RESEARCH ARTICLE

MENet: A Mitscherlich function based ensemble of CNN models to classify lung cancer using CT scans

Surya Majumder[1,2☯], Nandita Gautam[2☯]*, Abhishek Basu[3☯], Arup Sau[2☯], Zong Woo Geem[4]*, Ram Sarkar[2]

1 Department of Computer Science and Engineering, Heritage Institute of Technology, Kolkata, India, **2** Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, **3** Department of Computer Science and Engineering, National Institute of Technology Durgapur, Durgapur, India, **4** College of IT Convergence, Gachon University, Seongnam, South Korea

☯ These authors contributed equally to this work.
* nanditag.cse.rs@jadavpuruniversity.in (NG); geem@gachon.ac.kr (ZWG)

Abstract

Lung cancer is one of the leading causes of cancer-related deaths worldwide. To reduce the mortality rate, early detection and proper treatment should be ensured. Computer-aided diagnosis methods analyze different modalities of medical images to increase diagnostic precision. In this paper, we propose an ensemble model, called the Mitscherlich function-based Ensemble Network (MENet), which combines the prediction probabilities obtained from three deep learning models, namely Xception, InceptionResNetV2, and MobileNetV2, to improve the accuracy of a lung cancer prediction model. The ensemble approach is based on the Mitscherlich function, which produces a fuzzy rank to combine the outputs of the said base classifiers. The proposed method is trained and tested on the two publicly available lung cancer datasets, namely Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) and LIDC-IDRI, both of these are computed tomography (CT) scan datasets. The obtained results in terms of some standard metrics show that the proposed method performs better than state-of-the-art methods. The codes for the proposed work are available at https://github.com/SuryaMajumder/MENet.

Introduction

Lung cancer is cancer that starts in the lungs, usually in the cells that line the airways. It is one of the main reasons for cancer-related fatalities globally. According to the World Health Organization (WHO), there were 2.21 million new cases of lung cancer in 2020 [1], making it the second most common type of cancer in terms of new cases. Smoking, exposure to certain chemicals and pollutants, and a family history of the disease are all risk factors for lung cancer. Cancer significantly affects a person's quality of life, as well as their physical and mental health. To reduce mortality and improve patient's quality of life, awareness of cancer prevention, early detection and effective treatment options must be increased. The use of computer-aided

detection (CAD) methods to help medical workers in the early detection of diseases like lung cancer has grown in popularity in recent years. These methods analyze medical images using machine learning algorithms. CAD systems can increase diagnostic precision and decrease the likelihood of human mistakes by automating the diagnosis process. Recent years have seen the publication of several study papers discussing the application of machine learning methods such as [2–4] for early detection of lung cancer. These techniques generally involve using a collection of the particular modalities of medical images to teach a machine learning algorithm and derive useful features that can be used for identifying the presence or absence of lung cancer, or the severity of the same. Over the last few decades, Convolutional Neural Networks (CNNs) [5] have been incredibly effective in overcoming several difficulties related to image classification and pattern recognition. Unlike conventional machine learning algorithms, CNNs do not need any feature-engineering methods, rather they can automatically extract the relevant features from the input data. CNNs, a subclass of deep neural networks, have transformed the field of computer vision by getting cutting-edge results on a variety of visual recognition tasks, including image segmentation, object detection, and classification. CNNs are a powerful tool used for different applications, but they are not a silver bullet and may require support from other methods to improve performance [6, 7].

Also, training deep neural networks from scratch requires large amounts of data and computational resources, and can take a long time to converge. In some cases, it may be impractical or impossible to form a deep network from scratch due to these limitations. Transfer learning provides a solution [8, 9] that allows us to use pre-trained models that have been trained on large datasets and learned useful feature representations. By using pre-trained models or transferring knowledge from one task to another, transfer learning can boost the performance of a single model. By lowering the variance and enhancing the models' capacity for generalization, ensembling multiple models can further enhance performance. Ensemble learning can assist in building more reliable and precise models that are better able to manage real-world data by combining the decision-making abilities of varied models. The use of component classifiers learned from various groups to create a composite categorization system was suggested by [10] in order to improve the performance of identification systems. The popularity of ensemble learning has grown steadily in recent years [11–17]. Nowadays, ensemble learning is considered a key technique in the machine learning toolkit and is widely used in industry and the academic world. There are several approaches like bagging [18], boosting [19], and staking [20] for classification tasks in ensemble learning based on the specific application and available data.

Considering these details, in this study, we have assumed that ensemble learning can be an effective way to detect lung cancer, which involves identifying different types of lung abnormalities from medical computed tomography (CT) scans. This approach may be helpful to tackle few challenges, such as image quality variations, noise, distortion, low-density contrast, etc., as shown in Fig 1. The overall pipeline of the proposed model is represented in the diagram presented in Fig 2. The main **contributions** of this work are as follows:

- We propose an ensemble model, called MENet, for lung cancer classification using CT scans.

- MENet combines confidence scores obtained from three transfer learning-based CNN models, namely Xception, InceptionResNetV2, and MobileNetV2.

- We also propose a fuzzy ranking system based on the Mitscherlich function to rank and combine the outputs of different base classifiers for forming an ensemble-based prediction model.
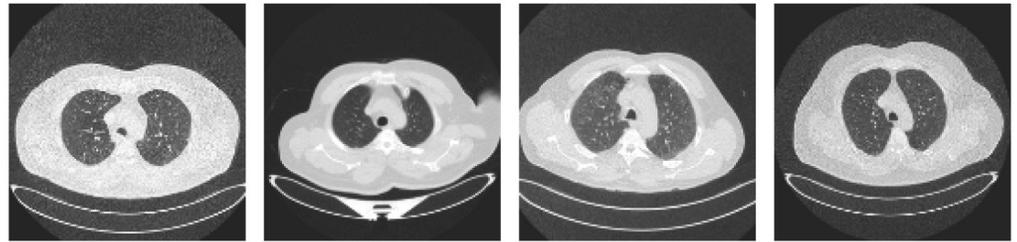
Fig 1. CT scan images showing noise and distortion, low-density contrast, and quality variation, respectively.

https://doi.org/10.1371/journal.pone.0298527.g001
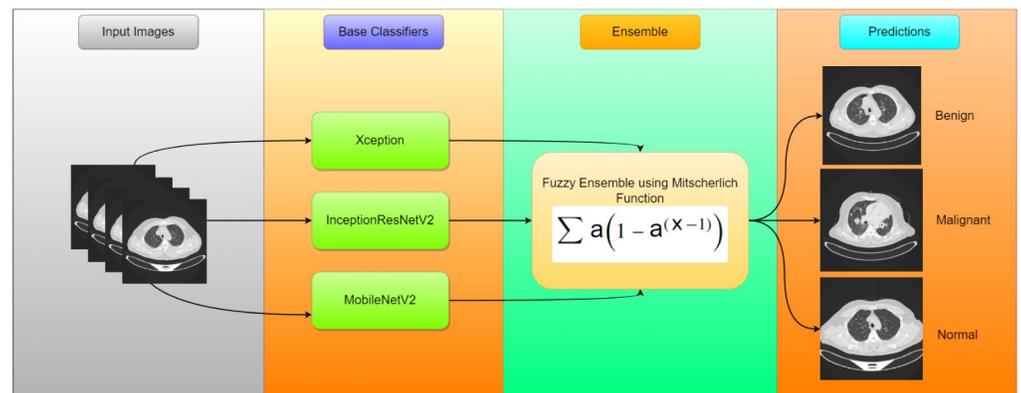
Fig 2. Overall pipeline of the proposed model, called MENet, highlighting the base models and the ensemble procedure.

https://doi.org/10.1371/journal.pone.0298527.g002

- Our proposed method is trained and tested on two publicly available lung CT scan datasets, namely IQ-OTHNCCD and LIDC-IDRI.

- MENet outperforms the existing results in lung cancer prediction with an accuracy of 99.54% and 95.75% on IQ-OTHNCCD and LIDC-IDRI datasets, respectively.

Literature review

The paper's literature review has two distinct segments. The first segment presents a comprehensive summary of previous studies conducted on lung cancer classification. The second segment concentrates on examining ensemble techniques, specifically in the realm of medical image processing.

A review of lung cancer classification

In the study, [21] presented an approach for the classification of lung nodules and non-nodules in CT images, which involved the utilization of texture features. They used three strategies to extract texture measurements: rose diagrams (RD), artificial crawlers (ACs), and a hybrid model that incorporates ACs & RD. The authors employed a support vector machine (SVM) classifier with a radial basis kernel to differentiate between nodules and non-nodules in the candidate categorization process. The study used the Lung Image Database Consortium (LIDC-IDRI) image database and achieved a mean specificity of 94.78%, mean sensitivity of

91.86%, and mean accuracy of 94.30%. However, the main limitation of this work is that the comparison with previous methods is only approximate due to differences in databases and test scenarios. The only precise comparison is with [22], and the proposed methodology achieved a lower specificity score, indicating its lesser ability to classify non-nodule cases accurately. Marcin et al. [23] introduced a novel approach to classify lung carcinomas. This method involves the localization and extraction of lung nodules by computing the local variance of pixels, detecting the maxima, and utilizing a probabilistic neural network as a classifier. The simplicity of the proposed method enables it to detect low-contrast nodules and reduce computational workload while maintaining performance. To explain the nodule appearance without ignoring spatial information, a 7th Order Markov Gibbs Random Field and a Local Binary Pattern are devised by Shaffie et al. [24].

Netto et al. [25] proposed a methodology for analyzing lung lesions using temporal evaluation, which can aid in the diagnosis of indeterminate lesions during treatment. The modified quality threshold clustering technique was employed to assign each voxel of the lesion to a cluster, and the alteration in the lesion was evaluated by analyzing the movement of voxels to other clusters over time. To differentiate between benign and malignant lesions, statistical features were extracted. The authors employed two databases of pulmonary lesions, one for malignant lesions under treatment and the other for undetermined cases, to develop their proposed methodology. By analyzing the density changes of lesions over time, the researchers achieved an accuracy of 98.41% in accurately identifying lung lesions. Xie et al. [26] proposed a distinctive approach for lung nodule classification, named Fuse-TSD. This method combines information on texture, shape, and deep model-based learning at the decision level. The approach utilizes a texture descriptor derived from a gray-level co-occurrence matrix (GLCM), a Fourier shape descriptor that captures the heterogeneity of nodules, and a deep convolutional neural network (DCNN) that learns the feature representation of nodules automatically, slice-by-slice. The obtained characteristics are trained with an AdaBoosted back propagation neural network (BPNN), and the judgments of three classifiers are combined to distinguish benign from malignant nodules.

An automated diagnosis classification method for CT lung images was introduced by [27]. The method employs an Optimal Deep Neural Network and Linear Discriminate Analysis to extract deep features and reduce dimensionality. To optimize the Optimal Deep Neural Network for classifying lung nodules as benign or malignant, the authors used a modified gravitational search algorithm. The automated approach not only enhances the classification accuracy but also reduces the time required for manual labeling and prevents human errors in recognizing normal and abnormal lung images. Shafi et al. [28] proposes deep learning supported SVM model for CT images and gain accuracy of 94%. A lightweight, multi-view sampling-based multi-section CNN architecture was introduced by [29], which efficiently captures the structural information of nodules from CT scans for lung cancer diagnosis. The method utilizes a view pooling layer to aggregate information from multiple cross-sections of the nodule. It encodes its volumetric information into a concise representation that is then utilized for nodule classification. Masood et al. [30] developed a decision support system using CT scans for lung nodule detection. Their approach utilized a 3D deep CNN, multi-region proposal network, and median intensity projection to automatically identify regions of interest. The performance of the system was evaluated on LUNA16, ANODE09, and LIDC-IDRI datasets. The main shortcoming of this work is that the accuracy of detecting micro-nodules with a diameter of less than 3 mm is relatively low.

