2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID 2021)

Deep Learning Approach for Auto-Detecting Idiopathic Pulmonary Fibrosis Prediction

Ziyuan Wang
School of Electrical Engineering and Computer Science
Kungliga Tekniska högskolan
Stockholm, Sweden
guanghua.ren@gecacademy.cn

***Abstract*—In the field of computer vision, Convolutional Neural Network has been the most mainstream method and has shown excellent performance in medical images. Among Convolutional Neural Networks, U-Net and DenseNet have demonstrated outstanding and robust performance in image recognition and image segmentation, respectively. In this paper, we proposed a neural network with DenseNet as the Encoder and Unet as the Decoder for lung image segmentation and feature extraction. With this neural network, we extracted features from patients' CT Scan images and combined them with patients' clinical records to predict lung function trends in the future. This predictive value will provide significant help in determining whether the patient has Idiopathic Pulmonary Fibrosis, which is the purpose of our study.**

Keywords-component; Auto-Detection, Computer Vision, IPF Prediction

I. INTRODUCTION

In modern society, with the skyrocketing development of Artificial Intelligence, the interdisciplinary usage of AI techniques and Convolutional Neural Network (CNN) applications are widely striking. In 1994, C.J.Vyborny et al. applied computer-vision techniques in the area of mammography, i.e., detecting the abnormal segments of given Computed Tomography(CT) Scan images[1]. And this type of application could also be valid on other sources of medical images for monitoring or diagnosing [2-5]. With this mentality, we are focusing on approaching a disease that is relatively difficult for normal approaches to diagnose. We focus on lung function disease named idiopathic pulmonary fibrosis (IPF) for this study.

Idiopathic pulmonary fibrosis (IPF) is one deadly disease of the family of idiopathic types of pneumonia. The symptom of IPF is idiopathic interstitial pneumonia, including short breath and dry cough [6]. Successful diagnosis for IPF within a relatively short amount of time is a challenging problem. Patients with IPF cannot have access to subspecialty care within the ideal treatment duration since the diagnosis for IPF often comes with a delay than the disease itself. Although advances have been made in antifibrotic therapy, which utilizes Nintedanib [7] for reducing lung function decline, the median survival time is only 2-5 years with a median diagnostic delay of 2.1 years [8]. A recent study indicated that IPF patients experience a precipitous course controlled in the early phase, and then the lung function severely declines within a short amount of time [7]. The logic behind the delay in diagnosis is that patients cannot be aware of the disease in the early stage, which leads to a 78 percent death rate. And another rationale for the delay is that symptoms are similar to other lung conditions. Through a CT Scan of patients' lungs, the visual image can only show the scarring of lungs, leading to multiple diseases.

D. A. Lynch et al. introduced an approach [9] to diagnosing IPF using High-Resolution Computed Tomography (HRCT). This approach uses a radiologic assessment on the HRCT images to determine if the patient is a "definite" or "probable" IPF patient. They matched the results given by radiologists with specific features in HRCT images of lungs. It turned out that the most important feature is the honeycombing pattern of the lung, and the other 13 features contain ground-glass attenuation, centrilobular nodules, etc. This approach needs the HRCT to have a honeycombing pattern as a foundation; however, the honeycombing pattern cannot be found within any common CT images. Another methodology brought by G. Raghu et al. is to extract possible interstitial pneumonia features from over 1000 HRCT images to calculate the predictive value, which shows around 95 percent correlation with IPF [10]. However, the usage of HRCT is relatively rare in the current stage of our society. Due to unknowing the potential of IPF, patients and doctors rarely take this approach due to the high cost of time and resources.

With the development of deep learning, computer-aided diagnosis for IPF using CT images is found more efficient [11-13]. In 2019, A. Christe et al. implemented deep learning techniques into the diagnosis process to test the possibility of using CT images and clinical records to extract the pattern of IPF [14]. After segmenting the lung CT images, they used CNN to identify the specific region's health condition. This model's accuracy achieved 0.81; however, the sensitivity is relatively low, at 0.76, which indicates their model's robustness could be possibly lower if the dataset shifts, which is one of the heating problems in the modern computer vision area [15].

