scientific reports

Check for updates

OPEN

An efficient hybrid artificial intelligence framework for lung cancer classification using CT images

L. Kamala✉ & K. G. Mohan

Lung cancer is the most dangerous type of cancer, and its affected rate increases gradually. The survival rate of lung cancer is very low compared to other cancers. Early prediction of lung cancer can help increase the survival rate by ensuring sufficient oxygen supply and eliminating carbon dioxide in the human body. Computed Tomography (CT) imaging is widely used by physicians for lung disease identification because it provides detailed cross-sectional views of the lungs and helps in detecting small nodules that may be missed by other imaging techniques. Manual visualization is time-consuming and sometimes leads to classification errors, making it difficult to identify lung cancer at an early stage. This limitation can be addressed through automated lung cancer prediction using Artificial Intelligence (AI). This research proposes a hybrid AI model to classify lung CT images as normal, benign, or malignant. Images from the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) dataset are used. These images undergo various pre-processing techniques, and features are extracted using both traditional methods such as Gray-Level Co-occurrence Matrix (GLCM) and Scale-Invariant Feature Transform (SIFT), and Deep Learning (DL) methods such as Visual Geometry Group (VGG-16) and MobileNet. The extracted features from both approaches are fused and processed through a Fully Connected Layer (FCL) for classification. Six combinations of feature extraction modules are evaluated for lung cancer prediction: GLCM + MobileNet, GLCM + VGG-16, SIFT + MobileNet, SIFT + VGG-16, GLCM + SIFT + MobileNet, and GLCM + SIFT + VGG-16. Among these combinations, the integration of GLCM and SIFT with MobileNet demonstrates superior performance compared to other AI combinations, achieving the highest accuracy, precision, recall, F1-score, and specificity. The proposed methodology is also compared with state-of-the-art methods, and the results indicate that the hybrid AI model surpasses existing approaches. The experimental findings confirm that the proposed model provides reliable results for lung cancer prediction from CT images and is suitable for real-time deployment.

Keywords Lung cancer, Computed tomography images, Visual geometry group-16, MobileNet, Augmentation, Feature extraction, Accuracy

Cancer has always been considered a potentially fatal disease. Lung cancer, the common cancer worldwide, accounts for the majority of cancer-related fatalities in both developed and developing countries, with a 5-year survival rate of only 18%[1]. Individuals with lung cancer are more likely to be in advanced stages of the disease, making it more difficult for medical science to improve survival rates than with other types of cancer. Cancer cells follow the natural lymphatic flow, migrating from the lungs to the lymph glands before entering the blood and moving toward the center of the chest. Quick identification is critical for preventing the spread of cancer to other organs. Radiation, chemotherapy, and monoclonal antibodies are common nonsurgical treatments for late-stage cancers[2]. Follow-up radiography is essential for assessing treatment response and monitoring changes in tumor progression over time. A pathology laboratory frequently employs various techniques to analyze cancer. Malignant tissue is examined using both microscopic and electronic modalities, including biopsies and CT scans. The most common and preferred diagnostic method in pathology tests is CT scanning. This method provides high-resolution and contrast images of the organ from multiple angles, enabling a 3D evaluation of the lesion[3].

Department of Computer Science and Engineering, GITAM School of Computer Science and Engineering, GITAM (Deemed to be university), Bengaluru 561203, India. ✉email: lllkamala5@gmail.com

The malignant cell's nucleus is visually conspicuous due to its unusual size and structure. The morphological elements of CT scans are employed to differentiate malignant and benign nodules. Patients have a better chance of survival if nodules are found early. There is considerable overlap in the features of nodules, and morphologic features must be carefully investigated for accurate classification. CT scans allow for the timely detection of concerning nodules. The body's internal structure scan images display aspects from the front, bottom, and top perspectives. Plain X-rays provide less information about soft tissues, blood vessels, and bones[4]. Chest CT scans, on the other hand, make it hard to diagnose initial-stage lung cancer since lung nodules resemble neighboring structures. As a result, there is a requirement for developing a computer-aided design (CAD) system for detecting malignant nodules[5]. DL is a promising strategy for distinguishing between malignant or benign nodules, thereby reducing the number of scans required to make a diagnosis. DL algorithms show excellent results for detecting benign and malignant nodules. The most recent CAD system uses DL algorithms to diagnose lung nodules and determine whether they are malignant or benign.

Many researchers have already worked on lung cancer prediction using CT images, but some issues remain unsolved. The accuracy of the models is not very high, and they do not extract all relevant features, leading to a drop in accuracy. Additionally, the identification of lung nodules is challenging at an early stage. Proper feature extraction can enhance the accuracy of lung cancer prediction. This research employs multiple feature extraction techniques, which are then fused to create a comprehensive feature set. The fused features are utilized for classification, and various combinations are systematically evaluated to determine the most effective feature extraction approach. This ensures that all critical features from the CT scan are captured, enhancing the accuracy of cancer prediction.

The research is outlined as follows: Section I details lung cancer, CT imaging, and the importance of CAD. Section II reviews some of the recent research on lung cancer prediction. Section III presents the theoretical background of the techniques used in this research for feature extraction. Section IV describes the experimental setup, data used, pre-processing steps, and experimental outcomes. Section V concludes the research with its limitations and future work.

Related research work

Medical imaging research is rapidly evolving to discover disease patterns, illustrating the advancement of computationally intelligent systems. Medical image research is among the most fascinating areas for establishing a comprehensive expert system that includes AI, computer vision, and pattern recognition applications. This section describes the associated research activities and findings.

The study[3] aimed to create a hybrid framework for diagnosing lung cancer using CT images by integrating a Discrete Local Binary Pattern (DLBP) with a Hybrid Wavelet Partial Hadamard Transform (WPHT). Adaptive median filtering is the first pre-processing step in the proposed approach for lung cancer CT images. Then, using DLBP and hybrid WPHT, features are extracted. Next, by reducing the dimensionality of features, the Harris-Hawk optimization technique is utilized to improve the feature selection. The classifier's parameters are adjusted using the modified weight-based beetle swarm approach and the SVM-based classification. The suggested approach classifies the input images as malignant, benign, and normal. The outcome says that the proposed approach is better at predicting cancer. The study[6] uses lung X-ray images to identify malignant development in its most basic form. Otsu's thresholding approach is used to segment the images. The number of black and white pixels, as well as the histogram properties of the segmented image, make up the feature vector. To determine whether the cancer is benign or malignant, the feature vector is subjected to the K-Nearest Neighbor and the Neural Network with Back Propagation. A Kaggle dataset served as the basis for the project. It has been demonstrated that the proposed strategy excels existing techniques in recognition accuracy. In the study[7], the Convolutional Neural Network (CNN) framework was developed for initial-stage lung cancer prediction. Various models, such as Xception, Inception, and ResNet, were also evaluated for comparison. The models were assessed based on favorable metrics, and CNN outperformed other models, proving to be more effective than previous approaches.

