INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 13-04-2024

Accepted: 06-08-2024

Published: 28-08-2024

Citation: Patil NC, Patil NJ (2024) A Hybrid VGG 19 and Capsule Network Based Deep Learning Model for Lung Cancer Diagnosis using CT S can Images. Indian Journal of Science and Technology 17(35): 3623-3635. https://doi.org/10.17485/IJST/v17i35.1234

* **Corresponding author**.

nand_patil@rediffmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

A Hybrid VGG 19 and Capsule Network Based Deep Learning Model for Lung Cancer Diagnosis using CT S can Images

Nandkishor Chhagan Patil[1]*, Nitin Jagannath Patil[2]

1 Research Scholar, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, Maharashtra, India
2 Professor, D N Patel College of Engineering, Shahada, Maharashtra, India

Abstract

Objectives: 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 and to achieve higher sensitivity and accuracy on large data sets, despite several advancements in this field. Our study builds upon an extensive review of existing lung cancer detection methods, highlighting their strengths and weaknesses. The performance of this model deteriorates when there are variances in Lung CT image features such as rotation, tiling, and other aberrant image orientations. This study aims to address the issues associated with variations in lung CT images. **Methods**: 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, 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. **Findings**: 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. **Novelty**: This research employing a hybrid deep learning model that combines the capabilities of the VGG-19 and capsule network could lead to more widespread utilization in cancer diagnosis by enhancing early lung cancer detection and developing the field of medical image analysis. The shortcomings of convolutional neural networks are addressed with Capsule Network. Leveraging the knowledge and insights gained from lung cancer detection, our research also investigates the potential applicability of our methodology to identify other types of medical image classification such as breast and ovarian cancer.

Keywords: Computed Tomography; Lung cancer; Deep Learning; CNN; VGG19; Capsule Network

1 Introduction

As lung cancer is difficult to detect in its early stages, it diminishes the possibility 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. The survival rate of five years is just 21% of the patients[1]. 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. Figure 1 depicts images of benign and malignant lung cancer. Artificial Intelligence (AI) approaches prove increasingly noteworthy in the diagnosis at the basic level and categorization of different cancer forms[2].

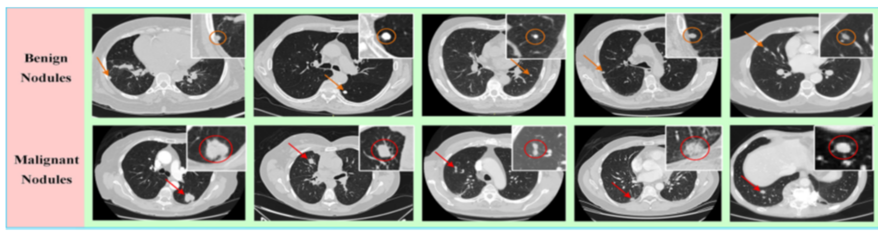

Fig 1. Benign and Malignant nodules in CT images

Machine and deep learning AI algorithms are critical in training a computer system to become an expert that can assist in making predictions and making decisions. Machine learning is a sub-category of AI that 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 objectsDeep learning models exercised in medical image analysis performed better than machine learning models in digital image localization, segmentation, anomaly detection, image registration, and diagnosis in a comparative evaluation[4].

CNN is a feed-forward technique that works exceptionally well for identification and classification. 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. Regardless of the overall amount of spatial data, an item is classified by 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. The capsule network notion originated by Geoffrey Hinton. A cluster of neurons that retains data regarding a specific feature in an image is called a capsule. 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)[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]. 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].

Shanchen Pang et al. [8] Clinical importance increases significantly if lung cancer is diagnosed and classified on time. 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. 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%.

Imdad Ali et al. [9] Proposed CNN with transferable textures to enhance the ability to classify lung nodules on CT images. 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. The LUNGx database and the publicly available LIDC-IDRI have been used to evaluate this work. Six-fold cross-validation was used to train the model, which yielded an accuracy score of 96.69% +/- 0.72%.

