# Tracking The Pulmonary Function Distortion in Smokers and Covid 19 Post Infectiuos Patients using Deep Learning

Hany S. Elnashar [1], Abdelrahim Koura [2]

Beni_Suef University, FCAI

*Abstract*:- **Technological evolution of the recent years has been growth rapidly to a greater extent by deep learning. Conventional classification systems rely on laborious feature engineering process which requires expertise in domain knowledge for data interpretation. On the contrary, deep learning without using a domain expert can extract distinct features from the data. In this study, conduct some preliminary experiments using the deep learning approach to classify Pulmonary Fibrosis based on human lung CT images, pulmonary fibrosis occurs when lung tissue becomes damaged and scarred. Stiff tissue makes it more difficult for the lungs to work properly. As it worsens, the breath becomes progressively more shortest. This study predicts the severity of the decline in lung function based on a CT scan of more than one patient. In other words, well predict the final three Forced vital capacity "FVC" measurements for each patient, as well as a confidence value.**

*Keywords: -Machine Learning, Deep learning, Classification, Image processing, Supervisory learning, Python, CT image, FVC, MRI*

## 1. INTRODUCTION

Artificial intelligence is a field that displays human intelligence in a machine, machine learning and deep learning are the domains that work under AI. Machine learning allows computers to learn without explicit programming, whereas deep learning a type of machine learning enables the system to read data and comprehend it. Machine learning categories include supervised, semi-supervised, unsupervised, active learning, and reinforcement algorithms. Deep learning methods are an advanced phase of these algorithms which using a neural network classifies data and allows automatic decision-making[1] . These methods enable computers to extract patterns corresponding to a specific data set and automatically reason them[2] to use for selection orin more action of prediction. Medical imaging owing to its impact in t h e early diagnosis and treatment of diseases is a rapidly growing research area. Image processing in medical imaging is relatively significant. Image processing concepts include image classification, object detection, pattern recognition, reasoning ..etc. These concepts allow t h e extraction of valuable patterns from specific data with increased accuracy. Machine and deep learning methods aid in classification and automatic decision-making for multi-dimensional medical data[3]. D iseases can be studied effectively in medical imaging using machine learning algorithms. A simple mathematical solution or model cannot accurately distinguish lesions and organs in medical image processing. Machine learning use pixel-based investigation of medical images[4]. ML models have been applied to a wide range of tasks. These tasks can be broadly divided into five main classes: Association, Classification, Regression, Clustering, and Optimization Predections tasks [5]. Prediction means that the output of a computer base algorithm has been trained on a known dataset and applied to unknown data when forecasting the likelihood of a particular outcome, Machine learning model predictions allow making highly accurate guesses as to the likely outcomes of a question based on this historical data, which can be about all kinds of things and cases. CT scan imaging uses X-rays to obtain structural and functional information about various organs of the human body. Based on the fact that different tissues and matter absorbs X-rays differently, X- rays are used in diagnosis. Bones appear white, while soft tissues appear gray in a CT image. CT performs as a supplement to Magnetic resonance imaging and ultrasonography in diagnosis. They are particularly used in imaging and diagnosis of the brain, liver, chest, abdomen, pelvis, and spine [7]. Cavities of lungs filled with air appear black. Pulmonary fibrosis is a lung disease that occurs when lung tissue becomes damaged and scarred This thickened, stiff tissue makes it more difficult for lungs to work properly. As pulmonary fibrosis worsens, lungs become progressively more short of breath. The lung damage caused by pulmonary fibrosis can't be repaired, but medications and therapies can sometimes help ease symptoms and improve quality of life. For some people, a lung transplant might be appropriate.

In this paper needs to predict a patient's severity of the decline in lung function based on a CT scan of their lungs and some additional tabular data fields. Now the challenge is to use machine learning techniques for a prediction from some not sufficient images, metadata, and baseline FVC. In this paper will look for prediction of predict a patient's severity of decline in lung function based on a CT scan of their lungs. In other words, Will predict the final three FVC measurements for each patient, as well as a confidence value in prediction [8]. This is not simple. Due to insufficient data availability, hardest hit use traditional Computer Vision approaches to model the dependency between values of patient FVC and CT scans.

This paper organized in subsectors started by literature review, then one partition of porposed methods used with our model structure and dataset representations and relation with reflect to each other, therd partition for Predection methods as per two method used started by Linear Decay Predictions model and Multiple Quantile Regression Predictions model, and finaly the Result discutions with enassembles of two metods.

## 2.   LITERATURE REVIEW

Many research groups have reviewed and have reported the emerging trends of deep learning applications in medical image processing. [9] The computational implementation difficulty of convolutional neural  networks owing to high memory bandwidth requirement and intensive computation resources has paved the way for the use of deep  learning algorithms. [10] for image classification that enables   a     significant reduction in the networks' model size and improves the accuracy and performance. [11] use proposed weakly supervised learning, in which labeled training data is improvised with image annotations and   scribbles. Semi-supervised learning method to train the network for cardiac MR image segmentation was reported in [12] They were able to effectively improve segmentation accuracy. In [13]   review has elaborated on the  general image data sets for supervised learning. And [14] in their article has proposed a supervised learning approach  for t h e   classification of lung cancer CT-scan images. They have combined classical feature-based SVM classifier and supervised  learning algorithm for the accurate classification and detection of lung nodules for early and efficient diagnosis of lung cancer.