The purpose of this study is to assess the capability of only using patients' lung CT images and their clinical records to predict if they have the potential of having IPF with deep learning techniques. If we can provide a robust model for predicting the Forced Vital Capacity (FVC) in the given future, it would become possible for doctors to know whether patients are prone to having IPF in the future. It is also possible to calculate the confidence of the FVC prediction.

978-0-7381-1083-7/21/$31.00 ©2021 IEEE

II. MATERIEL & METHOD

A. Dataset

The data is secured and provided by Open Source Imaging Consortium (OSIC). Based on ethical rules for medical research, the database provided 85 percent of the patients' clinical records with seven features, as shown in Table 1. While the clinical records stated the time of measurement, the dataset also contains the CT-Scan images of patients without statements of time of each image. An example of CT-Scan images of lung slices is shown in Fig. 1. The visual impact of the two slices in Fig. 1. is not considered significant to ensure the slices are coherent and can capture the small variances in each slice.

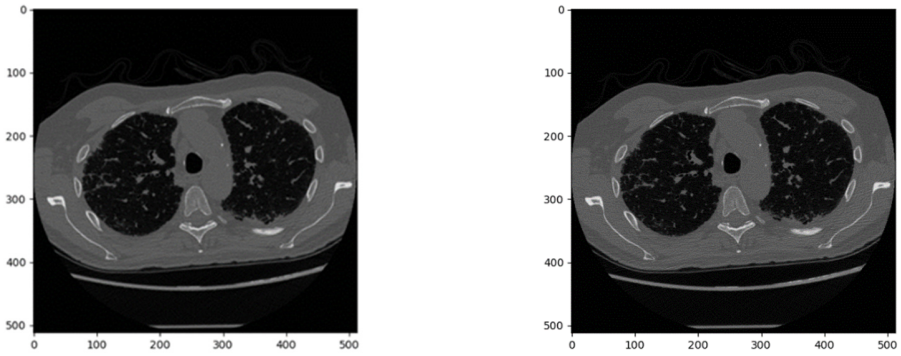

Fig 1. Lung CT Scan Slices from A Random Patient

B. Hypotheses

Definition 1: Forced vital capacity (FVC). One of our primary goals of predicting process is to predict the FVC value with a specific timestamp. FVC is the total amount of air exhaled when the lung reaches its capacity. We defined the FVC score in patients' clinical records as FVC_{true} and defined the predicted FVC score as FVC_{true}.

Definition 2: Error and Confidence. The evaluation is provided by OSIC and Kaggle, which defined the error of prediction in Equation (1). Equation (1) stated that set the threshold of Δerror as 1000 ml to avoid large errors adversely penalizing results. And the confidence value was defined with δ_{true}, which is the variance of our prediction. In Equation (2), we can see that the confidence value, δ_{conf}, set the threshold at 70 ml to reflect the approximate measurement uncertainty in prediction.

$$\Delta_{error} = \min\left(\left|FVC_{error} - FVC_{predict}\right|, 1000\right) \quad (1)$$

$$\delta_{conf} = \max(\delta_{true}, 70) \quad (2)$$

Based on various researches [16]–[18]and our own result, the distribution of FVC value across the patients could be simulated by Laplace distribution. This enabled us to combine Equation (1) and Equation (2) into a metric to represent the accuracy and confidence value of the result.

Definition 3: Normalized metric. According to Equation (1) and Equation (2), they defined a function to combine both scores to represent the accuracy and confidence. Hence, based on Laplace Log Likelihood, Equation (3) represent our definition of evaluation of prediction. Note that F_{metric} is negative, and a F_{metric} that is closer to 0 represents a better prediction.

$$F_{metric} = {\sqrt{2}\Delta_{error}}\Big/{\delta_{conf}} + ln(\sqrt{2}\delta_{conf}) \quad (3)$$

Table 1. Clinical Data Description

Feature	Description
Patient	Unique ID for Patients
Weeks	Relative number of weeks pre/post the baseline CT
FVC*	Recorded lung capacity
Percent*	Computed field which approximates the patient's FVC as a percent of the typical FVC for a person of similar characteristics
Age	Age of patients
Sex	Sex for the patient
SmokingStates	The Smoking status of a given patient

