

## RESEARCH ARTICLE

# Segmentation and Classification of Interstitial Lung Diseases Based on Hybrid Deep Learning Network Model

SURENDRA REDDY VINTA<sup>1</sup>, B. LAKSHMI<sup>2</sup>, M. ARUNA SAFALI<sup>3</sup>,  
AND G. SAI CHAITANYA KUMAR<sup>4</sup>

<sup>1</sup>School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh 522237, India

<sup>2</sup>Department of Computer Applications, Velagapudi Ramakrishna Siddhartha Engineering College, Kanuru, Vijayawada, Andhra Pradesh 520007, India

<sup>3</sup>Department of Computer Science and Engineering, Dhanekula Institute of Engineering & Technology, Ganguru, Andhra Pradesh 521139, India

<sup>4</sup>Department of Computer Science and Engineering, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh 521180, India

Corresponding author: Surendra Reddy Vinta (surendrareddy.vinta@vitap.ac.in)

**ABSTRACT** Interstitial lung diseases (ILD) are diverse diseases that share pathological, radiological, and clinical traits and involve interstitial fibrosis and inflammation. These have a significant impact on lung disease morbidity and mortality. From the lung High-Resolution Computed Tomography (HRCT) image, the region of interest (ROI) had to be manually identified for most of the early ILD classification investigations, which was time-consuming. Additionally, the clinical signs of various disorders are identical, which makes precise detection difficult. In recent studies, outstanding results were achieved in categorizing medical photos using deep learning techniques. For ILD classification, a hybrid deep learning network model has been developed in this research. The lung portion of the HRCT images was initially segmented using an improved U-Net++ model. The multi-scale improved U-Net++ module has been applied for effective lung segmentation with lung anomalies. The segmented lung image's features were extracted for categorization in the second stage using a Refined Attention Pyramid Network (RAPNet). Then, we developed a MobileUNetV3 to classify five ILD classes. The ILD database is used to test the proposed approach. Due to the stage-by-stage improvement in the DL method performance, the proposed hybrid deep learning network model's performance has significantly increased.

**INDEX TERMS** Interstitial lung diseases (ILD), deep learning, improved U-Net++, refined attention pyramid network (RAPNet), MobileUNetV3.

## I. INTRODUCTION

More than 200 chronic lung tissue inflammation types are grouped as interstitial lung disease (ILD), which can severely impact the pulmonary interstitium and potentially make it difficult for the patient to breathe [1]. Consequently, early detection of these disorders is crucial for developing decisions regarding decisions [1]. Finding the precise type of interstitial lung disease (ILD) is crucial to creating effective treatment regimens since individuals with ILD risk developing lung cancer [2]. Data-driven decision-making is growing

The associate editor coordinating the review of this manuscript and approving it for publication was P. Venkata Krishna<sup>1</sup>.

in popularity in the healthcare sector because it can quickly collect and analyze complete and reliable data [3]. It compels those in charge of making decisions to select an appropriate course of action, foresee future occurrences, and formulate long-term plans [4]. Diagnosing a specific form of ILD and adopting defensive processes are all related to problems with ILD classification [5], [6].

Image processing using HRCT images is the initial method for quickly identifying normal and pathological illness instances [7]. Additionally, making decisions based on data to classify ILDs can help with early ILD detection [8], [9]. The feature extraction and ILD class labelling steps in the image-based classification approach are done to train the

classifier [10], [11]. For spatial and frequency-based image analysis, feature extraction involves effectively extracting shape, texture, and color [12], [13]. The deep learning features that deep learning algorithms suggest, however, will not be captured by these features [14].

In medical image analysis, the reliability of feature extraction has increased due to the development of deep learning algorithms [15]. These techniques are employed in the classification, segmentation, and detection fields to address a variety of applications [16]. Deep feature vectors can be obtained using deep learning algorithms like GoogLeNet, VGG, and AlexNet [17]. For training and testing, the considerable amount of data required in this architecture makes it challenging in medical research and occasionally time-consuming and tedious [18]. From a data-rich source domain, the knowledge is inherited and preserved by transfer learning and addresses the problem of data scarcity, which may lead to over-fitting [19]. To implement the machine learning methods, the appropriate feature extraction results in labelling these characteristics.

Deep learning network models can identify many combinations of pathological patterns in HRCT pictures. Still, most of them have minimal inter-class distinctions and substantial intra-class variance. Even for many seasoned experts in this field, the complexity of the diagnosis presents a challenge and may cause up to 50% ambiguity in the radiological assessment. Radiologists must also examine several instances, which is time-consuming and complicated. This research suggests a novel hybrid deep-learning network model addressing these problems.

The research's essential contribution is

- To reduce the need for ILD-infected lung tissue manually extracted from ROI, the proposed approach employs whole HRCT images.
- In the hybrid deep learning network model, the first stage involved effective lung segmentation with lung anomalies using an improved U-Net++ technique with a multi-scale feature extraction module using HRCT images. By removing the undesirable background from HRCT images, lung segmentation enables this effective network model to concentrate on the ILD properties of the lung.
- After the lung regions are segmented, the deep features are extracted using the RAPNet. The segmented lung images from various ILD classes were also used to refine the RAPNet.
- Consolidation, micronodules, ground glass, fibrosis, emphysema, and normal are among the six ILD classes that were finally classified using the MobileUNetV3 using the in-depth features from RAPNet.
- The lung segmentation is effectively performed using improved U-Net++, and the deep feature extraction is accurately performed using RAPNet, improving ILD classification performance overall.
- A recent existing deep learning network model is used to compare the performance of the suggested algorithm.

The implemented methodology is explained in Section III. Section IV presents the experimental results and discussions, and the conclusions of the present research are provided in Section V.

## II. LITERATURE PRIOR WORKS

For Segmentation and Classification, we review some recent deep-learning and machine-learning techniques for Interstitial Lung Diseases in this section.

### A. MULTI-CLASSIFICATION OF LUNG CANCER USING CT IMAGES

To enhance ILD classification performance, Small kernel DenseNet (SK-DenseNet), an upgraded version of DenseNet, was presented by Guo et al. [20]. The high-level pathological features are extracted using the SKDenseNet network based on the HRCT feature's characteristics for ILD classification. Micronodules, fibrosis, ground glass, emphysema, and healthy tissue are the primary patterns that classify ILDs. The research showed that when feature patterns are small, utilizing a small convolution kernel helps to increase recognition effectiveness.

Sukanya et al. [21] developed an effective CNN model for ILD classification using HRCT images. The medical image classification is performed by using CNN because, at the same time, it performs both feature extraction and classification, producing excellent results. The CNN has three convolutional layers: a dense layer, a maximum pooling layer, and a leaky ReLU activation layer. Five outputs in the final Fully Connected (FC) layer correspond to the considered classes: Fibrosis, Micro Nodules, Normal, Emphysema, and Ground Glass (GG).

### B. DETECTION AND CLASSIFICATION OF LUNG NODULES IN CT IMAGES

Khan et al. [22] used nodule mining with VGG-SegNet support to facilitate automated lung nodule detection. The developed deep features are used to classify lung CT images, and the handcrafted features are successively concatenated with these features, including Pyramid Histogram of Oriented Gradients (PHOG), and Grey Level Co-Occurrence Matrix (GLCM) for improving the disease detection accuracy.

The LdcNet was proposed by Tran et al. [23] for long nodule detection. This approach solves the difficulty of classifying pulmonary nodule candidates as nodule or non-nodule in CT images. In this research, three convolutional blocks are used for feature extraction, and Classification is done using fully connected layers.

Khehrah et al. [24] established a fully automatic approach to detect nodules from lung CT scans. From the foundation, the lung region is automatically separated by generating a grayscale histogram of the CT image. Morphological operators are used to refine the results. The parenchyma's internal structures are then removed. A threshold-based technique is provided to separate possible nodules from other structures, such as blood vessels and bronchioles. These nodule

TABLE 1. Comparison details of related prior works on lung cancer classification using CT image datasets.

