



A hybrid VGG 19 and Capsule Network based deep learning model for lung cancer diagnosis using CT scan images

Nandkishor Chhagan Patil¹, Nitin Jagannath Patil²

¹Research Scholar, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India

²Principal, D N Patel College of Engineering, Shahada, India

E-mail address: nand_patil@rediffmail.com, nitinpatil2002@gmail.com

Abstract: Lung cancer is one of the most prevalent causes of carcinoma-related fatalities globally; precise and effective diagnostic instruments are desperately needed. Researchers and physicians continue to encounter difficulties when aiming to employ deep learning models in healthcare environments for recognizing lung cancer, in order to achieve higher sensitivity and accuracy on large data sets, despite a number of advancements in this field. In this work, we enhance the recognition and classification of lung cancer from CT images using a distinct deep learning approach. Here, a hybrid deep learning model that blends the potency of CNNs with the cutting-edge design of capsule networks is proposed. Inspired by recent advancements, particularly the VggCapNet model (VGG16 and Capsule network), our model integrates the VGG19 and Capsule Network architectures to address orientation-related challenges often encountered in traditional CNN-based approaches. The creation, training, and assessment of this hybrid model for lung nodule identification and categorization are our main research goals. We focus on achieving superior Performance indicators, such as high Accuracy, Specificity, F1 score, and Sensitivity with a significant emphasis on reducing false positive rates. We compare the outcomes of our suggested model with the most recent findings in the field to confirm its efficacy. We anticipate that our research has yielded several valuable outcomes, including improved nodule classification accuracy of 99.20%, reduced false positive rates, and minimized training times. This research could lead to more widespread utilization in cancer diagnosis by enhancing early lung cancer detection and developing the field of medical image analysis.

Keywords: Capsule Network, Convolutional Neural Network (CNN), Computed Tomography, Deep Learning, Lung cancer, Visual Geometry Group (VGG19).

1. INTRODUCTION

Lung cancer is difficult to detect in its early stages, thus it diminishes the possibilities of patient survival. Similarly, the treatment of cancer relies on how early the disease is detected so that treatment can prevent the disease from advancing (in stage) and extending to other bodily areas. Lung Cancer can be controlled with good medical care corresponding to the disease's stage, the individual's health, and other factors, medical science offers several therapies including surgery, chemotherapy, and radiography. The survival rate of five years is just 21% of the patient [1]. Numerous investigators are concentrating on methods of exploiting digital images, particularly computed tomography (CT), to identify lung cancer nodules. Numerous academics and scientists are developing automated solutions to assist physicians save time and money.

Image Processing and Artificial Intelligence approaches may be used to process medical field data

using technology solutions to discover and diagnose Lung cancer, which is most desirable in the intended way of preventing Lung Cancer and evaluating the initial stage of this disease which is most desirable for doctors for treatment.

Most of the time, the size and appearance of nodules which can be categorized as benign or malignant provide the first indication of malignancy. Lung nodules less than 3 cm are usually classified as benign nodules, while those greater than 3 cm are classified as lung lumps or malignant lesions. Figure 1 depicts images of benign and malignant lung cancer. The possibility of malignancy can be anticipated based on the nodule's categorization and further parameters. Artificial Intelligence (AI) approaches prove increasingly noteworthy in the diagnosis at the basic level and categorization of different cancer forms. AI is a prominent area for representing human intelligence in a machine [2].

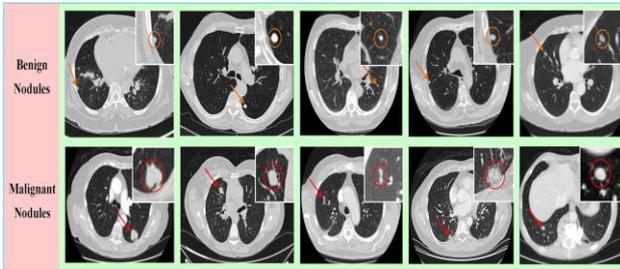


Figure 1. Benign and Malignant nodules in CT images

Machine and deep learning AI algorithms are critical in training a computer system to become an expert that can assist in making predictions and taking decisions. Machine learning is a sub-category of AI which allows a computer system to learn from previous data without having to program it explicitly. "Deep learning" is a kind of ML technique that enables computers to acquire information from data. These fields imbue a computer with intelligence, enabling it to extract patterns based on specific facts and then process them for autonomous reasoning [3].

DL models work well, especially for certain kinds of applications like segmenting images, classifying images, and identifying objects. Digital learning algorithms are not limited to preprogrammed instructions; instead, they may learn from training data. Deep learning models exercised in medical image analysis performed better than machine learning models in digital image localisation, segmentation, anomaly detection, image registration, and diagnosis in a comparative evaluation. Many researchers have already used deep learning to identify lung cancer [4].

An increasingly used form of ANN in computer vision and radiography are referred to as CNN. Neural networks with several layers are called CNNs. CNN is a feed-forward technique that works exceptionally well for identification and classification. The entire network is simpler, and there are smaller training parameters. Because of the network's weight-sharing architecture, CNN seems to be like biological brain networks. Convolution, Activation, Pooling, and a fully linked layer are all layers in the CNN network's structure. In the last layer of CNN, like in classical neural networks, a loss function, such as softmax, is used. CNN employs a sequence of pooling layers to categorize the image. Regardless of the overall amount of spatial data, an item is classified by the CNN's pooling layer, which lowers the dimension. This indicates the actual location of the item in the image. CNN's pooling feature has both benefits and drawbacks. It loses some crucial information that is vital

to object recognition and image segmentation during the pooling process.

The capsule network the notion originated by Geoffrey Hinton as an alternative to this challenge. A cluster of neurons retains data regarding a specific feature in an image is called a capsule. This data mostly relates to the position, rotation, scale, and other characteristics in highly dimensional vectors i.e. 8 or 16. There are three components in the architecture of the capsule network. (1) Upper level capsule (2) loss-related functions and (3) principal capsule (the convolution, shift, and squash operations). The primary goal of capsule networks, or CapsNets, is to acquire spatial data while mitigating the loss of information carried forth by CNN's pooling procedure. CapsNet's several convolutional layers are contained within capsules [5].

The LIDC/IDRI collection has 1018 cases. About 1010 distinct CT images are available altogether since eight patients were unintentionally scanned more than once during the acquisition process. Every bit of image data is recorded in DICOM format, with 512 x 512 pixels as its standard size [6].

The hybrid model VggCapNet for identifying lung nodules from LIDC CT images is discussed in this study. The hybrid system is developed by integrating the capsule network with the VGG-19 and used to classify and recognize lung cancer cases. Based on CNN computer vision algorithms; VGG19 is one of the most effective image model architectures. The number 19 in VGG stands for the number of weight layers it incorporates [7].

