Enhanced Pulmonary Pattern Classification in HRCT Images Using Advanced Data Augmentation Strategies and Fine-Tuned CNN Models

Zeynep Hafsa DİLMAÇ
Computer Engineering Department
Hacettepe University
Ankara,TURkey
zeynephafsadilmac@gmail.com

Alihan SAĞÖZ
Computer Engineering Department
Hacettepe University
Ankara,TURkey
sagozalihan@gmail.com

Hüseyin Yiğit ÜLKER
Computer Engineering Department
Hacettepe University
Ankara,TURkey
huseyinulker@hacettepe.edu.tr

Ebru AKÇAPINAR SEZER
Computer Engineering Department
Hacettepe University
Ankara,TURkey
ebruakcapinarsezer@gmail.com

2024 Innovations in Intelligent Systems and Applications Conference (ASYU) | 979-8-3503-7943-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ASYU62119.2024.10757029

***Abstract*—High-resolution computed tomography (HRCT) images are among the most essential and fundamental imaging modalities to diagnose and assess various pulmonary condition types. The current research focuses on classifying HRCT images into seven significant categories: ground-glass, fibrosis, micronodules, consolidation, healthy, reticulation, and emphysema. We first preprocess and clean the collected HRCT data and then fine-tune the state-of-the-art 2D pre-trained models like Inception-v3 and EfficientNet on the category-classified HRCT images. Further, we applied various data augmentation techniques that will improve the performance of these models. The results we obtained showed that this combination of models produced significant improvements in classification accuracy, as demonstrated by our experiment. This is expected to help doctors make more confident decisions when diagnosing pulmonary conditions. This approach not only enhances diagnostic precision but also paves the way for further research in automated medical image analysis.**

INTRODUCTION

Over the years, high-resolution computed tomography (HRCT) imaging has become an essential adjunct to the diagnosis and evaluation of many pulmonary conditions. Most importantly, HRCT scanning provides high-quality, high-resolution images of the lung parenchyma. [1] It is of remarkable value to recognize and evaluate interstitial lung diseases (ILDs), infections, and other changes in lung tissue. Accurate classification of these conditions is essential for patient management and treatment planning. Although effective, traditional diagnostic methods are often time-consuming and subject to interobserver variability.

Few studies to date have used machine learning via convolutional neural networks for the automated analysis of HRCT images to overcome this drawback and advance precision. The ILD database was developed at the University Hospitals of Geneva (HUG) and is one of the databases used for this current study. This publicly available collection of high-resolution CT scans includes annotated pathologic lung regions and complete clinical parameters of 128 patients diagnosed with one of 13 types of ILD. [5]

One notable research study on lung segmentation that involved the HUG ILD dataset Joana Sousa and Tania Pereira used the Residual U-Net method; they achieved a Dice similarity coefficient value of 0.9267 ± 0.1671 for the lung, depicting enormous improvements in segmentation accuracy from the conventional methods. [4] The second study, Pawar and Talbar proposed a two-stage method: the first for segmenting HRCT images and the second for classifying the segmented images into six classes of ILD, namely average, emphysema, fibrosis, ground glass, micronodules, and consolidation. The achieved overall accuracy was 89.39%. [3]

We have classified the HRCT images into seven categories: ground-glass, fibrosis, micronodules, consolidation, healthy, reticulation, and emphysema using two state-of-the-art CNN architectures—Inception and EfficientNet. These models are capable of handling various feature scales in the HRCT images and automatically optimizing the dimension of network capacity to support complex pulmonary conditions. Because of the minimal number of labeled samples in our dataset, we use data augmentation techniques further to enforce model generalization and performance on unseen images. [13]

Our work has tried to enhance this field, contributing with a solid automated approach to HRCT image classification, seeking high accuracy in clinical applications with high efficiency without using any segmentation technique.

METHODOLOGY AND STUDIES

I. DATA

In this project, we used the HUG-data, an extensive dataset with high-resolution CT images, annotated, including seg-

979-8-3503-7943-3/24/31.00 ©2024 IEEE

mentation for different types of pulmonary abnormalities. [5] Though this dataset has segmentation masks, for classifying, only image labels have been taken into consideration. There are 12 types of labels in the HUG-data. The dataset is exceptionally rich in variety and consists of the following class distribution, as shown in Table I.