<b>Reference</b>	Wang et al. [25]	Khehrah et al. [24]	Tran et al. [23]	Khan et al. [22]	Sukanya et al. [21]	Guo et al. [20]
<b>Year</b>	2020	2020	2019	2021	2020	2019
<b>Method</b>	CNN-LSTM	SVM	LdcNet	VGG-SegNet SVM-RBF	CNN	SK-DenseNet
<b>Results</b>	Accuracy = 97.1%, Sensitivity = 92.23% Specificity = 91.17%	Sensitivity = 93.75% Accuracy = 91%	Accuracy = 97.2%, Sensitivity = 96.0% Specificity = 97.3%	Accuracy=97.83% Precision=98.38% Sensitivity=97.20% Specificity=98.40%	F-Score = 94.65%, Accuracy = 94.67%	Accuracy=98.4%
<b>Merits</b>	It is appropriate for keeping significant information during detection and classification since it can capture fine and coarse picture features.	It improves the efficiency and effectiveness of the classification process.	The final lung nodule classification result is directly obtained without any additional parameter adjustment or settings when inputting original signals after the training process	(i) Accurate and quick diagnosis is performed using this approach. (ii) With excellent accuracy and precision, it produces reliable results.	(i) It is capable of capturing inter-scale variations (ii) It lowers the cost of computing.	(i) It prevents over-fitting and increases classification accuracy. (ii) It has a good generalization ability for multi-class classification
<b>Demerits</b>	(i) It requires large amounts of labelled data for training (ii) It is sensitivity to noise	(i) It is computationally intensive (ii) It is not suitable for large-scale image datasets or real-time applications	(i) It requires large amounts of labelled data for training (ii) It requires more time for model training and fitting	(i) High processing power needed to train on high-resolution data (ii) Ineffective when there are a large number of features	(i) It reduces the quality of the classifier in real-time applications. (ii) It achieves less classification accuracy than other models	(i) Hyperparameter tuning is required. (ii) When handling extremely large datasets, it is computationally costly.

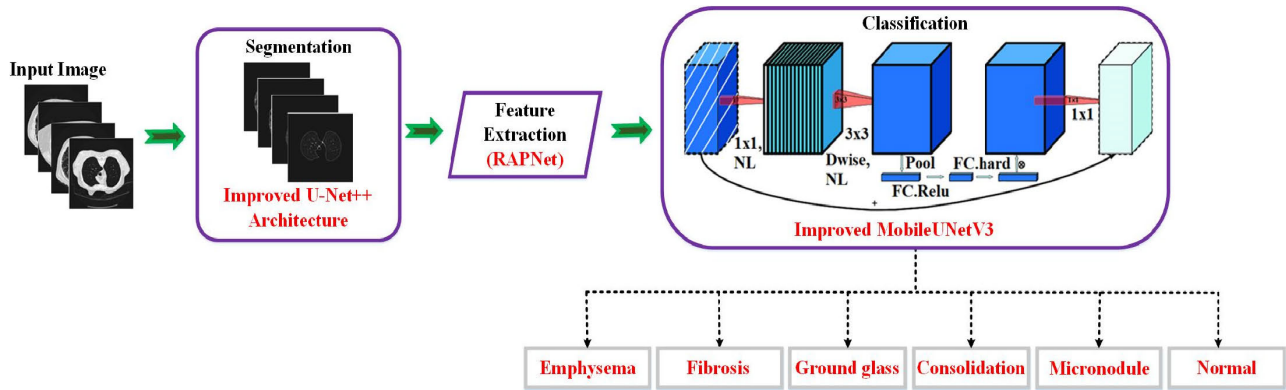


FIGURE 1. Schematic diagram of the proposed approach.

candidates' various statistical and shape-based properties are extracted to create nodule feature vectors, which are then categorized using SVMs.

A combined deep-learning algorithm was proposed by Wang et al. [25]. First, image clipping, normalization, and segmentation are used to preprocess CT images, and the range of negative and positive samples is balanced by increasing the positive samples. The characteristics of candidate lung nodule samples are learned using CNNs and residual networks, which are then imported into extended short-term memory networks accordingly. The network parameters are then continuously optimized during the training phase to produce the model with the best performance.

The summary of related prior works on lung cancer classification using different CT image lung cancer datasets. Better results are obtained by the majority of current research on ILD classification and segmentation; however, there are certain challenges, such as (i) The model may crash if the memory load becomes too high since model training requires a large number of images; hence a large amount of computer memory is needed, (ii) take longer to train and involve a more complicated method, and (iii) The quality of the classified image is not good. To address these issues, we present a novel, efficient hybrid deep learning network model in this study.

### III. PROPOSED METHODOLOGY

To screen for probable disease, the majority of the ILD classifiers now in use manually identify regions of interest (ROI). Two deep learning network models used for ILD classification are connected by the hybrid deep learning network model described in this research. The proposed method's schematic diagram is displayed in Figure 1. It accepts input from all HRCT images and outputs the ILD class label. From the HRCT images, the segmentation is done in the first stage using an improved U-Net++ model for segmenting the lung region. From the first stage's segmented lung HRCT images, another deep learning-based RAPNet model was used to extract the features. Additionally, in the second stage, ILDs have been classified using the features produced by the

RAPNet model using the memory-efficient network model MobileUNetV3.

The hybrid deep learning network model employed in the proposed strategy has been briefly described in the following sentences.

#### A. SEGMENTATION

The well-known U-Net architecture, which is frequently employed in medical image segmentation tasks, is extended by U-Net++. To enhance the original U-Net architecture, U-Net++ adds several dense, nested skip links between the encoder and decoder blocks. Better feature reuse is made possible by these skip connections, which also aid in addressing the issue of vanishing gradients that can arise in deep networks. The defective area of ILD is precisely segmented in the present research using the Improved U-Net++ model.

The modified version of U-Net++ is called Improved U-Net++. With a custom loss function, this research builds a custom U-Net++ model for the segmentation problem. Because of the dataset's small size, the class imbalance is addressed by the specially designed loss function. The traditional U-Net++ model does not solve the class imbalance problem; consequently, to address the class imbalance issue, we have added a custom loss function to enhance the structure of U-Net++ during segmentation. U-Net++ enhances U-Net by utilizing DenseNet's dense block concepts. Compared to the original U-Net, U-Net++ includes Deep supervision, dense skip connections, and redesigned skip pathways.

Redesigned skip pathways: A new skip connection architecture makes up the U-Net++. These redesigned skip pathways connect the semantic gap between the decoder and encoder. Skip connection reduces the feature map's semantic gap between the encoder and the decoder. Semantically disparate feature maps are fused in the U-Net encoder and decoder due to the direct feature map connection. Combining the previous layers' convolutional layers with the output of the matching up-sampled low, dense block is what U-Net++ does. This aids in quickly optimizing the encoder's feature