2. LITERATURE SURVEY

Shanchen Pang et.al. [8] Clinical importance increases significantly if lung cancer is diagnosed and classified on time. Typically, obtaining a complete diagnosis necessitates the complementary use of numerous imaging modalities. This study proposes a deep learning algorithm to detect lung cancer in patients at Shandong Provincial Hospital based on CT scans. It has two challenges: firstly, the limited amount of patient data that has been collected; secondly, artificial intelligence models based on public datasets are unable to meet such practical requirements. The two-pronged problem is solved by first expanding and balancing our training data using image rotation, translation, and transformation techniques. Next, we classify tumors that are malignant and lastly, we assemble multiple classification results to enhance classification performance using the adaptive boosting (adaboost) algorithm. Based on experimental data, this method outperforms DenseNet without adaboost, ResNet,

VGG16, and AlexNet in terms of identification accuracy, with a score of 89.85%. This enables a successful, non-invasive method for pathological diagnosis of this type of lung cancer.

Imdad Ali et. al. [9] Proposed CNN with transferable textures to enhance the ability to classify lung nodules on CT images. The integration of an Energy Layer (EL) allows for the extraction of textural characteristics from the convolutional layer. By integrating EL, the network's learnable parameter count is lowered, thus lowering memory needs and the complexity of computation. The suggested model only has 3 convolutional layers along with one EL in place of a pooling layer. Using several evaluation matrices, The LUNGx database and the publicly available LIDC-IDRI have been used to evaluate this work. Six-fold cross-validation is used to train model, which yielded an accuracy score of 96.69%, +-0.72%.

Shanchen Pang et. al. [10] First, compile a dataset of 125 early-stage lung cancer patients' CT scans. Revolving, altering, and replicating processes are used to improve the data collection in order to prevent its innate asymmetry. Next, a deep CNN called VGG16-T has been implied and many of these were used to train weak classifiers using a boosting technique. By using joint voting, this strategy significantly improves the performance of CT scans in diagnosing the pathological type of lung cancer. Trials on the expanded dataset of CT scans indicate that VGG16-T, a set consisting of three weak classifiers, can recognize pathological classes with 86.58% accuracy. Additionally, VGG16-T diagnoses 20 randomly chosen CT scans with an accuracy of 85%, whereas two respiratory physicians from Grade 3A hospitals diagnose patients with an accuracy of 55% and 65% using handcrafted diagnoses, respectively.

Cheng Wang et. al [11] the objective of this research was to categorize pulmonary images using the Inception-

v3 TL Model in order to provide a workable and realistic computer-aided diagnosis model. Here, the pulmonary image data was first enhanced. Next, features were automatically extracted using the refined classifiers to categorize the images using the Inception-v3 model, which is based on transfer learning. Subsequently, a number of models that were built on the original Deep CNN model were contrasted with it. This demonstrated the significance of the TL experiment for the categorization of lung images. At 95.41% and 80.09%, respectively, the maximum sensitivity and specificity are recorded.

Heng YU et.al. [12] The Adaptive Hierarchical Heuristic Mathematical Model (AHHMM) has been presented as an approach for deep learning. The system that is suggested is composed up of many steps, including preliminary processing, segmentation, binarization, thresholding, and image extraction, which are followed by extraction of features and DNN detection. The lung CT image has been divided and to extract any relevant properties, a particular feature extraction technique is applied on the fragmented image. The outcome of the evaluation showed that the suggested approach was able to identify 96.67% of lung cancer cases or non-cases.

Abdulrazak Yahya Saleh et. al. [13] in this study, a hybrid CNN and SVM model based image categorization technique is proposed. Every lung image can be automatically recognized and assessed by a machine to ascertain if cancerous cells exist or not. SVM has been used to eliminate unnecessary data that has a negative effect on accuracy. Recently, CNNs perform exceptionally well in a number of computer vision applications. The suggested CNN-SVM system has successfully categorized lung images with an accuracy of 97.91%.

TABLE I. AN OVERVIEW OF CNN'S RESEARCH ARTICLES ON THE USE OF THE DL MODEL TO IDENTIFY LUNG CANCER

Reference	Year	Method	Image used	Image dataset	Accuracy	Findings and limitation
Shanchen pang[8]	2019	Densely Connected CNN with Adaptive Boosting	CT	Shandong Provincial Hospital	89.85%	Densenet is designed to process and categorize the datasets related to lung cancer. In order to enhance classification performance, numerous classification results are finally aggregated using the Adaboost method.
Imdad Ali[9]	2020	Transferable Texture Convolutional Neural Network	CT	LIDC	96.69%	The model only has 3 convolutional layers with a single EL in place of a pooling layer. EL removes the general shape information and examines the texture characteristics. Additionally, the EL decreased the network's learnable parameter count.
Shanchen Pang[10]	2020	VGG16-T	CT	Shandong Provincial Hospital.	86.58%	In order to support lung cancer type, introduces a classification framework based on enhanced VGG16-T. A boosting-based classifier is trained for the second stage with the goal of lowering the FPs generated by the first stage.
Cheng Wang[11]	2019	Inception-v3	CT	JSRT	85.70%	The enhanced Inception-v3 approach employed transfer learning techniques to automatically acquire attributes, and it used multiple classifiers (Soft max, Logistic, and SVM) to classify the pulmonary images.



Heng Yu[12]	2020	Adaptive Hierarchical Heuristic Mathematical Model	CT	DIAG	90%	Divides images into slices within the same image using the Modified K-means method, with the DNN focusing on classification of identical images.
Abdulrazak Yahya Saleh[13]	2021	Hybrid CNN-SVM	CT	LIDC-IDRI	97.9%	SVM has been used to eradicate irrelevant information that adversely impacts accuracy.
Hongtao Xie [14]	2019	Two-Dimensional CNN	CT	LUNA16	86.42%	To help with CT learning, a novel automated approach for pulmonary nodule diagnosis using 2D convolutional neural networks (CNNs) is suggested. Consisting of two phases: (1) enhanced Faster R-CNN for nodule candidate recognition; (2) boosting-based classifier for false positive minimization.
Neal Joshua[15]	2021	3D-AlexNet architecture	CT	LUNA 16	97.17%	It utilizes the capabilities of the 3D AlexNet architecture. To detect lung nodules, a 3D multiple dimensions convolutional neural network is employed.
Parnian Afshar [16]	2020	3D-MCN	CT	LIDC-IDRI	83.00%	To detect lung nodules, a 3D multiple dimensions convolutional neural network is employed.
Zhou Tao[17]	2020	DenseNet	CT	LIDC-IDRI	93.26%	DenseNet (DenseNet-NSCR) nonnegative, sparse, and collaborative representation classification is presented.