Shanchen Pang et al. [10], 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 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. 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 of many steps, including preliminary processing, segmentation, binarization, thresholding, and image extraction, which are followed by extraction of features and DNN detection. 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], a hybrid CNN and SVM model-based image categorization technique 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 hurts accuracy. Recently, CNNs have performed exceptionally well in several computer vision applications. The suggested CNN-SVM system has successfully categorized lung images with an accuracy of 97.91%.

Table 1. An overview of CNN's research articles on the use of the DL model to identify lung cancer

Reference no.	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. 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%	To support lung cancer type, introduces a classification framework based on enhanced VGG16-T. A boosting-based classifier is trained for the second stage to lower the FPs generated by the first stage.

Continued on next page

Table 1 continued

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 the 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%	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.

Although several researchers as shown in Table 1 have proposed several deep learning architectures 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. Some researchers employed either small online datasets or proprietary datasets gathered from hospitals, which resulted high degree of accuracy.

When it comes to identifying an image's location, texture, and distortions, CNN performs an inadequate performance. The CNNs are implied to be invariant with this. CNN's pooling stage might result in invariance. Also, most established CNN models just take into account the primary nodule location, ignoring the surrounding tissues i.e. it cannot recognize the spatial relationships between multiple components in an image. Hence it is difficult to predict lung nodule malignancy with good sensitivity [18]. Instead of scalar values, capsules are equivariant networks of neurons that input and output vectors. It can learn about the characteristics, viewing situations, and deformations of the image by using this capsule capability.CapsNets are better equipped to manage rotations and transformations via the capsule layers and the ”routing by agreement” method. Hence, Capsule Networks integrates and amends for each deficit of CNN.

In this paper, we propose a novel and innovative hybrid deep learning architecture by integrating the capsule network with the VGG-19 CNN as a solution to overcome CNN drawbacks. VGG19 is one of the most effective image model architectures while CapsNet can recognize the spatial relationships between multiple components in an image. The hybrid model is used to classify and recognize lung cancer from the standard LIDC-IDRI dataset as more than half of the recent studies on lung cancer detection have made use of the LIDC/IDRI dataset [19]. The LIDC-IDRI dataset is a completed reference database of lung nodules on CT scans developed by the National Cancer Institute (NCI), the Foundation for the National Institutes of Health (FNIH), and the Food and Drug Administration (FDA).

Based on major contributions and considerations several models our works focus on

- 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 with completely connected layers and sigmoid functions of activation.
- The LIDC dataset will be used to assess VggCapNet's performance with state of art models.

2 Methodology

2.1 Model for Lung Cancer Detection

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.

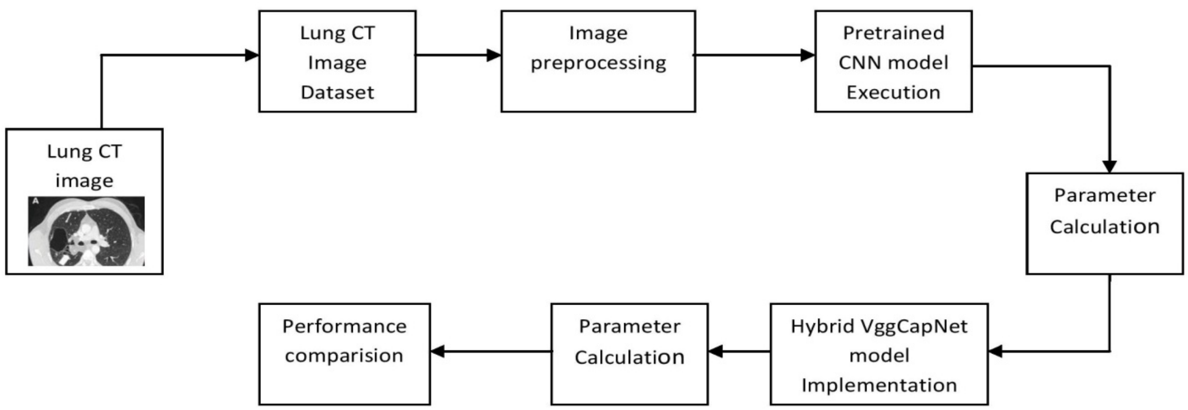

Fig 2. Basic Model for Lung Cancer Detection

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, are 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. Such as specificity, sensitivity, 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 performance of the image.