The article[8] explores the application of pre-trained CNNs for lung cancer diagnosis. It also investigates the efficiency of CNNs in analyzing cancer images. The pre-trained networks were modified with a few additional layers and adjustments to hyperparameters. The pre-trained transfer learning networks excel the state-of-the-art techniques in detecting lung cancer, achieving accuracy rates between 78 and 86%. The pre-trained VGG-16 model achieved the highest detection accuracy. The work[9] proposed an autonomous nodule recognition strategy in CT images based on a modified AlexNet and the LungNet-SVM model. The proposed network has 7 convolutional, 3 pooling, and 2 FCLs for feature extraction. The SVM classifier is used to categorize nodules as benign or malignant. The experimental investigation utilizes the Lung Nodule Analysis 2016 (LUNA16) dataset. A comparison was made between the proposed and existing cutting-edge approaches for lung cancer detection. The comparison outcomes demonstrate that the proposed network performed exceptionally well on the LUNA16 dataset. An ensemble technique was developed in the study[10] for lung nodule prediction. Rather than using a single DL network, the performance of two or more CNNs was combined to improve prediction accuracy. The LUNA16 dataset was utilized, comprising CT scans with annotations to aid in data interpretation. CNNs were trained on the dataset to detect malignant and benign images. This model contains three unique CNNs with distinct kernels, layers, and pooling algorithms. The proposed CNN produced excellent results, outperforming the baseline technique.

The paper[11] proposes a novel CNN for automatic lung cancer detection. To achieve optimal CNN design, the Snake Optimizer was developed to fine-tune the CNN architecture. The created network was then tested using the IQ-OTH/NCCD dataset. This approach was also validated by comparing it to several other published methods in the literature. The study[12] aims to process CT images for lung cancer detection. In the first stage, the inputs are processed with CNN and classified using an Artificial Neural Network (ANN). The second stage

involves pre-processing and segmenting the images before applying CNN and ANN. The third stage uses specific feature extraction methods, where all processed images are converted to numerical format. Additionally, feature selection and dimensionality reduction techniques are used for classification with three techniques: Random Forest (RF), Gradient Boosting, and SVM. The evaluation was conducted to identify the best approach. The results indicate that using either the SVM or RF algorithms resulted in excellent performance in classifying lung cancer.

The recent literature on lung cancer detection highlights significant advancements in deep learning–based imaging analysis[13]. developed a hybrid diagnostic model that integrates handcrafted and deep-learning features to improve CT-based lung cancer detection and stage classification. Article[14] provided an extensive review of deep learning techniques including CNNs, transfer learning, and hybrid models emphasizing challenges such as limited datasets and the need for better model generalization. The authors in[15] proposed Mask-EffNet, a fusion of Mask R-CNN and EfficientNet, achieving improved lesion segmentation and malignancy prediction from CT scans. In paper[16]applied transfer learning to classify lung anomalies, demonstrating enhanced accuracy and computational efficiency for practical medical screening. Complementing these works[17], integrated a deep-learning framework with the Ebola Optimization Search Algorithm to refine model parameters and boost diagnostic reliability in CT-based lung cancer classification. Together, these studies demonstrate the growing effectiveness of hybrid, transfer-learning, and optimizationaugmented deep learning models in advancing automated lung cancer detection and clinical decision support.

Proposed methodology

The proposed research primarily focuses on feature extraction from lung CT images, emphasizing that effective feature extraction enhances accuracy in disease prediction. Four techniques are utilized for feature extraction: GLCM, SIFT, VGG-16, and MobileNet. The working principles of each technique are detailed in this section.

GLCM

The GLCM technique is a texture feature extracted method that shows spatial correlations between pixel values[18]. A co-occurrence matrix, known as GLCM, indicates which combinations of various pixel contrast values appear and how frequently they occur in an image's grayscale representation. The texture created from the GLCM depicts the relationship between reference and nearby pixel pairs in a given scenario. When employing GLCM, each pixel in an image function as both a reference and a neighboring pixel. Instead of precisely representing the actual image and its pixel score, the entries in the GLCM display the frequencies of adjacent pairs of pixels[19]. The function of figuring out the probabilistic convergence score of the frequency of occurrence of specific pixel pairings throughout the whole image is known as normalization in GLCM metrics. The normalized GLCM formula is given in Eq. (1).

$$P_{(i',j)} = \frac{V_{(i',j)}}{\sum_{i,j=0}^{N-1} V_{i,j}} \tag{1}$$

where N represents the total rows and columns, V represents the value of pixel pairs i and j, and $P_{(i',j)}$ denotes the probabilistic convergence score of pixel pairs i and j. The co-occurrence matrix in GLCM texture analysis should be symmetric around the diagonal axis. When the cells on either side of the diagonal have identical values, the matrix is said to be symmetric. Standardized values for various texture features are designed to manage the normalized symmetrical GLCM from many perspectives. The integration of multiple features results in a complex pattern that can be utilized to make broad generalizations about both healthy and diseased samples. Each texture feature describes a specific aspect of lung texture. Furthermore, the weighted averages of normalized GLCM computations, derived from surrounding pixel pairs, make up the majority of texture features. In this study, GLCM analysis was utilized to detect nine unique texture features in each image. Table 1 lists the formulas used to calculate the features, which are energy, Angular Second Moment (ASM), correlation, homogeneity, variance, contrast, difference variance, difference entropy, and joint entropy[20]. In this study, 1 is set as the default value for the pixel-to-pixel spacing during the GLCM computation step. Additionally, only the horizontal direction (0°) was considered while performing the GLCM calculations.