Lin et al. [31] put forth a framework for classifying lung cancer using Generative Adversarial Networks (GANs) and discriminator networks. They utilized a deconvolution layer, leaky Rectified Linear Unit (ReLU) activation function, and batch normalization to obtain a

64×64×3 image. The main objective of their research was to overcome the challenge of sparse image data and generate precise predictions. Zhao et al. [32] proposed a forward and backward GAN with multi-scale VGG16. Yuan et al. [33] proposed a CAD method to enhance the detection of pulmonary nodules on CT scans. Their proposed method utilized a 3D Residual U-Net model along with a multi-branch classification network to achieve a high detection sensitivity of 94.0% and a competition performance metric (CPM) score of 0.959 on the LUNA 2016 dataset through multi-task learning. Bhatia et al. [34] obtained an accuracy of 84% using the same model on the LIDC-IDRI dataset. Halder et al. [35] presented an end-to-end system for detecting and classifying lung nodules from high-resolution CT images using atrous convolution. They achieved high-performance indices, with the proposed architecture, ATCNN2PR, consisting of a two-layer atrous pyramid and residual connections, demonstrating the highest classification accuracy. The results showed that the system outperformed other competing frameworks with an accuracy, specificity, and sensitivity of 95.97%, 96.89%, and, 95.84%, respectively.

Survey of ensemble techniques

Ensemble learning is a technique that combines many learning models to increase the overall accuracy and resilience of the learning framework. In this section, we will give a quick overview of various recently proposed ensemble learning approaches with an emphasis on medical picture analysis. Maji et al. [36] introduced a computational imaging system for the detection of blood vessels in fundus color images using ensemble learning and deep learning. Their approach involved training an ensemble of deep CNNs to segment vessel and non-vessel regions of a color fundus image. During inference, the responses of individual ConvNets within the ensemble were combined by averaging to generate the final segmentation. The method was evaluated using the DRIVE database, and the results showed a maximum average accuracy of 94.7%. Kassan et al. [37] presented an ensemble deep learning-based approach for binary classification of breast histopathology images. They utilized three pre-trained CNNs for feature extraction and a multi-layer perceptron classifier for the final classification. The proposed technique outperformed individual classifiers and other machine learning algorithms in terms of prediction accuracy on four benchmark datasets. Bhowal et al. [38] developed a CAD system for breast cancer detection using a Choquet Integral fusion technique that considers subsets of classifiers. For the categorization of breast cancer histology images, their proposed method included InceptionV3, Xception, VGG19, VGG16, and InceptionResNetV2. To address the complexity of calculating fuzzy measurements, the authors used the Coalition Game, Information Theory, and constructed a novel mathematical function. The ICIAR 2018 Grand Challenge on Breast Cancer Histology (BACH) photos were utilized for evaluation, which included 2-class and 4-class tasks. The fusion method outperformed all individual models with a test accuracy of 96% for the two-class problem. Similarly, the fusion approach achieved a test accuracy of 95% for the four-class problem. Table 1 shows some recent applications of ensemble learning in the field of medical imaging.

In our proposed ensemble approach, we use the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) [54] lung cancer dataset taken over a period of three months (in the fall of 2019 and published in 2020). The dataset contains CT scan images of healthy and lung cancer patients who were diagnosed through different stages of lung cancer. This dataset is publicly available [55–57] and contains three classes of images. Among the three classes, the benign class contains 120 images, the malignant class contains 561 images, and the other 461 images from normal patients as shown in Table 2. We took 70% of the data to train the models, 20% data for validation, and the remaining 10% for testing.

Table 1. Applications of ensemble learning in medical image analysis.

Application Domain	Reference
Glioma detection	[39]
COVID-19 detection	[40–43]
Breast cancer	[44–48]
Colorectal polyp classification	[49]
Brain cancer detection	[50]
Lymphoblastic leukemia cell image analysis	[51]
Alzheimer's disease classification	[52]
Monkeypox detection	[53]

https://doi.org/10.1371/journal.pone.0298527.t001

Table 2. Distribution of images found in the IQ-OTHNCCD dataset into Benign, Malignant, and Normal classes.

Dataset	Class	No. of images
IQ-OTHNCCD	Benign	120
	Malignant	561
	Normal	461

https://doi.org/10.1371/journal.pone.0298527.t002

Proposed methodology

The suggested ensemble-based lung cancer classification model (MENet) is described in-depth in this section. In order to determine the correct class of the test data, we first provide a brief description of every base learner that produces scores of confidence for an incoming lung image. These scores are then further fused via the suggested ensemble methodology.

Deep neural based classifiers

CNNs are preferred over other types of machine learning algorithms for image classification tasks, including the detection of lung cancer from medical imaging datasets. One of the main reasons for this is that CNNs are specifically designed to handle spatial data such as images. They can learn features from the input data by using convolutional layers and pooling layers combination, which can effectively capture local patterns and spatial relationships between pixels. This makes them well-suited for analyzing medical images, which often contain complex structures and patterns. Additionally, CNNs can automatically learn and adapt to the features of the input data, without the need for manual feature engineering. Overall, these characteristics make CNNs a powerful tool for accurately classifying medical images, including those of lung cancer and have the potential to increase diagnosis precision and speed. Hereafter an exhaustive set of experiments, we have successfully settled down with Xception, InceptionResNetV2, and MobileNetV2 models. The choice of these architectures likely balances a trade-off between accuracy and computational efficiency.

Xception. Xception is a convolutional neural network architecture proposed by [58]. It is an extension of the Inception architecture and is named "Extreme Inception" because it uses depthwise separable convolutions instead of standard convolutions used in Inception. The depth-wise separable convolution layer blocks make up the bulk of the Xception architecture, which is then followed by batch normalization and ReLU activation. The depthwise separable convolution layers are made up of two distinct layers: a pointwise convolution layer that applies a 1x1 convolutional filter to combine the output channels of the depthwise convolution and a depthwise convolution layer that applies a single convolutional filter to each input

Fig 3. Architecture of the Xception model.

https://doi.org/10.1371/journal.pone.0298527.g003

channel. A global average pooling layer and a fully connected layer for classification, constitute the final layer of the network. The depthwise separable convolutions are more computationally efficient than conventional convolutions and aid in lowering overfitting, and the residual connections enable deeper network topologies while addressing the vanishing gradient issue. This makes Xception a powerful and effective model for image classification tasks. The Xception architecture is illustrated in Fig 3.

InceptionResNetV2. Introduced by [59], InceptionResNetV2 combines the strengths of Inception and ResNet architectures. It utilizes residual connections with Inception modules, which not only address the vanishing gradient problem but also enable more complex network designs. Additionally, the model requires fewer parameters to achieve high accuracy and can quickly learn and extract features from input data. Overall, the InceptionResNetV2 architecture is a powerful tool for image classification tasks, offering both high accuracy and efficient computation. The InceptionResNetV2 architecture is shown in Fig 4.

MobileNetV2. MobileNetV2, proposed by [60] is a convolutional neural network designed for mobile and embedded vision applications. It uses depthwise separable convolution, inverted residuals, and linear bottlenecks to improve efficiency while preserving representational power. It includes two types of blocks: residual blocks with a stride of 1 and blocks with a stride of 2 for downsizing. Both types of blocks consist of three layers. The first layer in each block is a 1×1 convolution with ReLU6 activation. The second layer is a depthwise convolution, which applies a separate convolutional filter to each input channel. The final layer in each block is another 1×1 convolution but without any non-linearity. A width multiplier is used in this network to optimize the network for different hardware and resource constraints. MobileNetV2 is a highly efficient and lightweight model, making it suitable for deployment on

Fig 4. Architecture of the InceptionResNetV2 model.

https://doi.org/10.1371/journal.pone.0298527.g004

Fig 5. Architecture of the MobileNetV2 model.

https://doi.org/10.1371/journal.pone.0298527.g005

mobile devices and other resource-constrained environments. Despite its small size, Mobile-NetV2 achieves high accuracy on a wide range of vision tasks, including image classification, object detection, and semantic segmentation. The architecture of MobileNet is shown in Fig 5.

Proposed fuzzy-ensemble method

The main goal of our proposed approach is to provide greater adaptability and freedom in handling datasets with varying degrees of complexity. Using a fuzzy ranking-based approach, we can consider the uncertainty of each classifier's predictions and assign different levels of importance to each classifier based on its performance on a particular test case. In the proposed methodology, the three CNN-based classifiers (Xception, InceptionResNetV2, and MobileNetV2) are used to detect lung cancer cases from CT scans, and fuzzy ranks are generated for each of them using the re-parameterized Mitscherlich function. To increase the overall accuracy of the classification, these fuzzy ranks are then combined using an ensemble method. The Mitscherlich function, which has been applied to the field of machine learning by [61, 62], is based on the idea of crop yield response [63] to various levels of fertilizer application [64]. Here, it is used to combine the outputs of different models with various strengths and weaknesses.

Significance of Mitscherlich function

The Mitscherlich function plays a pivotal role in our rank-based fuzzy ensemble logic. Its selection was made based on a multitude of factors, which we shall delve into, in this section. The Mitscherlich function serves as a crucial tool for illustrating the performance of each model within our ensemble as a function of input strength. The function has a similarity to linear mapping within the (0,1) range. To address this concern, we will provide a more comprehensive explanation of our rationale and support it with relevant discussions and experiments.

In essence, the Mitscherlich function allows us to create a dose-response curve that portrays each model's performance concerning the input strength. This curve represents the relationship between the weighted sum of the models' predictions and the actual outcomes, yielding a value that quantifies the fit between predictions and actual results. Here's where its unique value becomes apparent: we employ this function to generate fuzzy ranks for each class from the ensemble's individual models. The process involves ranking the confidence scores of each class provided by the base classifiers. These fuzzy ranks are then employed in ensemble predictions on a validation set, specifically within the top K ranks. Why is this important? Well, it enables us to construct an adaptable ensemble model capable of dynamically adjusting the ranking of individual models based on the specific characteristics of each input instance. As the decision score for a class, accurately predicted by a classifier, approaches one, the Mitscherlich function exhibits a steeply dropping nature within the domain range of 0 to 1. This characteristic proves highly advantageous when forming an ensemble of decision scores from various learning models. It ensures that as we approach a confident prediction (score near one), the Mitscherlich function's response is highly sensitive, effectively distinguishing between models with slight performance variations in this critical region.

Implementation of the ensemble model

There should be M different decision scores (also known as confidence scores of classifiers) such as $CoF^{(1)}$, $CoF^{(2)}$, ..., $CoF^{(M)}$ for each input image P. As we utilized three different CNN-based models to produce the scores of confidence on the dataset, M, in our case, equals 3. In the Eq 1, the decision scores from the dataset are normalized, where C is the total count of classes in the dataset under consideration.

$$\sum_{c=1}^{C} CoF_c^{(i)} = 1; \forall i, \ i = 1, 2, 3, \ldots, M \tag{1}$$

Fuzzy ranks are generated using the scores of confidence of each sample in the dataset, which are divided into three different classes. The Mitscherlich function produces the fuzzy rank for a class c using the i_{th} classifier's scores of confidence as shown Eq 2.

$$R_c^{(i)} = 2(1 - 2^{CoF_c^{(i)} - 1}); \forall i, c, \ i = 1, 2, 3, \ldots, M; c = 1, 2, \ldots, C \tag{2}$$

The Fig 6 shows a pictorial depiction of the modified Mitscherlich function graph, as mentioned in Eq 2. The value of $R_c^{(i)}$ is in-between 0 and 1, with the lowest value 0 being equivalent to rank 1 (best rank), i.e., a greater score of confidence results in a lower (better) value of rank. The fuzzy rank sum (FRS_c) and complement of the confidence factor sum ($CCFS_c$) are computed as follows if $K^{(i)}$ is used to represent only the top k ranks, that is, rank 1, 2,..., k, that corresponds to class c are as follows:

$$FRS_c = \sum_{i=1}^{M} \begin{cases} R_c^{(i)}, & \text{if } R_c^{(i)} \in K^{(i)} \\ P_c^R, & \text{otherwise} \end{cases} \tag{3}$$

$$CCFS_c = \frac{1}{M} \sum_{i=1}^{M} \begin{cases} CoF_c^{(i)}, & \text{if } R_c^{(i)} \in K^{(i)} \\ P_c^{CoF}, & \text{otherwise} \end{cases} \tag{4}$$
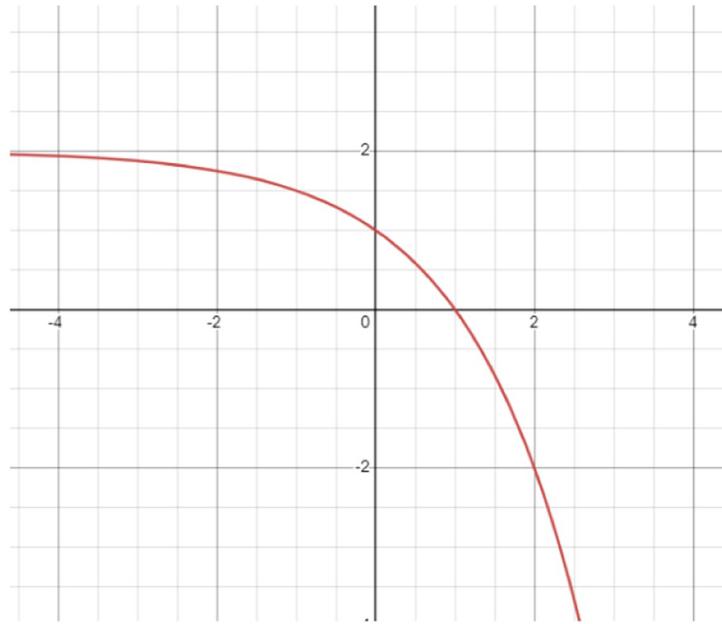
Fig 6. Graphical representation of the modified Mitscherlich function used in the present work.

