature representations directly from CT images without relying on handcrafted feature engineering, thus significantly enhancing the diagnostic performance and eliminating reliance on feature engineering. Although these success stories come with great promise, there are still several challenges to address, such as the availability of limited datasets, the lack of class balance, the computational complexity, the decreased interpretability of models, and the poor generalizability of the models across clinical datasets.

Initial research on automatic diagnosis of lung cancer was mostly based on traditional machine learning methods that relied on manually designed radiomic, morphological and texture-based features of CT images. The used feature description techniques were Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), histogram statistics and shape description, and the classification methods were Support Vector Machines (SVM), Random Forests, Decision Trees and Artificial Neural Networks. Singh and Gupta [13] compared the performance of various machine learning algorithms on the detection of lung cancer and found that the manually extracted features in the images were a crucial factor in the performance of the prediction. Likewise, Podolsky et al. [16] examined machine learning techniques that utilize gene expression data to classify lung cancer and found promising predictive power with molecular markers. Lynch et al. [20] continued their study of supervised machine learning approaches to forecasting the survival for lung cancer patients, demonstrating the possible use of artificial intelligence in personalized medicine. These traditional methods helped identify key clinical features but were limited by relying on manually extracted features, which failed to fully capture the complex nature of pulmonary lesions and their imaging characteristics.

Nowadays, medical image analysis has been revolutionized by the advent of deep learning, where Convolutional Neural Networks (CNNs) are trained to learn what discriminative hierarchical representations are directly from medical images. The CNN based model not only extracts features but also optimizes the classification process, which is in contrast to the traditional ML models that separately optimize the feature extraction and classification process, leading to better predictive accuracy and robustness. Wang [2] had reviewed the deep learning methods used in diagnosing lung cancer and found CNN based techniques outperform traditional ML techniques due to their outstanding feature learning. In a similar vein, Asuntha and Srinivasan [3] showed that the deep learning architectures offer a major boost in the accuracy of lung cancer detection and classification by capturing the complex spatial representation from CT images. Moreover, Javed et al. [4] conducted an extensive review and pointed out the transfer learning, explainable artificial intelligence, and ensemble learning as the key research fields for enhancing automated lung cancer diagnosis. These studies are all evidence of the current trend towards the use of deep learning for lung cancer classification via computed tomography.

Because annotated healthcare datasets can be small and costly, one of the most popular approaches used for medical image classification is transfer learning. Transfer learning promotes the transfer of knowledge from a large-scale dataset, e.g., ImageNet, to a specific medical imaging domain, e.g.,

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

medical imaging tasks, by fine-tuning the feature representations. For lung cancer classification from CT images, an optimized deep learning model was proposed by Lakshmanaprabu et al. [1] and it was proved that transfer learning method is very effective in improving the performance of the classification and reducing the computational complexity. Similarly, Riquelme and Akhloufi [8] demonstrated that, for medical imaging, pretrained CNN architectures are able to learn discriminative pulmonary features even when relatively small medical imaging datasets are available. Recently, Faizi et al. [9] reported that deep learning models for automated classification of CT images can be highly robust and generalizable by leveraging transfer learning. Combined, these studies have shown that transfer learning is a crucial approach for building reliable computer-aided diagnosis system to achieve high diagnostic accuracy using a limited amount of annotated data.

The advancement of CNNs has further improved automated lung cancer classification with the continuous evolution of the architecture. By using shortcut connections that avoid the vanishing gradient problem, residual learning proposed by ResNet has been used to successfully train substantially deeper neural network. EfficientNet then developed the idea of compound scaling for network depth, width, and input resolution to boost computational efficiency. Raza et al. [17] introduced the Lung-EffNet, an EfficientNet-based architecture specifically tailored to lung cancer classification, with the results showing a good performance and having relatively low computational complexity. Likewise, Ren et al. [15] proposed a hybrid deep learning architecture that combines complementary feature extraction mechanisms to enhance the classification accuracy. Pandit et al. [12] further improved deep learning optimization by optimizing the parameters using better optimizing strategies, which led to faster convergence and more accurate diagnosis. It is evident that these architectures have done a great deal to improve the automated diagnosis of lung cancer; however, feature reuse, computational complexity, interpretability and model generalization are still active research areas that drive the exploration of more sophisticated transfer learning architectures for multi-class lung cancer classification.

The multi-source complementary information integration, which can enhance the robustness and reliability of automated lung cancer diagnosis, has received more and more research attention recently. Multimodal deep learning frameworks leverage multi-omics data, genomic biomarkers, histopathology and clinical data, alongside radiological imaging, to create comprehensive diagnostic models instead of relying solely on CT images. Mohamed and Ezugwu [7] introduced a deep learning framework that uses multi-omics information combined with imaging data and showed that multimodal learning achieves better classification accuracy and helps to provide personalized health care. Likewise, Sangeetha et al. [10] proposed an improved multimodal fusion neural network that can fuse the multimodal representations of features derived from different sources, thus improving the diagnostic performance compared to traditional single-modality models. The studies showed that multimodal learning can enhance disease characterization, but in many cases, it is challenging to implement in clinical routine practices due to the need for large heterogeneous datasets and the complexity of the models.

The other significant research direction is weakly supervised learning and models for pathology-based deep learning, which minimize the need for extensive manual annotations. In order to reduce the cost of pixel-level annotation, Kanavati et al. [11] proposed a weakly-supervised approach for lung carcinoma classification based on image-level annotation. Similarly, Chen et al. [18] presented an annotation-free whole slide image classification method that can accurately classify lung cancer types, without requiring tedious manual annotation. Chaunzwa et al. [5] have shown that deep learning models could be used to classify the lung cancer histological subtypes directly from CT images, which would give clinically relevant information beyond mere disease detection. Moreover, deep learning was successfully applied to small-cell and non-small-cell lung cancer diagnosis by Kriegsmann et al. [6] that further demonstrates the potential of AI in various stages of diagnosis, such as radiological and pathological image analysis. All of these studies illustrate the progress of deep learning from binary classification to more comprehensive clinical decision support tools.

Additionally, deep learning has demonstrated impressive results in various other medical imaging tasks, which highlights the versatility and ability of CNN to generalize. Recent works have shown that transfer learning, data augmentation, and regularization techniques can greatly enhance disease classification accuracy for brain MRI analysis. CNN-based methods that use optimized transfer learning have yielded superior performance for automated brain tumor classification, thus demonstrating that powerful feature extraction models successfully trained on one imaging modality can be transferred to other medical imaging tasks, as reported in [14,19]. The results further confirm the generalizability of the sophisticated CNN architectures and give further impetus to the use of DenseNet201 for automated lung cancer classification. Results obtained for transfer learning across various medical imaging domains also illustrate that the judicious design of a pretrained architecture can achieve success in learning discriminative representations under the condition of relatively small domain-specific datasets.

