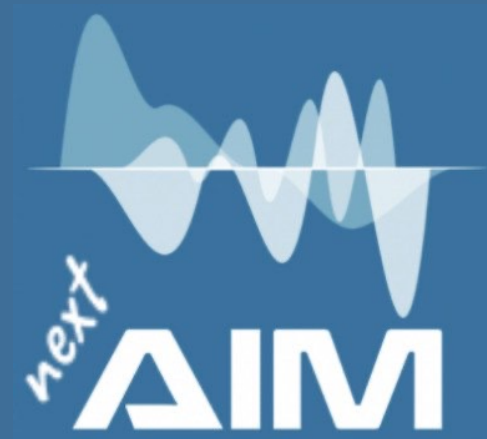


Artificial Intelligence in Medicine: next steps



**COVID-19 severity prediction based on radiomics
features obtained by
exploiting the LungQuant segmentation output**

C. Scapicchio for the Covid19-WG
INFN Sezione di Pisa



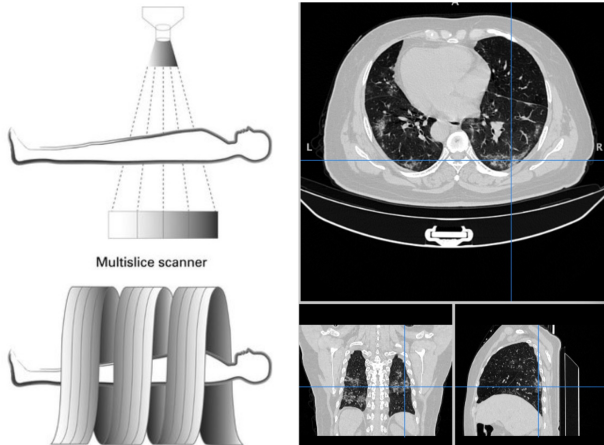
Outline

- AIM - Covid19-WG
- Il software di segmentazione LungQuant
- Predizione di gravità
 - Dataset
 - Machine Learning pipeline
 - Analisi delle features radiomiche
 - Combinazione dei dataset
 - Effetto del sito di acquisizione
- Prossimo step e limiti dello studio
- Conclusioni

AIM - Covid19-WG [PI,MI,PV,GE,FI,PA,CA]

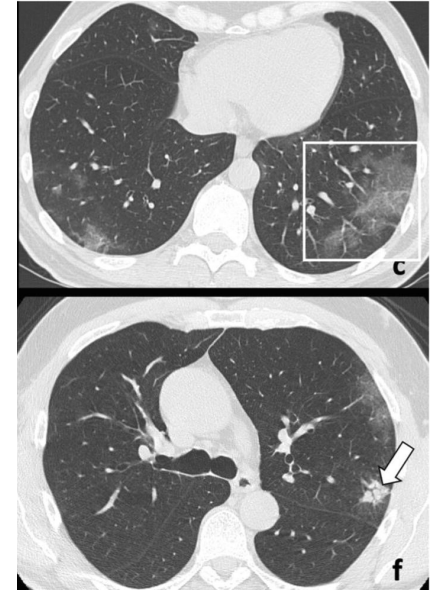
Caratteristiche sia qualitative che quantitative delle CT toraciche possono essere usate per definire la gravità della polmonite da COVID-19

CT images (3D)



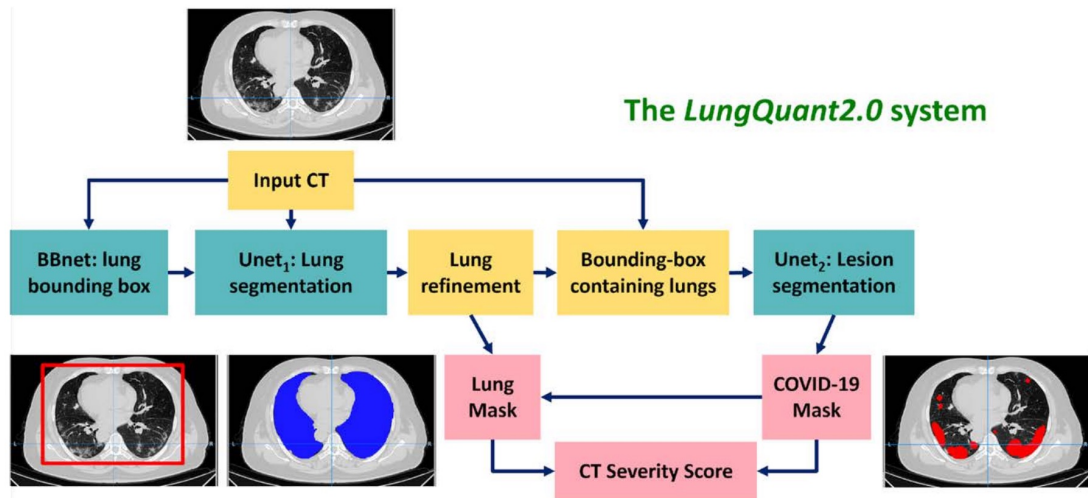
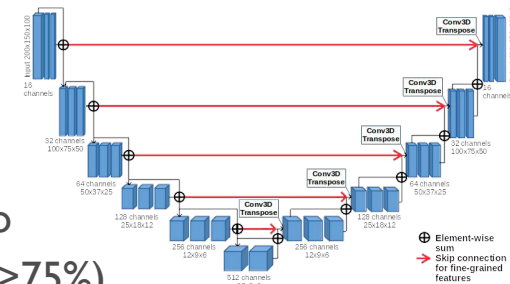
Segni di progressione critica della malattia in pazienti COVID-19:

- **ground glass opacities (GGO)**
- **consolidations**



Il software di segmentazione LungQuant

- U-net in cascata allenate per la segmentazione
 - Maschera del polmone
 - Maschera della lesione COVID-19
 - CT-Severity Score (CTSS), indice di gravità del polmone coinvolto dall'infezione: 1 (<5%), 2 (5%-25%), 3 (25%-50%), 4 (50%-75%), 5 (>75%)



$$L = Dice_{loss} + CE_{weighted}$$

$$CE_{weighted} = w(x) \sum_{x \in \Omega} \log(M_{true}(x) \cdot M_{pred}(x))$$

[F.Lizzi et al. Quantification of pulmonary involvement in COVID-19 pneumonia by means of a cascade of two U-nets: training and assessment on multiple datasets using different annotation criteria. International Journal of Computer Assisted Radiology and Surgery, 2021.]