∗ Dependent Variable. Representing the quality of lung function, and the greater the FVC, the better the lung function
∗∗ Dependent Variable. The result of dividing the FVC value by the average FVC of people with the same attributes

C. Theoretical Framework

1) Feature Extraction

Based on [9, 10, 14], feature extraction from CT and HRCT images is significantly important in diagnosing process for IPF. The fundamental problem in medical image feature extraction is the segmentation of medical images since it allows the data

source to present in-depth features. And current methods to segment lung CT images are prone to deep neural networks compared to traditional methods such as the Watershed algorithm [19], since deep neural networks could have a much more robust performance and a lighter, faster model. The most astonishing achievement in segmenting CT images of deep neural networks is proposed by O. Ronneberger et al. in 2015, named as U-Net [20]. This model's structure is shown in Fig. 2; their study is based on a contracting path and an expansive path. The contracting path is a typical neural network containing two 3×3 convolutions followed by the ReLU linear unit and 2×2 max-pooling layers for downsampling. The expansive path contains an upsampling layer followed by a 2×2 convolution layers and two 3×3 convolutions and the ReLU linear unit. The architecture enables two advantages: one is in the feature extraction part, that is, there is one scale every time it passes through a pooling layer. Plus, the original image scale, there are a total of 5 scales. And another one is in the up-sampling part, that is, each up-sampling is fused at the same scale as the number of channels corresponding to the feature extraction part, but it must be cropped before fusion. These bestow U-Net's ability to manage different input sizes, which is ideal for managing medical images. And with these advantages, the method won the ISBI cell tracking challenge 2015 [21] outstandingly.

U-Net is one of the earliest algorithms that use multi-scale features for semantic segmentation tasks, and its U-shaped structure also inspired many subsequent algorithms. However, some flaws are lying in U-Net. One of U-Net's downsides is that sufficient convolution layers increase the difficulty and universality of model design. To cope with that, in the year 2018, a structure based on ResNet [22] is published by T. G. Dietterichet et al. This article suggested the concept of denseblock whose form is shown in Fig. 3, in the network G. Huang et al. proposed in [23], named DenseNet. The design of denseblock enables the input directly to be connected to the loss function, which boosted the transaction of features between layers to allow a more efficient feature utilization and significantly ease the problem of vanishing-gradient.