Although deep learning has made great leaps forward in the diagnosis of lung cancer, there are several significant research gaps in the current literature. Firstly, most of the studies focus only on binary class classification while in clinical diagnosis the accuracy of discrimination between several disease categories is required. Second, many studies test their frameworks with small or institution-specific datasets, and so do not have broad generalization ability for other patient populations and imaging protocols. Third, while some studies have high classification accuracy, relatively few studies make detailed assessment with clinically relevant parameters such as sensitivity, specificity, Receiver Operating Characteristic (ROC) curve analysis, Area Under the Curve (AUC) and class-wise analysis. In addition, while there has been increasing emphasis on the need for explainable AI systems that clinicians can trust and explain, there are currently few existing systems that give clinicians sufficient confidence to use explainability. However, comparisons of state-of-the-art architectures of transfer learning are generally carried out in different experimental setups, which make it challenging to evaluate objectively the superiority of the different models.

To overcome these drawbacks, a new deep transfer learning framework is proposed in this study for automated multi-class classification of lung cancer from CT images of the IQ-OTHNCCD dataset, based on the DenseNet201 network. The main reason for choosing DenseNet201 is its dense connectivity mechanism which provides efficient feature reuse, better gradient propagation and maintains low level and high level feature representations during the learning process. The proposed framework combines the image preprocessing techniques, extensive data augmentation, transfer learning, two-stage fine-tuning, batch normalization, dropout regularization, label smoothing, adaptive learning-rate scheduling, early stopping, and class-weight balancing to enhance the robustness of the classification algorithm and minimize overfitting. Further, the proposed framework conducts a thorough quantitative and qualitative assessment through accuracy, precision, recall, F1 score, confusion matrix, sensitivity, specificity, ROC and AUC and also using explainability (Grad-CAM). To ensure a fair comparison of the proposed DenseNet201 architecture, comparative experiments using ResNet50 and EfficientNetB0 with same experimental setup are also conducted. Based on this, this study is motivated to building an accurate, robust and interpretable computer-aided diagnosis framework to support early lung cancer detection and help clinicians to make reliable diagnoses.

3. Methodology

The overall methodology used for the automated multi-class lung cancer classification with computed tomography (CT) images is described completely here. The proposed system combines image preprocessing, transfer learning, feature extraction using a DenseNet201 network, two-stage training, and explainable artificial intelligence methods to create a powerful computer-aided diagnostic system for identifying the lung nodules. The overall workflow of the proposed methodology is shown in Figure 1 and the training configuration and hyperparameter settings are provided in Table 1.

3.1 Overview of the proposed framework

The framework proposed in this paper follows a systematic deep learning pipeline for multi-class classification of lung cancer from the CT image. First, CT image data is retrieved from the IQ-OTHNCCD lung cancer database, and then pre-processed to standardize the input dimensions and enhance the quality of the image data. Then, the diversity of training samples is enhanced by data augmentation techniques to prevent overfitting. The pre-processed images are then fed to a DenseNet201 back bone pre-trained on ImageNet for deep feature extraction. A customized classification head, which includes

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Global Average Pooling, Batch Normalization, fully connected layers, dropout regularization is added at the end, and a Softmax output layer is used for three-class classification. A two-stage transfer learning strategy, which involves freezing feature extractor and fine-tuning higher convolutional layers, is adopted for training the network. Lastly, the trained model is tested with traditional classification accuracy and the Gradient-weighted Class Activation Mapping (Grad-CAM) is used to explain the network's predictions visually. The detailed process of the proposed methodology is shown in Figure 1.

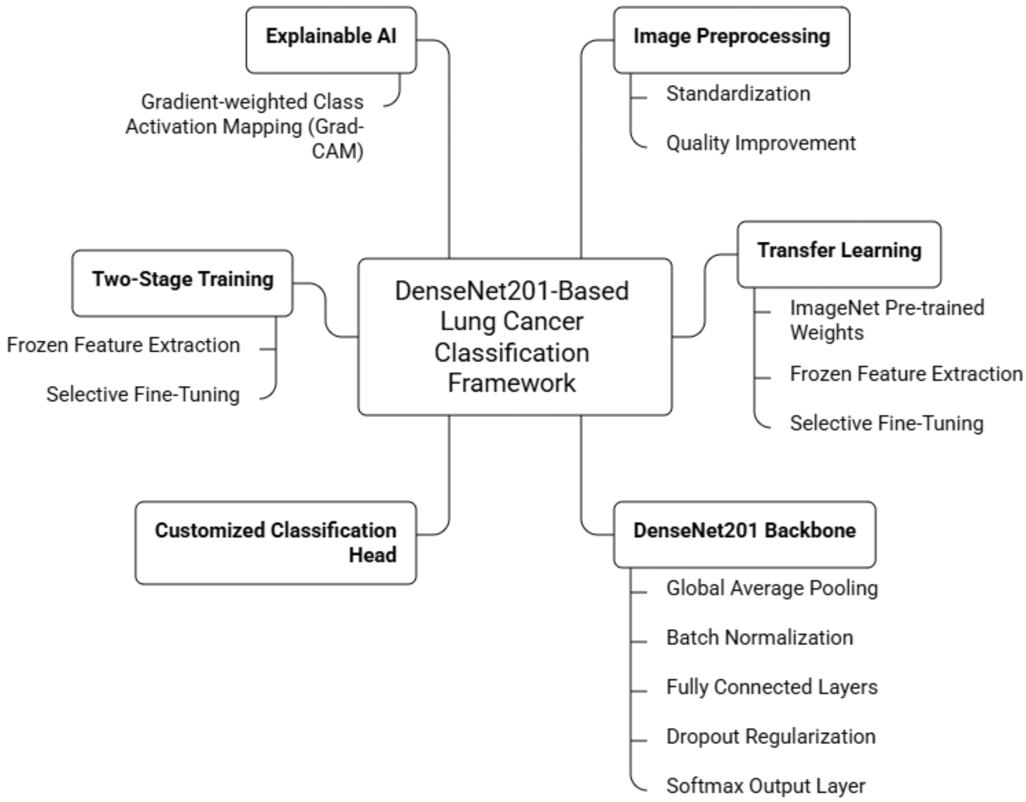


Figure 1. Overall workflow of the proposed DenseNet201-based framework for multi-class lung cancer classification.

3.2 Dataset Description

The proposed framework uses publicly available IQ-OTHNCCD lung cancer dataset consisting of thoracic CT images of three clinically relevant diagnostic categories: benign, malignant, and normal. The dataset contains a wide variety of anatomical structures, lesions, and image characteristics, which allow for the assessment of the ability of deep learning algorithms to classify them. The data set is meticulously divided into independent subsets of training, validation, and testing data before the model is developed, ensuring an unbiased evaluation of the model's performance and avoiding information leakage among the different partitions. It is proposed that a discriminative model be learned from the selected dataset, which provides a representative example of both cancerous and non-cancerous lung conditions to obtain discriminative imaging features that are necessary for accurate multiple class classification.

3.3 Image Preprocessing and Data Augmentation

The consistency is ensured for all the CT images by performing image preprocessing before supplying to the deep neural network. All the CT images are resized to 224 × 224 pixels, which is the input size

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

needed by DenseNet201. To make the preprocessed images better suited to feature representations trained on ImageNet, each pixel's intensity is normalized via the ImageNet-specific pretrained DenseNet201 preprocessing function. In order to increase model generalization and reduce overfitting, several online data augmentation techniques are used during training such as rotation, zoom, horizontal translation and vertical translation and horizontal flipping. These augmentation operations simulate the realistic anatomical variation and imaging conditions without changing the diagnostic characteristics of the CT scans, enhance the robustness of the proposed classification model.

