

# Hybrid model detection and classification of lung cancer

Rami Yousef<sup>1</sup>, Eman Yaser Daraghmi<sup>2</sup>

<sup>1</sup>Department of Computer Systems Engineering, Palestine Technical University–Kadoorie, Tulkarm, Palestine

<sup>2</sup>Department of Computer Science, Palestine Technical University–Kadoorie, Tulkarm, Palestine

## Article Info

### Article history:

Received Jun 21, 2024

Revised Nov 4, 2024

Accepted Nov 14, 2024

### Keywords:

Artificial intelligence  
Convolutional neural network  
Ensemble learning  
Lung cancer  
Transfer learning

## ABSTRACT

Lung cancer ranks among the most prevalent malignancies worldwide. Early detection is pivotal to improving treatment outcomes for various cancer types. The integration of artificial intelligence (AI) into image processing, coupled with the availability of comprehensive historical lung cancer datasets, provides the chance to create a classification model based on deep learning, thus improving the precision and effectiveness of detecting lung cancer. This not only aids laboratory teams but also contributes to reducing the time to diagnosis and associated costs. Consequently, early detection serves to conserve resources and, more significantly, human lives. This study proposes convolutional neural network (CNN) models and transfer learning-based architectures, including ResNet50, VGG19, DenseNet169, and InceptionV3, for lung cancer classification. An ensemble approach is used to enhance overall cancer detection performance. The proposed ensemble model, composed of five effective models, achieves an F1-score of 97.77% and an accuracy rate of 97.5% on the IQ-OTH/NCCD test dataset. These findings highlight the effectiveness and dependability of our novel model in automating the classification of lung cancer, outperforming prior research efforts, streamlining diagnosis processes, and ultimately contributing to the preservation of patients' lives.

This is an open access article under the [CC BY-SA](#) license.



## Corresponding Author:

Rami Yousef

Department of Computer Systems Engineering, Palestine Technical University

Jaffa Street, Tulkarm, Palestine

Email: r.yousuf@ptuk.edu.ps

## 1. INTRODUCTION

Cancer is a disease that manifests in many ways and is mostly linked to aberrant cell populations. These cancer cells keep dividing and expanding to become tumors. Lung cancer is cancer that poses the greatest risk to human life globally. As per the World Health Organization [1], lung cancer is the leading cause of mortality worldwide. In 2008, lung cancer accounted for 1.37 million deaths globally [2]. Based on the available data, lung cancer comprises the majority of new cancer diagnoses worldwide, with 1,350,000 new cases, representing 12.4% of all new cancer cases. Additionally, it constitutes the majority of cancer-related fatalities, with 1,180,000 deaths, accounting for 17.6% of all cancer deaths [3]. Lung cancer ranked first among causes of death for men and third among causes of death for women in the Global Cancer Observatory database created by the International Agency for Research on Cancer (IARC) in 2018. The database encompassed rates of both incidence and mortality for 36 cancer types across 185 countries. Nearly 1.8 million fatalities from cancer were recorded in 2018, accounting for about 18.4% of all cancer-related deaths [4]. Due to the alarming increase in lung cancer fatalities and the disease's excessively high incidence by nature, many studies focusing on cancer control and prompt identification methods have emerged to reduce mortality.

The potential of a successful cure for lung cancer is contingent upon the timely detection of the disease and the accuracy of diagnostic procedures. Effective diagnostic methods contribute to a reduction in the incidence of lung cancers, and the early identification of the ailment is fundamentally imperative for achieving a favorable prognosis in lung cancer treatment. Presently, there exist seven modalities for the management of lung cancer, which encompass cytology sputum and breath analyses, positron emission tomography (PET), magnetic resonance imaging (MRI), and chest radiographs (CXRs) [2]. It is noteworthy that CXRs, sputum cytology, and computed tomography (CT) scans entail exposure to ionizing radiation, while MRI and PET scans impose certain limitations on the precise identification and staging of lung cancer. It is essential to acknowledge that these diagnostic techniques are not without their inherent limitations. Furthermore, it is essential to consider that the administration of a serum test is an invasive medical procedure, and its limited capacity for early detection sensitivity and specificity renders it unsuitable as a primary diagnostic tool [5]. Conversely, the assessment of sputum necessitates additional investigation due to the presence of gene promoter methylation, as indicated by a study on [www.ijeec.ir](http://www.ijeec.ir) [5]. Despite this need for additional scrutiny, sputum analysis shows potential to facilitate timely identification of lung cancer. Additionally, volatile organic compounds (VOCs) detected in urine have demonstrated noteworthy sensitivity and specificity, although a larger sample size is necessary for more robust results [5]. Conversely, chest X-rays (CXR) exhibit relatively low sensitivity and are prone to producing false-negative outcomes, as reported in previous studies [6], [7]. Presently, the most dependable approach to detecting lung cancer is the utilization of CT imaging. This imaging modality offers precise information regarding the location and size of pulmonary nodules, enabling the early detection of cancerous growths. Low-dose CT screening has proven effective in identifying early-stage cancer tumors, resulting in a notable 20.0% reduction in mortality when compared to conventional radiographic techniques, and an increased rate of positive screening results [8].

Deep learning is a specialized branch of machine learning (ML), which itself falls within the larger field of artificial intelligence (AI). The overarching objective of AI is to furnish a collection of algorithms and methodologies designed to address problems that humans effortlessly and intuitively undertake but pose significant computational challenges. ML, as a discipline, is harnessed for the purpose of pattern recognition, with deep learning comprising a category of ML algorithms conceived by drawing inspiration from the structural and operational principles of the human brain. Deep learning endeavors to emulate the human perceptual process by establishing artificial neurons or nodes within layered architectures, which are capable of feature extraction from objects. This implies that when applied to image classification tasks, deep learning aims to discern patterns from a set of images for the purpose of distinguishing between diverse classes or objects. Significantly, the neural network's training process involves the automatic extraction of image features [9].