A. Problem Statement

Although a number of researchers have already published several deep learning architecture for recognizing lung cancer, attaining the required precision on public datasets continues a significant problem for radiologists assessing malignant nodules and using these models in clinical applications. Although they employed either small online datasets or proprietary datasets gathered from hospitals, multiple investigators have produced work with a high degree of accuracy. Shanchen Pang et al. [8],[10] proposed a DCN was implemented using an unreleased database of 125 individuals diagnosed with lung cancer from Shandong Provincial Hospital. Specifically, VGG16-T with multiple VGG16-T was utilized as a weak classifier trained with a boosting approach. The accuracy achieved by a model, known as "DL based Model for identifying Lung Cancer by Densely Connected Convolutional Networks with Adaptive Boosting algorithm," is 89.85%.

Cheng Wang et al. [11] The Japanese Association of Radiology Technologies conventional digital image database, JSRT, was the source of the dataset for the inception-v3 machine learning model that was suggested. It attained 80.09% specificity and 95.41% sensitivity, respectively. Nevertheless, there were only 154 images with tumours in the lungs and 93 images without in the collection. Heng Yu et al. [12] proposed a technique to use DL technology to predict the existence of lung cancer. Using the unpublished dataset Diagnostic Image Analysis Group, the AHHMM yielded an accuracy of 96.67%. The accuracy of the CNN and Two-Dimensional CNN 3D AlexNet architectures in detecting lung nodules

using the LUNA-16 dataset was 97.17% and 86.42%, respectively [14].

Most of established CNN models just take into account the primary nodule location, ignoring the surrounding tissues. It is difficult to predict lung nodule malignancy with good sensitivity [18]. A novel and innovative machine learning architecture called a capsule network (CapsNet) is suggested as a solution to CNN drawbacks. CapsNet-based architecture can handle a limited quantity of training data. CapsNet can recognize the spatial relationships between multiple components in an image. CNNs require a constant stream of input to learn the translations as they are not resilient to rotation and transformation. Conversely, CapsNets are better equipped to manage rotations and transformations via the capsule layers and the "routing by agreement" method [5].

Based on major contribution and considerations of above several models our works is

- To propose a hybrid model namely VggCapNet combining the features of VGG19 and Capsule networks (CapsNet).
- For CT image pre-processing and feature extraction, a pretrained VGG19 model will be employed.
- The transformation, rotation, and spatial relationship issues with the existing deep learning algorithms will be addressed by a capsule network.
- To prevent overfitting and provide a more complete model, these methods are also used in conjunction to completely connected layers and sigmoid functions of activation.



- The LIDC dataset will be used to assess VggCapNet's performance.

3. MATERIAL AND METHOD

Medical image analysis is no longer unfamiliar with deep learning. It's a developing trend, and there's an increasing need for deep learning to produce precise and accurate outcomes. Figure 2 illustrates the general methodology for using CT scans to diagnose cancer in its primary stages based on its pathological kind. Figure 3 illustrates the suggested framework developed with a capsule network and VGG 19. Getting the lesion data from various lung CT scan classes from the LIDC IDRI dataset is the first step. Pre-processing images, such as image acquisition, augmentation, and color transformation, is included in stage 2. The study and application of many pre-trained CNN models, including VGG16, Xception, Resnet, Inception V3, Mobilenet, etc., comprise Stages III. Stage IV entails computing the previously trained model's parameters.

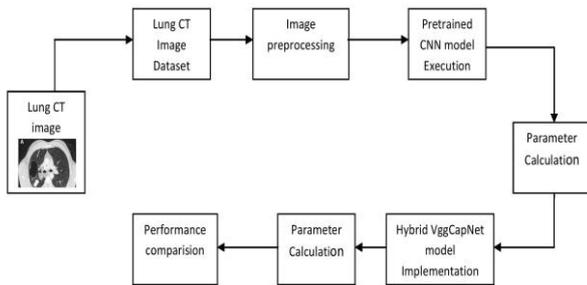


Figure 2. Basic Model for lung cancer detection

Such as specificity, sensitivity, and accuracy and F1 score. The hybrid VggCapNet model is implemented in stages V and VI, and its parameters are estimated. In Stage VII, which is the last stage, VggCapNet uses a typical pretrained model for determining the benign or cancerous kind of CT scan images and evaluates the images performance.

A. Dataset Description

LIDC/IDRI refers to the Lung Image Database Consortium and Image Database Resource Initiative. It is an international resource that may be used online for the development and evaluation of CAD methods for the detection and evaluation of lung cancer.

This data collection was developed in partnership with eight medical imaging organizations and seven academic institutions. It consists of 1018 cases and 244,527 images. Each section contains images as a file in XML format with the annotation process findings, which were finalized in two steps by four expert radiologists in thoracic analysis. At the initial blinded-read stage, every radiologist evaluated every CT image independently,

categorizing lesions into three groups: "nodule with lesion of 3 mm or more," "nodule with lesion less than 3 mm," and "non-nodule lesion with 3 mm or more". Each radiologist in the subsequent unblinded-read phase, evaluated their anonymized ratings and those of the other three radiologists independently in order to make a final conclusion. Four kinds of lung nodules are identified in the LIDC-IDRI dataset. (1) Metastatic Lesion whose underlying malignancy was not lung cancer; (2) Benign (3) Malignant and (4) Unknown (No Label) [11].

The LIDC/IDRI dataset has 1,018 instances in total. Because eight cases were unintentionally duplicated throughout the CT scan collection process, there are really only 1,010 distinct CT scans. Details acquired in image data are saved in DICOM format. The collection's average CT scan width is 240, while the DICOM pictures range from 64 to 764 slices with a consistent dimension of 512 × 512. The thickness of the pictures varies from 0.5 to 5 mm image thicknesses. It is interesting that more than half of the recent studies on lung cancer detection have made use of the LIDC/IDRI dataset.

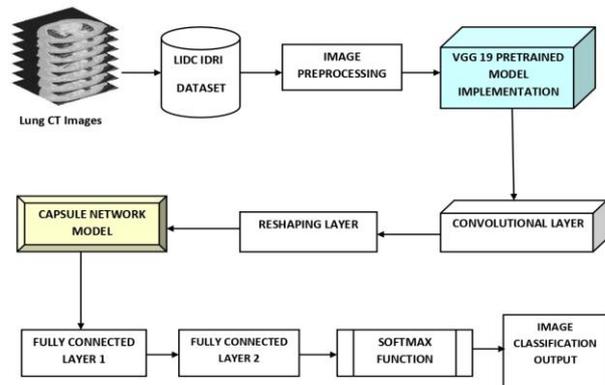


Figure 3. VggCapNet Model using VGG 19 and Capsule Network

B. Data Pre-processing

In order to reduce computing complexity and network overhead, pre-processing of the images is necessary. Deep learning models cannot use a whole CT scan directly because of image dimensionality, format, size, and other issues. There are a few stages involved between acquiring the CT scan and integrating the data into the deep learning model. The LIDC-IDRI dataset includes a variety of CT scan pictures of the lungs, with labeling included in a CSV file. Assigning of malignant and benign images is carried out based on details provided in a CSV file to execute the suggested model. Typically, a CT scan is saved in DICOM format. The Keras ImageDataGenerator package was leveraged to accomplish the rescaling, rotation, flipping both horizontally and vertically, and other data augmentation techniques. Rescaling reduces computing complexity by limiting the picture pixels to a range of 0 to 1. To improve visualization, photos are color

transformed from grayscale to RGB. Rotation of an image by 0 to 10 degrees is possible [18].

