International Journal of Research in Engineering, Science and Management
Volume 5, Issue 4, April 2022
https://www.ijresm.com | ISSN (Online): 2581-5792

A Survey on the Approaches to Detect Pulmonary Fibrosis

Pranav Pradeep[1], H. S. Mansi[2], Likitha Keerthi[3*], Dev Narayanan[4], K. Sumithra Devi[5]
[1,2,3,4]Student, Department of Information Science and Engineering, Dayanand Sagar Academy of Technology and Management, Bengaluru, India
[5]Professor & HoD, Department of Information Science and Engineering, Dayanand Sagar Academy of Technology and Management, Bengaluru, India

***Abstract*: Pulmonary fibrosis is an incurable, fatal, and debilitating disease that damages the patient's respiratory system, making it difficult to live with. Despite the fact that the situation appears to be hopeless, modern medicine can help to postpone the disease's prognosis. The ability of the doctor to determine the severity of the sickness becomes critical for appropriate therapy, yet this is a highly risky decision. We suggest a unique way to address this bottleneck problem by constructing a system that can accurately anticipate disease progression for a given week by measuring the patient's FVC value. This saves the pulmonologist time and effort while potentially extending a person's life. The suggested approach predicts FVC output for a given week by combining image and tabular data.**

***Keywords*: Machine Learning, Deep Learning, CNN, vanilla quantile regression, image augmentation, SVR, efficientNet-b3, resnet, adam optimizer.**

1. Introduction

Pulmonary Fibrosis is a variant/form of revolutionary lung contamination that belongs to the interstitial lung illness own circle of relatives of disorders. Idiopathic pulmonary fibrosis (IPF) is the maximum common form of Pulmonary Fibrosis, and its etiology is unclear. Pulmonary fibrosis is as a result of repeated lesions to the alveolar epithelium or endothelium, which reasons the innate and adaptive immune structures to reconstruct the tissue structure of the wounded tissue. Pulmonary Fibrosis is a result of a lot of factors. Long-time period publicity to environmental and occupational risks along with asbestos, silica, coal dust, beryllium, tough metals, and radiation remedies are just a few of the recognized reasons. Lung scarring can also be as a result of gastrointestinal flux disorder, autoimmune disorders (along with rheumatoid arthritis and systemic lupus erythematosus), sarcoidosis, and diverse muscle diseases (along with dermatomyositis, polymyositis, and the anti-synthetase syndrome).

Deep Neural Network refers to Artificial Neural Networks (ANN) with multi layers. One of the maximum broadly used deep neural networks is the convolutional neural network (CNN). Convolutional layer, non-linearity layer, pooling layer, and completely related layer are a number of the layers of CNN. CNN plays properly in programs which include huge photograph class dataset (photograph net), laptop vision, and in herbal language processing (NLP). The cause of this undertaking is to increase a utility that makes use of tabular statistics containing affected person statistics and FVC values with CT scans of the sufferers to expect the destiny FVC values, such that in addition decline in lung feature may be predicted and moves may be taken accordingly.

2. Literature Survey

M. Anthimopoulos, S. Christodoulidis et al. [6] considered a dataset made using 2 databases. One of the databases consists of 109 HRCT scans with 512x512 pixels per slice. The second database comprises 26 512x512 resolution HRCT scans. The CNN includes 5 convolutional layers along with 2x2 kernels [6], leaky ReLu [6], which is a variant of ReLu, was used for activating every convolutional layer. The training of the artificial neural network is done by combining a loss function and an optimization algorithm. This study uses adam optimizer [6] to reduce categorical cross entropy. Average pooling is performed with a size equal to the final feature map and three dense layers. The last thick layer has 7 exits. The classification performance is about 85.5% [6], indicating the possibility of CNN in the analysis of lung patterns. Many parameters are the drawbacks of this model. Relatively slow training times and slight variations are the drawbacks of this approach.