2.2 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 to conclude. 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]. 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 $\times$ 512. The thickness of the pictures varies from 0.5 to 5 mm image thicknesses.

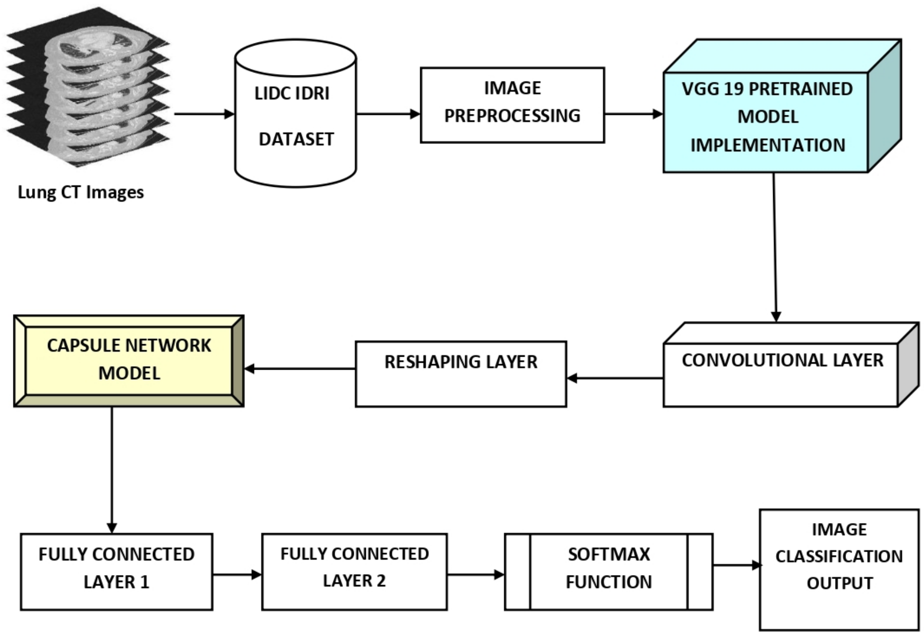

Fig 3. VggCapNet Model using VGG 19 and Capsule Network

2.3 Data Preprocessing

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[20].

Grayscale images may be converted to RGB images by using the Python flow_from directory's colour mode approach. The Keras ImageDataGenerator 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 the 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 512*512 for the picture size.

2.4 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 a couple of levels to correctly recognize images: firstly, it preprocesses for feature extraction from a 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[21]. 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 the total count of training samples. It may be calculated as Equation (1).

$$C = -\frac{1}{n}\sum \times \sum_{j} \left(y_j a_j^1\right)\left(1+y_j\right) l_n \left(1-a_j^1\right) \tag{1}$$

2.4.1 Convolutional Layer

The product of the interconnected neurons weights and the constrained region of input to which they are connected determine the 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 k^{th} 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.

$$Y^K = f\left(\sum\nolimits_k x^k * \omega^k + b^k\right) \tag{2}$$

2.4.2 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 the 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 [Equation (3)].

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

$$H_2 = \left(\frac{H_{1+}1}{S}\right) + 1 \tag{3}$$

2.4.3 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. The equation below, where l and (I - 1) are designated as completely linked layers, gives the fully connected operations. The weight connections between the j^{th} unit of layer (I-1) in (r, s) location and the i^{th} unit in layer l and y [Equation (4)].

$$Y_i^1 = f\left(Z_i^1\right).with\_Z_i^1 = \sum\nolimits_{j=1}^{m1^{I-1}} \sum\nolimits_{r=1}^{m2^{I-1}} \sum\nolimits_{s=1}^{m3^{I-1}} W_{i,j,r,s}\left(Y_i^{I-1}\right).r,s \tag{4}$$

2.4.4 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. Several popular activation functions are available, such as the Sigmoid, ReLU, Softmax, and tanH functions.

