Deep Neural Network and Multiple Quantile Regression for Prediction of Pulmonary Fibrosis Progression

Abstract

Pulmonary fibrosis is a chronic lung disease that has no known cause. This disease can cause permanent scarring and damage of the lung tissue. It can range from a decline in lung capacity over a period of time or rapidly.There is no such concrete cure for this disease as of now. However, it can be efficiently managed by understanding how the computed tomography (CT) scan can give insight into lung damage. The motivation behind this project is to find a way to detect the extent of decline of lung damage of the patients so that it could be understood better and accelerate immediate action.

In this project, we try to solve the Kaggle challenge: OSIC Pulmonary Fibrosis [1]. The challenge is to predict future week's Forced Vital Capacity(FVC) which is measured using the device spirometer and confidence value. We leverage exploratory data analysis to help us understand the data better. We propose an ensemble architecture consisting of ResNet and Multiple Quantile Regression on a dense neural network which predicts FVC value based on the clinical metadata, the CT scan image and the initial spirometry measurement. After measuring the modified Laplace Log Likelihood score we find that our proposed framework gives a score better than the winning solution: [2] .

1. Introduction

In this project our problem statement is to predict a patient's severity of decline in lung function based on a CT scan of their lungs. A spirometer will help us determine the lung decay and its severity. The spirometer helps us by measuring the volume of air inhaled and exhaled. Our challenge in this project is to use machine learning and computer vision techniques to make a prediction by combining the dataset of the image, metadata and the baseline FVC as the input.

If the model is successful, patients and their families would better understand their prognosis when they are first diagnosed with this incurable lung disease. This detection of the severity of the disease will also impact the treatment design of Pulmonary Fibrorsis and accelerate the clinical development in this field. The major step in assessment of Pulmonary Fibrosis to understand how to measure the lung function decline and take immediate action for its treatment. Some of the limited treatments that act as a temporary cure include oxygen therapy and pulmonary rehabilitation to pharmacological agents (pirfenidone [15] and nintedanib [11]) and lung transplantation [9].Other medical techniques such as Surgical Lung Biopsy can also be used in order to cure Pulmonary Fibrosis. However, this treatment possess major risk to the health and lungs of the patients.

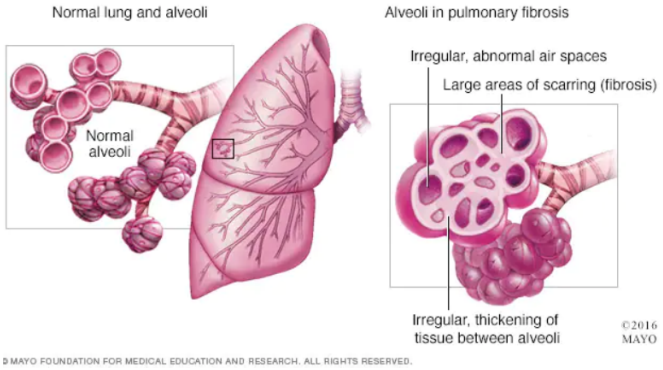


Figure 1: **Pulmonary Fibrosis**

In [12], the authors of the paper talk about other solutions for decline in lung cancer and how CT Scans can tell us a lot about how to surgically treat Pulmonary Fibrosis. They say that a high-resolution computed tomography scan pattern of probable UIP(usual interstitial pneumonia), indeterminate, or an alternative diagnosis, conditional recommendations were made for performing BAL(Bronchoalveolar lavage)- a diagnostic test and surgical lung biopsy; because of lack of evidence, no recommendation was made for or against performing transbronchial lung biopsy or lung cryobiopsy. In contrast, for patients with newly detected interstitial lung disease who have a high-resolution computed tomography scan pattern of UIP, strong recommendations were made against performing surgical lung biopsy, transbronchial lung biopsy, and lung cryobiopsy, and a conditional recommendation was made against performing BAL.