3.4 Proposed DenseNet201 Architecture

The DenseNet201 network is used as the main feature extractor because it uses a dense connectivity mechanism that allows for efficient re-use of features and better propagation of gradients across the network. DenseNet creates direct connections between the consecutive layers, enabling the current layer to access the feature representations produced by all the previous layers, which is different from the traditional CNN. This architectural design is useful for extracting fine-grained texture and structural information from the CT images of the lungs while minimizing redundant feature learning. The DenseNet201 backbone is initialized with ImageNet-pretrained weights to take advantage of the generic visual representations acquired from using a large-scale image dataset. The pretrained back bone is coupled with a customized classification head: a Global Average Pooling layer is used to reduce the dimensionality of the features, and Batch Normalization is used to stabilize the distribution of the features during optimization. Two fully-connected layers with Rectified Linear Unit (ReLU) activation are added to train high-level discriminative representations and dropout layers are added to prevent overfitting by randomly silencing neurons during the training process. The last Softmax layer is used to generate the probability scores associated with the three diagnostic classes.

3.5 Transfer Learning and Model Training Strategy

A two-stage transfer learning is used to maximize learning efficiency and improve the convergence of the model. The convolutional layers of DenseNet201 are kept frozen and only the newly-designed classification layers are trained using CT scans of the lungs. This technique allows the classifier to learn the feature representations from the ImageNet data without interfering with the general feature representations. During the second training phase, some convolutional layers of DenseNet201 backbone are unfrozen and retrained with a lower learning rate.During the second training phase, some convolutional layers of DenseNet201 backbone are unfrozen and retrained with a lower learning rate. By fine tuning, the network can learn the high level semantic features while retaining the low level visual features based on the features of lung CT images. In the optimization, adaptive learning-rate scheduling, early stopping, model checkpointing, dropout regularization, label smoothing and class-weight balancing are used to enhance the stability of convergence, mitigate overfitting, and handle class imbalance.

3.6 Hyperparameter Configuration

The proposed DenseNet201 framework is optimized to obtain a trade-off between classification performance and computational efficiency in the training process. The model is trained with the Adam optimization algorithm and categorical cross-entropy loss with label smoothing. A mini-batch training technique is used and early stopping and adaptive learning-rate reduction are used to avoid executing unnecessary iterations beyond convergence. Table 1 gives a summary of all the hyperparameter values used during the development of the model.

Table 1. Hyperparameter configuration of the proposed DenseNet201 framework.

Hyperparameter	Value
Backbone Network	DenseNet201
Input Image Size	224 × 224 pixels
Transfer Learning	ImageNet Pretrained
Batch Size	16
Optimizer	Adam

Initial Learning Rate	0.0001
Fine-Tuning Learning Rate	0.00001
Frozen Training Epochs	20
Fine-Tuning Epochs	20
Loss Function	Categorical Cross-Entropy
Label Smoothing	0.10
Dropout Rates	0.50, 0.30
Output Activation	Softmax
Number of Classes	3

3.7 Explainability Using Gradient-weighted Class Activation Mapping

The predictive performance of deep convolutional neural networks is high, but the reasoning of making decisions is difficult to explain. In order to improve the transparency of the proposed framework, one post hoc explainability method, namely Gradient-weighted Class Activation Mapping (Grad-CAM), is included. The Grad-CAM approach is based on the observation that the gradients of the prediction on the final convolutional feature map in the DenseNet201 backbone can be used to create localization heatmaps. These heatmaps identify image regions that are most important for the classification decision and thus serve as visual proof of the focus of the model during the inference. By incorporating Grad-CAM, the interpretability of the proposed framework was enhanced and this tool could be utilized to qualitatively validate whether the network is focusing on a diagnostically meaningful region of the anatomy, which could lead to better potential use for computer-aided diagnosis.

4. Experimental Results and Discussion

In this section, the proposed DenseNet201-based framework for automated multi-class lung cancer classification from computed tomography (CT) images is evaluated thoroughly through experiments. The effectiveness of the proposed approach is evaluated using an independent test data set by quantitative and qualitative analysis. Firstly the experimental setup and the environment of the implementation is described and then the model learning behavior is analyzed using training and validation curves. The performance of the classification, Receiver Operating Characteristic (ROC) analysis, sensitivity and specificity evaluation, comparison with the state-of-the-art transfer learning model, and explainability evaluation with Gradient-weighted Class Activation Mapping (Grad-CAM) are subsequently presented. The combination of these experiments offers a thorough assessment of the robustness, generalization and clinical applicability of the proposed framework.

4.1 Experimental Setup

The lung cancer CT image data for all experiments were the publicly available IQ-OTHNCCD. There are 1097 CT scans of the thorax, each of which is categorized as benign, malignant or normal. In order to fairly assess the proposed framework, the data set was divided into the mutually exclusive training, validation, and testing sets with 767, 164, and 166 images, respectively. In order to minimize sampling bias for the process of model development and maintain the class distribution in all partitions, stratified data partitioning was used.

The framework was built using Python, deep learning library TensorFlow and Keras. To transfer learn, the DenseNet201 pretrained on ImageNet was used. All the CT images were resized into 224 × 224 pixels and normalized with the DenseNet201's preprocessing function. To make the model more robust and prevent overfitting, during the training process, some of the online data augmentation techniques were employed, such as random rotation, zoom transformation, horizontal translation, vertical translation and horizontal flipping. The categorical cross-entropy loss was used with label smoothing applied to boost the confidence of the predictions and decrease overfitting, while the Adam optimization algorithm was used to optimize the classifier.

A two-stage transfer learning approach was used during model development. The DenseNet201 back bone was frozen and the new classification layers were optimized in the first phase, using the training data. Selected higher convolutional layers of the DenseNet201 backbone were then unfrozen and trained with a decreased learning rate to learn high-level feature representations that are suitable for lung CT

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

images. To increase convergence stability and handle the class imbalance, early stopping, adaptive learning-rate scheduling, model checkpointing, dropout regularization, and balancing the class weights were implemented. The final model was tested only on the test images of the independent test set to prevent any knowledge learnt from the test set images from affecting the training process.

4.2 Training Performance Analysis

The learning characteristics of the proposed DenseNet201 framework were investigated by analyzing both the training and validation accuracy and loss throughout the optimization process. The monitoring of these learning curves gives insight into convergence behavior, learning stability and the generalization capability of the proposed model.

The training and validation accuracy are shown in Figure 2 during the entire training. Initially, both accuracy curves show a fast improvement during the training of the network with representative low-level features and high-level features from the CT images. The accuracy of the training set also rises gradually over the course of the successive epochs, with the validation set accuracy rising in a similar pattern, albeit with some minor fluctuations. Once about the middle of the training, both curves flatten out and start to merge, a near optimal solution was found by the optimization. More important, the validation accuracy is very close to the training accuracy, and the difference is not significant, indicating that the proposed regularization methods do not overfit significantly. The small difference between the two curves is yet another sign of the learned features generalizing across the test set of CT images, rather than memorizing the learning set.