2.4.5 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 types of activation functions are beneficial. We can see the Relu function in Equation (5).

$$f(y) = max(0, x) \tag{5}$$

2.4.6 Softmax Activation

The final activation function in a neural network 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 identified will produce an output value between 0 and 1 is calculated using Softmax, and the outcomes are then translated into numerical form [Equation (6)].

$$F(z)i^k = \frac{e^{Z_j}}{\sum_{k=1}^{k} e^{K_k}} \tag{6}$$

2.5 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 19 layers, including 16 convolution layers, 5 MaxPool layers, 1 SoftMax layer, 3 completely connected layers, and so forth. 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[(22)].

2.6 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.

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.
- 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.

2.6.1 Capsule Network Architecture

The capsule network's design (Figure 4) is a shallow network with the three components as a 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.

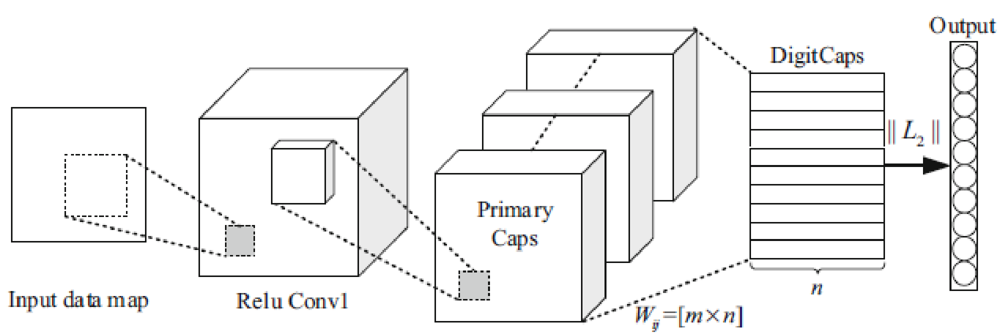

Fig 4. Capsule Network Architecture

2.7 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 (Equation (7)).

True-positive (TP) results, as seen from a medical standpoint, signify that individuals with abnormal lesions are examined following a predefined assessment 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 are 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 when a patient has an abnormal lesion and is mistakenly classified as a normal sample, misleading for diagnosis.

Sensitivity/recall or TP rate is the probability that patients with abnormal lesions are identified as medically positive. It must be highly valued practically for sensitive identification (Equation (8)).

Specificity rate is the probability that an individual will receive a negative diagnosis 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 (Equation (9)).

According to Aurelien Geron in Hands-on Machine Learning, the F1 Score relates to the "harmonic mean of precision and sensitivity." (Equation (10))

$$accuracy = \frac{TN + TP}{TN + FP + FN + TP} \tag{7}$$

$$sensitivity = \frac{TP}{FN + TP} \tag{8}$$

$$specificity = \frac{TN}{FP + TN} \tag{9}$$

$$F1score = \frac{TP}{\frac{1}{2}(FN + FP) + TP} \tag{10}$$

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

2.8 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.

2.8.1 Mobilenet

MobileNets generally use depthwise separable convolutions rather than the regular ones found in earlier structures to provide lighter models. The width multiplier and resolution multiplier, two new global hyper parameters introduced by MobileNets, Permit tradeoffs for model creators in accuracy or speed-related delay to achieve a 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[23].

2.8.2 Xception

"Extreme Inception" is what Architecture Xception stands for. The network for feature extraction is made up of 36 convolutional layers. The architecture of Xception is made up of several stacks 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.

2.8.3 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[23].

2.8.4 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 include 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.

2.8.5 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 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.

2.9 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 the ReLu activation function based on dense layer and global average pooling layers. 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 efficiently than the MobileNet, Xception, Inception V3, and VGG 16 and 19 models.