Sometimes, physical tests are also a good way of telling

Dhuri Shrivastava, Sakshi Paresh Doshi

us if the patient has pulmonary fibrosis. Hence authors in papers [16][14][5] enhance the importance of spirometry tests to understand the measurement of the FVC of the lung. However, one of the most optimal ways to measure the lung function decline is by analysing Computed Tomography Imaging. The major challenge in these types of Images is to properly predict the progression of the disease. The reason why this becomes so inaccurate is because different people have variable rate of progression. In this study, they try to use exploratory data anaylsis to help us understand and formulate a strong model which gives important insights for Lung function Analysis. This model gives accurate prediction of forced vital capacity (FVC) based on a patient's CT scan, initial spirometry measurement, and clinical metadata. We used exploratory data analysis to derive insights into our data which helped us create a network that predicts FVC values.

In this paper the organization is as follows. The related work section gives us a background of the algorithms that we used such as ResNet and Quantile Regression. It also talks about different tools in AI which we can use for more analysis on Pulmonary Fibrosis. The Approach sections provides a detailed description about our framework and architecture. It contributes to our major understanding of the project. The Experiment section tells us more about our Dataset, Evaluation Metric, Exploratory Data Analysis and our results. The Conclusion discusses about the broader impact of the proposed solution to the challenge and our future work.

2. Background/Related Work

The Open Source Imaging Consortium (OSIC) Pulmonary Fibrosis Kaggle Challenge created awareness regarding lung function decline in the Data Science and Artificial Intelligence Community. This motivated thousands of enthusiasts to come together and form solutions to this intricate problem. In this section we will talk about some solutions to the Pulmonary Fibrosis Progression problem and different neural networks proposed to predict lung decline. We will also discuss about the methods we are using.

Amongst so many participants that really tried their best at this competition, the winner of the Kaggle Competition used a deep learning weighted ensemble of Convolution Neural Networks. The model took benefit of the State of the Art model of EfficientNet B5 along with a multiple Quantile Regressor[2] to predict the lung function decline of the patient with patient's CT scans, initial spirometry measurement, and clinical metadata as the input.

Besides the above project, another paper that highlighted lung decline function using deep learning was "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network" [7]. In this paper, the authors talk about using Convolutional Neural Networks to visualize 2D patches from the given image cohort(Figure 2). Their deep learning architechture consists of 5 convolutional layers with 2×2 kernels and LeakyReLU activations, which is followed by average pooling with size equal to the size of the final feature maps and three dense layers. The last dense layer has 7 outputs, equivalent to the classes considered: healthy, ground glass opacity (GGO), micronodules, consolidation, reticulation, honeycombing and a combination of GGO/reticulation. Such classification from the above method makes it easy to identify what kind of damage is occurring in the patient's lungs.

