



## Adaptive feature selection based on machine learning algorithms for Lung tumors diagnosis and the COVID-19 index

Bashar Talib AL-Nuaimi<sup>1</sup>, Ruaa Azzah Suhail<sup>2</sup>, Sanaa adnan abbas<sup>3</sup>, El-Sayed M. El-kenawy<sup>4,\*</sup>

<sup>1</sup>Diyala University, College of Science, Department of Computer Science,  
Directorate General of Education, Diyala, Iraq

<sup>2</sup>Diyala University, College of Science, Department of Computer Science,  
Directorate General of Education, Diyala, Iraq

<sup>3</sup>Diyala University, College of Science, Department of Computer Science,  
Directorate General of Education, Diyala, Iraq

<sup>4</sup>Department of Communications and Electronics, Delta Higher Institute of Engineering and  
Technology, Mansoura 35111, Egypt

Emails: [alnuaimi\\_bashar@uodiyala.edu.iq](mailto:alnuaimi_bashar@uodiyala.edu.iq); [ruaaizzat@gmail.com](mailto:ruaaizzat@gmail.com); [sanaalasadi2000@yahoo.com](mailto:sanaalasadi2000@yahoo.com);  
[skenawy@ieee.org](mailto:skenawy@ieee.org)

### Abstract

Early detection of Lung tumors, which is lethal and equally affects men and women, is challenging. In order to decrease mortality rates and raise survival rates, early detection and classification of Lung tumors is essential. However, at the start of 2020, the entire planet would be afflicted with a coronavirus that causes a fatal sickness (COVID-19). CT imaging is a good tool to detect illness among the various COVID-19 screening techniques available. On the other hand, alternative methods of disease detection take a lot of time. Deep learning, a type of machine learning, opens up a wealth of opportunities for investigating and assessing tumor features using CT scans, allowing for improved disease prediction, diagnosis, and classification. Using CNN, DNN, and VGG-16 models, the suggested approach in this research gives unambiguous and accurate categorization.

Received: August 07, 2023 Revised: November 19, 2023 Accepted: January 08, 2024

**Keywords:** Lung tumors; CT; COVID-19; DNN; CNN; VGG-16.

### 1. Introduction

Many diseases can affect the human lungs, including pneumonia, Lung tumors, and, most recently, COVID-19. Deep learning-based systems, for example, have been developed using computer science technology to assist with disease diagnosis. When all other malignancies are taken into account, in both men and women, lung carcinoma is the most prevalent type of cancer. It is brought on by smoking and consuming tobacco products, which is a cancerous substance that affects the tissues of the lungs. Malignant tumors are cancerous nodules that develop as a result of excessive cell proliferation in lung tissues [1]. Despite advances

in detection, diagnosis, and treatment, many people die from Lung tumors.

Various machine learning algorithms have proved the efficiency and outstanding performance of artificial intelligence in automated photo classification challenges, and it is now being used to automate the diagnosis of a number of diseases. Machine learning models are models that can learn and make decisions based on a huge number of examples of input data. AI performs computations and develops predictions based on incoming input in order to accomplish tasks that typically require human intelligence.

The purpose of many ML and DL algorithms might be to increase the quality, consistency, and/or diagnostic capability of data interpretation [2]. This cleared the path for the creation of Deep Learning, a sub-field of machine learning (DL).

The spread of COVID-19, sometimes known as the Zika virus, around the world as 2019 drew to a close was cause for concern. Infections of this organism typically cause respiratory symptoms. Diagnosis is challenging for even the most COVID-19 [3, 4]. The condition is diagnosed using an RT-PCR is a real-time polymerase chain reaction kit. Patients who are suspect cannot be treated immediately due to a paucity of kits, allowing sickness to spread. CT scans and other forms of cutting-edge radiological imaging were used by doctors to remedy this situation, which produce detailed, high-resolution pictures of the human body. A lot of research has using deep learning to analyse radiological pictures. Deep learning algorithms, particularly CNN, have sparked a lot of interest in data processing. Lung tumors, pneumonia, TB, emphysema, and other pleural illnesses are all detected with CNN technology.