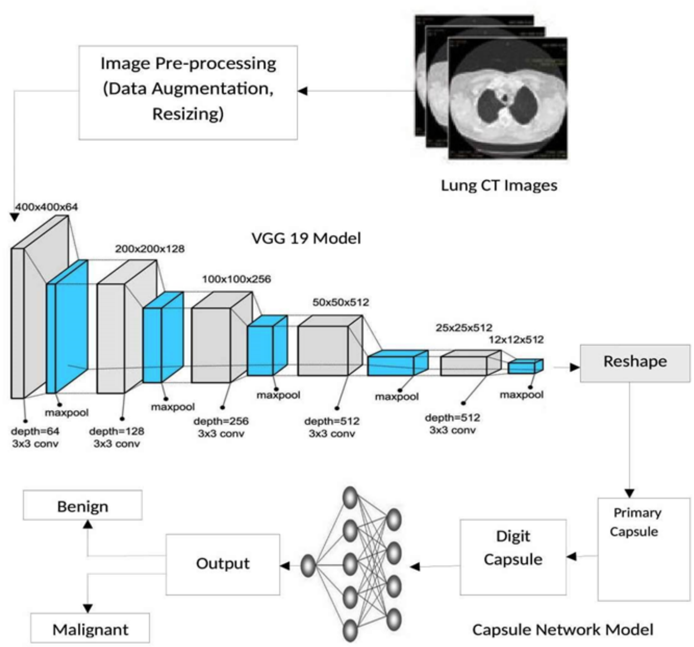

Fig 5. The architecture of the VggCapNet

The capsule network implementation uses the pooling layer based on the global median as its input [25]. The features are then extracted using a 2-dimensional convolution layer with a kernel size of 9, 9,512. Then, 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 5.

3 Results and Discussion

This paper presents a unique hybrid Deep learning model, named VggCapNet, to detect lung cancer from CT images. The suggested model for classifying lung nodule malignancy was effectively trained and tested. LIDC-IDRI dataset CT images are processed using the VggCapNet model, 70 % of the said dataset is used as a training set and the remaining 30 percent is used to evaluate the model. Among the various state of art models in use, the proposed VggCapNet architecture has the best accuracy at 99.20%. Models such as MobileNet [20], Xception [21], VGG-16 [10], VGG-19 [18] and InceptionV3 [11] had accuracy percentages of 98.00%, 97.97%, 96.95%, 95.60%, and 94.35%, respectively.

Table 2 displays the Performance comparison proposed VggCapNet model with state of art methods. It suggests that VggCapNet, the hybrid model, outperformed the other models and achieved the best accuracy of 99.20 percent in the classification of lung CT images as Benign or Malignant. It can also be observed the proposed model outperforms in terms of F measure, sensitivity, specificity, and AUC respectively.

Figure 6**a** displays a confusion matrix that represents the Predicted label and actual label for lung CT images. Figure 6**b, c ,** and **d** display the validation accuracy, loss curves, and ROC curve for the VggCapNet model respectively. 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.

Table 2. Performance comparison of proposed VggCapNet model with state of art methods

Model	Accuracy	F-Measure	Sensitivity	Specificity	AUC
MobileNet [20]	98.00 $\pm$ 0.40	98.00 $\pm$ 0.55	98.15 $\pm$ 0.65	98.00 $\pm$ 0.40	98.00 $\pm$ 0.35
Xception [21]	97.97 $\pm$ 1.02	96.87 $\pm$ 1.82	97.47 $\pm$ 1.05	97.97 $\pm$ 1.02	96.97 $\pm$ 1.02
VGG-16 [10]	96.95 $\pm$ 1.75	96.95 $\pm$ 1.55	95.95 $\pm$ 1.75	96.95 $\pm$ 1.25	95.95 $\pm$ 1.84
VGG-19 [18]	95.60 $\pm$1.00	95.55 $\pm$ 1.05	95.10 $\pm$ 1.00	95.60 $\pm$ 0.75	95.50 $\pm$ 1.10
Inception V3 [11]	94.35 $\pm$ 1.1	94.35 $\pm$ 1.15	94.35 $\pm$ 1.2	94.35 $\pm$ 1.10	94.35 $\pm$ 1.1
Proposed VggCapNet	**99.20** $\pm$ **0.2**	**99.79** $\pm$ **0.5**	**99.50** $\pm$ **0.5**	**99.20** $\pm$ **0.6**	**99.15** $\pm$ **0.6**

As many deep learning models can assist in lung cancer diagnosis, accuracy in publically accessible datasets continues to be a significant challenge. The researcher's findings were correct overall, even though some of them used tiny datasets that were freely accessible online or private hospital datasets [8,10–12]. Despite recent advances, the clinical implementation of DL models for malignancy detection is still challenging. Almost more than 50% of recent research utilizes the LIDC-IDRI lung CT scan dataset, thus we utilized the same for our suggested model. After pre-processing the LIDC-IDRI dataset, there were no problems while using the models.