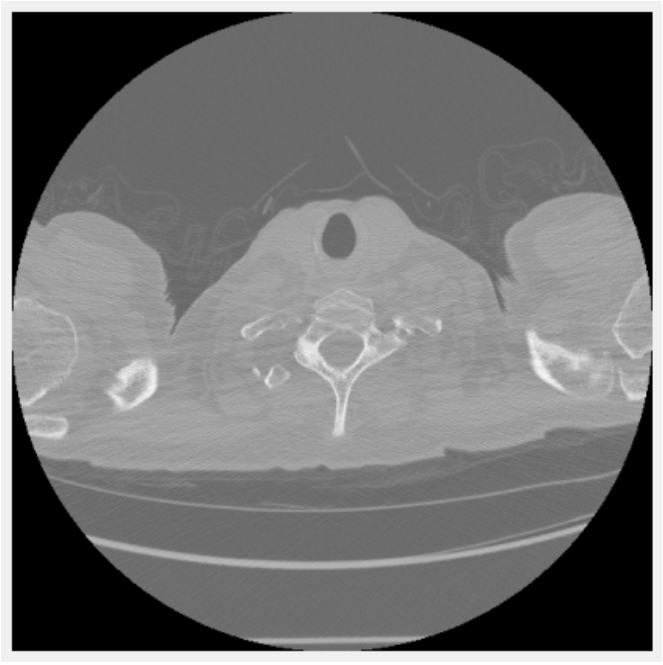


Figure 2: **A CT Scan Image from the OSIC Cohort**

The same authors from the above paper then introduce a new improved and advanced methodology [6]for this problem statement that talks about transferring knowledge from the similar domain of general texture classification. In this method, they use Six publicly available texture databases to pretrain networks with the proposed architecture, which are then fine-tuned on the lung tissue data. The resulting CNNs are combined in an ensemble and their fused knowledge is compressed back to a network with the original architecture. The proposed approach resulted in an absolute increase of about 2 percent in the performance of the proposed CNN.

In the paper "Classification of Interstitial Lung Abnormality Patterns with an Ensemble of Deep Convolutional Neural Networks" [4] the authors talks about how they use an ensemble of deep convolutional neural networks (CNNs) that detect more discriminative features by incorporating two, two-and-a-half and three- dimensional architectures. This helps the model to make more accurate classification.

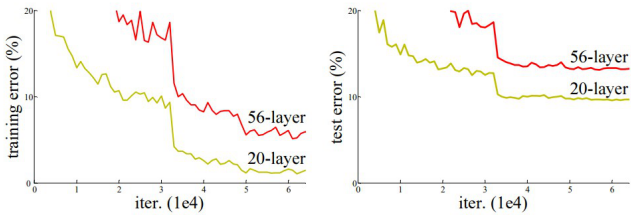


Figure 3: **Training and Test Error for deep layers [8]**

This technique is implemented by first training each individual CNN, and then combining its output responses to form the overall ensemble output. Hence the work of these authors shows that the proposed method is a viable solution to identify radiographic patterns that precede the development of ILD.

We decided to use a deep CNN to extract features from the image. As we make a deeper network the complexity of the model increases. It has been found that there is a maximum threshold for depth with the traditional Convolutional neural network model(Figure 3). As we can see in the figure that 56 layer has more error than 20 layer network. This could be for variety reasons like, optimization function, initialization of the network and vanishing gradient problem. The solution for this problem is a ResNet CNN architecture which contains residual blocks to skip connections in addition to heavy batch normalization. [8]. This identity mapping does not have any parameters and is just there to add the output from the previous layer to the layer ahead(Fig 4). The skip connection provides a shortcut for the gradient to flow and also the identity function makes sure that the higher layer perform as well as lower layers and not worse.

To improve our model's performance we added Quantile Multiple Regressor. Quantile regression models the relationship between a set of predictor (independent) variables and specific percentiles (or "quantiles") of a target (dependent) variable, most often the median [13]. Unlike linear regression, quantile regression does not make any assumptions about the distribution of residuals. Quantile regression uses pinball loss known as quantile loss. Quantile regression estimates any quantile or percentile value of the response value. A quantile is the value below which a fraction of observations in a group falls.

3. Approach

We plan to approach this problem by first performing some exploratory data analysis. This will give us an idea on how the data actually looks and what kind of techniques can be used in order for us to get better prediction. We did data analysis on both the types of dataset that we received from Kaggle.

The observations that helped us form this approach are:

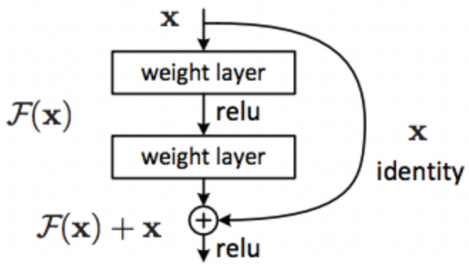


Figure 4: **Residual Block [8]**

- There are no Null values in the Training or the Testing Dataset
- The total patient ids are 1549, from those the unique ids are 176.
- There are total 33,026 images of 176 patients
- Hence we have 187 average images per patient and 1,018 max files per patient
- In the given dataset, number of Ex-smokers are 118 ;