TABLE I
CLASS DISTRIBUTION AND INDIVIDUALS WITH ABNORMALITIES

Class	Number of Instances	Number of Individuals
Ground Glass	427	37
Fibrosis	473	38
Micronodules	297	16
Consolidation	196	14
Healthy	100	7
Reticulation	131	10
Emphysema	66	5
Bronchiectasis	44	8
Macronodules	37	7
Bronchial wall thickening	15	1
Cysts	11	3
Others	149	14

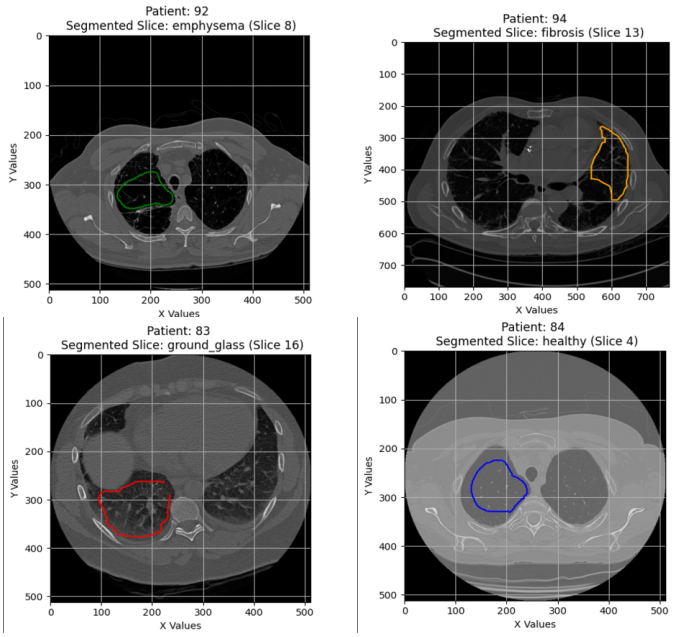

Fig. 1. Examples from the dataset showing various conditions.

This diversity in the dataset, both in terms of the number of images and the variety of conditions, can be ideal for training robust deep-learning models. In this work, we applied Inception and EfficientNet individually or via ensembling methods to classify these images. We will also use many different augmentation techniques to make the training data more variable, which should help the model generalize better. In the next sections, specific items of our methodology are elaborated on with the model architectures, training procedures, and results of our comparative analysis. The class examples are shown in Figure 1.

II. DEEP LEARNING MODELS

A. EFFICIENT-B0

EfficientNet [6] enhances the performance of convolutional neural networks (CNNs) by optimizing model scaling methods. Traditional approaches often scale only one dimension of the model (depth, width, or resolution), leading to suboptimal performance. EfficientNet introduces a compound scaling method that balances and scales these three dimensions uniformly, improving efficiency and accuracy. The EfficientNet family includes models from EfficientNet-B0 to EfficientNet-B7. EfficientNet-B7 achieves 84.3% top-1 accuracy on ImageNet, being 8.4 times smaller and 6.1 times faster than the best existing CNN model. EfficientNet models also perform exceptionally well on transfer learning datasets such as CIFAR-100 and Flowers. [11]

a) Development of EfficientNet-B0: EfficientNet-B0 was developed using a multi-objective neural architecture search (NAS) that optimizes both accuracy and floating point operations per second (FLOPS). The optimization objective is expressed as $ACC(m) \times \left[\frac{\text{FLOPS}(m)}{T}\right]^{w}$, where ACC(m) denotes the accuracy of the model m, FLOPS(m) represents the FLOPS of the model, T is the target FLOPS, and w=-0.07 is a hyperparameter controlling the trade-off between accuracy and FLOPS. Unlike approaches targeting specific hardware, this method optimizes FLOPS rather than latency. This search process resulted in the efficient network named EfficientNet-B0.

B. INCEPTION-RESNETV2

Recent advances in image recognition have been driven by deep convolutional networks like Inception and the use of residual connections to mitigate the vanishing gradient problem. Combining these led to Inception-ResNet networks, which train faster and sometimes perform better than non-residual versions. [7]

Inception-ResNet-v2 Architecture and Features: Inception-ResNet-v2 is an advanced iteration that builds upon the Inception-ResNet architecture by introducing several refinements aimed at improving performance and stability. Key features and improvements include:

- Streamlined Architecture: Inception-ResNet-v2 incorporates a more streamlined and efficient architecture, designed to optimize the balance between computational cost and recognition performance.
- Residual Inception Modules: The network employs residual connections within the Inception modules, facilitating faster training and improved gradient flow.
- Activation Scaling: Proper scaling of activations is used to stabilize the training of very wide residual Inception networks, addressing potential issues of gradient instability.
- Single-Frame Recognition Performance: The architecture significantly enhances single-frame recognition performance on the ILSVRC 2012 classification task.

