Prediction of Immunotherapy response in NSCLC based on RW evidence and radiomic features

Author: Mariagrazia Monteleone 952133

Advisor: Prof. Elena De Momi

Co-advisor: Dott.ssa Arsela Prelaj M.D., Dott. Aldo Marzullo, Prof. Alessandra Laura Giulia Pedrocchi, Prof. Giuseppe Baselli

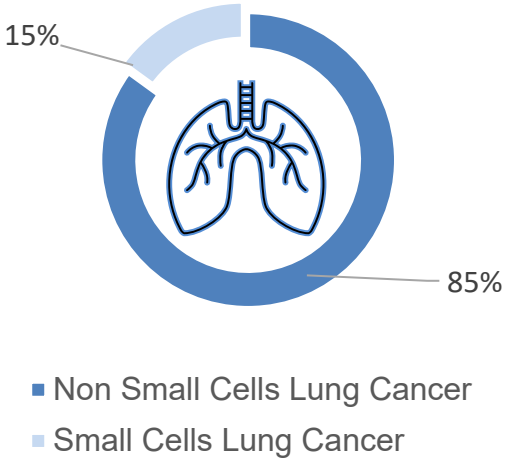
Academic Year: 2021-22

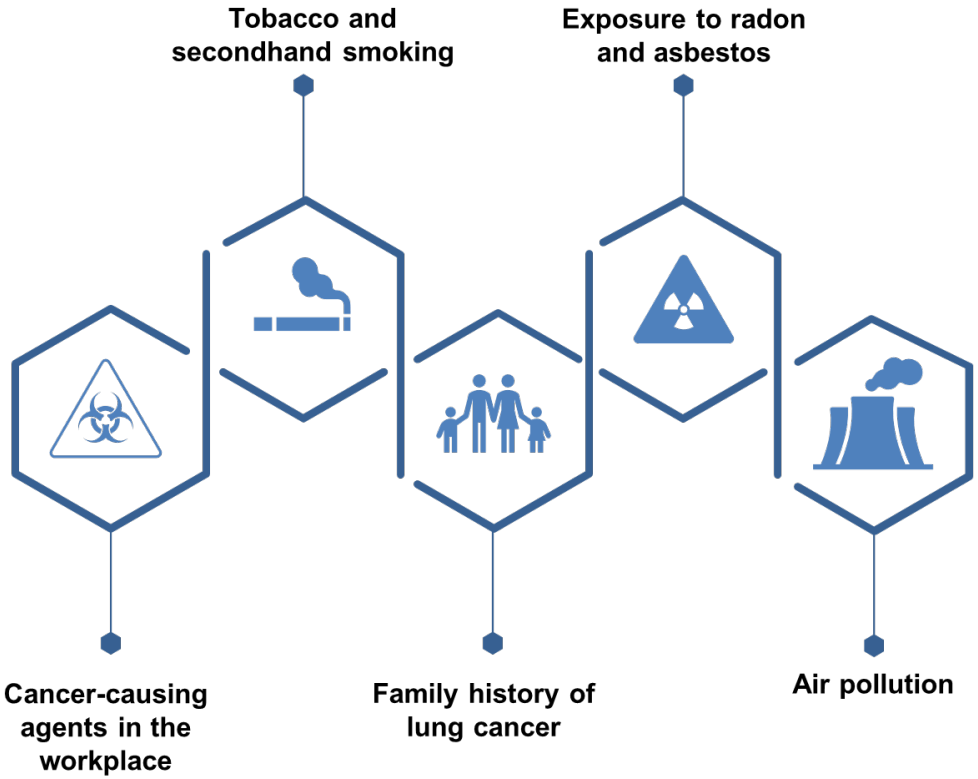
Clinical Background: Lung Cancer

Lung cancer, with the most frequent histology of Non-Small Cells Lung Cancer (NSCLC), is still on top of the rankings as the leading cause of tumor-related deaths (H. Sung et al., 2021).