SIFT

SIFT is a region descriptor that extracts features from images that are largely invariant to camera viewpoint, lighting effect, image rotation, and scaling[21]. Each feature has a descriptor, which is a vector of dimension 128. This vector allows features to be compared between image pairings that undergo the geometrical changes mentioned above. The theoretical and experimental findings[22] justify the use of a tiered filtering technique for feature detection. It is assumed that a Gaussian function provides the scale-invariant image space that exists. The Gaussian kernel's convolution $G(i, j, \sigma)$ with the input CT scan $I(i, j)$ results in a function $L(i, j, \sigma)$, which defines the scale-space is given in Eq. (2):

$$L(i, j, \sigma) = G(i, j, \sigma) * I(i, j) \tag{2}$$

Lowe introduced a novel technique for discovering scale-invariant key points. SIFT uses the Difference-of-Gaussian (DoG) process and it is derived from the difference of two neighboring scalessplitted by a fixed multiplication factor (k), to choose key points as the optimal response for the Lindeberg algorithm.

$$D(i, j, \sigma) = (G(i, j, k\sigma) - G(i, j, \sigma) * I(i, j)) \tag{3}$$

Texture features	Formula
Energy	$\sum_{i,j}^{K-1} P(i,j)^2$
ASM	$\mathrm{ASM} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\{P(i,j)\}^2$
Correlation	$\mathrm{Correlation} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{\{i\times j\}\times P(i,j)-\{\mu_x\times\mu_y\}}{\sigma_x\times\sigma_y}$
Contrast	$\mathrm{Contrast} = \sum_{i=1}^{G}\sum_{j=1}^{G}\lvert i-j\rvert^2 P(i,j)$
Homogeneity	$\sum_{(i,j=0)}^{K-1}\frac{P(i,j)}{1+\lvert i-j\rvert^2}$
Variance	$\mathrm{Variance} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-\mu)^2 P(i,j)$
Difference variance	$Varianceof P_{(x-y)}$
Difference entropy	$\sum_{i=0}^{K-1} P_{(x-y)}(i)\, log\left\{P_{(x-y)}(i)\right\}$
Joint entropy	$-\sum_{i,j} P(i,j)\, log\,(P(i,j))$

Table 1. Texture features and its formula.

$$= L(i,j,k\sigma) - L(i,j,\sigma) \tag{4}$$

The merit of Lowe's detector over the Lindeberg technique is minimal computational cost. This is because a difference in Gaussian functions, which can be determined using simple image subtraction, is employed instead of Gauss's scale-normalized Laplacian. The DoG identifies feature points as the local minima and maxima of $D(i,j,\sigma)$. All the image points are analyzed against the nine neighbors above and below it, as well as the eight nearby pixels. Key points are defined as local peaks or minima. To choose stable key points, the local extrema score of $D(i,j,\sigma)$ should exceed a threshold (Th_key). The research[23] developed a subpixel localization approach to improve local detection. To find local maxima, a Taylor expansion is used to fit the quadratic function $D(i,j,\sigma)$ to the sample points. The removal of low-contrast key points increases the stability of local detection. The DoG key points are invariant to transformation and dependent on rotation. To overcome these difficulties in feature-mapping approaches, the SIFT technique uses the radiometric characteristics of surrounding pixels to compute a canonical orientation at each key point.

The descriptor is an n-dimensional vector that represents the gradient orientation and magnitude pattern near the key point location. Lowe's approach generates an orientation histogram for each of the $4\mathrm{x}\ 4$ sub-regions, also known as bins, that comprise the region. Each histogram contains eight orientation bins that collect gradient magnitude data relative to the canonical orientation. Each pixel in the region has its magnitude value allocated to the appropriate sub-region after being weighted by a Gaussian function. Furthermore, the orientation histogram accumulates the weighted magnitude value. The Gaussian is used to prevent significant shifts in the descriptor caused by slight modifications in the window's position and to reduce the emphasis on gradients located far from the descriptor's center. Lastly, to make the descriptor insensitive to contrast fluctuations, the contrast variations are normalized to unit length. The shortest distance between the n-dimensional vectors is calculated to determine the relationship between two candidate locations. It makes use of Euclidean distance.

VGG-16

The network structure of VGGNet[24], which finished second in the 2014 ImageNet image categorization competition, is shown in Fig. 1. VGG16 is one of VGGNet's most powerful categorization networks. The VGG16 network consists of six blocks: five convolutional layers and one FCL. Five of these blocks contain thirteen convolutional layers, and the final three FCL are referred to as a single block. As a result, the VGG16 network has 16 layers, calculated as 13 + 3. Image features from the low, mid, and high feature layers are retrieved using a 5-block convolution[25]. Each section contains two or three convolutional layers. To make the network more nonlinear, accelerate network training, reduce the gradient vanishing, and overfitting, the final convolution in each block is followed by the ReLU function. Each convolutional layer uses a 3×3 kernel to capture fine details, generate stronger nonlinear effects, and use fewer parameters. Stacking 3×3 kernels to replace bigger kernels (5×5 and 7×7), simplifying network construction.

The last 5-block convolution is integrated into the top pooling layer, which uses 2×2 pooling kernels. This helps to minimize the predicted mean deviation produced by inaccuracies in the convolutional parameters.

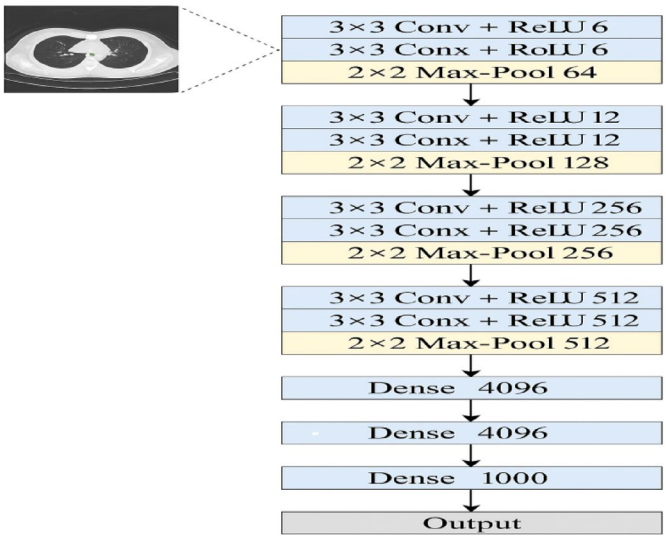

Fig. 1. VGG-16 architecture.