## 3.   MATERIAL AND METHODS:

### 3.1.  Data source

Data downloaded as dataset from Kaggle google, its world's largest data science community .dataset contains a baseline chest CT scan and associated clinical information for a set of patients. The data contains 176 patients for each have CT scan image. Most patients have FVC measurements at nine different timesteps, but this number can vary between 6 and 10. Thus, the number of FVC measurements is not consistent for different patients.A patient has an image acquired at time Week = 0 and has numerous follow-up visits over approximately 1-2 years, at which time their FVC is measured.

### 3.2.  Method:

Various steps are performed on the Data of medical images before the detection of output as elaborated in figure 1. Initially, the  medical images are given as input to the machine and deep learning algorithms and pre-processed for the removal of distortion and   noise. After that, the images are divided into different segments to zoom the interested area (ROI). Then, the features are extracted   from these segments through information retrieval techniques. The desired features are selected and the noise is removed.After the process of feature extraction, a database is created based on the feature parameters. And then the created dataset is ready to use. In current research data started by reference data a baseline chest CT scan and associated clinical information for a set of patients. A patient has an image acquired at time Week = 0 and has numerous follow-up visits over approximately 1-2 years, at which time their FVC is measured.  As this  data distributed to the training set and test set

- In the training set, are provided with anonym zed, baseline CT scan, and the entire history of FVC measurements.
- In the test set, are provided with a baseline CT scan and only the initial FVC measurement. Now are asked to predict the final three FVC measurements for each patient, as well as a confidence value in prediction.
- There are no null values train and test sets.

  The timing of the initial measurement relative to the CT scan and the duration to the forecasted time points may be different for each patient. To avoid potential leakage in the timing of follow-up visits, one must predict every patient's FVC measurement for every possible week. This data is distributed on more than one files:

- train.csv - the training set, contains the full history of clinical information
- test.csv - the test set, contains only the baseline measurement
- train/ - contains the training patients' baseline CT scan in DICOM format
- test/ - contains the test patients' baseline CT scan in DICOM format

The train.csv file contains a unique id for each patient, And in Columns formats as Patient- a unique Id for each patient (also the name of the patient's DICOM folder), Weeks- the relative number of weeks pre/post the baseline CT (may be negative), FVC - the recorded lung capacity in ml, Percent- a computed field which approximates the patient's FVC as a percent of the typical FVC for a person of similar characteristics, Age, Sex, Smoking Status. Figure-1show block diagrams of all steps to run this prediction method by Python was chosen as the development platform due to its vast set of APIs and modules. The syntax was very easy to code and implement hence coding these algorithms using python helped to create logic. Human Lung CT scan images were given as inputs to the deep learning algorithm, which has been developed in python language. Figure 1 shows steps from reading data folders and python library running. Table 1show data records which contain parameters and features that are used to run Machine learning and Deep learning algorithms named in the block diagram in step two as the Data Exploration step.   In seven columns with a total number of records 1549,   if start with classify Patient duplicate data rows in the train where the Patient and Weeks elements match, Well find that there are 14 duplicate rows table 2 show these duplications.  So first step to droop this duplication and remove it from Patient data.

Using pythons library to present data relations, starts by unique patients data distributions over  Age, Sex, and  SmokingStatus grouped by Patient table
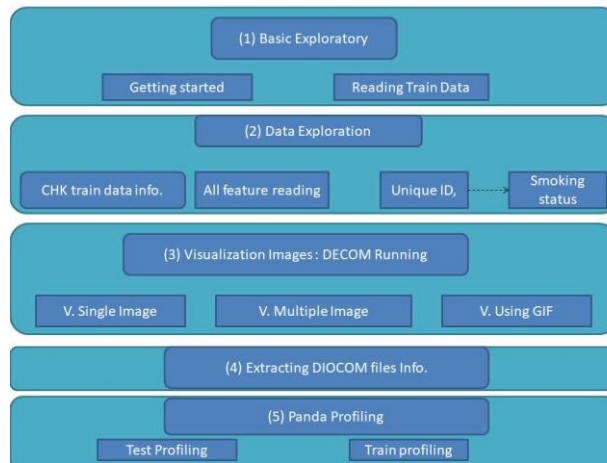
Figure-1 Block diagram of the proposed system

### Table1: Training data shape

|   | 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 |

### Table 2: train data Duplications

|   | Patient | Weeks | FVC | Percent | Age | Sex | SmokingStatus |
|---|---------|-------|-----|---------|-----|-----|---------------|
| 166 | ID00048637202185016727717 | 9 | 1375 | 60.059404 | 70 | Female | Never smoked |
| 167 | ID00048637202185016727717 | 9 | 1477 | 64.514720 | 70 | Female | Never smoked |
| 226 | ID00068637202190879923934 | 11 | 2794 | 82.633385 | 73 | Male | Ex-smoker |
| 227 | ID00068637202190879923934 | 11 | 2827 | 83.609369 | 73 | Male | Ex-smoker |
| 306 | ID00089637202204675567570 | 7 | 2478 | 57.967624 | 63 | Male | Never smoked |
| 307 | ID00089637202204675567570 | 7 | 2539 | 59.394592 | 63 | Male | Never smoked |
| 465 | ID00125637202218590429387 | 8 | 2349 | 53.757781 | 65 | Male | Never smoked |
| 466 | ID00125637202218590429387 | 8 | 2584 | 59.135848 | 65 | Male | Never smoked |
| 750 | ID00199637202248141386743 | 5 | 2672 | 64.728682 | 62 | Male | Ex-smoker |
| 751 | ID00199637202248141386743 | 5 | 2683 | 64.995155 | 62 | Male | Ex-smoker |
| 920 | ID00240637202264138860065 | 5 | 2991 | 78.421605 | 63 | Male | Ex-smoker |
| 921 | ID00240637202264138860065 | 5 | 3132 | 82.118511 | 63 | Male | Ex-smoker |
| 1521 | ID00421637202311550012437 | 70 | 2628 | 78.720345 | 68 | Male | Ex-smoker |
| 1522 | ID00421637202311550012437 | 70 | 2719 | 81.446202 | 68 | Male | Ex-smoker |