If class c does not fall among the top k class ranks, penalties P_c^R and P_c^{CoF} are applied to it. Using the aforementioned Mitscherlich function, the value of P_c^R is set to 1, which is determined by setting $CoF_c^{(i)} = 0$; and the value of P_c^{CoF} is set to 0. These penalty values prevent class c from emerging as an improbable winner. The combination of FRS_c and $CCFS_c$, which is used to produce the ensemble model's final predictions, yields the decision score in question. Eq 5 determines the final decision score (FDS).

$$FDS_c = FRS_c * CCFS_c \tag{5}$$

Finding the class with the lowest FDS value yields the final projected class for data instance **I**, which is provided as shown in Eq 6.

$$class(\mathbf{I}) = \arg \min_{c=1,2,\ldots,C} FDS_c \tag{6}$$

A dry run of the proposed fuzzy rank-based ensemble method is shown under S1 Appendix.

Results and discussion

In this section, we report the detailed results and the corresponding analysis of the proposed ensemble of CNN models used for lung cancer detection in CT scans. The distribution of images in the dataset is already provided in Table 2. The implications of the obtained results are also discussed. In addition, we present a comparative evaluation to show that the proposed method outperforms other models and commonly used ensemble techniques implemented in the literature.

Table 3. Evaluation matrix used in our classification methods.

Metric	Formula
Accuracy(Acc)	$\frac{T^P+T^N}{T^P+T^N+F^P+F^N}$
Precision(Pre)	$\frac{T^P}{T^P+F^P}$
Recall(Re)	$\frac{T^P}{T^P+F^N}$
F1-Score(F1)	$\frac{T^P}{T^P+\frac{1}{2}(F^P+F^N)}$

System configuration

The entire set of experiments has been run in Jupyter equipped with a 12 GB NVIDIA Tesla T4 GPU that has been made available by Google Colab. Python 3 environment, along with open-source modules such as Tensorflow, Keras, Matplotlib, Scikit, Numpy, and Pandas is utilized for the implementation.

Evaluation metrics

Evaluation metrics are important in assessing the effectiveness and strength of a predictive/ learning model. It is important to use a variety of standard evaluation metrics to get a complete picture of the model's performance and to ensure that it meets the requirements of the problem under consideration. In classification problems, these metrics are used to evaluate the performance in predicting the correct class label for a given input. Table 3 shows various performance measures used in our classification method. Consider there is a two-class classification problem, where one class is termed as 'positive' and another one is termed as 'negative'. Most of these measures are computed using a confusion matrix employing four fundamental components, such as true positive (T^P) rate, true negative (T^N) rate, false positive(F^P) rate, and false negative(F^N) rate.

Implementation

Initially, we perform extensive experimentation with different combinations of CNN models to determine the best combination of base learners for our proposed ensemble technique. The hyperparameters selected for this experiment are mentioned in Table 4. The results of the experimentation are shown in Table 5.

The best accuracy is obtained by selecting Xception, InceptionResnetV2, and MobileNetV2 as the base classifers for the ensemble approach. The metric scores given by these models along with some additional information is shown in Table 6. These three models, i.e., Xception, InceptionResNetV2, and MobileNetV2 give us an accuracy score of 99.02%, 97.26%, and 99.45%, respectively. The results show that these models are very reliable for being chosen for this fuzzy ensemble.

Table 4. Hyperparameters of the base classifiers.

Hyperparameter	Value/Name
Optimizer	Adam
Loss function	Sparse Categorical Cross Entropy
Learning rate	0.001
No. of epochs	60

Table 5. Results of experiments implemented to determine the base classifiers for forming the ensemble in this study.

Model 1	Model 2	Model 3	Accuracy (%)
NasNetMobile	MobileNet	ResNet-152	95.98
NasNetMobile	MobileNetV2	ResNet-50	94.87
NasNetMobile	MobileNetV2	DenseNet-169	95.21
NasNetMobile	MobileNetV2	InceptionResNetV2	92.56
MobileNetV2	ResNet-152	DenseNet-169	96.41
MobileNetV2	ResNet-50	DenseNet-169	92.99
MobileNetV2	ResNet-50	InceptionResNetV2	97.17
MobileNetV2	ResNet-152	InceptionResNetV2	96.67
NasNetMobile	ResNet-152	DenseNet-169	97.00
NasNetMobile	ResNet-152	InceptionResNetV2	96.41
NasNetMobile	ResNet-50	DenseNet-169	96.06
NasNetMobile	ResNet-50	InceptionResNetV2	96.92
Xception	ResNet-152	DenseNet-121	97.77
Xception	ResNet-152	InceptionResNetV2	95.04
Xception	ResNet-18	InceptionResNetV2	98.20
Xception	MobileNet	DenseNet-121	98.29
Xception	ResNet-18	MobileNetV2	98.54
Xception	MobileNet	NasNetMobile	98.12
Xception	**InceptionResNetV2**	**MobileNetV2**	99.54

Table 6. Performance measure of each model along with their total number of parameters.

Model	No. of parameters	Acc (%)	Pre (%)	Re (%)
Xception	20,867,627	99.02	99.38	99.59
InceptionResNetV2	54,341,347	97.26	88.33	93.80
MobileNetV2	2,261,827	99.45	99.59	99.79
Proposed Ensemble	**77,470,801**	**99.54**	**99.62**	**98.61**

Each of these base models generates the confidence scores for each class for every single image in the dataset. These confidence scores of all the base models for each image are generated and stored for each individual classifier. For the results obtained in Table 6, the three transfer-learned models have been trained for 60 epochs with the Adam optimizer individually. Fig 7 shows their respective confusion matrices on the dataset used. Although there are some misclassifications, the ratio of misclassifications to that of the correct classifications is very low, indicating that these models are reliable for the work. The loss curves obtained by each of the base learners(or classifiers) are shown in Fig 8. As our models are all transfer learning-based models that have been pre-trained on the ImageNet dataset, we just had to fine-tune the models on our IQ-OTHNCCD lung cancer dataset. From the loss curves of the three models, we can see that there is hardly any problem with overfitting in the three models. The accuracy curves as shown in Fig 9 display the corresponding accuracy scores as shown in Table 6. After fusing the confidence scores of the three models using the proposed Mitscherlich function-based ensemble model, we get the results.

From Figs 10 and 11 and Table 7, we can observe that the overall accuracy achieved is 99.54% after combining the results of the three base classifiers. The class-wise results have also been given in the Table 7. All these high classification accuracies indicate that the model is
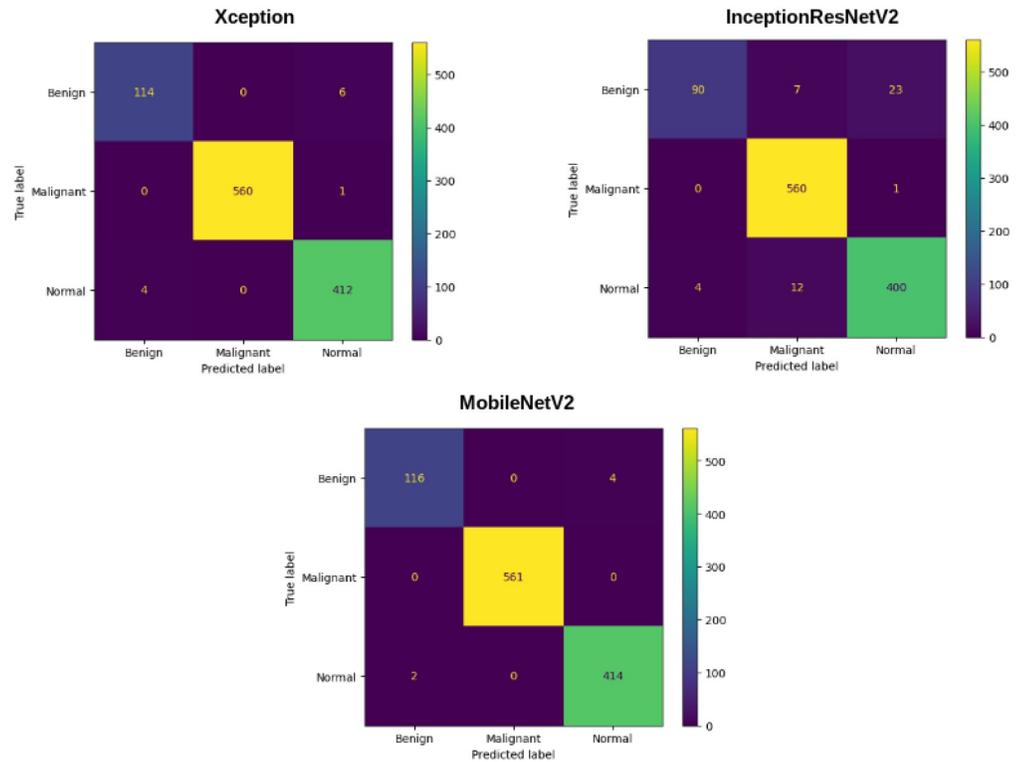
Fig 7. Confusion matrix obtained on the IQ-OTHNCCD dataset by Xception, InceptionResNetV2, and MobileNetV2 respectively.

https://doi.org/10.1371/journal.pone.0298527.g007
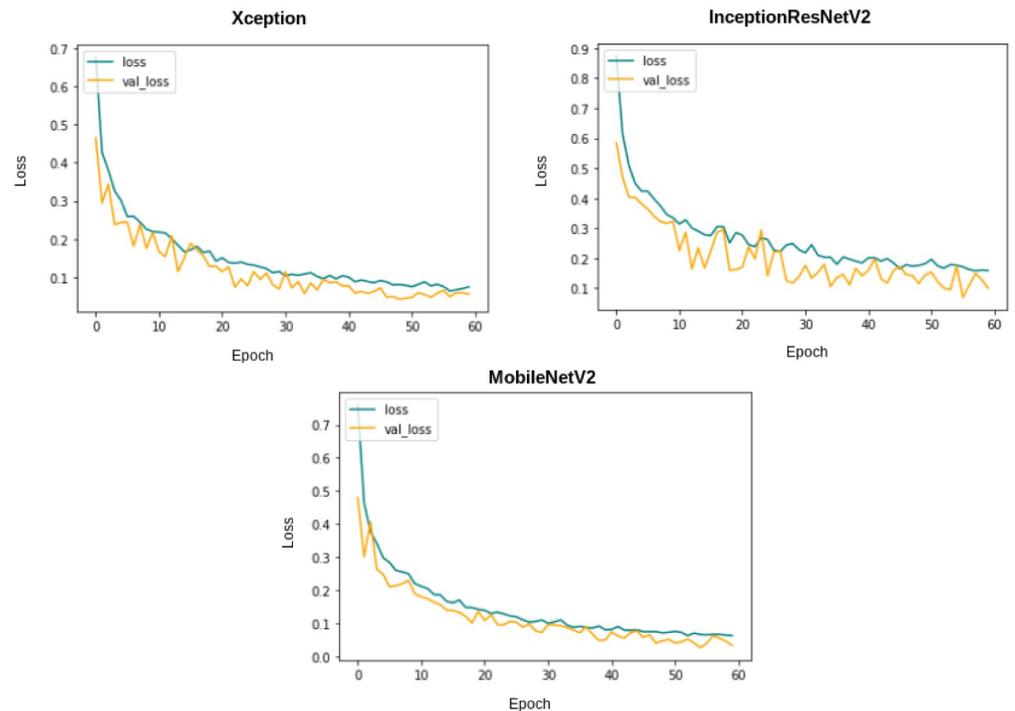
Fig 8. Training and validation loss plots for 60 epochs on the IQ-OTHNCCD dataset.