Changes in images and gradients are easier to capture, which is useful for preserving tiny features like textures. The final block of the network is made up of three fully connected networks. The previous layer's output features are combined to form a classifier. Simply, the network has 16 layers of depth. Layer-by-layer abstraction allows this model to progressively learn the features of each layer, from local to global. It can express more features, fit more complex input features, and has a higher nonlinear expression capability. The network initially used 64-3 $\times$ 3 convolute kernels. As the VGG-16 grows deeper, the number of convolute kernels increases from 64 to 128, 256, and 512, increasing its capacity. As the network's width expands, each layer can learn more detailed observational features.

MobileNet

The first mobile DL model based on TensorFlow is known as MobileNet[26]. The name "Mobile" implies that the network may be appropriate for mobile applications due to its simplified architecture and lower processing cost compared to previous transfer learning models. By comparing with the networks that employ traditional convolutions with equal depth, MobileNet's approach considerably reduces computational costs by reducing the number of parameters. As a result, MobileNet is a lightweight DL network. Depth-wise convolution (DWC) and point-wise convolution (PWC) are two fundamental methods combined to create the depth-wise separable convolution method[27]. DWC is based on the premise that the filter's depth and spatial components can be separated. The filter's width and height dimensions are divided, and then the depth is isolated. The PWC changes the previous layer's dimension and is a 1×1 convolution. This network acts as the backbone for extracting features from $224 \times 224 \times 3$ CT images.

To reduce the 2D channel to a single value, the global pooling layer averages all the data from each channel. This layer helps reduce the model's computational expenses and parameter count. This layer, implemented after the MobileNet model, averages the values of each (7×7) channel to modify the retrieved features from $(7 \times 7 \times 1024)$ to (1×1024). The Dense layer is the main component that represents a fully connected layer (FCL), making it an important and frequently used layer in DL networks. It extracts intricate patterns and representations from input data by first applying a transformation followed by an activation function. Following the global average layer, this layer type is used several times until the final classification output is reached. During training, a layer called batch normalization normalizes its inputs by calculating the standard deviation and mean of the current batch. Furthermore, it helps to reduce the network's dependence on learning rates and initialization, accelerates training, and prevents overfitting. To take advantage of its benefits, the batch layer is included next to the dense layers. To eliminate overfitting, a regularization layer known as dropout randomly assigns a fraction of the input to 0. The rate is a fraction used to accomplish this adjustment. The dropout layer enhances the rate, or the units' proportion to drop, by multiplying the remaining units by $1/(1 - rate)$.

Figure 2 depicts the architecture of MobileNet. RGB images $(224 \times 224 \times 3)$ are utilized as input, with features extracted as $7 \times 7 \times 1024$ output. The average pooling function averages each extracted feature channel, reducing the collected features to a size of (1×1024). For feature extraction from MobileNet, the output from the dense layer is taken and considered as MobileNet features.

Code availability

The code used to support the findings of this study is openly available to anyone at the following link: [https://drive.google.com/drive/folders/1l4aq1_AH5gdfbFznEWRL226tS1IR8zJT?usp=sharing]. The repository contains all scripts and documentation necessary to reproduce the analyses and results presented in this paper. No specific software license has been applied to this code.

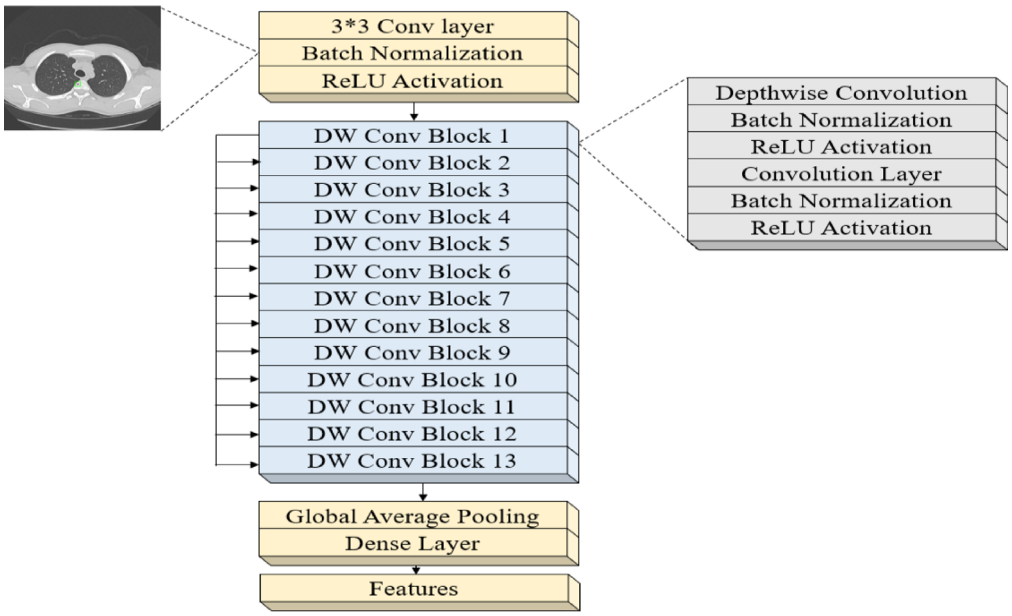

Fig. 2. MobileNet architecture.