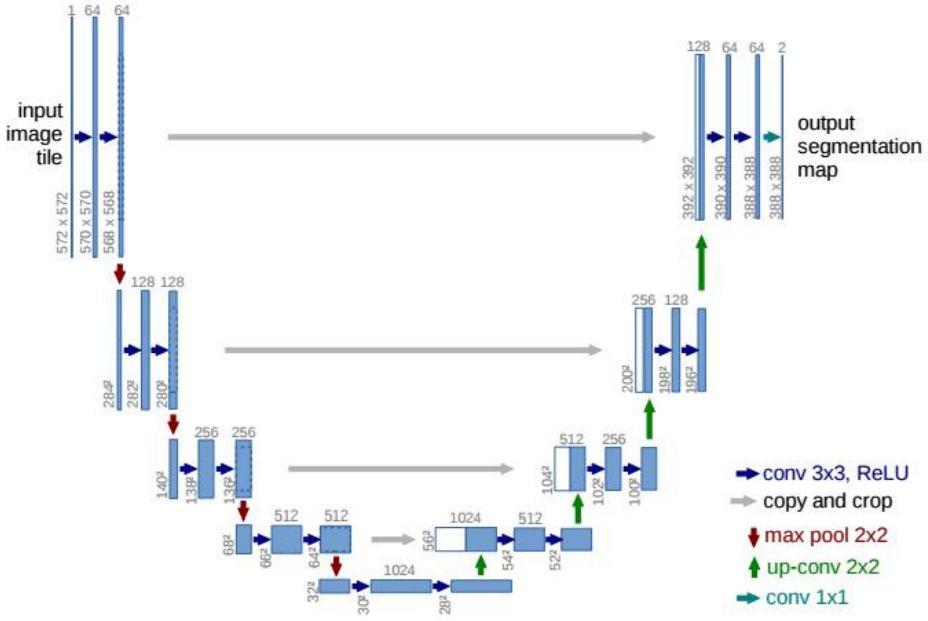

Fig 2. The U-Shape Structure of DecoderNetwork

2) Progression

With the supervised learning algorithm of machine learning, our goal is to learn a stable model that performs well. Still, the actual situation is often not ideal, and sometimes we can only get multiple models with preferences. Ensemble learning is to combine various weakly-supervised models to obtain a better and more comprehensive strong-supervised model. The underlying idea of ensemble learning is that even if a specific weak classifier gets a wrong prediction, these weak learners' combinations could be correct and robust. In the year 2016, T. Chen et al. proposed Extreme Gradient Boosting [24], which performs well in prediction, regression, and classification. And the machine learning idea, gradient boosting, published by J. H. Friedman et al. in 2001 [25], is based on combining weak learners to obtain a strong learner. To combine weak learners, J. H. Friedman proposed an approach that uses the negative gradient to measure the error of the previous round of the weak learners. In each round, weak learners pay more attention to the samples of the prior round of learning errors during the training process. The weak learners could linear regression, binary tree, or simple neural network. T. Chen et al. used first-order Taylor expansion to simplify the gradient descent process and enable parallel computing technique in the algorithm [24], which have great practical value.

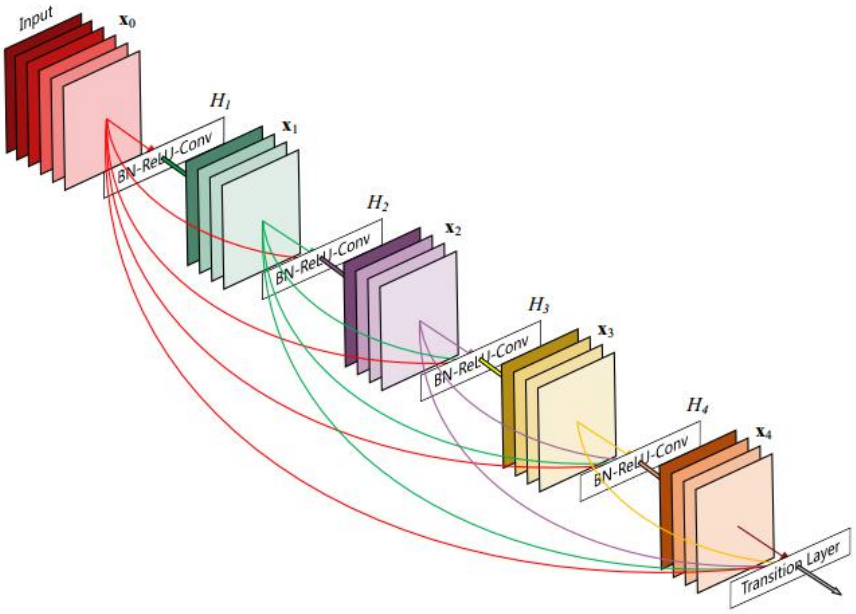

Fig 3. The Structure of DenseBlock in DesenNet as Encoder

Another well-known and well-performed method proposed in the ensemble learning area is named Random Forest [26], brought by L. Breiman in 2001. With a similar mentality and assumption of gradient boosting, the random forest is based on decision trees. The model has made improvements to the establishment of decision trees [27]. Ordinary decision trees select an optimal feature from all features at each node to divide it into left and right subtrees. The random forest randomly selects a part of the features at each node and then selects an optimal one. This method further enhanced the generalization ability of the model. Google has brought a more robust model in 2019 named Efficient Net [28]. Convolutional Neural Networks are usually developed at a fixed resource cost and then scaled up when more resources are added to achieve higher accuracy. For example, ResNet [22] can extend ResNet-18 to ResNet-200 by increasing the number of layers, and GPipe [29] has achieved 84.3\% accuracy on ImageNet by extending the CNN baseline by 4 times. The traditional model scaling practice is to arbitrarily increase CNN's depth or width or use a larger input image resolution for training and evaluation. Although these methods improve accuracy, they usually require long manual tuning and often produce sub-optimal performance. Efficient Net proposed a new scaling method that uses a simple and efficient composite coefficient to enlarge CNNs in a more structured way. Unlike traditional methods that arbitrarily scale the network dimensions, such as width, depth, and resolution, Efficient Net uses a series of fixed scale scaling factors to scale the network dimensions uniformly.

