**Appendix**

Methodology of the machine learning instruction

The methodology implements the following steps:

1. Transform the categorical variables (features #2 and #4) in dummy variables. Append feature #1 and #2 to the dummy variables. Call this Input Space
2. Round the values of the target variable to the closest integer. Call this Target Variable
3. Extract the list of unique values of the Target Variable. Call this set Target Set = [40, 42, ..., 99]
4. Create a set of machine learning classifiers:
	1. Make a temporary copy of the dataset that is formed by the tuples (Input Space, Target Variable)
	2. Iterate over the values in the Target Set. Say θ the current value (e.g., θ = 45)
	3. Threshold the Target Variable by comparing it with θ: 0 if Target Variable < θ, 1 otherwise. This casts the problem into a 2-class classification.
	4. Split the dataset into two subsets by randomly shuffling the rows. The former set contains the 70% of rows and it is called training set, the latter dataset contains the remaining 30% and it is called Test Set
		1. Use the training set for training a Machine Learning Classifier with a k-fold cross validation approach for avoiding over-fitting. Compute training accuracies
		2. Use the test set for evaluating the performances of the Classifier.
5. Given a new unseen patient, apply the trained classifiers for predicting the probability of nodule’s reduction percentage.

### Dummy variable

Machine learning algorithm must be trained on numerical features, i.e. quantity that can be compared with some operators (<,>). For this purpose, we used the one-hot encoding process for transforming categorical variables into number, i.e. dummy variables. A dummy variable is an artificial variable created to represent an attribute with two or more distinct categories/levels. The dummy variable represents the original value as a tuple of binary values. For example, if the original categorical variable has n values it will be represented by n-1 new binary variables [16].

### Automatic Machine Learning - TPOT

Automatic Machine Learning (AutoML) systems aim to assist data scientist on the selection of the proper algorithms and their automatic tuning. In this work, we have challenged the Tree-based Pipeline Optimization Tool (TPOT) AutoML framework.TPOT automatically designs and optimizes machine learning pipelines for a given problem domain, without any need for human intervention. In brief, TPOT optimizes machine-learning pipelines using a version of genetic programming (GP), a well-known evolutionary computation technique for automatically constructing computer. TPOT can deal with both regression and classification problems. In the following, we list the main algorithms handled by TPOT and that are useful for our scope:

• Supervised Classification Operators. Decision Tree, Random Forest, eXtreme Gradient Boosting Classifier, Logistic Regression, and K-Nearest Neighbor Classifier.

• Feature Preprocessing Operators. StandardScaler, RobustScaler, MinMaxScaler, MaxAbsScaler, RandomizedPCA, Binarizer, and Polynomial Features

• Feature Selection Operators. VarianceThreshold, SelectKBest, SelectPercentile, Select and Recursive Feature Elimination (RFE) [17].

## Accuracy metric

Classifiers accuracy is computed with the F1 metric. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The formula for the F1 score is [18]:

$F1=2\* \frac{precision\*recall}{precision+recall}$

Where

$Precision= \frac{True Positive}{True Prediced Positive}$

$$Recall= \frac{True Positive}{Total Actual Positive}$$

### K-fold cross validation

Cross-validation is a common technique used to evaluate machine-learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into [17]. The general procedure is as follows:

* Shuffle the dataset randomly.
* Split the dataset into k groups
* For each unique group:
	+ Take the group as a hold out or test data set
	+ Take the remaining groups as a training data set
	+ Fit a model on the training set and evaluate it on the test set
	+ Retain the evaluation score and discard the model
	+ Summarize the skill of the model using the sample of model evaluation scores