Only in 2020:

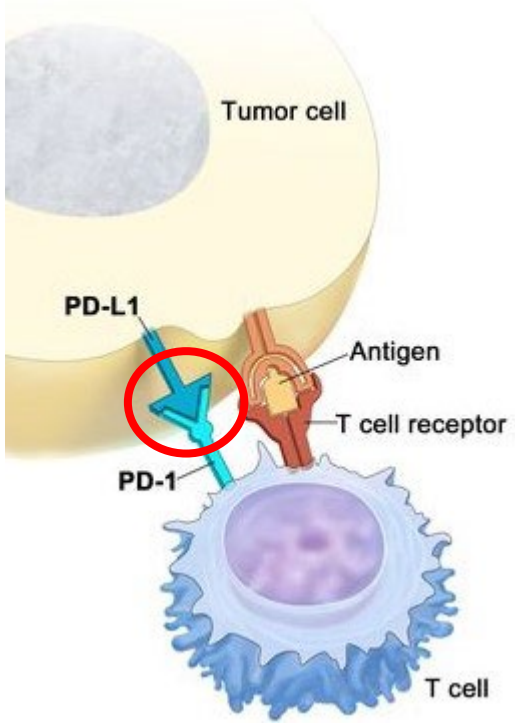
- 2.2 million new estimated cases
- 1.8 million deaths

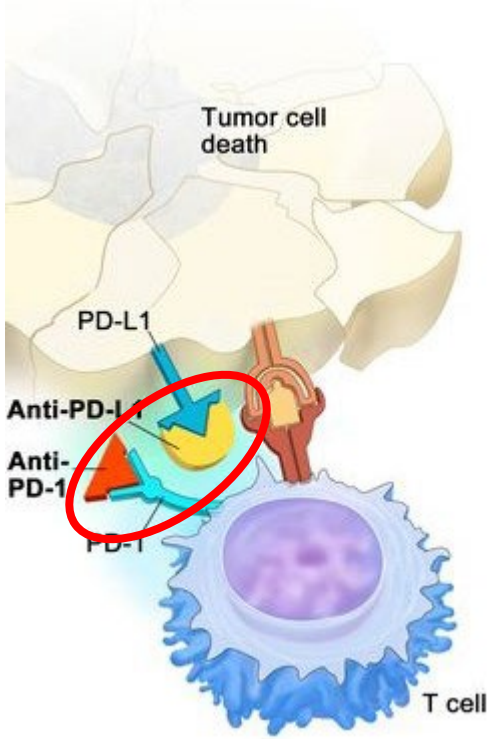




Clinical Background: Immunotherapy

Immunotherapy (IO) completely changed the NSCLC landscape reaching a median survival for a median OS of 24 months compared to the 12 months obtained with chemotherapy (CT) alone (H. Borghaei et al., 2019).



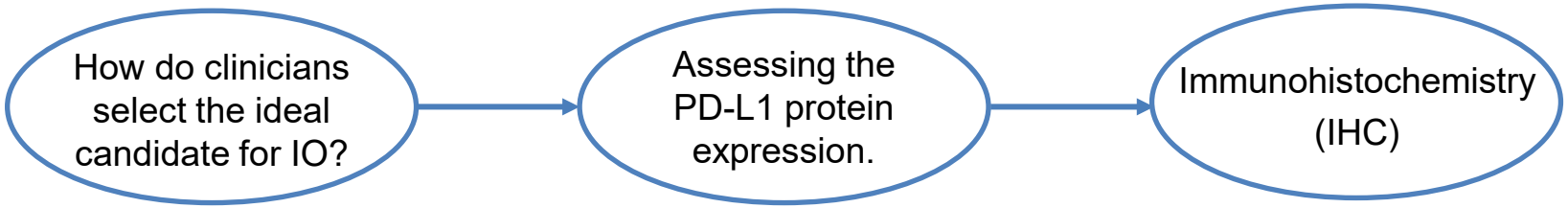


Some Patients Who respond to IO continue with the Duration of Response over 5-years:

This Never happened before

Clinical Background: Biomarker State of the Art

Only a small percentage of patients benefits from IO making the selection of the ideal candidate very challenging.