https://doi.org/10.1371/journal.pone.0298527.g008
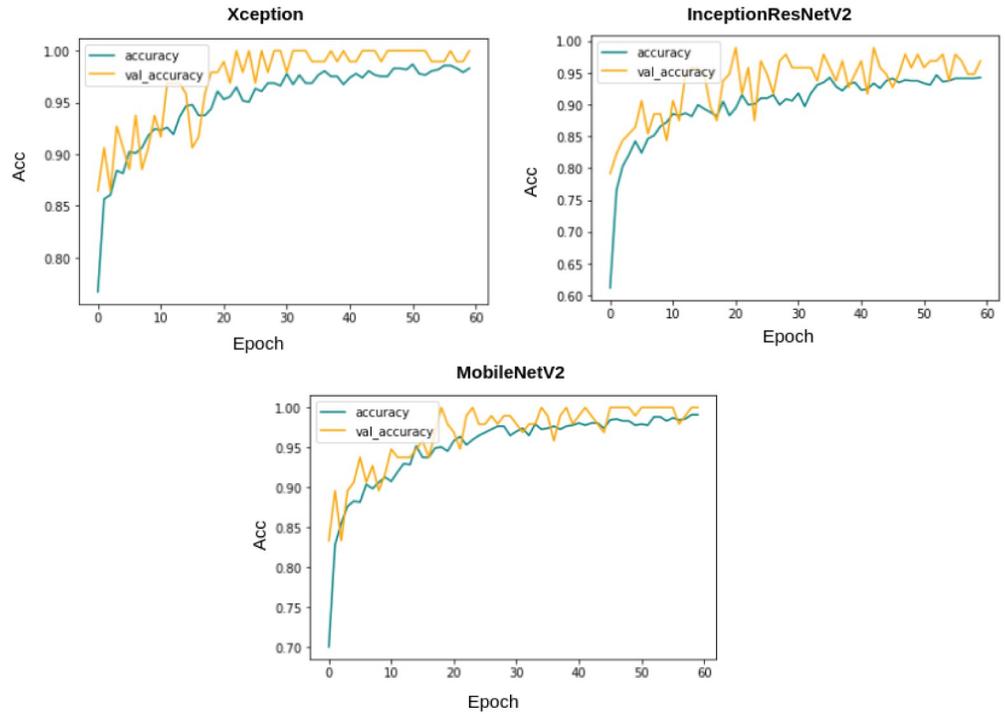
Fig 9. Training and validation accuracy plots for 60 epochs on the IQ-OTHNCCD dataset.

highly reliable. The final ROC curve obtained on this dataset shows that the False Positive rate is significantly low. This indicates that this proposed fuzzy ensemble method is very efficient for the problem under consideration. The Confusion Matrix of this proposed fuzzy ensemble method, as shown in Fig 10 also shows very few mispredictions than all the three base
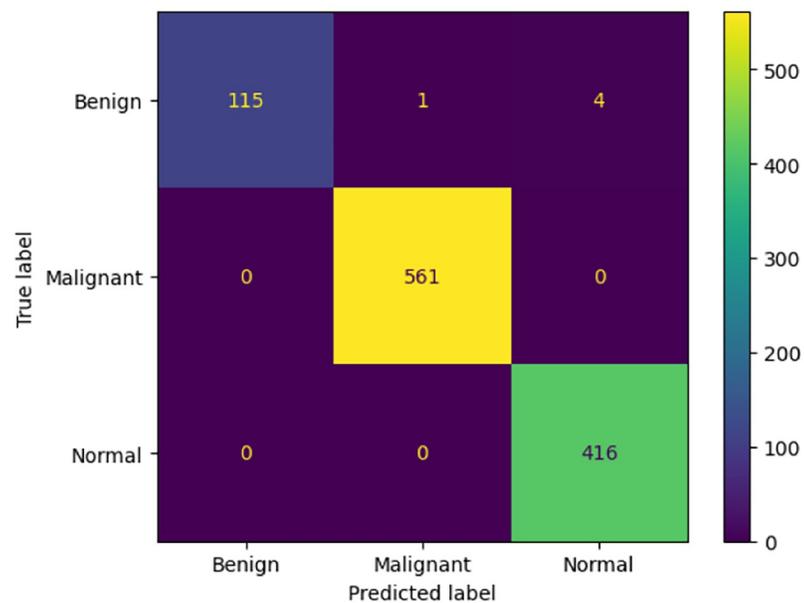
Fig 10. Confusion matrix of proposed ensemble using Mitscherlich function (MENet).
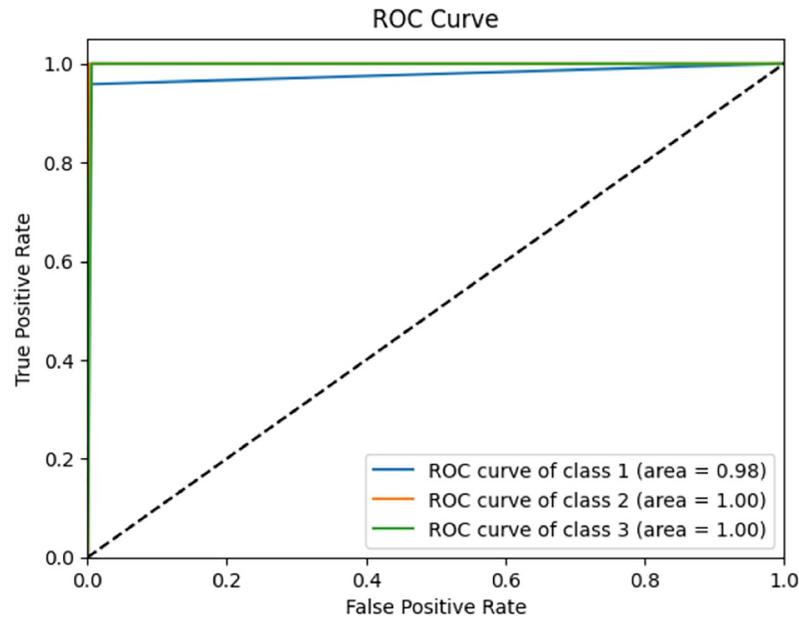
Fig 11. ROC-AUC curve obtained by the proposed ensemble model (MENet).

classifiers. The area under the curve (AUC) is shown in Fig 11. Here, we can observe for class 1, i.e., Benign: 98%, for class 2, i.e., Malignant: 100%, and for class 3, i.e., Normal: 100%. As the graph in Fig 11 is shifted to the top left corner, it indicates the highly efficient ability of the final model to classify test images into one of the three classes. Fig 12 provides a detailed pictorial representation of the performances of each base model and the final ensemble method.

Comparison with state-of-the-art methods

In our proposed method, we have evaluated our model using the Mitscherlich function-based fuzzy ranking-based ensemble approach. Fuzzy ranks are generated for each of the base classifiers, and then these ranks are combined using the ensemble method. The proposed method is compared with the methods found in the literature, and it has been observed that the proposed method achieves the highest accuracy score among others. In comparison, the dataset we have used and compared with others is the same dataset (i.e., the IQ-OTHNCCD lung cancer dataset). Table 8 provides an illustration of the comparative study of our suggested model with others.

Our proposed ensemble approach achieves the best performance accuracy of 99.54% as compared to the other methods or individual classifiers. Out of these, the methods proposed by [57, 65, 66] are based on the deep CNN methods and achieve accuracy of 95.71%, 97.00%, and 99.45%, respectively. [67] have attained 98.83% accuracy by using three transfer learning

Table 7. Class-wise and overall results obtained by the proposed ensemble model.

Class	Pre	Re	F1	Support	Acc	Overall Acc	Overall Re	Overall F1
Benign	1.0000	0.9583	0.9787	120	0.95833333	0.9954	0.9861	0.9910
Malignant	0.9982	1.0000	0.9991	561	1.			
Normal	0.9905	1.0000	0.9952	416	1.			

Overall Accuracy

(a)

Overall Precision

(b)

Overall Recall

(c)
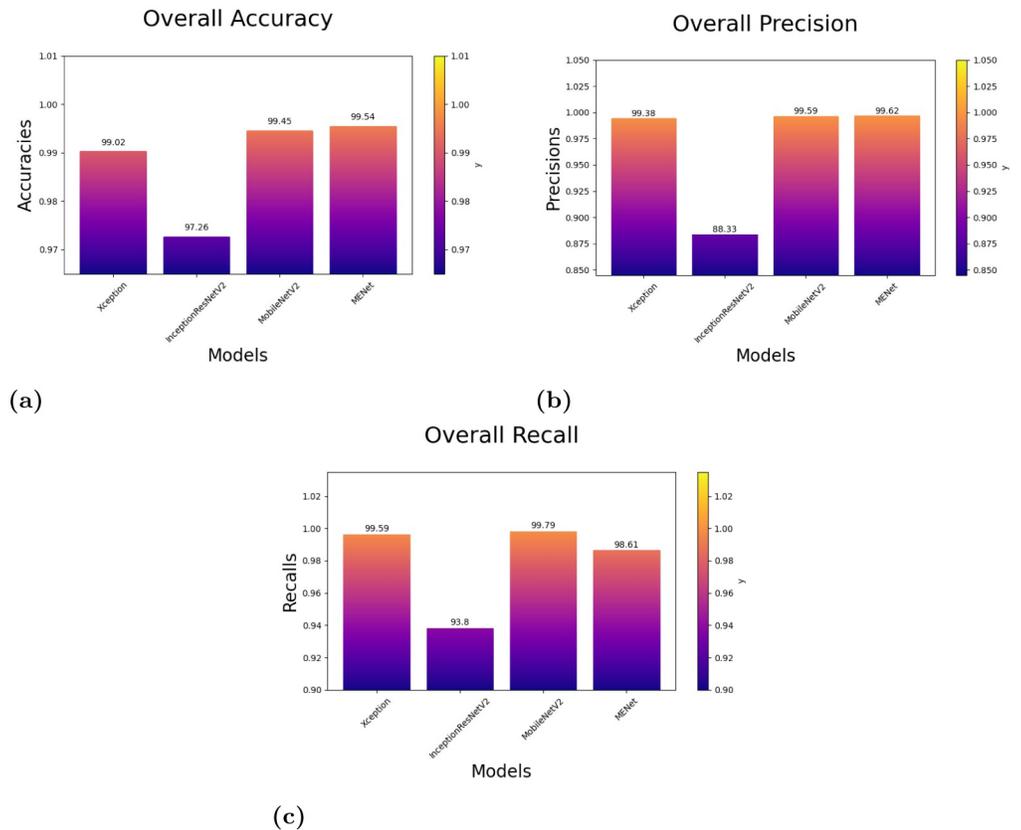
Fig 12. Results of the Models depicting their overall accuracies, precisions and recalls on **the IQ-OTHNCCD dataset**: (a) Overall Accuracies of the models, (b) Overall Precisions of the models, (c) Overall Recalls of the models.

models, VGG16, VGG19, and Xception, with CNN for the classification of lung cancer on the same dataset. [68, 69] achieve accuracy of 98.58% and 97.38%.

Comparison with other ensemble techniques

In this section, experimental results are compiled in Table 9 to demonstrate the proposed ensemble scheme's superiority to well-known traditional ensemble techniques. The ensembles used the identical three base CNN learners, Xception, InceptionResNetV2, and MobileNetV2. The suggested ensemble method performed better than several popularly used ensemble

Table 8. Performance comparison of the proposed ensemble model with state-of-the-art methods on the IQ-OTH/NCCD dataset.

Method	Approach	Acc (%)
Al-Yasriy et al. [57]	AlexNet	95.71
Humayun et al. [67]	VGG16,VGG19 and Xception	98.83
Narin et al. [68]	Alexnet and ResNet50	98.58
Al-Huseiny et al. [69]	Transfer Learning with GoogleNet	94.38
Begum et al. [65]	SMOTE based on 2D-CNN	97.00
Abunajm et al. [66]	CNN	99.45
Proposed (MENet)	**An ensemble of CNN models**	**99.54**

Table 9. Performance comparison of the proposed ensemble method with some commonly used ensemble methods evaluated on the IQ-OTHNCCD dataset. Scores are in %.