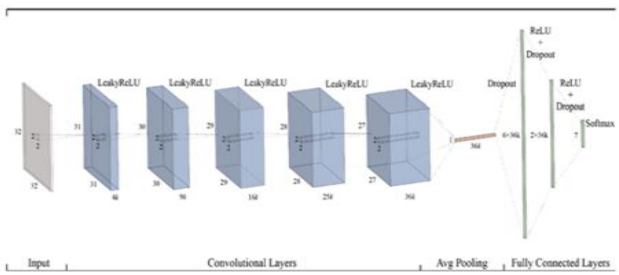

Fig. 1. Proposed model for ILD classification

Chenshuo Wang, XianXiang Chen et al. [5] considered a dataset consisting of 5 variables: Age, Gender, FEV1, PEF and FVC.FVC was the target feature. Rbf SVR algorithm was used for predicting FVC values. The prediction model was built using the data of 354 participants who underwent traditional spirometry. [5] 3 prediction models were established which are mixed model, normal model and abnormal model. To validate

*Corresponding author: likithakeerthi47@gmail.com

the performance of the developed prediction model, data from 35 subjects evaluated with both standard spirometers and low-grade spirometers were used. Data from 35 subjects characterized by both conventional spirometers and low-standard spirometers were used. [5] The performance of the normal and abnormal models in FVC predictions was excellent. The FEV1 / FVC ratio cutoff in this study was 0.70, limiting flexibility in identifying airflow boundaries. The number of training samples with an FEV1 / FVC ratio greater than 0.70 is 82, significantly less than the 272 training samples with an FEV1 / FVC ratio of less than 0.70. As a result, the training set for the mixed model becomes imbalanced.

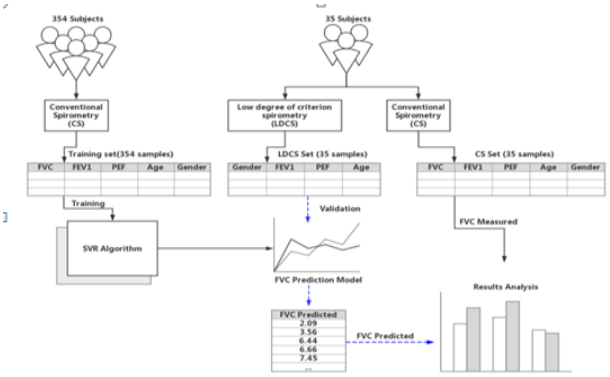

Fig. 2. Model construction procedure for FVC prediction

Z. Zhang et al. [2] recommended a model which gives better generalized performance than that of Adam. For variants of SGD refer [2]. The normalized directional storage Adam (ND-Adam) algorithm more accurately specifies the updated direction and step size. The ND-Adam algorithm, which improves the optimization of input weights for hidden units, is the normalized directional storage Adam (ND-Adam) algorithm, while the vanilla Adam algorithm is used to update other parameters. To bridge the gap between Adam and SGD in terms of generalization, ND-Adam is Adam's tailor-made implementation for training DNN. ND-Adam is designed to maintain the gradient direction of each weight vector and generate the L2 weight drop regularization effect in an accurate and efficient way. This works with the CIPHER100 dataset and shows an error rate of 18.48. The main drawback of this method is the problem of weight decay.

Q. Wang et al. [3] developed a method for explicitly modeling the relationships among nodule properties using transfer learning which is discussed in the paper. In computed tomography (CT) images, texture, lobulation, malignancy, and other characteristics etc., can provide useful information for identifying early lung cancer. The proposed method has been tested using a set of 2632 nodules from the public Lung Image Database Consortium and Image Database Resource Initiative dataset. The recommended higher-order transfer learning approach included three steps: semantic attribute-specific modeling, semantic attribute transfer modeling, and pathological attribute generalization, and was used to improve the predictive performance of node attributes. We also create transfer charts to highlight the combination of attributes that provide the most useful information for each purpose and the combination of the most relevant of the 11 attributes. The accuracy is 0.8194 ± 0.02.