The focus of this research is to build a multi-classification deep learning model that can identify chest infections. Datasets termed "lung" and "Covid" were compiled from Iraqi hospitals, and their ability to classify cases into "benign," "malignant," and "normal" was tested using the "Covid-19" and "Nocovid" datasets, respectively. Convolutional neural networks (CNNs), deep neural networks (DNNs), and the VGG-16 model are just a few examples of possible implementations.

## 2. Related Work and background

Several projects are routinely constructed using a neural network to improve Lung tumors detection precision. Chen et al. [5] developed a technique for distinguishing between totally benign, lung nodules that are unknown and most likely cancerous using a neural network ensemble (NNE) architecture. The scheme outperformed the individual classifier in classification accuracy (78.7%), (LVQNN: 68.1%). Kuruvilla and Gunavathi suggested a textural feature based artificial neural network (ANN) methodology that has a (93.30 percent) accuracy [6]. For detection and classification, combining texture and shape features may increase classification accuracy [7].

Kumar and colleagues developed a system based on stacked auto encoders (SAE), a deep learning algorithm, that has a (75.01%) accuracy rate[8]. In most countries, computed tomography (CT) scans can be used to discover COVID-19 since its quick and inexpensive. [7]. COVID-19 was identified utilizing chest X-ray radiographs, pneumonia in infected persons, and three separate CNN models: (ResNet50, InceptionV3, and Inception-ResNetV2). There was no need for a feature extraction or selection phase. The classification accuracy of the ResNet50 model was (98 percent), The categorization accuracy of InceptionV3 was (97 percent), and InceptionResNetV2 had an accuracy of (87 percent). They ran roughly 30 epochs for all of the models in their training phase to avoid overfitting. In their study, they did, however, employ a few photographs that were available at the time. Deep CNNs were used in some studies for medical picture identification and classification. Shen et al. [8] used a multiscale two-layer CNN to identify Lung tumors using the LIDC database in 2015, with an 86.84 percent accuracy.

Deep learning produces more accurate predictive models. and diagnosing illnesses based on images, all without the need for human participation, Breast cancer [9], liver illness [10], colon cancer [11], brain tumor [12], skin cancer [13], Lung tumors [14], pneumonia [15], and, more recently, COVID-19 diagnosis. As a consequence, the model gathers data on its own and generates more exact results. In contrast to traditional machine learning methods, deep learning approaches improve model accuracy by combining a series of nonlinear functions. For personality factor categorization, the CNN + LSTM framework combines CNN and LSTM to aid in the classification of input text into several personality traits. Psychopaths are also detected using the integrated framework, which has a classification accuracy of (91.67 percent) [16, 17]. Based on a poorly supervised multiscale deep learning framework motivated by the VGG architecture [19], Hu et al. [18] proposed a COVID-19 detection approach using chest CT data.

## 3. System Architecture Proposed

Chest tomography images CT scans are utilized as a source of data. CT images are preprocessed before being used. As shown in Figure (1), this research investigates the performance of different classifiers in Lung tumors and covid-19 detection, as well as, preprocessing techniques. Then, by comparing their individual performance measurements, the usefulness of various classifiers and preprocessing methods in cancer diagnosis may be assessed. The accuracy of the predictions for each class, as well as the time it takes the classifier to predict the class of each input, are used to determine the model's complexity. Increase the amount of training images in order for our system to provide accurate and efficient results, image preprocessing method used.

The accuracy of the system will be affected by the usage of pre-processing. Figure (3) shows the model block diagram for Lung tumors and Covid19 dataset respectively. The three major steps of the model are represented in the diagram: Preprocessing of data, feature extraction using deep learning models, and categorization. Using chest CT scans, the proposed approach divides the input picture into five categories: Lung tumors data has (Benign, Malignant, Normal) categories, while the Covid19 data has (Covid-19, Nocovid) categories. In the first step, data is split between training and testing sets; training receives 80% of the data while the latter receives 20%. The photographs in the dataset are arbitrarily divided into two groups (training and testing) to guarantee visual variety. The image's pixel value is rescaled to the intervals (0,1) and (0,1,2) using the picture scaling and enlargement for Lung tumors and Covid19 datasets. Later, we'll talk about the second and third steps, feature extraction and image categorization using deep learning algorithms.