  Number of people who never smoked are 49 and those who currently smoke are 9

We first extract the image and meta data from the dicom files. Our approach is a simple ensemble model of two techniques, ResNet and Multiple Quantile Regressor. We started with the ResNet model to predict the linear decay coefficient which in turn predicts FVC values. Inspired by the winning notebook of the competition [2] which uses EfficientNet to calculate the coefficient for lung decay and blends it with another model which is the Multiple Quantile Regressor, we also added quantile regression model to our ensemble.

3.1. ResNet

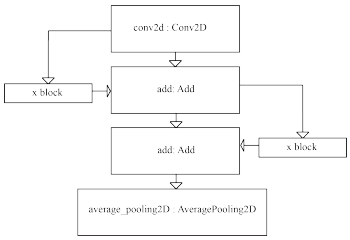
The decay in the lungs is given by:

$$FVC = a.quantile(0.75) * (week - week_{test}) + FVC_{test}$$

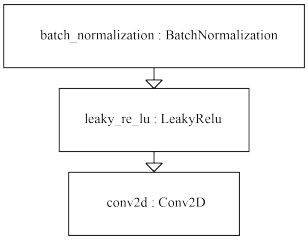
$$Confidence = Percent + a.quantile(0.75)$$

$$* abs(week - week_{test})$$

We are predicting the lung decay by predicting the coefficient 'a'. The aim is to find the FVC of the patient in any given week which can be computed using the formual given. Here we are assuming a linear relationship, with this assumption we build our model and get good results.To compute the coefficient 'a', our model is combining two types of inputs, tabular data and CT scan image with it's meta data. We have used a ResNet architecture for this



(a) **Convolution Block**



(b) **x Block**

Figure 5: **Convolutional and Residual Block**

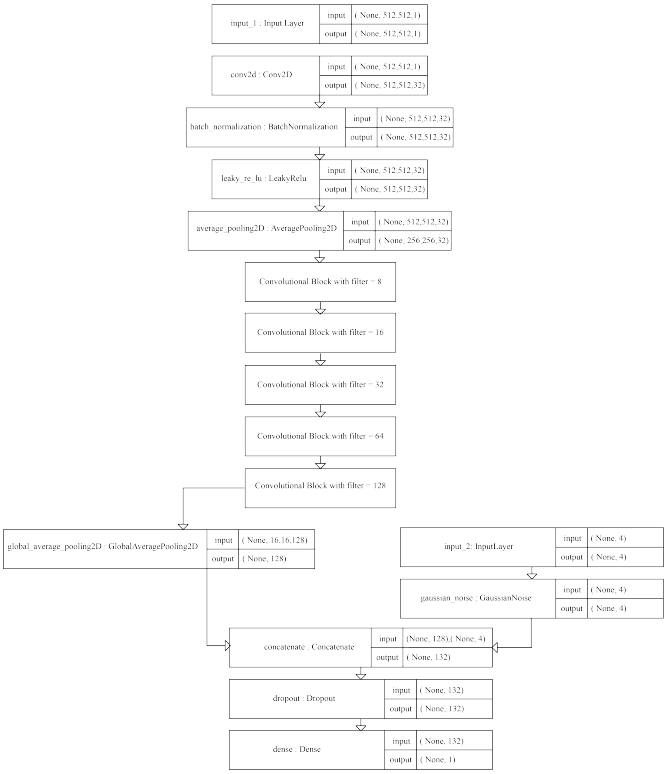


Figure 6: **ResNet Architecture**

part. Figure 6 shows our network architecture. In this diagram, every convolutional block consists of a convolutional layer,Batchnormalization,Leaky ReLU, convolutional layer with residual identity mapping. Figure 5(a) shows the convolutional block.