In the medical domain, specifically within the context of lung cancer diagnosis, the principal diagnostic technique relies on the examination of tissue samples. However, it is worth acknowledging that this diagnostic procedure entails a time-consuming process. Utilizing an array of deep learning models and transfer learning-based models, including ResNet50, EfficientNetB7, DenseNet169, VGG16, VGG19, Xception, and InceptionV3, applied to the lung cancer dataset IQ-OTH/NCCD, we introduce a deep learning methodology in this research to train and test the models at: <https://www.kaggle.com/code/kerneler/starter-the-iq-oth-nccd-lung-cancer-09c3a8c9-4/data>.

## 2. RELATED WORK

The identification of pulmonary irregularities constitutes a substantial hazard to human well-being, and the timely recognition thereof assumes a pivotal role in risk mitigation. Timely diagnosis facilitates expeditious and efficacious intervention, thereby reducing potential complications and enhancing patient outcomes. Among the diagnostic modes, CT emerges as a noteworthy tool for detecting pulmonary abnormalities. Nevertheless, the interpretation of lung CT scans presents challenges, even for seasoned radiologists [10].

Over the past few years, investigators have delved into the application of deep learning methodologies to automate the diagnostic process for pulmonary irregularities, aiming to enhance diagnostic precision and potentially save lives. As an illustration, Asuntha and Srinivasan [11] introduced a pioneering approach termed fast and power-efficient system-on-chip convolutional neural network (FPSOCNN), showcasing the ongoing exploration of innovative methodologies in this domain. The FPSOCNN developed by [11] seeks to alleviate the computational intricacies inherent in conventional CNNs. In their study, they examined various feature extraction techniques, including Zernike moment, histogram of oriented gradients (HoG), wavelet transform-based features, local binary pattern (LBP), wavelet transform-based features, and scale invariant feature transform (SIFT). The proposed FPSOCNN methodology not only demonstrated outstanding achievement but also effectively addressed the computational complexities associated with traditional CNNs. In a separate study by [12] a multi-path CNN was introduced, leveraging contextual features, both local and broad, to identify lung cancer autonomously. By integrating this method, their model

demonstrates increased adaptability in handling variations in nodule size and shape. This characteristic leads to enhanced detection results compared to modern state-of-the-art techniques [12]. A comprehensive demonstration of the use of computer-aided diagnosis (CAD) methods for the identification of early-stage lung cancer is presented in the study by [13]. Convolutional neural networks (CNNs), among various deep learning approaches, have been extensively applied to the tasks involving computer vision. The authors emphasize the superiority of 3-dimensional CNNs over 2-dimensional CNNs for improved effectiveness in detecting lung cancer. In contrast, Shyni and Chitra [14] highlight the widespread use of CNN as the primary deep learning algorithm for detecting COVID-19 from medical images. These articles not only promote the extensive adoption of CNN but also provide valuable insights, inspiring emerging researchers to create highly effective CNN models that use medical images to detect diseases early. In a distinct investigation, Rahman *et al.* [15] conducted a study utilizing CNN for tasks related to classification that involve two, three, and multiple classes. They employed electrocardiogram (ECG) signals as input, achieving promising outcomes in their experiments. Notably, they used the Grad-CAM method to pinpoint critical and particular areas in the input signals, thereby facilitating informed decision-making during the classification process [15]. Gifani *et al.* [16] conducted research wherein they devised an assemblage deep learning model for the automated identification of COVID-19 from CT scans. Employing CNN alongside transfer learning techniques, the researchers integrated 15 pre-trained CNN models, leveraging the collective expertise and capabilities. The ensemble access resulted in enhanced robustness and accuracy in the identification of COVID-19 from CT scans. In their study, Dorj *et al.* [17] concentrated on lung cancer classification using error-correcting output codes support vector machine (ECOC SVM) and deep convolutional neural network (DCNN). An ECOC SVM classifier was employed for categorizing several types of lung cancer, with the algorithm applied to a dataset comprising 3753 images representing four lung cancer types. The implementation yielded notable accuracy, sensitivity, and specificity, with peak values reported as 94.17% for squamous cell carcinoma, 98.9% for actinic keratosis, and, 95.1% for squamous cell carcinoma, respectively. A DCNN model was proposed by [18], employing a deep learning approach for precise classification between benign and malignant lung lesions. With a testing accuracy of 91.93% and a training accuracy of 93.16%, the evaluation of the HAM10000 dataset revealed remarkable results. These results highlight how well the DCNN model distinguishes benign from malignant lung lesions.

Pal *et al.* [19] created an interpretable ML model for lung cancer detection called AI CAD. Integrating explainable artificial intelligence (XAI) mechanisms, the model furnishes comprehensive explanations for crucial features identified by the AI/ML algorithms. Encouraging confidence in its application, the model of AI CAD demonstrated excellent interpretation in detecting lung cancer. The models were constructed using diverse algorithms, including k-nearest neighbors (KNN), support vector machine (SVM), gradient boosting machine (GBM), XGBoost, random forest classifier (RFC), and feed-forward architectures. The interpretability of models with superior performance in neural networks was highlighted through XAI outputs, revealing that these models prioritized a consistent set of input features with elevated importance.

