International Journal of Research in Engineering, Science and Management
Volume 5, Issue 6, June 2022
https://www.ijresm.com | ISSN (Online): 2581-5792

Predicting Pulmonary Fibrosis Progression

Pranav Pradeep[1], H. S. Mansi[2], Likitha Keerthi[3*], Dev Narayanan[4], K. A. Sumithra Devi[5]

[1,2,3,4]*Student, Department of Information Science Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore, India*

[5]*Head of the Department, Department of Information Science Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore, India*

***Abstract*: Fibrosis is an incurable, fatal, and debilitating disease that damages the patient's respiratory system, making it difficult to live with. 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 will present how to use vital capacity (FVC) as a measure of lung condition to predict regression of idiopathic pulmonary fibrosis from CT images and tabular data characteristics. Features Combines the features of quantile regression extracted from a convolutional neural network to predict a decline in FCV values. We have also developed a web-based application that allows pulmonologists to enter data, retrieve results, and distribute reports.**

***Keywords*: Convolutional Neural Network, Quantile regression, Relu, Adam optimizer.**

1. Introduction

Pulmonary fibrosis is a condition that progressively worsens over time. The degree of fibrosis (scarring) in the lungs is linked to this deterioration. As a result, a person's breathing becomes more complicated, leading to shortness of breath even when at rest. The illness progresses at varying rates in patients with pulmonary fibrosis. Some individuals have a slow development and might stay with PF for plenty years, even as others have a quicker decrease. Although there is no cure for pulmonary fibrosis, several therapies can help to decrease the disease's course.

The loss in forced vital capacity (FVC) is used to track disease progression in pulmonary fibrosis (PF) (FVC). Vital capacity measurement measures the amount of air that can be forcibly expelled from the lungs after inhaling as deeply as possible. Mortality is linked to a drop in percent expected FVC of less than 10%, either in absolute or relative terms. In pivotal section three research of anti-fibrotic medicines, measures of FVC lower had been selected as major endpoints. Despite steady FVC lower styles withinside the PF population, the tempo of ailment development in people is unpredictable and extraordinarily variable: there's good sized heterogeneity in FVC throughout time, and the latest reductions are a bad predictor of destiny FVC decline.

Deep learning is an AI feature that mimics the way the human brain processes data to recognize objects, recognize voice, translate languages, and make decisions. Deep learning helps healthcare professionals and researchers uncover hidden potential data and better serve the healthcare industry. Medical images such as MRI and CT scans are used to diagnose severe illnesses such as heart disease and cancer. Deep learning helps doctors better analyze their condition and provide the best treatment for their patients. Traditional machine learning techniques can be used to predict the prognosis of a particular patient using CT scans and FVC data.

Predicting regression can provide valuable information, and improved detection of severity has a positive impact on the design of treatment studies and accelerates the clinical development of new therapies.

2. Literature Survey

Chenshuo Wang, Xian Xiang Chen et al. [5] used RBF SVR algorithm for predicting FVC values. [5] 3 Predictive models of mixed models, normal models, and anomalous models have been built. It was concluded that the performance of FVC prediction was better using the normal and abnormal models.

S. Mandal V. E. Balas et al. [1] The paper presents a comparison analysis of numerous programs by predicting the final forced volume measurements and confidence in lung function for each patient. Describes various regression algorithms such as convolutional neural networks and multiple division point regression, which is a useful statistical technique for determining the relationship between response variables and covariates using ridge regression. ElasticNet is used to penalize the coefficients of the regression model. ElasticNet regularization uses both L1 and L2 normalization regularizations.

[7] Jack Lu, Adam Dorfman et al introduced Fibrosis-Net is a deep convolutional neural network developed to predict the progression of pulmonary fibrosis from chest CT scans. Finally, they used descriptive-driven performance validation techniques to examine the behavior of the fibrosis decision-making network and confirm that predictions were retained based on relevant visual indicators on CT images. The proposed Fibrosis-Net has a modified Laplace Log Likelihood score of −6.8188 which is greater few states of the art proposed architectures.