The first step of Efficient Net is to perform a grid search to find the relationship between the different scaling dimensions of the baseline network under fixed resource constraints. This determines the appropriate scale factor for each dimension mentioned above. These coefficients are then applied to expand the baseline network to achieve the desired model size or resource requirements. The effectiveness of model extension depends to a large extent on the baseline network. To further improve performance, the author also developed a new baseline network, which optimizes accuracy and efficiency by using the AutoML MNAS framework to perform neural structure search.

D. Methodology

1) Segmentation

The network we proposed contains two parts: DenseNet encoder and U-Net decoder, based on the method proposed by Li X. et al. [30]. The structure of our proposed DenseNet encoder could be seen in table 2. The structure of U-Net decoder we took is consistent with the original paper [16], followed by the final Segmentation, whose structure consists of a layer of 2-D, 16×16 convolutional layer, an identity layer, and an activation layer with the sigmoid function.

To segment our CT-Scan images and extract features, as mentioned above, we proposed a method to combine U-Net design with the concept of denseblock based on H-DenseUNet. Therefore, our network structure has a skeleton based on a U-Net and a DenseNet, which takes the U-Net input encoder. In this structure, we use dilated convolutions of different apertures to form a block and use dense connections between blocks to enhance its feature reuse. Considering the U-net design, to reconstruct detailed features in subsequent layers, the front-end features will be transferred to the corresponding reconstruction layer. Simultaneously, considering the dense connection network, the dense connection within the block will directly transfer the features to the subsequent layers, similar to the feature transfer of U-net. That is to say, feature reuse or reconstruction may have some common reasons to improve the performance of the network.

Table 2. DenseNet Encoder Structure

Encoder	Layers	Output Size	Parameter
DenseNet	Convolution	112×112	$7 \times 7\ conv.stride\ 2$
	Pooling	56×56	$3 \times 3 maxpool, stride 2$
	Dense Block (1)	56×56	$[1 \times 1\ conv, 3 \times 3 conv] \times 6$
	Transition Layers (1)	56×56	$1 \times 1\ conv$
		28×28	$2 \times 2\ averagepool, stride\ 2$
	Dense Block (2)	28×28	$[1 \times 1 conv, 3 \times 3 conv] \times 12$
	Transition Layers (2)	28×28	$1 \times 1\ conv$
		14×14	$2 \times 2\ averagepool, stride\ 2$
	Dense Block (3)	14×14	$[1 \times 1 conv, 3 \times 3 conv] \times 24$
	Transition Layers (3)	14×14	$1 \times 1\ conv$
		7×7	$2 \times 2\ averagepool, stride\ 2$
	Dense Block (4)	7×7	$[1 \times 1 conv, 3 \times 3 conv] \times 16$

Table 3. Parameter of Light Gradient Boosting Machine

n_estimators	30000
leave_size	4
objective	quatile
max_depth	5
num_leaves	7
min_data_in_lead	9
learning_rate	0.008

2) Progression

In the prediction process, we applied Light Gradient Boosting and Efficient Net on the training set. We also applied the five-fold cross-validation to tune hyperparameters for each model to the fullest. The best set of parameters of Light Gradient Boosting Machine could be seen in Table 3. And the parameter of Efficient Net we deployed into this project is the Efficient Net b5 [28], the detailed structure of the prediction network could be seen in Fig 4.Result Analysis

E. Segmentation

These images are comprised by pixels with a Hounsfield Unit (HU) value, respectively. HU is a quantitative scale for radiodensity in the medical area, usually used in CT-Scan images. HU values in CT-Scan images range from -1024 HU to 3071 HU. -1024 HU is defined to represent air. 0 HU represents water. 3071 HU, which is the highest, represents the densest tissue in a human body. HU values for lung range from -600HU TO -700HU.