Grayscale images may be converted to RGB images by using the Python flow from directory's colour mode approach. The Keras Image Data Generator class, which offers a fast and simple method of augmenting pictures, is used to resize the image from 512x512 to CNN input size. The size of the input image is target size, which is the method's most crucial contention. This property is responsible for converting photos to the specified size for the input. In this instance, we set the target size attribute value of 512x512 for the picture size.

C. Convolutional Neural Network

CNNs are extremely proficient at handling complex problems in many different computer vision fields, like image assessment and categorization. The CNN model takes couple of level to correctly recognize images: firstly, it preprocesses for features extraction from given image, and then it inputs those features into the neural network to classify the image. Convolution, non-linearity pooling, and classification are the primary operations that comprise these two phases. Every CNN is built upon these operations. The widely used architectures for CNN classification are LeNet, AlexNet, VGG, ZF-Net, Resnet, Inception V3, Mobilenet, Xception, etc [19].

The back propagation method is among the best algorithms because CNN is easy to calculate and use. Back propagation makes a distinction between the labels predicted and the ground truth labels ascertained by the loss function during training. The output layer represents the j referred to as output neurons as inputs (a) and n represents total count of training samples.

$$C = -\frac{1}{n} \sum_j \sum_j (y_j a_j^1) (1 + y_j)_n (1 - a_j^1) \quad (1)$$

1) Convolutional Layer

The product of the interconnected neurons weights and the constrained region of input to which they are connected determine the end result of each convolutional layer. Equation (2) provides the output of the cells at the initial convolutional layer. Where b is the bias term and Y^k is the convolutional layer's output feature map for the kth input. * stands for the action of 2D convolution. Let ω represent the weights and x be the input feature map. Usually, the CNN splits the weight of the convolutional layer to combine the chaotic, dense features. These variables are combined by CNN to categorize images.

$$Y^k = f \left(\sum_k x^k * \omega^k + b^k \right) \quad (2)$$

2) Pooling Layer

Pooling facilitates the network's ability to handle massive data volumes. Additionally, it frees the network from minor deviations, distortions, and translations in the input by collecting the key information from it. This minimizes the size of feature map. This in turn reduces the calculations and weights, causing the network to become overfit. Different forms of pooling exist: The name of the pooling depends on the operation Max Pooling, Average Pooling, or Sum Pooling (3) (4).

$$W_2 = \left(\frac{W_1 + 1}{S} \right) + 1 \quad (3)$$

$$H_2 = \left(\frac{H_1 + 1}{S} \right) + 1 \quad (4)$$

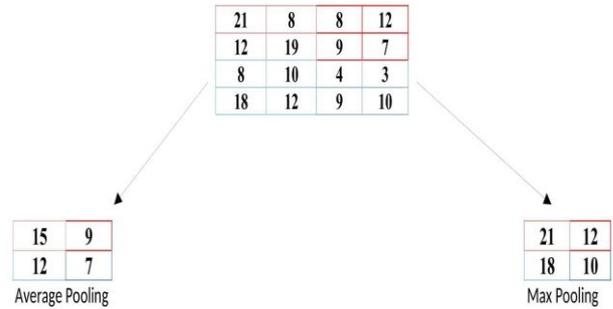


Figure 4. Average Pooling and Max Pooling

3) Fully-connected layer

Because of the flattened input, every neuron in the FC is coupled to every other neuron. The flattened vector is then sent over a few more FC levels, which are typically where the functional operations related to mathematics are carried out. This is the moment where the classifying process begins. If FC layers are present, they are usually located close to the end of CNN architectures. Equation below, where l and (l - 1) are designated as completely linked layers, gives the fully connected operations. The weight connections between the jth unit of layer (l-1) in (r, s) location and the ith unit in layer l and y (5)

$$Y_i^l = f(Z_i^l) \text{ with } Z_i^l = \sum_{j=1}^{m^{l-1}} \sum_{r=1}^{m^{l-1}} \sum_{s=1}^{m^{l-1}} W_{i,j,r,s} (Y_j^{l-1})_{r,s} \quad (5)$$

4) Activation Function

A neuron's function of activation determines whether or not to activate it. This suggests that, at the prediction stage, it will ascertain the relevance of the neuron's input to the network. There are several popular activation functions that are available, such as the Sigmoid, ReLU, Softmax, and tanH functions.



5) Relu Activation

Rather than a local connection, a nonlinear activation function governs the output of neurons in CNN. Because of their improved performance, ease of learning, and simple construction, the logistic sigmoid and hyperbolic tangent type of activation functions are beneficial.

$$f(y) = \max(0, x) \tag{6}$$

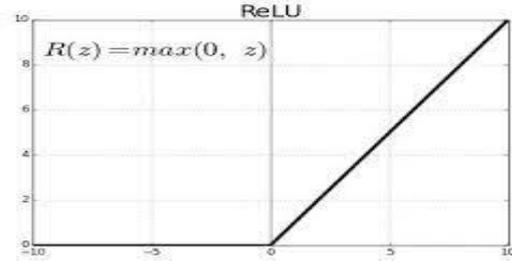


Figure 5. Relu Activation

6) Softmax Activation

Final activation function in a neural network, it is frequently utilized in multinomial logistic regression to transform the network's output into a distribution of likelihood over the anticipated output classes. The likelihood that any ground truth identify will produce an output value between 0 and 1 is calculated using Softmax, and the outcomes are then translated into numerical form (7).

$$F(z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{K_k}} \tag{7}$$

D. VGG 19 model

The acronym for Visual Geometry Group is VGG. The convolution performed by the 19 deep trainable layers that make up the VGG-19 model is completely coupled to the max pooling and dropout layers. It comprises of 19 layers, including 16 convolution layers, 5 MaxPool layers, 1 SoftMax layer, 3 completely connected layers, and so forth. In order to predict 1000 labels, VGG is composed of two completely connected layers; each one has a total of 4096 channels. There is one more completely linked layer with one thousand channels after that. The final fully interconnected layer uses the Softmax layer for classification [18].

E. Capsule Network

The architecture of a CapsNet is made up of capsules rather than neurons. Capsules produce vectors with direction, as opposed to neurons, which only produce scalars. CNN orienting issues are resolved in part by this capsule feature [5].

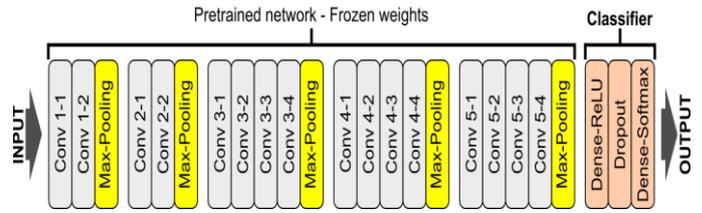


Figure 6. VGG 19 model

A cluster or collection of neurons known as a capsule is used to store data on a particular object in an image. This data mostly consists of the object's position, rotation, scale, and other attributes.