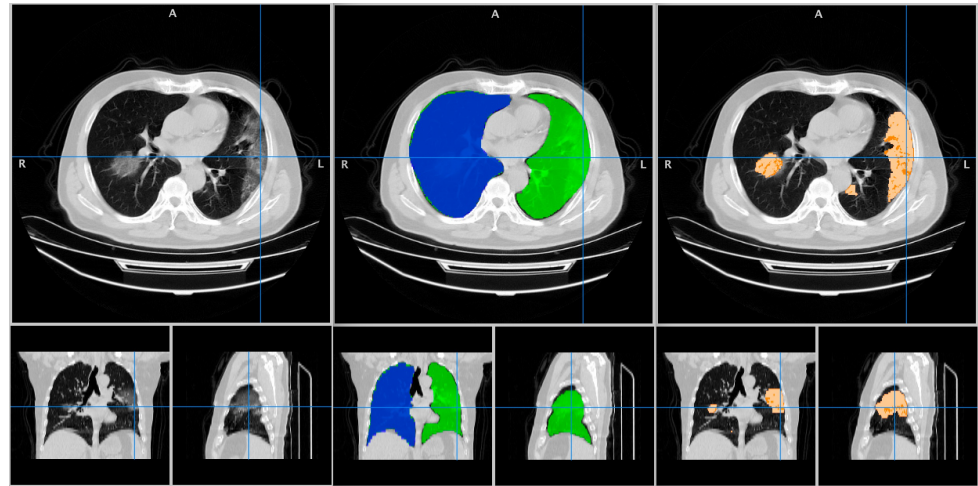
Il software di segmentazione LungQuant

● Versione 2.0

- CNN basata su regressione all'inizio della pipeline per predire il bounding box attorno ai polmoni.
- Funzione per separare polmone destro e sinistro.
- Soglia per distinguere GGO da consolidamenti.

	sDSC (5mm)	vDSC
Lung	0.97 ± 0.01	0.96 ± 0.01
<u>COVID Lesion</u>	0.83 ± 0.07	0.69 ± 0.08

CT-SS
prediction:
Accuracy: 80%

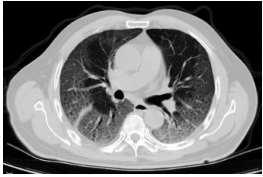


table_results

ID	lung_volume_mm3	lesions_volume_mm3	Lesions_to_Lung_ratio	CTSS	consolidation_vol	ground_glass_vol	ground_glass_R_ratio	ground_glass_L_ratio	consolidation_R_ratio	consolidation_L_ratio
0 volume-covid19-A-0041	6354312.42	305574.50	0.04	1	15576.48	311896.54	0.01	0.08	0.02	0.00
0 volume-covid19-A-0319	4396079.51	514796.34	0.11	2	50270.53	929051.62	0.27	0.15	0.01	0.01
0 volume-covid19-A-0120	3378643.30	48209.30	0.01	1	6962.34	82493.92	0.00	0.04	0.00	0.00
0 volume-covid19-A-0003	2778919.56	413522.35	0.14	2	45464.49	736089.43	0.02	0.43	0.00	0.02
0 volume-covid19-A-0251	4230179.70	560512.83	0.13	2	25525.16	1069975.33	0.16	0.32	0.00	0.01

Il software di segmentazione LungQuant

120 CT scans
(TCIA public
database)



LungQuant

	lung_volume_mm3	lesions_volume_mm3	Lesions_to_Lung_ratio	CTSS
-A-0041	6354312.42	305574.50	0.04	1
-A-0319	4396079.51	514796.34	0.11	2
-A-0120	3378643.30	48209.30	0.01	1
-A-0003	2778919.56	413522.35	0.14	2
-A-0251	4230179.70	560512.83	0.13	2

Quantification

ID	LESION_TYPE_INDEX	BILATERAL_INDEX	BASAL_INDEX
A-0037	0,137	0,447	37
A-0311	0,198	0,041	61
A-0291_0	0,224	0,193	31
A-0327	0,292	0,351	60

$$\text{LESION_TYPE_INDEX} = \frac{\text{consolidation_volume}}{\text{lesion_volume_mm3}}$$

$$\text{BILATERAL_INDEX} = \frac{|(R_{\text{consolidation}} + R_{\text{ggo}}) - (L_{\text{consolidation}} + L_{\text{ggo}})|}{\text{lesion_volume_mm3}}$$

BASAL_INDEX = the percentile of the distribution of the lung in which lies the median of the distribution of the lesion

Valutazione visiva da
parte di 14 radiologi
da 5 centri clinici
(Milano, Firenze,
Palermo, Pavia e Pisa)

Quality	AffVol(R/L)	Peripheral	Basal Predominant	Pleural Effusion	Type	Bilateral
Acceptable Low quality Optimal	% on lung volume	yes/no	yes/no	yes/no	CoGG MCo MGG CoO GGO	yes/no