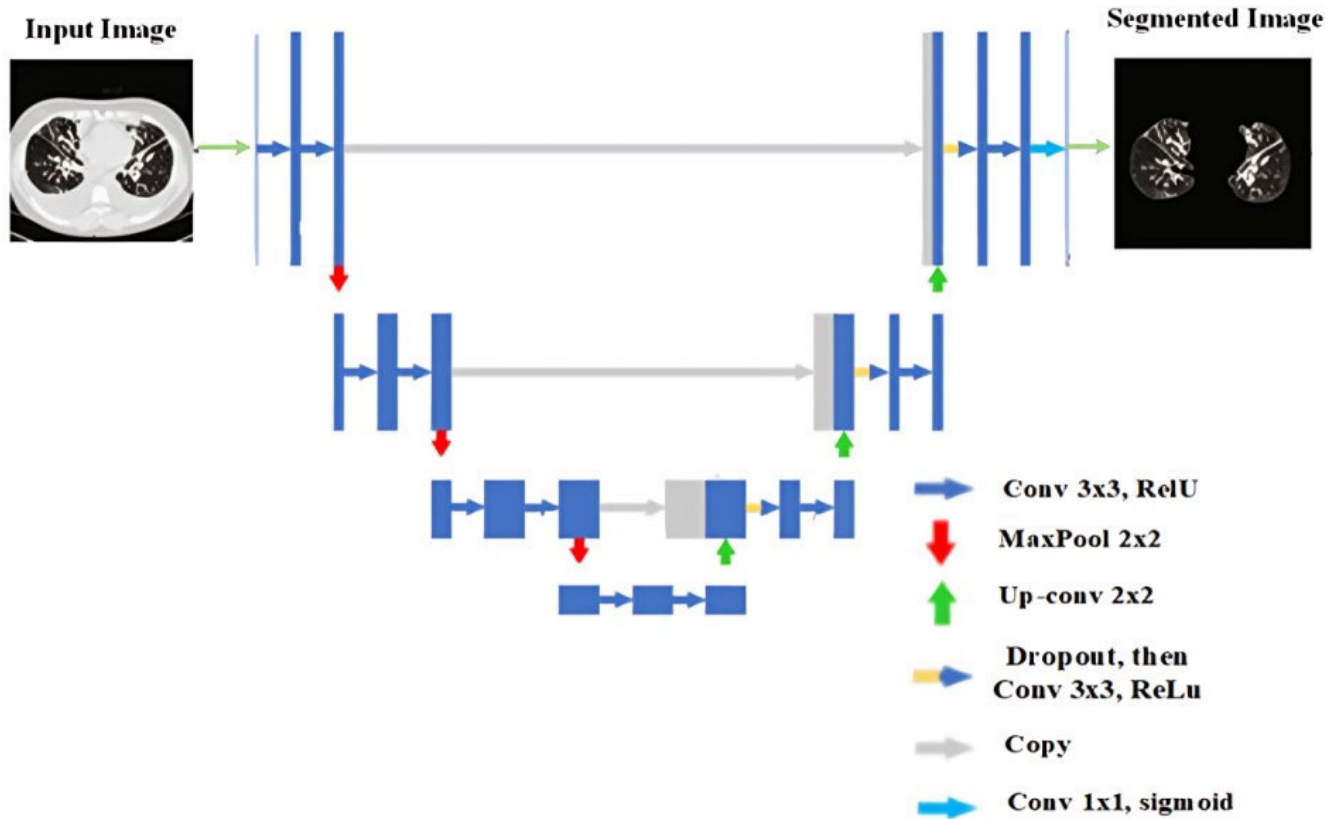


FIGURE 2. Improved U-Net++ architecture.

maps by bringing them closer to feature maps that are waiting in the matching decoder.

Dense skip connections: The U-Net ++ contains dense skip connections modelled after DenseNet to implement skip paths among the encoder and decoder. This enhances the gradient flow and segmentation accuracy. A deep skip connection also aggregates previous feature maps and delivers them to the appropriate node.

Deep supervision: The architecture’s performance and speed are balanced by adjusting the model complexity with the help of deep supervision in U-Net ++. The output from each segmentation branch must be averaged for a CNN model to be correct. Figure 2 displays the proposed segmentation network model.

When a tiny portion of the image’s total pixels are the pixels that need to be separated, class imbalance is evident. Utilizing dice coefficient and cross-entropy, a custom mixed loss function is used in the proposed segmentation network to solve the problem. (We use the flexibility provided by binary cross entropy and dice loss to smooth the gradient curve during training.)

$$L_{mixed} = \alpha * L_{BCE} (y, y') - L_{Dice} (y, y') \quad (1)$$

where.

$$L_{BCE} (y, y') = (y \log (y') + (1 - y) \log (1 - y')) \quad (2)$$

$$L_{dice} = \frac{2 * y * y' + 1}{y + y' + 1} \quad (3)$$

Here, the inconsistencies between samples are balanced using a constant  $\alpha$ , the correct value is depicted as  $y$  and the prediction value as  $y'$ . We set  $\alpha$  to 0.5 for segmentation to observe the best results. At a dropout rate of 0.1, the model worked as well as possible. Our improved U-Net++ model significantly enhances the loss function by boosting segmentation performance.

### B. FEATURE EXTRACTION

For deep feature extraction, the stage 1 segmented lung image was processed by RAPNet. The network extracts various deep features relating to classification at each layer. To prevent gradient vanishing and enable the training of considerably deeper networks than those previously employed, RAPNet is used. Using refined attention pyramid networks (RAPNets), we created an effective multi-scale building extraction technique in this research. RAP-Net mainly focuses on specific regions of interest at different scales, improving the discriminative power of features. Refined attention can help the model better understand the semantic content of an image by emphasizing important features and suppressing irrelevant ones. We constructed a pyramid pooling module, deformable convolution, atrous convolution, combined attention mechanism, and encoder-decoder structure to enhance the effectiveness



of feature extraction. The convolutional block attention module is also used to extract salient multi-scale features. The multi-scale features are fused by adopting the refined feature pyramid structure to obtain the final extraction results in the decoding path. For feature extraction, attention mechanism (AM), pyramid pooling module (PPM), deformable convolution (DC), and atrous convolution (AC) are included in the feature extraction network model.

In the feature extraction network, the encoder-decoder structure is the core network. RAPNet comprises a decoding path, encoding path, and lateral connections with CBAM. ResNet101, a cutting-edge feature extraction architecture, serves as the framework for the encoding path. As per the output feature map's dimensions, ResNet101's convolutional layers are divided into five stages, designated *conv1*, *conv2*, *conv3*, *conv4*, and *conv5*, and the final outputs of the final residual block in *conv2*, *conv3*, *conv4*, and *conv5* are represented by  $C_2$ ,  $C_3$ ,  $C_4$ ,  $C_5$ . In the decoding path, the feature map's channel dimension is decreased using  $1 \times 1$  convolutional layers.

The big feature size is maintained, and the receptive field is increased by applying the AC with rates of 2 and 4 to each stage's three  $3 \times 3$  convolutional layers (*conv4* and *conv5*, respectively). Because of this, feature maps in *conv4* and *conv5* provide output that is 1/8 of the input image rather than 1/16 and 1/8 of the input image rather than 1/32, respectively. The loss of spatial information can be reduced by feature fusion, which is more advantageous with higher feature sizes.

This DC fuses the neighboring similar pixel's structural information, and buildings with various shapes are extracted more effectively by adding this DC to residual blocks. Between feature channels, the AC is then implemented in the residual block's stages *conv2*, *conv3*, *conv4*, and *conv5* to discover the distinctions and provide important characteristics with a lot of weight, which can boost prominent features and cut out background noise. It can improve conspicuous features and remove background noise.

In the proposed model, the feature maps are refined layer by layer by embedding the CBAM, unlike the FPN structure. The CBAM processed the feature maps  $C_2$ ,  $C_3$ ,  $C_4$ ,  $C_5$  are indicated as  $M_2$ ,  $M_3$ ,  $M_4$ ,  $M_5$ . As a result, several features with a wealth of geographical and semantic data are processed to provide significant features that can be fused.

The context information is gathered by applying the PPM to the *conv5* stage's last residual block and a global average pooling operation.  $C'_5$  refers to the feature map  $C_5$  that PPM processed.

Then,  $M_2$ ,  $M_3$ ,  $M_4$ ,  $M_5$  the refined feature maps from the appropriate levels are concatenated in the decoding network, and the lateral connection's fused feature maps are indicated  $\{P_2, P_3, P_4, P_5\}$ . By concatenating  $P_3$ ,  $P_4$ , and a dense connection, we also improve the feature pyramid structure and create the fusion feature map  $P_2$ . With rich semantic and spatial data, this method allows for the acquisition of multilayer fused feature maps for multi-scale building extraction. Rather than element-wise addition in FPN, concatenation operation

has been employed for dense and lateral connections between up-sampled feature maps.

The final fused feature map's up-sampling process produces the results of the building extraction. It is identical in spatial size to the input image's original version.

### C. CLASSIFICATION

The MobileUNetV3 classifier is a viable choice because it produces high classification performance, as the proposed approach used multiple deep features retrieved from RAPNet. From RAPNet, the MobileUNetV3 classifier receives the deep features obtained for ILD classification. The proposed classification MobileUNetV3 model effectively classifies the input images as Consolidation, Micronodules, Normal, Ground glass, Fibrosis, and Emphysema.

To benefit from the strong feature extraction capabilities of MobileNetV3, the suggested system integrates the MobileNetV3 architecture with the UNet model for multiple ILD classifications, hence the name MobileUNetV3. The main contribution of our study effort here concerns the optimal way to merge these two models. MobileUNetV3 architectures aim to strike a balance between model efficiency and classification accuracy. MobileUNetV3 has a smaller model size than larger, more complex architectures.