An interpretable system for diagnosing lung cancer was created in the work of [20] utilizing numerous models of ML, including naive Bayes classifier, logistic regression, decision tree, and random forest. Based on a dataset of lung cancer cases, the analysis yielded a 97% accuracy rate. However, it is noteworthy that the validation of the model was restricted to a CSV file, lacking image validation. In order to increase interpretability, XAI techniques like lime and SHAP were used. Bhandari *et al.* [21] introduced a deep learning model that was validated using a dataset to predict four categories comprising visuals of 7132 CXR. The model, leveraging Grad-CAM, SHAP, and LIME methods for interpretation through 10-fold cross-validation, achieved an average of 94.54% ( $\pm 1.33\%$ ) validation accuracy and 94.31% ( $\pm 1.01\%$ ) of test accuracy. To determine biomarkers in non-small-cell lung carcinoma (NSCLC) subtypes, Dwivedi *et al.* [22] suggested a framework of XAI-based deep learning. The framework incorporated an autoencoder, a biomarker discovery module, and a classification neural network. 52 related biomarkers were discovered using XAI techniques; of these, 14 were druggable and 28 were survival predictive. 95.74% accuracy in NSCLC subtype classification was attained with the multilayer perceptron. We recommend utilizing CNN models in our approach and several transfer learning-based architectures, including ResNet50, VGG19, DenseNet169, and InceptionV3, for lung cancer classification tasks. Additionally, we employ an ensemble approach by combining various model combinations to enhance the overall performance of cancer detection.

### 3. DATASET

To assess the effectiveness of our proposed model, we utilized the IQ-OTH/NCCD lung cancer dataset [23] which comprises 2073 CT images from 110 patients, including both those in good health and those who have been diagnosed with lung cancer. This dataset was gathered at the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases during three months in 2019. A Siemens scanner in DICOM

format was used to obtain the CT scans and each scan contained 80–200 slices with a 1mm slice thickness. In order to guarantee diversity, the dataset was collected from diverse regions in Iraq, which represent a range of demographics. The study was approved ethically by the institutional review board, ensuring the rights and privacy of participants are upheld during the process of collecting data.

#### 4. METHOD

The thorough methodology employed in this study is described in this section. Figure 1 shows the schematic workflow of our procedure, giving detailed steps of the experimental procedure. Several preprocessing techniques are compared to structural data and taken into consideration when discussing image data. By scaling data and reducing computational complexity, these techniques help to decrease modeling costs while increasing the quantity or quality of the dataset. Since the use of deep learning has increased, scientists have worked on this problem, and nowadays there are a variety of approaches to dealing with this aspect. Preprocessing procedures are used in this work to get the data in Figures 2(a) and 2(b) ready for use in our deep learning models. The following actions were required to guarantee the data's compatibility and enhance its quality:

- Data augmentation: to address the imbalanced nature of the dataset, data augmentation is employed. This involves the creation of new images using techniques such as cropping, rotating, flipping, and zooming. In this study, a two-phase augmentation approach is implemented. Initially, various techniques, including rotating, zooming, random distortion, contrast and brightness adjustments, random cropping, and flipping, are applied to images in the malignant class across the training, validation, and test sets. Subsequently, random flipping in both horizontal and vertical directions is applied to images in both classes, ensuring a diverse set of images to aid model generalization.
- Image size standardization: the chosen CNN architecture can accommodate images of varying sizes; however, for transfer learning, it is advisable to align with pre-trained models that typically process images in  $224 \times 224$  dimensions. Therefore, the image size is reduced to  $224 \times 224$ , enabling a direct comparison with pre-trained models and reducing computational overhead. Subsequently, normalization is applied to further streamline computational complexity.
- Data normalization: in order to prepare datasets for further analysis and modeling, data normalization is an essential step. A spectrum of techniques is available for normalization, encompassing min–max normalization, z-score normalization, and decimal scaling normalization [24]. Data normalization is primarily used to improve data quality, make data comparable across various records and fields, and improve entry type uniformity and consistency.
- Conversion of data: the LIDC-IDRI dataset's CT scans were converted from their original DICOM format to the more commonly used NIfTI format. This conversion was imperative to ensure harmonization between our deep learning framework and the data, facilitating smooth integration and processing [25].

In the realm of modeling, two approaches are available: the implementation of a CNN architecture from scratch or the utilization of transfer learning. The former is typically favored when an ample amount of data is available, ensuring that the risk of overfitting is minimized. Conversely, the latter is employed in scenarios characterized by data scarcity or the imperative to reduce modeling costs. Transfer learning leverages pre-trained models, benefiting from weights derived from extensive training on diverse datasets encompassing thousands of classes, thereby reducing the requisite number of training epochs.