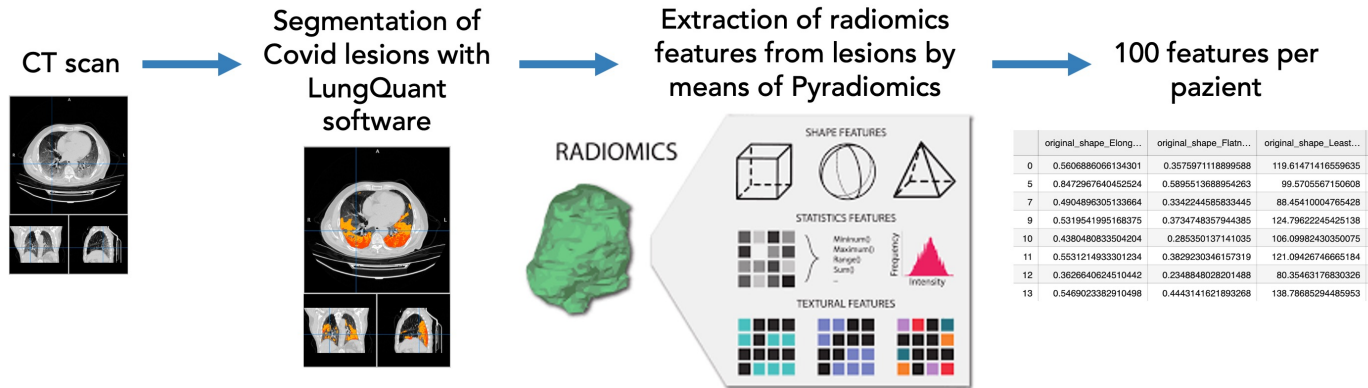
Analisi statistica
dell'accordo: analisi
AUC e regressione
non lineare.

[Chincarini A, Scapicchio C., et al., A multicenter evaluation of the LungQuant software for lung parenchyma characterization in COVID-19 pneumonia, submitted to European Radiology Experimental, UNDER REVIEW]

Metric	AUC	Youden cutoff	Inflection linear constrain [95% CL]
CTSS	0.98	0.10	0.20 [0.19 0.21]
Bilateral	0.85	0.60	0.64 [0.52 0.76]
Basal Predominant	0.90	0.34	0.32 [0.31 0.33]
Lesion Type	0.81	0.15	0.18 [0.12 0.25]

Predizione di gravità

Obiettivo: Sfruttare LungQuant per segmentare le lesioni da COVID-19 sulle CT polmonari ed estrarre da queste regioni delle quantità caratteristiche (feature radiomiche) per costruire modelli predittivi.



- Predire la prognosi clinica e quindi il decorso della malattia (grave/non grave) per i pazienti COVID-19 dopo un certo tempo dall'acquisizione della CT.
- Utilizzare un classificatore ML tradizionale allenato sulle features, quindi più semplice di una rete deep e più «spiegabile».

Dataset

Dataset privati di immagini CT di pazienti COVID e relativo esito clinico sulla gravità a circa un mese dall'acquisizione della CT.

Classe 0: NON Grave

Classe 1: Grave (intubato o deceduto)

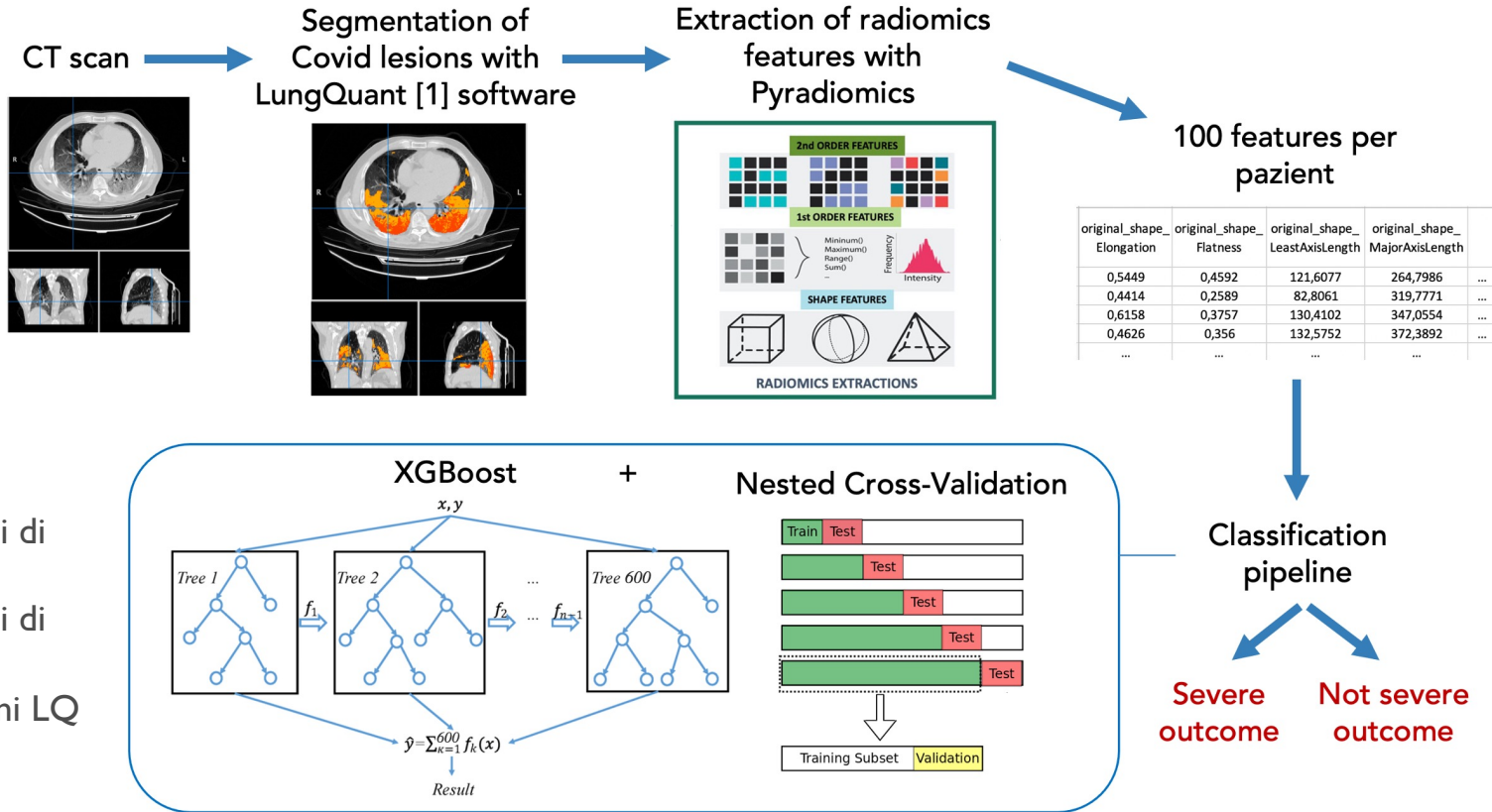
DATASET	# OF CASES	# NOT-SEVERE (CLASS 0)	# SEVERE (CLASS 1)
Palermo (B20f e B80f)	78	48	30
Pisa	69	45	24
Pavia	25	18	7
Firenze	100	50	50
Milano	369	171	198
Milano (25d < ΔT < 35d)	160	98	62
	432	259	173