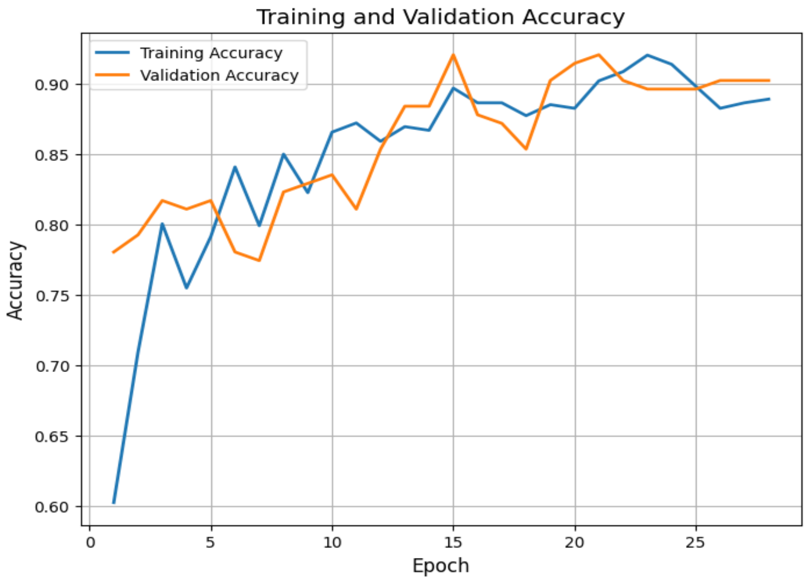


Figure 2. Training and validation accuracy curves of the proposed DenseNet201 framework.

The smooth convergence shown in Figure 2 is possible because of the use of transfer learning, dropout regularization, label smoothing, batch normalization, adaptive learning-rate scheduling and two-stage fine-tuning. The transfer learning process starts with powerful initial feature representations and subsequent fine-tuning allows the network to learn the features specific to lung CT images without disrupting the optimization process. Additionally, when convergence is reached, early stopping will help to avoid unnecessary training, minimizing the risk of overfitting.

Likewise, the curve of training loss and validation loss in Figure 3 illustrate a consistent optimization process during learning. The loss values both decrease gradually as the training progresses, demonstrating the network's gradual improvement in reducing the number of classification mistakes. While some oscillations are noted for a few epochs, they are considered as noise in stochastic optimization and gradually fade as the optimizer approaches a stable minimum. The validation loss

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

curve shows a similar shape to the Training loss curve and demonstrates the stability of the proposed model in its predictive performance on validation data not seen during the training process.

Figure 3. Training and validation loss curves of the proposed DenseNet201 framework.

The fact that validation accuracy has improved while validation loss has decreased throughout the training process suggests that the network is still learning discriminative representations that are meaningful, not just fitting the random noise in the training set. Furthermore, the level of validation loss does not significantly increase in later epochs, indicating that overfitting has been well managed by using dropout layers, label smoothing, adaptive learning-rate reduction and early stopping. The trends of the convergence shown in Figures 2 and 3 also validate the stability of the optimization process and the generalization ability of the proposed DenseNet201 framework, which provides a foundation for quantitative evaluation in the following sections.

4.3 Classification Performance Analysis

Multiple metrics, such as accuracy, precision, recall, F1 score, and confusion matrix analysis were employed to assess the classification ability of the proposed DenseNet201 framework with the independent test set. These evaluation metrics give a holistic evaluation of the model's predictive potential by assessing the general classification performance and its ability to accurately categorize types of lung cancer.

The quantitative performance results measured from the proposed DenseNet201 framework is summarized in Table 2. The proposed model gave a classification accuracy of 96.39%, proving the efficiency in the classification of benign, malignant and normal CT images. Moreover, weighted precision, recall and F1-scores were 96.49%, 96.39% and 96.30%, respectively. The values of these performance measures are very close, which suggests that the proposed framework performs well in terms of achieving a balance between the classification capability for each class. High precision means that most (if not all) of the classes that were predicted were correct, and high recall means that most (if not all) of the clinically relevant cases were identified as such. Likewise, the F1-score indicates a good balance between precision and recall scores, thereby showing that the proposed classifier is strong in all the diagnostic categories.

Table 2. Overall classification performance of the proposed DenseNet201 framework.

Performance Metric	Value
Accuracy	96.39%
Precision	96.49%
Recall	96.39%
F1-Score	96.30%

The consistent high scores of all the evaluation metrics show that DenseNet201 is able to learn discriminative representations from lung CT images and can distinguish between benign, malignant

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

and normal cases. These results also suggest that the proposed transfer learning approach, fine-tuning and regularization methods play a significant role in enhancing the generalization ability of the model.

A confusion matrix was also created to explore further the classification behavior of the proposed framework as shown in Figure 4. The confusion matrix presents the actual classification results in a detailed way for each diagnostic category, allowing a detailed knowledge of the classification strengths and the current difficulties.

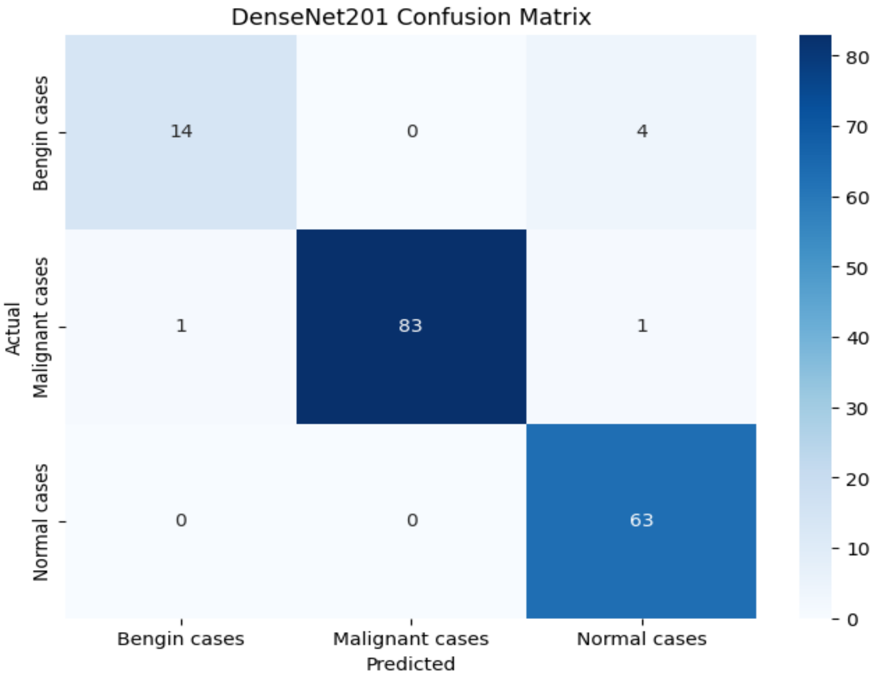


Figure 4. Confusion Matrix

The proposed DenseNet201 model was able to correctly classify most of the CT images as shown in Figure 4 across all three classes. Of the 18 benign CT images in the test set, 14 were correctly identified as benign and four were wrongly identified as normal. Most importantly, none of the benign images were misclassified as malignant, which demonstrates the capability of the proposed model to distinguish non-malignant lung abnormalities from malignant tumors. While several benign cases were misclassified as normal tissue, there is a clinical reason for this as benign lesions have imaging characteristics that are similar to healthy lung tissue.