Several DL CNN models are utilized to detect the lung's nodule in most situations but relative spatial correlations between features in scanned images are not stored by CNN. CNN's flaws are entirely accounted for by Capsule Networks. To compensate for the information loss caused by aggregations, capsule networks, or CapsNets, can obtain spatial information. This paper presents the VggCapNet architecture for the detection and classification of lung nodules as benign or malignant based on lung CT images in comparison with standard CNN MobileNet, Xception, VGG-16 VGG-19, Inception V3 models for lung cancer classification using CT images. 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.

We can infer from the graphs that there hasn't been any overfitting or underfitting. Accuracy and losses in training and validation continue to converge, peaking after 50 epochs. The confusion matrix proves a strong relationship between the expected and real numbers. The confusion matrix for lung nodules also shows that the predicted values of Specificity and Sensitivity are close to 1, suggesting that the model has a high true positive and true negative rate. The low rates of false positives and negatives suggest that there is very little chance of data misclassification using the proposed VggCapNet model.

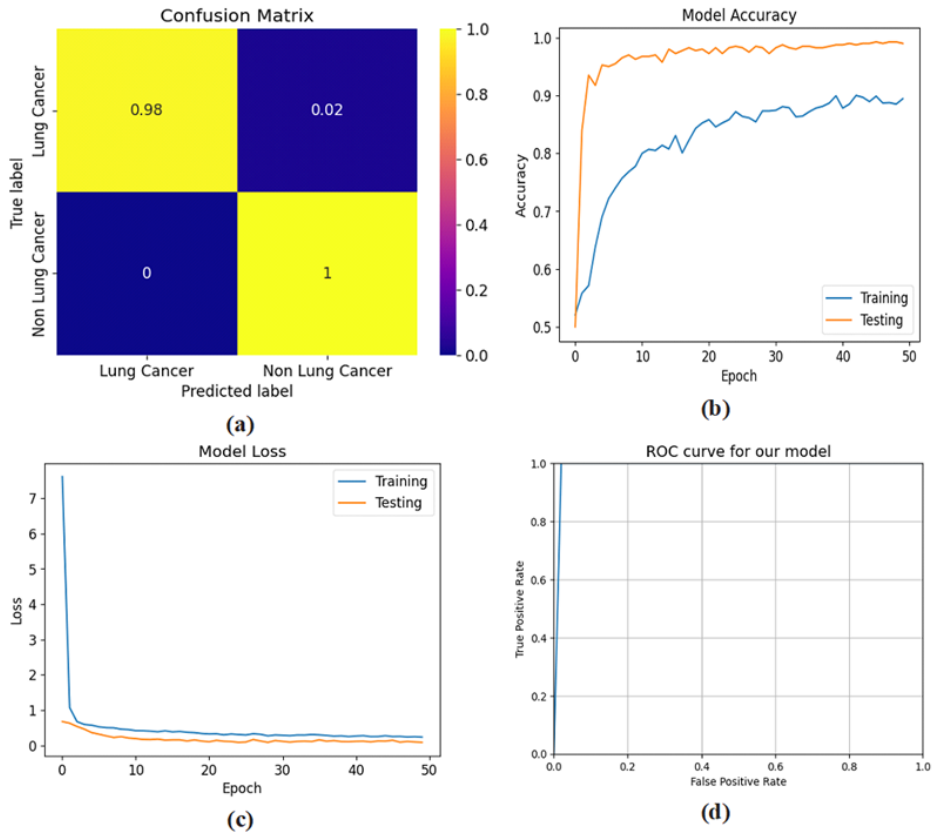

Fig 6. (a) Confusion Matrix, (b) Model Accuracy, (c) Model loss, (d) ROC Curve

4 Conclusion

Relative spatial correlations between features in scanned images are not stored by CNN. Considering the high spatial resolution of CT scan images and their sensitivity to misalignments during the scanning process, a method that can assist in incorporating spatial information of image elements into consideration is necessary. The most effective method for identifying real-world images is the capsule network, which relies on the spatial relationship of the image's characteristics. The shortcomings of convolutional neural networks are addressed with Capsule Network. 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. A suggested hybrid architecture called VggCapNet combines the capabilities of the VGG-19 and capsule network architectures to address orientation-related challenges often encountered in traditional CNN-based approaches. Although the existing approaches of MobileNet, Xception, VGG-16, 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.

Future Scope

In the future, more cancers such as breast, liver, cervical, brain, and skin may be detected using this proposed VggCapNet hybrid model. Different CNN optimization strategies can also be used to optimize the model's performance. The system performance might be enhanced by optimizing the Capsule Network hyperparameter.

Limitation

Compared to CNNs, capsule networks certainly have more complicated deployments. The dynamic routing algorithm's inner loop leads to the latency algorithm's speed.