Three components make up the capsule network.

Convolution, reshape, and squash functions comprise the primary capsule.

Dynamic routing is found in the higher level capsule.

(3) Margin and reconstruction loss are the loss functions.

The main purpose of capsule networks, or CapsNets is to acquire geographical information while reducing the loss of information brought on by CNN's pooling procedure.

F. Capsule Network Architecture

The capsule network's design is a shallow network with the three components as Convolutional layer via the ReLU activation function, this layer extracts the feature map of the input data via a traditional convolution operation.

Primary caps layer: Using a linear combination, this layer divides the whole feature map into capsules, which are then covered by a convolution layer and reshaped. Following convolution, all maps undergo reshaping to classify them into groupings known as capsules. The calculated and enlarged dimensionality size of n is computed by multiplying each capsule output by a weight matrix of m * n, where each capsule vector's dimension is set to m. When the main capsules are moved to the digit capsules, the dimensionality rises because m is smaller than n.

Digitcaps layer: Relationships between the various levels of the hierarchical capsule are established. Weights are updated by applying the dynamic routing algorithm and squash operation to the capsule layers. The vectors of the output capsules that show the dimension

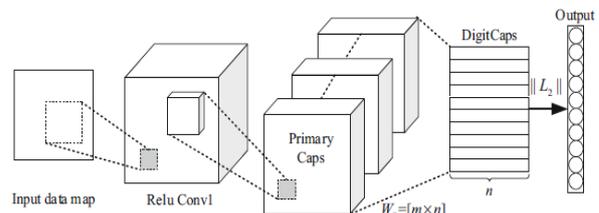


Figure 7. Capsule Network Architecture



G. Metrics for Performance Evaluation

Numerous indicators of efficiency such as accuracy, sensitivity, specificity and F1-score, were used to assess the suggested hybrid DL model for identifying and categorizing lung cancer. The total number of valid findings divided by the overall number of lung image sample counts yields the overall accuracy of this study (8).

True-positive (TP) results, as seen from a medical standpoint, signify that individuals with abnormal lesions are examined following a predefined assessing protocol and are exactly categorized as having weird lesions. Conversely, a positive result means that the patient either carries the virus or has abnormal lesions. FP diagnosis is the medical misinterpretation of normal lesions that is wrongly recognized as abnormal lesions. A patient that tests negative (i.e., not abnormal) or true-negative (TN) means that they do not have any abnormal lesions and are considered normal. A false-negative (FN) scenario occurs where a patient has an abnormal lesion is mistakenly classified as a normal sample, misleading for diagnosis [17].

Sensitivity/recall or TP rate is the probabilities of patients with abnormal lesions are identified as medically positive. It must be high valued practically for much sensitive identification (9).

Specificity rate is the probability that an individual will receive a negative diagnostic if they do not have abnormal lesions. Likewise, this figure ought to be as high as is practical. Stated differently, a diagnosis's accuracy increases with its value (10).

According to Aurelien Geron in Hands on Machine Learning, the F1 Score relates to the "harmonic mean of precision and sensitivity."(11)

$$accuracy = \frac{TN+TP}{TN+FP+FN+TP} \quad (8)$$

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

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

$$F1\ score = \frac{TP}{\frac{1}{2}(FN+FP)+TP} \quad (11)$$

AUC: The acronym AUC refers to "Area under the ROC Curve." A ROC curve, also known as the operational characteristic of the receiver curve, it displays effectiveness of a classification models across all criteria.

4. EXPERIMENTAL SETUP

The Models such as MobileNet, Xception, VGG-16, VGG-19 and Inception V3 that were trained on the LIDC dataset to detect lung cancer and its classification are used for the comparison. Additionally, the new model VggCapNet is created and put to use in experiments. Based on a hybrid design that blends the VGG19 and Capsule Network, comes the VggCapNet architecture.

A. Mobilenet

MobileNets generally use depthwise separable convolutions rather than the regular found in earlier structures to provide lighter models. The width multiplier and resolution multiplier, two new global hyper parameters introduced by MobileNets, Permit tradeoff for model creators in accuracy or speed-related delay to achieve compact size network, depending on the demands. Each depthwise separable CN layer consists of a pointwise and a depthwise convolution respectively. MobileNet has 28 layers [20].

B. Xception

"Extreme Inception" is what Architecture Xception stands for. The network's for feature extraction is made up of 36 convolutional layers. The architecture of Xception is made up of a several stack of convolutional layers that may be separated based on depth and include residual connections. With the same amount of parameters as Inception V3, the Xception architecture performs better because it makes better use of its model parameters rather than having greater capacity [21].

C. VGG-16

VGG is the abbreviation for Visual Geometry Group. Extensive Image Identification using Very Deep Convolutional Networks is known as VGG. One of the most often used pre-trained models for image categorization is the VGG-16. In ImageNet, the VGG16 model has around 92.7% top-5 test accuracy. An extensive collection of over 14 million photos from about 1000 types makes up ImageNet. With a succession of smaller 3x3 filters in place of the huge filters, VGG16 outperforms AlexNet. For the first convolutional layer in AlexNet, the kernel size is 11, while for the second layer, it is 5. VGG16, with 138 million parameters overall, is thus a reasonably large network. It comprises 13 convolution layers, 5 max-pooling layers, and 3 dense layers in the VGG16 pre-trained design. The dropout rate for both thick layers is 0.5 [10].



D. VGG-19

For the VGG-19 model, the convolution is fully coupled to the max pooling and dropout layers by the 19 deep trainable layers which includes nineteen layers: sixteen convolution layers, five MaxPool layers, one SoftMax layer, and three completely connected layers. There are two completely linked layers in VGGNet, each having 4096 channels, to predict 1000 labels. There is one more completely linked layer with 1000 channels after that. The final completely connected layer uses the Softmax layer for classification [18].

E. Inception V3

Reducing the computational burden without affecting the deeper network's generalization was the goal of Inception-V3. It used asymmetric small-size filters in place of large-size filters. Due to these modifications, cross-channel correlation and the conventional convolution operate quite similarly. Using the one by one convolutional technique the input data are mapped into 3 or 4 separate spaces that are tiny than the original input areas. This model can achieve an accuracy of more than 78.1 percent using the ImageNet dataset [11] [33].

F. VggCapNet Model Building and Implementation

Tensorflow and Keras, two Python libraries, are used to implement the suggested hybrid model on the Google Colab platform. The CNN model based on transfer learning [24] on pre-trained design is fitted to VGG-19 by including ReLu activation function based on dense layer and global average pooling layers [33]. Route 03 has been added to a capsule network that comes after the VGG19. Our findings indicate that our proposed hybrid model VggCapNet functions more efficient than the MobileNet, Xception, Inception V3 and VGG models.