Ensemble method	Acc	Pre	Re	F1-score	AUC
Maximum Probability	98.11	98.43	97.79	97.76	97.79
Average Probability	98.82	97.81	97.79	97.78	97.81
Sum Rule	96.55	96.11	96.79	97.47	96.81
Majority Voting	98.11	98.13	98.12	98.10	97.85
Weighted Average	98.92	98.74	98.52	98.20	98.13
Proposed (MENet)	**99.54**	**99.62**	**98.61**	**99.10**	**98.88**

schemes. It is clear from the results that the performance of the weighted average ensemble, which only takes the accuracy metric into account when determining the weights, came the closest to matching the performance of the proposed ensemble technique. In the majority voting-based ensemble, the class that received the highest votes from the base learners is predicted to be the class of the sample.

Data visualization

In this section, we take help from two data visualization tools to visually show some results of the proposed method, namely, GradCAM and t-SNE plots.

GradCAM analysis. In this study, we have utilized GradCAM, a technique that creates a gradient-weighted class activation map as proposed by [70], to produce visual explanations of the model predictions. These visualizations help to demonstrate how neural networks arrive at decisions. As shown in Figs 13–15, we have used GradCAM to generate visualizations for the Malignant, Benign, and Normal lung images, respectively from the IQ-OTHNCCD dataset, using the three base models that have been used to construct the ensemble model. It is evident that the various models concentrate on distinct areas of the lung CT scans, thereby suggesting that individual learners capture unique and complementary information that is required to
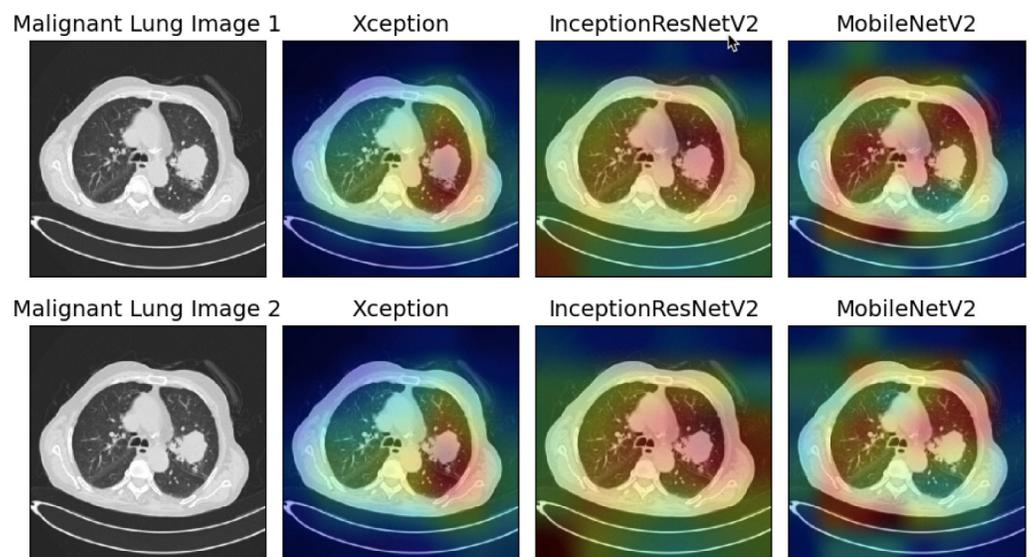
Fig 13. GradCAM visualization of the Malignant lung images taken from the IQ-OTHNCCD dataset with the three base models used to form the ensemble.
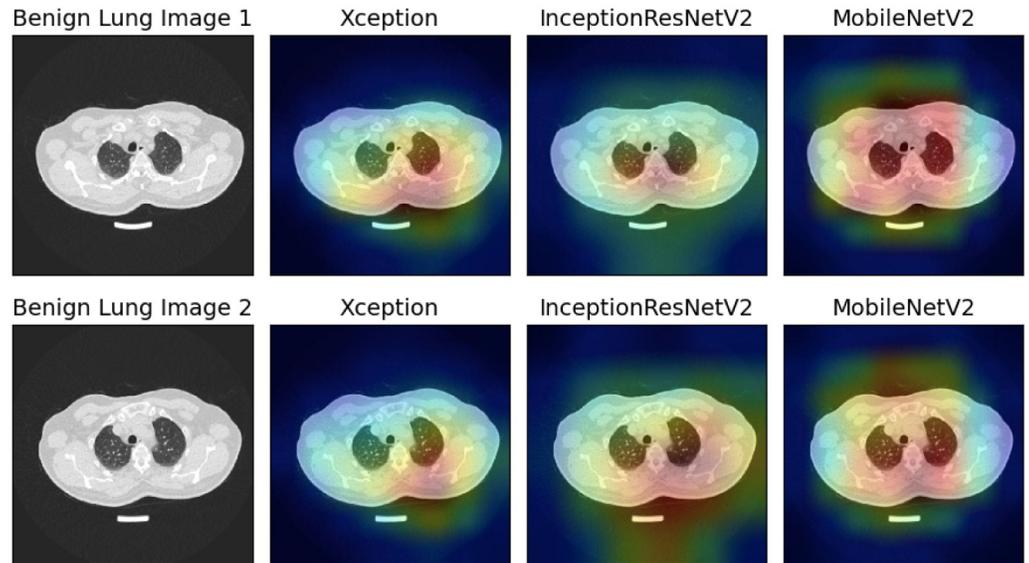
Fig 14. GradCAM visualization of the Benign lung images taken from the IQ-OTHNCCD dataset with the three base models used to form the ensemble.

form an ensemble-learning-based model as observed in Fig 13, the Xception model appears to concentrate on the right segment of the lung, the InceptionResNetV2 model on the lung nodules, and the MobileNetV2 model on the corners of the lung. Due to this, the ensemble technique proved to be effective in our present work.

t-SNE plots. t-SNE (t-Distributed Stochastic Neighbor Embedding) is a popular dimensionality reduction technique that can be used to visualize high-dimensional data in a lower-dimensional space as proposed by [71]. At first, the algorithm transforms the high-
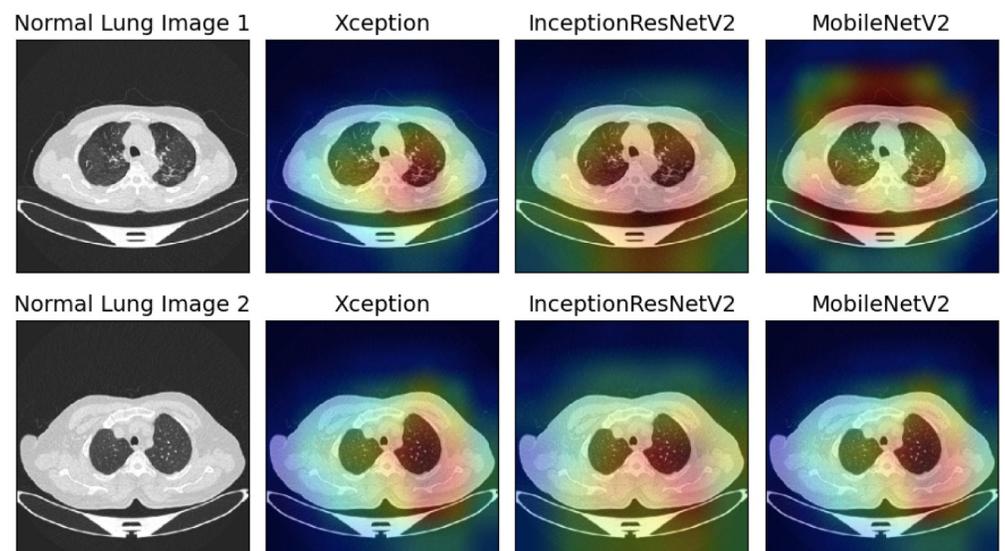
Fig 15. GradCAM visualization of the Normal lung images taken from the IQ-OTHNCCD dataset with the three base models used to form the ensemble.
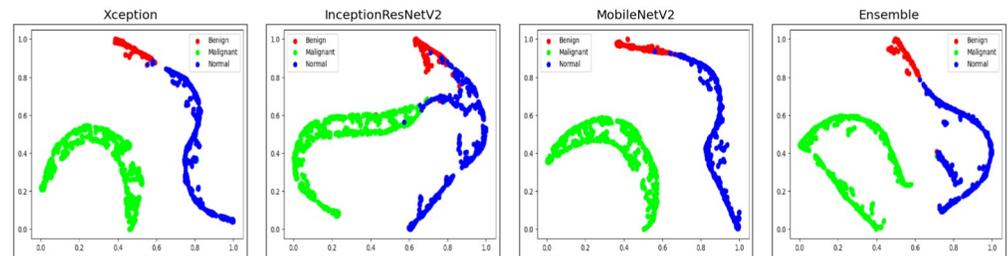
Fig 16. Graphical representation of the t-SNE plots of samples of the IQ-OTHNCCD dataset with the three base models and the final ensemble model.

https://doi.org/10.1371/journal.pone.0298527.g016

dimensional Euclidean distances among the data points into conditional probability scores that indicate similarities. This is achieved using SNE (Stochastic Neighbor Embedding) on the data points. The conditional probability $P_{j|i}$, which denotes the similarity of data point x_j to data point x_i, is defined using the Eq 7.

$$P_{i|j} = \frac{exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma_i^2}\right)} \tag{7}$$

We observe that the images corresponding to the different stages of lung cancer are clearly separated into different clusters of points. The t-SNE plot visualizations for the Benign, Malignant, and Normal classes of lung nodule CT images from the IQ-OTHNCCD dataset are presented in the first three images of Fig 16. These visualizations have been generated using the three selected base models that form the ensemble. The t-SNE plot of the ensemble model is presented in the fourth image of Fig 16.

Empirical significance of the ensemble model

In this section, we report few outlier cases identified during the model testing and analysis. There are cases where the proposed model has performed well, although one or two base classifiers have wrongly classified the sample. Figs 17 and 18 highlight the cases when for a benign image and a malignant image, respectively, two of our base models have given correct answers, and one has not, but our proposed ensemble model is still able to give the correct result. Fig 19 highlights the condition where for a normal image, only one of our base models is able to give the correct result, while the other two could not. In this case, our proposed ensemble method can also give the correct results. These empirical results ensure the effectiveness of the proposed ensemble model in varied scenarios.

Statistical analysis

We performed McNemar's statistical test, a non-parametric test, to determine the statistical significance of the results obtained by the proposed ensemble method. The ensemble result is compared against three base models' results, namely Xception, InceptionResNetV2, and MobileNetV2, on the IQ-OTHNCCD dataset used in this study. The probability scores are used to design the ensemble model. In this case, the null hypothesis is that there is no significant difference in the performance of the three base CNN models in predicting lung cancer outcomes. The results of McNemar's test are presented in Table 10, and the p-value for the dataset used in this study is found to be less than 0.05 (5%), indicating statistical evidence against the null
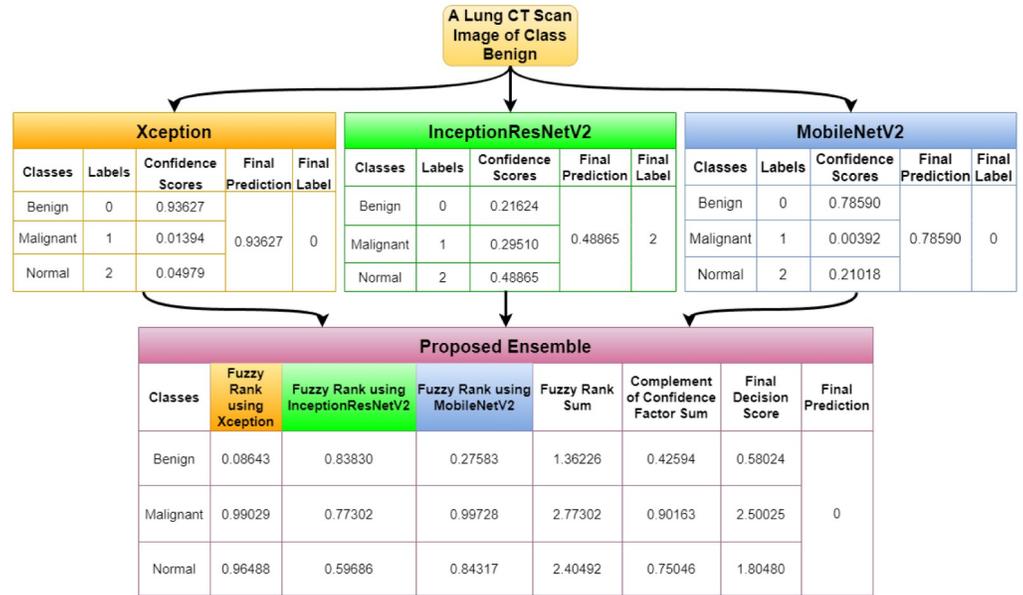
Fig 17. Demonstration for the case when two base models give the correct results, but the third one gives an erroneous result for a Benign image.

hypothesis. The statistical test confirms that the proposed ensemble model captures complementary information from the base models and performs better than individual base models. Consequently, none of the contributing base models are statistically equivalent to the ensemble classifier.