Giving a smaller weight decay gives better results [3]. L2 kernel regularizer with regularization hyperparameter 4e-5 gave us better results. We have trained the model with Adam Optimizer for 30 epochs with learning rate of 0.03. The optimizer updates on the basis of mean absolute error loss.

3.2. Multiple Quantile Regressor

Quantile regression can be applied to neural networks, instead of the mean squared error used in linear regression we use quantile(pinball) loss function. We have built a dense neural network. The architecture of our model is given in Figure 7.

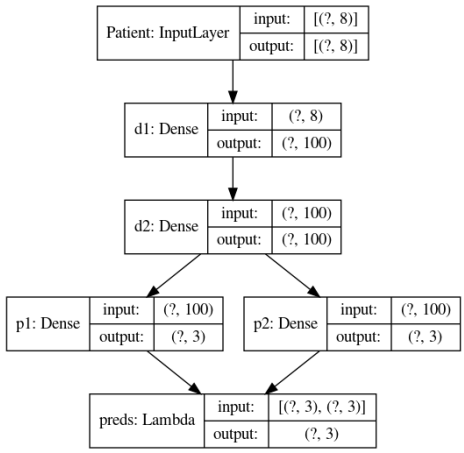


Figure 7: **Dense Neural Network Architecture**

The input to this model was tabular data only. The model learns using a combination of pinball loss and metric score defined by the challenge. The model learns 20th,50th and 80th quantile. The loss function defined by us takes 80 percent from quantile loss and 20 percent from the score metric. We use Adam Optimizer with learning rate 0.3 and run the model for 600 epochs. The equation for quantile regression is given by:
$\hat{f} = \text{argmin}\, f \in \mathrm{F}(\alpha, \beta) \sum i = 1^{n} \rho_{\tau}\left(y_i - f\left(x_i; \alpha, \beta\right)\right) + J_{\lambda}(f)$

The equation for multiple quantile regression is given by [10]:
$\left(\overline{f_{\tau_1}}, \ldots, \overline{f_{\tau_1}}\right) =$
$\text{argmin}\, f\tau_1, \ldots, f_{\tau_r} \in \mathrm{F}(\alpha, \beta) \sum_{t=1}^{r} \sum_{i=1}^{n} \rho_{\tau_t}\left(y_i - f_{\tau_t}\left(x_i\right)\right)$

3.3. Simple Ensemble

As both these models take different kinds of data, we have taken a simple ensemble of the two models. We are taking the weighted average of the FVC values computed using ResNet and Multiple Quantile Regression. After trying different values for the weights, we got the best results with 60 percent score from Quantile Regression model and 40 percent from ResNet.

4. Experiment

4.1. Dataset

For this project, we will use the dataset provided to us on Kaggle. This dataset can be found over here: https://www.kaggle.com/c/osic-pulmonary-fibrosis-progression/data.

The Open Source Imaging Consortium (OSIC) has provided the dataset for this kaggle challenge. The dataset is divided into two parts:

- Tabular data which contain the patient's clinical history i.e the forced vital capacity(FVC) which is the volume of air exhaled and its progression through time, this information is extracted from the output from a spirometer. People with pulmonary fibrosis have increasing level of fibrotic tissue which decrease their FVC.Our main aim is to calculate the decline in FVC value through time. Another column is week which gives the week when the FVC was recorded with respect to the week when CT scan was taken. Example: week -2 refers to 2 weeks before the CT scan. The other two columns are age, smoking status and sex of the patient. The final column shows percent FVC that a person has compared to a healthy person. Figure 8 shows some rows of the training csv file. There are a total of 176 patients in the training set with a total of 1549 datapoints. The "test" csv is just a representation of how the data will look like but the whole test data is not visible to us and we only get the score after submitting our model in Kaggle. The test csv has initial week data and we are required to predict the FVC values for a given week.