Filtri di ricostruzione diversi tra i vari dataset, ma il più possibile sharp

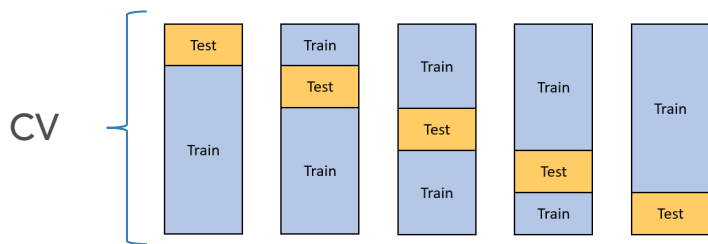
Machine Learning pipeline



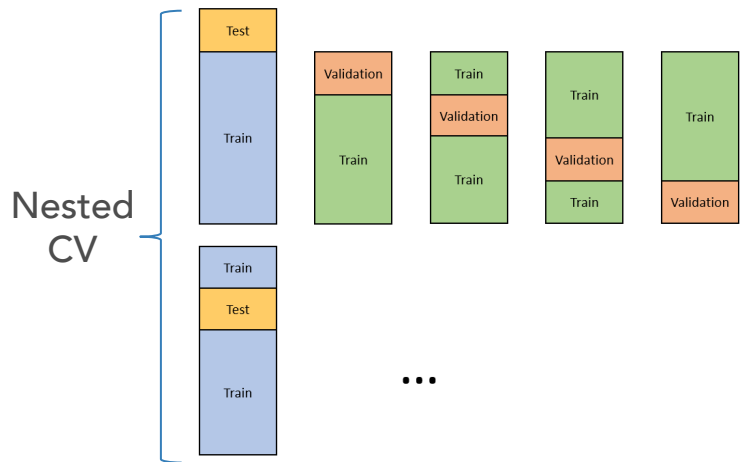
ML-INFN cloud



Machine Learning pipeline: Nested CV



La Cross-validation può essere usata sia per l'ottimizzazione degli iperparametri sia per stimare la generalizzabilità del modello. Tuttavia, usarla per entrambi gli scopi contemporaneamente può portare a una sottostima dell'overfitting dalla procedura stessa di ottimizzazione.

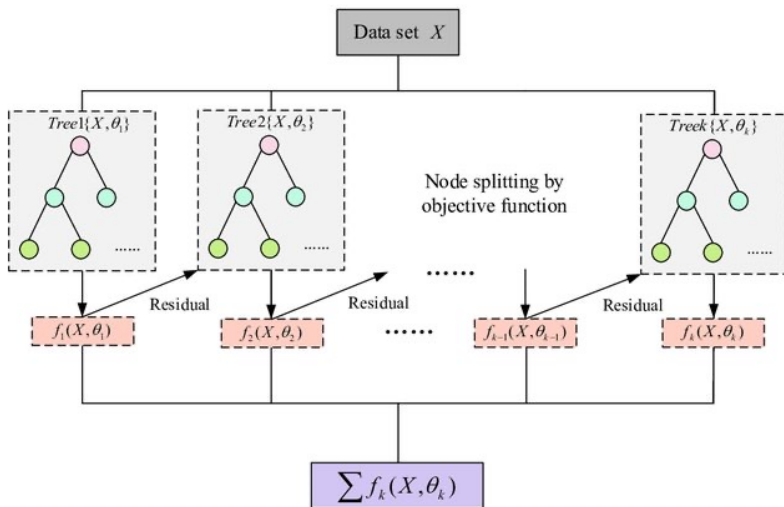


Serve un altro loop esterno di cross-validazione per valutare la generalizzabilità della performance.

- 10 trials
- 15% del dataset totale in Test
- 4-fold CV del restante 85% dei dati

Machine Learning pipeline: XGBoost classifier

EXTREME GRADIENT BOOSTING



Alberi decisionali aggiunti in forma sequenziale per correggere gli errori di predizione del modello precedente.

Il peso delle variabili predette erroneamente dall'albero viene aumentato e queste variabili vengono poi date in pasto al modello successivo.

L'ensemble di tutti gli alberi dà un modello più robusto e preciso.

Gradient Boosting perché usa un algoritmo gradient descent per minimizzare la loss quando vengono aggiunti i modelli.

	XGBoost	Logistic Regression
Interpretability	Good	Fair
Performance in a dataset with a linear relationship	Fair	Good
Performance in a large number of features	Good	Poor
Performance in a small training dataset	Good	Poor
Performance in a dataset with lot of outlier	Good	Poor (or remove the outlier from the dataset)
Performance in a skewed dataset	Good	Poor (or rebalance weight to the minor)
Performance in a continuous dataset	Bad	Good
Missing values handling in dataset	Good	Poor (or remove/patch the missing values from the dataset)
Ease of Decision Making	Automatically handles	Threshold can be set
Commonality	Rapidly trending up	Widely adopted

- gamma: [0.5, 1, 1.5, 2, 5]
- max_depth: [3, 4, 5]
- subsample: [0.6, 0.8, 1.0]
- colsample_bytree: [0.6, 0.8, 1.0]
- min_child_weight: [1, 5, 10]
- learning_rate=0.02
- n_estimators=600
- objective='binary:logistic'
- nthread=1

ML pipeline: Nested CV sui singoli siti

Nested CV – 100 features

	Palermo B20f	Palermo B80f	Pisa	Pavia	Firenze
roc_AUC score	0.72 ± 0.14	0.69 ± 0.14	0.58 ± 0.18	0.47 ± 0.29	0.91 ± 0.06

	Milano	Milano ($\Delta T > 25d$)	Milano ($25d < \Delta T < 35d$)
roc_AUC score	0.64 ± 0.06	0.80 ± 0.07	0.88 ± 0.04