- Tumour Mutational Burden (TMB) (KEYNOTE 158 , A. Marabelle et al., 2020)
- Tumour Infiltration Lymphocytes (TILs) (D.Q. Zeng et al.,2016)
- Neutrophil to Lymphocyte Ratio (NLR) (S. Diem et al.,2017)
- Tumour Microenvironment (TME) (Huang et al., 2020)

After 10 years of basic & translational researches, PD-L1 remains the only confirmed biomarker approved for clinical practice for NSCLC patients.

Aim of the Thesis

An alternative solution might be to use a Machine Learning (ML) approach to exploit different type of information to predict the treatment outcome.

Design a ML algorithm to predict response to IO → Four binary classificators: Response and Survival Outcomes → Three feature sets

- Real World Evidence set (RW)
- Radiomics set (RD)
- Combination Set (RW + RD)

Real World Evidence and Radiomic Data

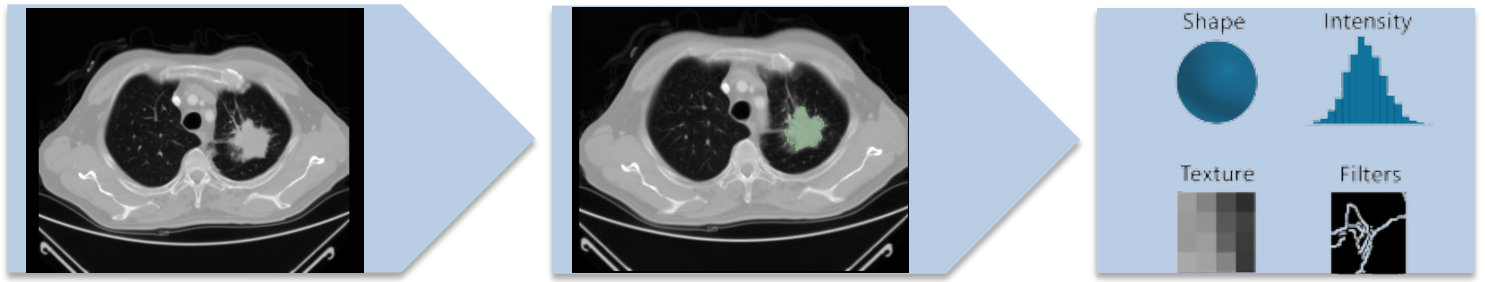
Thanks to the implementation of the Electronic Health Record (EHR), a large set of patient information are available in clinical institution databases.

Fondazione IRCCS Istituto Nazionale dei Tumori

Sistema Socio Sanitario Regione Lombardia

RW evidence data

Radiomics is a quantitative approach to medical imaging (J. E. v. Timmeren et al.,2020) .



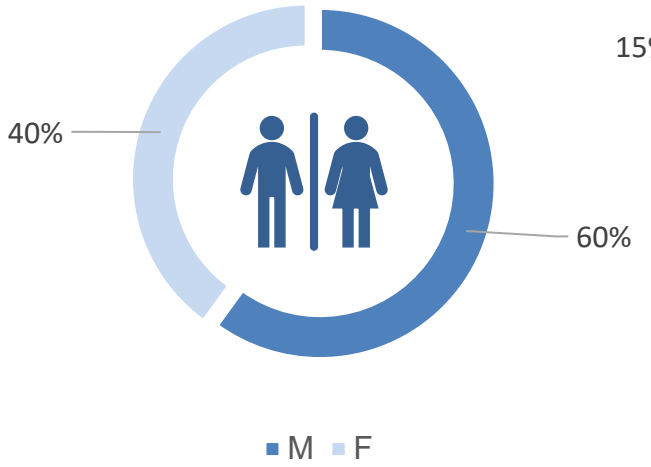
Radiomic data

Study Population

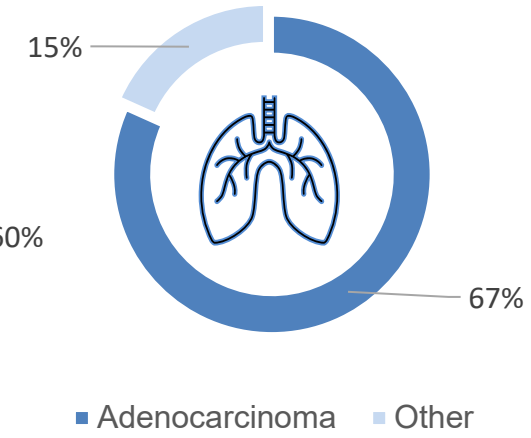
Sistema Socio Sanitario
Fondazione IRCCS Istituto Nazionale dei Tumori
Regione Lombardia

The study was conducted on 184 patients with IO, alone or in combination with chemotherapy, at the National Cancer Institute of Milan between 2015 and 2020 under the APOLLO study.