G. Capsule Network Deployment

The capsule network implementation uses the pooling layer based on global median as its input. The features are then extracted using a 2-dimensional convolution layer with a kernel size of 9, 9,512. Then, using the reshape function is utilized for normalization these characteristics are transformed into an array of one-dimensional. Next, lambda is utilized to show that the lung image features had a normal value of 0. The fundamental architecture of the capsule network (CapsNet), which is utilized to analyze lung nodules, is seen in Figure 7

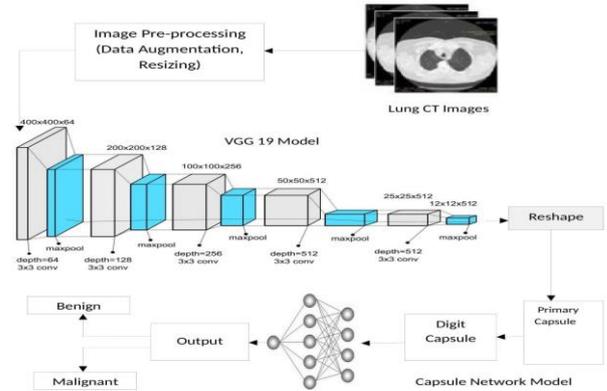


Figure 8. Architecture of the VggCapNet

5. RESULT AND DISCUSSION

This paper presents a unique hybrid DL model, named VggCapNet, to detect lung cancer from CT images. LIDC-IDRI dataset CT images are processed using the VggCapNet model. Out of all the pretrained models that are currently in use, the VggCapNet architecture has the best accuracy at 99.20%. Pretrained models such as MobileNet, Xception, VGG-16, VGG-19 and Inception V3 had accuracy percentages of 98.00%, 97.97%, 96.95%, 95.60%, and 94.35%, in that orders. An evaluation set of lung imaging data is used to evaluate the constructed hybrid model.

Table II displays the classification matrix. It suggests that VggCapNet, the hybrid model, outperformed the other models. Consequently, the hybrid VggCapNet that was recommended achieved the best accuracy of 99.20 percent.

TABLE II. CLASSIFICATION MATRIX UTILIZING THE LIDC DATASET FOR SEVERAL MODELS.

Model	Accuracy	F-Measure	Sensitivity	Specificity	AUC
MobileNet	98.00 ± 0.40	98.00 ± 0.55	98.15 ± 0.65	98.00 ± 0.40	98.00 ± 0.35
Xception	97.97 ± 1.02	96.87 ± 1.82	97.47 ± 1.05	97.97 ± 1.02	96.97 ± 1.02
VGG-16	96.95 ± 1.75	96.95 ± 1.55	95.95 ± 1.75	96.95 ± 1.25	95.95 ± 1.84
VGG-19	95.60 ± 1.00	95.55 ± 1.05	95.10 ± 1.00	95.60 ± 0.75	95.50 ± 1.10
Inception V3	94.35 ± 1.1	94.35 ± 1.15	94.35 ± 1.2	94.35 ± 1.10	94.35 ± 1.1
VggCapNet	99.20 ± 0.2	99.79 ± 0.5	99.50 ± 0.5	99.20 ± 0.6	99.15 ± 0.6

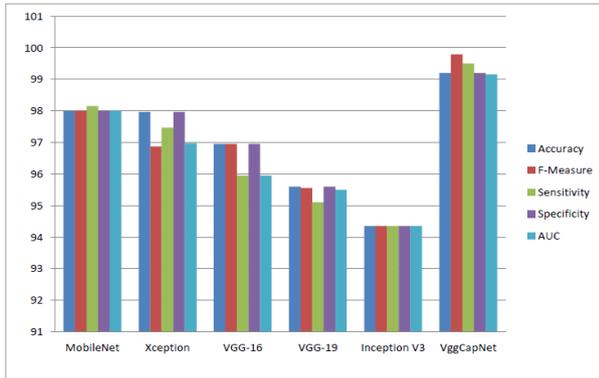


Figure 9. Comparison of efficiency with existing methodologies

• Discussion

As many deep learning models can assist in lung cancer diagnosis, accuracy in publically accessible datasets continues to be a significant challenge. The researchers findings were correct overall, despite the fact that some of them used tiny datasets that were freely accessible online or private hospital datasets. It takes a great deal of diligence to identify malignant lung nodules, particularly if you have never done this before. Despite recent advances, the clinical implementation of DL models for malignancy detection is still challenging. DL models are utilized to detect the lung's core nodule in most situations, disregarding the surrounding tissue. This section presents the outcomes of the VggCapNet architecture for the detection and classification of lung nodules as benign or malignant based on lung CT images in comparison with pre-trained CNN MobileNet, Xception, VGG-16 VGG-19, Inception V3 models. After preparing the LIDC-IDRI dataset, there were no problems while using the models.

In this study 30% of the dataset is utilized to assess the proposed VggCapNet designs, while the remaining 70% is used as a training set. The performance parameters such as Accuracy, F1 scores, sensitivity, specificity and AUC for each model are summarized in Table II.

Figure 10 displays confusion matrix for Predicted label and True label for images. Figure 11, Figure 12 and Figure 13 displays the validation accuracy, loss curves and ROC curve for the VggCapNet model respectively. Accuracy and losses in training and validation continue to converge, peaking after 50 epochs. A strong connection between the actual and predicted values was found in the networks with less than 0.1% error rate and test accuracy of less than 0.1% for the proposed VggCapNet model when evaluated.

In comparison to existing methods MobileNet, Xception, VGG-16, VGG19 and inception V3 the suggested hybrid VggCapNet model assessed on the CT images collected dataset demonstrated greater accuracy. Any medical system must be able to generate precise results since this framework serves therapeutic objectives and may be used to identify lung nodules in lung cancer.