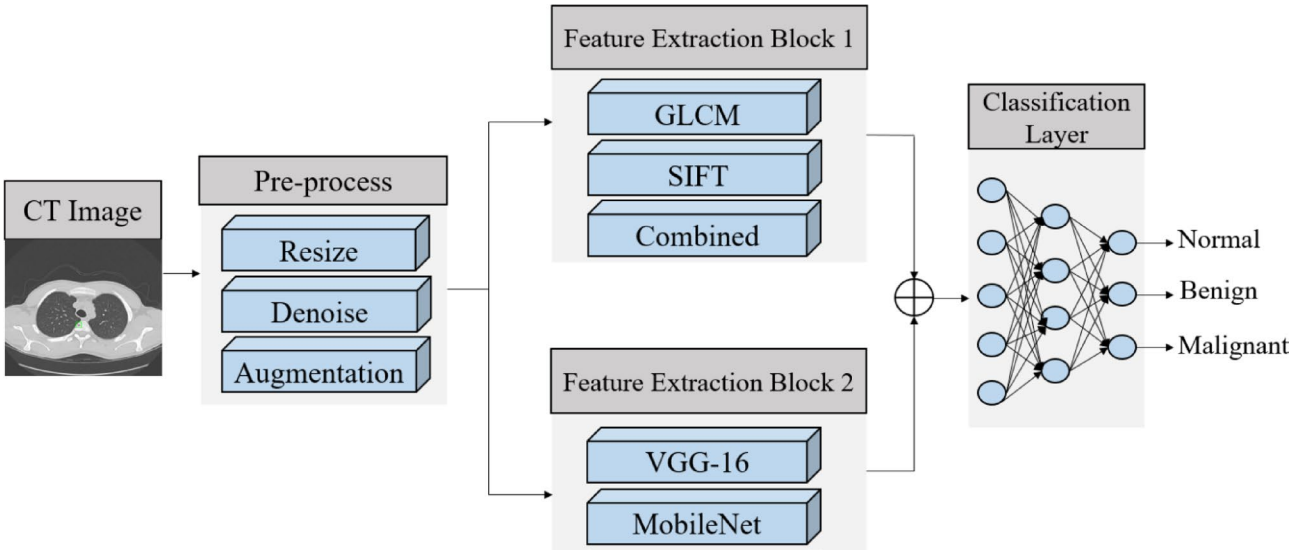

Fig. 3. Proposed Methodology for lung cancer prediction from CT images.

Result and discussion

Experimental setup

For lung cancer prediction using CT images, the experiment was performed on Jupyter Notebook. The programming language used is Python. Six different combinations of feature extraction techniques are evaluated under the same criteria: all models are trained, validated, and tested with the same set of images, using a cross-entropy loss function, the Stochastic Gradient Descent optimizer, 100 epochs, and a batch dimension of 32. The implementation was run on an Intel(R) Core(TM) i5-9500T CPU with 8 GB of RAM, using Windows OS. The proposed methodology workflow for lung cancer prediction from CT images is illustrated in Fig. 3. The methodology begins with collecting image data, followed by processing using various techniques such as resizing, denoising, and augmentation. The processed images are then passed to feature extraction models. Feature extraction is performed using two approaches: traditional methods and DL models. Extracted features from both approaches are fused and fed into an FCL for categorizing CT images into normal, benign, and malignant. Since the extracted feature dimensions are large, the dense layer in the FCL down samples the features to 32, and the output layer is further down sampled to 3. Various combinations of traditional and DL models are evaluated, and the best-performing combination is identified based on performance metrics.

All experiments use the stated split (train:val:test = 7:2:1) with augmentation applied only to the training set (zoom = 0.3, shear = 0.3, vertical_flip = True). Training employs categorical cross-entropy with SGD (learning rate 1×10^{-3}, momentum 0.9, weight decay 1×10^{-4}), 100 epochs, batch size 32, and early stopping (patience = 10; monitor = val-loss). We fix random_seed = 42 for NumPy and framework backends.

Feature normalization and fusion.

(1) Compute GLCM features using distances = {1} and angles = {0°, 45°, 90°, 135°}; aggregate by averaging across angles to form a 20-D texture vector per image (contrast, correlation, energy, homogeneity, entropy, etc.).
(2) Extract SIFT key points; form a bag-of-visual-words (k = 256) histogram with L1 normalization. (3) From MobileNet, take the penultimate layer activations (global average pooled) as a deep feature vector.
(3) Standardize each block independently on training statistics (μ, σ) → z = (x − μ)/σ, then concatenate in the order [GLCM, SIFT, MobileNet].
(4) Apply PCA to 256 dims (fit on training only) to reduce redundancy and stabilize optimization before classification. The classifier head is Dense(32) → ReLU → Dropout(0.3) → Dense(3) → Softmax. Inference uses the same preprocessing; no test-time augmentation is applied.

Data acquisition and processing

This study employs openly accessible IQ-OTH/NCCD dataset[7]. The Iraq-Oncology Teaching Hospital and the National Center for Cancer Diseases acquired 1,097 CT scans from 110 individuals in 2019. The data comprises CT scans of malignant, benign, and healthy individuals with a range of demographic features. Oncologists and radiologists helped categorize the images as benign, malignant, or normal. Following the initial recovery of the images in DICOM format, JPG files of 512×512 pixels were produced. The distribution of images in the dataset by class is shown in Table 2.

The images come in various sizes. The resize technique is used to make all images the same size. The resized images are all 224×224. Let $I_{original}$ be the original image, which is represented as I_{resize} in Eq. (5) after being scaled to $224 \times 224 \times 3$.

$$I_{resize} = Resize\left(I_{original}\right) \tag{5}$$

Several factors can contribute to noise in CT images, including patient size, slice thickness, electronic noise, and radiation exposure. Image denoising is the process of applying one or more filters to an image to reduce noise. The Gaussian filter is used to minimize probable lung CT noise and prevent image distortion. It convolves the CT image using a Gaussian kernel. The Gaussian filter's bell-shaped response gradually reduces the weight of pixels farther away while increasing the weight of pixels closer to the kernel's center. The blurring effect created by this weighting scheme reduces high-frequency noise while maintaining the image's basic structure. The formula for the Gaussian filter is shown in Eq. (6).

$$G\left(x, y\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{x^2}{2\sigma^2}} \tag{6}$$

Assuming the mean is zero at $x = 0$, σ represents the distribution's standard deviation. The augmentation phase increases the size of the dataset images. When employing the DL structure, the original image is altered in a unique way to make it larger. For the built model to be reliable, it must also include variations in the images. The augmentation technique was employed in this research, and the training data had the following values: zoom = 0.3, shear = 0.3, and vertical_flip = True. Table 2 provides a detailed count of actual images, augmented images, and the final images obtained after augmentation.

Experimental outcome

After pre-processing, including resizing, denoising, and augmentation, the images are split in a ratio of 7 (Train):2 (Validate): 1 (Test). Feature extraction is performed using two approaches: one with traditional imaging techniques such as GLCM and SIFT and the other with DL models such as VGG16 and MobileNet. Among several available CNN architectures, VGG-16 and MobileNet were specifically chosen based on a balance between computational efficiency, interpretability, and feature diversity. VGG-16 provides deep hierarchical representations with straightforward layer connectivity, making it suitable for extracting stable and interpretable visual features from limited medical datasets. In contrast, MobileNet employs depthwise separable convolutions that significantly reduce the number of parameters and memory usage, allowing fast inference on CPU hardware without sacrificing accuracy. Architectures such as ResNet, DenseNet, and EfficientNet, although powerful, introduce higher parameter counts and training complexity that are less suitable for moderate-sized datasets like

Class	Number of images	Augmented images	Total images
Malignant	561	439	1000
Benign	120	880	1000
Normal	416	584	1000
Total	1197	1803	3000

Table 2. Detail information on data distribution.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)
97	95.15	98.00	96.55	97.50
	98.97	96.00	97.46	99.50
	97.00	97.00	97.00	98.50
Average	97.04	97.00	97.00	98.50

Table 3. Performance metrics of GLCM and MobileNet features.

Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)
96.33	97.00	97.00	97.00	98.50
	96.91	95.92	96.41	98.51
	95.15	96.08	95.61	97.47
Average	96.35	96.33	96.34	98.16

Table 4. Performance metrics of GLCM and VGG16 features.

Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)
96.67	97.00	96.04	96.52	98.49
	95.15	95.15	95.15	97.46
	97.94	98.96	98.45	99.02
Average	96.69	96.71	96.70	98.32

Table 5. Performance metrics of SIFT and MobileNet features.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)
94.33	93.20	96.00	94.58	96.50
	96.94	95.00	95.96	98.50
	92.93	92.00	92.46	96.50
Average	94.36	94.33	94.33	97.17

Table 6. Performance metrics of SIFT and VGG16 features.

IQ-OTH/NCCD. Thus, the combination of VGG-16 and MobileNet ensures a practical trade-off between model generalization, computational cost, and clinical interpretability.

A total of six different feature combinations are evaluated using 300 images. First, GLCM is combined with MobileNet and VGG16, Out of 300 CT images, GLCM with MobileNet correctly identifies 291 cases of lung health conditions, while 9 are misclassified. Similarly, GLCM with VGG16 predicts 289 cases correctly and misclassifies 11 images. The performance measures computed from the confusion matrix are presented in Tables 3 and 4. The accuracy for GLCM with MobileNet is 91%, with an average precision of 96.04%, recall of 90.00%, F1 of 96.00%, and specificity of 96.50%. GLCM with VGG16 achieves an accuracy of 96.66%, with a precision of 96.44%, recall of 96.33%, F1 of 96.34%, and specificity of 98.16%. A comparison of Tables 3 and 4 indicates that GLCM with MobileNet provides better performance in lung disease prediction.

Next, SIFT features are combined with MobileNet and VGG16, and their performance is evaluated using 300 CT images. The confusion matrix for SIFT with MobileNet and VGG16 is presented. SIFT with MobileNet correctly identifies 290 cases, misclassifying 10 images. SIFT with VGG16 correctly classifies 283 images while misclassifying 17. The performance metrics for both combinations are provided in Tables 5 and 6. The accuracy for SIFT with MobileNet is 96.67%, whereas SIFT with VGG16 achieves 94.33%. SIFT with MobileNet has a precision of 96.69%, recall of 96.71%, F1 of 96.70%, and specificity of 98.32%. SIFT with VGG16 achieves a precision, recall, F1 of 94.3%, and specificity of 97.17%.

Finally, a combination of GLCM and SIFT features is used along with MobileNet and VGG16 for further evaluation. The same set of 300 images is used for testing. The combined features with MobileNet correctly classify 298 cases with only 2 incorrect predictions, achieving an accuracy of 99.33%. The combination with VGG16 correctly identifies 294 cases while misclassifying 6 images, resulting in an accuracy of 98%. The detailed performance metrics for both feature combinations are presented in Tables 7 and 8.

From the analysis of all six feature combinations, it is observed that the models incorporating MobileNet outperform those using VGG16. The combination of GLCM and SIFT features with MobileNet provides the best

Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)
98.33	97.00	98	97.50	98.00
	98.02	98.00	9700	98.01
	97.00	99.00	98.50	100.00
Average	98.34	98.34	97.33	98.67

Table 7. Performance metrics of combined (GLCM + SIFT) and MobileNet features.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)
98	96.12	99.00	97.54	98.00
	100.00	97.98	98.98	100.00
	98.00	97.03	97.51	98.00
Average	98.04	98.00	98.01	98.00

Table 8. Performance metrics of combined (GLCM + SIFT) and VGG16 features.

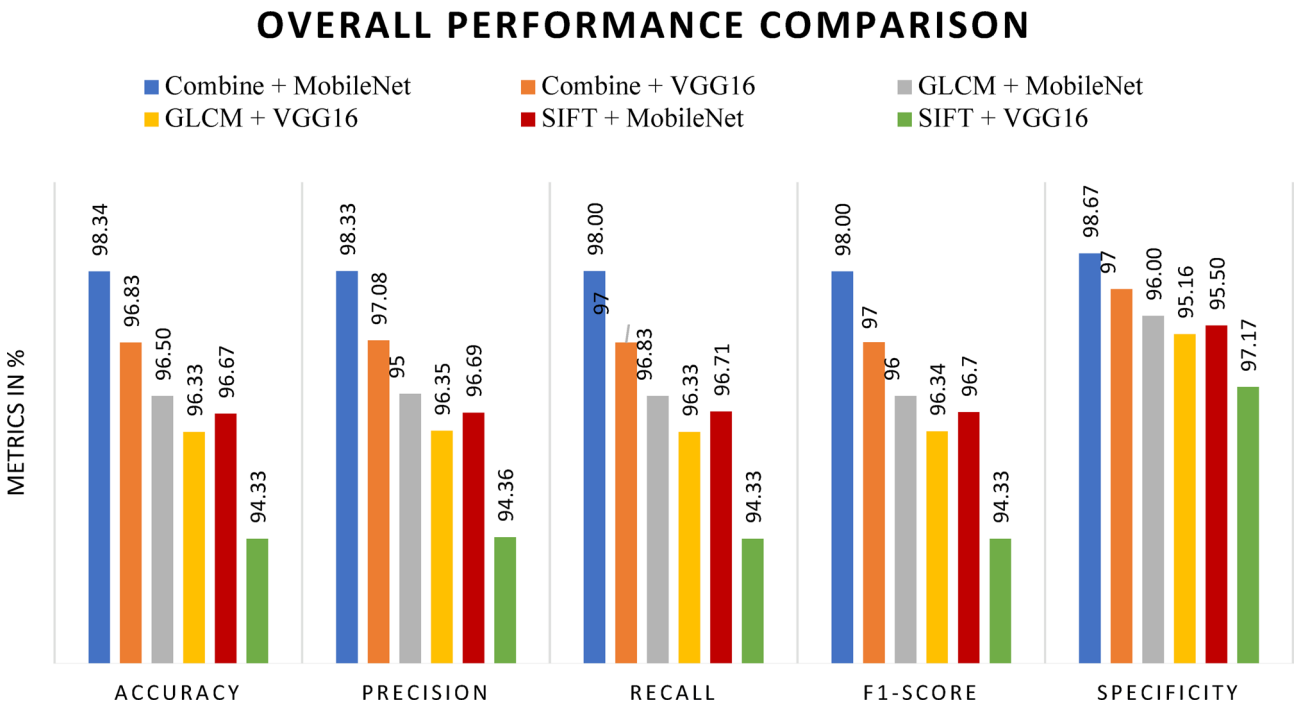

Fig. 4. Overall performance comparison of six feature combinations.

performance in predicting lung health conditions. Figure 4 presents an overall comparison of the performance of all six feature combinations.