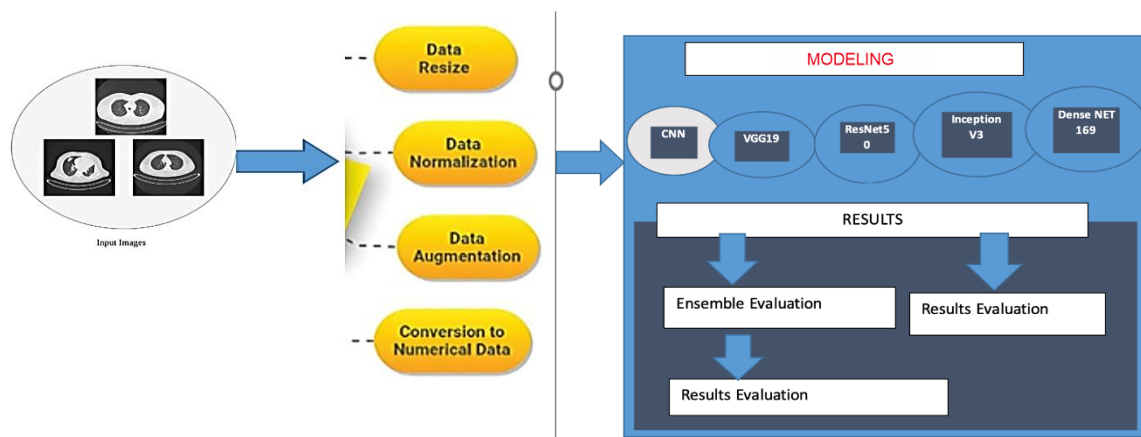


Figure 1. The proposed methodology

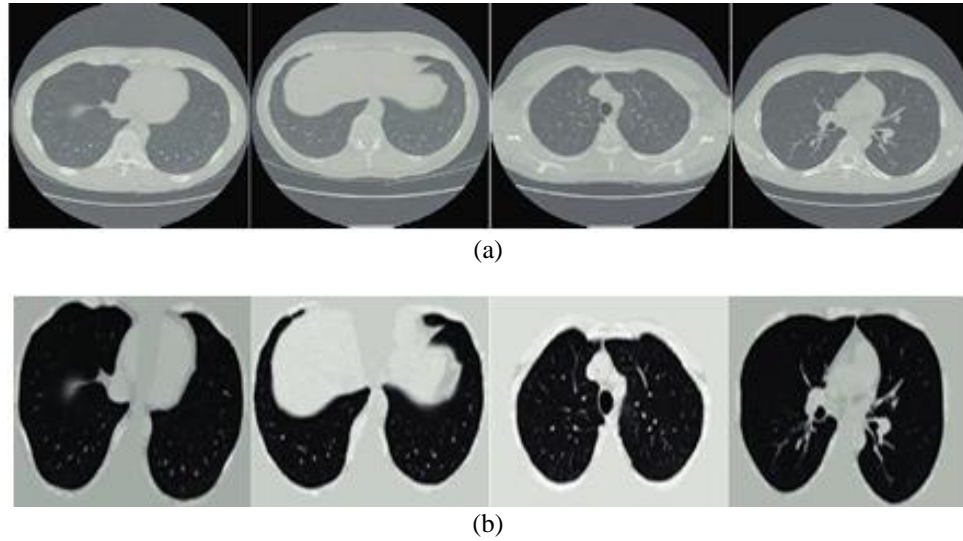


Figure 2. Diagnostic imagery: (a) input CT images and (b) preprocessed images

Since the advent of AlexNet in 2012, which demonstrated superior accuracy over traditional ML algorithms, CNNs have become pervasive. This deep feed-forward architecture finds wide applications in image classification, image segmentation, and object detection. In medical imaging datasets, where precision is crucial for human life, conventional ML algorithms often fall short. Subsequent to the introduction of AlexNet, substantial research efforts have been dedicated to innovatively investigating and implementing CNNs, resulting in models exhibiting superior performance even in comparison to AlexNet.

A standard CNN model comprises one or few convolutional and pooling layers for feature extraction as shown in Figure 3, culminating in one or more fully connected layers for generating classified outputs. To extract features from every visual, the convolutional layer convolves it using a learnable kernel. The kernel, represented by a matrix of discrete weights, is initialized randomly and updated iteratively to minimize errors. The stride parameter governs the kernel's movement through the image, with values updated through a computational process. The output of a convolutional layer, known as a feature map, is subsequently forwarded to the next layer as input.

The output size ( $h_{new} \times w_{new} \times d$ ) of a convolutional layer is computed using the formula:

$$h_{new} = \frac{h - f + 2p}{s} + 1 \quad (1)$$

$$w_{new} = \frac{w - f + 2p}{s} + 1 \quad (2)$$

Where  $h$  and  $w$  are the height and width of the input image, respectively.  $f$  is the size of the filter (or kernel).  $p$  is the amount of zero-padding applied to the input image (if any).  $s$  is the stride of the convolution.

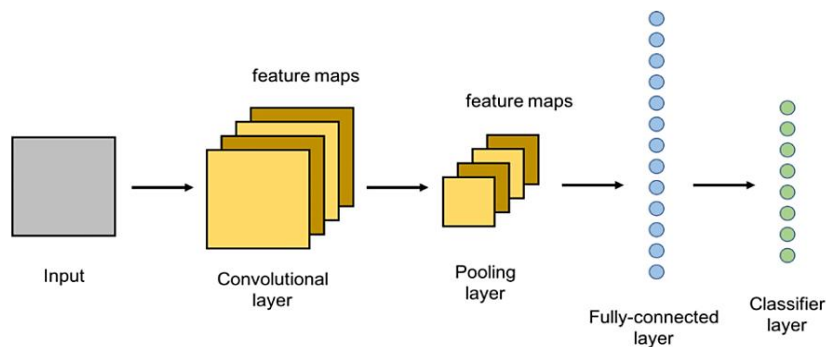


Figure 3. Sample convolutional neuron network

As the depth of a CNN increases, more details can be extracted, resulting in increased accuracy. However, a trade-off exists, as deeper architectures demand more computational processes, thereby incurring higher costs. The optimal number of layers necessitates careful consideration, as excessive depth may not always translate to improved performance compared to shallower architectures.

In the subsequent stages of neural network operations, error calculation becomes imperative. Post-output generation, the loss function is employed to compare estimated labels with true labels, facilitating error assessment. Common loss functions include cross entropy, euclidean, and hinge. Weight updates for subsequent epochs are orchestrated by optimizer functions, with Adam being a widely adopted choice. This iterative process is repeated across epochs, allowing for error comparison with the best previous epoch and saving model improvements when observed. An illustrative figure of a CNN with two hidden layers, featuring kernel sizes of  $(9 \times 9 \times 1)$  and  $(5 \times 5 \times 4)$  in the first and second layers, respectively, is provided for reference.

## 5. TRANSFER LEARNING

Deep learning algorithms, notably CNNs, demonstrate exceptional performance when confronted with a substantial volume of images per class. However, the computational demands imposed by extensive layer usage can result in protracted training times, spanning days or weeks on contemporary hardware. This becomes particularly impractical for numerous problem domains, especially those pertaining to medical applications. Consequently, adopting a transfer learning paradigm, wherein precomputed weights from a pre-trained model on analogous data are repurposed, proves advantageous in terms of cost efficiency. This approach leverages knowledge gleaned during the prior training of a model on comparable datasets. For example, a model initially trained for histopathology image-based cancer diagnosis can be repurposed for lung cancer diagnosis by transferring the acquired weights.

