

Machine Learning for Adaptive RT

Dott. Gabriele Guidi, PhD

Dott. Nicola Maffei

Azienda Ospedaliero - Universitaria di Modena - Policlinico, Modena

guidi.gabriele@aou.mo.it - Phone: +39 059 422 5699

Machine Learning



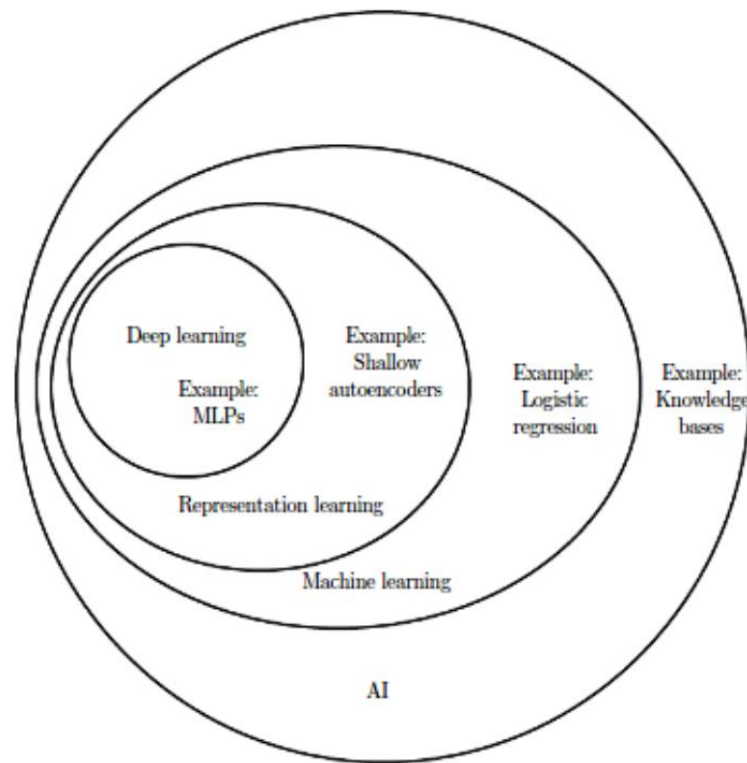


Figure 1.4: A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning, which is used for many but not all approaches to AI. Each section of the Venn diagram includes an example of an AI technology.

An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside [...]. The learning process may be regarded as a search for a form of behaviour which will satisfy the teacher (or some other criterion)."
A. Turing (1950)

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E ."
T. Mitchell (1997)

Tasks "T"

$$f = ?$$

- Classification: to specify which of k categories some input belongs to.

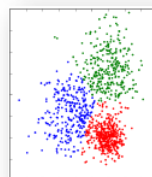
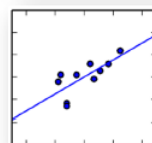
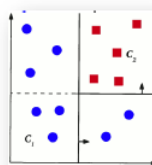
$$f: \mathbb{R}^n \rightarrow \{1, \dots, k\}$$

- Regression: to predict a set of numerical values given some input.

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

- Clustering: to specify the probability that a input belongs to one of the k clusters

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^k$$



Performance "P"

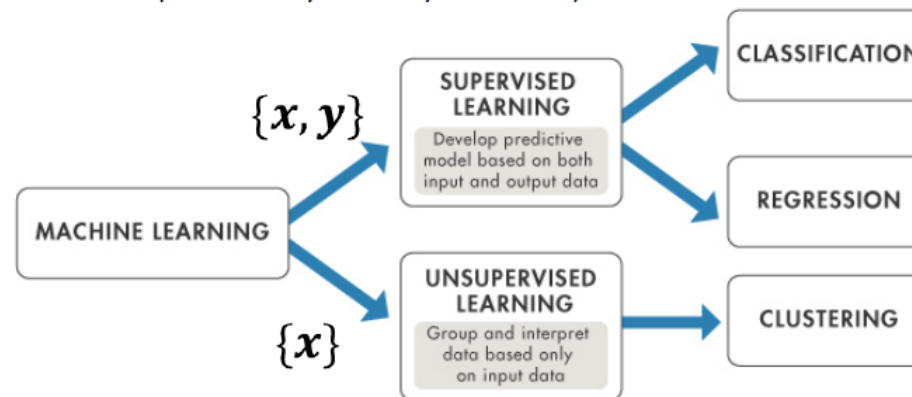
- Usually this performance measure P is specific to the task T being carried out by the system.
e.g. T =classification $\rightarrow P$ = accuracy (proportion of examples for which the model produces the correct output)
- It is important to know how well the machine learning algorithm performs on data that it has not seen before, since this determines how well it will work when deployed in the real world. These performance measures have to carry out using a **test set** of data that is separate from the data used for training the machine learning system
- The choice of performance measure may seem straightforward and objective, but it is often difficult to choose a performance measure that corresponds well to the desired behaviour of the system.

Artificial intelligence inside
radiotherapy treatment
planning systems

Experience “E”

Input $x \rightarrow f(x) = y$ Output

- Is a collection of features that have been quantitatively measured from some object or event that we want the system to process
- Machine learning can be divided, according to the nature of the data, into:
 - supervised learning
 - is used to estimate an unknown mapping from known (input, output) samples, where the output is labelled (e.g., classification and regression)
 - unsupervised learning
 - only input samples are given to the learning system (e.g., clustering and estimation of probability density function)

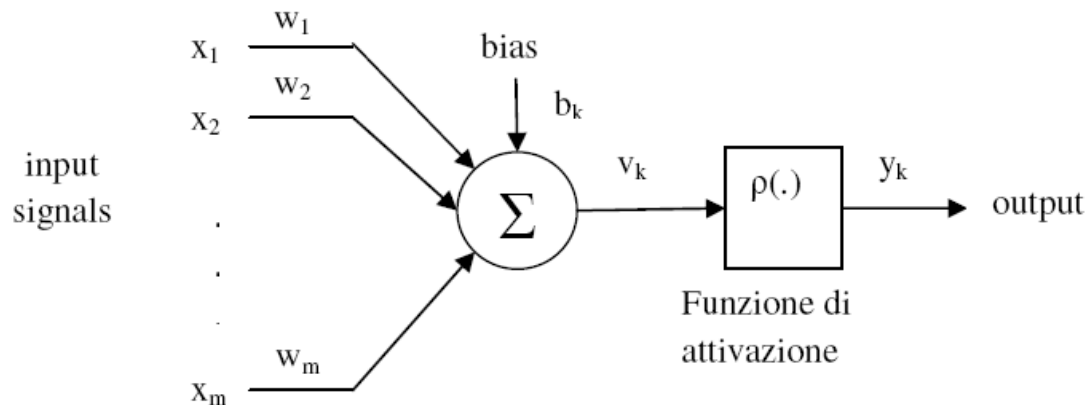


Artificial intelligence inside
radiotherapy treatment
planning systems

DEFINITION of Neural Network

expert systems that simulate the biological nervous system. They consist of an arbitrary number of nerve cells (neurons), connected together in a complex network, in which intelligent behavior emerges from the numerous interactions between interconnected units. In most cases, a neural network is an adaptive system that changes its structure based on external or internal information during the learning phase.

Some nodes receive information from the environment, others emit responses in the environment and others still communicate only with the units within the network: they are defined respectively input units (input), output units (output) and hidden units (hidden).



4 Fundamental Elements of a Neuron:

- 1) set of synapses (or links), characterized by their own "weight";
- 2) bias, with the purpose of raising or lowering the activation threshold of the function;
- 3) adder, which performs the weighted sum of the input signals of the neuron;
- 4) activation function, which limits the extent of neuron output.

In mathematical terms, the behavior of a **neuron** can be described by:

$$u_k = \sum_{j=1}^m w_{kj} x_j$$

$$y_k = \varphi(u_k + b_k)$$

x_1, x_2, \dots, x_m are the inputs,

$w_{k1}, w_{k2}, \dots, w_{km}$ are the synaptic weights of the connections between the inputs and the neuron k

u_k is the linear combination of the input signals

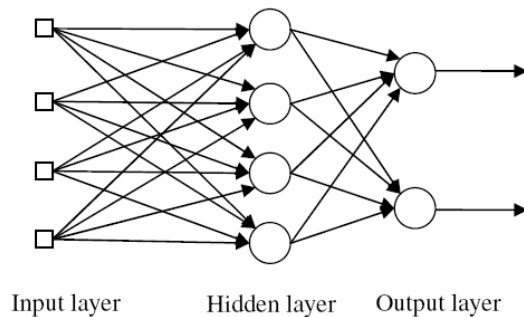
b_k is the bias

φ is the activation function

y_k is the output of the neuron.

Each unit becomes active if the total amount of signal it receives exceeds a certain threshold; each connection point also acts as a filter, which transforms the message received into an inhibitory or excitatory signal, increasing or decreasing its intensity, according to its individual characteristics.

- **Supervised learning.** If there is a set of data for training, including typical examples of inputs with their corresponding outputs, the network can learn to infer the relationship that links them. If the training is successful, the network learns to recognize the unknown relationship that links the input variables to the output variables and is therefore able to make predictions even where the output is not known a priori.
- **Unsupervised learning.** It is based on training algorithms that modify the weights of the network, referring exclusively to a set of data that includes only the input variables. These algorithms try to group the input data and therefore identify appropriate clusters representative of the same.
- **Reinforcement learning** (reinforcement learning). The algorithm aims to identify a modus operandi based on a process of observing the external environment; every action has an impact on the environment and the environment produces a feedback, which guides the algorithm itself in the learning process, providing in response an incentive or a disincentive as appropriate. The learning with reinforcement differs from the supervised one, since no input-output pairs of known examples are ever presented, nor is there any explicit correction of suboptimal actions.



Input layer: receive information from the environment

Hidden layer: communicate with the units within the network

Output layer: emit responses in the environment

Neural Networks (NN) learn from the external environment through an iterative process of adaptation of the weights of synaptic connections

Supervised learning

- **Known:** training data set
- **NN:** learn to recognize the relationship between input and output

Unsupervised learning

- **Known:** set of data containing input variables
- **NN:** identifies representative clusters

Reinforcement learning

- **Known:** observation of the external environment
- **NN:** identifies a modus operandi through feedback

An example: Nonlinear Autoregressive with External (Exogenous) Input (NARX)

Perceptron Multi Layer (MLP) networks implement a static mapping between input and output. Defining with $y(t)$ the output of the network at a given instant t , this depends solely on an input vector $x(t)$ at that instant of time:

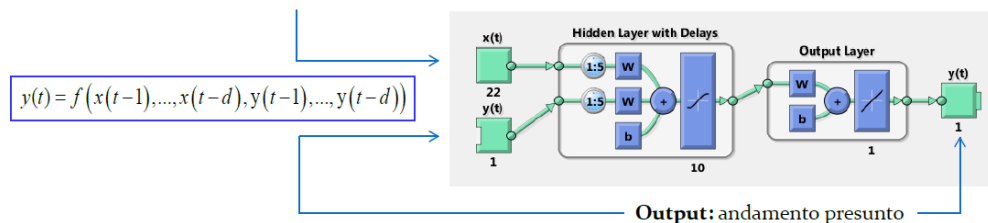
$$y(t) = f(x(t))$$

Recurrent Neural Networks (RNN) differ from the previous ones due to the presence of one or more cycles of local or global feedback allowing to implement a system dynamic with memory.

The Nonlinear Autoregressive with External (Exogenous) Input (**NARX**) is a network model with input / output architecture with feedback connections, in which the output is given by the non-linear function depending on the value of the output considered in the previous instants (with a delay d) and from the value of the exogenous variable, also observed in the previous instants:

$$y(t) = f(x(t-1), \dots, x(t-d), y(t-1), \dots, y(t-d))$$

Input: cell array di V o D di una ROI, di tutti i pazienti per tutte le sedute



Input e output NARX
(Volume - Parotide dx)

Input Data division			Network Architecture			
Training	Validation	Testing	num. Hidden neurons	num. Delay	Autocorrelation Error	Regression
70%	15%	15%	10	3	5,1%	0,98
				5	9,8%	0,97
50%	25%	25%	10	3	12,2%	0,95
				5	8,1%	0,97
30%	35%	35%	10	3	4,7%	0,89
				5	3,1%	0,88

Input e output NARX
(Dose - Parotide dx)

Input Data division			Network Architecture			
Training	Validation	Testing	num. Hidden neurons	num. Delay	Autocorrelation Error	Regression
70%	15%	15%	10	3	3,0%	0,97
				5	0,9%	0,95
50%	25%	25%	10	3	0,8%	0,83
				5	0,4%	0,94
30%	35%	35%	10	3	2,2%	0,82
				5	1,5%	0,84

open loop mode advantages
(compared to the close loop):

- since the forecast is available during the training phase, the use of the latter rather than a feedback with an estimated output makes the input more accurate
- the network thus presents a purely feed-forward architecture, which allows training based on a static backpropagation.

Newton's algorithm allows for convergence to local minima, as the weights are updated according to:

$$W(t+1) = W(t) - H^{-1}(t) \cdot g(t)$$

W is the matrix of the weights

H is the Hessian matrix of the error and g is the gradient.

This algorithm requires a significant computational capacity since, in the training phase, it is necessary to calculate at each step the matrix of the second derivatives of the error with respect to the weights (H).

Iterative Levenberg-Marquardt (L-M) provides the approximation of the Hessian matrix and the error gradient in the following way:

$$H = J^T J$$

$$g = J^T e$$

J is the Jacobian matrix, whose elements are the first derivatives of the error with respect to weights e is the error vector.

Finally, these approximations allow you to rewrite the weight matrix update law as follows:

$$W(t+1) = W(t) - [J^T J]^{-1} J^T e$$

In the NARX network, some features have **to be defined**:

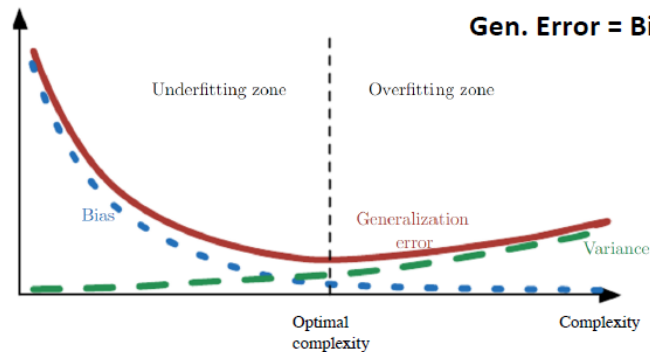
- Timesteps division:
 - Training: percentage of days chosen to train the Neural Network;
 - Validation: percentage of days used to verify the generalization of the network;
 - Testing: percentage of days used as evidence of the NARX on "new" data.
- Number of days of delay to be considered in the input feedback
- Number of hidden layers
- Number of nodes for each layer

Theorem II (Siegelmann et al., 1997): «NARX networks with a layer of hidden neurons having limited and saturation activation on one side and a layer of linear output neurons can simulate any completely connected recurrent network built with neurons having limited activation function and saturation on one side, except for a linear slowdown. »

Principle of minimization of structural risk: «if the number of neurons present in the hidden layers is increased excessively, there is the risk of undergoing an overfitting process (over-training), if instead it is reduced beyond a certain limit, there is the risk of looming in an underfitting (under training) »

Complexity

- The objective is to create a balance between complexity (increased model order) and the model ability to generalize to unseen data (**Occam's razor principle**)
- The complexity of a learning model increases with the number of input features (i.e., the dimension of the input feature vector space) and model parameters
- A straightforward approach is to make an educated guess based on experience and domain knowledge and then apply feature transformation (e.g., principal component analysis (PCA))



Keep in mind that:

- you need to characterize properly the nature of problem, in terms of the input data and the desired outputs.
- a good model cannot substitute for bad data, models are primarily built on approximations.