The empirical results demonstrate that Inception-ResNet-v2 outperforms similarly expensive non-residual Inception networks by a thin margin. Specifically, an ensemble of three residual networks and one Inception-v4 achieved a top-5 error rate of 3.08% on the test set of the ImageNet classification challenge. This highlights the efficacy of combining residual connections with the Inception architecture, resulting in improved training dynamics and overall performance. [12]

C. METHOD

a) Data and Pattern Labeling: In this study, CT images from the HUG dataset were labeled based on patterns, ignoring region-specific labels. Each CT image was annotated with the pattern(s) present. If a single image contained multiple patterns, the image was duplicated and saved with distinct labels corresponding to each pattern. All images were stored in PNG format. The final distribution of pattern labels and the number of images for each pattern are presented in the Table I.

b) Classification Using Different Patterns: Using data encompassing seven distinct patterns, an image classification task was undertaken. Based on the number of patterns, separate classification tasks were conducted for the first five patterns, the first six patterns, and all seven patterns, resulting in 5-class, 6-class, and 7-class classification tasks, respectively. For these classification tasks, the aforementioned models were fine-tuned according to the specific number of classes required for each task. The outcomes of these classification tasks are summarized in the results section, with detailed tables displaying the performance metrics for each task. as shown in Table III.

c) Data Sampling Strategies: To enhance the classification results, several data augmentation strategies were employed.

- Augmentation with Random Transformations: In this strategy, data augmentation was performed by applying random transformations to the images, based on specified augmentation ratios. In order to produce augmented data, augmentation parameters were used. Images may have their width and height adjusted by up to 1% of their overall width or height, with the parameters "width_shift_range" and "height_shift_range" set to 0.01. Random zooming was regulated by a 'zoom_range' of 0.01, while the shear transformation's strength was set with a 'shear_range' of 0.01. To improve dataset diversity, 'Horizontal_flip' was enabled, allowing photos to be flipped horizontally at random. To fill in missing pixels using the value of the nearest pixel, the 'fill_mode' argument was set to 'nearest'. In addition, the range for random brightness modifications was defined by the 'brightness_range' of [0.5, 1.5], which improved model resilience and dataset variability.
  The final number of images for each pattern after applying these augmentations is provided in the Table II.

- Balancing the Number of Images: In this strategy, the number of images for the "consolidation" pattern was increased to 390, the maximum number for any pattern, through image duplication.

TABLE II
NUMBER OF IMAGES FOR EACH PATTERN

Pattern	Number of Images
Ground Glass	468
Fibrosis	463
Micronodules	328
Consolidation	237
Healthy	120
Reticulation	97
Emphysema	72

Improvements were observed with both data augmentation strategies, and the results for each strategy are presented separately in the results section with detailed tables. These results demonstrate the effectiveness of the data augmentation approaches in enhancing classification performance.

III. EXPERIMENTAL RESULTS

A methodology, consisting of several tests, was carefully developed in our methodical approach to improving the identification of interstitial pulmonary fibrosis (IPF) using deep learning architectures, namely EfficientNet and Inception. The HUG dataset, comprising high-resolution CT images carefully annotated with diseased lung tissue areas and clinical information from 108 patients diagnosed with ILDs, was kindly supplied by the University Hospitals of Geneva. This dataset was utilized to craft a rigorously constructed research methodology. Our models' classification precision was systematically improved via several iterative stages in order to enhance the precision of IPF detection in clinical settings. A distinct increase in model performance was observed at different stages of optimization in our test findings. Initially, when the original dataset with directly labeled slices was utilized, promising accuracy was demonstrated by our models. For instance, EfficientNet and Inception models showed substantial results with varying class numbers, with EfficientNet achieving a test accuracy of 0.918 for 5 classes and Inception reaching 0.91 for the same. As we progressed, a significant increase in classification accuracy was noticed after augmentation techniques were introduced to the data. Specifically, the Inception model with 5 classes improved to a test accuracy of 0.86 from 0.91, showcasing the impact of data augmentation on model performance. Furthermore, the implementation of data duplication techniques to ensure uniform consolidation pattern representation resulted in further improvement in the performance of our models, particularly in distinguishing subtle differences between lung patterns. For example, the EfficientNet model for 5 classes achieved a test F1 Score of 0.90 after balancing the dataset with duplications, reflecting enhanced model precision. These outcomes underscore the effectiveness of our method in enhancing IPF detection precision, thereby opening up the possibility of more precise diagnosis and treatment planning in clinical settings.

A. Original Data Classification Results

The original data classification results for both EfficientNet and Inception models are summarized in Table III. The models were trained for 20 epochs using SGD optimizer with a learning rate of 0.005.

TABLE III
ORIGINAL DATA CLASSIFICATION RESULTS

Model	Class Number	Loss Train	Accuracy Train	Loss Test	Accuracy Test	F1 Score Test
EfficientNet	5	0.094	0.95	0.2746	0.918	0.92
Inception	5	0.082	0.96	0.40	0.91	0.90
EfficientNet	6	0.12	0.94	0.84	0.76	0.76
Inception	6	0.17	0.93	0.71	0.85	0.79
EfficientNet	7	0.035	0.9866	0.6131	0.84	0.83
Inception	7	0.126	0.942	0.4768	0.86	0.83

B. Augmented Data Classification Results

To improve the results, we applied data augmentation techniques. The augmented data classification results are shown in Table IV. The models were trained for 20 epochs using SGD optimizer with a learning rate of 0.005.