**A. Datasets for the study**

We used CT images from a variety of sources in our investigations. Normal scans, COVID-19 scans, and CT images of Lung tumors are all included here. Our analysis of COVID-19 began with CT scan data from SARS-CoV-2, which includes (1252) scans from people infected with COVID-19 and (1230) scans from healthy individuals. Second, the Lung tumors CT images used to compile the dataset were specifically collected for this purpose. The pictures, numbering (1315) was collected from various Iraqi medical facilities. The suggested deep learning system (CNN, DNN, and VGG) can be trained to distinguish between benign, malignant, and normal Lung tumors (Covid-19, NoCovid) using these images.

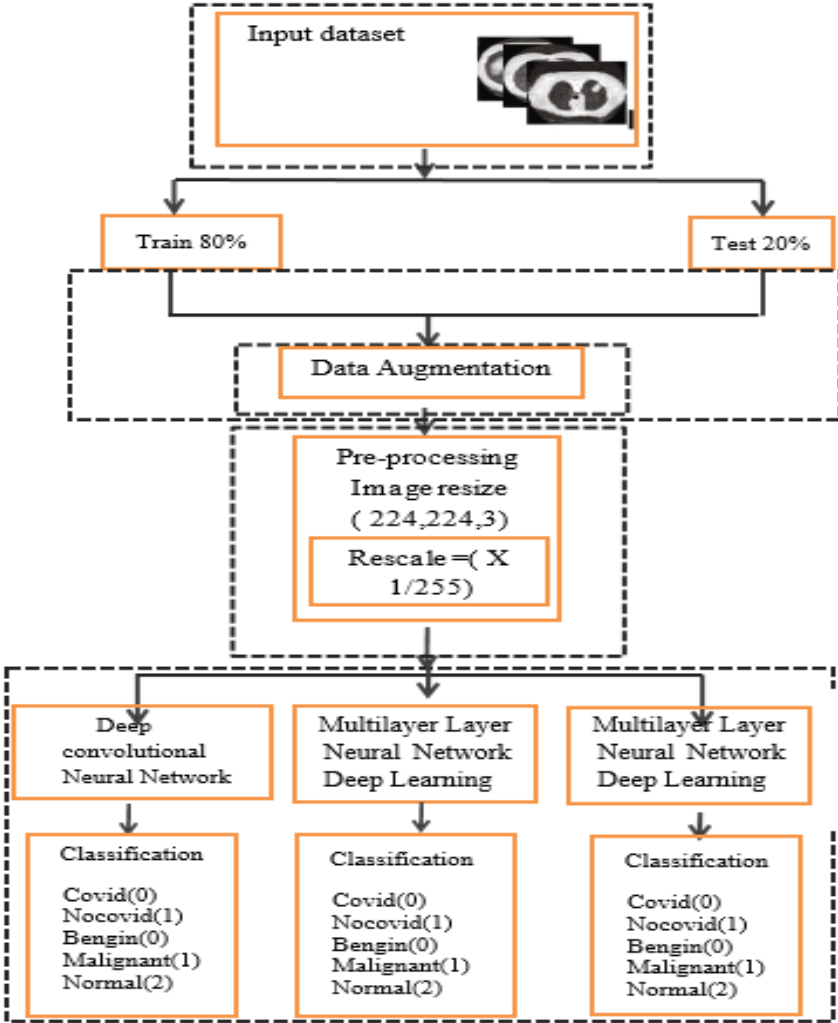


Figure 1: General block diagram of (CNN, DNN AND VGG-16) model for Lung tumors and Covid dataset.

## **B. Preparing the dataset**

The collections are first supplemented to increase the number of photos used and balance the various dataset kinds. The five training classes will divide the first application of augmentation techniques: (Covid-19, Nocovid) and (Benign, Malignant, Normal) for Lung tumors datasets. Following that, these photos are sent into the stage of data preparation. The initial stage of processing is frequently used to prepare data for deep learning models. Various pre-processing procedures were employed in our approach to prepare the input photos.: reducing photos to (224×224 ×3) adding augmentation methods like rotate, flip, skewing, and normalizing to the resized images.

### **3.1. Deep Learning Techniques**

#### **3.1.1 Neural Networks using Convolution (CNNs).**

Convolutional neural networks (CNNs) are multi-layer neural networks that combine fully-connected layers with convolutional layers. As a result, the CNN technique was used for the following reasons:

1- High and accurate performance, outperforming other recognition algorithms implemented on the same dataset.

2- The CNN can be taught to construct elaborate structures.