Advantages

- it can substitute for laborious and repetitive human effort.
- more significantly, it can potentially learn more complicated and subtle patterns in the input data than the average human observer is able to do, especially with large amounts of data.

.... both of these advantages are important to radiation therapy.

Artificial intelligence inside
radiotherapy treatment
planning systems

Andrea Botti
Medical Physics - AUSL - IRCCS Reggio Emilia

Automatic classification can be defined as the 'scientific discipline whose goal is the classification of *objects* into a number of categories or *classes*' (Theodoridis and Koutroumbas, 2006).

The following terms are equivalent:

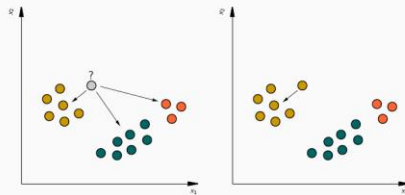
- *Automatic classification*
- *Pattern recognition*
- *Pattern classification*

Automatic classification is part of the broader discipline of *Machine learning*

F. Bianconi and B. Palumbo

Supervised classification (learning by examples)

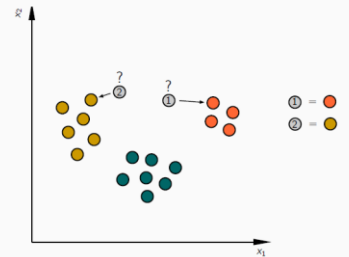
- Input:
 - A set of labelled patterns (*train set*)
 - A set of patterns to classify
- Output:
 - The labels (classes) of the patterns to classify



F. Bianconi and B. Palumbo

Nearest neighbour (1-NN) classifier

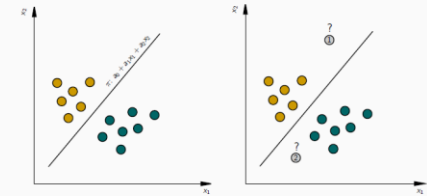
- Assign the pattern to classify to the class of the nearest pattern among those of the train set



F. Bianconi and B. Palumbo

Linear classifier

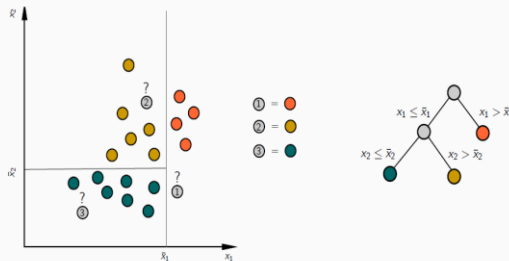
- Find the hyperplane (line when $d = 2$) that best separates the train patterns in two classes
- Assign the element to classify to the class of the corresponding half-space
- Suitable for one-class classification (can be extended to multi-class too)



F. Bianconi and B. Palumbo

Classification trees

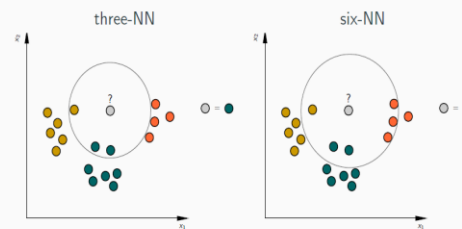
- Compute a set of thresholds that best split the feature space into sub-regions
- Determine the class of the pattern to classify by a cascade of binary (or n -ary) questions



F. Bianconi and B. Palumbo

k -nearest neighbours (k -NN) classifier

- Draw a bubble (circle when $d = 2$) centred on the pattern to classify that embraces k train patterns
- Assign the pattern to the most represented class



F. Bianconi and B. Palumbo

Other types of classifiers

There are many different types of classifiers, among them:

- Bayesian classifiers
 - Bayesian networks
 - Naïve Bayes (NB) classifier
- Neural networks
 - Multilayer perceptron (MLP)
- Random Forest (RF)
- Support Vector Machines (SVM)

F. Bianconi and B. Palumbo

Receiver Operation Characteristic (ROC)

In decision theory, the Receiver Operation Characteristic (**ROC**) curves are graphical schemes for a binary classifier and study the relationship between true alarms and false alarms, relating according to two axes: sensitivity (y) and 1-specificity (x).

Considering a 2 class prediction problem, and choosing a threshold value, which discriminates the positive and negative class, 4 possible solutions are possible, depending on the threshold value:

- True Positive (TP): the result of the prediction and the true value are positive;
- False Positive (FP): the result of the prediction is positive while the true value is negative;
- True Negative (TN): the result of the prediction and the true value are negative;
- False Negative (FN): the result of the prediction is negative while the true value is positive.

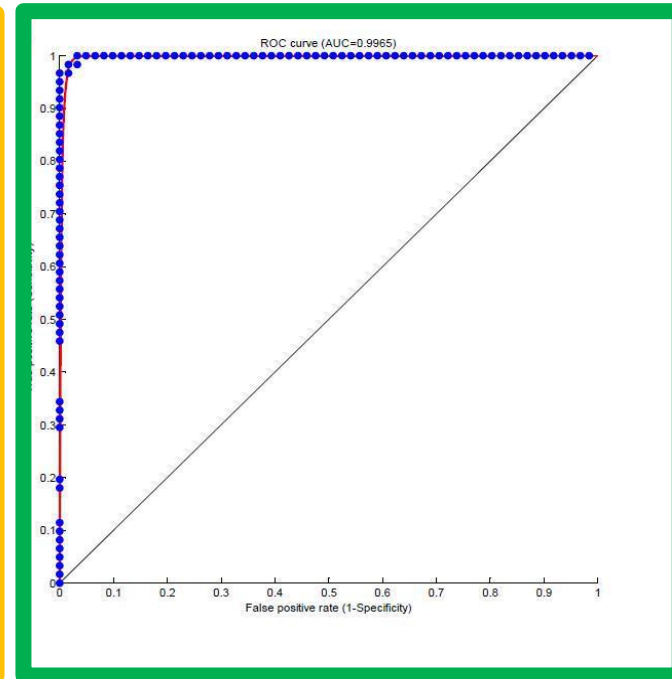
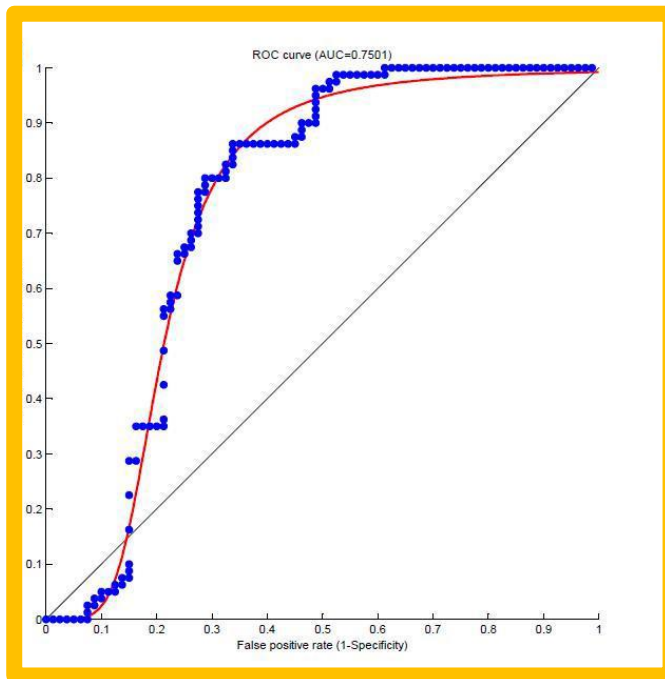
		Predicted class	
		○	●
True class	○	TP	FN
	●	FP	TN

- TP = *True positive*
- TN = *True negative*
- FP = *False positive*
- FN = *False negative*
- Sensitivity = $\frac{TP}{TP+FN}$
- Specificity = $\frac{TN}{TN+FP}$

A **ROC** curve is the graph of the set of pairs (FP, TP) for each possible threshold value, whose initial and final constraints are the pairs (0,0) and (1,1). The test carried out by analyzing the ROC curves has the ability to discriminate, for example, between a group of healthy and sick people.

Analyzing the area subtended by the curve (AUC), we obtain the probability that the test result carried out on an individual randomly extracted from the group of patients is higher than the one randomly extracted from the group of healthy people.

- if the ROC curve has a comparable trend with the 45 ° diagonal it is comparable to a random classifier
- if the ROC curve is systematically above the diagonal, we are better able to correctly classify the 2 cases



Data split

Problem: In most cases we do not have independent datasets, but just one

Solution: Split the dataset into two subsets for train and test.

Possible approaches:

- Full sampling
- Stratified sampling
- Leave-one-out

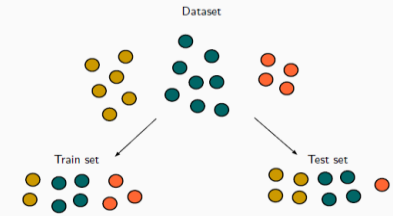
F. Bianconi and B. Palumbo



Data split: full sampling

Select a random fraction of the samples (e.g. 50%) for the train set and use the rest for the test set.

For a stable evaluation repeat the subdivision N times ($N \approx 50 - 500$) and average the results

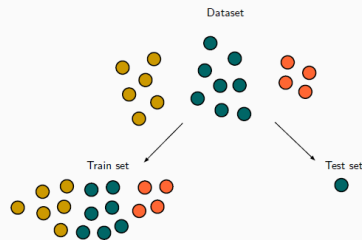


F. Bianconi and B. Palumbo

Data split: leave-one-out

Use all the samples but one for the train set and the remaining samples for the test set.

Repeat the procedure for all the samples (leaving one out each time) and average the results.



F. Bianconi and B. Palumbo

Data split: stratified sampling

Similar to full sampling, but with the constraint that the fraction of samples used for the train set (e.g. 50%) is the same for each class.



F. Bianconi and B. Palumbo

Physica Medica 32 (2016) 1659–1666



Original paper

A machine learning tool for re-planning and adaptive RT: A multicenter cohort investigation



G. Guidi^{a,b,*}, N. Maffei^a, B. Meduri^c, E. D'Angelo^c, G.M. Mistretta^a, P. Ceroni^a, A. Ciarmatori^{a,d}, A. Bernabei^a, S. Maggi^d, M. Cardinali^e, V.E. Morabito^d, F. Rosica^f, S. Malara^g, A. Savini^f, G. Orlandi^f, C. D'Ugo^g, F. Bunkheila^h, M. Bonoⁱ, S. Lappi^j, C. Blasi^h, F. Lohr^c, T. Costi^a

^aMedical Physics Department, Az. Ospedaliero-Universitaria di Modena, Italy

^bPhysics Department, Alma Mater Studiorum University of Bologna, Italy

^cRadiation Oncology Department, Az. Ospedaliero-Universitaria di Modena, Italy

^dMedical Physics Department, Az.Ospedaliero-Universitaria Ospedale Riuniti di Ancona, Italy

^eRadiation Oncology Department, Az.Ospedaliero-Universitaria Ospedale Riuniti di Ancona, Italy

^fMedical Physics Department, AUSL4 Teramo, Italy

^gRadiation Oncology Department, AUSL4 Teramo, Italy

^hRadiation Oncology Department, Az.Osp.Ospedali Riuniti Marche Nord di Pesaro, Italy

ⁱMedical Physics Department, Az.Osp.Ospedali Riuniti Marche Nord di Pesaro, Italy

^jRadiation Oncology Unit, Altnagelvin Hospital, Londonderry, United Kingdom

Physica Medica 31 (2015) 442–451



Original paper

A support vector machine tool for adaptive tomotherapy treatments: Prediction of head and neck patients criticalities



Gabriele Guidi^{a,c,*}, Nicola Maffei^{a,c}, Claudio Vecchi^c, Alberto Ciarmatori^{a,d}, Grazia Maria Mistretta^a, Giovanni Gottardi^a, Bruno Meduri^b, Giuseppe Baldazzi^c, Filippo Bertoni^b, Tiziana Costi^a

^a Medical Physics Department, Az. Ospedaliero-Universitaria di Modena, Italy

^b Radiation Oncology Department, Az. Ospedaliero-Universitaria di Modena, Italy

^c Physics Department, University of Bologna, Italy

^d Post-graduate School in Medical Physics, University of Bologna, Italy

Adaptive radiation therapy (**ART**) is an advanced field of radiation oncology.

Image-guided radiation therapy (**IGRT**) methods can support daily setup and assess anatomical variations during therapy, which could prevent incorrect dose distribution and unexpected toxicities. A **re-planning** to correct these anatomical variations should be done daily/weekly, but to be applicable to a large number of patients, still require time consumption and resources. Using unsupervised **machine learning** on retrospective data, we have developed a predictive network, to identify patients that would benefit of a re-planning.

Machine learning methods, for early cancer diagnosis, prediction of clinical complications and biological outcomes, could improve the effectiveness of RT with the aim to develop a daily personalized plan based on automatic validated processes:

Expert system classifier for adaptive radiation therapy in prostate cancer

Gabriele Guidi, Nicola Maffei, Claudio Vecchi, Giovanni Gottardi, Alberto Ciarmatori, Grazia Maria Mistretta, Ercole Mazzeo, et al.



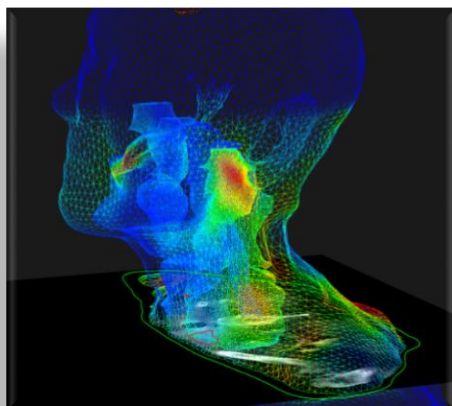
Australasian Physical & Engineering Sciences in Medicine
The Official Journal of the Australasian College of Physical Scientists and Engineers in Medicine

ISSN 0158-9938

Australas Phys Eng Sci Med
DOI 10.1007/s13246-017-0535-5

Australasian
Physical &
Engineering Sciences
in Medicine

CHALLENGES



Patient's anatomical variations:
Body – OARs - Target

Would need



Daily **re-evaluation** of the initial plan

But...