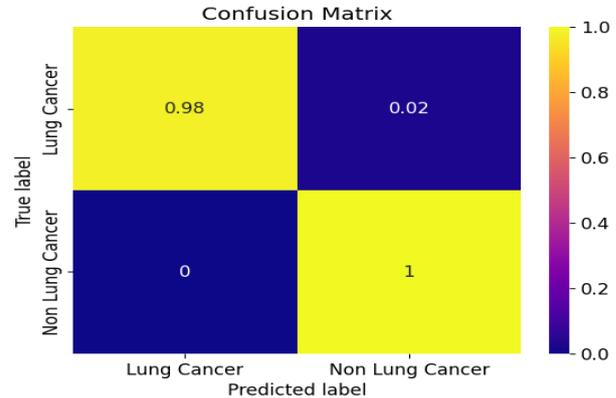


Figure 10. Confusion Matrix

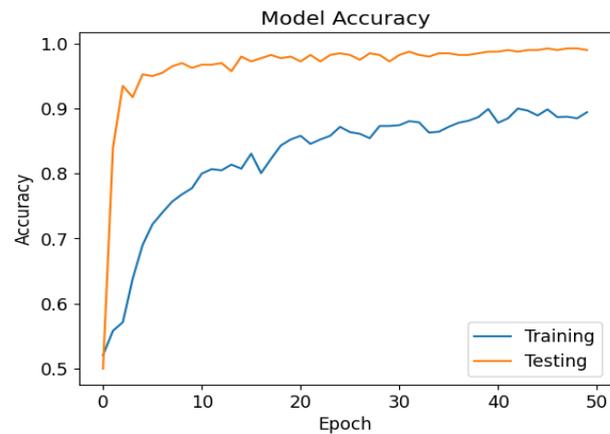


Figure 11. Model Accuracy

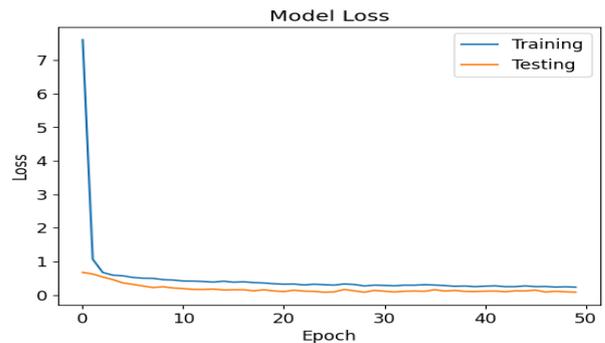


Figure 12. Model loss

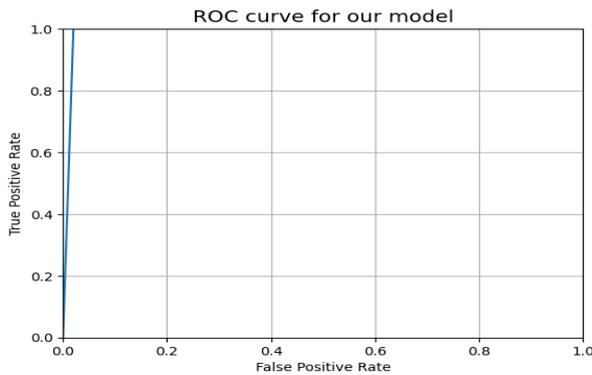


Figure 13. ROC Curve

6. CONCLUSION

Considering the remarkable outcomes of deep learning techniques, this research examined lung cancer detection based on CT images from the LIDC-IDRI dataset using CNN-based hybrid VggCapNet architecture, one of the medical research's most effective DL models. For this, before the model was used, an image dataset was used to create a series of CT images. The proposed hybrid VggCapNet model is applied on this series of CT images of LIDC-IDRI dataset. Although the existing approaches of MobileNet, Xception, VGG-16, and VGG-19 and inception v3 obtained an accuracy of 98.00%, 97.97%, and 96.95%, 95.60% and 94.35% respectively, the suggested VggCapNet architecture offered a greater accuracy of 99.20%. Radiologists can anticipate our hybrid VggCapNet framework to be extremely useful and time-saving when employing it for clinical investigations to identify lesions in lung carcinomas.

REFERENCES

- [1] R. L. Siegel, K. D. Miller, H. E. Fuchs and A. Jemal, "Cancer statistics 2021", *CA Cancer J. Clin.*, vol. 71, no. 1, pp. 7-33, 2021. doi.org/10.3322/caac.21654
- [2] Agrawal S, Jain SK. "Medical text and image processing: applications, issues and challenges", *Machine Learning with Health Care Perspective*. Cham: Springer (2020). p. 237–62. doi: 10.1007/978-3-030-40850-3_11
- [3] Dakhaz Mustafa Abdullah, Nawzat Sadiq Ahmed, "A Review of most Recent Lung Cancer Detection Techniques using Machine Learning", *International Journal of Science and Business*, Volume 5, issue 3, pp.159-173,2021. doi:10.5281/zenodo.4536818
- [4] Sarfaraz Hussein, Pujan Kandel, Candice W. Bolan, Michael B. Wallace, and UlasBagci, "Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches", *IEEE Transactions on Medical Imaging*, Volume: 38, Issue: 8, pp.1777-1787,2019. <https://arxiv.org/pdf/1801.03230.pdf>
- [5] Sabour S, Frosst N, Hinton GE. "Dynamic routing between capsules. In: *Advances in Neural Information Processing Systems*", p. 3856–66,2020. <https://arxiv.org/pdf/1710.09829.pdf>
- [6] Armato Samuel G III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. "Data From LIDC-IDRI". (2015). Available online at: [https:// wiki.cancerimagingarchive.net /display/Public/LIDC-IDRI](https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI) (accessed January 7, 2020).
- [7] Gaurav Kumar, Shraban Chatterjee, Chiranjoy Chattopadhyay, "DRISTI: a hybrid deep neural network for diabetic retinopathy Diagnosis", *Signal, Image and Video Processing Springer Nature* 15:1679–1686. 2021. doi:10.1007/s11760-021-01904-7
- [8] Shanchen Pang, Yaqin Zhang, Mao Ding, Xun Wang, and XianjinXie, "A Deep Model for Lung Cancer Type Identification by Densely Connected Convolutional Networks and Adaptive Boosting", *IEEE Access*, Volume 8, pp. 4799-4805, 2020. doi:10.1109/ACCESS.2019.2962862
- [9] Imdad Ali , Muhammad Muzammil ,Ihsan Ul Haq , Amir A. khaliq ,and Suheel Abdullah, "Efficient Lung Nodule Classification Using Transferable Texture Convolutional Neural Network",*IEEE access*, 2020, doi:10.1109/ACCESS.2020.3026080
- [10] Shanchen Pang, Fan Meng, Xun Wang, Jianmin Wang, Tao Song, Xingguang Wang and Xiaochun Cheng, "VGG16-T: A Novel Deep Convolutional Neural Network with Boosting to Identify Pathological Type of Lung Cancer in Early Stage by CT Images", *International Journal of Computational Intelligence Systems*, Vol. 13(1), pp. 771–780, 2020, doi :10.2991/ijcis.d.200608.001
- [11] Cheng Wang, Delei Chen1, Hao Lin, B. Xuebo Liu, Yu Zeng, Jianwei Chen, and E. Guokai Zhang, " Pulmonary Image Classification based on Inception-v3 Transfer Learning Model" , *IEEE Access*, Volume 7, pp. 146533 – 146541, 2019, doi: 10.1109/access.2019.2946000
- [12] Heng Yu, Zhiqing Zhou, Qiming Wang, "Deep Learning Assisted Predict of Lung Cancer on Computed Tomography Images Using the Adaptive Hierarchical Heuristic Mathematical Model, Deep Learning Algorithms for Internet of Medical Things", *IEEE Access*, volume 8, pp.86400-86410,2020. doi:10.1109/ACCESS.2020.2992645
- [13] Abdulrazak Yahya Saleh , Chee Ka Chin, Vanessa Penshie , Hamada Rasheed Hassan Ali, " Lung cancer medical images classification using hybrid CNN-SVM", *International Journal of Advances in Intelligent Informatics*, Vol. 7, No. 2, July 2021, doi.org/10.26555/ijain.v7i2.317
- [14] Xie H, Yang D, Sun N, Chen Z, Zhang Y., "Automated Pulmonary Nodule detection in CT images using Deep Convolutional Neural Networks", *Pattern Recognition*, Elsevier ,85 pp.109–19, 2019 doi: 10.1016/j.patcog.2018.07.031
- [15] Neal Joshua ES, Bhattacharyya D, Chakkravarthy M, Byun YC. "3D CNN with visual insights for early detection of lung cancer using gradient-weighted class activation", *J Healthcare Eng.* (2021). doi: 10.1155/2021/6695518.
- [16] Afshar P, Oikonomou A, Naderkhani F, Tyrrell PN, Plataniotis KN, Farahani K, and Arash M, "3D-MCN: a 3D multi-scale Capsule Network for lung nodule malignancy prediction", *Scientific Report*, 10:7948, 2020, doi: 10.1038/s41598-020-64824-5
- [17] Zhou Tao ,Huo Bingqiang,Lu Huiling ,Yang Zaoli and Shi Hongbin, "NSCR-Based DenseNet for Lung Tumor Recognition Using Chest CT Image" , *BioMed Research International*, Hindawi, Volume 2020, Article ID 6636321, 2020, <https://doi.org/10.1155/2020/6636321>
- [18] Muhammad Attique Khan , Venkatesan Rajinikanth , Suresh Chandra Satapathy , David Taniar ,Jnyana Ranjan Mohanty , Usman Tariq and Robertas Damaševičius , "VGG19 Network Assisted Joint Segmentation and Classification of Lung Nodules in CT Images",*MDPI Diagnostics* ,Volume 11, 2021, doi.org/10.3390/diagnostics11122208
- [19] Hamdalla F. Al-Yasriy, Muayed S. AL-Husieny, Furat Y. Mohsen, Enam A. Khalil , Zainab S. Hassan, "Diagnosis of Lung Cancer Based on CT Scans Using CNN", *2nd International*