Figure 3 depicts the architecture of the proposed MobileUNetV3 model, together with the various building elements and the input feature maps. The image size is reduced by applying the down-sampling operation to the selected MobileUNetV3 layers on the encoder part. The decoder part uses transposed convolution and up-sampling for each image classification.

The model's input image is an MRI slice with a dimension of  $244 \times 244$ . The UNet model's encoder structure sends the input image to MobileNetV3. With 64 bands, the image size to  $112 \times 112$  is changed by layer 16, for example.  $56 \times 56$  with 64 bands is the new image size introduced by layer 20. The image size is changed by layer 38 to  $28 \times 28$  with 78. Following that, each layer of MobileNetV3 is concatenated with the preceding output layer and up-sampled using the UNet decoder. The output of Layer 93 is concatenated with Layer 214, which performed the first up-sampling process. The output is then obtained using a de-convolution layer, a Softmax activation function, and a transposed convolution layer.

The proposed MobileUNetV3 model efficiently classifies the input images into multiple lung disease classifications with high accuracy. There is also a list of all the parameters, both trainable and untrainable. When working with tiny datasets, 8,045,347 parameters are present in the classification model, which is simple to train and less likely to overfit.

By categorizing each pixel according to the class that has the highest output probability, the channels are localized and distinguished in the proposed architecture. Based on cross-entropy (CE) loss, the MobileUNetV3 model is effectively trained by optimizing the loss function and it is denoted by

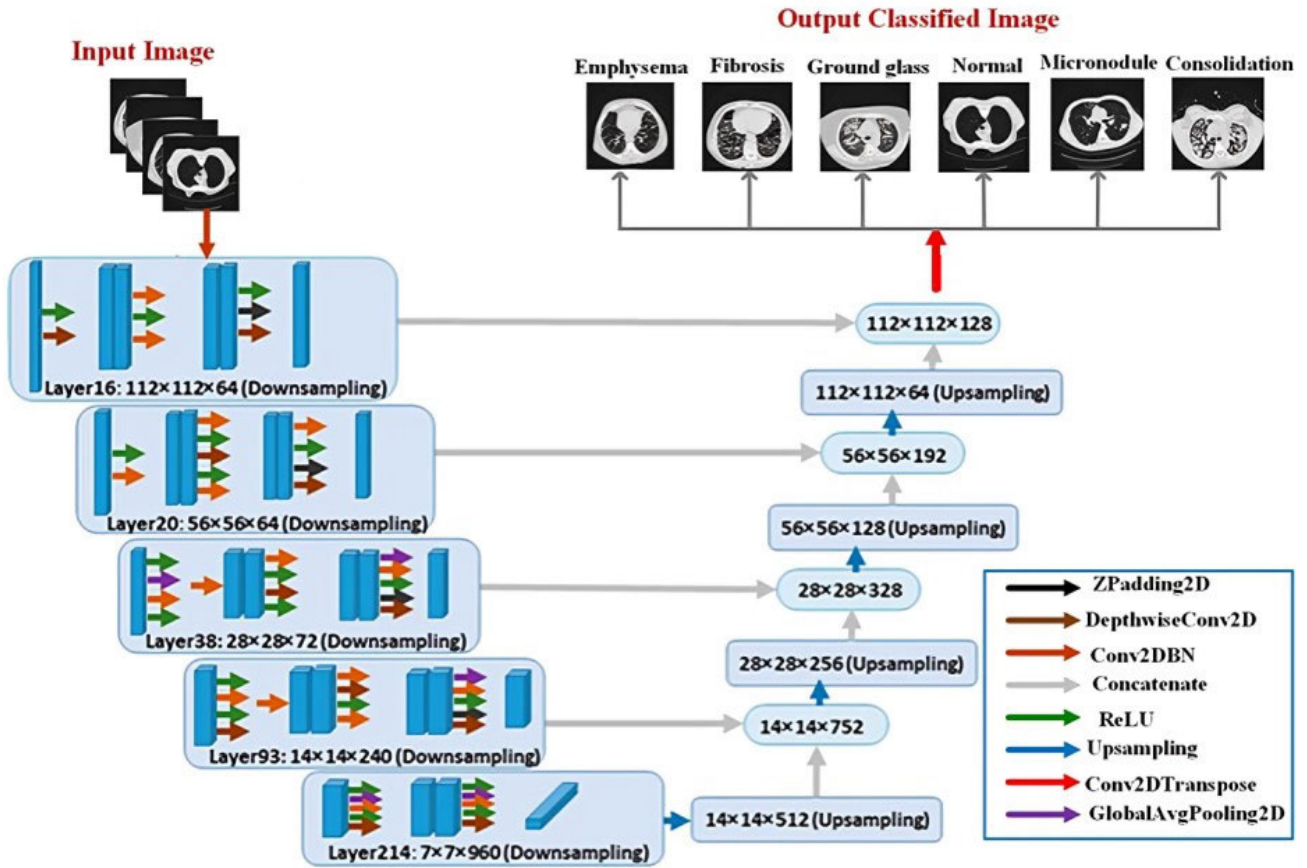


FIGURE 3. The proposed MobileUNetV3 model architecture.

$L_{ce}$ . The training set's  $L_{ce}$  loss is calculated as follows:

$$L_{ce} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^C 1(y_i = k) \ln [p(y = k|x)], \quad (4)$$

The number of classes is depicted as C, and the number of training samples is represented as N. Also included is  $p(y = k|x)$  the element x's class k prediction probability. A value of  $y_i$  is the true label of sample x.

UNet++'s parameter number can change depending on depth, width, and other architectural decisions. However, because of its complex design and extensive skip connections, it usually has a lot of parameters. For instance, depending on how it is configured, a normal UNet++ model may include millions of parameters, ranging from tens of millions to over a hundred million.

VGG network models are well-known for their uniform architecture and simplicity, particularly the deeper variations like VGG16 and VGG19. There are about 138 million parameters in VGG16 and 144 million in VGG19. Because of their depth and completely connected layers, these models are rather massive and memory-intensive.

Because of their success in deep learning tasks, ResNet models—in particular, ResNet50, ResNet101, and ResNet152—have become increasingly well-known. Over 25.6 million parameters make up ResNet50, 44.6 million

parameters make up ResNet101, and over 60 million parameters make up ResNet152. ResNet models are notable for their effective training and memory utilization because they use residual connections despite being deeper than VGG models.

In our proposed research, we have used the MobileUNetV3 model for Interstitial Lung Diseases classification, it is a lightweight and efficient model for classification. Compared to the other models listed above, it usually has a substantially smaller number of parameters. The number of parameters of the classification models is 8,045,347 parameters. However, MobileUNetV3 models are generally designed to have fewer parameters while maintaining competitive performance in image classification tasks.

#### IV. RESULTS AND DISCUSSION

This section compares the suggested technique to currently employed state-of-the-art methods by examining the various performance measures used to assess the method. The impact of cost-sensitive learning on classification performance is demonstrated through experimental data.

The Multi-class ILD classification structures, including Micronodule, Consolidation, Emphysema, Ground glass, and Fibrosis, are effectively classified and identified extensively and systematically in this research. The IRCT images dataset

**TABLE 2.** Database description.

ILD patterns	Total number of ROIs	Total number of an image slice
Ground glass	415	241
Normal	98	69
Fibrosis	479	293
Emphysema	108	71
Consolidation	194	116
Micronodule	277	154
Total	1571	944

was classified and segmented with hybrid deep learning models. The MobileUNetV3 model categorizes disorders that fall under multiple ILD classes. The defective region of ILD images is segmented by developing an Improved U-Net++ model, with 30% of the dataset used for validation and 70% for training. Results were obtained using a PC with an i5 processor with the Python platform. The performance of deep learning networks for classification is measured using various techniques.

#### A. DATABASE DESCRIPTION

The present research makes use of the publicly accessible MedGIFT database. A complete set of 1946 ROIs was offered from 108 HRCT imaging series. 17 different ILD patterns are comprised in this research, each measuring  $512 \times 512$ . The five most common health and ILD pattern combinations are considered in the present research. Since the database is multi-pattern, there is a chance that multiple patterns will appear in a single slice. Table 2 includes statistical information on the trend, as mentioned earlier. The MedGIFT database is used to obtain the total number of picture slices for each type of pattern.