- Time Consuming
- Cost for the (staff + technological resources)



PREDICTIVE ANALYSIS



RE-PLANNING

Not, generally, **sustainable** for all patients in clinical practice
(...very soon we could do it...)

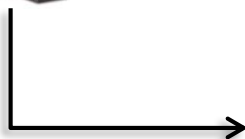
5 External Beam RT · **40** pts. = **200** re-plans/day

Actual Standard: **2·200** plans/year



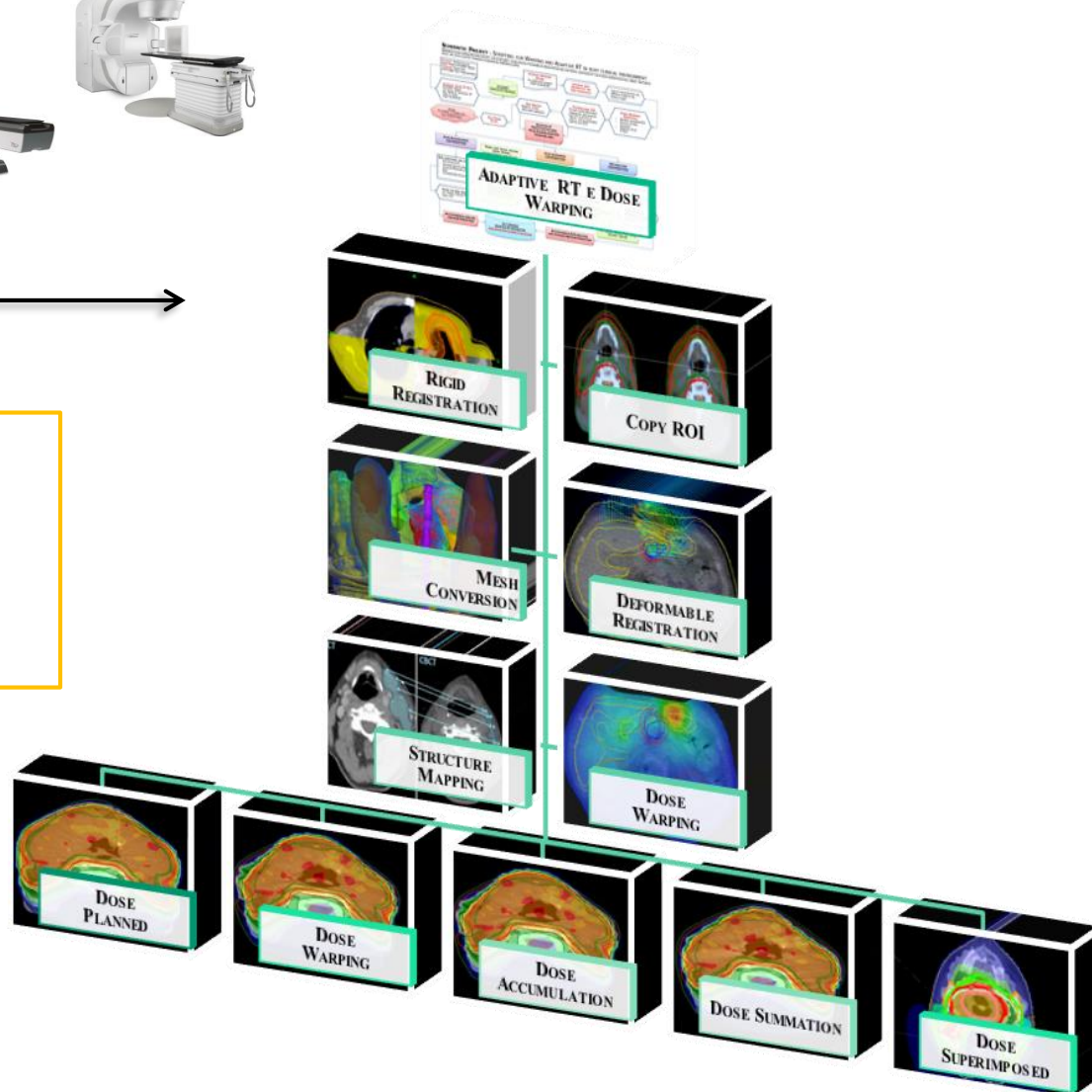
Clinical workload increment
= **2·000 %**





Needs for Clinical Practice:

1. IGRT
2. TPS with DIR
3. Clinical Validation





Off-line ART simulation:

ART tasks	M (min)
DICOMRT import + RIR + Copy ROI	50
DIR + Map ROI	240
Dose deformation	50
Statistical data extraction	350
TOTAL	690min

A – CPU (min)	CPU time gain
30	40%
180	25%
40	20%
60	83%
310min	168%

A – GPU (min)	GPU time gain
12	76%
62	74%
30	40%
28	92%
132min	282%

2 days for
1 pts. (30Fx)

=

30 pts. (1Fx)



3 pts. (30Fx)/day

=

90 pts. (1Fx)

Parallel calculation

Nightly batch mode

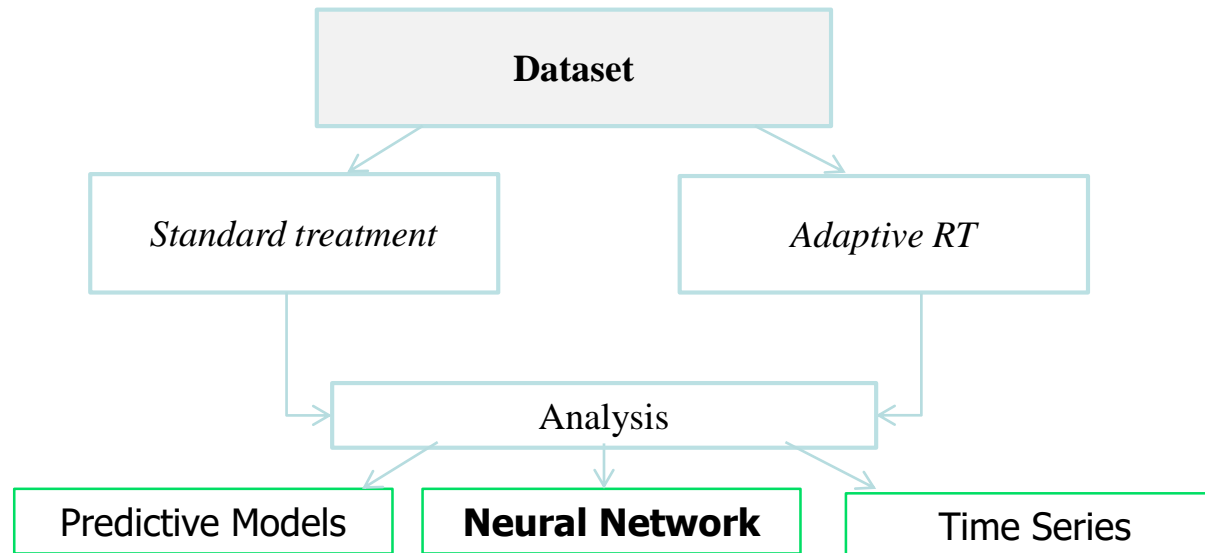


**>8 pts.
(30Fx)/day**

=

>240 pts. (1Fx)

Our approach



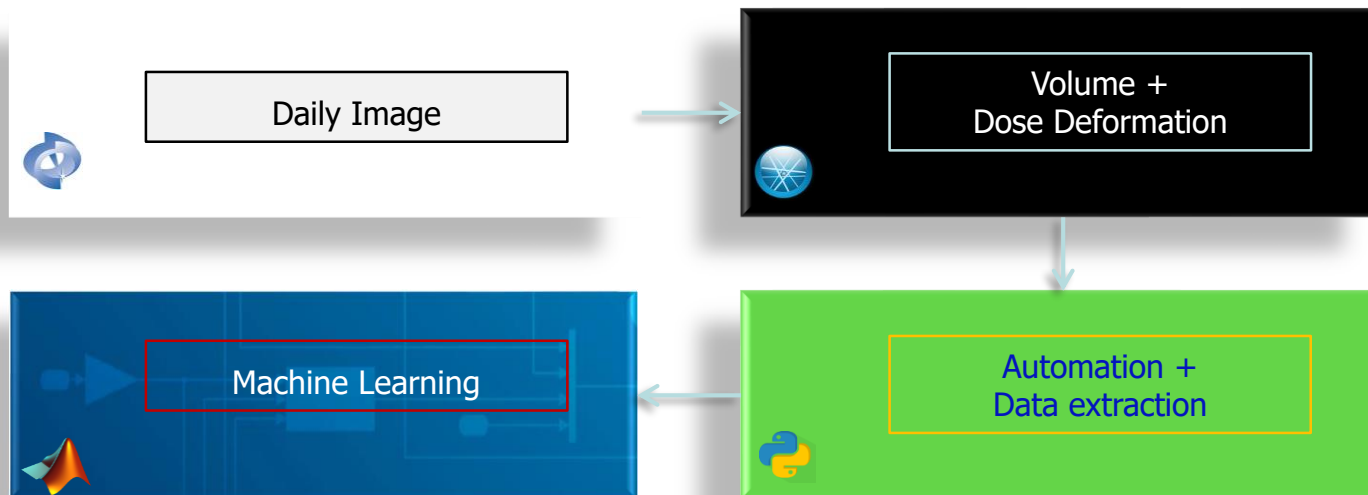
To quantify divergence between
Standard RT vs. ART

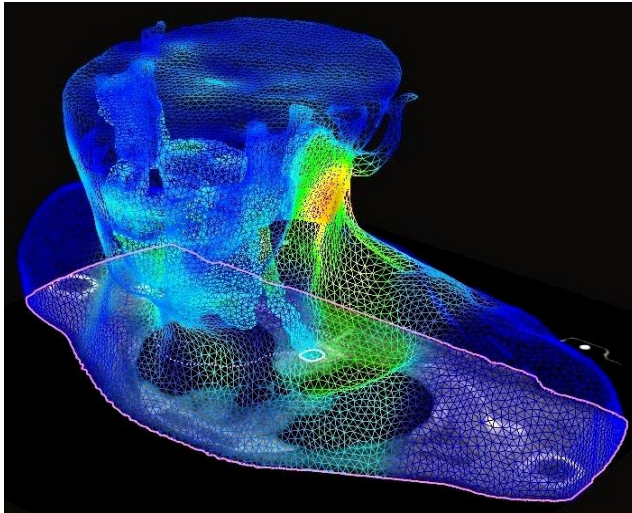
To identify cases not in line with
the average trend

To get information about ROIs
most affected by V/D deviation

To detect a temporal range for
re-planning

- IDENTIFY & FOLLOW ROIs MOST AFFECTED BY WARPING
- SELECT CASES ELIGIBLE FOR RE-PLANNING
- DETECT A TEMPORAL RANGE FOR RE-PLANNING

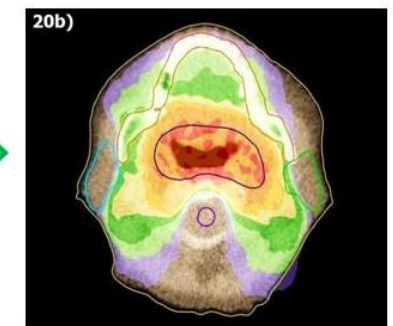
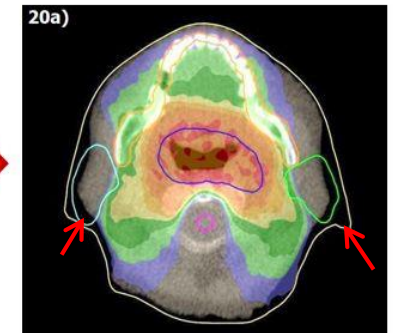
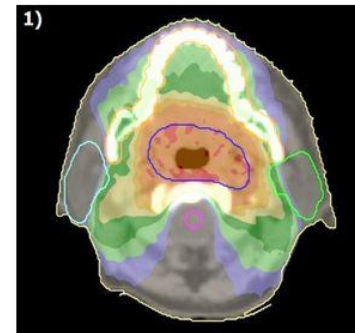




"Parotid glands are the ROI more susceptible to warping in H&N region"

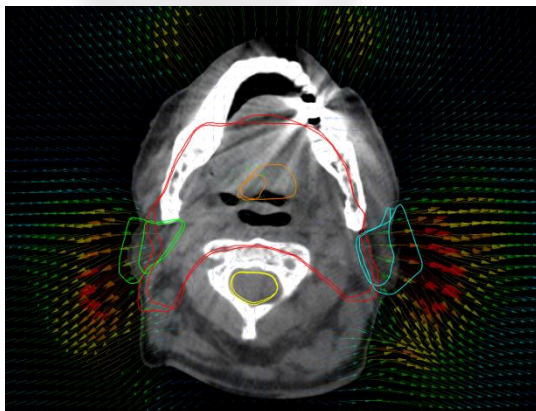
cit. [1]Lee; [2]Brouwer; [3]Fiorino; [4]Lee; [5]Olteanu;
[6]Scalco; [7]Marzi; [8]Guidi

LEARNING PHASE	PTS	PTS (%)	DAILY STUDIES	IMAGES
TRAINING	29	48.3%	870	≈52000
VALIDATION	12	20%	360	≈21000
Total	41	68.3%	≈1200	>70000



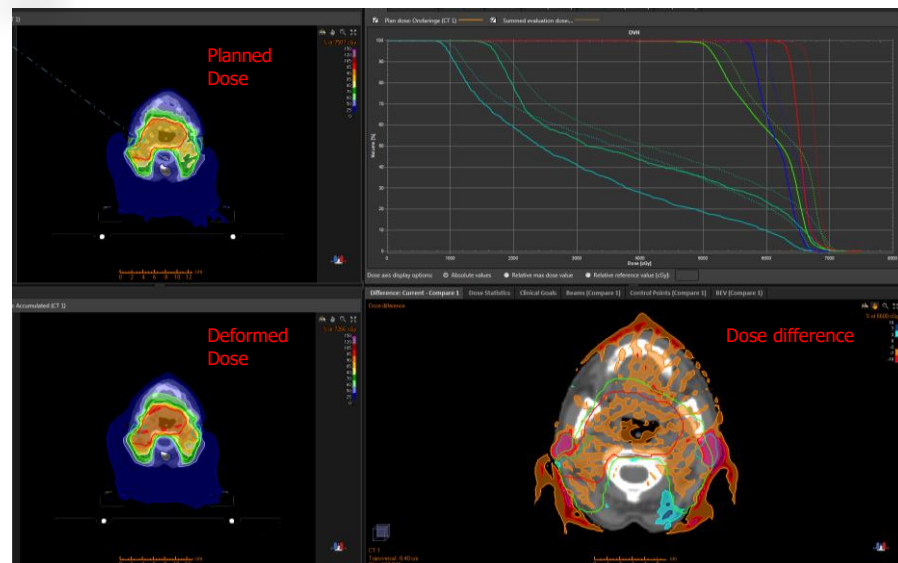
- ROI
- Selected week of therapy
- Daily anatomical + dosimetric variations of the ROI