Transfer learning finds utility across multiple domains, including natural language processing (NLP), sound, image, and video processing. In the context of CNNs, the training process begins by discerning image edges and borders, progressing to shape identification. Deeper layers capture increasingly intricate details, culminating in attempts to classify images based on assigned labels. Subsequently, employing transfer learning involves utilizing the initial layers responsible for feature extraction while retraining the model's latter layers with new data. In this study, CNN models and several transfer learning-based architectures, including ResNet50, VGG19, DenseNet169, and InceptionV3.

### 5.1. ResNet50

The architecture is separated into two blocks, each with a skip connection. The output of the previous block is summed with the current block's output, and the activation function is applied to  $f(x) + x$  rather than just  $f(x)$ . This generates the block's output, which is then forwarded to the subsequent block. In ResNet 50, three convolutional layers are utilized with a specific configuration, as illustrated in Figure 4, instead of the typical two convolutional layers.

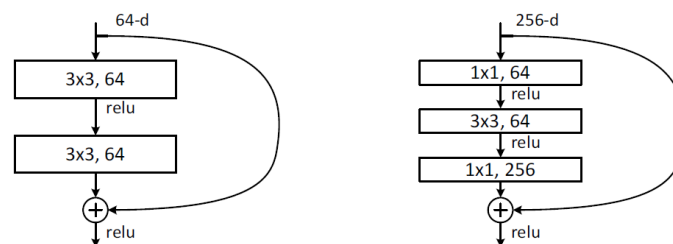


Figure 4. Residual block

### 5.2. VGG19

The VGG19 neural network architecture comprises 19 layers and features a substantial number of parameters. The model's size, specifically in terms of fully connected nodes, is 574MB. An increase in the number of layers typically correlates with enhanced accuracy in deep neural networks (DNNs). The VGG19 architecture includes 19 trainable convolutional layers, which are interspersed with max pooling and dropout layers, as depicted in Figure 5.

VGG19 is highly advantageous due to its use of a series of  $3 \times 3$  ConvNet to increase network depth. To reduce feature map dimensions, max-pooling layers are incorporated. The fully connected network (FCN)



comprises two layers, each containing 4096 neurons. VGG19 was trained on individual lesion samples, while for testing, all lesion types were included to minimize the occurrence of false positives.

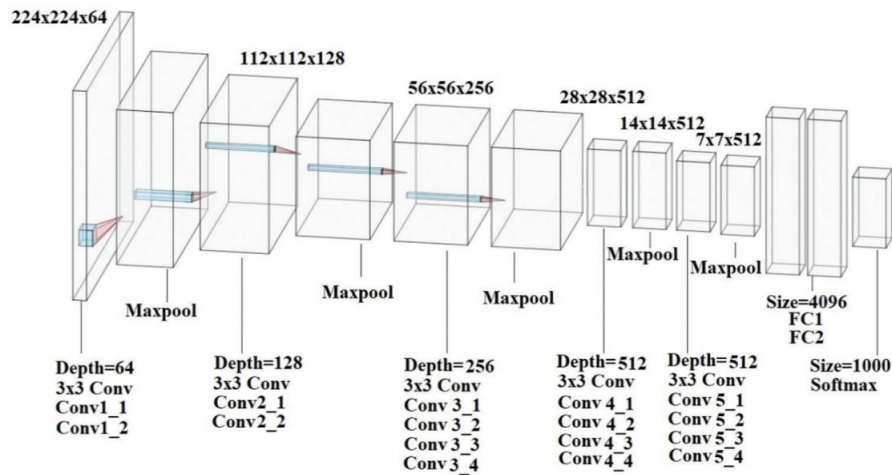


Figure 5. VGG19 architecture

### 5.3. DenseNet169

The DenseNet architecture is composed of multiple dense blocks, which are interconnected by transition blocks that consist of convolutional and pooling layers. Each dense block is structured from several units containing convolutional layers, where each unit receives the outputs of all preceding units with the same feature map size. Unlike ResNet, which performs summation on the outputs, DenseNet concatenates them. Figure 6 illustrates the structure of a dense block with a growth rate of  $k=4$ . In this context, 'k' represents the number of feature maps generated by the function H. If each H function produces k feature maps, then the n-th layer will have  $k_0 + k \times (n-1)$  input feature maps, where  $k_0$  denotes the number of channels in the input layer.

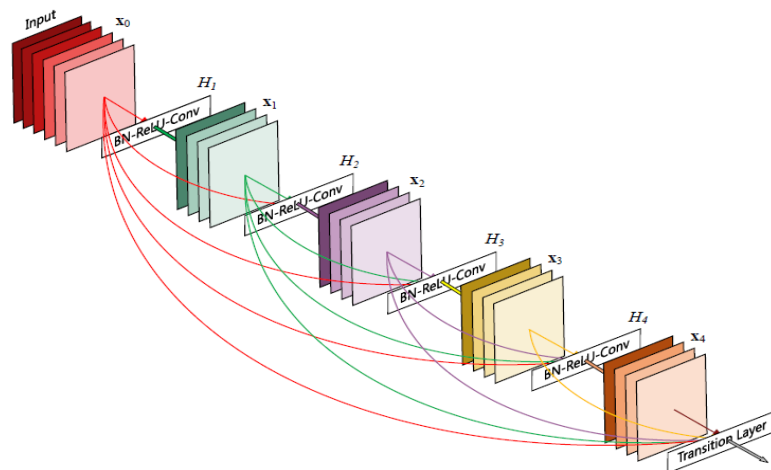


Figure 6. DenseNet block with a growth rate of  $k=4$

### 5.4. InceptionV3

As shown in Table 1, this architecture includes a stem block consisting of six convolutional layers, each with a  $3 \times 3$  kernel size, with a pooling layer following the third convolutional layer. Padding is set to zero for the specified convolutional layers, while for the others, no padding is applied. To reduce grid size, a reduction method is employed between Inception blocks. The architecture has a depth of 42 layers and

incurs a computational cost 2.5 times higher than GoogleNet, though it remains significantly more efficient than the VGG architecture. The primary distinction between Inception v2 and v3 lies in the factorization of convolutions, which reduces the number of parameters without affecting network performance. For example, replacing a 5×5 convolutional layer with two 3×3 layers allows the network to extract features similarly while reducing computational complexity and model cost.