Fig 18. Demonstration for the case when two base models give the correct results, but the third one gives an erroneous result for a Malignant image.
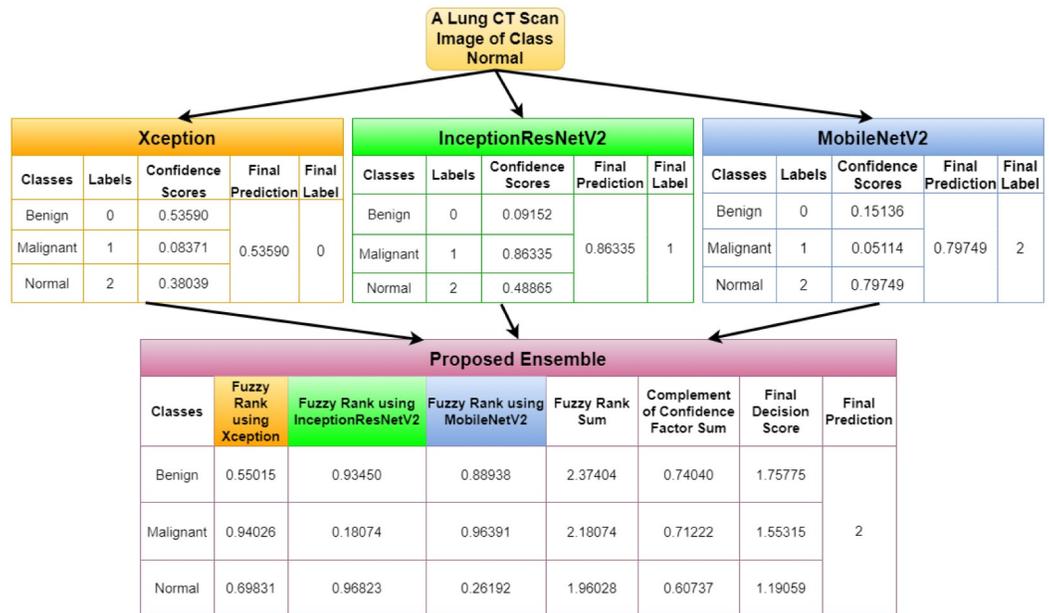
Fig 19. Demonstration for the case when one base model gives the correct results, but the remaining two give erroneous results for a Normal image.

Additional experiments

For further generalization of results, we have executed our proposed model on the LIDC-IDRI [72] dataset, which is also a popular publicly available dataset. The dataset has already been pre-split into train, test and validation sets, each containing images for two classes Benign and Malignant, both of which further contain the CT images of the respective category. Table 11 provides a detailed description of the distribution of the dataset.

Table 10. Results of McNemar's statistical test show that the null hypothesis is rejected for the base models on the IQ-OTHNCCD dataset when compared to the proposed ensemble model, as the p-value is found to be less than 0.05.

Base model	p-value
Xception	0.0014
InceptionResNetV2	0.0317
MobileNetV2	0.0005

Table 11. Distribution of images found in the LIDC-IDRI dataset into Benign and Malignant classes.

Dataset	Division	Classes	No. of Images
LIDC-IDRI	Train	Benign	4342
		Malignant	845
	Test	Benign	1340
		Malignant	282
	Validation	Benign	1073
		Malignant	224

We use class weights to address the class imbalance in the dataset. To prevent overfitting, we use early stopping while training the model. Fig 20 depicts the loss and accuracy curves of the individual base models along with their confusion matrices and ROC-AUC curves on the test dataset. We can observe from Fig 20 that the loss curves still show a slight tendency of overfitting but overall it shows that our models are able to learn properly with decreasing loss values. The accuracy curves show the high scores obtained by our model. The confusion
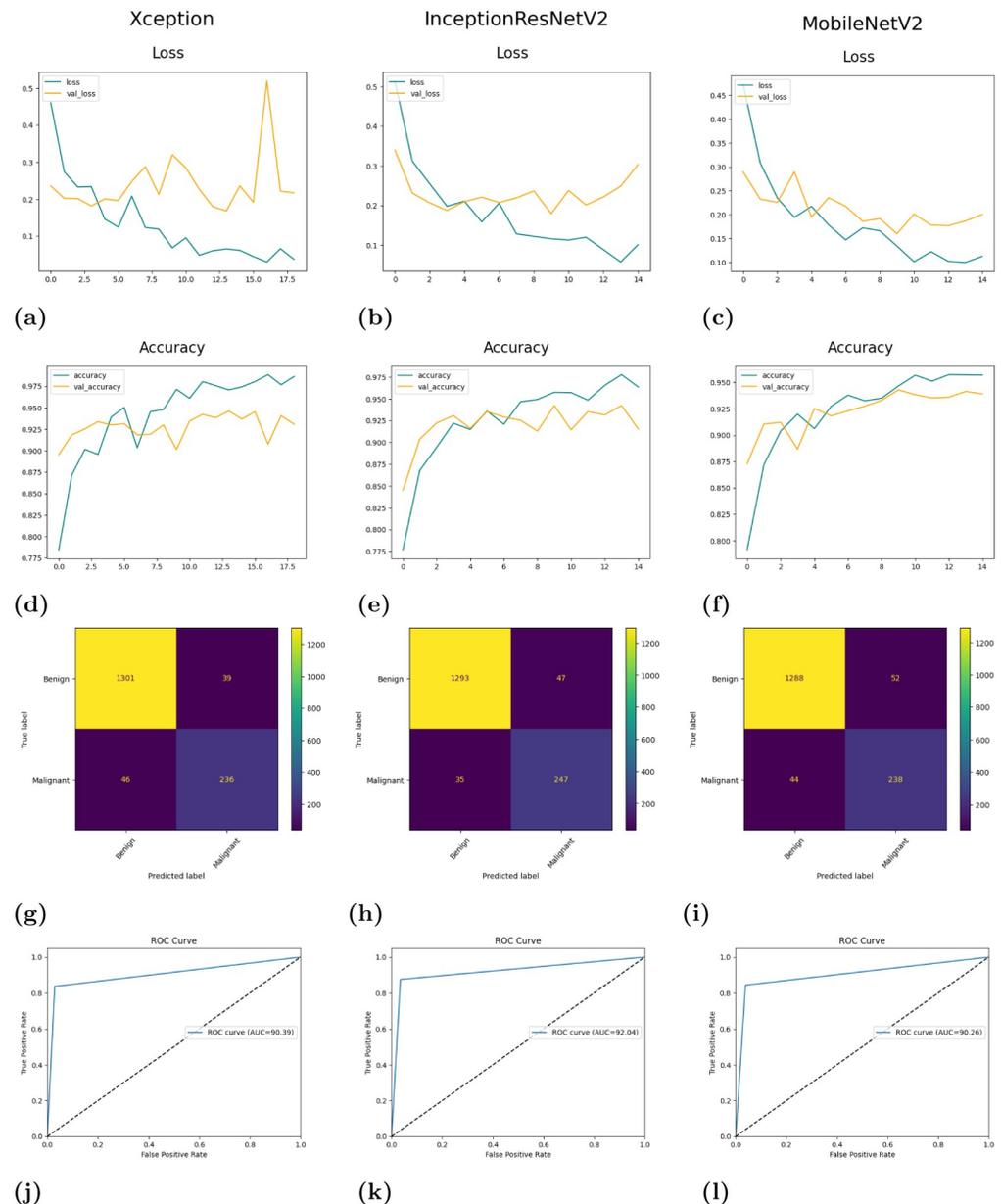
Fig 20. Results obtained from the base models on **LIDC-IDRI dataset**: (a) Loss curve of Xception, (b) Loss curve of InceptionResNetV2, (c) Loss curve of MobileNetV2, (d) Accuracy curve of Xception, (e) Accuracy curve of InceptionResNetV2, (f) Accuracy curve of MobileNetV2, (g) Confusion matrix of Xception, (h) Confusion matrix of InceptionResNetV2, (i) Confusion matrix of MobileNetV2, (j) ROC-AUC curve of Xception, (k) ROC-AUC curve of InceptionResNetV2, (l) ROC-AUC curve of MobileNetV2.
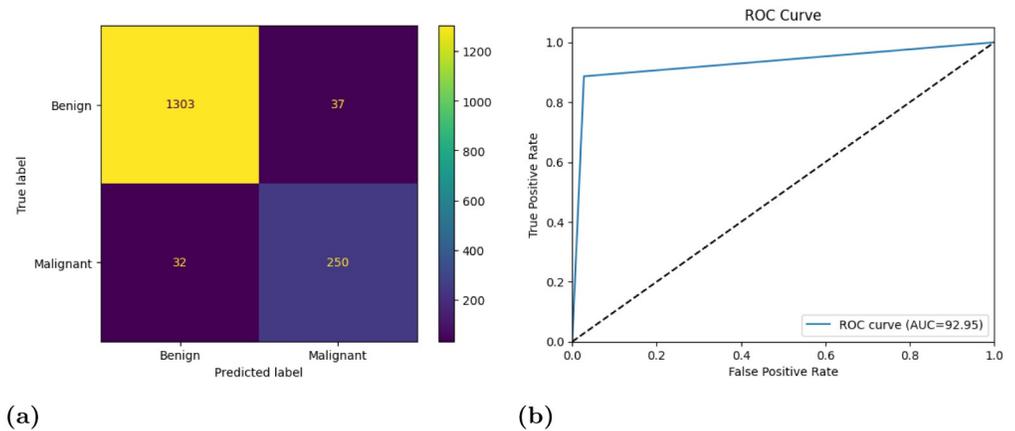
Fig 21. (a) Confusion matrix and (b) ROC-AUC curve, obtained by the proposed ensemble model (MENet) on the LIDC-IDRI dataset.

matrices highlight that although there are some wrong classifications, the ratio of wrong classifications to that of the correct classifications is very low, indicating that these models are reliable for the work. The final ROC curves obtained on the test dataset by each of the models, show that the False Positive rate is significantly low. As the ROC-AUC curves in Fig 20 is shifted towards the top left corner, it indicates the highly efficient ability of the models to classify test images into one of the two classes. Fig 21 show the final ensemble results confusion matrix and ROC-AUC curve.

Each of these base models generates the confidence scores for each class for every single image in the dataset. These confidence scores of all the base models for each image are generated and stored for each individual base models. For the results obtained in Table 12, the three transfer-learned models have been trained for 60 epochs with the Adam optimizer individually.

From Table 13, we can observe that the overall accuracy achieved is 95.75% after combining the results of the three base classifiers. The class-wise results have also been given in the Table 13. All these classification scores indicate that the model is highly reliable. Fig 22 provides a detailed pictorial representation of the performances of each base model and the final ensemble method on the LIDC-IDRI dataset.

Table 12. Performance measure of each model along with their total number of parameters while experimented on the LIDC-IDRI dataset.

Model	No. of parameters	Acc (%)	Pre (%)	Re (%)
Xception	20,867,627	94.76	91.20	90.39
InceptionResNetV2	54,341,347	94.94	90.69	92.04
MobileNetV2	2,261,827	94.08	89.38	90.26
Proposed Ensemble	**77,470,801**	**95.75**	**92.36**	**92.95**

Table 13. Class-wise and overall results obtained by the proposed ensemble model on the LIDC-IDRI dataset.