Normalized ΔV



Normalized ΔD :

- D_{average}
- D_{99}
- D_{98}
- D_{95}
- D_{50}
- D_2
- D_1



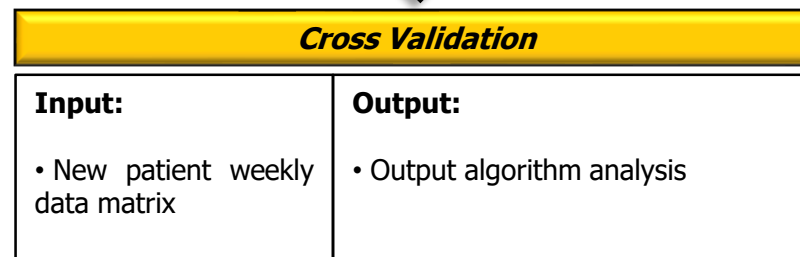
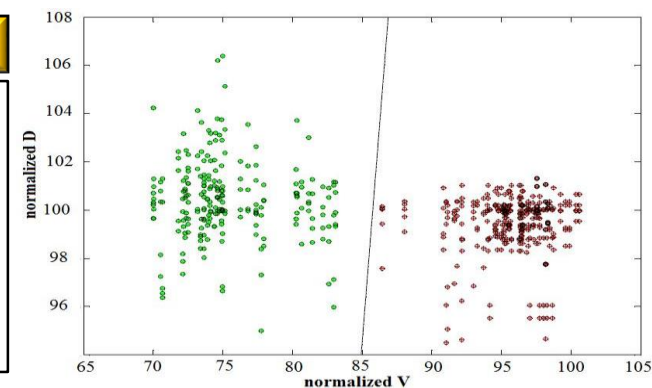
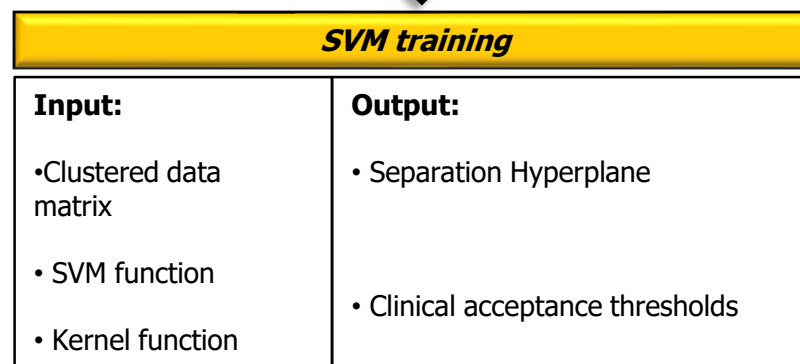
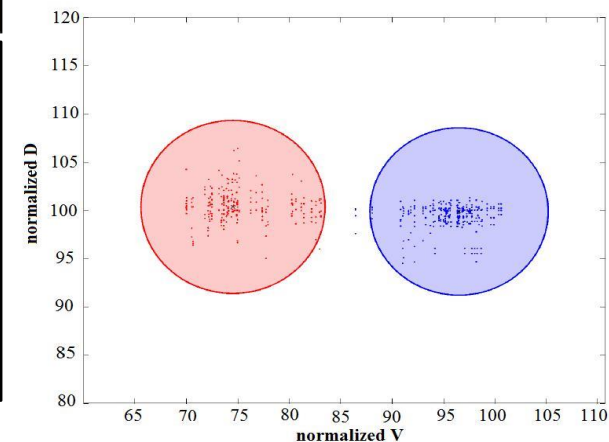
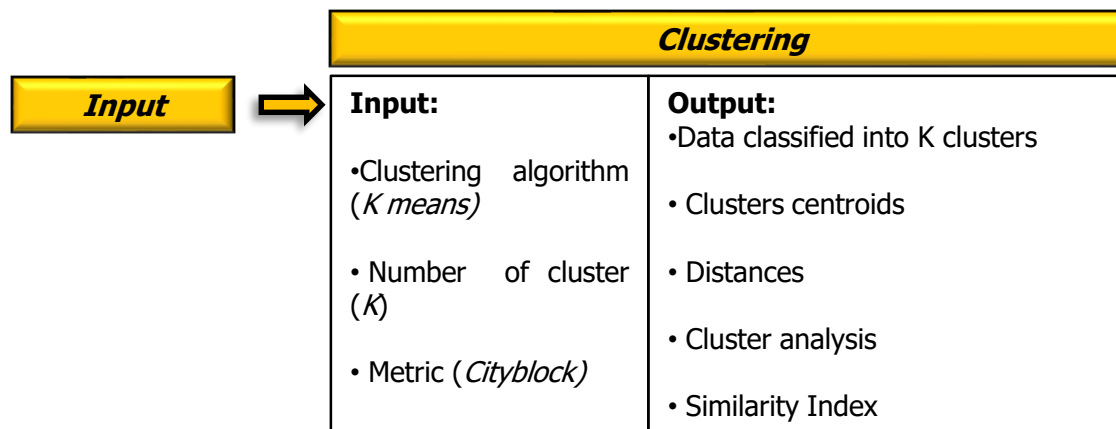
Algorithm architecture

The purpose of our study is to quantify, through an unsupervised learning on retrospective data, anatomical/dosimetric divergences that may occur during the weeks of treatments.

Simulations allowed monitoring each patient during the radiotherapy period, highlighting changes in planning treatment and its daily optimization.

The **input** nodes were obtained from dose-volume histogram (DVH) curves:

- The first network input variable to select is the ROI.
- The second input is the week of treatment in which carry out the analysis.
- The total volume (V) of the selected structure and doses (D) at different volume percentage: D99, D98, D95, Daverage, D50, D2, D1 of the selected structure must be uploaded into the algorithm. To avoid an incorrect training of the classifier, thresholds were applied to the input data. Clinical thresholds of serial organs are more restrictive than the limits applied to parallel OARs and target.



Cluster analysis, through k-means clustering algorithm, plays a research role of latent structures in order to deduce the most probable partition between the reference gold standard and the ART plan. Starting from an initial weekly set of input elements (x_1, \dots, x_N) , the aim of the algorithm is to cluster input data into K set $S \cup \{S_1, \dots, S_K\}$ in order to minimize the within-cluster sum of squares (WCSS):

$$\min \sum_{i=1}^K \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

with μ_i the mean value of the S_i cluster.

Initialization (2), assignment step (3) and update step (4) are imposed to achieve global optimization on the base of the following relationships:

$$\mu_1^{(1)}, \dots, \mu_K^{(1)} \quad (2)$$

$$S_i^{(t)} = \left\{ x_j : \|x_j - \mu_i^{(t)}\| \leq \|x_j - \mu_{i^*}^{(t)}\| \quad \forall i^* = 1, \dots, K \right\} \quad (3)$$

$$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad (4)$$

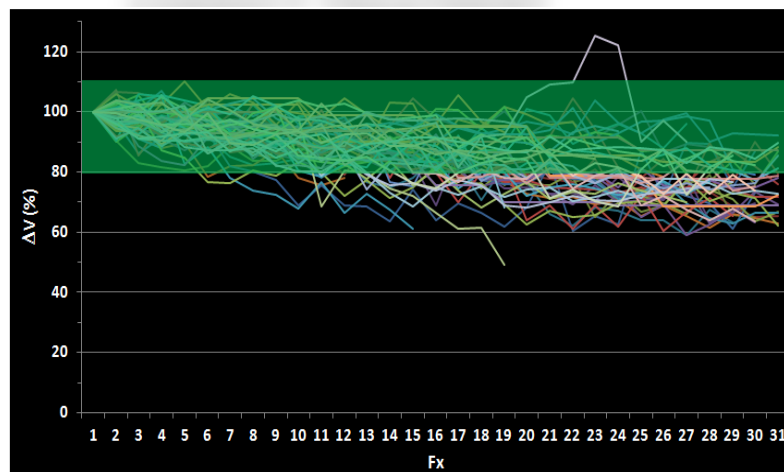
Data were divided into two macro clusters and a different initial seed set was used during reiterated runs. At the end of the first step, k-means algorithm returns a n-by-1 vector containing the cluster indices of each point, the WCSS and distances of each point from centroids. Cityblock metric (sum of absolute differences) is considered the most adequate distance from our data, considering each centroid like the median component wise of the points in that cluster.

A **SVM** function was then used to individuate the optimum hyper-plane between the 2 macro-samples: RT plan and ART simulations. Svmtrain MATLAB function uses an optimization method to identify support vectors s_i , weights a_i , and bias b to classify vectors x , according to Eq.

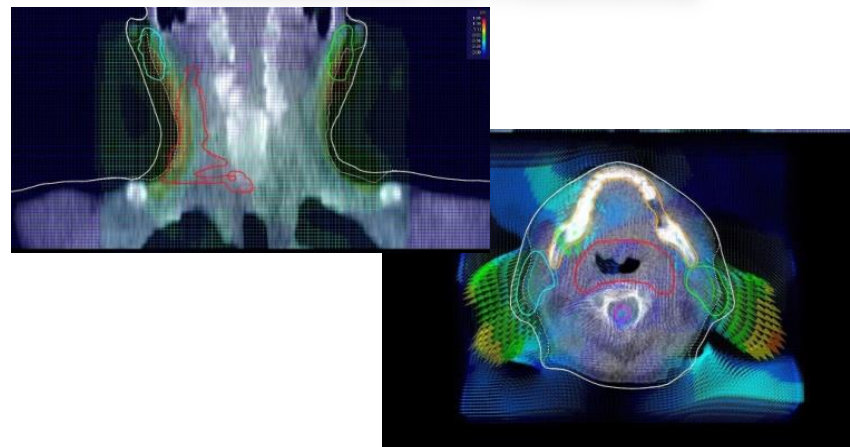
$$c = \sum_i a_i k(s_i, x) + b$$

The kernel function k was assumed to be linear, by considering a compromise between decision rules complexity and the generalized algorithm performance extended to the unanalyzed cases. A cross-correlation approach was used during the learning phase to improve statistical power of our sample and to ensure amore accurate clinical range identification around cluster centroids.

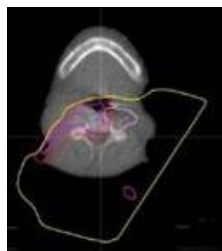
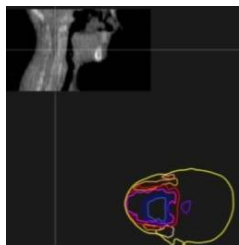
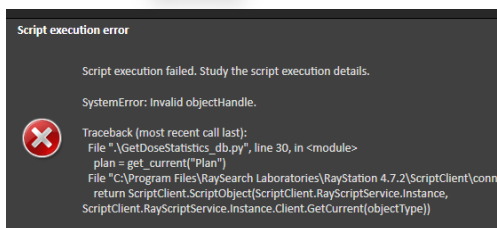
CORRECT TREATMENT



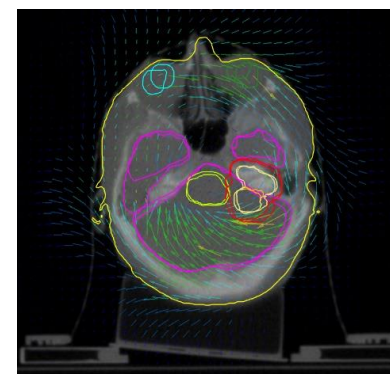
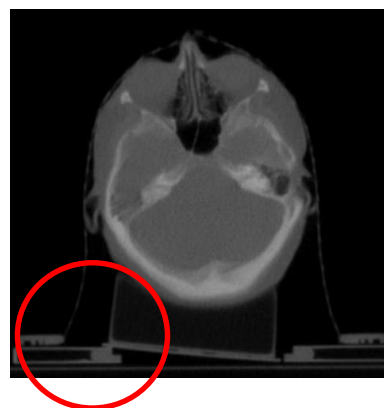
SUGGESTED RE-PLANNING



BIAS



WARNING



Classifier output cases

After the learning phase, normalized data of each new patient was initialized as input into the predictive tool.

From the V and D data postprocessor by the machine learning, it is identifiable the proximity of the daily treatments conditions with the cluster centroid; four solutions, representatives of the coherence with original planning, are classifiable:

Correct treatment (points are closer to the V/D initial values cluster): New patient has a weekly trend comparable (within a predefined threshold as detailed below) with V and D mean values obtained from training patients. re-planning is not needed because there are not significant discrepancies between the initial and the weekly status.

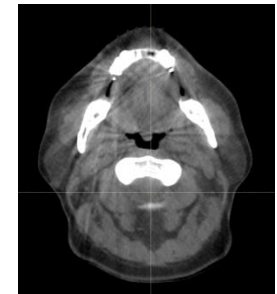
Suggested re-planning (points are closer to the V/D deformed values cluster): The patient is recommended for replanning. DIR shows morphological/dosimetric differences that remain undetected in the standard RT approach looking only at the planning CT.

Bias (points do not have clinically reasonable values): incorrect analysis attributable to a software bias (e.g. inappropriate daily image, limited FOV, uncorrected RIR and/or DIR).

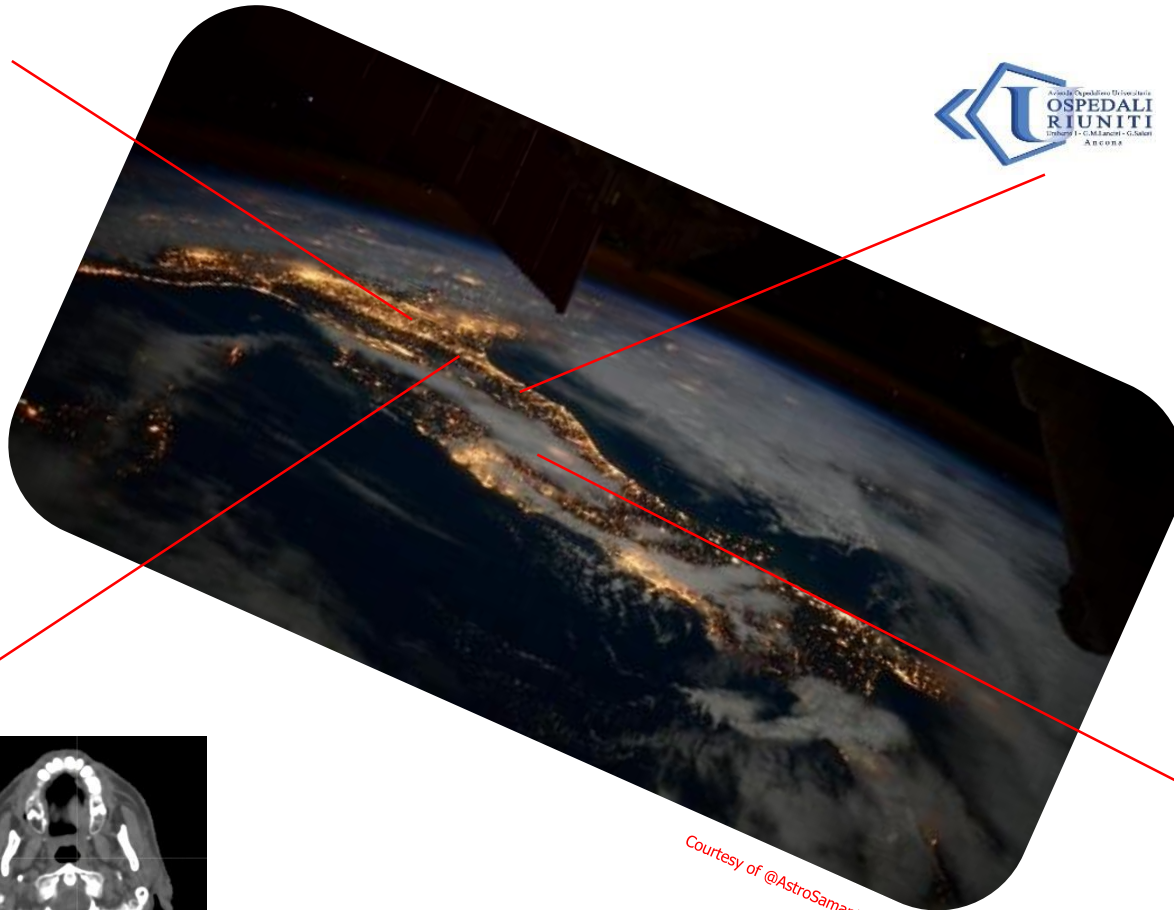
Warning (points are far from both clusters): The patient dataset must be investigated; abnormal variations may have happened during both the actual treatment and the ART simulation process such as improperly delivered dose due to a systematic setup error during treatment day(s) or improper data handling.