In the case of malignant category the proposed model has shown excellent classification result with the correct classification of 83 out of 85 malignant CT images. A single normal image was wrongly classified as malignant and one image was wrongly classified as normal. The result indicates DenseNet201 can learn highly discriminative tumor properties, which is essential to ensure that the cancer diagnosis is not missed in clinical practice. The extremely low percentage of misclassifications that were malignant is a good sign of the high sensitivity of the proposed framework to malignant lesions.

In a similar fashion, the proposed framework showed superb performance on the normal class with a perfect accuracy of 100 percent with no false negatives on the 63 normal CT images. This ideal performance in terms of classification proves the efficiency of the proposed model in the detection of healthy pulmonary anatomy and the avoidance of unnecessary false detection. This attribute is especially advantageous in computer-assisted diagnosis systems, where the excessive follow-up examination of healthy people can be avoided.

In general, the confusion matrix shows that the classification performance of the proposed DenseNet201 framework is very reliable for all the diagnostic categories. The errors of prediction are

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

mainly made between the classes benign and normal, while malignant tumors are diagnosed with good accuracy. This observation implies that the extracted feature representations from the images are indeed clinically relevant and effective in characterizing the lung lesion types that are malignant, and also have high discriminatory power for the rest of the lung lesion types.

4.4 Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC) Analysis

In addition to evaluating the proposed DenseNet201 framework using the traditional classification accuracy, Receiver Operating Characteristic (ROC) analysis was employed to assess the discriminative power. ROCs, unlike accuracy-based ones, explore how the performance of a model changes as a function of the classification threshold. The visual representation of all ROC curves can be found in the following figure (Figure 5) and the quantitative values of the AUC is summarized in table 3.

ROC curves shown in Figure 5 are quite close to the top-left corner of the ROC space, thus demonstrating good classification performance across all three diagnostic categories. The sharp increase in every ROC curve shows that the proposed framework is able to attain high true positive rates while keeping extremely low false positive rates. The behavior has been validated by DenseNet201's excellent discriminative power for distinguishing subtle imaging characteristics among benign, malignant and normal CT scans.

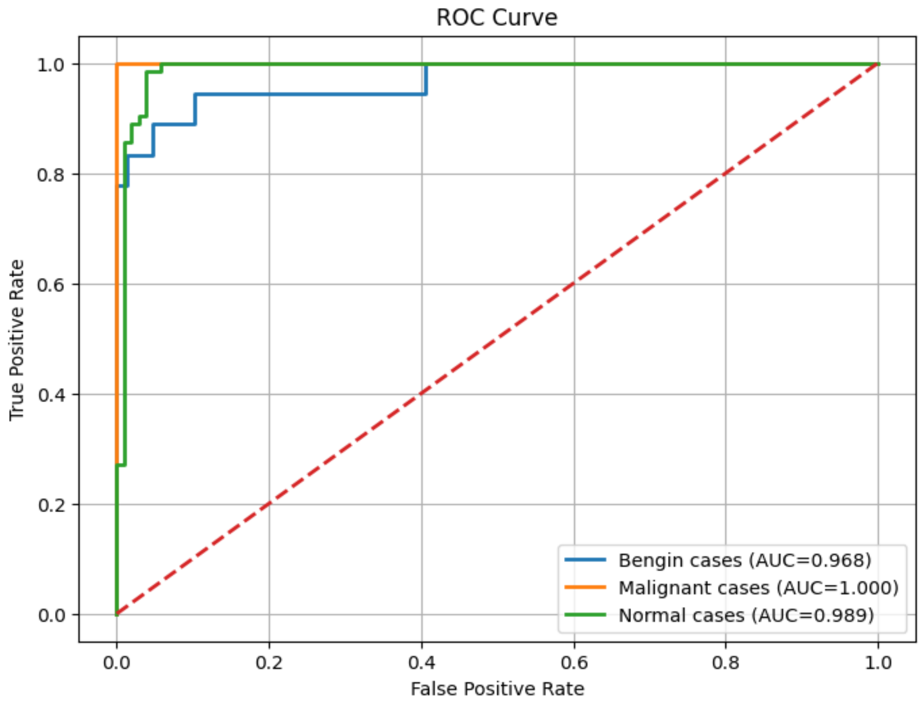


Figure 5. ROC Curve

The excellent classification performance of the proposed framework is further supported by the quantitative AUC values. The benign class yielded excellent AUC value of 0.9685, despite the limited dataset of benign CT images available in the test set due to the small number of CT images available. Complete discrimination between the remaining diagnostic classes and the category of malignant tumors was seen for the malignant category with a perfect AUC of 1.0000. Similarly, the AUC of the normal class was 0.9892, demonstrating the model's proficiency in correctly labelling healthy lung CT images when there is very little overlap. AUCs for all the diagnostic categories were well above the threshold value of 0.5, with an overall macro-average reaching 0.9859, which demonstrated the excellent generalization performance of DenseNet201 over all the diagnostic categories.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Table 3. Receiver Operating Characteristic (ROC) analysis of the proposed DenseNet201 framework.

Class	AUC
Benign Cases	0.9685
Malignant Cases	1.0000
Normal Cases	0.9892
Macro Average	0.9859

The obtained AUC values for all three classes are extremely high, highlighting the high discriminative performance of the proposed framework over a wide range of classification thresholds. The excellent AUC achieved on CT images for the malignant class in particular, underscores the ability of DenseNet201 to effectively classify cancerous regions from non-cancerous tissue, a crucial aspect for minimizing false negative diagnoses in clinical settings. In addition, the high macro-average AUC shows that the proposed framework does not favour any particular class but provides a balanced classification performance across all classes. These results further confirm the success of the transfer learning proposed approach, and make DenseNet201 an extremely reliable backbone for automated multi-class lung cancer classification on CT images.

4.5 Sensitivity and Specificity Analysis

While overall classification accuracy is useful, there is a need for other metrics to evaluate clinical decision support systems that measure the model's accuracy in both diagnosed and non-diagnosed cases. Hence, the sensitivity (TPR) and specificity (TNR) of the proposed DenseNet201 framework were calculated at a class-wise level to check the diagnostic reliability of the framework. The sensitivity measures what % correctly identified positive cases, while the specificity measures what % correctly identified negative cases. In medical image analysis, these metrics are relevant as they directly affect the clinical decision-making process and diagnostic confidence. The results of sensitivity, specificity, precision, and F1-score obtained are tabulated in Table 4.

Table 4. Class-wise sensitivity, specificity, precision, and F1-score of the proposed DenseNet201 framework.

Class	Sensitivity	Specificity	Precision	F1-Score
Benign Cases	0.7778	0.9932	0.9333	0.8485
Malignant Cases	0.9765	1.0000	1.0000	0.9881
Normal Cases	1.0000	0.9515	0.9265	0.9618

As shown in Table 4, the proposed DenseNet201 framework shows excellent diagnostic ability for all three types of lung CT images. This is the highest overall performance of all classes where almost all the CT images of the malignant class were identified correctly (97.65%) and all healthy and benign class were identified correctly (100%). This finding is clinically relevant because it is important to reduce false-negative cancer diagnosis, which would otherwise hinder early cancer treatment and consequently affect patient prognosis.