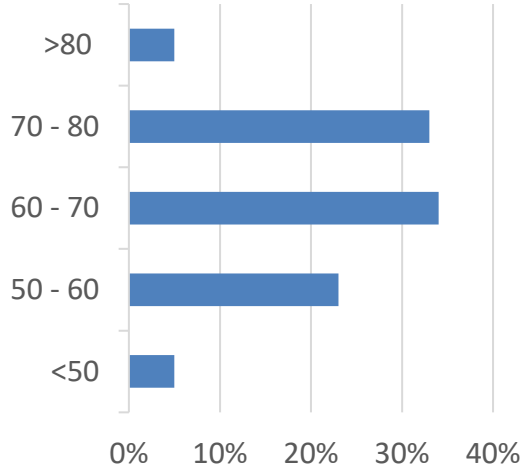
Gender



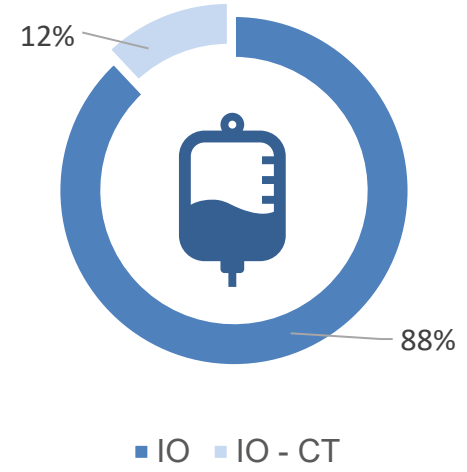
Histology



Age



Treatment



Project Workflow

Examination

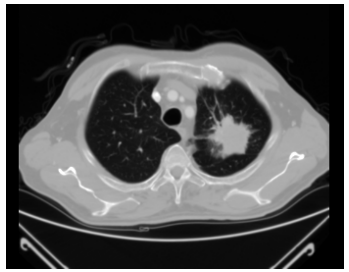


Image Segmentation

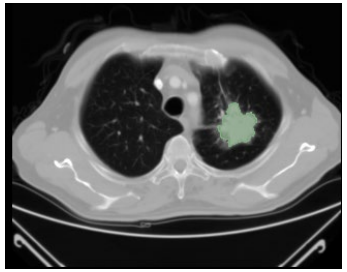
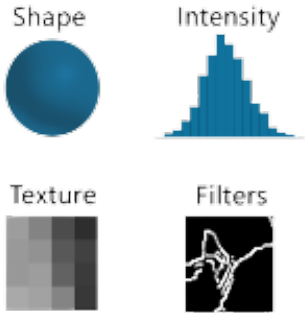


Image Processing

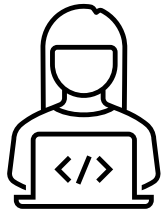
pyradiomics
python
+
RADIOMICS

Radiomic features



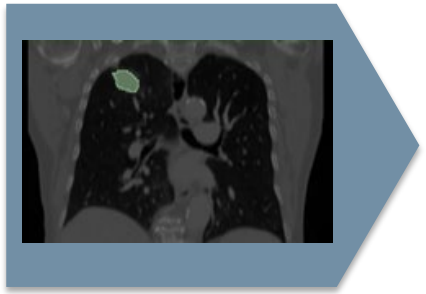
Real World features

Predictive models



Materials and Methods: Image and Features Processing

RD features



Grey Level Normalization

Spatial Resampling

Feature Extraction

RW features

Uniform Encoding

Iterative Imputer algorithm to fill missing values

Standardization

Materials and Methods: Features selection

RW and RD features

Filtering of highly correlated variables → Recursive Feature Elimination (RFE) → Removal of extremely noisy features