### Table 3:  smoker status as per contend

|   | Weeks | FVC | Percent | Age | Sex | SmokingStatus |
|---|-------|-----|---------|-----|-----|---------------|
| 0 | -4 | 2315 | 58.253649 | 79 | Male | Ex-smoker |
| 1 | 8 | 3660 | 85.282878 | 69 | Male | Ex-smoker |
| 2 | 0 | 3523 | 94.724672 | 60 | Male | Ex-smoker |
| 3 | 6 | 3326 | 85.987590 | 72 | Male | Ex-smoker |
| 4 | 33 | 3418 | 93.726006 | 65 | Male | Never smoked |

And figure 2 show the Distribution of unique patients data based on (sex, age, and somking_status)
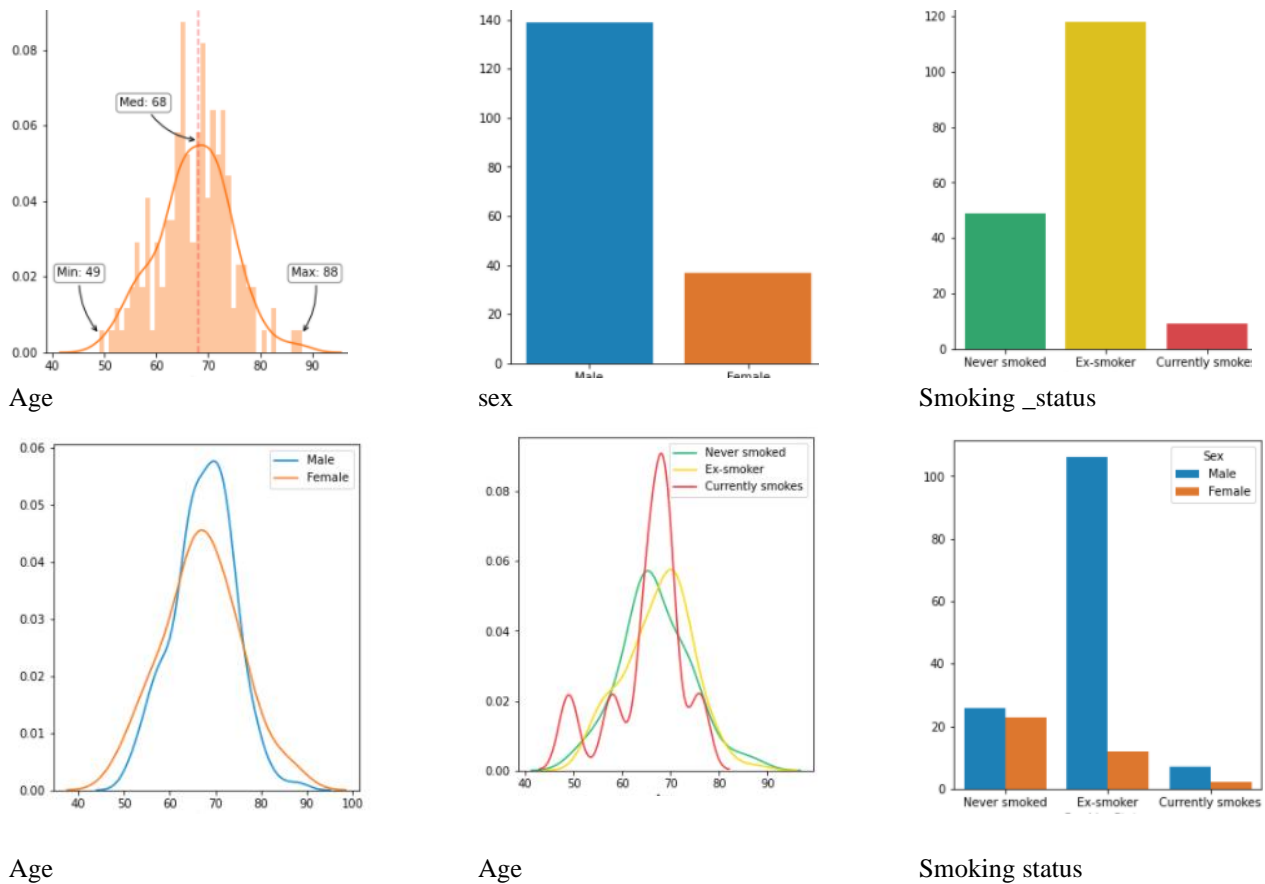
Figure 2: Distribution of unique patients data

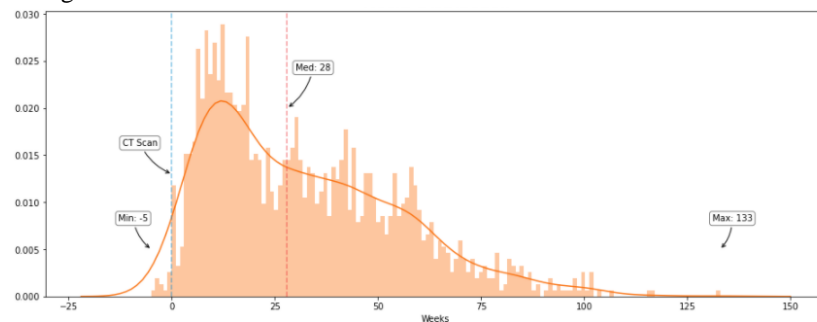As per weeks distributions figure 3 show data distributions