The results showed that the proposed framework has a sensitivity of 100% for the normal class, indicating that all the normal CT images in the testing dataset were correctly identified. The specificity of the normal class was slightly lower (95.15%), but the value achieved was still excellent, in terms of discrimination between healthy and abnormal lung tissue. The precision and F1-score are also presented, which further validate the consistent high reliability of the proposed classifier to differentiate the normal pulmonary anatomy from pathological cases.

The sensitive scores for the benign class were relatively lower at 77.78% compared to other diagnosis classes. This behavior is similar to the confusion Matrices shown earlier where a few of the benign CT images were classified as normal. This is not surprising since some less severe lung abnormalities may have similar imaging features to the normal lung, which makes them difficult to classify. However, a very high specificity of 99.32% was achieved, meaning that samples of other categories were not often confused with the benign category. The proposed DenseNet201 framework, therefore, succeeds in achieving high diagnostic precision with just a few benign false-negative predictions.

The overall class-wise sensitivity/specificity analysis shows, that the proposed framework classifies the various categories of clinical patients with good class-wise performance with a high level of discrimination of malignant lesions. The results further confirm the appropriateness of DenseNet201 model for the automated classification of lung cancer and also strengthen the proposed use of DenseNet201 model in computer-aided diagnostic systems.

4.6 Comparative Analysis with Baseline Deep Learning Models

To prove the efficacy of the proposed DenseNet201 framework, two popular transfer learning architectures ResNet50 and EfficientNetB0 were used for comparative analysis. Training and testing for all baseline models were done under the same experimental conditions, such as split of data into train/val/test sets, data preprocessing pipeline, data augmentation strategy, optimizer configuration, learning-rate schedule, and evaluation protocol. This experimental design helps isolate the performance differences into the feature extraction ability of each of the backbones without the influence of differences in training.

The results of the performance of the models evaluated are presented in Table 5.

Table 5. Comparative performance of different deep transfer learning models.

Model	Accuracy	Precision	Recall	F1-Score
ResNet50	87.95%	88.46%	87.95%	88.04%
EfficientNetB0	72.89%	74.95%	72.89%	73.64%
Proposed DenseNet201	96.39%	96.49%	96.39%	96.30%

The results clearly show that the proposed DenseNet201 framework outperforms both baselines. DenseNet201 outperformed ResNet50 (87.95%) and EfficientNetB0 (72.89%) with the best classification accuracy of 96.39%. It is observed that the improvements are consistent across all evaluation metrics – precision, recall and F1-score – suggesting the proposed framework achieves a balanced classification performance without any specific advantage to one metric over another.

The dense connectivity mechanism enables each convolutional layer to have access to feature information from all the previous layers, which could be the reason for the superior performance of DenseNet201. This architecture enables effective reuse of features, better gradient propagation in the optimization process, and maintains lower-level texture details and higher-level semantic representations. In this way, DenseNet201 is able to detect subtle anatomical variations and lesion characteristics better than traditional convolutional networks, which in turn benefits the discrimination between benign, malignant, and normal lung CT images.

While ResNet50 has achieved good classification accuracy, its residual learning method relies mostly on additive shortcut connections, thus offering relatively less reusability of features than DenseNet201. Consequently, ResNet50 achieved poor classification performance with poor discriminative ability for complex lung patterns in the lung CT images. EfficientNetB0 had the worst overall scores when tested against the models. Although EfficientNetB0 is very efficient for computation, its light architecture does not allow it to learn from the various imaging characteristics found in the IQ-OTHNCCD dataset to learn highly discriminative features.

The positive trend seen in all the evaluation metrics further validates the proposed DenseNet201 framework to be a more effective feature representation for multi-class lung cancer classification than the selected baseline transfer learning models. Moreover, the significant performance difference when taking the same experimental conditions, shows that the proposed architecture possesses a better generalization ability and an equal performance under different diagnostic categories. These results

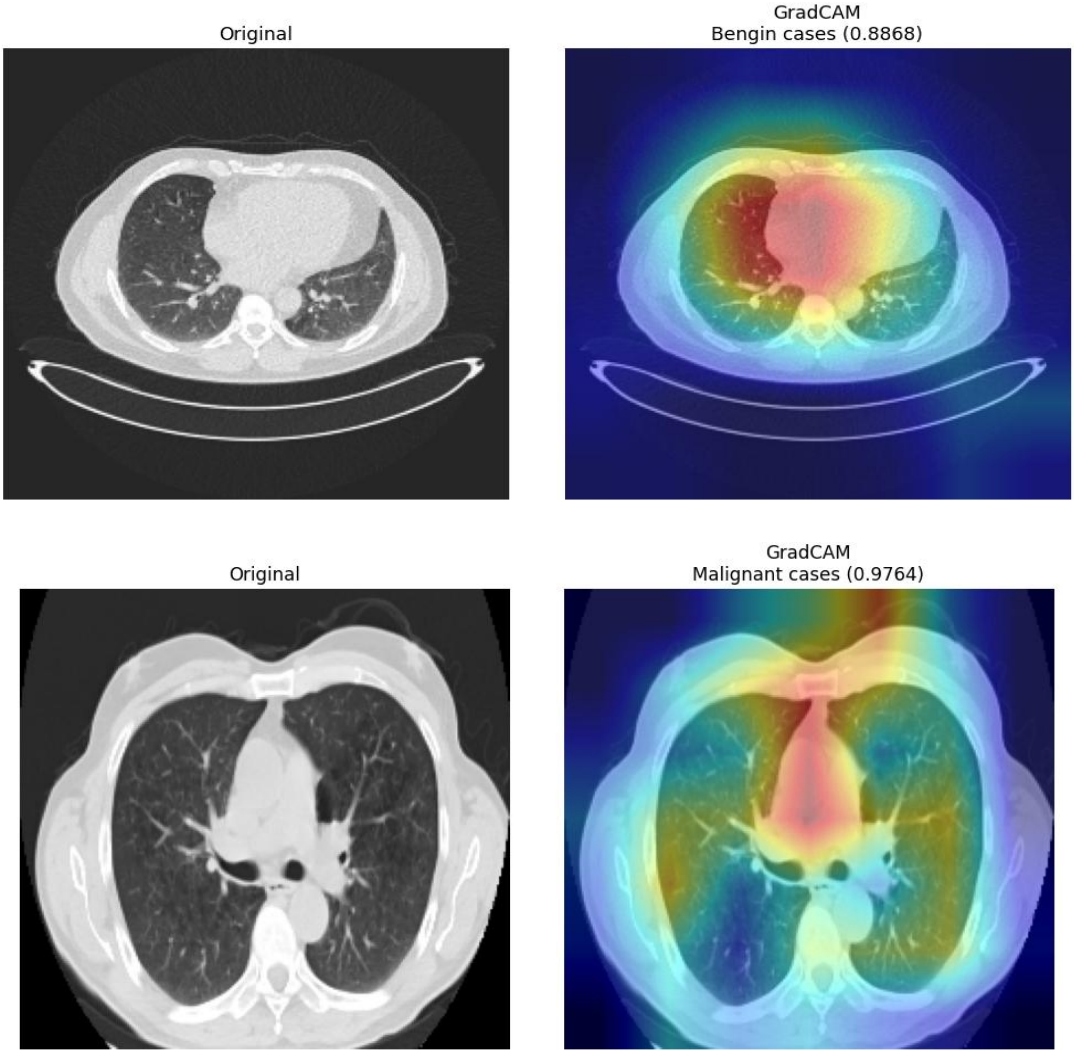
Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

show that DenseNet201 is the most appropriate backbone network of the transfer learning architecture evaluated for automated classification of lung cancer from computed tomography (CT) images.