3- It does not necessitate any pre-processing operations (excepting resize and normalize of the input image).

The samples are presented to the CNN as images of a constant resolution. Image input dimensions are (224×224×3) for this network. One of the fundamental efforts for the proposed model is finding an appropriate size of the image. The process of choosing an efficient size of the image represents a trade-off between high-speed classification (smaller size) and accuracy (larger size), i.e., the larger size of the image indicates more information and could be more accurate. On the contrary, the image with a smaller size indicates a fewer amount of information in real performance, and so, faster classification.

The second task is to determine which architecture is ideal for CNN. So, first and foremost, the network's input and output structures have been specified. Then, based on the outcomes of some experiments, the optimal design has been provided. For optimal performance, a CT picture size of (224×224×3) was chosen. For the COVID dataset, we use a sigmoid layer to select a final classification (Covid-19 or NOCOVID) and for the Lung Cancer dataset, we select a final classification (Benign, Malignant, Normal).

#### **3.1.2. Neural Networks in Depth (DNN)**

A basic neural network with a large number of hidden nodes is called a (DNN). More sophisticated input computations are possible with a "deep" neural network as opposed to a "shallow" neural network due to the fact that each hidden layer may represent a nonlinear modification of the output layer. Throughout the training phase, the original picture (50×50×3) is used as the input layer settings to maintain a high level of performance while maintaining a large quantity of precise information about the image. (DNN) architecture couples an input layer, a hidden layer, and an output layer. The two necessary phases of the system are the training phase and the testing phase. The results of these two steps are presented so that Lung tumors and COVID images can be detected and classified. The DNN technique provided 80% of the data required during training. Random weights are assigned during training network configuration; the network then trains these weights over multiple epochs until they stabilize. In both the Lung tumors and Covid datasets, the optimal training performance occurs at epoch 50. Table (1) below provides the (DNN) parameters with a fully connected layer used for the Covid dataset.

Table (1): show Covid dataset.

Layer (type)	Output Shape	Param#
img_input (InputLayer)	[(None, 50, 50, 3)]	0
fc_1 (Flatten)	(None, 7500)	0
layer_3 (Dense)	(None, 1024)	7681024
dropout (Dropout)	(None, 1024)	0
layer_4 (Dense)	(None, 512)	524800
dropout_1 (Dropout)	(None, 512)	0
layer_5 (Dense)	(None, 256)	131328
layer_6 (Dense)	(None, 128)	32896
layer_7 (Dense)	(None, 64)	8256
layer_8 (Dense)	(None, 32)	2080
dropout_2 (Dropout)	(None, 32)	0
predictions (Dense)	(None, 2)	66
The total number of parameters is 8,380,450. 8,380,450 trainable parameters 0 non-trainable parameters		

There is a large layer of (512) neurons in the classification phase before the dropout layer. In order to classify images into one of four chest sickness categories ((Covid-19, NOCOVID) for the Covid dataset, and ((Benign, Malignant, and Normal) for the Lung tumors dataset), the final output is produced using a dense layer made up of two neurons with a SoftMax activation function. The general parameters of the dataset, which are divided into two groups, are as follows: The COVID dataset and the VGG Lung tumors dataset, respectively, contain (35,474,498) trainable parameters and (28), and (35,475,139) non-trainable parameters. Time-varying trainable parameters require training, but time-invariant non-trainable values do not.

### 3.3.1 Visual Geometry Group (VGG16).

For time-varying trainable parameters, training is required to achieve the best potential value, however this is not the case for static non-trainable parameters. The suggested model has layers for ingesting images, extracting features, and classifying them. The model's input layer is fed a total of (224×224×3) CT scans. The first CNN block consists of a convolution layer and a ReLU layer. Two convolution layers, two ReLU layers, a batch normalizing layer, a maximum pooling layer, and a dropout layer make up the second block, as shown in Figure (1).

The parameters of a model must either be explicitly defined or given as inputs if they cannot be learned. Because of this, classification won't be affected by the non-trainable. Parameters of a model that cannot be trained in any way, shape, or form must be defined a priori or supplied as inputs. As a result, classification will be unaffected by the non-trainable.