The combined features with MobileNet are selected as the proposed network. Table 9 compares the proposed network with existing models that utilize the same IQ-OTH/NCCD dataset. All the studies classified lung CT images into normal, benign, and malignant. The highest accuracy achieved in research[29–31] was within the range of 98–99%. However, the proposed network gives an accuracy of 98.33%. Moreover, existing studies did not evaluate all performance metrics, whereas the proposed research examines the model comprehensively across multiple metrics. This comparison highlights the reliability of the proposed network in lung cancer prediction, demonstrating its superior performance in accurately classifying lung CT images.

To evaluate the practical feasibility of the hybrid framework, the computational complexity was analyzed in terms of runtime and memory utilization. The fusion of handcrafted (GLCM and SIFT) and deep features (MobileNet) naturally increases the dimensionality of the input to the fully connected layer, which could influence execution time and resource consumption. Empirical profiling on the test system (Intel Core i5-9500 T CPU, 8 GB RAM) revealed an average inference time of 0.82 s per CT image and a peak memory usage of approximately 1.6 GB. These results indicate that the proposed hybrid configuration achieves high accuracy with moderate computational overhead, maintaining a balance between precision and efficiency. While suitable

References	Accuracy (%)	Sensitivity (%)	Precision (%)	Specificity (%)	F1 score (%)
[28]	95.43	93.40	–	97.09	–
[6]	97.88	93.14	–	95.91	–
[29]	96.54	96.35	96.63	–	96.36
[2]	93.67	93.67	–	–	94%
[30]	96.98	96.71	95.97	–	96.77
[31]	96	96.3	–	96.6	96
	96.21	90.71	–	–	92.72
Ours	98.33	98.34	98.33	98.67	98.00

Table 9. Texture features and its formula.

Feature combination	Model size (MB)	Peak RAM (GB)	Inference time (s/image)	Throughput (images/min)
MobileNet only	32	1.2	0.61	98
GLCM + MobileNet	44	1.4	0.74	81
GLCM + SIFT + MobileNet (Proposed)	48	1.6	0.82	73
GLCM + VGG-16	92	2.3	1.45	41
SIFT + MobileNet	40	1.5	0.79	76

Table 10. Runtime and resource utilization analysis.

for desktop-based clinical decision support, further optimization—such as model pruning, quantization, or lightweight fusion—is planned to enable real-time inference on portable and mobile platforms. The results are provided in Table 10.

The superior performance of the proposed GLCM + SIFT + MobileNet framework can be theoretically explained by the complementary nature of its fused features. GLCM provides second-order statistical descriptors that capture spatial gray-level dependencies and fine-grained texture heterogeneity of pulmonary tissue, making it effective for differentiating smooth benign nodules from irregular malignant patterns. SIFT contributes scale- and rotation-invariant key point descriptors, encoding structural edges and geometric contours that remain stable under variable CT slice orientations. MobileNet, in turn, extracts deep hierarchical abstractions—including contextual and semantic cues—through its depth wise separable convolutions that preserve spatial locality while reducing redundancy. When combined, the handcrafted descriptors (GLCM and SIFT) supply low-level radiomic texture evidence, whereas MobileNet's deep features supply high-level contextual semantics. The fusion therefore forms a multi-scale representation that unites micro-texture with macro-context, enabling the model to discriminate subtle intra-class variations and improve robustness to illumination and contrast differences. This synergy explains the consistent rise in accuracy, F1-score, and specificity observed for the GLCM + SIFT + MobileNet combination over single- or dual-feature configurations.

Conclusion

The research aims to predict lung cancer from CT images with high accuracy using a hybrid AI model. CT lung images from three categories—normal, benign, and malignant—are collected and processed. Since the dataset is limited, data augmentation is applied to enhance its size and diversity. Processed images pass through two feature extraction blocks. Feature Block 1 consists of GLCM and SIFT, while Feature Block 2 includes MobileNet and VGG-16. Extracted features from both blocks are fused and fed into an FCL for classification. Three feature combinations are evaluated using 300 test images: GLCM with MobileNet and VGG-16, SIFT with MobileNet and VGG-16, and a combination of GLCM and SIFT with MobileNet and VGG-16. The combination of GLCM and SIFT with MobileNet achieves the highest accuracy, correctly predicting 298 out of 300 images, resulting in an accuracy of 98.33 percent, surpassing other combinations. The proposed hybrid AI network outperforms existing research works and can assist physicians in classifying lung cancer from CT scans.

To assess the generalizability of the proposed hybrid framework beyond the IQ-OTH/NCCD dataset, a small-scale external validation was additionally conducted using ten anonymized CT lung images obtained from an independent open-access repository (Kaggle Lung CT Dataset). The same preprocessing and classification pipeline were applied without retraining or fine-tuning the network. The model achieved an accuracy of 97.2% and an F1-score of 0.96 on this external subset, demonstrating that the performance trend remained consistent with the primary evaluation results. Although the sample size is limited, this brief validation reinforces the robustness and adaptability of the hybrid feature-fusion approach. A comprehensive multi-institutional validation will be pursued as part of our future work to further confirm the model's real-world applicability.

The model is evaluated using only one dataset, and its generalizability needs to be tested on other datasets or real-time data. Future work will involve the validation of other publicly available datasets. Model complexity and processing time were not measured in this study, which are essential factors for hardware deployment and will be addressed in future research.

Data availability

The dataset used in this study, the IQ-OTH/NCCD Lung Cancer Dataset, is publicly available via Kaggle at [https://www.kaggle.com/datasets/adityamahimkar/iqothnccd-lung-cancer-dataset].