References

1) Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics 2021. *CA: A Cancer Journal for Clinician*. 2021;71(1):7–33. Available from: https://doi.org/10.3322/caac.21654.
2) Agrawal S, Jain SK. Medical text and image processing: applications, issues and challenges. In: Machine Learning with Health Care Perspective;vol. 13 of Learning and Analytics in Intelligent Systems. Cham. Springer. 2020;p. 237–262. Available from: https://doi.org/10.1007/978-3-030-40850-3_11.
3) Abdullah DM, Ahmed NS. A Review of most Recent Lung Cancer Detection Techniques using Machine Learning". *International Journal of Science and Business*. 2021;5(3):159–173. Available from: https://ideas.repec.org/a/aif/journl/v5y2021i3p159-173.html.
4) Hussein S, Kandel P, Bolan CW, Wallace MB, Bagci U. Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches. *IEEE Transactions on Medical Imaging*. 2019;38(8):1777–1787. Available from: https://doi.org/10.1109/TMI.2019.2894349.
5) Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017;p. 3859 –3869. Available from: https://dl.acm.org/doi/10.5555/3294996.3295142.
6) III SGA, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. Data From LIDC-IDRI. 2020. Available from: https://www.cancerimagingarchive.net/collection/lidc-idri/.
7) Kumar G, Chatterjee S, Chattopadhyay C. DRISTI: a hybrid deep neural network for diabetic retinopathy Diagnosis. *Signal, Image, and Video Processing*. 2021;15:1679–1686. Available from: https://doi.org/10.1007/s11760-021-01904-7.
8) Pang S, Zhang Y, Ding M, Wang X, Xie X. A Deep Model for Lung Cancer Type Identification by Densely Connected Convolutional Networks and Adaptive Boosting. *IEEE Access*. 2019;8:4799–4805. Available from: https://doi.org/10.1109/ACCESS.2019.2962862.
9) Ali I, Muzammil M, Haq IU, Amir M, Abdullah S. Efficient Lung Nodule Classification Using Transferable Texture Convolutional Neural Network. *IEEE Access*. 2020;8:175859 –175870. Available from: https://doi.org/10.1109/ACCESS.2020.3026080.
10) Pang S, Meng F, Wang X, Wang J, Song T, Wang X, et al. 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*. 2020;13(1):771–780. Available from: https://doi.org/10.2991/ijcis.d.200608.001.
11) Wang C, Chen D, Hao L, Liu X, Zeng Y, Chen J, et al. Pulmonary Image Classification based on Inception-v3 Transfer Learning Model. *IEEE Access*. 2019;7:146533 –146541. Available from: https://doi.org/10.1109/ACCESS.2019.2946000.
12) Yu H, Zhou Z, Wang Q. Deep Learning Assisted Predict of Lung Cancer on Computed Tomography Images Using the Adaptive Hierarchical Heuristic Mathematical Model. *IEEE Access*. 2020;8:86400–86410. Available from: https://doi.org/10.1109/ACCESS.2020.2992645.
13) Saleh AY, Chin CK, Vanessa Penshie, Al-Absi HRH. Lung cancer medical images classification using hybrid CNN-SVM. *International Journal of Advances in Intelligent Informatics*. 2021;7(2):151–162. Available from: https://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*. 2019;85:109–119. Available from: https://doi.org/10.1016/j.patcog.2018.07.031.