Class	Pre	Re	F1	Support	Acc	Overall Acc	Overall Re	Overall F1
Benign	0.9760	0.9724	0.9742	1340	0.9724	0.9575	0.9295	0.9265
Malignant	0.8711	0.8865	0.8787	282	0.8865			
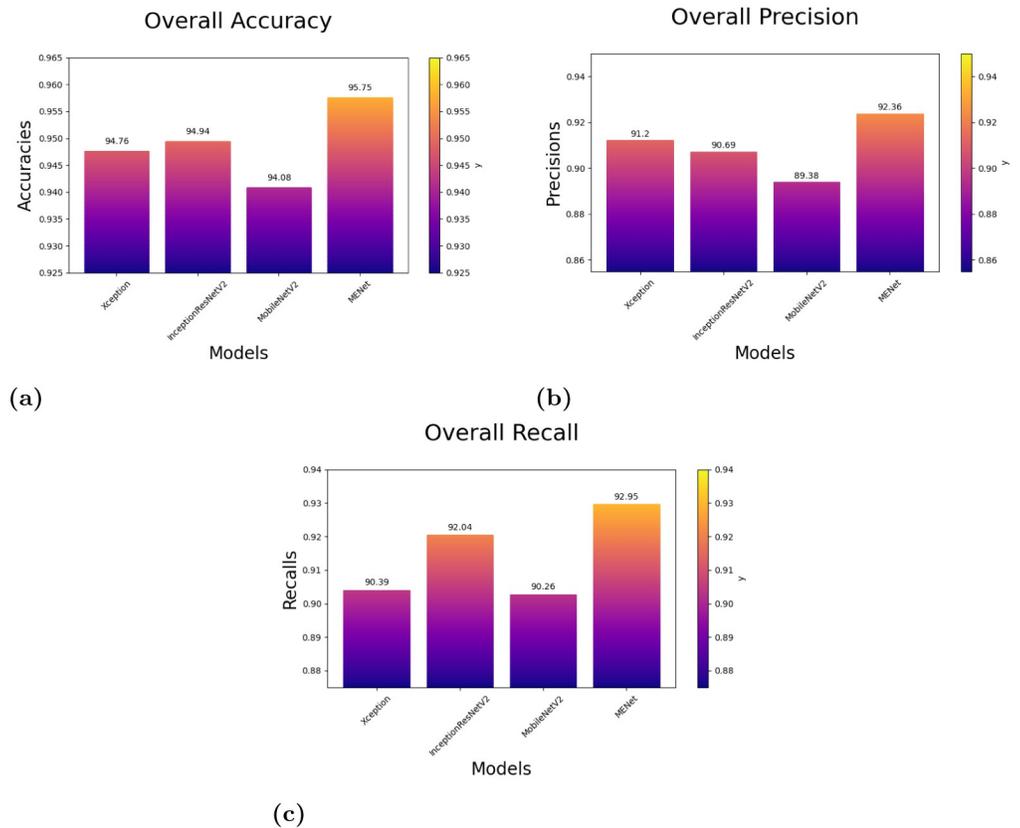
Fig 22. Results of the models depicting their overall accuracy, precision and recall on the **LIDC-IDRI dataset**: (a) Overall accuracy of the models, (b) Overall Precision of the models, (c) Overall Recall of the models.

https://doi.org/10.1371/journal.pone.0298527.g022

Comparison with past methods

The proposed method is compared with the methods found in the literature review, and it has been observed that the proposed method achieves the highest accuracy score among others. In comparison, the dataset we have used and compared with others is the same dataset (i.e., the LIDC-IDRI lung cancer dataset). Table 14 provides an illustration of the comparative study of our suggested model with others.

Table 14. Performance comparison of the proposed ensemble model with state-of-the-art methods for the LIDC-IDRI dataset.

Work Ref.	Method	Acc (%)
Zhao et al. [32]	Forward and Backward GAN and Multi-scale VGG16	95.20
Bhatia et al. [34]	Deep Residual Learning	84.00
Shaffie et al. [24]	Resolved Ambiguity Local Binary Pattern	94.90
Xie et al. [73]	Semi-supervised Adversarial Classification mode	92.50
Baoxian et al. [74]	Novel DCNN	94.00
Hanliang et al. [75]	Convolutional block attention module	90.70
Shafi et al. [28]	Deep Learning-Assisted SVM-Based Model	94.00
AR et al. [76]	LCD-CapsNet	94.00
Proposed (MENet)	**An ensemble of three CNN models**	**95.75**

https://doi.org/10.1371/journal.pone.0298527.t014

Conclusion and future scope

In recent times, it has been observed that CAD systems can help diagnostic precision and decrease the likelihood of human errors by automating the diagnosis process. In the present work, we have designed an ensemble model, called MENet, using three transfer learning-based CNN models, namely Xception, InceptionResNetV2, and MobileNetV2, to enhance the accuracy of a lung cancer prediction model. In doing so, we have used a fuzzy ranking-based approach, which considers the uncertainty of each classifier's predictions and assigns different levels of importance to each classifier. The fuzzy ranking system is designed based on the Mitscherlich function, which combines the outputs of the base classifiers to form a final prediction model that is more accurate than each classifier's individual prediction ability. The proposed method has been evaluated on the two lung CT scan datasets, namely IQ-OTHNCCD and LIDC-IDRI, and the obtained results are better than many recently proposed methods.

However, there are some false positives and false negatives, and these are significant challenges in the medical field as it would directly affect the treatment of patients. Hence, in the future, we need to reduce such errors. For this purpose, we may apply some attention mechanisms to the base CNN models that might help to generate better feature maps by focusing on the important regions, which, in turn, might produce a better prediction model. In the future, we would like to explore some lightweight CNN models to make the overall model to be useful in practical cases. Moreover, we recognize the importance of thoroughly analyzing and mitigating potential problems associated with our proposed method. Future work will involve a comprehensive examination of challenges, ranging from interpretability issues to scalability concerns. Developing strategies to address these challenges will contribute to the robustness and dependability of the MENet across diverse medical settings. Considering the evolving nature of medical datasets, the model's adaptability to new information and potential concept drift can also be assessed as future work.

Supporting information

S1 Appendix. Dry run of the entire proposed method.
(PDF)

Author Contributions

Conceptualization: Surya Majumder, Nandita Gautam.

Formal analysis: Surya Majumder, Nandita Gautam, Abhishek Basu, Arup Sau.

Funding acquisition: Zong Woo Geem.

Investigation: Ram Sarkar.

Supervision: Nandita Gautam, Zong Woo Geem, Ram Sarkar.

Validation: Abhishek Basu, Arup Sau.

Writing – original draft: Surya Majumder.

Writing – review & editing: Nandita Gautam, Zong Woo Geem, Ram Sarkar.

References

1. WHO. World Health Organisation. https://wwwwhoint/news-room/fact-sheets/detail/cancer. 2020;.

2. Sujitha R, Seenivasagam V. Classification of lung cancer stages with machine learning over big data healthcare framework. Journal of Ambient Intelligence and Humanized Computing. 2021; 12:5639–5649. https://doi.org/10.1007/s12652-020-02071-2

3. Yang Y, Xu L, Sun L, Zhang P, Farid SS. Machine learning application in personalised lung cancer recurrence and survivability prediction. Computational and Structural Biotechnology Journal. 2022; 20:1811–1820. https://doi.org/10.1016/j.csbj.2022.03.035 PMID: 35521553

4. Alsinglawi B, Alshari O, Alorjani M, Mubin O, Alnajjar F, Novoa M, et al. An explainable machine learning framework for lung cancer hospital length of stay prediction. Scientific reports. 2022; 12(1):1–10. https://doi.org/10.1038/s41598-021-04608-7 PMID: 35022512

5. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems. 2021;.

6. Zhan X, Long H, Gou F, Duan X, Kong G, Wu J. A convolutional neural network-based intelligent medical system with sensors for assistive diagnosis and decision-making in non-small cell lung cancer. Sensors. 2021; 21(23):7996. https://doi.org/10.3390/s21237996 PMID: 34884000

7. Faruqui N, Yousuf MA, Whaiduzzaman M, Azad A, Barros A, Moni MA. LungNet: A hybrid deep-CNN model for lung cancer diagnosis using CT and wearable sensor-based medical IoT data. Computers in Biology and Medicine. 2021; 139:104961. https://doi.org/10.1016/j.compbiomed.2021.104961 PMID: 34741906

8. Lakshmi D, Thanaraj KP, Arunmozhi M. Convolutional neural network in the detection of lung carcinoma using transfer learning approach. International journal of imaging systems and technology. 2020; 30 (2):445–454. https://doi.org/10.1002/ima.22394

9. Saikia T, Kumar R, Kumar D, Singh KK. An automatic lung nodule classification system based on hybrid transfer learning approach. SN Computer Science. 2022; 3(4):272. https://doi.org/10.1007/s42979-022-01167-0

10. Dasarathy BV, Sheela BV. A composite classifier system design: Concepts and methodology. Proceedings of the IEEE. 1979; 67(5):708–713. https://doi.org/10.1109/PROC.1979.11321

11. Shakeel PM, Tolba A, Al-Makhadmeh Z, Jaber MM. Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks. Neural Computing and Applications. 2020; 32:777–790. https://doi.org/10.1007/s00521-018-03972-2

12. Ayaz M, Shaukat F, Raja G. Ensemble learning based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors. Physical and Engineering Sciences in Medicine. 2021; 44 (1):183–194. https://doi.org/10.1007/s13246-020-00966-0 PMID: 33459996

13. Pramanik R, Biswas M, Sen S, de Souza LA Júnior, Papa JP, Sarkar R. A fuzzy distance-based ensemble of deep models for cervical cancer detection. Computer Methods and Programs in Biomedicine. 2022; 219:106776. https://doi.org/10.1016/j.cmpb.2022.106776 PMID: 35398621

14. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences. 1997; 55(1):119–139. https://doi.org/10.1006/jcss.1997.1504

15. He P, Deng Z, Wang H, Liu Z. Model approach to grammatical evolution: theory and case study. Soft Computing. 2016; 20:3537–3548. https://doi.org/10.1007/s00500-015-1710-9

16. Zhou Y, Kong L, Wu Z, Liu S, Cai Y, Liu Y. Ensemble of multi-objective metaheuristic algorithms for multi-objective unconstrained binary quadratic programming problem. Applied Soft Computing. 2019; 81:105485. https://doi.org/10.1016/j.asoc.2019.105485

17. Wang Z, Wang L, Huang C, Luo X. A Hybrid Ensemble Learning Model for Short-Term Solar Irradiance Forecasting Using Historical Observations and Sky Images. IEEE Transactions on Industry Applications. 2022;.

18. Mamun M, Farjana A, Al Mamun M, Ahammed MS. Lung cancer prediction model using ensemble learning techniques and a systematic review analysis. In: 2022 IEEE World AI IoT Congress (AIIoT). IEEE; 2022. p. 187–193.

19. ALzubi JA, Bharathikannan B, Tanwar S, Manikandan R, Khanna A, Thaventhiran C. Boosted neural network ensemble classification for lung cancer disease diagnosis. Applied Soft Computing. 2019; 80:579–591. https://doi.org/10.1016/j.asoc.2019.04.031

20. Berliana AU, Bustamam A. Implementation of stacking ensemble learning for classification of COVID-19 using image dataset CT scan and lung X-Ray. In: 2020 3rd International Conference on Information and Communications Technology (ICOIACT). IEEE; 2020. p. 148–152.

21. Froz BR, de Carvalho Filho AO, Silva AC, de Paiva AC, Acatauassú Nunes R, Gattass M. Lung nodule classification using artificial crawlers, directional texture and support vector machine. Expert Systems with Applications. 2017; 69:176–188. https://doi.org/10.1016/j.eswa.2016.10.039

22. de Carvalho Filho AO, de Sampaio WB, Silva AC, de Paiva AC, Nunes RA, Gattass M. Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. Artificial Intelligence in Medicine. 2014; 60(3):165–177. https://doi.org/10.1016/j.artmed.2013.11.002 PMID: 24332156

23. Woźniak M, Połap D, Capizzi G, Sciuto GL, Kośmider L, Frankiewicz K. Small lung nodules detection based on local variance analysis and probabilistic neural network. Computer Methods and Programs in Biomedicine. 2018; 161:173–180. https://doi.org/10.1016/j.cmpb.2018.04.025 PMID: 29852959

24. Shaffie A, Soliman A, Khalifeh HA, Taher F, Ghazal M, Dunlap N, et al. A Novel CT-Based Descriptors for Precise Diagnosis of Pulmonary Nodules. In: 2019 IEEE International Conference on Image Processing (ICIP); 2019. p. 1400–1404.