Nested CV con Mutual Information feature selection in ogni trial (15 features)

	Palermo B20f	Palermo B80f	Pisa	Pavia	Firenze
roc_AUC score	0.78 ± 0.15	0.76 ± 0.13	0.51 ± 0.25	0.50 ± 0.25	0.93 ± 0.04

	Milano	Milano ($\Delta T > 25d$)	Milano ($25d < \Delta T < 35d$)
roc_AUC score	0.59 ± 0.05	0.73 ± 0.07	0.83 ± 0.05

ML pipeline: Nested CV sui singoli siti

Nested CV – 100 features

	Palermo B20f	Palermo B80f	Pisa	Pavia	Firenze
roc_AUC score	0.72 ± 0.14	0.69 ± 0.14	0.58 ± 0.18	0.47 ± 0.29	0.91 ± 0.06
	Milano	Milano ($\Delta T > 25d$)	Milano ($25d < \Delta T < 35d$)		
roc_AUC score	0.64 ± 0.06	0.80 ± 0.07	0.88 ± 0.04		

Nested CV con Mutual Information feature selection in ogni trial (15 features)

	Palermo B20f	Palermo B80f	Pisa	Pavia	Firenze
roc_AUC score	0.78 ± 0.15	0.76 ± 0.13	0.51 ± 0.25	0.50 ± 0.25	0.93 ± 0.04
	Milano	Milano ($\Delta T > 25d$)	Milano ($25d < \Delta T < 35d$)		
roc_AUC score	0.59 ± 0.05	0.73 ± 0.07	0.83 ± 0.05		

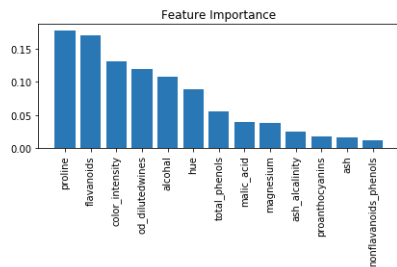
Analisi delle features radiomiche

Analisi delle features più significative per la predizione della gravità confrontando i diversi dataset e diversi metodi di selezione delle features

mR·MR
EFFICIENT FEATURE SELECTION

**minimum Redundancy –
Maximum Relevance**

$$score_i(f) = \frac{relevance(f | target)}{redundancy(f | features \text{ selected until } i - 1)}$$



Feature Importance

Mutual Information

`sklearn.feature_selection.mutual_info_classif`

Per poter realizzare un ranking delle features è stato dato un nuovo score a ciascuna feature, poichè le migliori features selezionate possono cambiare nei vari trials. Lo score totale su tutti i trials sarà dato dalla somma degli score per i singoli trial.

Misura la riduzione di incertezza per una variabile dato il valore noto di un'altra variabile.

Considero l'unione delle features selezionate in tutti i 10 trials.

Analisi delle features radiomiche

**minimum Redundancy –
Maximum Relevance
(k=15)**

Features in comune tra
FI – MI – PA (B80f)

gldm_DependenceNonUniformity
gldm_GrayLevelNonUniformity
glrlm_GrayLevelNonUniformity
glrlm_RunLengthNonUniformity
glszm_GrayLevelNonUniformity
glszm_SizeZoneNonUniformity

Features in comune tra
FI – MI – PA (B80f) - PI

gldm_GrayLevelNonUniformity
glrlm_GrayLevelNonUniformity

**Gray Level
Non-Uniformity:**
Measures the
similarity of
gray-level intensity
values in the image

**Feature importance
(k=15)**

glrlm_GrayLevelNonUniformity

**Mutual Information
(unione sui trial)**

gldm_DependenceNonUniformity
gldm_LowGrayLevelEmphasis
gldm_SmallDependenceLowGrayLevelEmphasis
glrlm_GrayLevelNonUniformity
glrlm_RunLengthNonUniformity
glrlm_ShortRunLowGrayLevelEmphasis
glszm_GrayLevelNonUniformity

Features in comune tra
FI – MI – PA (B80f) – PI - PV

gldm_SmallDependenceLowGray
LevelEmphasis

Analisi delle features radiomiche

Features in comune tra mRMR – Feature Importance – Mutual Information

Milano

firstorder_Energy
firstorder_Maximum
firstorder_Range
gldm_DependenceNonUniformity
glrlm_GrayLevelNonUniformity
glrlm_LongRunHighGrayLevelEmphasis
glrlm_RunLengthNonUniformity
glszm_SizeZoneNonUniformity

Firenze

firstorder_10Percentile
firstorder_Median
firstorder_RootMeanSquared
gldm_GrayLevelNonUniformity
glrlm_GrayLevelNonUniformity
glszm_GrayLevelNonUniformity
glszm_LowGrayLevelZoneEmphasis
glszm_SmallAreaLowGrayLevelEmphasis
shape_MeshVolume
shape_SurfaceArea
shape_VoxelVolume

Palermo B80f

gldm_DependenceNonUniformity
glrlm_GrayLevelNonUniformity
glrlm_RunLengthNonUniformity
glszm_HighGrayLevelZoneEmphasis
glszm_SizeZoneNonUniformity
shape_SurfaceArea

Pisa

firstorder_InterquartileRange
glcm_Idmn
glcm_MaximumProbability
gldm_DependenceVariance
shape_Flatness

Pavia

firstorder_90Percentile
firstorder_RobustMeanAbsoluteDeviation
glcm_ClusterTendency
shape_Flatness

Combinazione dei dataset

Nested CV su dataset Firenze + Milano (100 features)

roc_auc	accuracy	precision	recall	f1-score
0.83 ± 0.06	0.75 ± 0.07	0.71 ± 0.11	0.71 ± 0.11	0.75 ± 0.07

Nested CV con Mutual Information feature selection in ogni trial (26 features) (13 comuni a tutti i trial)

roc_auc	accuracy	precision	recall	f1-score
0.83 ± 0.07	0.76 ± 0.07	0.73 ± 0.10	0.69 ± 0.11	0.75 ± 0.07

Nested CV su dataset Firenze + Milano + Palermo B80f (100 features)