#### B. PERFORMANCE METRICS

F-score, precision, accuracy, and recall are performance metrics used to assess classification performance.

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$Precision = \frac{Tp}{TP + FP} \quad (6)$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \quad (7)$$

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

For ILD classification, the number of true positives is represented as TP. Similarly, the true negative is depicted as TN, the false positive as FP, and the false negative as FN.

#### C. PARAMETER SETTINGS

In this research, segmentation and classification stages can be trained separately in a sequential fashion. This method involves first training the segmentation network (improved UNet++) to separate the input images into relevant lung

**TABLE 3.** The optimized hyperparameters of the classification network.

Hyperparameters	Range
Initial learning rate	0.001
Mini-batch size	64
Maximum number of epochs	100
Maximum number of iterations	400
Gradient decay factor	0.9000
L2 regularization	$1.0000e^{-0.4}$

regions. Following the training of the segmentation network, the retrieved lung regions are input into a different classification network (MobileUNetV3), which uses the segmented regions to carry out classification tasks.

For training and testing purposes, the data is divided into 70% training and 30% testing. We presented a hybrid deep learning-based approach in this research paper for the segmentation and classification of interstitial lung diseases. Because Adam Optimizer is one of the best SGD algorithms and combines the best features of Ada-Grad and RMS-Prop, it is used to fine-tune the parameters in each network step during training. During training, it may handle noise and sparse gradients. Optimizing the proposed hybrid network's hyperparameters yields the best performance outcomes. In Table 3, the optimized parameter settings are provided.

#### D. EXPERIMENTAL RESULTS

A generated ILD database has been used to show the performance analysis of the proposed approach. As a result of discussions with computer scientists, research physicians, and radiologists throughout the project's four years, annotated lung HRCT slices are now available in this database. For the analysis, consolidation, micronodules, fibrosis, ground glass, emphysema, and normal pictures from the six classes considered in the study were used. Utilizing the data augmentation method, around 4000 HRCT slices were used to train the network. The training dataset size is increased using data augmentation methods, including flip up-down, flip right, and flip left.

In the second stage, deep features were extracted using improved U-Net++ mapping to the six ILD classes on segmented images labelled normal, consolidation, micronodules, ground glass, fibrosis, and emphysema. The results of the proposed approach's experiment are shown in Table 4. To reduce the over-fitting on training for image recognition, augmentation is performed with more images. Also, an effort has been made to protect the enhanced images' labels. The  $224 \times 224 \times 3$  CT scans that were originally  $512 \times 512 \times 3$  have been reduced throughout the augmentation procedure. RAPNet has received these images for feature extraction. When the feature layer selection is made correctly, the classifier's performance will improve at this stage. Earlier network layers' weights were frozen because they weren't updated throughout the fine-tuning phase. Table 5 contains the proposed classifier's confusion matrix.



TABLE 4. Experimental results of the proposed approach.

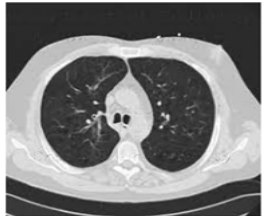
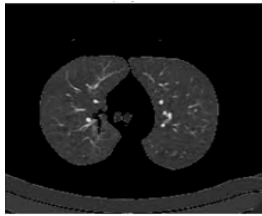

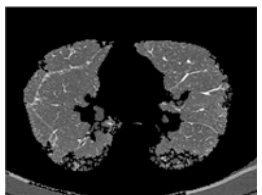

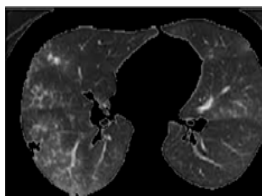

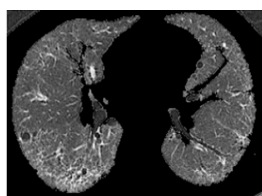
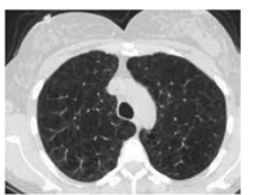
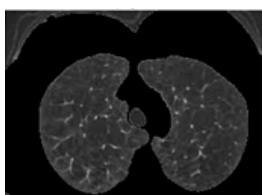

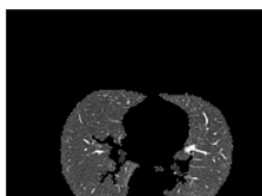

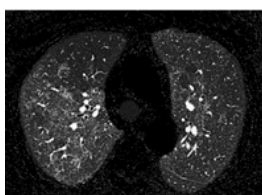
Input Image	Segmented Image	Classification
		Emphysema
		Fibrosis
		Ground glass
		Fibrosis
		Emphysema
		Normal
		Ground glass

TABLE 4. (Continued.) Experimental results of the proposed approach.


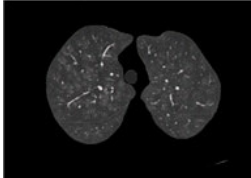

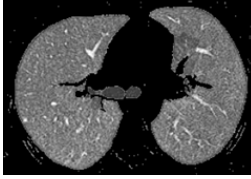

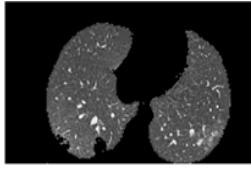
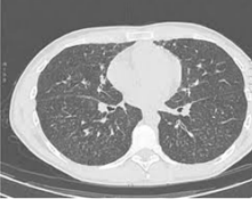
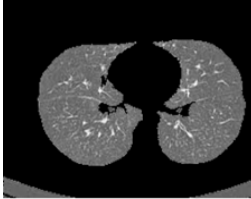
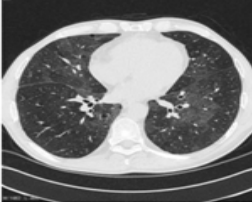
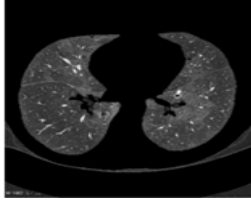
		Normal
		Micro Nodules
		Consolidation
		Micro Nodules
		Consolidation

TABLE 5. Proposed model's confusion matrix.

Actual cases	Consolidation	Micronodules	Normal	Ground glass	Fibrosis	Emphysema
Consolidation	84.12 ± 129	1.07 ± 0.43	0.21 ± 0.19	4.67 ± 0.67	9.84 ± 1.21	0.10 ± 0.08
Emphysema	0.46 ± 0.46	0.37 ± 0.34	0.00	0.37 ± 0.34	5.56 ± 1.36	93.24 ± 1.66
Fibrosis	4.76 ± 0.77	0.77 ± 0.46	0.22 ± 0.14	4.43 ± 0.91	89:26 ± 1:54	0.55 ± 0.23
Ground glass	3.08 ± 0.66	2.23 ± 0.84	4.70 ± 0.87	84.44 ± 2.17	5:27 ± 1:37	0.28 ± 0:26
Normal	0.02 ± 0.05	2.55 ± 0.70	94.65 ± 1.59	2.57 ± 1.14	0.16 ± 0.13	0.05 ± 0.06
Micronodules	1.06 ± 0.41	90.64 ± 1.42	3.45 ± 0.85	4.56 ± 0.73	0.26 ± 0.19	0.02 ± 0.04

The statistical consistency of the proposed approach is analyzed by randomly selecting 30 different cycles of HRCT images. Table 5 shows a 95% confidence interval (CI) for the network model's 30-cycle confusion matrices. This research shows the mean values and variances considered for the six ILD classes and the 95% confidence interval shown in this table.

The interactions between the six ILD classes are shown in Table 6. F-score, recall, and precision values are 98.13%, 98.05%, and 98.65%, respectively. The proposed network model's classification performance has been much enhanced for multiple ILD classifications. For the training and validation datasets, the accuracy and cross-entropy loss performance are shown in Figures 4 and 5.