Training data shape: (1549, 7)

	Patient	Weeks	FVC	Percent	Age	Sex	SmokingStatus
0	ID00007637202177411956430	-4	2315	58.253649	79	Male	Ex-smoker
1	ID00007637202177411956430	5	2214	55.712129	79	Male	Ex-smoker
2	ID00007637202177411956430	7	2061	51.862104	79	Male	Ex-smoker
3	ID00007637202177411956430	9	2144	53.950679	79	Male	Ex-smoker
4	ID00007637202177411956430	11	2069	52.063412	79	Male	Ex-smoker

Figure 8: **Training Data**

- The second part contains patient's CT scan which is in Dicom format and is a 3D image with metadata. Each patient's complete CT scan at week 0 is given in the dataset. There are multiple diacom files for each patient which contains slices of CT scan. There is a huge imbalance between the number of CT scan per patient. There are total 33,026 images of 176 patient.Hence we have 187 average images per patient and 1,018 max files per patient. Due to this reason and computational restrictions we have only incorporated one CT scan per patient. Figure 9 shows the CT scan images.

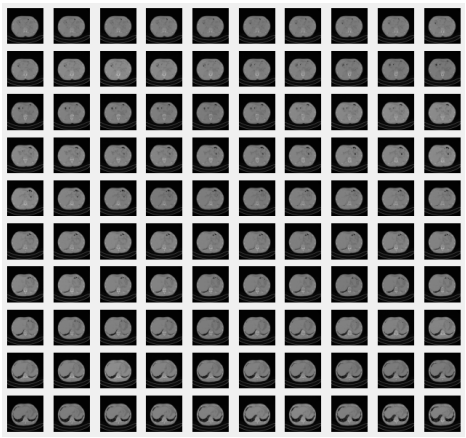


Figure 9: **A grid of CT Scans (10x10)**

4.2. Evaluation Metric

We use a modified version of the Laplace Log Likelihood. In medical applications, it is useful to evaluate a model's confidence in its decisions. Accordingly, the metric is designed to reflect both the accuracy and certainty of each prediction. For each true FVC measurement, we will predict both an FVC and a confidence measure. The metric is computed as:

$$\sigma_{clipped} = max(\sigma, 70),$$
$$\Delta = min(|FVC_{true} - FVC_{predicted}|, 1000),$$
$$metric = -\frac{\sqrt{2}\Delta}{\sigma_{clipped}} - \ln(\sqrt{2}\sigma_{clipped}).$$

The confidence value gives us the uncertainity in our prediction(standard deviation). The uncertainity plays an important role in the scoring of our model hence we needed a loss which gives it equal importance as the predicted value. We used a pinball loss for this. Pinball loss in a combination with the metric score is the loss function for quantile regression. The formula for quantile loss is:

$$\mathrm{L}(y_i^p, y_i) = \max\left[q\left(y_i^p - y_i\right), (q-1)\left(y_i^p - y_i\right)\right]$$

Method	Laplace Log Likelihood Score
Kaggle 1st place	-6.8305
Kaggle 2nd place	-6.8311
Kaggle 3rd place	-6.8336
ResNet	-6.8971
ResNet+Quantile Regression	**-6.8303**

Table 1: Comparison of Laplace Log Likelihood scores for the Kaggle winning methods and our proposed methods

Model	Total Parameters
Kaggle 1st place(EfficientNet B5)	28,514,709
ResNet	691,021

Table 2: Comparison of total number of parameters

	Patient_Week	FVC	Confidence
0	ID00419637202311204720264_-12	3083.362912	133.549767
1	ID00421637202311550012437_-12	2880.289970	223.335261
2	ID00422637202311677017371_-12	2046.809969	193.482462
3	ID00423637202312137826377_-12	3517.852148	303.111051
4	ID00426637202313170790466_-12	2994.186379	141.011347

Figure 10: **A snippet of Submission.csv**

4.3. Results

Our results will be stores in file called "Submission.csv". Figure 10 shows us how a sample submission file will look like. It has three columns Patient ID with the week, FVC value and confidence in that value. We evaluate our results quantitatively. The Laplace Log Liikelihood score is calculated. We submitted our notebook to the kaggle competition and received a private score of -6.8303 which is better than the score of the winning solution. The winner of the competition has the score of -6.8305 (Table 1). Our method has fewer parameters than the winning model (Table 2).