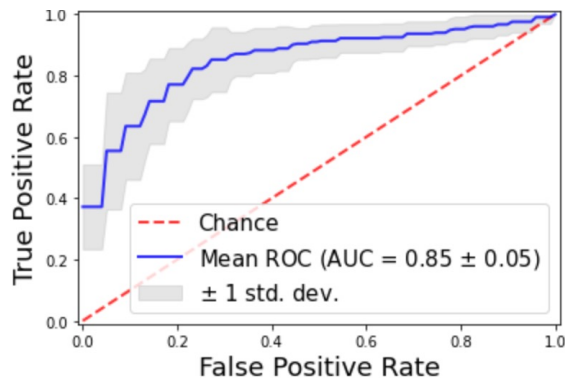
roc_auc	accuracy	precision	recall	f1-score
0.82 ± 0.05	0.77 ± 0.05	0.75 ± 0.08	0.68 ± 0.06	0.77 ± 0.05

Combinazione dei dataset

Cross-Validazione (Firenze + Milano)
con best parameters ricavati dalla nested CV

roc_auc

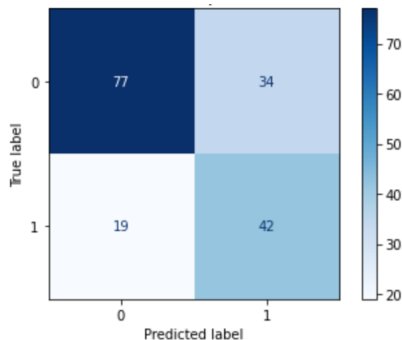
0.85 ± 0.05



Train (Firenze + Milano)
Test (Palermo B80f + Pisa + Pavia)

roc_auc

0.71



Train (Firenze + Milano)
Test (Palermo B80f)

roc_auc

0.77

Train (Firenze + Milano)
Test (Pisa)

roc_auc

0.64

Train (Firenze + Milano)
Test (Pavia)

roc_auc

0.65

Combinazione dei dataset

Nested CV su dataset Firenze + Milano + Palermo B20f + Pisa + Pavia (100 features)

roc_auc	accuracy	precision	recall	f1-score
0.78 ± 0.06	0.73 ± 0.05	0.65 ± 0.06	0.66 ± 0.10	0.73 ± 0.05

Nested CV con Mutual Information feature selection in ogni trial (35 features) (14 comuni a tutti i trial)

roc_auc	accuracy	precision	recall	f1-score
0.77 ± 0.03	0.72 ± 0.04	0.65 ± 0.06	0.65 ± 0.10	0.71 ± 0.04

Features in comune tra
mRMR – Feature Importance – Mutual Information

gldm_DependenceNonUniformity
 gldm_GrayLevelNonUniformity
 glrlm_GrayLevelNonUniformity
 glrlm_RunLengthNonUniformity
 glszm_GrayLevelNonUniformity

Combinazione dei dataset

Firenze + Milano + Palermo B20f + Pisa + Pavia (100 features)



Split Train set (90%) – Test set (10%)

Stratificando sulla classe di gravità e sul sito



**Nested CV su Train set
(388 casi) (0:232, 1:156)**

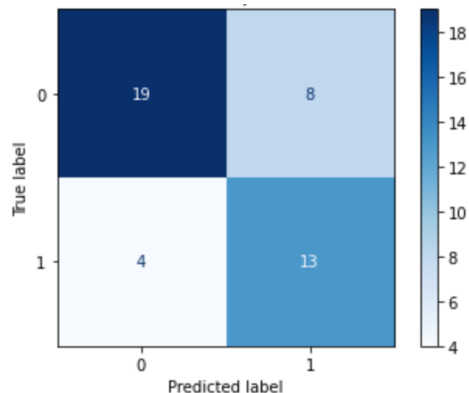


**Training su tutto il Train set
(388 casi) e test su Test set
indipendente (44 casi)**

con best parameters ottenuti dalla
nested CV

roc_auc

0.77 ± 0.06



roc_auc	accuracy	precision	recall
0.80	0.73	0.62	0.76

Effetto del sito di acquisizione

Poiché i dataset sono stati uniti senza precedente pre-processing o armonizzazione, verifico se il sito può essere un fattore confondente, costruendo un classificatore di sito.

Creo un dataset completamente bilanciato nella classe sito e nella classe gravità : 97 casi

Li seleziono dai singoli dataset in maniera random.

DATASET	# OF CASES	# NOT-SEVERE (CLASS 0)	# SEVERE (CLASS 1)
Palermo (B20f)	20	10	10
Pisa	20	10	10
Pavia	17	10	7
Firenze	20	10	10
Milano ($25 < \Delta T < 35$)	20	10	10
	97	50	47

Effetto del sito di acquisizione

Classificatore a 5 classi (FI – MI – PA – PI – PV)

Due modelli indipendenti usati entrambi in Cross-Validation

ARTIFICIAL NEURAL NETWORK

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 10)]	0
dense (Dense)	(None, 64)	704
dense_1 (Dense)	(None, 64)	4160
dropout (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 5)	325

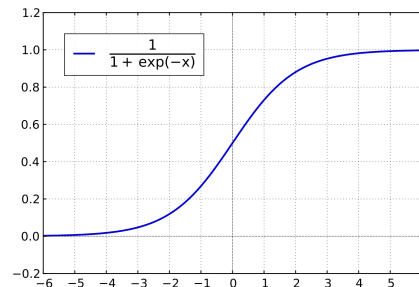
=====

Total params: 5,189
Trainable params: 5,189
Non-trainable params: 0

roc_auc

0.92 ± 0.04

LOGISTIC REGRESSION



roc_auc

0.87 ± 0.03

Il sito di acquisizione viene riconosciuto facilmente e potrebbe quindi avere un effetto confondente sulla predizione della gravità!