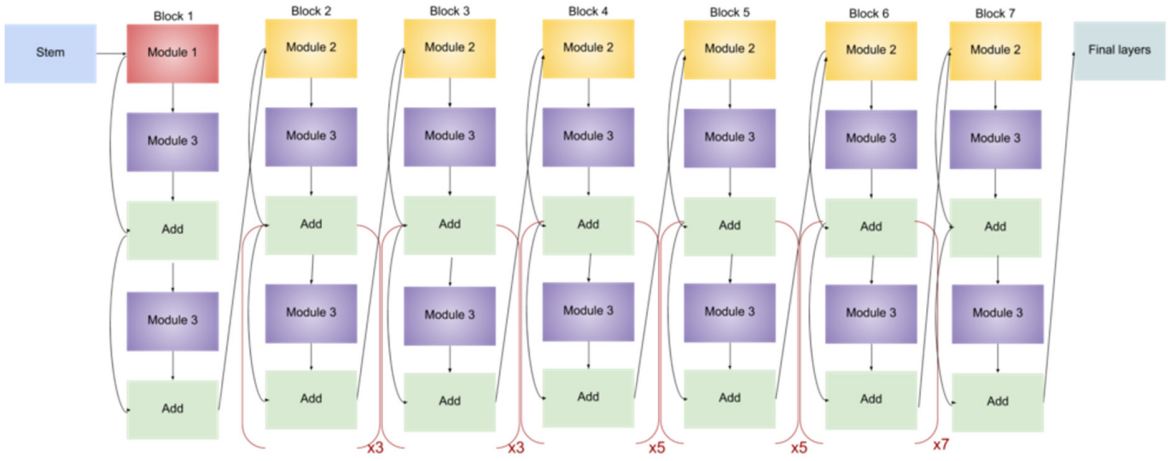

Fig. 4. Structure of Efficient Net b5

Table 4. Extracted Features from A Random Patient

Feature	Description	Value
Mean	Mean of HU values in CT-Scan images	-694.15
Median	Median of HU values in CT-Scan images	-777
Skew	Skewness of HU values in CT-Scan images	1.43
Kurtosis	Kurtosis of HU values in CT-Scan images	1.86
HAA	High-attenuation area, i.e., percentage of lung voxels between -600 and -250 HU	0.97
HU_x^*	Proportion of pixels with HU value $\in (x-100, x]$ in CT-Scan images	0.109

$^*\ x \in \{x\ mod\ 100 = 0\}$, e.g., HU_-900represents the proportion of pixels with HU value $\in (-1000, -900]$ in a CT-Scan image. The sample HU_x value is HU_-600 value of a random patient.

We then extracted further features based on HU values from these segmented lung images, as shown in Table 4. The formula for skewness is shown in Equation (4), where $\bar{x}$ is the mean, s is the standard deviation, and N is the number of data points. The formula for kurtosis is shown in Equation (5), where $\bar{x}$, s, and N has the same meaning as in Equation (4)

$$skewness = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^3 / N}{s^3} \quad (4)$$

$$kurtosis = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4 / N}{s^4} \quad (5)$$

F. Prediction

To ensure our tuned model's robustness and universality, we applied ordinary least squares into the prediction process to visualize the model. Ordinary least squares is mainly used for parameter selection in linear regression. Its idea is very simple. It is to find parameters that minimize the sum of squares of the difference between the actual value and the prediction. The least-square method can quickly obtain proper parameters and can be used for curve fitting. Some other optimization problems can also be expressed by the least square method by minimizing energy or maximizing entropy.

We set the partition value at $0.25, 0.5, 0.75$ to see the correlation of each prediction with the truth value. The visualization of the OLS regression of our model is given in Fig. 7. We could see that Efficient Net b5 has a much more trusted prediction. Based on the model, the final prediction of five-fold cross-validation could be seen in Table \ref{features}, and the final prediction of our purposed metric is at 6.811

G. Prediction

To ensure our tuned model's robustness and universality, we applied ordinary least squares into the prediction process to visualize the model. Ordinary least squares is mainly used for parameter selection in linear regression. Its idea is very simple. It is to find parameters that minimize the sum of squares of the difference between the actual value and the prediction. The least-square method can quickly obtain proper parameters and can be used for curve fitting. Some other optimization problems can also be expressed by the least square method by minimizing energy or maximizing entropy.