5. Conclusion and Future Work

We provide a method to predict the FVC values with a model which is not computationally that expensive. The model can help the patients to better understand the disease and see it's pattern. With the input of medical community, the proposed methods can be used to identify patients at risk of pulmonary fibrosis. Our work can be extended to provide deep learning solution for different clinical scenarios like examining CT scans for pneumonia,Covid-19 etc . Although the method gives decent results, we can not rely on it completely for clinical assessment. It gives a base to further research with the help of medical professionals. Due to the computation restrictions we could not try deeper models and had to restrict our number of layers. Our future work will be to try deeper networks. We also want to explore ResNet-RS which improves training and scaling strategies to make ResNet faster than EfficientNet and improved performance over ResNet.[3]

References

[1] Osic pulmonary fibrosis progression. https://www.kaggle.com/c/osic-pulmonary-fibrosis-progression. 1

[2] Artkulak. Inference 45/55 600 epochs tuned effnet b5 30 ep. https://www.kaggle.com/artkulak/inference-45-55-600-epochs-tuned-effnet-b5-30-ep, Oct 2020. 1, 2, 3

[3] Irwan Bello, William Fedus, Xianzhi Du, Ekin D. Cubuk, Aravind Srinivas, Tsung-Yi Lin, Jonathon Shlens, and Barret Zoph. Revisiting resnets: Improved training and scaling strategies, 2021. 4, 6

[4] David Bermejo-Peláez, Samuel Ash, George Washko, Raúl Estépar, and María Ledesma-Carbayo. Classification of interstitial lung abnormality patterns with an ensemble of deep convolutional neural networks. *Scientific Reports*, 10:338, 01 2020. 2

[5] Roland Bois, Derek Weycker, Carlo Albera, Williamson Bradford, Ulrich Costabel, Alex Kartashov, Talmadge King, Lisa Lancaster, Paul Noble, Steven Sahn, Michiel Thomeer, Dominique Valeyre, and Athol Wells. Forced vital capacity in patients with idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*, 184:1382–9, 09 2011. 2

[6] Stergios Christodoulidis, Marios Anthimopoulos, Lukas Ebner, Andreas Christe, and Stavroula Mougiakakou. Multisource transfer learning with convolutional neural networks for lung pattern analysis. *IEEE Journal of Biomedical and Health Informatics*, PP:1–1, 12 2016. 2

[7] Stergios Christodoulidis, Marios Anthimopoulos, Lukas Ebner, Andreas Christe, and Stavroula Mougiakakou. Multisource transfer learning with convolutional neural networks for lung pattern analysis. *IEEE Journal of Biomedical and Health Informatics*, 21(1):76–84, 2017. 2

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 3

[9] Kristin Kistler, Luba Nalysnyk, Philip Rotella, and Dirk Esser. Lung transplantation in idiopathic pulmonary fibrosis: A systematic review of the literature. *BMC pulmonary medicine*, 14:139, 08 2014. 1

[10] Sampurna Mandal, Valentina Balas, Rabindra Shaw, and Ankush Ghosh. Prediction analysis of idiopathic pulmonary fibrosis progression from osic dataset. pages 861–865, 10 2020. 4

[11] Luca Richeldi, Roland Bois, Ganesh Raghu, Arata Azuma, Kevin Brown, Ulrich Costabel, Vincent Cottin, Kevin Flaherty, David Hansell, Yoshikazu Inoue, Dong Soon Kim,

Martin Kolb, Andrew Nicholson, Paul Noble, Moisés Selman, Hiroyuki Taniguchi, Michèle Brun, Florence Le Maulf, Mannaïg Girard, and M. Luisetti. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *New England Journal of Medicine*, 370:2071–82, 05 2014. 1