4.7 Explainability Analysis Using Gradient-weighted Class Activation Mapping (Grad-CAM)

Deep convolutional neural network (CNN) has been widely used to achieve outstanding performance on medical image classification, but the mechanisms behind its classification are considered as a "black box" and therefore hinder the use of the CNN in clinical practice. To give DenseNet201 framework a transparency and interpretability, the Gradient-weighted Class Activation Mapping (Grad-CAM) method was used as a post hoc explainability technique. Grad-CAM is based on gradients of the predicted class with respect to the final convolutional feature maps of the network to produce visual attention maps. These attention maps indicate which parts of the image are most relevant to making the classification, providing a way to learn how the model arrived at its predictions.

The representative examples of the DenseNet201 framework-generated visualizations are shown in Fig. 6. The figure shows the original CT image and the heatmap created by Grad-CAM overlaid on the image. Regions with higher temperature colors (red and yellow) show that the model paid more attention to the prediction of those regions, while cooler colors (blue and green) show that the model paid relatively less attention to those regions.



Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

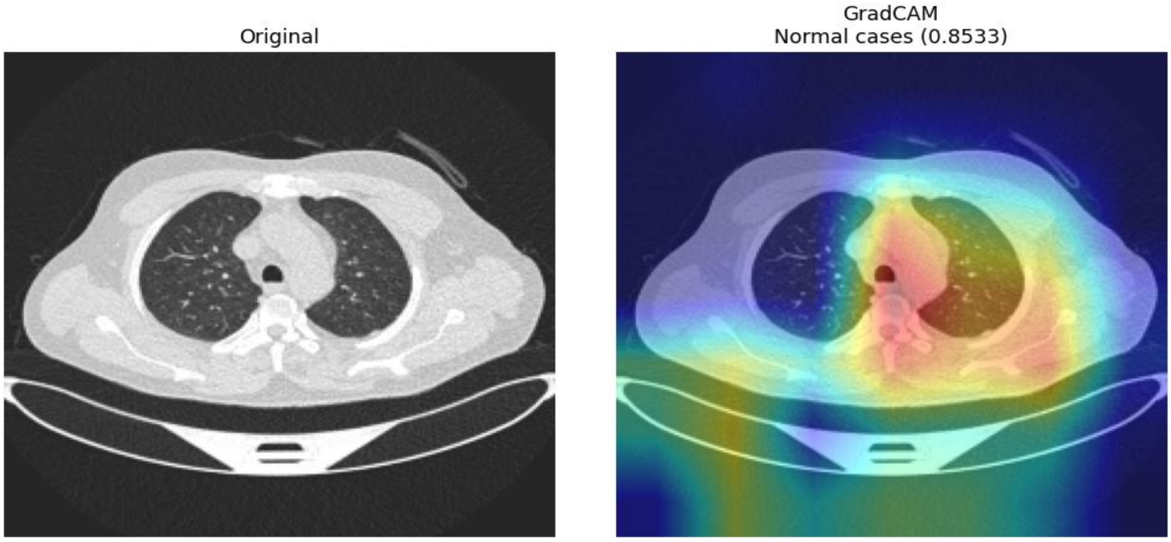


Figure 6. Representative Grad-CAM visualization generated by the proposed DenseNet201 framework for lung CT image classification.

The Grad-CAM visualization reveals that the DenseNet201 model mainly focuses on clinically relevant thoracic regions, instead of spreading its attention over the whole CT image, during the classification. The highlighted activation regions show that the network is able to learn discriminative structural information in the thoracic anatomy that is useful for its diagnostic decision. The localized attention patterns also show that the proposed model does not solely use a global image feature, but also uses a feature that is meaningful at the local level in imaging that is learned during transfer learning and fine-tuning.

Grad-CAM is incorporated to greatly improve interpretability of the proposed framework by visual evidence of the internal decision-making process. AI systems, especially in the field of computer-aided diagnosis, can be particularly beneficial, as it can boost the confidence of the clinicians, aid in qualitative validation of the models' predictions, and help in the transparent use of artificial intelligence for medical diagnosis. Therefore, the combination of Grad-CAM enhances the usefulness of the proposed DenseNet201 framework for better supporting radiologists in the context of lung cancer screening and diagnosis.

4.8 Discussion

The experimental results show that, the proposed DenseNet201 based Transfer learning framework is highly reliable for automatic multi-class lung cancer classification using CT images. To provide efficient features propagation as a dense network, DenseNet201 is used as the backbone network to preserve low-level and high-level features during training. Because of this architectural feature, it brings about better feature reusability, better gradient flow and better representation learning, which are more discriminative than the typical CNN architecture.

The training and validation learning curves show a consistent optimization behavior across training, suggesting good generalization performance without significant overfitting or convergence issues. The proposed framework's generalization capability was enhanced by a combination of the above mentioned techniques which minimized the contribution of unnecessary model complexity. Such optimisation strategies made it possible to learn optimally the features learnt from the ImageNet data to suit the characteristics of lung CT data, which enabled the network to adapt well to lung CT image data without compromising the network's predictive stability.

Finally, quantitative assessment of the proposed framework was comprehensive, which further verified the robustness of the proposed framework. An accuracy value of 0.93, precision of 0.94, and recall of 0.96 indicate that the model has excellent classification performance for most diagnostic categories.The accuracy of 0.93, precision of 0.94 and recall of 0.96 suggest that the model has good classification ability in most diagnostic categories. Confusion matrix analysis showed that most of the images in the CT images were correctly classified and ROC analysis showed that the images had good discriminative power with consistency high AUC values. In a similar way, the evaluation of sensitivity and specificity showed good detection of malignant lesions as well as good discrimination of healthy lung tissue, thus demonstrating the potential of the proposed framework for clinical computer-aided diagnosis.

For the comparative experiments, these further confirmed the effectiveness of DenseNet201. When implementing the same preprocessing, training strategy, and same evaluation conditions, the proposed framework always surpasses both ResNet50 and EfficientNetB0 on all the evaluation metrics. The results showed that as DenseNet201 has a dense connectivity mechanism, allowing for the reusing of a large number of features and retaining significant diagnostic information along the network, this method offers better feature extraction capability for lung CT image classification. As a result, the proposed architecture has better generalization and robustness to classification than the selected baseline transfer learning models.

The Grad-CAM visualizations also give qualitative evidence for the understandability of the proposed framework. Grad-CAM provides the clinicians with an improved understanding of how the network arrived at the classification decision, and adds to the confidence the network could have in making its automated diagnosis. The explainability of these systems is also acknowledged as crucial for the trustworthiness of AI systems in healthcare due to its transparency and its ability to enable clinical validation of the model's predictions.