25 Real Evidence features (RW)
5 Radiomics features (RD)

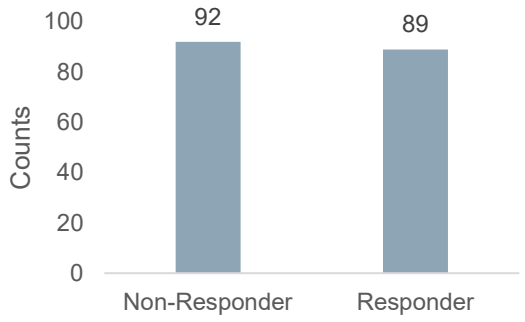
Materials and Methods: Classification Outcomes

Disease Control Rate (DCR)

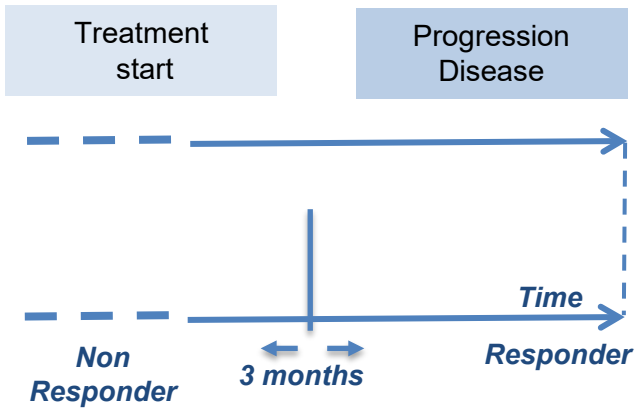
Percentage of patients whose disease shrinks or remains stable over a certain time period according to the geometric criteria illustrated by the RECIST guidelines (S. Litière et al.,2017)

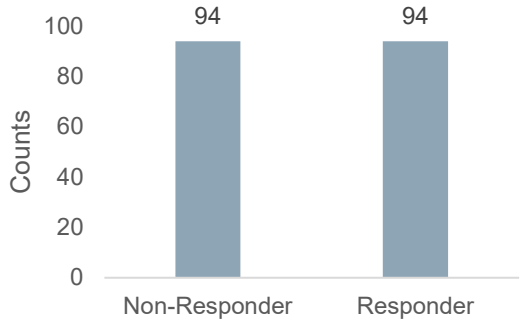
Responder

Non-Responder

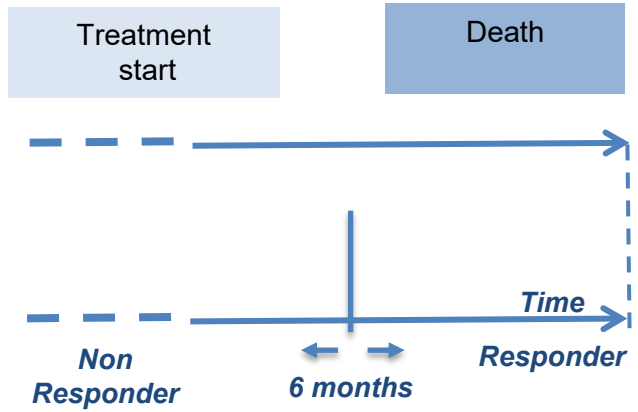


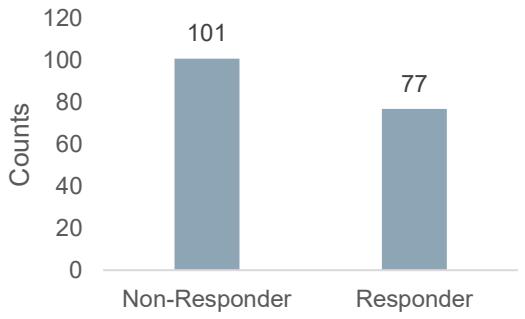
Progression-Free Survival (PFS)





Overall Survival (OS)