Figure: 3 weeks of data distributions

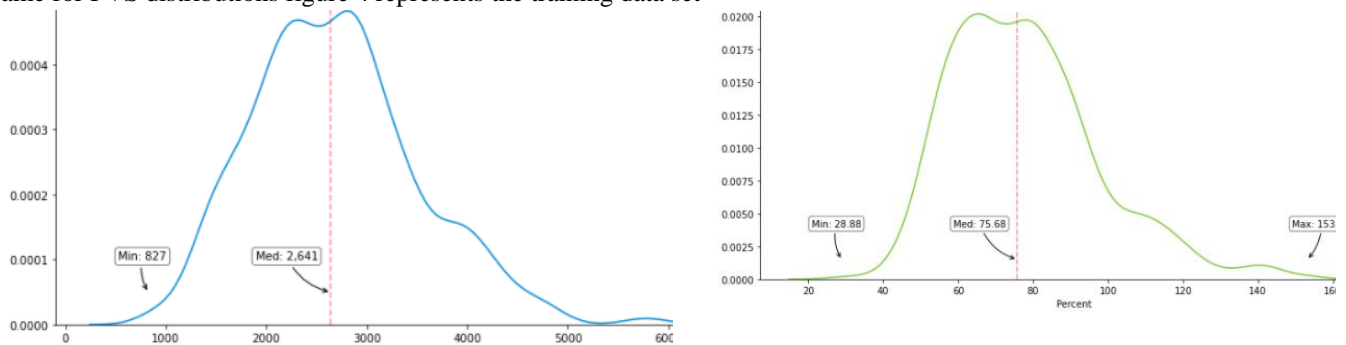Same for FVS distributions figure 4 represents the training data set



Figure:4 FVC and Persent Distributions

Now and from last representations building data relations between all values in training data sets, find the relation between main variable FVC and all other variables age, weeks,….ets from data set. Frist between FVC and sex find that as in figure 5.a and figure 5.b show the same between FVC and Smoking_status.
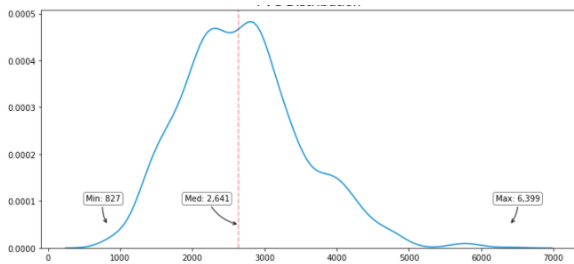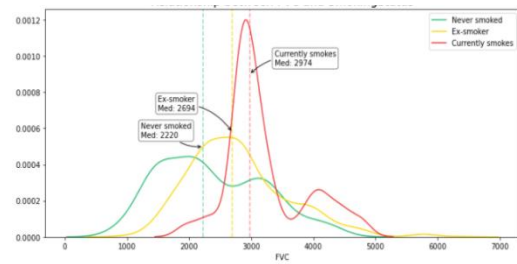
Figure 5.a: FVC per sex



Figure 5.b: FVC per Smoking

From that, all FVC of the Female gender appears to be less than of Male, and there is also a difference in FVC by SmokingStatus, this could be based on there are more Females who Never smoked. Let check it by looking deeply into data as shown in figure 6.a and 6.b
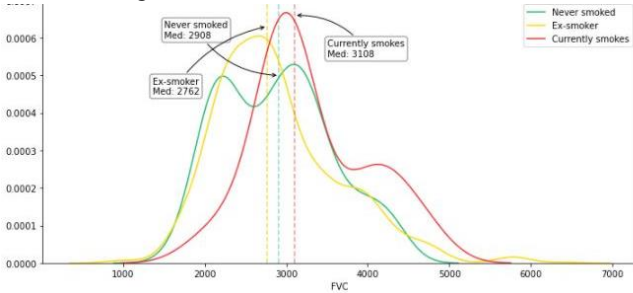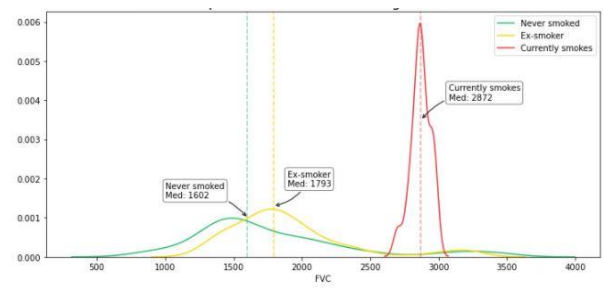