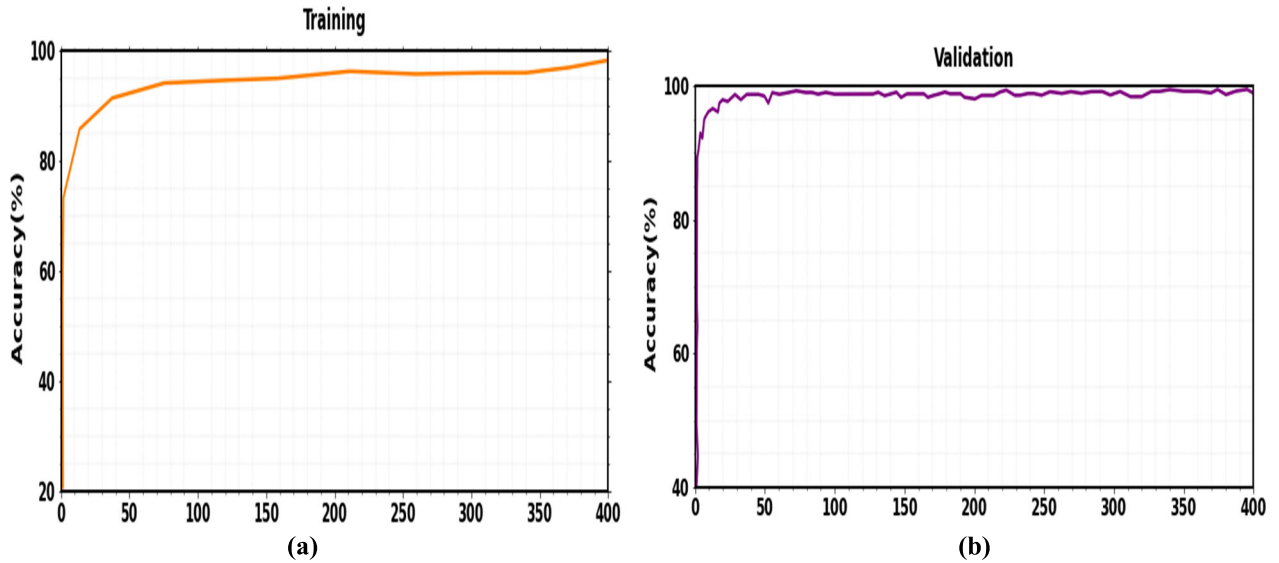


FIGURE 4. Accuracy is plotted against the iterations for both training and validation.

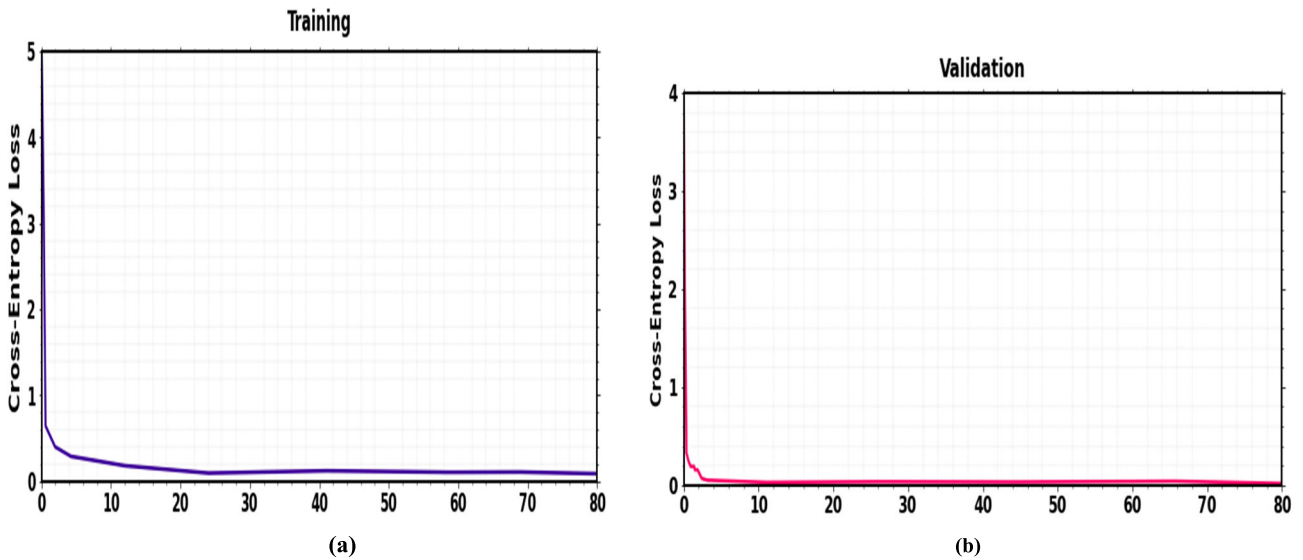


FIGURE 5. Cross-entropy loss is plotted against the iterations for both (a) training and (b) validation.

The training and validation accuracy rates were relatively high, at 99.10% and 98.40%. The six classes’ ROC curves are shown in Figure 6, and Table 6 contains their AUC values. The classifier has been appropriately trained, as shown by the ROC plots and AUC values, and accurate classification results have been obtained.

**E. TIME COMPLEXITY ANALYSIS**

Table 7 displays the estimated processing time for this research, an essential component in the image retrieval procedure. The entire time process consists of each image’s processing, training, and testing times. Reading the image is the initial stage in the processing time, followed by segmenta-

tion, feature extraction, and classification. Similarly, the time needed to train the entire dataset is represented by the training time for each network. Each network’s prediction determines the only testing time. However, compared to state-of-the-art techniques, this research displayed a processing time of about 10 seconds. The mean power used for calculating every sample was noted as 25.656 W.

**F. MODEL COMPLEXITY**

Additionally, we looked at both the computing and transmission costs associated with the proposed approach. The floating point operations (FLOPs), storage consumption, and parameters are used to quantify the computing cost of the

**TABLE 6.** Proposed approaches’ interactive performance analysis.

	Consolidation	Micronodules	Normal	Ground glass	Fibrosis	Emphysema	Avg
Precision (%)	98.96	98.84	98.68	98.75	97.52	98.94	98.65
Recall (%)	97.12	98.64	98.65	98.44	97.26	98.24	98.05
F-score (%)	98.94	97.73	97.14	97.65	94.87	97.50	98.13
AUC	0.9793	0.9948	0.9969	0.9811	0.9769	0.996	0.987

**TABLE 7.** The entire image processing time of the research.

Item	Time Complexity (ms)	Other Info
Processing time (Segmentation+ Feature extraction+ Classification)	9.56356	Per image
Training time	0.437288	Per Dataset
Testing time	0.037923	Per image

**TABLE 8.** Model complexity.

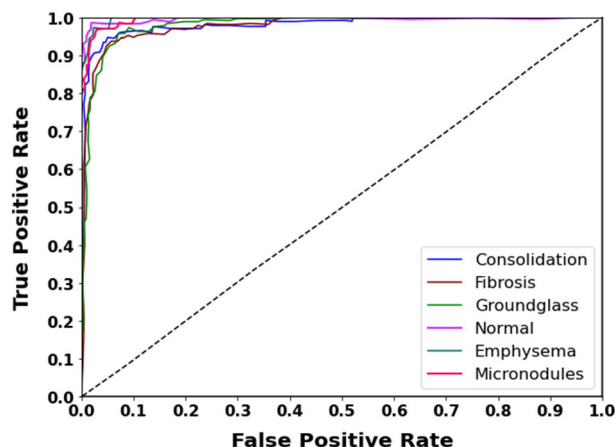
Techniques	Parameter	Flops	Memory consumption	Communication cost
improved U-Net++	3.6M	30M	0.85GB	1.2GB
MobileUNet V3	8M	32M	1.02GB	2.1GB

method. The transmission cost of the approach is also given in Table 8. According to the table, it can be used in lung disease classification processes and requires less computational work.

**G. PERFORMANCE COMPARISON**

In this section, we compare the categorization outcomes of our suggested approach with outcomes from earlier studies using IRCT images. The performance of the proposed approach is analyzed regarding F1-score, recall, precision, and accuracy. Table 9 represents the comparison of the proposed approach vs. existing approaches.

