



PROGRESS PREDICTION

6° WEEK

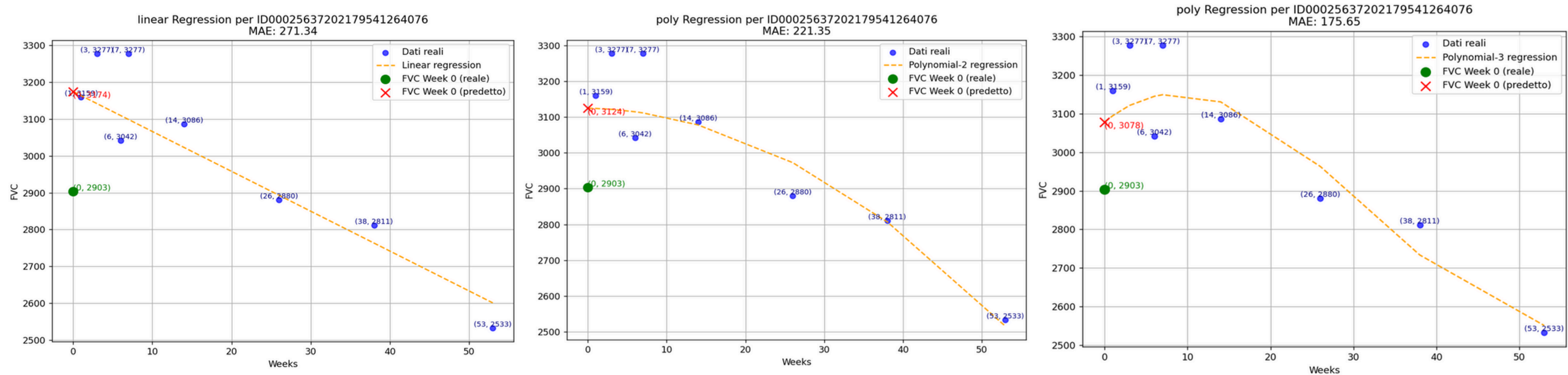
Baseline

3 models:

- **Linear Regression**
- **Polynomial regression (degree 2)**
- **Polynomial regression (degree 3)**

Metric (MAE)	Linear Regression	Polynomial 2	Polynomial 3
Average	127.62	93.95	67.76
Max value	557.10	359.04	175.65
Min value	1.08	2.27	6.88
Std	172.49	114.96	60.51
Median	30.70	32.08	41.26

Baseline comparison



Highlighting how data is nonlinear , therefore a cubic model approximates it substantially better than the linear one.

Feature Extraction

How are features extracted in Kaggle solution?

- **11 slices**
- **Chosen beween the 30% - 60%**

Is it the best way?

Problem → Different number of slices for patient

- **Low margin : 12 slices**
- **Upper margin : 1018 slices**
- **Average # slices : 187**

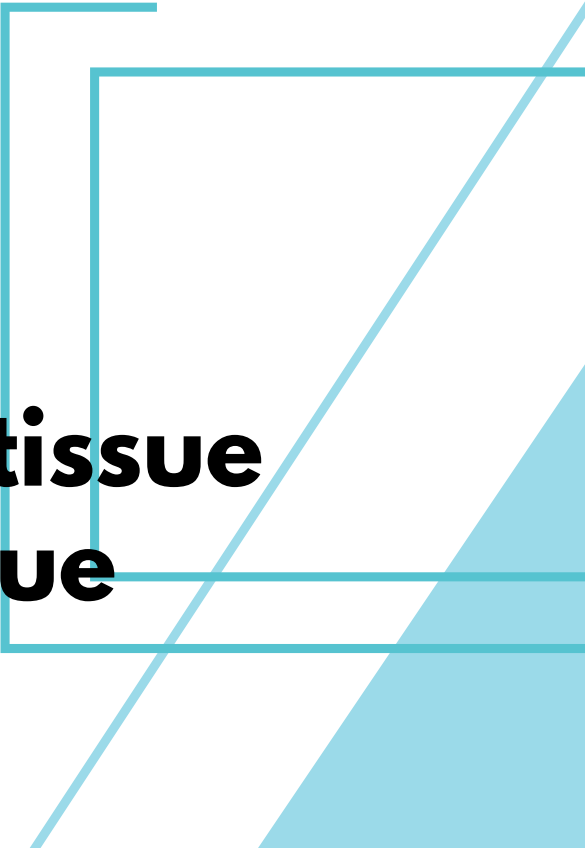


Feature Extraction

Kaggle

Combination of handcrafted features + CNN extracted features

Extracted features (based on the kaggle solution) :

- **SliceThickness (through metadata)**
 - **PixelSpacing (through metadata)**
 - **NumImgBw5Prec (number of slices between percentile)**
 - **ApproxVol_30_60**
 - **Avg_NumTissuePixel_30_60**
 - **Avg_Tissue_30_60**
 - **Avg_Tissue_thickness_30_60**
 - **Avg_TissueByTotal_30_60** → % of the entire image that is tissue
 - **Avg_TissueByLung_30_60** → % of the lung area that is tissue
- 

Feature Extraction

Possible adding of more features :

- **Mean, Median, Skew, Kurthosis, HAA of non zero pixels in image**
- **midMean, midMedian, midSkew, midKurthosis, midHAA of pixels in the largest CT slice of lung mask**

Mean → average value higher if fibrous tissue is present

Skew → fibrous lung is skewed to the right (normal lung to the left)

Kurthosis → peak of the low attenuation pixels is much lower

HAA → high-attenuation area - percentage of lung voxels between -600 and -250 Hounsfield Units, associated to inflammation

Feature Extraction

In a discussion, it was highlighted:

- **FVC highly correlated to age,height, sex**

Possibility to add other informations like:

- **Height → derived from age and gender assuming that in the competition a european cohort had been diagnosed**
 - **BMI → based on the chest circumference**
 - **Chest circumference as standalone criteria**
- 

CNN

Use a pretrained model (EfficientNet/U-Net):

- **extract complex patterns from images**
- **concatenate handcrafted features**
- **Predict Radiological progression and FVC decline**
 - **Ground Truth?**
 - **Option 1 : From CNN predict FVC \rightarrow % difference**
 - **Option 2: Leverage Survival Analysis ,without FVC prediction , just Event occurred or not.**
Example, if through data of patient decline %FVC > 10% that patient signed as progressed \rightarrow Event = 1 , so correlate features to if the Event happens or not (Bad for %FVC decline)