15) Joshua ESN, Bhattacharyya D, Chakkravarthy M, Byun YC. 3D CNN with visual insights for early detection of lung cancer using gradient-weighted class activation". *Journal of Healthcare Engineering*. 2021;2021:1–11. Available from: https://doi.org/10.1155/2021/6695518.
16) Afshar P, Oikonomou A, Naderkhani F, Tyrrell PN, Plataniotis KN, Farahani K, et al. 3D-MCN: A 3D Multi-scale Capsule Network for Lung Nodule Malignancy Prediction. *Scientific Reports*. 2020;10:1–11. Available from: https://doi.org/10.1038/s41598-020-64824-5.
17) Tao Z, Bingqiang H, Huiling L, Zaoli Y, Hongbin S. NSCR-Based DenseNet for Lung Tumor Recognition Using Chest CT Image. *BioMed Research International*. 2020;2020:1–9. Available from: https://doi.org/10.1155/2020/6636321.
18) Khan MA, Rajinikanth V, Satapathy SC, Taniar D, Mohanty JR, Tariq U, et al. VGG19 Network Assisted Joint Segmentation and Classification of Lung Nodules in CT Images. *Diagnostics*. 2021;11(12):1–16. Available from: https://doi.org/10.3390/diagnostics11122208.
19) Tandon R, Agrawal S, Chang A, Band SS. VCNet: Hybrid Deep Learning Model for Detection and Classification of Lung Carcinoma Using Chest Radiographs. *Frontiers in Public Health*. 2022;10:1–13. Available from: https://doi.org/10.3389/fpubh.2022.894920.
20) Crasta LJ, Neema R, Pais AR. A novel Deep Learning architecture for lung cancer detection and diagnosis from Computed Tomography image analysis. *Healthcare Analytics*. 2024;5:1–17. Available from: https://doi.org/10.1016/j.health.2024.100316.
21) Vani Rajasekar, Vaishnnave MP, Premkumar S, Sarveshwaran V, Rangaraaj V. Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques. *Results in Engineering*. 2023;18:1–9. Available from: https://dx.doi.org/10.1016/j.rineng.2023.101111.
22) Al-Yasriy HF, AL-Husieny MS, Mohsen FY, Khalil EA, Hassan ZS. Diagnosis of Lung Cancer Based on CT Scans Using CNN. In: 2nd International Scientific Conference of Al-Ayen University (ISCAU-2020) ;vol. 928 of IOP Conference Series: Materials Science and Engineering. IOP Publishing. 2020;p. 1–10. Available from: https://dx.doi.org/10.1088/1757-899x/928/2/022035.
23) Hareesh PM, Bellamkonda S. Deep Learning-Based Classification of Lung Cancer Lesions in CT Scans: Comparative Analysis of CNN, VGG-16, and MobileNet Models. In: Fourth International Conference on Image Processing and Capsule Networks ;vol. 798 of Lecture Notes in Networks and Systems. Singapore. Springer. 2023;p. 373–387. Available from: https://doi.org/10.1007/978-981-99-7093-3_25.
24) Bushara AR, Kumar RSV, Kumar SS. An ensemble method for the detection and classification of lung cancer using Computed Tomography images utilizing a capsule network with Visual Geometry Group. *Biomedical Signal Processing and Control*. 2023;85. Available from: https://doi.org/10.1016/j.bspc.2023.104930.
25) Mohite A. Application of Transfer Learning Technique for Detection and Classification of lung cancer using CT Images. *International Journal of Scientific Research and Management*. 2021;9(11):621–634. Available from: https://doi.org/10.18535/ijsrm/v9i11.ec02.