Table 1. Architecture of inception V3

Type	Patch size/stride or remarks	Input size
conv	3×3/2	299×299×3
conv	3×3/1	149×149×32
conv padded	3×3/1	147×147×32
pool	3×3/2	147×147×64
conv	3×3/1	73×73×64
conv	3×3/2	71×71×80
conv	3×3/1	35×35×192
3×Inception	Inception blok A	35×35×288
5×Inception	Inception blok B	17×17×768
2×Inception	Inception blok C	8×8×1280
pool	8×8	8×8×2048
linear	logits	1×1×2048
softmax	classifier	1×1×1000

## 6. ENSEMBLE APPROACH

Ensemble learning involves amalgamating the outcomes of multiple models, which may employ identical or distinct algorithms. This field encompasses various techniques categorized into three primary groups: bagging, boosting, and stacking. Bagging entails aggregating the results of a single model trained iteratively on diverse datasets. Boosting combines the outputs of multiple models, while stacking integrates elements of both approaches.

In this study, the boosting technique will be implemented. Following the training of CNN, ResNet50, VGG19, DenseNet169, and Inception v3 models, predictions will be made on images within the validation and test datasets. Subsequently, a max voting technique will be applied, leveraging the sum function, as the outcomes are confined to the [0, 1] class range. The class with the highest vote count will be selected. This approach serves to mitigate variance and enhance error generalization.

## 7. EVALUATION

When dealing with supervised problems, we can evaluate the model in two different ways. First, data must be divided into training and testing sets. Next, the model must be trained using the training sets, and test dataset labels must be predicted using the trained model. This allows us to compute the errors by comparing the predicted results with the true values. K-Fold cross-validation is the second process. To train the model in this scenario, the data will be split into K subsets, with the exception of one that will be kept out for assessment. After each training round, we compute matrices, and the optimal model will ultimately be chosen. Though it requires a lot more computation and money than deep learning, this approach is superior to the previous one. Due to the huge amount of computation in networks, it is not advised unless it is affordable.

## 8. RESULTS AND DISCUSSION

For results analysis, to make the results more appropriate for our purpose, which is minimizing the incorrect classifications, we computed the confusion matrix, which displays the number of correct and incorrect classifications. Additionally, the confusion matrix-assisted in determining all of the significant metrics; precision, accuracy, F1 scores, and recall were computed for each architecture.

### 8.1. Ensemble 5 models CNN, ResNet50, VGG19, DenseNet169, and InceptionV3

An ensemble model is a combination of multiple individual models to achieve better predictive performance. The idea is that by combining the strengths of multiple models, the ensemble can capture a broader range of features and reduce errors. In this method, CNN, ResNet50, VGG19, DenseNet169, and InceptionV3 models were chosen. After averaging their predictions, the performance metrics for the test dataset were as follows: 97.5% accuracy, 96.8% precision, 96.9% recall, and 97.77% F1-score.



## 8.2. Comparisons with other works

This section compares our proposed model's performance with a number of state-of-the-art methods that have been put forth for using microscopic images to identify lung cancer. Table 2 compares the results of the corresponding introduced method with several state-of-the-art techniques described in the literature review. The study compares the proposed model's performance to the state-of-the-art models for the detection and categorization of lung cancer.

Table 2. Performance comparison between our proposed model and existing models across diverse datasets

References	Dataset	Model	Recall	Accuracy
Maftouni <i>et al.</i> [26]	COVID19-CT	Ensemble + SVM	90.80	95.31
Gifani <i>et al.</i> [16]	COVID19-CT	CNN + LSTM	85.50	85.50
Bhandar <i>et al.</i> [21]	CR	Custom CNN	96.56	94.31
Chen <i>et al.</i> [27]v	IQ-OTH/NCCD	CNN + NLP	87.5	88.0
Ali <i>et al.</i> [18]	HAM10000	DCNN	93.66	91.93
Al-Yasriy <i>et al.</i> [28]	IQ-OTH/NCCD	CNN: AlexNet architecture	93.23	93.54
Proposed model	IQ-OTH/NCCD	Ensemble 5 Models CNN, ResNet50, VGG19, DenseNet169, and InceptionV3	97.77	97.5

This work aimed to improve performance by using four different ensemble approaches and a customized CNN trained on microscopic images of lung cancer. The accuracy results of all models are compared in Table 2. On test data, the ensemble of the top four models had the highest accuracy. A COVID-19-CT dataset comprising 7,593 images sourced from seven publicly available datasets, encompassing data from 466 patients is provided by Maftouni *et al.* [26] through the use of an ensemble deep learning model utilizing pre-trained residual attention and DenseNet architectures, Gifani *et al.* [16] employed an ensemble deep transfer learning system, leveraging diverse pre-trained CNN architectures, to achieve effective diagnosis of COVID-19 from CT scans.