19 pts.



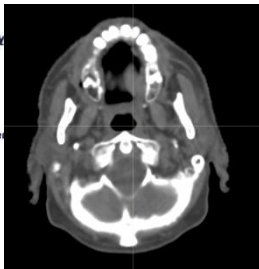
10 pts.



Courtesy of @AstroSamantha



8 pts.

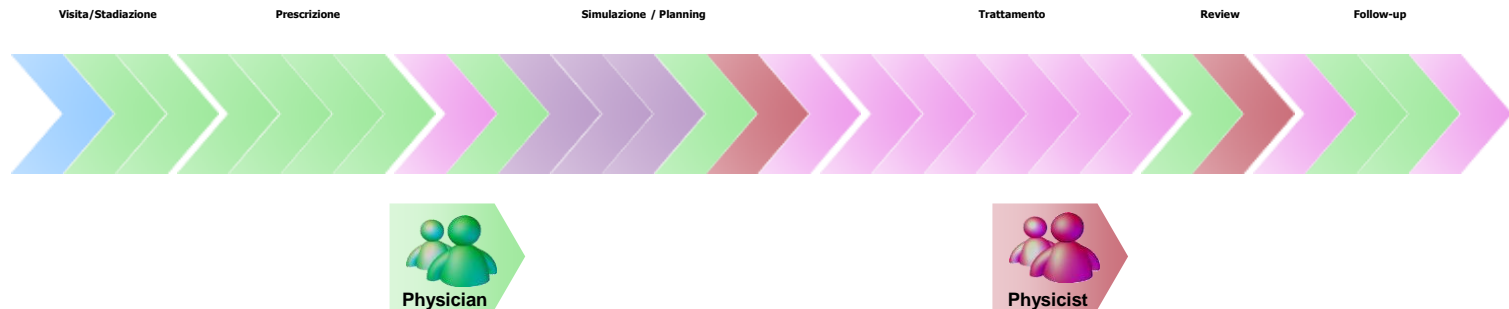


12 pts.



TOT. = 49 pts. \approx 1500 daily studies

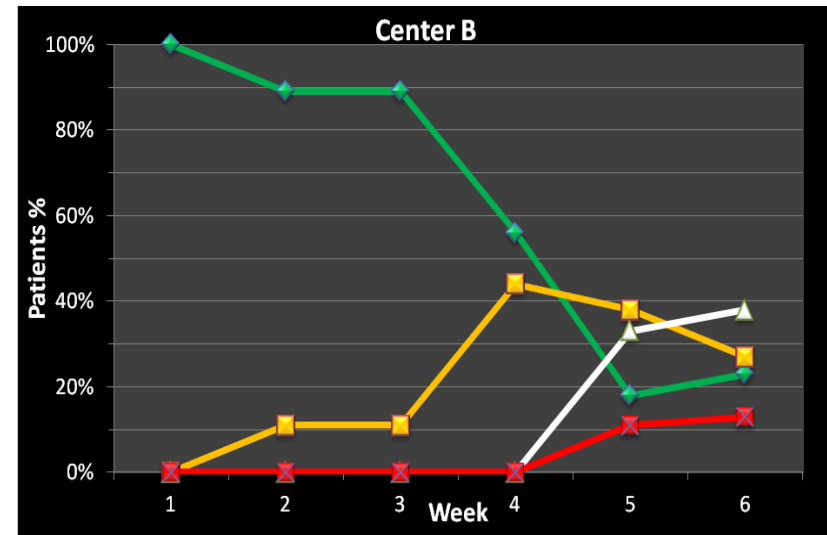
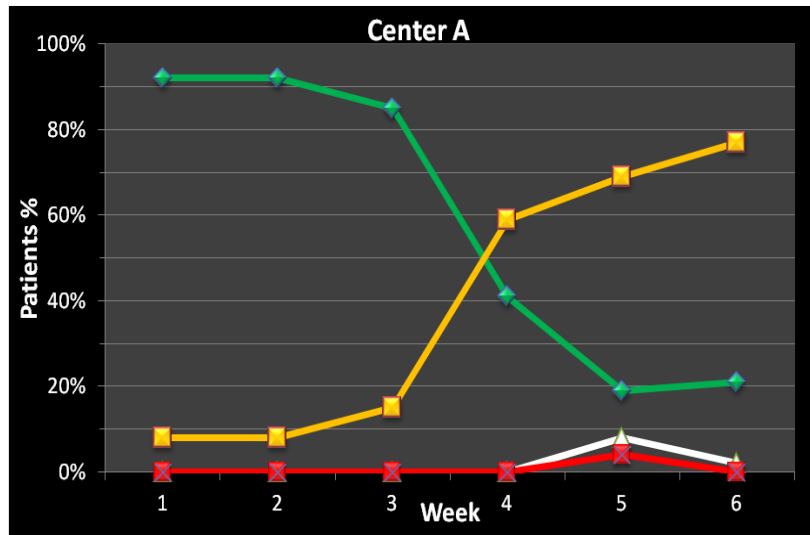
- **Validate** our model on different daily images (i.e. MVCT and CBCT)
- Increase **patients cohort** for analyzed pathologies
- Promote a *RO + Physics dpt.* **collaboration** to evaluate the automatic process of hybrid deformable algorithm



- Develop a national **Data-Mining**



SPECIFIC RESULTS

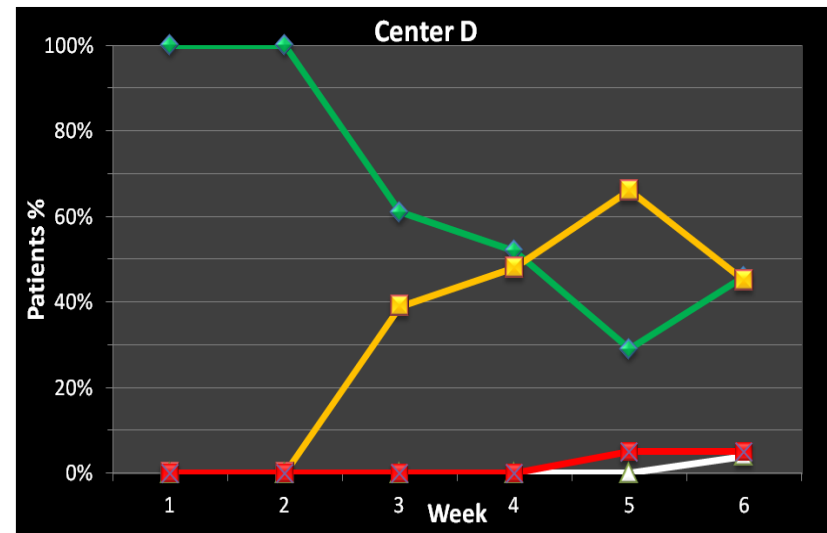
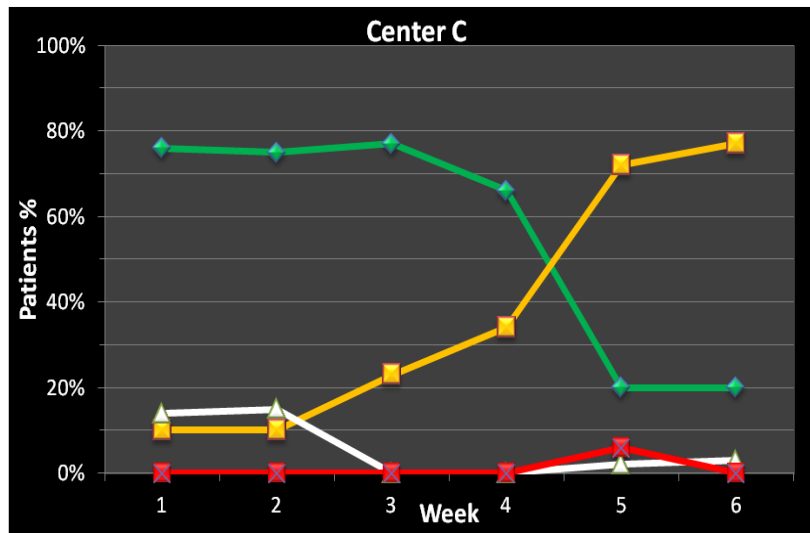


Correct treatment

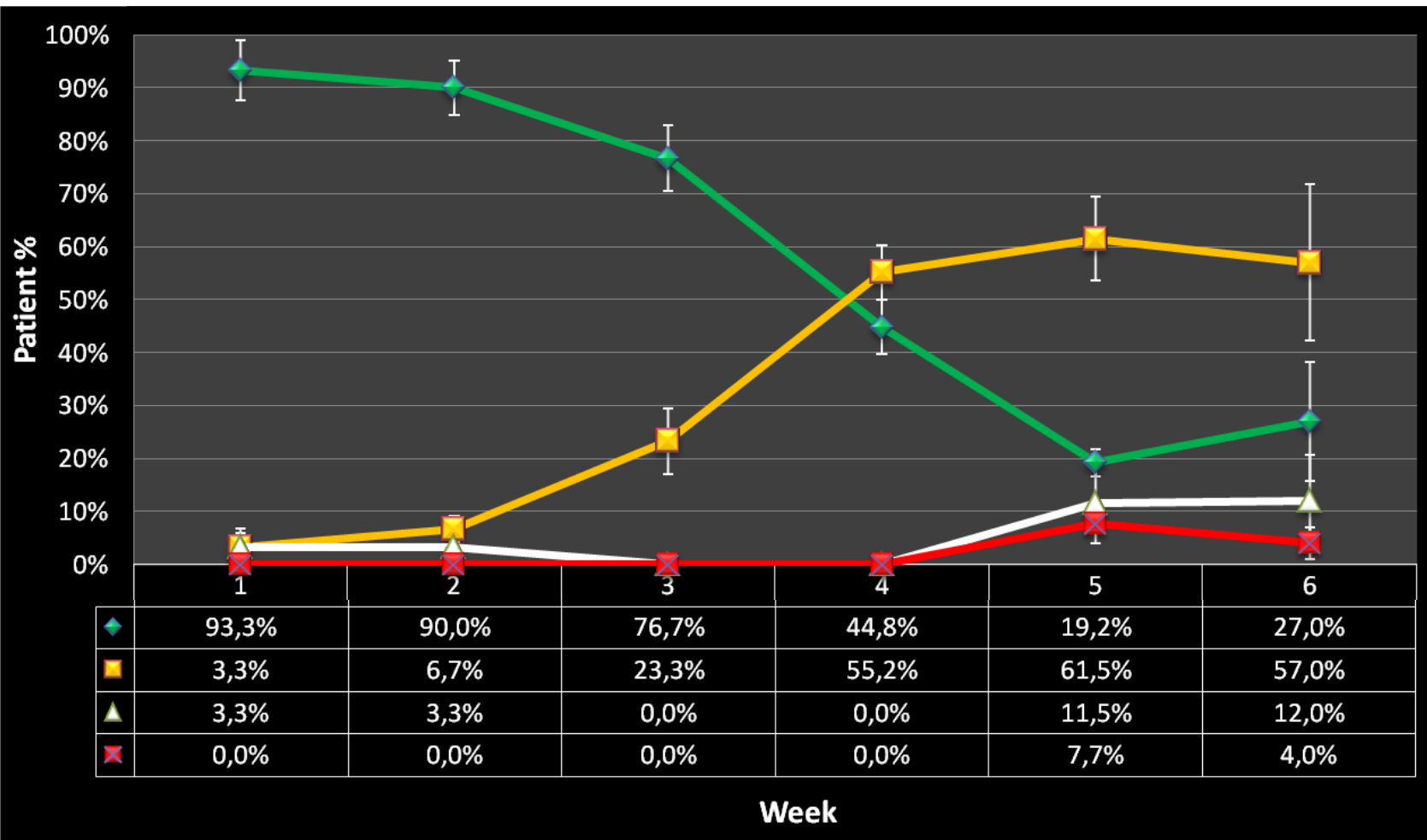
Suggested Re-planning

Bias

Warning



GLOBAL RESULTS

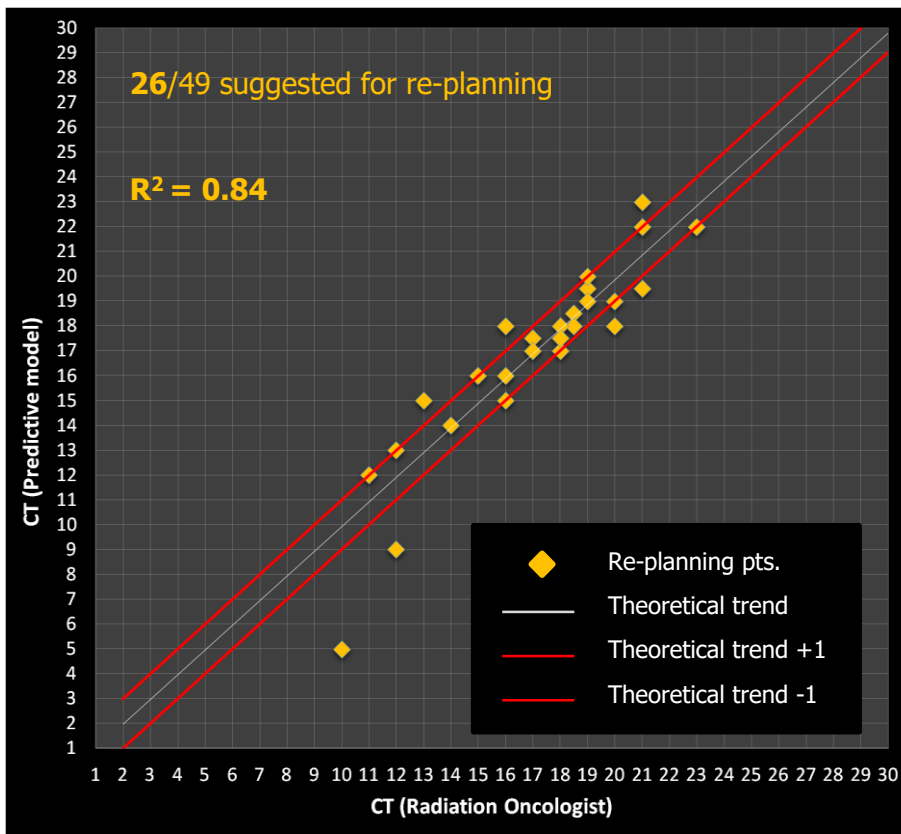


■ Correct treatment

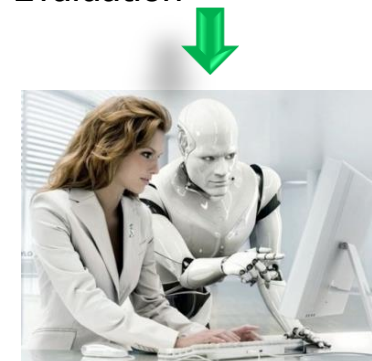
■ Suggested Re-planning

□ Bias

■ Warning



Hypothesis: **1** allowed Re-plan



Sensitivity = 23.1%
(Perfect matching with Theoretical trend)