3. Methods

A. Datasets

The data consist of chest CT scans and forced vital capacity (FVC) measurements from frequent visits over approximately 1-3 years. The training set contains a complete history of

*Corresponding author: likithakeerthi47@gmail.com

clinical information from 172 patient cases, and the test set contain baseline measurements consisting of approximately 28 patient cases. Each patient's CT scan in DICOM format was contained in a separate folder named after the patient ID. Patients' relative FVC readings compared to smoking status, weeks before and after baseline CT, FVC readings, and common FVC readings for patients with similar characteristics.

B. Algorithms Used

1) Quantile Regression is an alternative method to linear regression, which predicts numerical outputs. This regression minimizes the quantile loss in predicting a certain quantile. A quantile is the value below which a fraction of observations in a group falls. The idea behind Quantile Regression is to divide the data into quantiles containing inputs such as True targets, predicted values, Quantile value (q) and the loss obtained from the output depends on the quantile in which the actual output lies. It is better than Linear Regression when the data is multimodal than bimodal. To optimize the Quantile regression - define a function that takes the predicted value, the actual value, and the quantile value (q), define a Neural Network (NN) which takes the tabular features as the input layer and gives three values as output. The loss function for this neural network is the quantile regression loss. Train the NN for the decided number of epochs and backpropagate the loss obtained. Feed the output to .csv file.

2) Convolutional Neural Networks are feedforward, and ANN is inspired by biological processes and aims to identify patterns directly from pixel images (or other signals). Convolution, activation, pooling, and fully connected (or dense) layers are the four layers included in a standard CNN.

Sparse local connections and weight sharing are features of the convolution layer. Each neuron in the layer is connected only to a limited local area of input, similar to the receptive fields of the human visual system. Different neurons respond to different local parts of the input, which overlap to provide a better visual representation. Weight distribution significantly reduces the number of network parameters, improves efficiency, and prevents overfitting. The convolutional layer is often followed by a non-linear activation layer to capture the more complex features of the input signal. The pooling layer is also used to under-sample the previous layer by aggregating small rectangular data samples. The pooling layer mitigates output vulnerability to small input changes. Finally, one or more thick layers are used, each followed by an activated layer, resulting in classification results. CNNs trained in the same way as other ANNs. That is, it uses a gradient descent-based approach to minimize the loss function and back propagate the error.

4. Proposed Approach

Empty patient ID folders are checked with tabular training data and these IDs are removed from the dataset. The tabular data is then aggregated by each Patient ID. Then encode the category value. The numbers are also normalized. The FVC values are predicted using a hybrid technique. We begin by aggregating the data based on each patient. Then used FVC as the dependent variable and week as the independent variable in the linear regression model. Get the slope and intercept for each patient equation. The target variable to predict is the slope of the equation. These figures are derived only from tabular elements. Use a convolutional neural network (CNN) to extract fixed-size feature vectors and include image features. These vectors record the image's semantic properties, it can then be used to distinguish CT scans. Image dimensions (NXN) are displayed in matrix format and the depth or number of channels in these images is 1 because they are grayscale. Once you have the input image (NXN), filter all pixels (rows, columns) to collect the data. This information is passed to the pooling layer, which performs mathematical calculations and returns a specific result. The filter in this case is a weight matrix of the form nxn (33). To provide regularly distributed values, the weights are initialized as random integers with one standard deviation. This filter is applied to all values in the matrix and the result is the product of the weights and pixels. For our weights and biases, we build a function. TensorFlow is a computational graph that aids in the creation of these functions as well as the variables that aid in the storage of them as objects. Create a function that returns a convolutional layer. Before we create it, we need to think about a few things. The first is the weight matrix, or the filter`s size, in order to drift over all the matrix's values. Every generation, your stride is the quantity of pixels you leap or slide over. Then there may be the range of channels, which determines the image's intensity. The intensity is 1 for the reason that pix are grayscale. After that, we will use the Rectified Linear Unit (ReLU) activation characteristic to spark off this layer. When as compared to the conventional sigmoid method, the usage of the ReLU characteristic has been proven to noticeably accelerate the education process. A TensorFlow enter need to be a 4-dimensional vector. For example, [A, B, C, D]. The range of samples to be taken into consideration in every generation is denoted via way of means of A. (the range of images, on this case). The shape of the pix, or the dimensional description of the image, is represented via way of means of B and C. (B, C). And D is the input image's channel count, which is one in this example. We need to flatten the CNN layers into an array after they've been built. The form of the picture the second and third elements in our convoluted layer, as well as the 4th element, the number of filters are used for this.