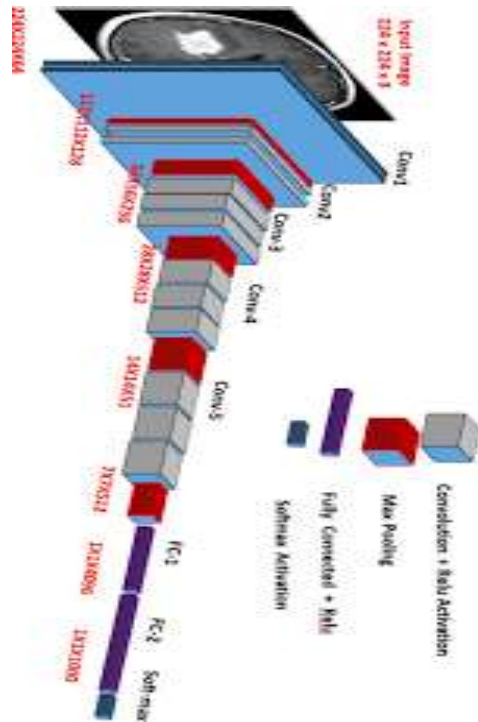


Figure (2): VGG Neural Network Architecture.

## 6. Results

There was a count of successful and unsuccessful classification model predictions in the confusion matrix as well.

Table 2: All-Architecture Outcomes.

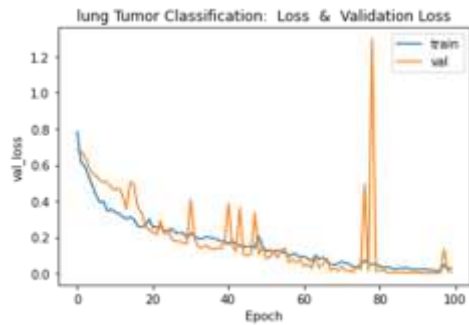
Models	Accuracy	Sensitivity
(CNN) Lung tumors	100%	1.00
(CNN) Covid	93%	0.94
(VGG) Lung tumors	100%	1.00
(VGG) Covid	100%	1.00
(DNN) Lung tumors	86%	86
(DNN) Covid	75%	75

Table 3: Comparison with other articles

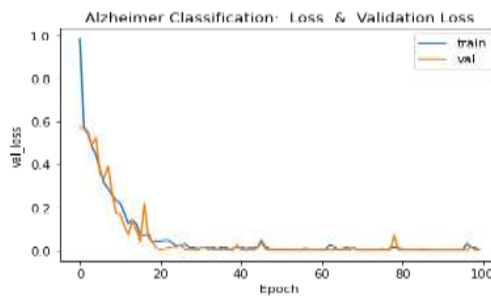
Work	Diagnostic picture	Accuracy (%)
Nascimento and al. [21]	LIDC (73)	92.78%
Krewer and al. [7]	LIDC-IDRI (33)	90.91%
Dandil et al. [22]	Private (128)	90.63%
Kuruvilla and Gunavathi, [6]	LIDC (110)	93.30%
Gupta and Tiwari [23]	Private (120)	90%
Kumar and al. [8]	LIDC (4323)	75.01%
da Silva [24]	LIDC-IDRI (8296)	82.3%
CNN [25]	LIDC-IDRI (5024)	84.15%

DNN [25]	LIDC-IDRI (5024)	82.37%
SAE [25]	LIDC-IDRI (5024)	82.59%

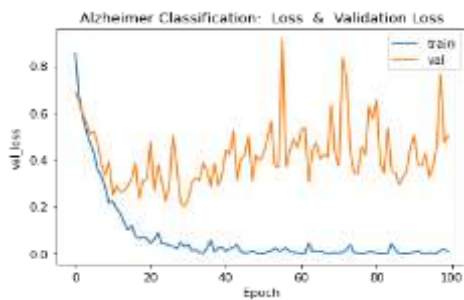
With 100% accuracy on the Lung tumors dataset, 94% accuracy on the COVID dataset, and 100% accuracy on both datasets, the CNN and VGG deep learning models clearly outperform the DNN models, which show overfitting when epochs are increased to 100%. The recommended model's analysis of epoch versus loss and vs accuracy is depicted in Figures (4) and (5), respectively. The epoch represents the data's iteration count. The term "loss" refers to the average mistake on the training set, which can be expressed in regression or classification terms. Table (3) shows the results of a comparison between the suggested method and the current system's accuracy.