Sensitivity = 73.3%
(Theoretical trend ± 1 day)

**Predictable
trend**

89.6% Center-A
92.7% Center-B
76.0% Center-C
87.0% Center-D

- Predictive approach can support ART in clinical routine
- Machine learning tool can identify patients and days for Re-planning
- The multicenter study has validated the model, highlighting a common trend for RT patients
- The forecasting algorithm is in line with the clinical end-points

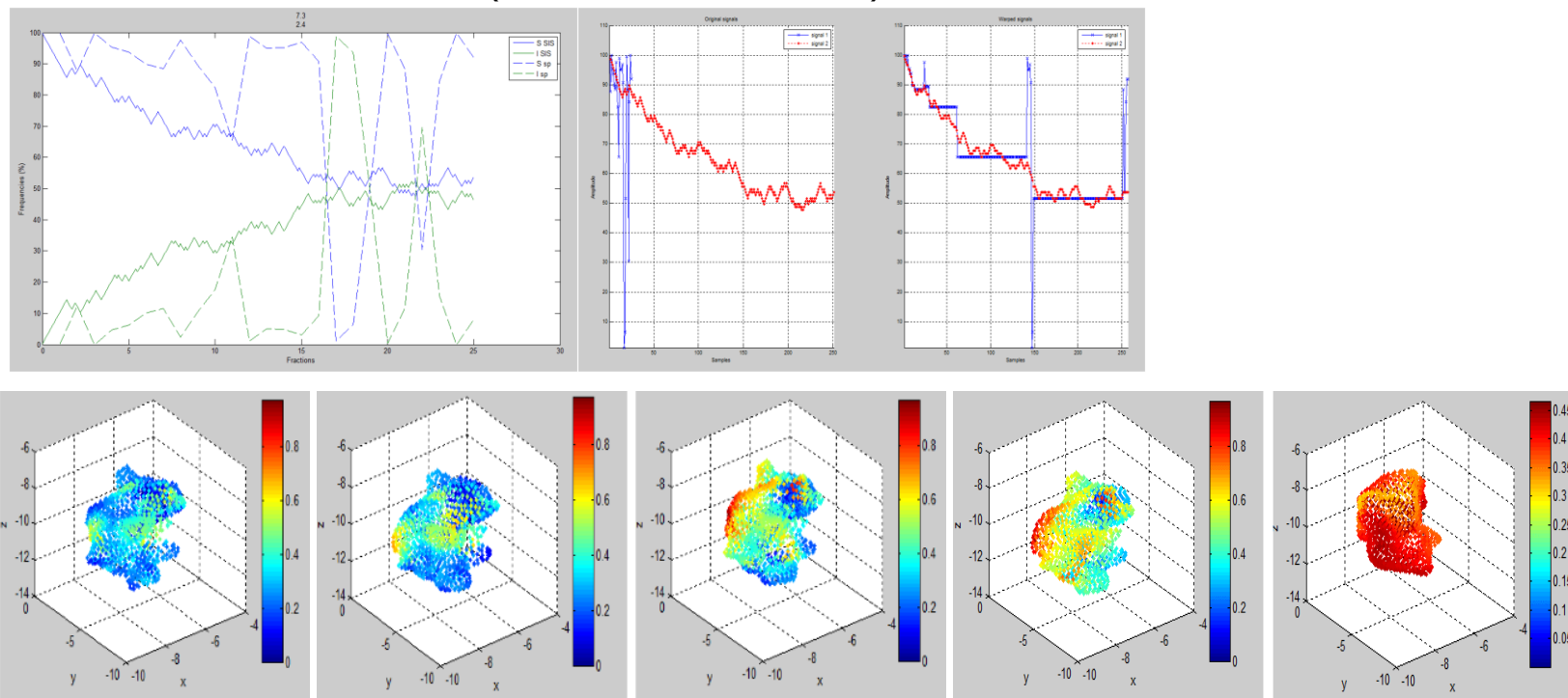


A common clinical ART workflow could be defined for national trials ...

Data Mining in developing ...

New collaborations
? ? ?

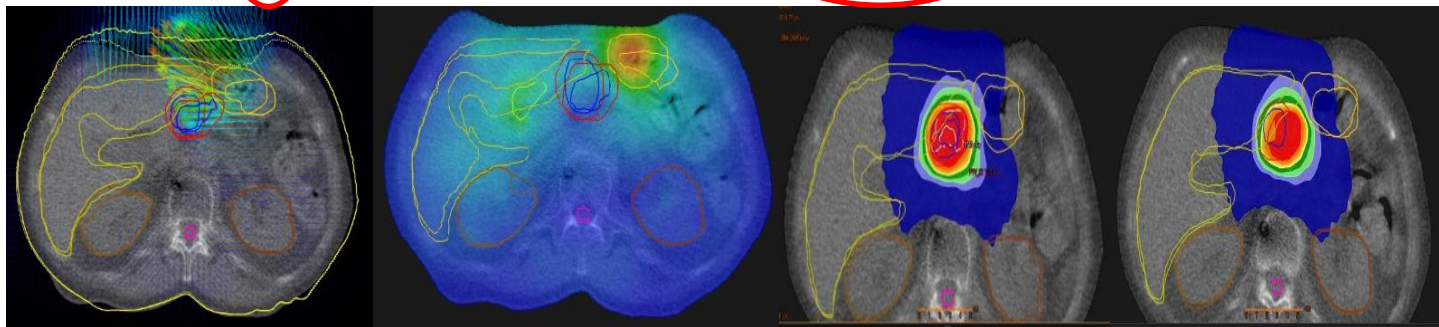
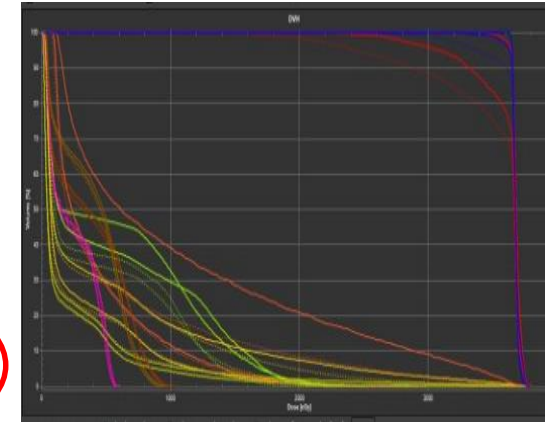
PREDICTIVE MODEL (EPIDEMIOLOGICAL MODEL) APPLIED TO DEFORMABLE IMAGE REGISTRATION



1. WE CAN PREDICT SINGLE VOXEL EVOLUTION USING DIR...
2. ORGAN VARIATION CAN DEPEND BY THE MOTION..
3. LOCAL TISSUE SHOULD BE INCLUDED IN DIR
4. MODEL SHOULD BE ROBUST
5.IT IS EMBRIONAL, BUT IT WORKS

IN REAL WORLD - PANCREAS

Pazienti	Volume GTV (cc)	Soglia (cm)	S ₀ (Susceptible)	I ₀ (Infected)	DTW	I ₀	α	β
1	26,16	1,2	75,6%	24,4%	0,16	0	1,65	1,55
2	17,07	1	87,7%	12,3%	1,44	0	2,30	0,05
3	16,75	0,8	97,2%	2,8%	8,09	0	2,85	0,05
4	3,56	0,7	15,1%	85,0%	5,58	0	1,50	1,90
5	43,61	1,4	75,5%	24,5%	1,30	0	1,50	0,30
6	31,95	0,8	87,4%	12,6%	1,89	0	1,55	0,25
7	28,32	0,8	100,0%	0,0%	7,56	0	2,05	0,05
Range (min-max)	[3,56 - 43,61]	[0,7 - 1,4]	[15,1 - 100]	[0 - 85]	[0,16 - 8,09]	[0 - 0]	[1,50 - 2,85]	[0,05 - 1,90]
Media \pm STD	23,9 \pm 12,8	1,0 \pm 0,3	76,9 \pm 28,9%	23,1 \pm 28,9%	3,72 \pm 3,28		1,91 \pm 0,51	0,59 \pm 0,79



WE NEED TO WORK, BUT SOME IS NOT COMPLYTLY CLEAR...

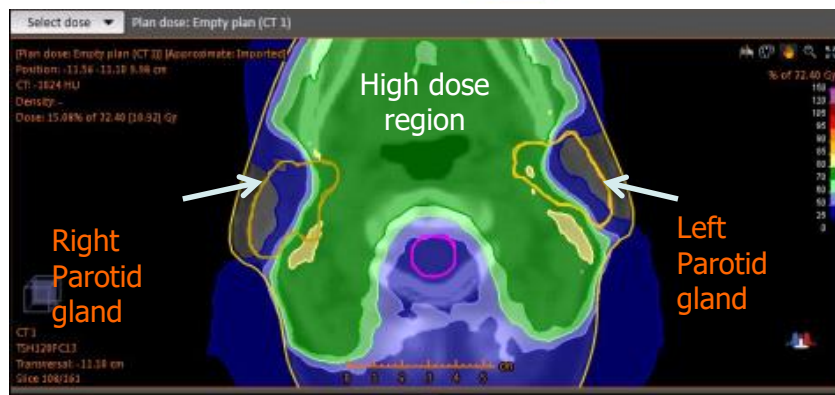
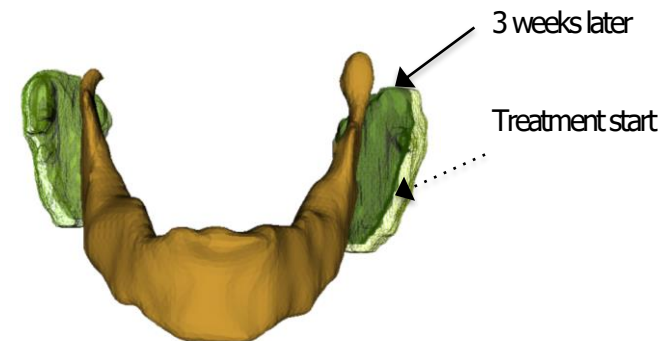
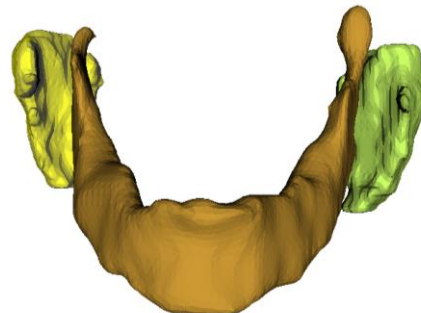
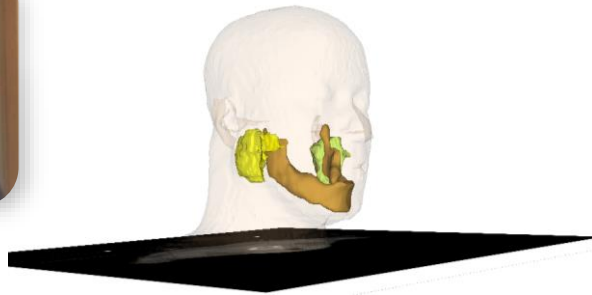
1. ANATOMICAL VARIATION ARE PREDICTED BY THE MODEL
2. ORGAN MOTION HAS HIGH IMPACT IN PRECTIVE DATA
3. INTERNAL ORGANS CONDITION (GAS, FILL, ECC..) CAN CHANGE «A LITTLE BIT» THE MODEL
4. ...BUT THIS IS THE WORST AREA WHERE YOU CAN WORK

THE HUMAN BODY IS A DINAMIC SISTEM

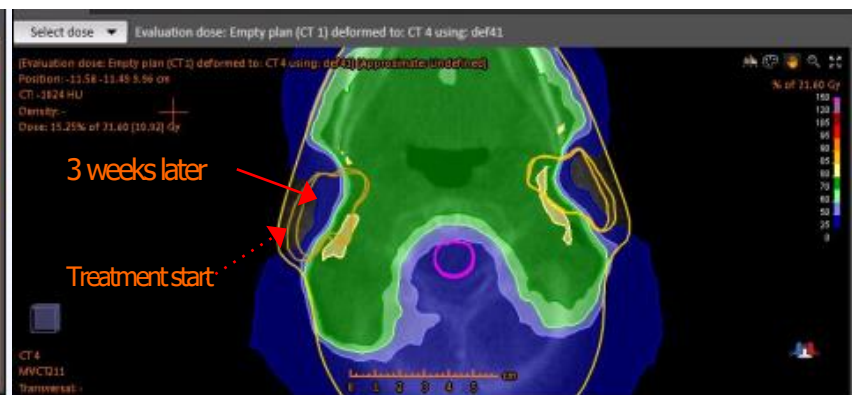
- ☐ Weight loss
- ☐ Tumour shrinkage
- ☐ Alterarion muscle mass