I siti che vengono confusi più facilmente sono Firenze e Milano

Effetto del sito di acquisizione

mRMR: features più significative per la predizione del sito di acquisizione sul dataset bilanciato ed eterogeneo di 97 casi

firstorder_Minimum
glcm_Autocorrelation
glcm_Id
glcm_Idm
glcm_Imc1
glcm_InverseVariance
glrlm_LongRunEmphasis
glrlm_RunLengthNonUniformityNormalized
glrlm_RunPercentage
glrlm_ShortRunEmphasis

Performance dei classificatori di sito selezionando solo queste 10 features:

Tutte features legate alla distribuzione dei livelli di grigio o alla texture



Presumibilmente legate al diverso filtro di ricostruzione utilizzato nei diversi siti di acquisizione

ANN	LR
roc_auc	roc_auc
0.92 ± 0.03	0.87 ± 0.02

Effetto del sito di acquisizione

Performance dei classificatori di sito selezionando solo le features di **shape**:

Classificatore a 5 classi (FI – MI – PA – PI – PV)

Due modelli indipendenti usati entrambi in Cross-Validation

ARTIFICIAL NEURAL NETWORK

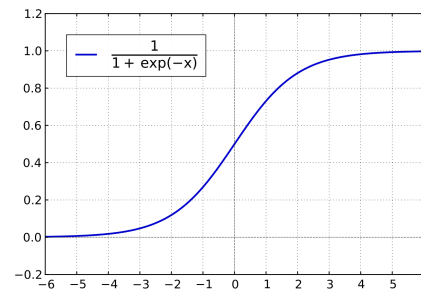
Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 10)]	0
dense (Dense)	(None, 64)	704
dense_1 (Dense)	(None, 64)	4160
dropout (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 5)	325

=====
 Total params: 5,189
 Trainable params: 5,189
 Non-trainable params: 0

roc_auc

0.45 ± 0.05

LOGISTIC REGRESSION



roc_auc

0.42 ± 0.04

Usando infatti solo le features di shape (14 features), il sito non viene identificato!

Prossimo step

- Il prossimo step è ridurre la variabilità dovuta ai diversi protocolli di acquisizione/ricostruzione prima di predire la gravità (ARMONIZZAZIONE)



Un rescaling o un filtering sulle immagini richiede tecniche di post-processing e spesso una riduzione nella risoluzione spaziale dell'immagine e quindi una qualità dell'immagine non ottima per successivi studi radiomici e quantitativi.



Armonizzazione delle feature estratte: calcolo degli **z-score indipendentemente per ogni sito, ComBat** (dopo aver verificato che siano soddisfatte le condizioni per utilizzarlo, per es. Kolmogorov-Smirnov (KS) test).



Explainability a seguito dell'armonizzazione

Limiti dello studio

- LungQuant allenato su dati pubblici non permette di ottenere segmentazioni perfette sulle immagini cliniche degli ospedali.

	OP	name	LungOK	LungBad	LungOUT	NoHTinLung	NoGGOinLung	GGOok	GGOmiss	GGOfail	GGOnotLung
1	LB	DECESSI_100021741_SMDC_3.0_B31f_105030833_20201214132236_e1.nii.gz	0	1	1	1	1	0	1	1	1
2	LB	DECESSI_100034676_Torace_SMDC_3.0_B70f_106004282_20210103173553_e1.nii.gz	1	0	0	1	1	0	1	0	0
3	LB	DECESSI_100056692_Torace_SMDC_3.0_B70f_106052343_20210319145935_e1.nii.gz	0	1	1	1	0	0	1	1	1
4	LB	DECESSI_10006517_Torace_3.0_BI57_1_103990483_20200320144418_e1.nii.gz	1	0	0	0	0	0	1	0	0
5	LB	DECESSI_10007645_SMDC_3.0_B70f_104018485_20200523001943_e1.nii.gz	0	1	1	0	0	0	1	1	0
6	LB	DECESSI_100134346_Torace_SMDC_3.0_B70f_105020038_20201124221220_e1.nii.gz	1	0	0	0	0	1	0	0	0
7	LB	DECESSI_100162391_Torace_1.0_BI57_1_103989664_20200318152401_e1.nii.gz	1	0	0	0	0	0	1	0	0
8	LB	DECESSI_100162421_PARENCHIMA_53685_20200511162707_e1.nii.gz	1	0	0	0	0	0	1	0	0

- In alcuni casi la segmentazione del polmone si estende oltre l'organo di interesse e le GGO possono essere così individuate in tessuti non polmonari, o il polmone sano non è segmentato.
 - Tendenzialmente, l'algoritmo di segmentazione sembra più "restrittivo" nel classificare GGO.
- Si tratta di dati acquisiti a inizio pandemia.

Conclusioni

- Nonostante le limitazioni, le **performance** sulla predizione di gravità sono particolarmente **buone**, soprattutto considerando singolarmente i dataset più numerosi di **Firenze** e **Milano**.
- L'**analisi multicentrica** effettuata mediante la combinazione dei diversi dataset porta a **risultati interessanti**, considerando che per ora **i dati non sono stati armonizzati** o pre-processati in alcun modo e avendo evidenziato che il sito di acquisizione ha un **effetto confondente**.
- Alcune **features** che descrivono principalmente la **disomogeneità spaziale** della lesione appaiono **robuste** rispetto al sito e al metodo di selezione, e potrebbero quindi essere specifiche della patologia. Stesse features individuate anche in altri lavori in letteratura.

Grazie per l'attenzione!

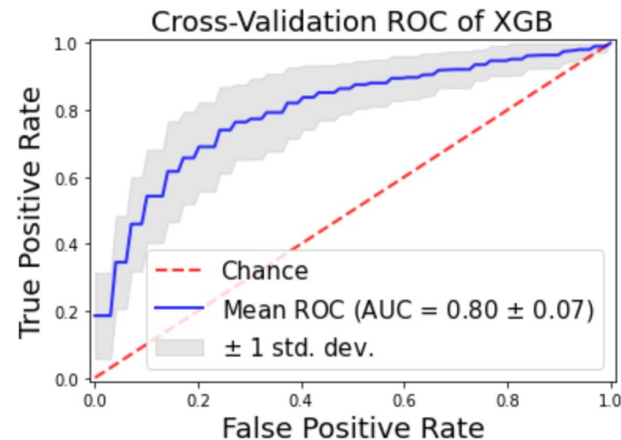
Back-up slides

Combinazione dei dataset

Cross-Validazione (Firenze + Milano + Palermo B80f)
con best parameters ricavati dalla nested CV