Figure 6.a: FVC and SmokingStatus in Male



Figure 6.b: FVC and SmokingStatus in Female

Whereas shown in figure 6.a changes are so limited FVC seem to no changes much with smoking status. But there is a difference with female status; this is probably due to the small sample size sets (Currently smokes). It may also be important to consider that patients who are Currently smokes are likely to be less severely affected. From this data looking for a relation between FVC and patient age by show its distribution from the data set as shown in figure 7, a and b.
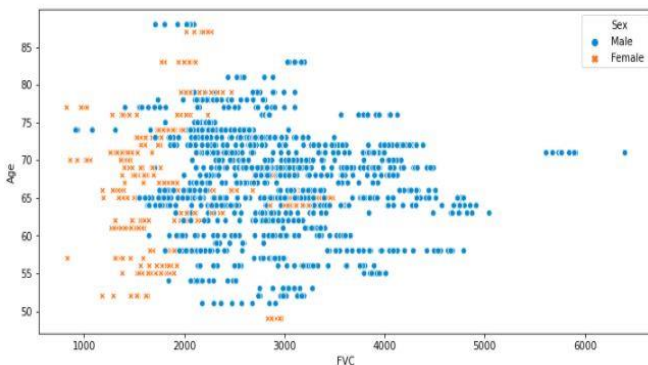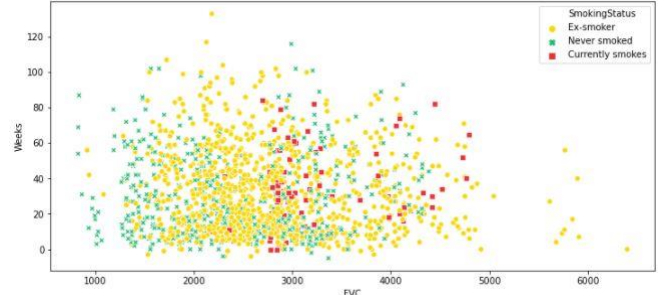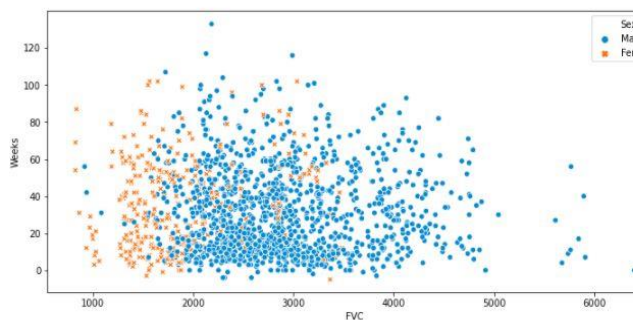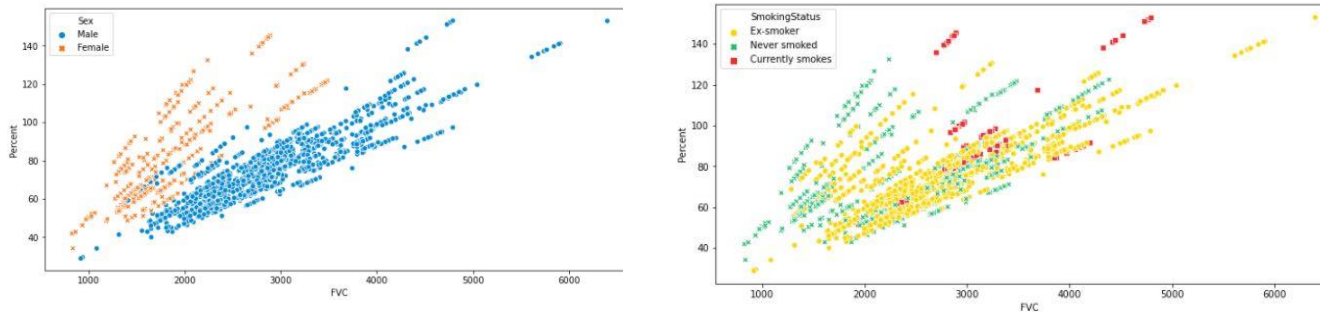


A
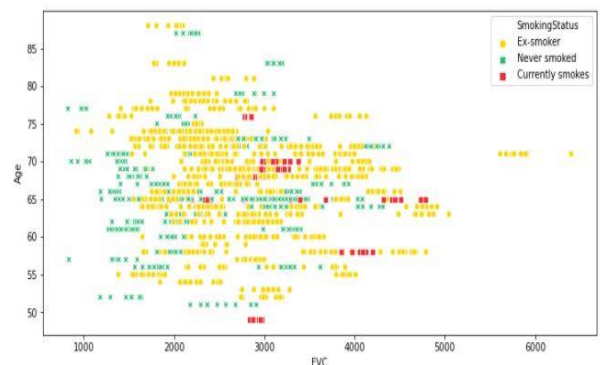


b

Figure7: Correlation Between Fvc And Age.

As same as for FVC and weeks it will show from data as in next figure 8 a,b,c, and d, and other feature from data set like percent's and weeks it will show more details about the relations





Correlation between FVC and Weeks

Correlation between FVC and Percent

Figure 8: FVC with other values

As shown in figures related to percent's There is a positive correlation between FVC and Percent, which is not surprising Since Percent is a value calculated from FVC and other data.

## 4. PREDICTIONS MODELS:

### 4.1. Linear Decay Predictions model:

Now start using DICOM [15] and other data by Create a function based on putting `Age`, `Sex`, and `Smoking_Status` into array_1, and then reorganize the data by the unique ID of Patients. And extract all of the Weeks from weeks and FVC from fvc in array_ 2 formats Respectively, join weeks and the array with all 1's in the vertical direction, then transpose them, and assign to c. Then find the least-squares of fvc and c, and add the slope to A. then add the value obtained from the last function that represents array_1for example notate it as function_1tab to TAB and the unique ID of Patient to p. by creating a new function for reading from a library of DICOM and to optimize ranges it could be divided by the value of 2048 and resize the image to 512*512 then crop it, using tensor through EfficientNetB6[16] and GlobalAveragePooling2D [17], and start pre-processed CSV data and adding Gaussian noise to the tensor and concatenate all of them together to the model output. And train The model weights. By run this modification over the model, And by creating a modified version of the Laplace Log-Likelihood function [18]. And then using a new model shon in figure 9 for predictions.
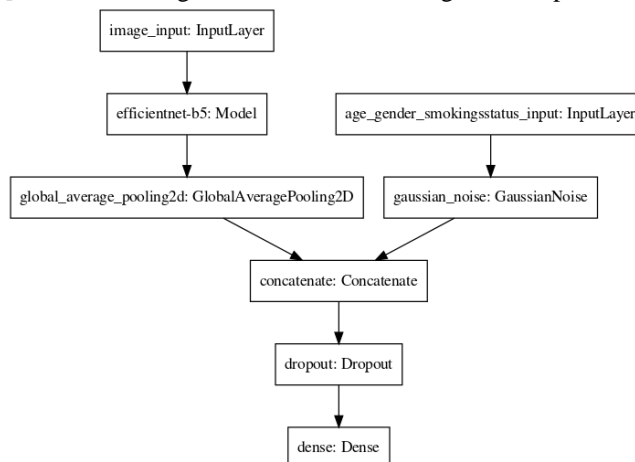