We acquire the form of the flattened output by multiplying all of them. All of the photos in this data collection have the same dimensions. Passing one of these images through the first convolution layer results in 16 channels in a reduced dimension. If you apply two layers of zeros to the outer layer of the image and pass it through the pooling layer, the result will be exactly half the size of the input. We obtain a 7x7x32 result when we convolve two of these layers. An Adam Optimizer is used in this example to assist reduce the cost determined by cross entropy. Mean absolute error was used to evaluate the model's performance.

5. System Architecture

A system architecture is a conceptual model that defines the structure, behavior, and other aspects of a system. First, collect and save the data. The data is then split into training data and validation / test data. The next step involves cleaning the pre-processing the data. The tabular information is cleaned concerning the columns. The model is then built by running the tabular data the on a Regression model. It is done to get the FVC values. The Image data is then run through a custom CNN architecture. This model is used to predict FVC levels over the next 3 weeks.

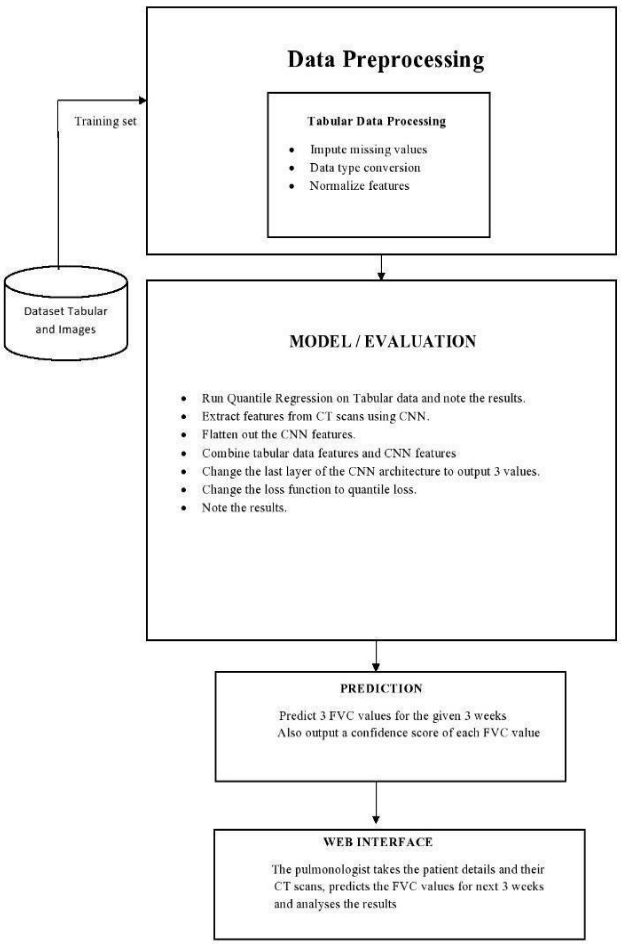

Fig. 1. System architecture

6. Output

Fig. 2. Output

The designed internet site includes a shape for amassing patient`s information together with Patient Id and precise Id for every patient, Number of weeks the relative variety of weeks pre/put up the baseline CT, Percent a computed subject which approximates the patient`s FVC as a percentage of the everyday FVC for someone of comparable characteristics, Age, Gender, Smoking status. On submitting these details, the model predicts the FVC values for the next three weeks. weeks.

7. Conclusion

In this paper, the approach used in predicting the Pulmonary Fibrosis are convolutional neural network with ReLu as activation function and Adam as optimizer and quantile regression to predict the FVC values for the next three weeks, the doctor receives these FVC values for the given weeks. Upon analyzing the result, the doctor concludes the progress of the disease. The performance of the model has been determined by mean absolute error. For further enhancement of the project IOT can be used. When the fall in FVC starts to accelerate, Internet of Things (IoT) solutions offer an alarm to the patients. To Making the Project patient centric for the patient to access his health history, medication history.

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] Wong, A., Lu, J., Dorfman, A., McInnis, P., Famouri, M., Manary, D., Lee, J. R. H. and Lynch, M., Fibrosis-Net: a tailored deep convolutional neural network design for prediction of pulmonary fibrosis progression from chest CT images. Frontiers in Artificial Intelligence, pp. 161, 2021.