[12] Luca Richeldi, Ganesh Raghu, Martine Remy-Jardin, Jeffrey L. Myers, Christopher J. Ryerson, David J. Lederer, Juergen Behr, Vincent Cottin, Sonye K. Danoff, Ferran Morell, Kevin R. Flaherty, Athol Wells, Fernando J. Martinez, Arata Azuma, Thomas J. Bice, Demosthenes Bouros, Kevin K. Brown, Harold R. Collard, Abhijit Duggal, Liam Galvin, Yoshikazu Inoue, R. Gisli Jenkins, Takeshi Johkoh, Ella A. Kazerooni, Masanori Kitaichi, Shandra L. Knight, George Mansour, Andrew G. Nicholson, Sudhakar N.J. Pipavath, Ivette Buendía-Roldán, Moisés Selman, William D. Travis, Simon Walsh, and Kevin C. Wilson. Diagnosis of idiopathic pulmonary fibrosis an official ats/ers/jrs/alat clinical practice guideline. *American Journal of Respiratory and Critical Care Medicine*, 198:e44–e68, 2018. 1

[13] Robert N. Rodriguez. Five things you should know about quantile regression. 2017. 3

[14] Anne-Marie Russell, Huzaifa Adamali, Philip Molyneaux, Pauline Lukey, Richard Marshall, Elisabetta Renzoni, Athol Wells, and Toby Maher. Daily home spirometry: An effective tool for detecting progression in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 194, 04 2016. 2

[15] H. Taniguchi, M. Ebina, Y. Kondoh, T. Ogura, A. Azuma, M. Suga, Y. Taguchie, H. Takahashi, K. Nakata, A. Sato, M. Takeuchi, G. Raghu, S. Kudoh, T. Nukiwa, T. Betsuyaku, Y. Sugawara, S. Fujiuchi, K. Yamauchi, K. Konishi, M. Munakata, Y. Kimura, Y. Ishii, Y. Sugiyama, K. Kudoh, T. Saito, T. Yamaguchi, A. Mizoo, A. Nagai, A. Ishizaka, K. Yamaguchi, K. Yoshimura, M. Oritsu, Y. Fukuchi, K. Takahashi, K. Kimura, Y. Yoshizawa, T. Nagase, T. Hisada, K. Ohta, K. Yoshimori, Y. Miyazawa, K. Tatsumi, Y. Sasaki, M. Taniguchi, Y. Sugita, E. Suzuki, Y. Saito, H. Nakamura, K. Chida, N. Kasamatsu, H. Hayakawa, K. Yasuda, H. Suganuma, H. Genma, R. Tamura, T. Shirai, J. Shindoh, S. Sato, O. Taguchi, Y. Sasaki, H. Ibata, M. Yasui, Y. Nakano, M. Ito, S. Kitada, H. Kimura, Y. Inoue, H. Yasuba, Y. Mochizuki, S. Horikawa, Y. Suzuki, N. Katakami, Y. Tanimoto, Y. Hitsuda, N. Burioka, T. Sato, N. Kohno, A. Yokoyama, Y. Nishioka, N. Ueda, K. Kuwano, K. Watanabe, H. Aizawa, S. Kohno, H. Mukae, H. Kohrogi, J. Kadota, I. Tokimatsu, E. Miyazaki, T. Sasaki, and M. Kawabata. Pirfenidone in idiopathic pulmonary fibrosis. *Scandinavian Journal of Respiratory Diseases*, 35(4):821–829, Apr. 2010. 1

[16] L. C. Watters, T. E. King, M. I. Schwarz, J. A. Waldron, R. E. Stanford, and R. M. Cherniack. A clinical, radiographic, and physiologic scoring system for the longitudinal assessment of patients with idiopathic pulmonary fibrosis. *American Review of Respiratory Disease*, 133(1):97–103, 1986. 2