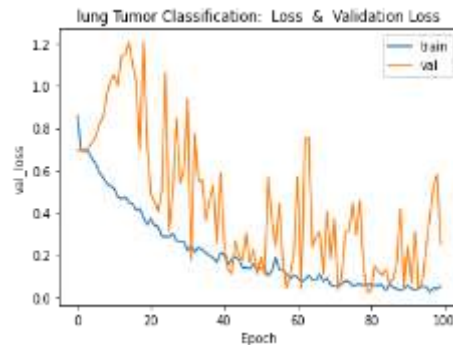
a- CNN Lung tumors



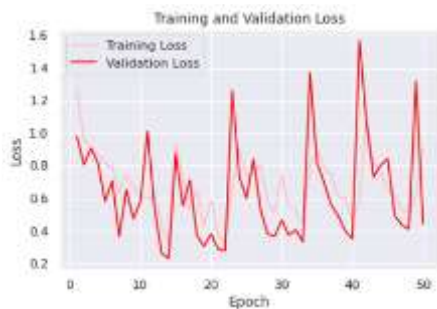
b- CNN Covid



c-VGG Lung tumors: The loss curves in both training and validation are decreasing towards zero.



d-VGG Covid: As a result, the VGG16 model outperforms other models in terms of accuracy.

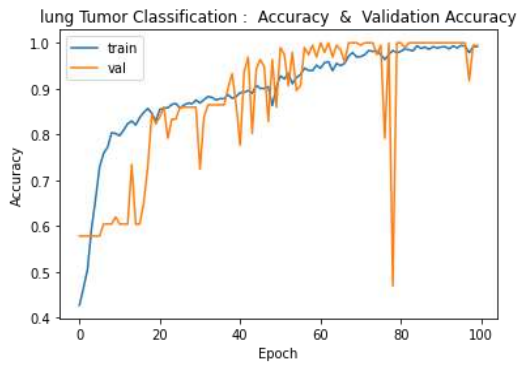


e- DNN Lungcancer

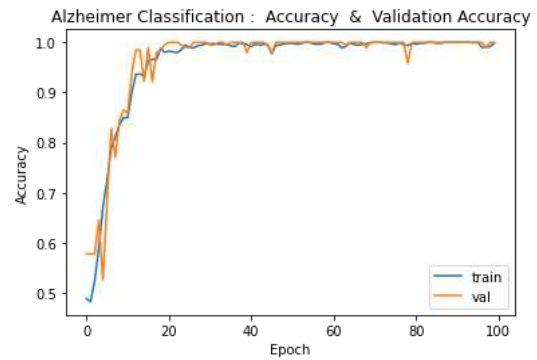


f- DNN Covid

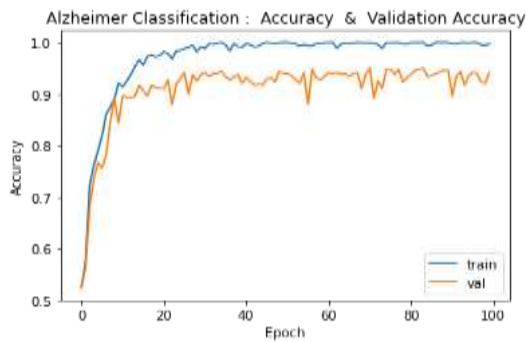
Figure (4) All deep models' epochs vs. loss graph