Received: 18 July 2025; Accepted: 2 December 2025
Published online: 24 December 2025

References

1. Zappa, C. & Mousa, S. A. Non-small cell lung cancer: Current treatment and future advances. *Transl. Lung Cancer Res.* **5**(3), 288 (2016).
2. Dash, S., Padhy, S., Suman, P. & Das, R. K. Enhancing lung cancer diagnosis through CT scan image analysis using mask-EffNet. *Eng. Access* **11**(1), 92–107 (2025).
3. Venkatesan, N., Pasupathy, S. & Gobinathan, B. An efficient lung cancer detection using optimal SVM and improved weight-based beetle swarm optimization. *Biomed. Signal Process. Control* **88**, 105373 (2024).
4. Williams, H. J. & Davies, A. M. The effect of X-rays on bone: A pictorial review. *Eur. Radiol.* **16**, 619–633 (2006).
5. Yeasmin, M. N., Al Amin, M., Joti, T. J., Aung, Z. & Azim, M. A. Advances of AI in image-based computer-aided diagnosis: A review. *Array* **23**, 100357 (2024).
6. Qadir, A. M., Abdalla, P. A. & Abd, D. F. A hybrid lung cancer model for diagnosis and stage classification from computed tomography images. *Iraqi J. Electr. Electr. Eng.* **20**(2), 266–274 (2024).
7. Abe, A. A. et al. A robust deep learning algorithm for lung cancer detection from computed tomography images. *Intell.-Based Med.* **11**, 100203 (2025).
8. Chiet, C. C., How, K. W., Han, P. Y. & Yen, Y. H. A lung cancer detection with pre-trained CNN models. *J. Inform. Web Eng.* **3**(1), 41–54 (2024).
9. Naseer, I. et al. Lung cancer detection using modified AlexNet architecture and support vector machine. *Comput. Mater. Continua* **74**(1), 2039–2054 (2023).
10. Shah, A. A., Malik, H. A. M., Muhammad, A., Alourani, A. & Butt, Z. A. Deep learning ensemble 2D CNN approach towards the detection of lung cancer. *Sci. Rep.* **13**(1), 2987 (2023).
11. Yan, C. & Razmjooy, N. Optimal lung cancer detection based on CNN optimized and improved Snake optimization algorithm. *Biomed. Signal Process. Control* **86**, 105319 (2023).
12. Maleki, N. & Akhavan Niaki, S. T. An intelligent algorithm for lung cancer diagnosis using extracted features from computerized tomography images. *Healthc. Anal.* **3**, 100150 (2023).
13. Melo-Alvim, C. et al. Radiotherapy, chemotherapy and immunotherapy—current practice and future perspectives for recurrent/metastatic oral cavity squamous cell carcinoma. *Diagnostics* **13**(1), 99 (2022).
14. Jung, H. Basic physical principles and clinical applications of computed tomography. *Progress Med. Phys.* **32**(1), 1–17 (2021).
15. Avinash, S., Naveen Kumar, H. N., Guru Prasad, M. S., Mohan Naik, R. & Parveen, G. Early detection of malignant tumor in lungs using feed-forward neural network and K-nearest neighbor classifier. *SN Comput. Sci.* **4**(2), 195 (2023).
16. Mamun, M.; Mahmud, M. I.; Meherin, M.; Abdelgawad, A. LCDctCNN: Lung cancer diagnosis of CT scan images using CNN based model. In 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), 205–212, IEEE, 2023.
17. Al-Yasriy, H. F. *The IQ-OTH/NCCD Lung Cancer Dataset* (Kaggle, 2023).
18. Barburiceanu, S., Terebes, R. & Meza, S. 3D texture feature extraction and classification using GLCM and LBP-based descriptors. *Appl. Sci.* **11**(5), 2332 (2021).
19. Mohanaiah, P., Sathyanarayana, P. & GuruKumar, L. Image texture feature extraction using GLCM approach. *Int. J. Sci. Res. Publ.* **3**(5), 1–5 (2013).
20. Chaki, J. & Dey, N. *Texture Feature Extraction Techniques for Image Recognition* (Springer, 2020).
21. Burger, W. & Burge, M. J. Scale-invariant feature transform (SIFT). In *Digital Image Processing: An Algorithmic Introduction* 709–763 (Springer, 2022).
22. Lindeberg, T. Feature detection with automatic scale selection. *Int. J. Comput. Vision* **30**, 79–116 (1998).
23. Brown, M. & Lowe, D. G. Invariant features from interest point groups. *In BMVC* **4**(September), 398–410 (2002).
24. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. https://arxiv.org/abs/1409.1556 (2014).
25. Tao, J., Gu, Y., Sun, J., Bie, Y. & Wang, H. Research on VGG16 convolutional neural network feature classification algorithm based on Transfer Learning. In *2021 2nd China International SAR Symposium (CISS)* 1–3 (IEEE, 2021).
26. Howard, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. https://arxiv.org/abs/1409.1556 (2017).
27. Sinha, D., El-Sharkawy, M. Thin MobileNet: An enhanced MobileNet architecture. In *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* 280–285 (IEEE, 2019).
28. Abdollahi, J. Evaluating LeNet algorithms in classification lung cancer from Iraq-oncology teaching hospital/national center for cancer diseases. Preprint at https://arxiv.org/abs/2305.13333 (2023).
29. Talukder, M. A. R., Singho, A. P., Mia, M. M., Jahan, N. & Islam, M. R. Unveiling hidden lesions: A deep dive into deep learning advancements for lung cancer imaging analysis. Thesis (AIUB, 2024).
30. Mostafa, B., Keshk, A. & Sakr, M. A deep-learning model based on transfer-learning technique for detecting and classifying anomalies in lungs images. *Int. J. Comput. Inf.* **10**(3), 63–72 (2023).
31. Mohamed, T. I. A., Oyelade, O. N. & Ezugwu, A. E. Automatic detection and classification of lung cancer CT scans based on deep learning and Ebola optimization search algorithm. *PLoS ONE* **18**(8), e0285796 (2023).

Author contributions

Kamala L: Conceived and designed the study, Writing, Methodology, Collection of datasets. K. G. Mohan: contributed to the development of the methodology, reviewed the manuscript for intellectual content, pseudocode, supervising of full article.

Funding

No funding was received for this study.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to L.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.