We set the partition value at $0.25, 0.5, 0.75$ to see the correlation of each prediction with the truth value. The visualization of the OLS regression of our model is given in Fig. 7. We could see that Efficient Net b5 has a much more trusted prediction. Based on the model, the final prediction of five-fold cross-validation could be seen in Table \ref{features}, and the final prediction of our purposed metric is at 6.811.

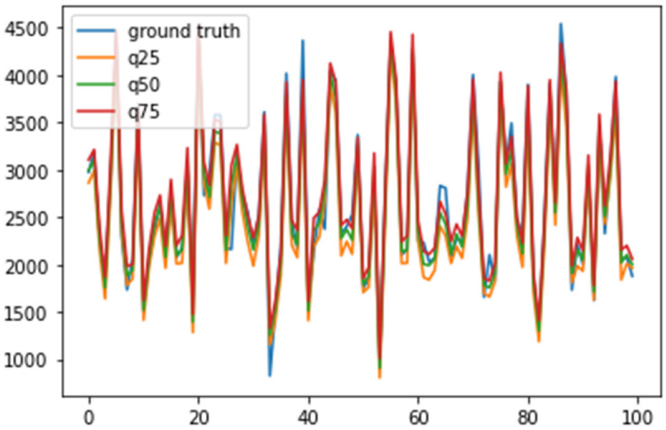

Fig 5. The performance of OLS Regression in Efficient Net

Table 5. Cross Validation Results

Validation_Sets	Train_mean_error	Train_mean_error	Val_mean_error	Val_mean_score
Fold_1	34.28	6.51	40.72	6.83
Fold_2	32.57	6.47	46.94	6.93
Fold_3	31.83	6.43	48.36	6.73
Fold_4	34.81	6.54	35.65	6.56
Fold_5	31.43	6.43	40.44	6.77

III. Discussion & Conclusion

The method we proposed is based on the structure mentioned in Section 2, a state-of-art method for segmentation in medical images. With the performance given in Section 4, we could see the potential performance lies in our proposed method for IPF detection for most patients. And from this perspective, we could see the potential usage for our method in the modern medical perspective.

Our study remains to have several risks. Firstly, the flaws that lie in DenseNet and U-Net are inevitable. The consumption for computational memory is significantly higher than other methods like the Watershed algorithm to encoder the images and

segment the images. With the structure we proposed, the network was conducted by our available computational platform, which could not provide the computational power for hyper tuning on our network.

Secondly, the lack of data could be a perspective of the flaw in our study. The total number of patients in our dataset is 177; even though the data size is up to 22.35GB, the number of patients could be considered as a small dataset. With this truth, we avoided overfitting to some degree by adding the regularization process into the structure. Due to the dataset, a certain degree of overfitting is inescapable.

REFERENCES

[1] C. J. Vyborny and M. L. Giger, "Computer vision and artificial intelligence in mammography," *American Journal of Roentgenology*, vol. 162, no. 3. 1994, doi: 10.2214/ajr.162.3.8109525.

[2] T. F. Cootes and C. J. Taylor, "<title>Statistical models of appearance for medical image analysis and computer vision</title>," in *Medical Imaging 2001: Image Processing*, 2001, vol. 4322, doi: 10.1117/12.431093.[3] J. Thevenot, M. B. Lopez, and A. Hadid, "A Survey on Computer Vision for Assistive Medical Diagnosis from Faces," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 5, 2018, doi: 10.1109/JBHI.2017.2754861.

[4] A. Esteva *et al.*, "Deep learning-enabled medical computer vision," *npj Digital Medicine*, vol. 4, no. 1. 2021, doi: 10.1038/s41746-020-00376-2.

[5] S. B. Kumar, "A Study to Evaluate the Knowledge Regarding Computer Vision Syndrome among Medical Students," *Biomed. Pharmacol. J.*, vol. 13, no. 1, 2020, doi: 10.13005/BPJ/1907.

[6] G. W. Gross, Thomas J and Hunninghake, "Idiopathic pulmonary fibrosis," *N. Engl. J. Med.*, vol. 345, no. 7, pp. 517--525, 2001.

[7] L. Richeldi *et al.*, "Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis," *N. Engl. J. Med.*, vol. 370, no. 22, 2014, doi: 10.1056/nejmoa1402584.