25. Netto SMB, Bandeira Diniz JO, Silva AC, de Paiva AC, Nunes RA, Gattass M. Modified Quality Threshold Clustering for Temporal Analysis and Classification of Lung Lesions. IEEE Transactions on Image Processing. 2019; 28(4):1813–1823. https://doi.org/10.1109/TIP.2018.2878954 PMID: 30387727

26. Xie Y, Zhang J, Xia Y, Fulham M, Zhang Y. Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. Information Fusion. 2018; 42:102–110. https://doi.org/10.1016/j.inffus.2017.10.005

27. K LS, Mohanty SN, K S, N A, Ramirez G. Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems. 2019; 92:374–382. https://doi.org/10.1016/j.future.2018.10.009

28. Shafi I, Din S, Khan A, Díez IDLT, Casanova RdJP, Pifarre KT, et al. An effective method for lung cancer diagnosis from ct scan using deep learning-based support vector network. Cancers. 2022; 14 (21):5457. https://doi.org/10.3390/cancers14215457 PMID: 36358875

29. Sahu P, Yu D, Dasari M, Hou F, Qin H. A Lightweight Multi-Section CNN for Lung Nodule Classification and Malignancy Estimation. IEEE Journal of Biomedical and Health Informatics. 2019; 23(3):960–968. https://doi.org/10.1109/JBHI.2018.2879834 PMID: 30418891

30. Masood A, Yang P, Sheng B, Li H, Li P, Qin J, et al. Cloud-Based Automated Clinical Decision Support System for Detection and Diagnosis of Lung Cancer in Chest CT. IEEE Journal of Translational Engineering in Health and Medicine. 2020; 8:1–13. https://doi.org/10.1109/JTEHM.2019.2955458 PMID: 31929952

31. Lin CH, Lin CJ, Li YC, Wang SH. Using Generative Adversarial Networks and Parameter Optimization of Convolutional Neural Networks for Lung Tumor Classification. Applied Sciences. 2021; 11(2). https://doi.org/10.3390/app11020480

32. Zhao D, Zhu D, Lu J, Luo Y, Zhang G. Synthetic Medical Images Using F&BGAN for Improved Lung Nodules Classification by Multi-Scale VGG16. Symmetry. 2018; 10:519. https://doi.org/10.3390/sym10100519

33. Yuan H, Wu Y, Cheng J, Fan Z, Zeng Z. Pulmonary Nodule Detection Using 3-D Residual U-Net Oriented Context-Guided Attention and Multi-Branch Classification Network. IEEE Access. 2022; 10:82–98. https://doi.org/10.1109/ACCESS.2021.3137317

34. Bhatia S, Sinha Y, Goel L. Lung cancer detection: a deep learning approach. In: Soft Computing for Problem Solving: SocProS 2017, Volume 2. Springer; 2019. p. 699–705.

35. Halder A, Dey D. Atrous convolution aided integrated framework for lung nodule segmentation and classification. Biomedical Signal Processing and Control. 2023; 82:104527. https://doi.org/10.1016/j.bspc.2022.104527

36. Maji D, Santara A, Mitra P, Sheet D. Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images. arXiv preprint arXiv:160304833. 2016;.

37. Kassani SH, Kassani PH, Wesolowski MJ, Schneider KA, Deters R. Classification of histopathological biopsy images using ensemble of deep learning networks. arXiv preprint arXiv:190911870. 2019;.

38. Bhowal P, Sen S, Velasquez JD, Sarkar R. Fuzzy ensemble of deep learning models using choquet fuzzy integral, coalition game and information theory for breast cancer histology classification. Expert Systems with Applications. 2022; 190:116167. https://doi.org/10.1016/j.eswa.2021.116167

39. Gupta N, Bhatele P, Khanna P. Glioma detection on brain MRIs using texture and morphological features with ensemble learning. Biomedical Signal Processing and Control. 2019; 47:115–125. https://doi.org/10.1016/j.bspc.2018.06.003

40. Dey S, Bhattacharya R, Malakar S, Schwenker F, Sarkar R. CovidConvLSTM: A fuzzy ensemble model for COVID-19 detection from chest X-rays. Expert Systems with Applications. 2022; 206:117812. https://doi.org/10.1016/j.eswa.2022.117812 PMID: 35754941

41. Pramanik R, Dey S, Malakar S, Mirjalili S, Sarkar R. TOPSIS aided ensemble of CNN models for screening COVID-19 in chest X-ray images. Scientific Reports. 2022; 12(1):15409. https://doi.org/10.1038/s41598-022-18463-7 PMID: 36104401

42. Banerjee A, Sarkar A, Roy S, Singh PK, Sarkar R. COVID-19 chest X-ray detection through blending ensemble of CNN snapshots. Biomedical Signal Processing and Control. 2022; 78:104000. https://doi.org/10.1016/j.bspc.2022.104000 PMID: 35855489

43. AlJame M, Ahmad I, Imtiaz A, Mohammed A. Ensemble learning model for diagnosing COVID-19 from routine blood tests. Informatics in Medicine Unlocked. 2020; 21:100449. https://doi.org/10.1016/j.imu.2020.100449 PMID: 33102686

44. Moon WK, Lee YW, Ke HH, Lee SH, Huang CS, Chang RF. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Computer methods and programs in biomedicine. 2020; 190:105361. https://doi.org/10.1016/j.cmpb.2020.105360 PMID: 32007839

45. Kadam VJ, Jadhav SM, Vijayakumar K. Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression. Journal of medical systems. 2019; 43 (8):263. https://doi.org/10.1007/s10916-019-1397-z PMID: 31270634

46. Luz DS, Lima TJ, Silva RR, Magalhães DM, Araujo FH. Automatic detection metastasis in breast histo-pathological images based on ensemble learning and color adjustment. Biomedical Signal Processing and Control. 2022; 75:103564. https://doi.org/10.1016/j.bspc.2022.103564

47. Gu D, Su K, Zhao H. A case-based ensemble learning system for explainable breast cancer recurrence prediction. Artificial Intelligence in Medicine. 2020; 107:101858. https://doi.org/10.1016/j.artmed.2020.101858 PMID: 32828461

48. Dey S, Mitra S, Chakraborty S, Mondal D, Nasipuri M, Das N. GC-EnC: A Copula based ensemble of CNNs for malignancy identification in breast histopathology and cytology images. Computers in Biology and Medicine. 2023; 152:106329. https://doi.org/10.1016/j.compbiomed.2022.106329 PMID: 36473342

49. Younas F, Usman M, Yan WQ. A deep ensemble learning method for colorectal polyp classification with optimized network parameters. Applied Intelligence. 2023; 53(2):2410–2433. https://doi.org/10.1007/s10489-022-03689-9

50. Brunese L, Mercaldo F, Reginelli A, Santone A. An ensemble learning approach for brain cancer detection exploiting radiomic features. Computer methods and programs in biomedicine. 2020; 185:105134. https://doi.org/10.1016/j.cmpb.2019.105134 PMID: 31675644

51. Liu Y, Long F. Acute lymphoblastic leukemia cells image analysis with deep bagging ensemble learning. In: ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging: Select Proceedings. Springer; 2019. p. 113–121.

52. An N, Ding H, Yang J, Au R, Ang TF. Deep ensemble learning for Alzheimer's disease classification. Journal of biomedical informatics. 2020; 105:103411. https://doi.org/10.1016/j.jbi.2020.103411 PMID: 32234546

53. Pramanik R, Banerjee B, Efimenko G, Kaplun D, Sarkar R. Monkeypox detection from skin lesion images using an amalgamation of CNN models aided with Beta function-based normalization scheme. PLOS ONE. 2023; 18(4):1–21. https://doi.org/10.1371/journal.pone.0281815 PMID: 37027356

54. Iraq-Oncology Teaching Hospital NCfCD. IQ-OTHNCCD Dataset. https://wwwkagglecom/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset. 2020;.

55. Alyasriy H, Muayed A. The IQ-OTHNCCD lung cancer dataset. Mendeley Data. 2020; 1(1):1–13.

56. Kareem HF, AL-Husieny MS, Mohsen FY, Khalil EA, Hassan ZS. Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset. Indonesian Journal of Electrical Engineering and Computer Science. 2021; 21(3):1731. https://doi.org/10.11591/ijeecs.v21.i3.pp1731-1738

57. Al-Yasriy HF, AL-Husieny MS, Mohsen FY, Khalil EA, Hassan ZS. Diagnosis of lung cancer based on CT scans using CNN. In: IOP Conference Series: Materials Science and Engineering. vol. 928. IOP Publishing; 2020. p. 022035.

58. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251–1258.

59. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-ResNet and the impact of residual connections on learning. arXi v: 1602.07261; 2018.

60. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottle-necks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 4510–4520.

61. Ghosal S, Sarkar M, Sarkar R. NoFED-Net: Nonlinear Fuzzy Ensemble of Deep Neural Networks for Human Activity Recognition. IEEE Internet of Things Journal. 2022; 9(18):17526–17535. https://doi.org/10.1109/JIOT.2022.3155560

62. Coulibali Z, Cambouris AN, Parent SÉ. Site-specific machine learning predictive fertilization models for potato crops in Eastern Canada. PloS one. 2020; 15(8):e0230888. https://doi.org/10.1371/journal.pone.0230888 PMID: 32764750

63. Harmsen K. A modified Mitscherlich equation for rainfed crop production in semi-arid areas: 1. Theory. NJAS: Wageningen Journal of Life Sciences. 2000; 48(3):237–250.

64. Sonar KR, Babhulkar VP. Application of Mitscherlich–Bray equation for fertilizer use in wheat. Communications in soil science and plant analysis. 2002; 33(15-18):3241–3249. https://doi.org/10.1081/CSS-120014519

65. Begum SH, Baig MI, Hussain MA, Muqeet MA. A Lightweight Deep Learning Model for Automatic Diagnosis of Lung Cancer. In: 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC). IEEE; 2022. p. 1–5.

66. Abunajm S, Elsayed N, ElSayed Z, Ozer M. Deep Learning Approach for Early Stage Lung Cancer Detection. arXiv preprint arXiv:230202456. 2023;.

67. Humayun M, Sujatha R, Almuayqil SN, Jhanjhi N. A transfer learning approach with a convolutional neural network for the classification of lung carcinoma. In: Healthcare. vol. 10. MDPI; 2022. p. 1058.

68. NARİN D, ONUR TÖ. The Effect of Hyper Parameters on the Classification of Lung Cancer Images Using Deep Learning Methods. Erzincan University Journal of Science and Technology. 2022; 15:258–268.

69. AL-Huseiny MS, Sajit AS. Transfer learning with GoogLeNet for detection of lung cancer. Indonesian Journal of Electrical Engineering and Computer Science. 2021; 22(2):1078–1086. https://doi.org/10.11591/ijeecs.v22.i2.pp1078-1086

70. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 618–626.

71. Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. 2008; 9(11).

72. Mader KS. LIDC-IDRI Dataset. https://ieee-dataportorg/documents/lung-image-database-consortium-image-collection-lidc-idri. 2011;.

73. Xie Y X Y Zhang J. Semi-supervised adversarial model for benign-malignant lung nodule classification on chest CT. In: Med Image Anal. 2019 Oct; 2019. p. 237–248. https://doi.org/10.1016/j.media.2019.07.004

74. Li B, Wang KCP, Zhang A, Yang E, Wang G. Automatic classification of pavement crack using deep convolutional neural network. International Journal of Pavement Engineering. 2020; 21(4):457–463. https://doi.org/10.1080/10298436.2018.1485917

75. Hanliang Jiang Shen F, Gao F, Han W. Learning efficient, explainable and discriminative representations for pulmonary nodules classification. Pattern Recognition. 2021; 113:107825. https://doi.org/10.1016/j.patcog.2021.107825

76. AR B, RS VK, SS K. LCD-Capsule Network for the Detection and Classification of Lung Cancer on Computed Tomography Images. Multimedia Tools and Applications. 2023; p. 1–20.