PAROTID GLANDS
INTER-FRACTION
DEFORMATION



Treatment start
 $D_{\text{mean}} = 25 \text{ Gy}$



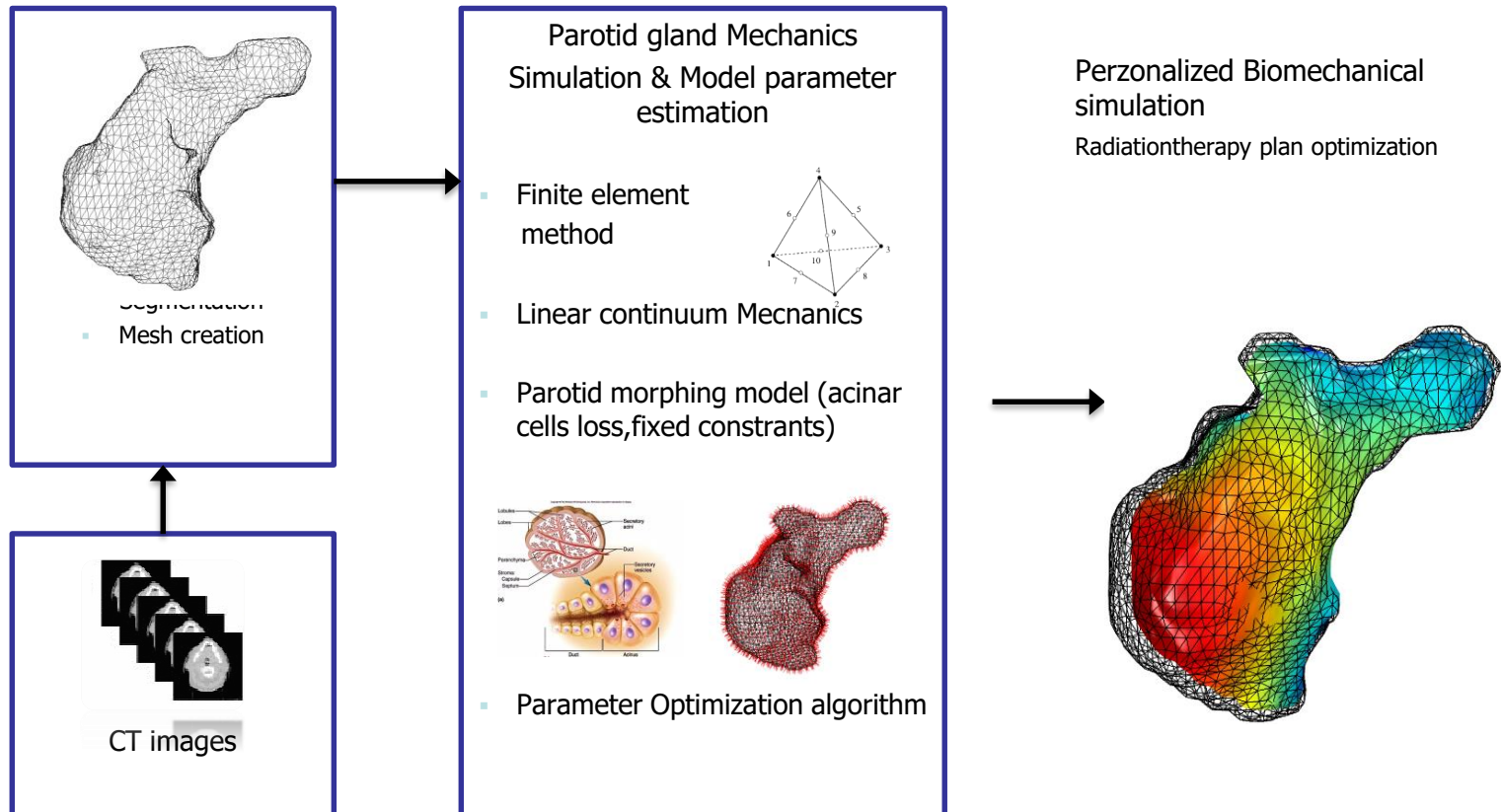
mid-time course (3 weeks later)
 $D_{\text{mean}} = 27 \text{ Gy}$

G.Guidi, N.Maffei, F.Itta



BIOMECHANICAL MODEL IMPLEMENTATION

- PHASE 1 : Image Acquisition of 8 H&N Patients
- PHASE 2: 3D Geometrical model creation from segmented structures
- PHASE 3 : Biomechanical model creation via Finite Element Method (FEM) software



PHASE 3 : BIOMECHANICAL MODEL IMPLEMENTATION

GEOMETRY

MATERIAL

- ☐ Linear elastic
- ☐ Isotropic
- ☐ Homogeneous
- ☐ Navier lamè equation
- ☐ Young's Modulus = ~ 10 kPa
- ☐ Poisson ratio = ~ 0.49
- ☐ Density = 1 (g/cm³)

PHYSICS

Load condition based on :

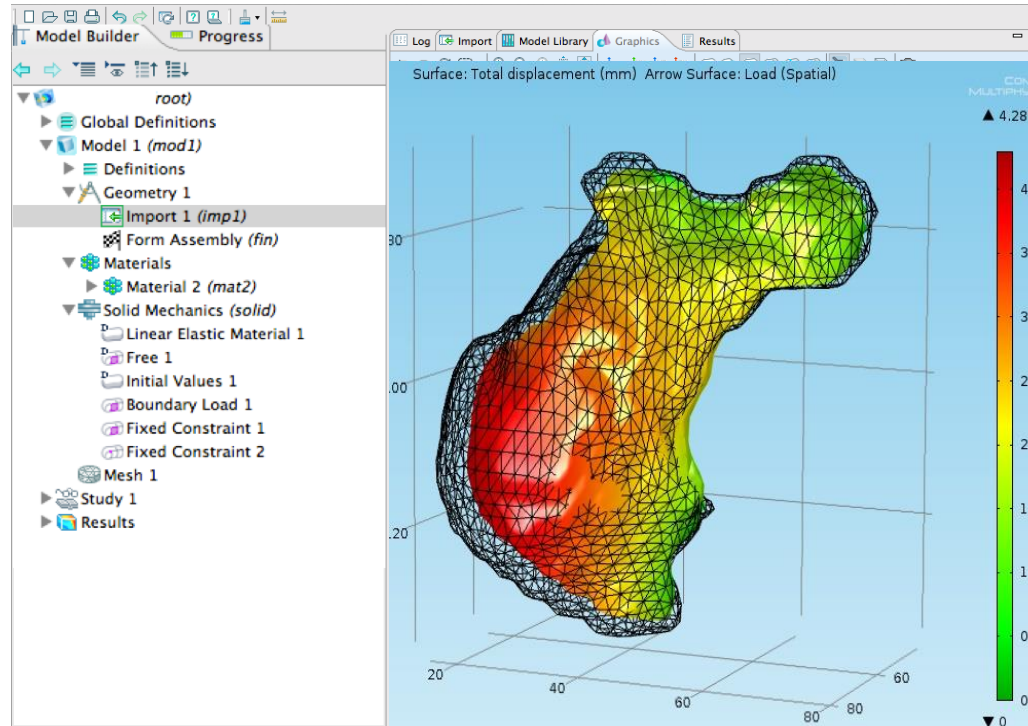
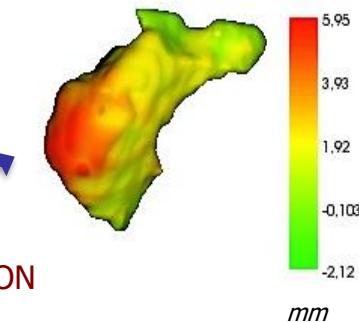
- ☐ Loss of acinar cells
- ☐ Swelling parotid lobuli

Boundary condition :

- ☐ Motion block carried out
by surrounding structures

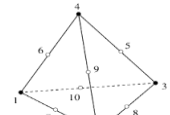


REAL
DEFORMATION



DOMAIN DISCRETIZATION

- ☐ Volumetric mesh creation
- ☐ Domain discretization with
250000 tetrahedral elements



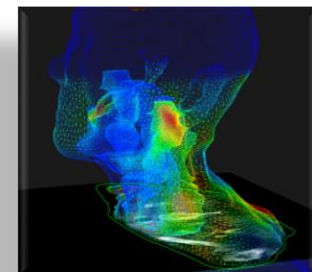
RUN STUDY

- ☐ Run simulation for different
load condition/Young modulus
values to find optimal model
parameter



$$RMSDE = \sqrt{\frac{1}{n} \sum (x_{sim} - x_{real})^2 + \frac{1}{n} \sum (y_{sim} - y_{real})^2 + \frac{1}{n} \sum (z_{sim} - z_{real})^2} \cdot \sqrt{\frac{1}{3}}$$

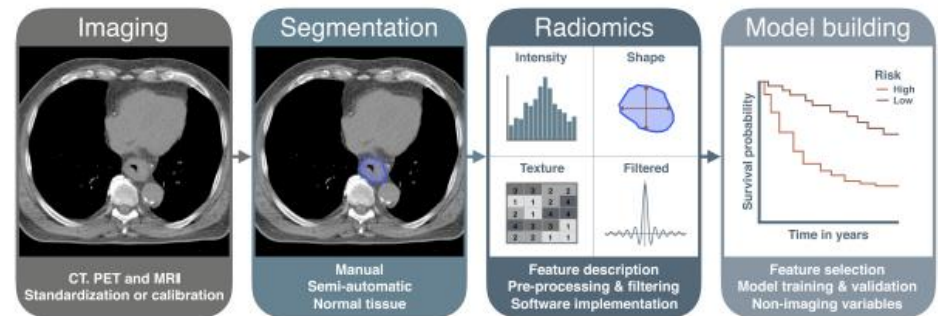
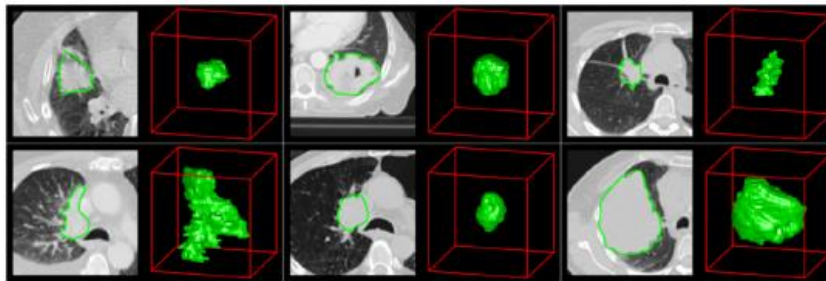
OPTIMAL MODEL PARAMETER ESTIMATION
COMPARING REAL AND SIMULATED
DEFORMATIONS



G.Guidi, N.Maffei, F.Itta

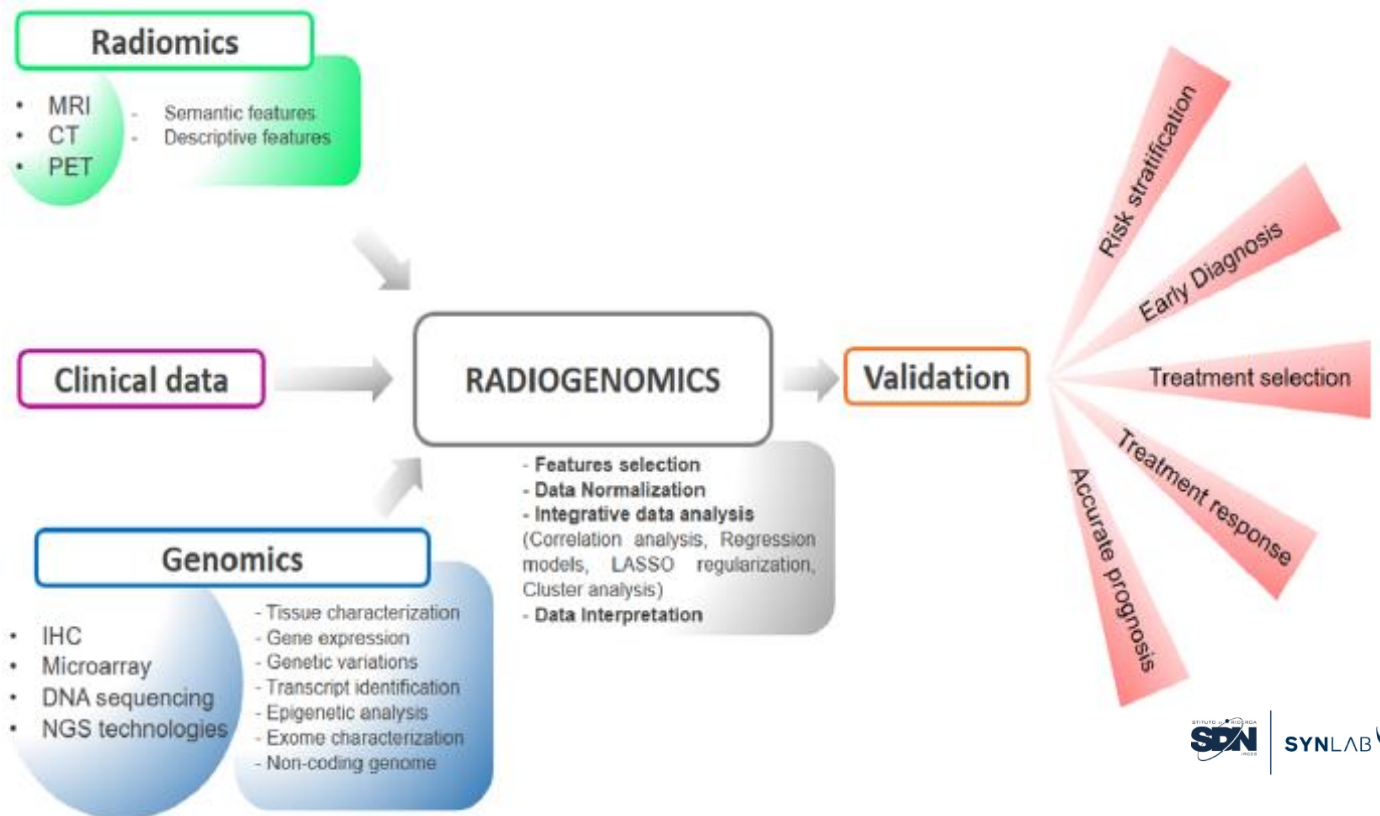
The «...omics» hera!

Images are not pictures, they are data



Larue, et al., Br J Radiol 2017

Gillies et al., Radiology 2016;278(2).



The study of genetic variation associated with response to radiation (Radiation Genomics)

The correlation between cancer imaging features and gene expression (Imaging Genomics)

RADIOGENOMICS

Radiogenomics


Technological improvements in the field of imaging and molecular biology have led to the “Radiogenomics” or “Imaging Genomics”. Literally, Radiogenomics refers to the **analytical processes aimed to correlate cancer imaging features (Radiomics) with Genomic data.**


Radiation genomics

Genetic variation, such as single nucleotide polymorphisms, is studied in relation to a cancer patient's risk of developing toxicity following radiation therapy. It is also used in the context of studying the genomics of tumor response to radiation therapy

Imaging genomics

In imaging genomics, radiogenomics can be used to create imaging biomarkers that can identify the genomics of a disease, especially cancer without the use of a biopsy.