The effectiveness of the presented approach for ILD categorization and segmentation is compared to that of earlier research in the literature in Table 5. For the categorization of HRCT images, Pawar and Talbar [26] introduced the LungSeg-Net model. This approach yields 96.25% accuracy in classifying fibrosis, ground glass, reticulation, consolidation, emphysema, and nodules. Agarwala et al. [27] provide research on deep learning-based techniques for three-class ILD categorizations. Employing HRCT images, they defined multiple-layer CNN with 86% accuracy. To accurately classify the HRCT images into fibrosis, emphysema, consolidation, micro-nodule, GGO, and normal lung, Agarwala et al. [28] employed Google Neural Network, combining batch normalization with each activity layer. A VGG-SegNet was developed by Khan et al. [22]; this framework obtains 97.83% accuracy for categorizing binary data. The accuracy of the Custom CNN model used by Lakshmi et al. [29] for classifying multi-class ILD classifications is 91.22% accu-



**FIGURE 6.** ROC plots of the proposed network model for the six ILD classes.

rate. Compared to earlier deep learning models, the proposed hybrid deep learning model performs better, successfully classifies the various ILDs, and segments the defective HRCT image region. The proposed hybrid deep learning technique has two main benefits: it does not cause over-fitting and does not negatively impact network performance because of the categorization and segmentation processes.

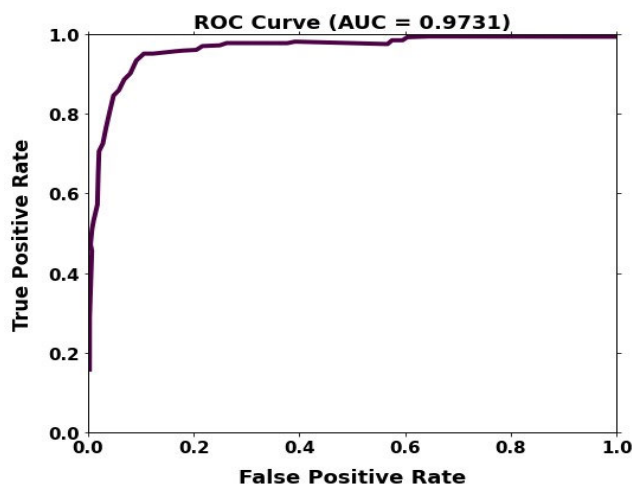
**H. EXTERNAL CLINICAL VALIDATION**

For clinical validation, we gathered information on Interstitial Lung Diseases from multiple sources. We collected data from 100 patients with lung disease to validate and evaluate the proposed framework for the classification of the disease. Consolidation, Micronodules, Normal, Ground glass, Fibrosis, and Emphysema are among the Interstitial Lung Diseases for which we gathered the multi-class classification of HRCT images. Three imaging professionals carefully examined the input data and combined it with pathology data to arrive at the final diagnosis. Once training was finished, the data were loaded into the MobileUNetV3 model after segmentation and feature extraction. The proposed hybrid system produced diagnostic results, such as the classification of lung disease into multiple classifications. The results were contrasted with the ground truth findings identified by experts. All lung cancer detection systems’ precision, recall, F1-score, accuracy, and AUC were evaluated. For clinical validation, the proposed hybrid system achieves a performance measure above 90%. Second, to compute the TPR and FPR, we categorized all of the labelled data from the test collection into TP and FP



TABLE 9. A comparative analysis between the proposed and cutting-edge approaches.

Reference	Lakshmi et al [29]	Khan et al [22]	Agarwala et al. [28]	Agarwala et al. [27]	Pawar et al. [26]
Year	2021	2021	2021	2020	2021
Train- test Splitting	training set:1000 images validation set:500 images testing set:273 images	10-fold cross-validation	Five-fold cross-validation	80% for training and 20% for testing	80% for training and 20% for testing
Dataset	Kaggle DICOM images	LIDC-IDRI, Lung-PET-CT-Dx	MedGIFT	MedGIFT, private ILD database	MedGIFT
Classifications	Carcinoma, Fibrosis, Healthy, Inflammation, and Necrosis	Normal, Nodule	Fibrosis, Emphysema, Consolidation, Micronodule, GGO, and Normal lung	Consolidation, Emphysema, Fibrosis	Fibrosis, Ground glass, Reticulation, Consolidation, Emphysema, Nodule
Technique	A hybrid deep learning model (improved UNet++-MobileUNetV3)	VGG-SegNet	GoogLeNet	CNN	LungSeg-Net
Accuracy (%)	91.22	97.83	93.03	86	96.25
Precision (%)	91.3	98.38	86.54	-	95.1
Recall (%)	88.2	97.02	87.01	-	-
F1-score (%)	87.8	97.79	86	-	-



**FIGURE 7.** ROC curve of proposed hybrid deep learning system in diagnosing lung diseases for clinical validation.

outcomes at various probability levels. This made it possible for us to keep a closer eye on and evaluate the results of the hybrid deep learning system's detection and classification. Figure 7 displays the ROC curve for the proposed diagnostic system based on the proposed methodology. When the AUC approaches 1, there is a discernible improvement in diagnostic effectiveness. When the AUC is between 0.5 and 0.7, it is not very accurate; between 0.7 and 0.9, it is very accurate; and above 0.9, it is very accurate. The accuracy of the test findings was shown by the AUC of 0.9731, which was determined using the trapezoidal rule. The clinical validation shows that the hybrid deep learning model (improved UNet++ - MobileUNetV3) has outstanding diagnostic abilities.

Furthermore, the proposed model allowed for a very accurate diagnostic. During our test, our proposed hybrid deep learning system automatically analyzed each data set in roughly 0.2 seconds. Given that each patient usually possesses fifteen features, the proposed method completes an automated diagnostic in three seconds as opposed to eight minutes for an imaging specialist. Based on the clinical data, the proposed hybrid deep learning system has an acceptable level of clinical viability. Consequently, our proposed AI diagnosis technique outperformed the conventional diagnostic approach in terms of effectiveness and accuracy.

## V. CONCLUSION

An innovative and efficient hybrid deep learning network model is suggested for screening ILD using whole HRCT images. The method's effectiveness has been enhanced by increasing the deep learning algorithms' accuracy at every level. From the HRCT images, the improved U-Net++ is employed in the initial stage of lung segmentation to remove the undesired background and enable the subsequent step to extract ILD, which precisely features deep feature extraction. The segmented lung images were utilized using RAPNet. The enhanced MobileUNetV3 uses deep learning characteristics

to categorize six ILD classes: consolidation, micro-nodules, ground glass, fibrosis, emphysema, and normal. The suggested algorithm's effectiveness has been contrasted with existing deep learning-based models. The proposed technique outperforms previous algorithms that take into consideration five ILD classes and outperforms an identical entire image-based algorithm by a significant margin. The proposed approach highlights the possibility of enhancing overall efficiency by selecting the most efficient CNN approach for a particular task and increasing accuracy at every level of the functions. The proposed method achieves an accuracy rate of 99.10% for various classifications of ILD disorders.

The proposed approach has a drawback because the dimension of combined characteristics is wider. It may be possible to minimize this set of features in the future by using a feature reduction strategy.

## ACKNOWLEDGMENT

They declare that this manuscript is original, has not been published before, and is not currently being considered for publication elsewhere.