On the other hand, a novel lightweight single CNN model for COVID-19 image classification using CXR images was proposed in [21]. Furthermore, on the test dataset, all of the ensemble models outperformed single models in terms of accuracy. The ensemble model of the top four models, which had an accuracy of 97.53%, was the best model in the dataset validation scenario. In addition to tuberculosis and pneumonia, an explanation generation (XAI) framework is used. The detection of COVID-19, pneumonia, and tuberculosis diseases using such an XAI-based single CNN model produced training accuracy of  $95.76 \pm 1.15\%$ , test accuracy of  $94.31 \pm 1.01\%$ , and validation accuracy of  $94.54 \pm 1.33\%$ .

Basic and benchmark CNN architectures were applied by Chen *et al.* [27], who were known to employ any parameter optimization technique. Even though the majority of his study's results are intriguing and achieve an accuracy of 88.0%, we observe that these models will perform poorly if performance tilting conditions are added. Additionally, a DCNN model based on a deep learning approach was proposed by Ali *et al.* [18] for precise categorization of benign and malignant skin lesions. The model was assessed using the HAM10000 dataset and attained impressive outcomes with a testing accuracy of 91.93% and a training accuracy of 93.16%. Al-Yasriy *et al.* [28] used the same dataset that we used in our proposed model using CNN-AlexNet architecture and achieved an accuracy of 93.548%, a sensitivity of 95.714%, and a specificity of 95%.

Our proposed hybrid models; CNN, ResNet50, VGG19, DenseNet169, and InceptionV3, gave performance metrics, after averaging prediction of selected models, of 97.5% accuracy. Consequently, this study has shown how crucial it is to solve the challenging task of choosing the ideal set of weights and biases needed for training a CNN model by utilizing the hybrid metaheuristic ensemble algorithm and CNN models [29]. Furthermore, the method demonstrates that amalgamating these approaches enhances the accuracy of classification and overall performance when classifying lung cancer in CT images.

## 9. CONCLUSION

Using a CNN model and four different ensemble approaches, this study presents a novel hybrid algorithm to increase the accuracy of lung cancer classification. To ensure a robust and accurate comparison of results, the experimental procedure involved partitioning the dataset into test sets, validation and training. During the training phase, model validation was performed using a dedicated validation dataset. Following the training process, predictions were generated for both the validation and test. The models were trained using a customized CNN model, along with several pre-trained transfer learning architectures, namely ResNet50, VGG19, DenseNet169, and InceptionV3. Subsequently, this ensemble approach was employed to address model weaknesses and enhance overall performance. Multiple models were combined and merged

using the average ensemble technique. The optimal combination, identified as a blend of the top five models, was chosen as the proposed hybrid model. This work can be extended to implement with different datasets and using 3D CNNs or volumetric segmentation for 3D medical images.

## ACKNOWLEDGEMENTS

The author/s would like to thank the Palestine Technical University-Kadoorie for their financial support to conduct this research.