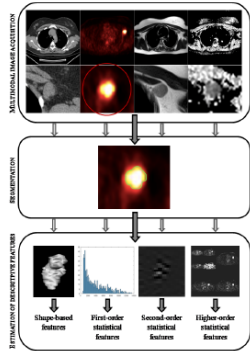





Radiogenomics


Radiomics refers to the comprehensive quantification of tumor phenotypes by the extraction of a large amount of quantitative features from medical images. This high-throughput extraction of quantitative imaging features is the result of a workflow composed of three main steps

- Acquiring the images
- Segmenting the regions of interest (ROIs)
- Estimating descriptive features



A great advantage of radiomic analyses is their feasibility with conventional clinical images (PET, CT, MRI). The first step of radiomic pipeline, in fact, involves the acquisition of images that are typically part of diagnostic or treatment planning protocols for oncological patients

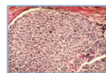




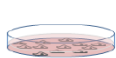
Radiogenomics

Genomics is the study of the entirety of an organism's genes actually performed by the combination of high-throughput molecular biology technologies with complex computing and math techniques (bioinformatic analysis). Generally, two technologies are critical for genomics analysis: 1) **microarray**; and 2) **next generation sequencing (NGS)**


BIOLOGICAL SAMPLES



Tumor Tissues



Cancer Cell Cultures



Nucleic acids (DNA and RNA)

↓

MeDIP Seq
ChIP Seq

↓

DNA/Chromosome isolation

↓ Sorption

Immunoprecipitation

↓ Sample preparation

Sequencing

↓

Epigenetic analyses

Microarray

↓

RNA isolation

Reverse Transcriptase labeling

cDNA + Fluorescent probes

↓

Microarray hybridization

RNASeq

↓

RNA isolation

Generate cDNA, fragmentation, size selection and adaptor addition

↓ Sequencing

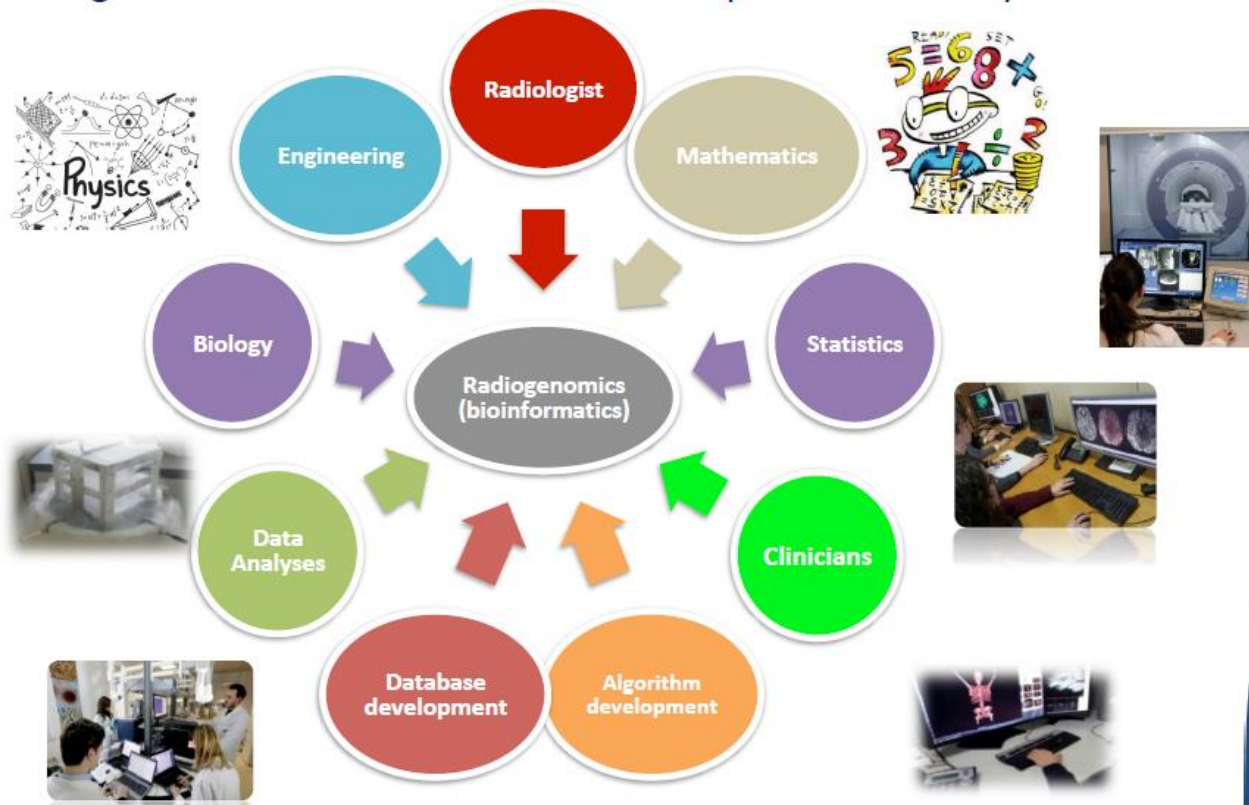
Quality control and mapping

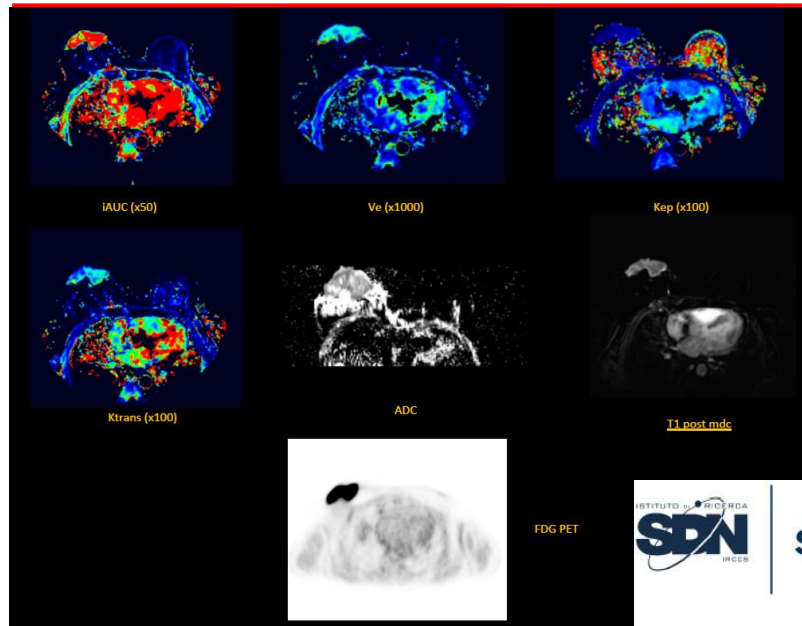
↓

Data processing and analysis



Once both radiomic and genomic features are extracted, radiogenomic analysis will be performed. Radiogenomic approaches are extensively based on numerical calculus and computer science methods, allowing the management and analysis of a huge number of variables for each sample and modality.





SYNLAB

Overall molecular landscape of the ERF score

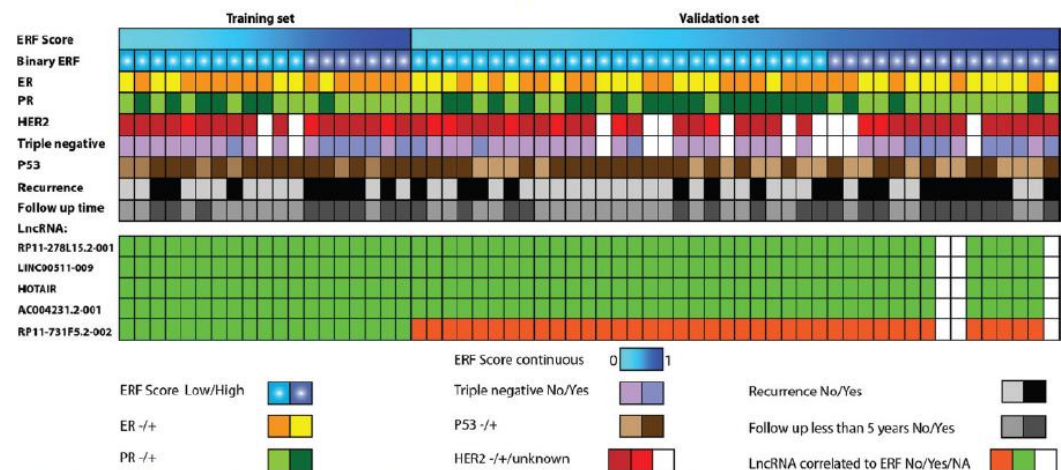
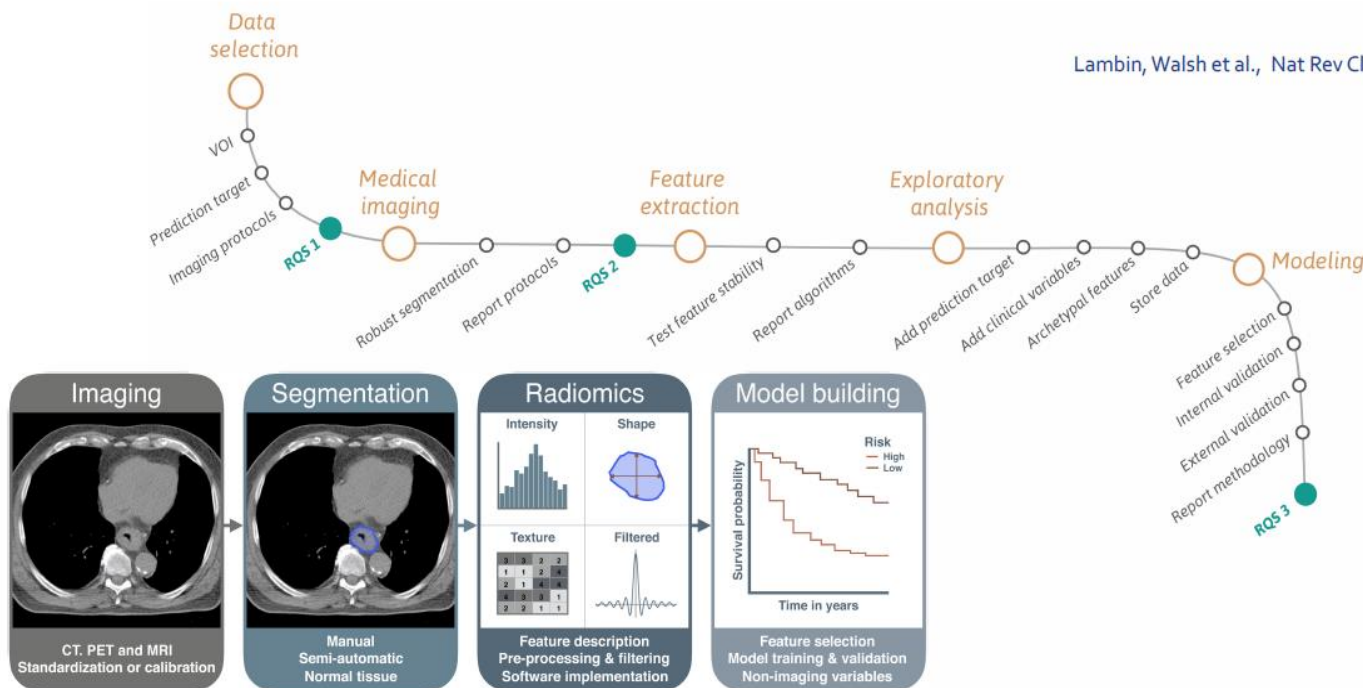


Figure 4: Graph shows molecular characteristics of the ERF score phenotype in training ($n = 19$) and validation ($n = 42$) sets. Continuous ERF scores for each patient are listed from low to high in the respective data sets. Status of estrogen receptor, progesterone receptor, epidermal growth factor receptor 2, triple negative receptor, tumor protein 53, and lncRNA expression are provided as labeled. Recurrence and follow-up data are also included.

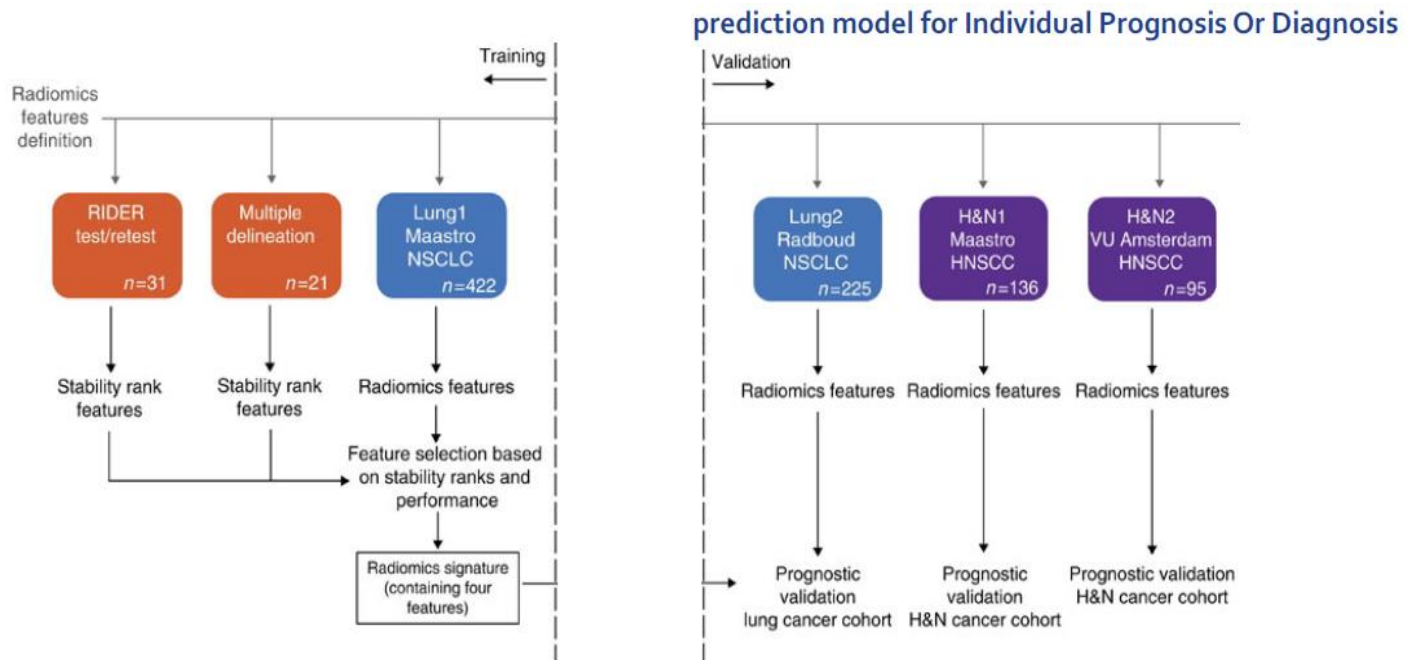
Radiomics Workflow

Lambin, Walsh et al., Nat Rev Clin Oncol (in-press)



Larue, et al., Br J Radiol 2017

So, Radiomics needs a lot of training data....



Aerts et al., Nature Communications 5, 4006

"That's too much!!!"

(Praha 2009 : Tomotherapy Meeting)



Un maestro in pensione...