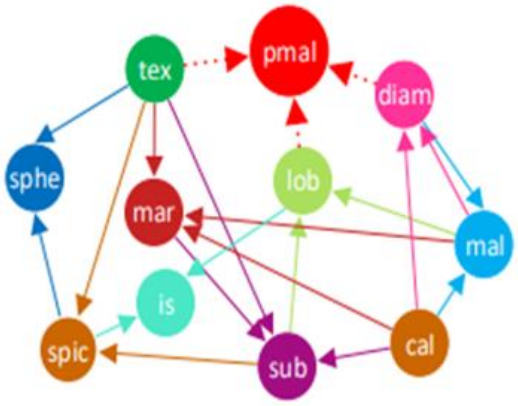

Fig. 3. Transfer graph of 11 attributes of pulmonary nodules

S. Mandal V. E. Balas et al. [1] computed a comparison analysis of numerous programs is presented in this paper by predicting the final forced volume capacity measurements for each patient and a confidence value. Lung function is assessed using a variety of regression algorithms, including: Multiple quantile regression, a useful statistical technique for determining the relationship between response variables and covariates. This model uses quantile regression using a multi-quantile convolutional neural network (CNN). Ridge regression is a regression method for estimating the results of equations that have no unique solution. With multicollinearity, the least squares estimation is unbiased, but the variance is large. As a result, the estimated value can deviate significantly from the actual value. ElasticNet is a machine learning model that uses a linear model. To penalize the coefficients of the regression model, ElasticNet regularization uses both L1 and L2 regularization regularizations.

Katarya, Rahul et al. [4] compared recent machine learning techniques for time series data analysis. There is a discussion of existing algorithms and their advantages and disadvantages. While the ARIMA model [4] laid the groundwork for precise trend learning and implementation, it was based on the assumption that the data set was linear, and hence failed to accurately map the trends. Neural networks revolutionized the way time series data dynamicity was assessed. BPA's gradient-based method is sometimes overlooked since it is impossible to measure. OSFELM uses a fuzzy inference framework to do temporal analysis. This can be changed to allow for the inclusion of exogenous variables in time-series analysis OSLA and OSELM, as well as their derivatives, currently provide the best results for time-series forecasts. Optimization approaches can help to improve these models even more. These techniques can be further improved by MMST and PSO. The drawback of time series requires extensive pre-processing, extensive computations and complex data structures to store the time series data.

Table 1
Comparative analysis

Paper	Date	Dataset	Methods	Accuracy
M. Anthimopoulos, S. Christodoulidis et al. [6]	2016	TWO DBS ILD CT SCANS FROM TWO SWISS UNIVERSITY HOSPITAL	LEAKYRELU, ADAM OPTIMIZER	85
Z. Zhang et al. [2]	2018	CIFAR 100	ND-ADAM, REGULARISED SOFTMAX	SGD 81.5 ND-Adam 81.58
Chenshuo Wang, Xian Xiang Chen et al. [5]	2018	The data of 354 subjects that underwent conventional spirometry. 35 other subjects that went through conventional spirometry and low degree of EOT criterion spirometry.	SVR	95%.
Q. Wang et al. [3]	2019	The public Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset are used to test our approach.	Pathologic attribute generalization, semantic attribute-specific modelling, and semantic attribute transfer modelling scheme of training	0.8194±0.02
Katarya, Rahul et al. [4]	2018		ARIMA Feedback Neural Network Recurrent Neural Network Long Short-Term memory (LSTM) Variants of OSLEM	Depends on how well the algorithm fits the new dataset
S. Mandal, V. E. Balas et al. [1]	2020	OSIC (Open-Source Imaging Consortium) Kaggle dataset	Multiple Regression Ridge Regression ElasticNet	OOF scores

3. Proposed Approach

Pulmonary fibrosis is a respiratory disease that causes scarring of the Lungs. It leads to various abnormalities in patients, ranging from rapid deterioration of lung function to mild cases of difficulty in breathing. There is no cure existing for this disease. The biggest hurdle in treating Pulmonary Fibrosis is in estimating the decline of the condition of the patients. Hence the scope of our system is to collect necessary medical information from patients diagnosed with pulmonary fibrosis and feed it into a system built using machine learning and deep learning techniques. The system performs image processing and regression to predict an output that the pulmonologist will use to decide the patient's further treatments.