Scientific Conference of Al-Ayen University, IOP Conf. Series: Materials Science and Engineering 928 (2020) 022035, doi:10.1088/1757-899X/928/2/022035

- [20] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, "Mobilenets: Efficient Convolutional Neural Networks for mobile vision applications", arxiv:1704.04861, 2017, doi: 10.48550/arXiv.1704.04861
- [21] Chollet F., "Xception: Deep Learning with Depthwise Separable Convolutions", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1800-1807, 2017, doi: 10.1109/CVPR.2017.195
- [22] Bharati S, Podder P, Mondal MRH, "Hybrid deep learning for detecting lung diseases from X-ray images", Informatics in Medicine Unlocked , 20:100391, 2020 doi: 10.1016/j.imu.2020.100391
- [23] Guobin Zhang , Zhiyong Yang, Li Gong, Shan Jiang1 and Lu Wang, "Classification of benign and malignant lung nodules from CT images based on hybrid features", Physics in Medicine and Biology, 64(19) 125011 , 2019, doi.org/10.1088/1361-6560/ab2544
- [24] Aashka Mohite, "Application of Transfer Learning Technique for Detection and Classification of lung cancer using CT Images", International Journal of Scientific Research and Management , Volume 09, Issue 11,pp. 621-634, doi: 10.18535/ijrm/v9i11.ec02
- [25] Asmaul Hosna, Ethel Merry, Jigmei Gyalmo, Zulfikar Alom, Zeyar Aung and Mohammad Abdul Azim, "Transfer learning: a friendly introduction", Hosna et al. Journal of Big Data,2022, doi.org/10.1186/s40537-022-00652-w
- [26] Nandita Gautam, Abhishek Basu, Ram Sarkar. "Lung cancer detection from thoracic CT scans using an ensemble of deep learning models", Neural Computing and Applications, 2023. <https://link.springer.com/article/10.1007/s00521-023-09130-7>.
- [27] Samuel G. Armato III, Geoffrey McLennan, Michael F. McNitt-Gray, Charles R. Meyer et al. "Lung Image Database Consortium: Developing a Resource for the Medical Imaging Research Community", Radiology, 2019. doi: 10.1148/radiol.2323032035
- [28] Bhagya Lakshmi Nandipati, Nagaraju Devarakonda. "Hybrid deep learning model for detection and classification of lung cancer fusion images using MCNet", Journal of Intelligent & Fuzzy Systems, 2023. doi: 10.3233/JIFS-231145
- [29] Huseyin Polat, Homay Danaei Mehr. "Classification of Pulmonary CT Images by Using Hybrid 3D-Deep Convolutional Neural Network Architecture", Applied Sciences, 2019. doi:10.3390/app9050940
- [30] Phuoc Pham, Huilan Zhang, Wenlian Gao, Xiaowei Zhu. "Determinants and performance outcomes of artificial intelligence adoption: Evidence from U.S. Hospitals", Journal of Business Research, 2024. doi.org/10.1016/j.jbusres.2023.114402
- [31] Bahauddin Taha, Fahmida Rahman Liza, Md Moniruzzaman, Md Samsuzzaman. "Comparative Analysis of Fine-Tuned MobileNet versions to Detect Brain Tumors from MRI images", 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), 2022. doi: 10.1109/STI56238.2022.10103254
- [32] Shereen El-Shekheby. "Chapter 33 Real-Time Facial Emotion Recognition Using Haar-Cascade Classifier and MobileNet in Smart Cities", Springer Science and Business Media LLC, 2024. doi:10.1007/978-3-031-46491-1_33
- [33] Himanshu B. Patel , Nitin J. Patil, "An Intelligent Grading System for Automated Identification and Classification of Banana Fruit Diseases Using Deep Neural Network", International Journal of Computing and Digital Systems, Vol. 15,1, 2024.



Nandkishor Chhagan Patil is a Ph.D research scholar at Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India in a subject of Electronics Engineering. Received a B. E. (Electronics and Telecommunication Engineering) from Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India in 2004 and M.E. degree in Electronics Engineering from Government College of Engineering, Aurangabad in 2011. Presently working as Assistant Professor at D N Patel College of Engineering, Shahada, Maharashtra, India since 2004. Research Interest Includes Artificial Intelligence in Image Processing.



Nitin Jagannath Patil received the B.E. degree in instrumentation engineering from the D. N. Patel College of Engineering, Shahada, Nandurbar Maharashtra, India, in 1997, and the M.E. and Ph.D. degrees in instrumentation engineering from Swami Ramanand Teerth Marathwada University, Nanded, India, in 2000 and 2013, respectively. He is currently a Principal and a Professor with the Department of Instrumentation Engineering, D. N. Patel College of Engineering. His areas of interest include fuzzy logic, control system, image processing, and signal processing.