[8] N. Hoyer, T. S. Prior, E. Bendstrup, T. Wilcke, and S. B. Shaker, "Risk factors for diagnostic delay in idiopathic pulmonary fibrosis," *Respir. Res.*, vol. 20, no. 1, 2019, doi: 10.1186/s12931-019-1076-0.

[9] D. A. Lynch *et al.*, "High-resolution computed tomography in idiopathic pulmonary fibrosis: Diagnosis and prognosis," *Am. J. Respir. Crit. Care Med.*, vol. 172, no. 4, 2005, doi: 10.1164/rccm.200412-1756OC.

[10] G. Raghu *et al.*, "Diagnosis of idiopathic pulmonary fibrosis with high-resolution CT in patients with little or no radiological evidence of honeycombing: Secondary analysis of a randomised, controlled trial," *Lancet Respir. Med.*, vol. 2, no. 4, 2014, doi: 10.1016/S2213-2600(14)70011-6.

[11] J. Jacob *et al.*, "Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures," *Eur. Respir. J.*, vol. 49, no. 1, 2017, doi: 10.1183/13993003.01011-2016.

[12] H. Ohkubo, H. Nakagawa, and A. Niimi, "Computer-based quantitative computed tomography image analysis in idiopathic pulmonary fibrosis: A mini review," *Respiratory Investigation*, vol. 56, no. 1. 2018, doi: 10.1016/j.resinv.2017.10.003.

[13] H. Robbie *et al.*, "Visual and automated CT measurements of lung volume loss in idiopathic pulmonary fibrosis," *Am. J. Roentgenol.*, vol. 213, no. 2, 2019, doi: 10.2214/AJR.18.20884.

[14] A. Christe *et al.*, "Computer-Aided Diagnosis of Pulmonary Fibrosis Using Deep Learning and CT Images," *Invest. Radiol.*, vol. 54, no. 10, 2019, doi: 10.1097/RLI.0000000000000574.

[15] N. Cesa-Bianchi, A. Conconi, and C. Gentile, "On the generalization ability of on-line learning algorithms," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, 2004, doi: 10.1109/TIT.2004.833339.

[16] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, no. 2, 2017, doi: 10.1148/rg.2017160130.

[17] E. Santermans *et al.*, "Modelling Forced Vital Capacity in Idiopathic Pulmonary Fibrosis: Optimising Trial Design," *Adv. Ther.*, vol. 36, no. 11, 2019, doi: 10.1007/s12325-019-01093-3.

[18] D. A. Lynch *et al.*, "Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper," *The Lancet Respiratory Medicine*, vol. 6, no. 2. 2018, doi: 10.1016/S2213-2600(17)30433-2.

[19] R. Shojaii, J. Alirezaie, and P. Babyn, "Automatic lung segmentation in CT images using watershed transform," in *Proceedings - International Conference on Image Processing, ICIP*, 2005, vol. 2, doi: 10.1109/ICIP.2005.1530294.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351, doi: 10.1007/978-3-319-24574-4_28.

[21] "IEEE International Symposium on Biomedical Imaging (ISBI)," *IEEE Trans. Med. Imaging*, vol. 34, no. 3, 2015, doi: 10.1109/tmi.2015.2407111.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December, doi: 10.1109/CVPR.2016.90.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January, doi: 10.1109/CVPR.2017.243.

[24] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, vol. 13-17-August-2016, doi: 10.1145/2939672.2939785.

[25] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, 2001, doi: 10.1214/aos/1013203451.

[26] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, 2001, doi: 10.1023/A:1010933404324.

[27] A. Liaw and M. Wiener, "Classification and Regression with Random Forest," *R News*, vol. 2, 2002.

[28] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *36th International Conference on Machine Learning, ICML 2019*, 2019, vol. 2019-June.

[29] Y. Huang *et al.*, "GPipe: Efficient training of giant neural networks using pipeline parallelism," in *Advances in Neural Information Processing Systems*, 2019, vol. 32.

[30] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes," *IEEE Trans. Med. Imaging*, vol. 37, no. 12, 2018, doi: 10.1109/TMI.2018.2845918.