The proposed system consists of using the image data and the tabular data to predict the FVC output for a particular week. The system analyzes the scarring of lung tissue on the CT scan and considers the FVC values from the tabular data. The image is preprocessed by converting them into standard sizes, converting them to tensors to be ready for CNN's. The system consists of using CNN's built upon a solid ImageNet architecture for Image data such as VggNet, InceptionNet, ResNet, and custom CNN models used for feature extraction. The tabular data will be used for Quantile Regression Analysis. We flatten the features so that they can be used along with tabular features. The flattened CNN features and Tabular features are combined to predict the FVC values for the given week. It makes use of Laplace Log-Likelihood function to predict the confidence of the prediction. The model predicts 3 FVC values for the given week. Where the loss function is a custom loss function used for regression and also outputs a confidence score for each FVC value. We are also designing a web-based application that provides an interface for pulmonologists to enter data, retrieve and provide results for the selected week. After analyzing the results, the doctor concludes the course of the illness.

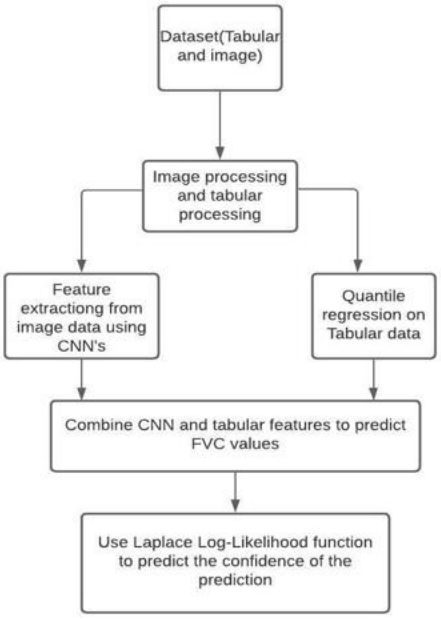

Fig. 4.

4. Conclusion

Pulmonary Fibrosis is an extremely critical issue in the present day and age. We require an effective programmed system that will be able to predict the early detection that would save countless lives. It will benefit the doctors and physicians diagnose patients proficiently and can track the progress through this system. Feature selection and prediction are principal for these mechanized systems. Choosing the right feature can help achieve better results in diagnosing and tracing pulmonary fibrosis. Search algorithms for selecting the features and using machine learning techniques are better used for giving accurate results.

References

[1] S. Mandal, V. E. Balas, R. N. Shaw and A. Ghosh, "Prediction Analysis of Idiopathic Pulmonary Fibrosis Progression from OSIC Dataset," 2020,

IEEE International Conference on Computing, Power and Communication Technologies (GUCON), 2020, pp. 861-865.
[2] Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), 2018, pp. 1-2.
[3] Q. Wang et al., "Higher-order Transfer Learning for Pulmonary Nodule Attribute Prediction in Chest CT Images," 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019, pp. 741-745.
[4] Katarya, Rahul, and Shubham Rastogi. "A study on neural networks approach to time-series analysis." In 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 116-119. IEEE, 2018.
[5] Wang, C., Chen, X., Zhao, R., He, Z., Zhao, Z., Zhan, Q., Yang, T. and Fang, Z., 2019. Predicting forced vital capacity (FVC) using support vector regression (SVR). Physiological measurement, 40(2), p.025010.
[6] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe and S. Mougiakakou, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1207-1216, May 2016.
[7] American Lung Association: Pulmonary Fibrosis, Nov. 18. http://www.lung.org/lung-health-and-diseases/lung-disease-lookup/pulmonary-fibrosis/.
[8] Nalysnyk L, Cid-Ruzafa J, Rotella P, Esser D. Incidence and prevalence of idiopathic pulmonary fibrosis: review of the literature. Eur Respir Rev. 2012;21(126):355-61.
[9] Verrecchia F, Mauviel A. Transforming growth factor-beta and fibrosis. World J Gastroenterol. 2007;13(22):3056-62.
[10] Wynn TA, Ramalingam TR. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. Nat Med. 2012;18(7):1028-40.
[11] King TE, Jr., Pardo A, Selman M. Idiopathic pulmonary fibrosis. Lancet. 2011;378(9807):1949-61.