Although good performance has been obtained in this study, there are some limitations to be noted. The proposed framework was tested on one publicly available dataset but further testing of the framework on larger multi-institutional datasets would enhance the generalizability of the framework to various imaging protocols and patient cohorts. In addition, although transfer learning significantly enhances classification accuracy, future research could explore modalities of transfer learning that combine clinical data, radiological reports, or genomic biomarkers with CT imaging data. The future of automated lung cancer diagnosis systems could be further improved by leveraging transformer-based architectures, self-supervised learning, federated learning, and advanced explainable artificial intelligence techniques to boost their robustness, interpretability, and clinical utility.

The experimental results show that the proposed DenseNet201 based framework is an accurate, robust and interpretable deep learning approach for multi-class lung cancer classification from CT images. An integrated system based on the proposed framework, with high predictive performance, complete quantitative evaluation, superiority in comparison with current transfer learning architectures, and improved explainability, has a great potential for future application in a computer-aided diagnosis system to support the radiologist in early detection and clinical decision making when treating lung cancer.

5. Conclusion

The objective of this study was to develop a DenseNet201-based deep transfer learning system to classify lung cancer in computed tomography (CT) images from the publicly available IQ-OTHNCCD dataset into multiple classes. The proposed framework combined image preprocessing, data augmentation, transfer learning, two stage fine-tuning, and a customized computer-aided diagnostic head to build an efficient computer-aided diagnostic system for distinguishing benign, malignant and normal lung computed tomography (CT) images. The proposed framework was able to extract robust features by combining the ImageNet-pretrained DenseNet201, batch normalization, dropout regularization, label smoothing, adaptive learning-rate scheduling, and early stopping to ensure the stability of the model convergence and prevent overfitting.

The proposed approach was evaluated through comprehensive experiments across various quantitative performance metrics and was shown to be effective. Overall, DenseNet201 framework demonstrated a high classification accuracy of 96.39% while retaining a precision of 96.49%, a recall of 96.39% and an F1-score of 96.30% which shows that the framework has good predictive power in the classification of lung cancer into its classes. Receiver Operating Characteristic (ROC) curve also reaffirmed the discriminative capability of the model with macro-average Area Under Curve (AUC) of 0.9859 and class-wise sensitivity and specificity analysis showed good detection performance in all the diagnostic categories. Comparative experiments under the same training conditions proved the performance of the proposed DenseNet201 framework to be more superior in feature extraction ability than ResNet50 and EfficientNetB0, which shows that the dense connectivity has better feature extraction ability in lung CT image analysis. In addition, Grad-CAM explainability enabled visualization of regions of the image that mattered for the model's prediction and thus enhanced transparency and interpretability of the proposed framework for potential clinical application.

Based on the results of this study, DenseNet201 proves to be an accurate, robust and interpretable method for automated lung cancer classification and shows great promise for incorporation into intelligent computer-aided diagnostic systems to assist radiologists in early detection of lung cancer and clinical decision making. However, there are certain limitations with the present study. This proposed framework was tested with one publicly available dataset, and should be tested with larger multi-center and multi-institutional datasets collected from a variety of imaging environments. Future research will involve external clinical validation of the proposed framework, multimodal data fusion that includes radiological and clinical data, hybrid architectures based on transformer models, self-supervised learning, federated learning for collaborative diagnosis while maintaining privacy, and the development of more sophisticated explainable artificial intelligence (XAI) techniques. In the future, these advancements will likely enhance the clinical relevance, generalizability and robustness of AI-driven deep learning systems in early lung cancer diagnosis and precision healthcare.

6. References

[1] Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. *Future generation computer systems*, *92*, 374-382.

[2] Wang, L. (2022). Deep learning techniques to diagnose lung cancer. *Cancers*, *14*(22), 5569.

[3] Asuntha, A., & Srinivasan, A. (2020). Deep learning for lung Cancer detection and classification. *Multimedia Tools and Applications*, *79*(11), 7731-7762.

[4] Javed, R., Abbas, T., Khan, A. H., Daud, A., Bukhari, A., & Alharbey, R. (2024). Deep learning for lungs cancer detection: a review. *Artificial Intelligence Review*, *57*(8), 1.

[5] Chaunzwa, T. L., Hosny, A., Xu, Y., Shafer, A., Diao, N., Lanuti, M., ... & Aerts, H. J. (2021). Deep learning classification of lung cancer histology using CT images. *Scientific reports*, *11*(1), 5471.

[6] Kriegsmann, M., Haag, C., Weis, C. A., Steinbuss, G., Warth, A., Zgorzelski, C., ... & Kriegsmann, K. (2020). Deep learning for the classification of small-cell and non-small-cell lung cancer. *Cancers*, *12*(6), 1604.

[7] Mohamed, T. I., & Ezugwu, A. E. S. (2024). Enhancing lung cancer classification and prediction with deep learning and multi-omics data. *IEEE Access*, *12*, 59880-59892.

[8] Riquelme, D., & Akhloufi, M. A. (2020). Deep learning for lung cancer nodules detection and classification in CT scans. *Ai*, *1*(1), 28-67.

[9] Sangeetha, S. K. B., Mathivanan, S. K., Karthikeyan, P., Rajadurai, H., Shivahare, B. D., Mallik, S., & Qin, H. (2024). An enhanced multimodal fusion deep learning neural network for lung cancer classification. *Systems and Soft Computing*, *6*, 200068.

[10] Kanavati, F., Toyokawa, G., Momosaki, S., Rambeau, M., Kozuma, Y., Shoji, F., ... & Tsuneki, M. (2020). Weakly-supervised learning for lung carcinoma classification using deep learning. *Scientific reports*, *10*(1), 9297.

[11] Pandit, B. R., Alsadoon, A., Prasad, P. W. C., Al Aloussi, S., Rashid, T. A., Alsadoon, O. H., & Jerew, O. D. (2023). Deep learning neural network for lung cancer classification: enhanced optimization function. *Multimedia Tools and Applications*, *82*(5), 6605-6624.

[12] Singh, G. A. P., & Gupta, P. K. (2019). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. *Neural Computing and Applications*, *31*(10), 6863-6877.

[13] Ren, Z., Zhang, Y., & Wang, S. (2022). A hybrid framework for lung cancer classification. *Electronics*, *11*(10), 1614.
[14] Podolsky, M. D., Barchuk, A. A., Kuznetcov, V. I., Gusarova, N. F., Gaidukov, V. S., & Tarakanov, S. A. (2016). Evaluation of machine learning algorithm utilization for lung cancer classification based on gene expression levels. *Asian Pacific journal of cancer prevention*, *17*(2), 835-838.
[15] Raza, R., Zulfiqar, F., Khan, M. O., Arif, M., Alvi, A., Iftikhar, M. A., & Alam, T. (2023). Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. *Engineering Applications of Artificial Intelligence*, *126*, 106902.
[16] Chen, C. L., Chen, C. C., Yu, W. H., Chen, S. H., Chang, Y. C., Hsu, T. I., ... & Chen, C. Y. (2021). An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning. *Nature communications*, *12*(1), 1193.
[17] Lynch, C. M., Abdollahi, B., Fuqua, J. D., De Carlo, A. R., Bartholomai, J. A., Balgemann, R. N., ... & Frieboes, H. B. (2017). Prediction of lung cancer patient survival via supervised machine learning classification techniques. *International journal of medical informatics*, *108*, 1-8.