TABLE IV
AUGMENTED DATA CLASSIFICATION RESULTS

Model	Class Number	Loss Train	Accuracy Train	Loss Test	Accuracy Test	F1 Score Test
EfficientNet	5	0.094	0.948	0.387	0.91	0.91
Inception	5	0.051	0.98	0.52	0.86	0.87
EfficientNet	6	0.157	0.92	0.32	0.88	0.88
Inception	6	0.1959	0.92	0.378	0.89	0.88
EfficientNet	7	0.019	0.99	0.4	0.91	0.90
Inception	7	0.134	0.93	0.45	0.87	0.88

C. Duplicate Consolidation Class Data Classification Results

We further balanced the dataset by duplicating samples in the consolidation class to have an equal number of samples (390) across all classes. The classification results for this approach are provided in Table V. The models were trained for 20 epochs using SGD optimizer with a learning rate of 0.005.

TABLE V
DUPLICATE CONSOLIDATION CLASS DATA CLASSIFICATION RESULTS

Model	Class Number	Loss Train	Accuracy Train	Loss Test	Accuracy Test	F1 Score Test
EfficientNet	5	0.12	0.932	0.243	0.90	0.90
Inception	5	0.1	0.945	0.456	0.88	0.88
EfficientNet	6	0.137	0.93	0.34	0.90	0.90
Inception	6	0.14	0.934	0.34	0.89	0.89
EfficientNet	7	0.1595	0.935	0.396	0.90	0.89
Inception	7	0.20	0.91	0.391	0.90	0.90

In the initial phase of our investigation, we evaluated baseline models on the original dataset. The results showed that EfficientNet and Inception had accuracies between 0.76 and 0.918 and 0.85 and 0.91, respectively. Experiments that combined data duplication and augmentation strategies resulted in much higher classification accuracy. Test accuracies were improved by augmentation to 0.88-0.91 for EfficientNet and 0.86-0.89 for Inception, and further increased to 0.88-0.90 for both architectures by data duplication.

IV. CONCLUSION

In radiology, an important diagnostic phase for lung diseases, detecting and examining patterns and their locations is a significant step. In this study, the detection of patterns related to major lung diseases was carried out for 7 different patterns in CT sections. When obtained results are considered, it is clear that the data augmentation increases the model performance while presenting the data to pre-trained models. The reason for this is reaching many examples that allow the discrimination of patterns with similar appearances and maintaining balance among the number of samples belonging to each class. Future studies can be conducted to develop decision support systems that serve the automatic diagnosis of diseases by increasing the number of classes and determining the intra-lung positions from sections containing lung patterns.

REFERENCES

[1] N. Sverzellati, "Highlights of HRCT imaging in IPF," *Respir Res*, vol. 14, Suppl 1, p. S3, 2013.

[2] H. P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, "Deep Learning in Medical Image Analysis," *Adv Exp Med Biol.*, vol. 1213, pp. 3-21, 2020. doi: 10.1007/978-3-030-33128-3-1. PMID: 32030660; PMCID: PMC7442218.

[3] S. P. Pawar and S. N. Talbar, "Two-Stage Hybrid Approach of Deep Learning Networks for Interstitial Lung Disease Classification," *BioMed Res. Int.*, vol. 2022, p. 7340902, 2022.

[4] J. Sousa, T. Pereira, F. Silva, M. C. Silva, A. T. Vilares, A. Cunha, and H. P. Oliveira, "Lung Segmentation in CT Images: A Residual U-Net Approach on a Cross-Cohort Dataset," *Appl. Sci.*, vol. 12, p. 1959, 2022.

[5] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P. A. Poletti, and H. Müller, "Building a reference multimedia database for interstitial lung diseases," *Computerized Medical Imaging and Graphics*, vol. 36, no. 3, pp. 227-238, 2012.

[6] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv preprint arXiv:1905.11946*, 2020.

[7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *arXiv preprint arXiv:1602.07261*, 2016.

[8] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[9] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[10] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[11] GeeksforGeeks, "EfficientNet Architecture," Available: urlhttps://www.geeksforgeeks.org/efficientnet-architecture/.

[12] Medium, "Understanding Residual Connections in Neural Networks," Available: urlhttps://readmedium.com/understanding-residual-connections-in-neural-networks-866b94f13a22.

[13] Y. S. Doddavarapu, V. N., G. B. Kande, and B. Prabhakara Rao, "Differential diagnosis of Interstitial Lung Diseases using Deep Learning networks," The Imaging Science Journal, vol. 68, no. 3, pp. 170–178, 2020. doi: 10.1080/13682199.2020.1781394.