roc_auc

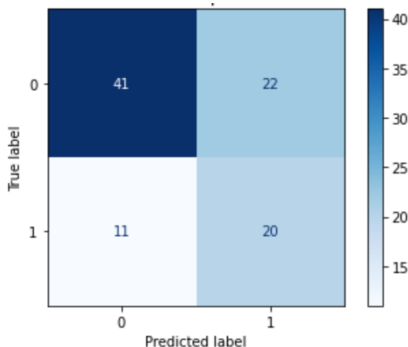
0.80 ± 0.07



Train (Firenze + Milano + Palermo B80f)
Test (Pisa + Pavia)

roc_auc

0.61



Train (Firenze + Milano + Palermo)
Test (Pisa)

roc_auc

0.61

Train (Firenze + Milano + Palermo)
Test (Pavia)

roc_auc

0.58

Parametri XGBoost



- `gamma`: parametro di regolarizzazione
- `max_depth`: profondità max dell'albero
- `subsample`: fraction of the training set that can be used to train each tree
- `colsample_bytree`: fraction of the features that can be used to train each tree.
- `min_child_weight`: minimum sum of instance weight (hessian) needed in a child.
- `n_estimators`: number of trees
- `objective`: objective function
- `nthread`: maximum number of threads available

Parametri per armonizzazione

	PALERMO	PISA	PAVIA	FIRENZE	MILANO
Filtro di ricostruzione	B20f B80f	LUNG	FC51 – B80f	Parenchima	Diversi filtri sharp
Slice thickness	1,0 mm	2,5 mm	[0,5 - 2,0] mm	[1,0 – 2,0] mm	[1,0 – 3,0] mm

kVp, Pixel Spacing, Manufacturer, FOV (acq – rec), sex, age

Pyradiomics Features

A Gray Level Run Length Matrix (GLRLM) quantifies gray level runs, which are defined as the length in number of pixels, of consecutive pixels that have the same gray level value. In a gray level run length matrix $\mathbf{P}(i, j|\theta)$, the $(i, j)^{\text{th}}$ element describes the number of runs with gray level i and length j occur in the image (ROI) along angle θ .

As a two dimensional example, consider the following 5x5 image, with 5 discrete gray levels:

$$\mathbf{I} = \begin{bmatrix} 5 & 2 & 5 & 4 & 4 \\ 3 & 3 & 3 & 1 & 3 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 2 & 2 & 2 & 3 \\ 3 & 5 & 3 & 3 & 2 \end{bmatrix}$$

The GLRLM for $\theta = 0$, where 0 degrees is the horizontal direction, then becomes:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \\ 4 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Let:

- N_g be the number of discrete intensity values in the image
- N_r be the number of discrete run lengths in the image
- N_p be the number of voxels in the image
- $N_r(\theta)$ be the number of runs in the image along angle θ , which is equal to $\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \mathbf{P}(i, j|\theta)$ and $1 \leq N_r(\theta) \leq N_p$
- $\mathbf{P}(i, j|\theta)$ be the run length matrix for an arbitrary direction θ
- $p(i, j|\theta)$ be the normalized run length matrix, defined as $p(i, j|\theta) = \frac{\mathbf{P}(i, j|\theta)}{N_r(\theta)}$

3. Gray Level Non-Uniformity (GLN)

$$GLN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_r} \mathbf{P}(i, j|\theta) \right)^2}{N_r(\theta)}$$

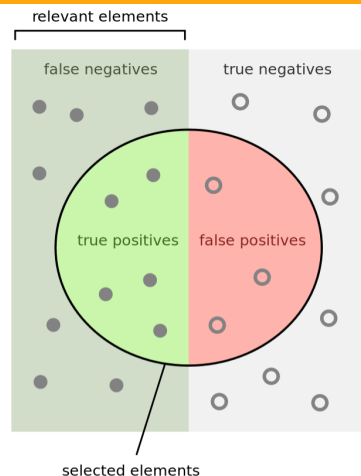
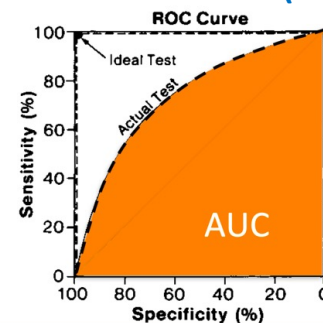
GLN measures the similarity of gray-level intensity values in the image, where a lower GLN value correlates with a greater similarity in intensity values.

Performance metrics

Performance evaluation

CONFUSION MATRIX		Predicted class	
		YES	NO
Actual class	YES	True Positive (TP)	False Negative (FN)
	NO	False Positive (FP)	True Negative (TN)

Area Under the ROC Curve (AUC)



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Sensitivity = True Positive Rate (TPR)

Specificity = 1 - False Positive Rate (FPR)

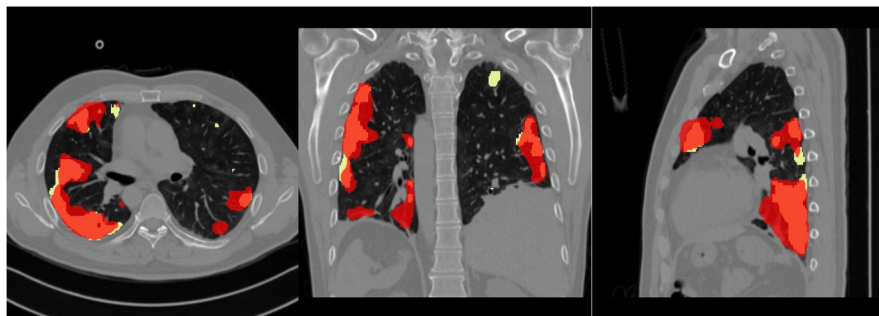
$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{accuracy} = \frac{TP + TN}{TP + FN + TN + FP}$$

Dipendenza dalla segmentazione



LungQuantSystem (LQ) (giallo)
LungQuantSystem2.0 (LQ2.0) (rosso)