Figure 9: new model modifred of EfficientNetB6

The score function[19] takes the true and predicted values of the target variable and returns a score based on the modified Laplace Log Likelihood.

Table:4 Prediction results based on Linear Decay Model

|  | Patient_Week | FVC | Confidence |
|---|---|---|---|
| 0 | ID00419637202311204720264_-12 | 3086.790210 | 136.977065 |
| 1 | ID00421637202311550012437_-12 | 2839.031197 | 182.076488 |
| 2 | ID00422637202311677017371_-12 | 1996.793283 | 143.465776 |
| 3 | ID00423637202312137826377_-12 | 3401.559275 | 186.818177 |
| 4 | ID00426637202313170790466_-12 | 2983.062881 | 114.887829 |

Where for data representation for one Patient noumber 4 intable 4 befor and after running linear decay shown in figure 10

**Published by :**

**http://www.ijert.org**

**International Journal of Engineering Research & Technology (IJERT)**
**ISSN: 2278-0181**
**Vol. 10 Issue 09, September-2021**

Figure 10: Patient 4 in table 4 befor and after running linear decay FVC per weeks

### 4.2. Multiple Quantile Regression Predictions model:

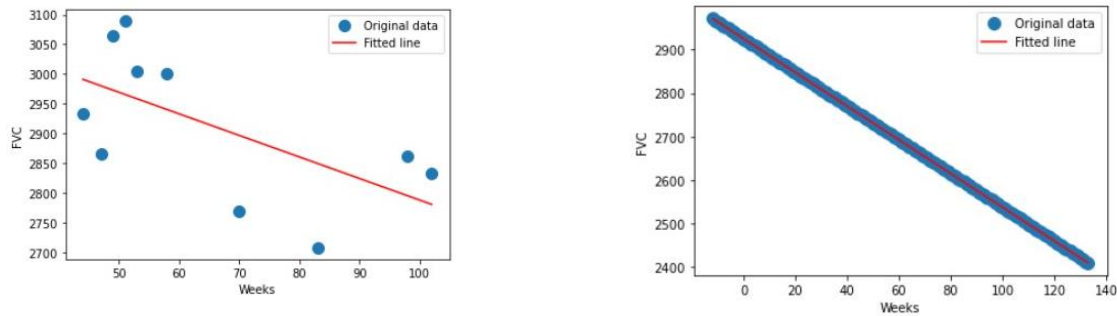And when data prepared to use multiple quantile regression with Splitting Patient_Week in sub into Patient and Weeks, according to the train and test formats. Then, attach the Patient to the Patient in a sub and merge it with the Patient. This makes it easier to handle the prediction as shown in table 5.

Table 5: change data attachments

| | Patient | Weeks | Confidence | Patient_Week | FVC | Percent | Age | Sex | Smoking Status |
|---|---|---|---|---|---|---|---|---|---|
| 0 | ID00419637202311204720264 | -12 | 100 | ID00419637202311204720264_-12 | 3020 | 70.186855 | 73 | Male | Ex-smoker |
| 1 | ID00419637202311204720264 | -11 | 100 | ID00419637202311204720264_-11 | 3020 | 70.186855 | 73 | Male | Ex-smoker |
| 2 | ID00419637202311204720264 | -10 | 100 | ID00419637202311204720264_-10 | 3020 | 70.186855 | 73 | Male | Ex-smoker |
| 3 | ID00419637202311204720264 | -9 | 100 | ID00419637202311204720264_-9 | 3020 | 70.186855 | 73 | Male | Ex-smoker |
| 4 | ID00419637202311204720264 | -8 | 100 | ID00419637202311204720264_-8 | 3020 | 70.186855 | 73 | Male | Ex-smoker |

And when selecting minimum weeks for each patient from the column for the minimum number of weeks per Patient. and calculate base_FVC (= the FVC of Patient at min_week) and the base_week (= how many weeks have passed since min_week). That by extract the rows in the data where Weeks is min_week and set them to base. Extract only the Patient and FVC columns from the base, and change the column name from FVC to base_FVC. Then create a new nb column and set all the values to 1. Group the base with the Patient and compute the cumulative sum with the nb column. Extract only the rows from the base that have nb columns of 1, and replace the base. This allows us to eliminate duplicate Patient rows from the base data frame with base_FVC in it. Let's remove the nb column. As shown in table 7:

Table 6: Min_weeeks for each Patient

| | Patient | Weeks | FVC | Percent | Age | Sex | Smoking Status | WHERE | Confidence | Patient_Week | min_week |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | ID00007637202177411956430 | -4 | 2315 | 58.253649 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 |
| 1 | ID00007637202177411956430 | 5 | 2214 | 55.712129 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 |
| 2 | ID00007637202177411956430 | 7 | 2061 | 51.862104 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 |
| 3 | ID00007637202177411956430 | 9 | 2144 | 53.950679 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 |
| 4 | ID00007637202177411956430 | 11 | 2069 | 52.063412 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 |
| 5 | ID00007637202177411956430 | 17 | 2101 | 52.868646 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 |
| 6 | ID00007637202177411956430 | 29 | 2000 | 50.327126 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 |
| 7 | ID00007637202177411956430 | 41 | 2064 | 51.937594 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 |
| 8 | ID00007637202177411956430 | 57 | 2057 | 51.761449 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 |
| 9 | ID00009637202177434476278 | 8 | 3660 | 85.282878 | 69 | Male | Ex-smoker | train | NaN | NaN | 8.0 |

Table 7: Patient per Base FVC

| | Patient | base_FVC |
|---|---|---|
| 0 | ID00007637202177411956430 | 2315 |
| 9 | ID00009637202177434476278 | 3660 |
| 18 | ID00010637202177584971671 | 3523 |
| 27 | ID00011637202177653955184 | 3326 |
| 36 | ID00012637202177665765362 | 3418 |

With some data preparation for the current model by concatenating some data fields and renew some as basic for use table 8 show these valves

Table8: new data formats using concatenation and selection from data set

| | Patient | Weeks | FVC | Percent | Age | Sex | SmokingStatus | WHERE | Confidence | Patient_Week | min_week | base_FVC | base_week |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | ID00007637202177411956430 | -4 | 2315 | 58.253649 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 | 2315 | 0.0 |
| 1 | ID00007637202177411956430 | 5 | 2214 | 55.712129 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 | 2315 | 9.0 |
| 2 | ID00007637202177411956430 | 7 | 2061 | 51.862104 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 | 2315 | 11.0 |
| 3 | ID00007637202177411956430 | 9 | 2144 | 53.950679 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 | 2315 | 13.0 |
| 4 | ID00007637202177411956430 | 11 | 2069 | 52.063412 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 | 2315 | 15.0 |
| 5 | ID00007637202177411956430 | 17 | 2101 | 52.868646 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 | 2315 | 21.0 |
| 6 | ID00007637202177411956430 | 29 | 2000 | 50.327126 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 | 2315 | 33.0 |
| 7 | ID00007637202177411956430 | 41 | 2064 | 51.937594 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 | 2315 | 45.0 |
| 8 | ID00007637202177411956430 | 57 | 2057 | 51.761449 | 79 | Male | Ex-smoker | train | NaN | NaN | -4.0 | 2315 | 61.0 |
| 9 | ID00009637202177434476278 | 8 | 3660 | 85.282878 | 69 | Male | Ex-smoker | train | NaN | NaN | 8.0 | 3660 | 0.0 |

For this model need it important to Normalize Percent, Age, base_FVC, and base_week. And let's split data into train, test, and sub using the WHERE column, and remove data. Now to run this model based on a modified version of the Laplace Log Likelihood. For each true FVC measurement, it will predict both an FVC and a confidence measure (standard deviation σ). The error is a threshold at 1000 ml to avoid significant errors adversely penalizing results, while the confidence values are clipped at 70 ml to reflect the approximate measurement uncertainty in FVC. The metrics are computed as follows:

$$\sigma clipped = \max(\sigma, 70)$$
$$\nabla = \min(|FVC_{true} - FVC_{predicted}|, 1000)$$
$$metrics = \frac{\sqrt{2\nabla}}{\sigma clipped} - \ln(2\,\sigma clipped)$$

And using this model to cross-validate of ( BATCH_SIZE = 128, EPOCHS = 804, and NFOLD = 5) and after run for each folds found that figure 11 :

```
FOLD 1
train [33.37018585205078, 6.500749111175537]
val [40.6624641418457, 6.9998555183410645]
FOLD 2
train [33.08328628540039, 6.500251770019531]
val [44.767337799072266, 6.851904392242432]
FOLD 3
train [31.155855178833008, 6.423372745513916]
val [46.81915283203125, 6.771785736083984]
FOLD 4
train [34.13739776611328, 6.529685974121094]
val [35.11758041381836, 6.559061527252197]
FOLD 5
train [31.01003646850586, 6.415341377258301]
val [40.26162338256836, 6.74094820022583]
```

Figure 11: Folds results of Model

Then by Read the original test.csv and overwrite the FVC and Confidence in the predicted data, it will give next table 9 show the modified outputs

Table 9: modified output as per model 2 runs

| | Patient_Week | FVC | Confidence |
|---|---|---|---|
| 1540 | ID00419637202311204720264_-12 | 2878.200500 | 228.700500 |
| 1541 | ID00419637202311204720264_-11 | 2875.794861 | 230.106323 |
| 1542 | ID00419637202311204720264_-10 | 2873.389343 | 231.511780 |
| 1543 | ID00419637202311204720264_-9 | 2870.983826 | 232.917480 |
| 1544 | ID00419637202311204720264_-8 | 2868.578247 | 234.322937 |

And by ensemble two models will find that results are shown in Table 10.