## REFERENCES

- [1] R. Yousef, "Identifying informative coronavirus tweets using recurrent neural network document embedding," *Palestine Technical University Research Journal*, vol. 10, no. 1, pp. 93–102, 2022, doi: 10.53671/pturj.v10i1.220.
- [2] W. Li *et al.*, "Advances in the early detection of lung cancer using analysis of volatile organic compounds: From imaging to sensors," *Asian Pacific Journal of Cancer Prevention*, vol. 15, no. 11, pp. 4377–4384, 2014, doi: 10.7314/APJCP.2014.15.11.4377.
- [3] C. S. Dela Cruz, L. T. Tanoue, and R. A. Matthey, "Lung cancer: epidemiology, etiology, and prevention," *Clinics in chest medicine*, vol. 32, no. 4, pp. 605–644, 2011, doi: 10.1016/j.ccm.2011.09.001.
- [4] F. Bray *et al.*, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 74, no. 3, pp. 229–263, 2024, doi: 10.3322/caac.21834.
- [5] S. A. Belinsky *et al.*, "Gene promoter methylation in plasma and sputum increases with lung cancer risk," *Clinical Cancer Research*, vol. 11, no. 18, pp. 6505–6511, 2005, doi: 10.1158/1078-0432.CCR-05-0625.
- [6] V. P. Doria-Rose, P. M. Marcus, E. Szabo, M. S. Tockman, M. R. Melamed, and P. C. Prorok, "Randomized controlled trials of the efficacy of lung cancer screening by sputum cytology revisited: A combined mortality analysis from the Johns Hopkins lung project and the memorial sloan-kettering lung study," *Cancer*, vol. 115, no. 21, pp. 5007–5017, 2009, doi: 10.1002/cncr.24545.
- [7] M. M. Oken *et al.*, "Screening by chest radiograph and lung cancer mortality: The prostate, lung, colorectal, and ovarian (PLCO) randomized trial," *Jama*, vol. 306, no. 17, pp. 1865–1873, 2011, doi: 10.1001/jama.2011.1591.
- [8] The National Lung Screening Trial Research Team, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395–409, 2011, doi: 10.1056/NEJMoa1102873.
- [9] K. Das *et al.*, "Machine learning and its application in skin cancer," *International Journal of Environmental Research and Public Health*, vol. 18, no. 24, 2021, doi: 10.3390/ijerph182413409.
- [10] M. M. Ahsan *et al.*, "Deep transfer learning approaches for Monkeypox disease diagnosis," *Expert Systems with Applications*, vol. 216, 2023, doi: 10.1016/j.eswa.2022.119483.
- [11] A. Asuntha and A. Srinivasan, "Deep learning for lung cancer detection and classification," *Multimedia Tools and Applications*, vol. 79, no. 11, pp. 7731–7762, 2020, doi: 10.1007/s11042-019-08394-3.
- [12] W. J. Sori, J. Feng, and S. Liu, "Multi-path convolutional neural network for lung cancer detection," *Multidimensional Systems and Signal Processing*, vol. 30, pp. 1749–1768, 2019, doi: 10.1007/s11045-018-0626-9.
- [13] L. N. Gumma, R. Thiruvengatanadhan, L. Kurakula, and T. Sivaprakasam, "A survey on convolutional neural network (deep-learning technique)-based lung cancer detection," *SN Computer Science*, vol. 3, no. 1, pp. 1–7, 2022, doi: 10.1007/s42979-021-00887-z.
- [14] H. M. Shyni and E. Chitra, "A comparative study of X-Ray and CT images in COVID-19 detection using image processing and deep learning techniques," *Computer Methods and Programs in Biomedicine Update*, vol. 2, 2022, doi: 10.1016/j.cmpbup.2022.100054.
- [15] T. Rahman *et al.*, "COV-ECGNET: COVID-19 detection using ECG trace images with deep convolutional neural network," *Health Information Science and Systems*, vol. 10, no. 1, 2022, doi: 10.1007/s13755-021-00169-1.
- [16] P. Gifani, A. Shalbaf, and M. Vafaezadeh, "Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 1, pp. 115–123, 2021, doi: 10.1007/s11548-020-02286-w.
- [17] U. O. Dorji, K. K. Lee, J. Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9909–9924, 2018, doi: 10.1007/s11042-018-5714-1.
- [18] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, 2021, doi: 10.1016/j.mlwa.2021.100036.
- [19] M. Pal, S. Mistry, and D. De, "Interpretability approaches of explainable AI in analyzing features for lung cancer detection," in *Frontiers of ICT in Healthcare*, Singapore: Springer, 2023, pp. 277–287, doi: 10.1007/978-981-19-5191-6\_23.
- [20] M. S. Ahmed, K. N. Iqbal, and M. G. R. Alam, "Interpretable lung cancer detection using explainable AI methods," in *2023 International Conference for Advancement in Technology, ICONAT 2023*, 2023, pp. 1–6, doi: 10.1109/ICONAT57137.2023.10080480.
- [21] M. Bhandari, T. B. Shahi, B. Sikur, and A. Neupane, "Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI," *Computers in Biology and Medicine*, vol. 150, 2022, doi: 10.1016/j.compbiomed.2022.106156.
- [22] K. Dwivedi, A. Rajpal, S. Rajpal, M. Agarwal, V. Kumar, and N. Kumar, "An explainable AI-driven biomarker discovery framework for non-small cell lung cancer classification," *Computers in Biology and Medicine*, vol. 153, 2023, doi: 10.1016/j.compbiomed.2023.106544.
- [23] W. N. Hennes, "Lung cancer dataset (IQ-OTH/NCCD)," *Kaggle*, 2022. [Online]. Available: <https://www.kaggle.com/datasets/waseemnagahhennes/lung-cancer-dataset-iq-othnccd>
- [24] M. Azizjon, A. Jumabek, and W. Kim, "1D CNN based network intrusion detection with normalization on imbalanced data," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, 2020, pp. 218–224, doi: 10.1109/ICAIIIC48513.2020.9064976.
- [25] L. Huang, X. Hong, Z. Yang, Y. Liu, and B. Zhang, "CNN-LSTM network-based damage detection approach for copper pipeline using laser ultrasonic scanning," *Ultrasonics*, vol. 121, 2022, doi: 10.1016/j.ultras.2022.106685.
- [26] M. Maftouni, A. C. C. Law, B. Shen, Z. K. Grado, Y. Zhou, and N. A. Yazdi, "A robust ensemble-deep learning model for COVID-19 diagnosis based on an integrated CT scan images database," in *IISE Annual Conference and Expo 2021*, 2021, pp. 632–637.




- [27] J. Chen, Q. Ma, and W. Wang, "A lung cancer detection system based on convolutional neural networks and natural language processing," in *Proceedings-2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology, AINIT 2021*, 2021, pp. 354–359, doi: 10.1109/AINIT54228.2021.00076.
- [28] H. F. Al-Yasriy, M. S. AL-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Diagnosis of lung cancer based on CT scans using CNN," *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 2, 2020, doi: 10.1088/1757-899X/928/2/022035.
- [29] E. Y. Daraghmi, S. Qadan, Y. A. Daraghmi, R. Yousuf, O. Cheikhrouhou, and M. Baz, "From text to insight: an integrated CNN-BiLSTM-GRU model for Arabic cyberbullying detection," *IEEE Access*, vol. 12, pp. 103504–103519, 2024, doi: 10.1109/ACCESS.2024.3431939.

## BIOGRAPHIES OF AUTHORS



**Rami Yousef**    is currently works as an Associate Professor at the Department of Computer Engineering at the Palestine Technical University-Kadoorie (Tulkarm, Palestine). He obtained his Ph.D. from the University Kebangsaan Malaysia (UKM), Malaysia in computer science–artificial intelligence in 2019, whereas his M.S. degree was from Birzeit University in scientific computing in 2005. His bachelor degree was in electrical engineering from An Najah National University in 1999. He has developed many professional disciplines that keep up with the local market and its need for professionals specialized in all technological and industrial fields. His primary research interests are in the fields of "artificial intelligence applications," remote control, expert systems, neural network, machine learning, and deep learning. He can be contacted at email: r.yousuf@ptuk.edu.ps.



**Eman Yaser Daraghmi**    is currently works as an Associate Professor at the Department of Computer Science, Palestine Technical University Tulkarm (PTUK). She received her B.S. degree in communication and information technology from Al Quds Open University in 2008, her M.S. degree in computer science from National Chiao Tung University, Taiwan in 2011, and her Ph.D. degree in computer science and engineering from National Chiao Tung University, Taiwan in 2015. Her current research interests include artificial intelligence, machine learning, distributed and cloud computing and blockchain. She is passionate about how artificial intelligence can support day-to-day activities such as digital assistants. She can be contacted at email: e.daraghmi@ptuk.edu.ps.