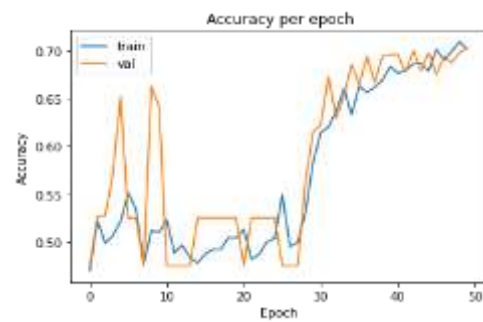
a- CNN Lung



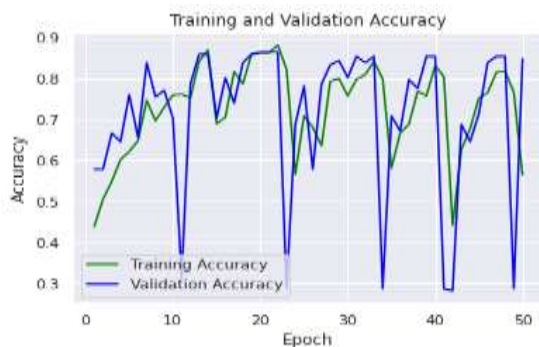
b- CNN covid



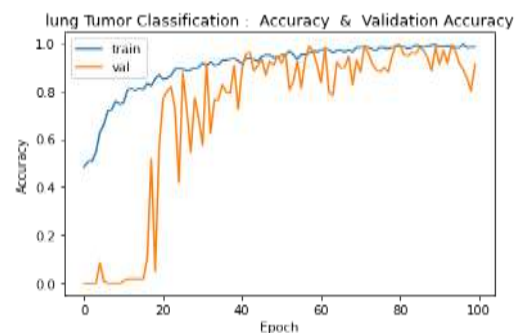
c-VGG Lung tumors: The VGG16 architecture



d-VGG Covid: improves both training and validation accuracy.



e- DNN Lung tumors



f- DNN Covid

Figure 5: Training and testing accuracy, and the total number of iterations for each model in the Lung and covid datasets, respectively.

## 7. Conclusion

This article discovers common and optimal deep learning architectures by analyzing CT scan pictures. To identify COVID-19 and Lung tumors (Benign, Malignant, Normal) in suspected patients. As a result, the suggested system identifies VGG16 and CNN as the most appropriate models for categorizing the CT scan pictures into (Benign, Malignant, Normal) for Lung tumors dataset. Deep learning at a high level of performance. This work employs architectures such as (CNN, VGG, and DNN). After preprocessing the samples and executing the architectures on the training and validation datasets, the models performed admirably. Deep learning algorithms have been developed to produce effective results by studying thoughts in order to detect COVID and Lung tumors in a person. Classifying COVID and NOCOVID images, as well as Lung tumors (Benign, Malignant, Normal) from a test dataset, should be a cakewalk for any of the models if they live up to the documentation's promises. The efficacy of a model can be evaluated with tools like (accuracy, recall, and the F1-score). For the Lung tumors dataset, the highest accuracy was achieved by (CNN with VGG-16), with a score of (100). The proposed approach, however, has limitations in that it cannot pinpoint specific areas of the lungs that are impacted by COVID. Locating the damaged regions of

the lungs requires a review of a larger dataset, in addition to novel preprocessing approaches and transfer learning models. As a result, the proposed models will be more precise.