## REFERENCES

- [1] A. H. N. Raju and P. Augustine, "Identification of interstitial lung diseases using deep learning," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 6, p. 6283, Dec. 2020.
- [2] R. Grassi, M. P. Belfiore, A. Montanelli, G. Patelli, F. Urraro, G. Giacobbe, R. Fusco, V. Granata, A. Petrillo, P. Sacco, M. A. Mazzei, B. Feragalli, A. Reginelli, and S. Cappabianca, "COVID-19 pneumonia: Computer-aided quantification of healthy lung parenchyma, emphysema, ground glass and consolidation on chest computed tomography (CT)," *La Radiologia Medica*, vol. 126, no. 4, pp. 553–560, Apr. 2021.
- [3] N. Nasrullah, J. Sang, M. S. Alam, M. Mateen, B. Cai, and H. Hu, "Automated lung nodule detection and classification using deep learning combined with multiple strategies," *Sensors*, vol. 19, no. 17, p. 3722, Aug. 2019.
- [4] H. Yar, N. Abbas, T. Sadad, and S. Iqbal, "Lung nodule detection and classification using 2D and 3D convolution neural networks (CNNs)," *Artif. Intell. Internet Things*, vol. 1, no. 1, pp. 365–386, 2021.
- [5] T. Saba, A. Sameh, F. Khan, S. A. Shad, and M. Sharif, "Lung nodule detection based on ensemble of hand crafted and deep features," *J. Med. Syst.*, vol. 43, no. 12, pp. 1–12, Dec. 2019.
- [6] S. M. Naqi, M. Sharif, and I. U. Lali, "A 3D nodule candidate detection method supported by hybrid features to reduce false positives in lung nodule detection," *Multimedia Tools Appl.*, vol. 78, no. 18, pp. 26287–26311, Sep. 2019.
- [7] S. A. Khan, S. Hussain, S. Yang, and K. Iqbal, "Effective and reliable framework for lung nodules detection from CT scan images," *Sci. Rep.*, vol. 9, no. 1, pp. 1–14, Mar. 2019.
- [8] M. Maqsood, S. Yasmin, I. Mehmood, M. Bukhari, and M. Kim, "An efficient DA-Net architecture for lung nodule segmentation," *Mathematics*, vol. 9, no. 13, p. 1457, Jun. 2021.
- [9] G. Singadkar, A. Mahajan, M. Thakur, and S. Talbar, "Deep deconvolutional residual network based automatic lung nodule segmentation," *J. Digit. Imag.*, vol. 33, no. 3, pp. 678–684, Jun. 2020.
- [10] B. Park, H. Park, S. M. Lee, J. B. Seo, and N. Kim, "Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks," *J. Digit. Imag.*, vol. 32, no. 6, pp. 1019–1026, Dec. 2019.
- [11] A. Kumar, S. Agarwala, A. K. Dhara, S. Mukhopadhyay, D. Nandi, M. Garg, N. Khandelwal, and N. Kalra, "Localization of lung fields in HRCT images using a deep convolution neural network," *Proc. SPIE*, vol. 10575, pp. 698–705, Feb. 2018.
- [12] Y. Wu and L. Lin, "Automatic lung segmentation in CT images using dilated convolution based weighted fully convolutional network," *J. Phys.: Conf. Ser.*, vol. 1646, no. 1, Sep. 2020, Art. no. 012032.

- [13] X. Zhang, X. Liu, B. Zhang, J. Dong, B. Zhang, S. Zhao, and S. Li, "Accurate segmentation for different types of lung nodules on CT images using improved U-Net convolutional network," *Medicine*, vol. 100, no. 40, Oct. 2021, Art. no. e27491.
- [14] S. Huang, F. Lee, R. Miao, Q. Si, C. Lu, and Q. Chen, "A deep convolutional neural network architecture for interstitial lung disease pattern classification," *Med. Biol. Eng. Comput.*, vol. 58, no. 4, pp. 725–737, Apr. 2020.
- [15] D. Bermejo-Peláez, S. Y. Ash, G. R. Washko, R. San José Estépar, and M. J. Ledesma-Carbayo, "Classification of interstitial lung abnormality patterns with an ensemble of deep convolutional neural networks," *Sci. Rep.*, vol. 10, no. 1, p. 338, Jan. 2020.
- [16] T. Pang, S. Guo, X. Zhang, and L. Zhao, "Automatic lung segmentation based on texture and deep features of HRCT images with interstitial lung disease," *BioMed Res. Int.*, vol. 2019, pp. 1–8, Nov. 2019.
- [17] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, and W. Cai, "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 991–1004, Apr. 2019.
- [18] C.-H. Chen, C.-K. Chang, C.-Y. Tu, W.-C. Liao, B.-R. Wu, K.-T. Chou, Y.-R. Chiou, S.-N. Yang, G. Zhang, and T.-C. Huang, "Radiomic features analysis in computed tomography images of lung nodule classification," *PLoS ONE*, vol. 13, no. 2, Feb. 2018, Art. no. e0192002.
- [19] M. Al-Shabi, B. L. Lan, W. Y. Chan, K.-H. Ng, and M. Tan, "Lung nodule classification using deep local-global networks," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 10, pp. 1815–1819, Oct. 2019.
- [20] W. Guo, Z. Xu, and H. Zhang, "Interstitial lung disease classification using improved DenseNet," *Multimedia Tools Appl.*, vol. 78, no. 21, pp. 30615–30626, Nov. 2019.
- [21] V. N. S. Doddavarapu, G. B. Kande, and B. P. Rao, "Differential diagnosis of interstitial lung diseases using deep learning networks," *Imag. Sci. J.*, vol. 68, no. 3, pp. 170–178, Apr. 2020.
- [22] M. A. Khan, V. Rajinikanth, S. C. Satapathy, D. Taniar, J. R. Mohanty, U. Tariq, and R. Damaševičius, "VGG19 network assisted joint segmentation and classification of lung nodules in CT images," *Diagnostics*, vol. 11, no. 12, p. 2208, Nov. 2021.
- [23] G. S. Tran, T. P. Nghiem, V. T. Nguyen, C. M. Luong, and J.-C. Burie, "Improving accuracy of lung nodule classification using deep learning with focal loss," *J. Healthcare Eng.*, vol. 2019, pp. 1–9, Feb. 2019.
- [24] N. Khehrah, M. S. Farid, S. Bilal, and M. H. Khan, "Lung nodule detection in CT images using statistical and shape-based features," *J. Imag.*, vol. 6, no. 2, p. 6, Feb. 2020.
- [25] W. Wang, F. Liu, X. Zhi, T. Zhang, and C. Huang, "An integrated deep learning algorithm for detecting lung nodules with low-dose CT and its application in 6G-enabled Internet of Medical Things," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5274–5284, Apr. 2021.
- [26] S. P. Pawar and S. N. Talbar, "LungSeg-Net: Lung field segmentation using generative adversarial network," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102296.
- [27] S. Agarwala, M. Kale, D. Kumar, R. Swaroop, A. Kumar, A. Kumar Dhara, S. Basu Thakur, A. Sadhu, and D. Nandi, "Deep learning for screening of interstitial lung disease patterns in high-resolution CT images," *Clin. Radiol.*, vol. 75, no. 6, pp. 481.e1–481.e8, Jun. 2020.
- [28] S. Agarwala, A. Kumar, A. K. Dhara, S. B. Thakur, A. Sadhu, and D. Nandi, "Special convolutional neural network for identification and positioning of interstitial lung disease patterns in computed tomography images," *Pattern Recognit. Image Anal.*, vol. 31, no. 4, pp. 730–738, Oct. 2021.
- [29] D. Lakshmi, J. Sivakumar, K. P. Thanaraj and N. Thendral, "Customized convolution neural network for multi-class lung abnormality classification from CT images," *Inf. Technol. Ind.*, vol. 9, no. 1, pp. 49–57, Feb. 2021.



**SURENDRA REDDY VINTA** is currently an Associate Professor with the School of Computer Science and Engineering, VIT-AP University, Amaravathi. He is dedicated to teaching field from the past 14 years. He has published more than 17 papers in national and international journals and conferences. He has guided 36 PG students and 50 UG students. His research interests include machine learning, image processing, data mining, network security, and artificial intelligence. He is highly passionate and enthusiastic about his teaching. He was a member of editorial board and a reviewer of various international journals and many more awards for his research.



**B. LAKSHMI** is currently an Assistant Professor with the Department of Computer Applications, Velagapudi Ramakrishna Siddhartha Engineering College. She has 15 years of teaching experience. She has participated in various international conferences and workshops held at different places. Her research interests include database security and data mining.



**M. ARUNA SAFALI** is currently an Associate Professor with the Department of CSE, Dhanekula Institute of Engineering & Technology, Amaravathi. She is dedicated to teaching field from the past 12 years. She has published more than 17 papers in national and international journals and conferences. Her research interests include machine learning and image processing. She was a member of the editorial board and a reviewer of various international journals and she has received many more awards for her research.



**G. SAI CHAITANYA KUMAR** received the master's degree from Acharya Nagarjuna University and the Ph.D. degree from JNTU Hyderabad. He is currently an Associate Professor with the Department of Computer Science and Engineering, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh. He has been dedicated to teaching for the past 14 years and has published more than 20 papers in national and international journals and conferences. He was a member of editorial board and a reviewer of various international journals. He has guided 25 PG students and 150 UG students. He is highly passionate and enthusiastic about his teaching. His research interests include image processing, data mining, and network security.

...