Table 10: concatenated results from two models

| | Patient_Week | FVC | Confidence |
|---|---|---|---|
| 0 | ID00419637202311204720264_-1 | 2890.586347 | 244.162292 |
| 1 | ID00419637202311204720264_-10 | 2914.585290 | 231.511780 |
| 2 | ID00419637202311204720264_-11 | 2917.251817 | 230.106323 |
| 3 | ID00419637202311204720264_-12 | 2919.918442 | 228.700500 |
| 4 | ID00419637202311204720264_-2 | 2893.252826 | 242.756836 |

## RESULTS DISCUTIONS:

There are around 200 cases in the public & private test sets, combined. Since this is real medical data, used and it will notice the relative timing of FVC measurements varies widely. The timing of the initial measurement relative to the CT scan and the duration to the forecasted time points may be different for each patient. This is considered part of the challenge of the paper. To avoid potential leakage in the timing of follow up visits, predictection for every patient's FVC measurements for every possible week. Those weeks which are not in the final three visits are ignored in scoring. Now from the Data represented in the last sections, the first point of view data correlations tables and figures represent that correlation between FVC and other features of data could be arranged as, There was no correlation between FVC and Age, no correlation between FVC and Weeks either, but There was a positive correlation between FVC and Percent, which is not surprising since Percent is a value calculated from FVC it selve and other patient data parameters. Form this corelations for each patient, there was given an initial CT scan that corresponded to the first week in the patients' metadata and additional weeks, which described how patients FVC changed during time. For the test set for each patient, only the first-week data was given, along with the initial CT scan. So the task was not only to predict the "FVC" values for the following weeks but also to display the "Confidence" score for each prediction. And comparing the results given and show the prediction modifications and its confidences. By ensembling two models outputs in one table it give optimum prediction representation values for each patient in form of FVC and its confidance.

## CONCLUTIONS:

Evidence suggests lungs as the organ most affected by smoking and in same time it is most in case of coronavirus infections. As world wide suffer from Covied_19 infections and its effects on human health and lungs characterstics espicaly FVC, this reasrsh couled be appled for measuring FVC value and predect paients lose of lung volume and its distortion effects in weeks post of patient invections.

## REFERENCES

[1] D. J. Norris, *Beginning Artificial Intelligence with the Raspberry Pi*. 2017.

[2] C. Robert, " 　　　　　　Machine Learning, a Probabilistic Perspective," *Chance*, vol. 27, no. 2, pp. 62–63, 2014, doi: 10.1080/09332480.2014.914768.

[3] P. D. Sugiyono, "済無No Title No Title," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2016, doi: 10.1017/CBO9781107415324.004.

[4] K. Suzuki, "Pixel-based machine learning in medical imaging," *Int. J. Biomed. Imaging*, vol. 2012, 2012, doi: 10.1155/2012/792079.

[5] A. C. P. L. F. de Carvalho and A. A. Freitas, "A tutorial on multi-label classification techniques," *Stud. Comput. Intell.*, vol. 205, no. July, pp. 177–195, 2009, doi: 10.1007/978-3-642-01536-6_8.

[6] J. Rutkowski, J. Siekmann, R. Tadeusiewicz, and L. A. Zadeh, *Artificial Intelligence and soft computing - ICAISC 2004*. 2004.

[7] M. Garg, N. Prabhakar, A. Gulati, R. Agarwal, and S. Dhooria, "Spectrum of imaging findings in pulmonary infections. Part 1: Bacterial and viral," *Polish J. Radiol.*, vol. 84, pp. e205–e213, 2019, doi: 10.5114/pjr.2019.85812.

[8] "osic-pulmonary-fibrosis-progression-basic-eda @ www.kaggle.com." [Online]. Available: https://www.kaggle.com/piantic/osic-pulmonary-fibrosis-progression-basic-eda#notebook-container.

[9] J. Latif, C. Xiao, A. Imran, and S. Tu, "Medical imaging using machine learning and deep learning algorithms: A review," *2019 2nd Int. Conf. Comput. Math. Eng. Technol. iCoMET 2019*, pp. 1–5, 2019, doi: 10.1109/ICOMET.2019.8673502.

[10] J. Latif, C. Xiao, A. Imran, and S. Tu, "Medical imaging using machine learning and deep learning algorithms: A review," *2019 2nd Int. Conf. Comput. Math. Eng. Technol. iCoMET 2019*, no. March, pp. 1–5, 2019, doi: 10.1109/ICOMET.2019.8673502.

[11] M. Zhang, Y. Zhou, J. Zhao, Y. Man, B. Liu, and R. Yao, "A survey of semi- and weakly supervised semantic segmentation of images," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4259–4288, 2020, doi: 10.1007/s10462-019-09792-7.

[12] A. Kornberg, *DNA replication*, vol. 263, no. 1. 1988.

[13] N. Nakata, "Recent technical development of artificial intelligence for diagnostic medical imaging," *Jpn. J. Radiol.*, vol. 37, no. 2, pp. 103–108, 2019, doi: 10.1007/s11604-018-0804-6.

[14] M. H. Mohana, "Reinforced concrete confinement coefficient estimation using soft computing models," *Period. Eng. Nat. Sci.*, vol. 7, no. 4, pp. 1833–1844, 2019, doi: 10.21533/pen.v7i4.947.

[15] D. Ps, "Ps3.1," pp. 1–34.

[16] M. Tan and Q. V Le, "EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks," 2019.

[17] B. Du, Š. A. N. Drev, I. Old, Ř. I. C. H. Kodym, and S. I. Sor, "BRNO UN I VERS I TY OF TECHNOLOGY FACULTY OF I NFORMAT I ON TECHNOLOGY DEEP LEARN I NG MODEL UNCERTA I NTY I N MED I CAL Master ' s Thesis Specification," 2019.

[18] F. I. Agu and C. E. Onwukwe, "Modified Laplace Distribution, Its Statistical Properties and Applications," *Asian J. Probab. Stat.*, no. June, pp. 1–14, 2019, doi: 10.9734/ajpas/2019/v4i130104.

[19] W. J. Braun, J. Stafford, and P. Brown, "Data sharpening via firth's adjusted score function," *Stat. Probab. Lett.*, vol. 165, p. 108831, 2020, doi: 10.1016/j.spl.2020.108831.