**Funding:** “This research received no external funding”

**Conflicts of Interest:** “The authors declare no conflict of interest.”

## References

- [1] A. Oulefki, S. Agaian, T. Trongtirakul, and A. Kassah Laouar, “Automatic COVID-19 lung infected region segmentation and measurement using CT- scans images,” *Pattern Recognition*, vol. 114, 107747, 2021.
- [2] D. Kumar, A. Wong, and D. A. Clausi, “Lung nodule classification using deep features in CT images,” in *12th Conference on Computer and Robot Vision (CRV)*, pp. 133–138, IEEE, 2015.
- [3] S. K. Mathivanan, P. Jayagopal, S. Ahmed et al., “Adoption of e-learning during lockdown in India,” *International Journal of System Assurance Engineering and Management*, pp. 1–10, 2021.
- [4] S. Rajendran and P. Jayagopal, “Assessing COVID19 epidemic outbreak in Tamilnadu and the impact of lockdown through epidemiological models and dynamic systems,” *Measurement*, vol. 169, article 108432, 2021.
- [5] H. Chen and W. Wu. H. Xia, J. Du, M. Yang, and B. Ma, “Classification of pulmonary nodules using neural network ensemble,” in *Advances in Neural Networks*, pp. 460–466, Springer, Guilin, China, 2011.
- [6] J. Kuruvilla and K. Gunavathi, “Lung tumors classification using neural networks for CT images,” *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 202–209, 2014.
- [7] A. Narin, C. Kaya, Z. Pamuk, Automatic Detection of Coronavirus Disease (Covid- 19) Using X-Ray Images and Deep Convolutional Neural Networks, 2020 arXiv preprint arXiv:2003.10849.
- [8] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, “Multi-scale convolutional neural networks for lung nodule classification,
- [9] D.A. Ragab, M. Sharkas, S. Marshall, J. Ren, Breast cancer detection using deep convolutional neural networks and support vector machines, *PeerJ* 7 (2019), e6201.
- [10] Z. Yao, J. Li, Z. Guan, Y. Ye, Y. Chen, Liver disease screening based on densely connected deep neural networks, *Neural Network*. 123 (2020) 299–304.
- [11] I. Pacal, D. Karaboga, A. Basturk, B. Akay, U. Nalbantoglu, A comprehensive review of deep learning in colon cancer, *Comput. Biol. Med.* (2020) 104003.
- [12] X.W. Gao, R. Hui, Z. Tian, Classification of ct brain images based on deep learning networks, *Comput. Methods Progr. Biomed.* 138 (2017) 49–56.
- [13] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 542 (2017) 115–118.
- [14] W. Ausawalaithong, A. Thirach, S. Marukatat, T. Wilaiprasitporn, Automatic Lung tumors prediction from chest x-ray images using the deep learning approach, in: *2018 11th Biomedical Engineering International Conference (BMEiCON)*, IEEE, 2018, pp. 1–5.
- [15] N.M. Elshennawy, D.M. Ibrahim, Deep-pneumonia framework using deep learning models based on chest x-ray images, *Diagnostics* 10 (2020) 649.
- [16] R. Kabilan, V. Chandran, J. Yogapriya et al., “Short-term power prediction of building integrated photovoltaic (BIPV) system based on machine learning algorithms,” *International Journal of Photoenergy*, vol. 2021, Article ID 5582418, 11 pages, 2021.
- [17] Ibrahim W, Abdullaev S, Alkattan H, Adelaja OA, Subhi AA. Development of a Model Using Data Mining Technique to Test, Predict and Obtain Knowledge from the Academics Results of Information Technology Students. *Data*. 2022; 7(5):67.
- [18] Shaymaa Adnan Abdulrahma, Abdel-Badeeh M. Salem, An efficient deep belief network for Detection of Coronavirus Disease COVID-19, *Fusion: Practice and Applications*, Vol. 2 , No. 1 , (2020) : 05-13 (Doi : <https://doi.org/10.54216/FPA.020102>)

- [19] Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
- [20] Hussein Alkattan, Sanjar Abdullaev, El-Sayed M. El-Kenawy. (2023). The «Climate in Weathers» Approach to Processing of Meteorological Series in Mesopotamia: Assessment of Climate Similarity and Climate Change using Data Mining. *Journal of Intelligent Systems and Internet of Things*, 10 (1), 48-65.
- [21] Firuz Kamalov, Said Elnaffar, Aswani Cherukuri, Annapurna Jonnalagadda, Forward feature selection: empirical analysis, *Journal of Intelligent Systems and Internet of Things*, Vol. 11 , No. 1 , (2024) : 44-54 (Doi : <https://doi.org/10.54216/JISIoT.110105>).
- [22] E. Dandıl, M. Çakiroğlu, Z. Ekşi, M. Özkan, Ö. K. Kurt, and A. Canan, “Artificial neural network-based classification system for lung nodules on computed tomography scans,” in 6th International Conference of Soft Computing and Pattern Recognition (soCPar), pp. 382–386, IEEE, 2014.
- [23] B. Gupta and S. Tiwari, “Lung tumors detection using curvelet transform and neural network,” *International Journal of Computer Applications*, vol. 86, p. 1, 2014.
- [24] Ehsan khodadadi, S. K. Towfek, Hussein Alkattan. (2023). Brain Tumor Classification Using Convolutional Neural Network and Feature Extraction. *Fusion: Practice and Applications*, 13(2), 34-41.
- [25] Develop application for prediction COVID-19 using artificial intelligence. Available from: [https://www.researchgate.net/publication/375998830\\_Develop\\_application\\_for\\_prediction\\_COVID-19\\_using\\_artificial\\_intelligence](https://www.researchgate.net/publication/375998830_Develop_application_for_prediction_COVID-19_using_artificial_intelligence) [accessed Dec 23 2023].
- [26] Song, QingZeng, Lei Zhao, XingKe Luo, and XueChen Dou. "Using deep learning for classification of lung nodules on computed tomography images." *Journal of healthcare engineering* 2017 (2017).