Materials and Methods: ML models

01 **K-Nearest Neighbors (KNN)**

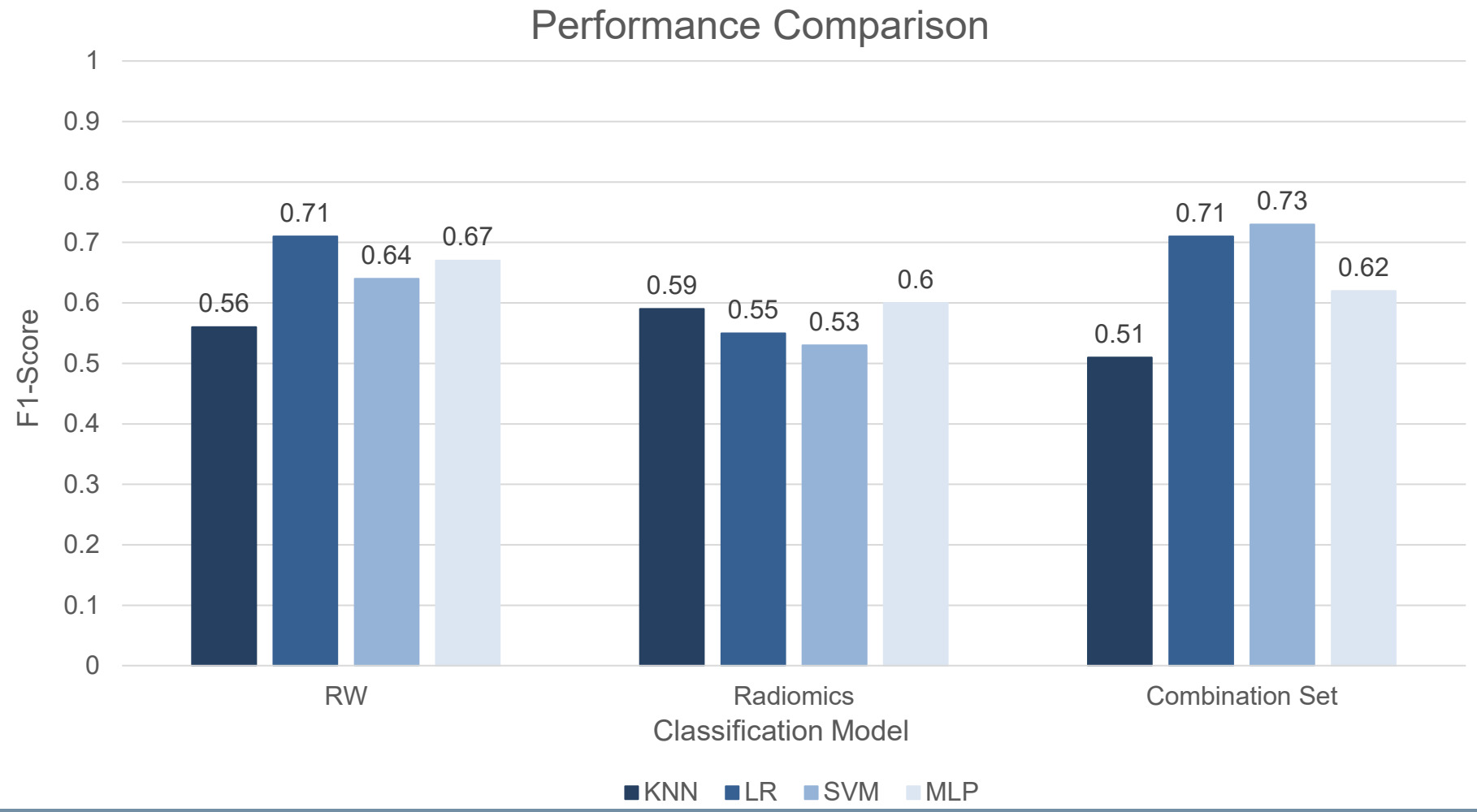
02 **Logistic Regression (LR)**

03 **Support Vector Machine (SVM)**

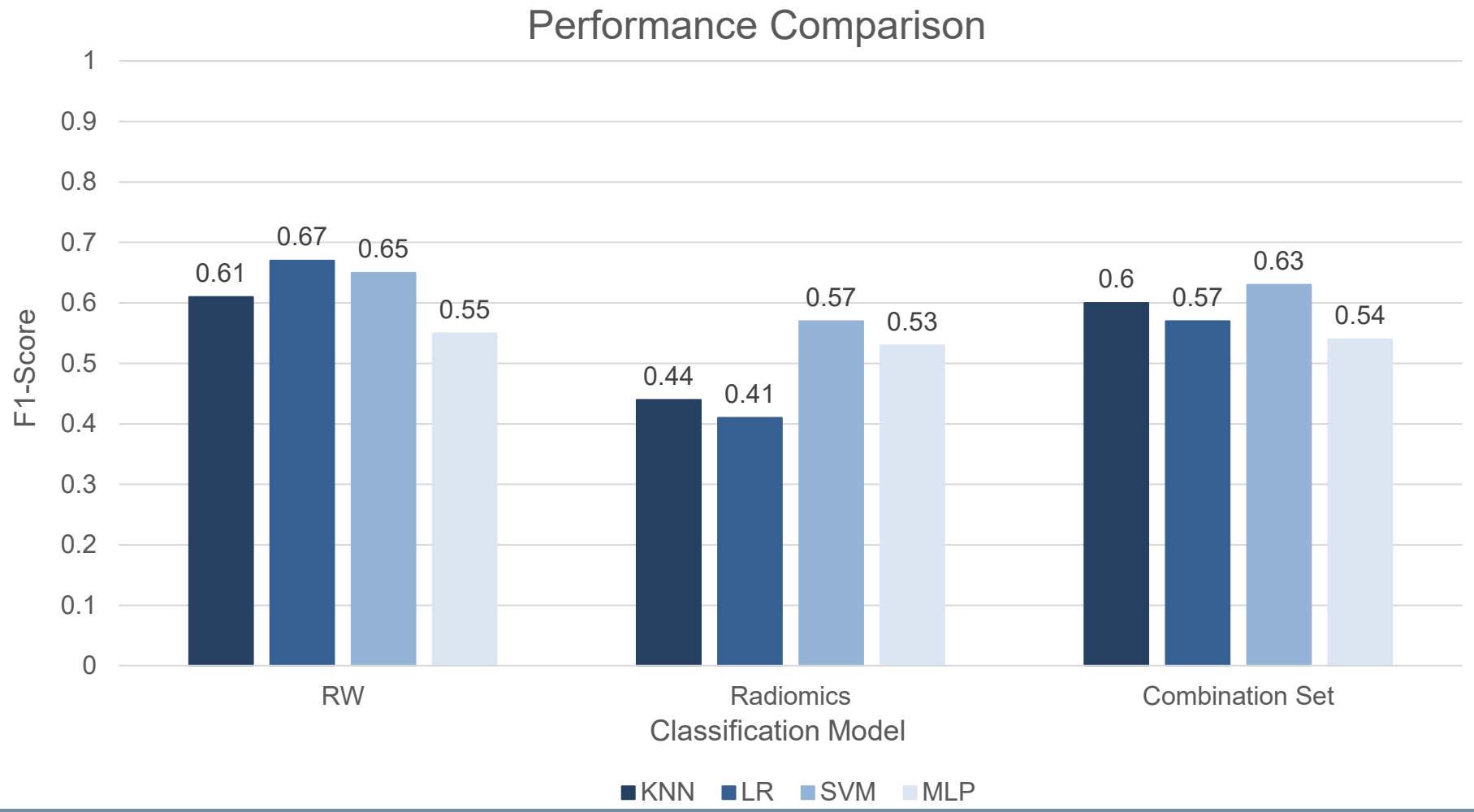
04 **Multi-Layer Perceptron (MLP)**

→ All the models were tuned by *GridSearch Cross Validation* to find the best hyperparameters
→ 70% data for training and 30% for external validation
→ On the training set was implemented a K-fold Cross-Validation (k = 5)

Experimental Results: Disease Control Rate



Experimental Results: Progression Free Survival

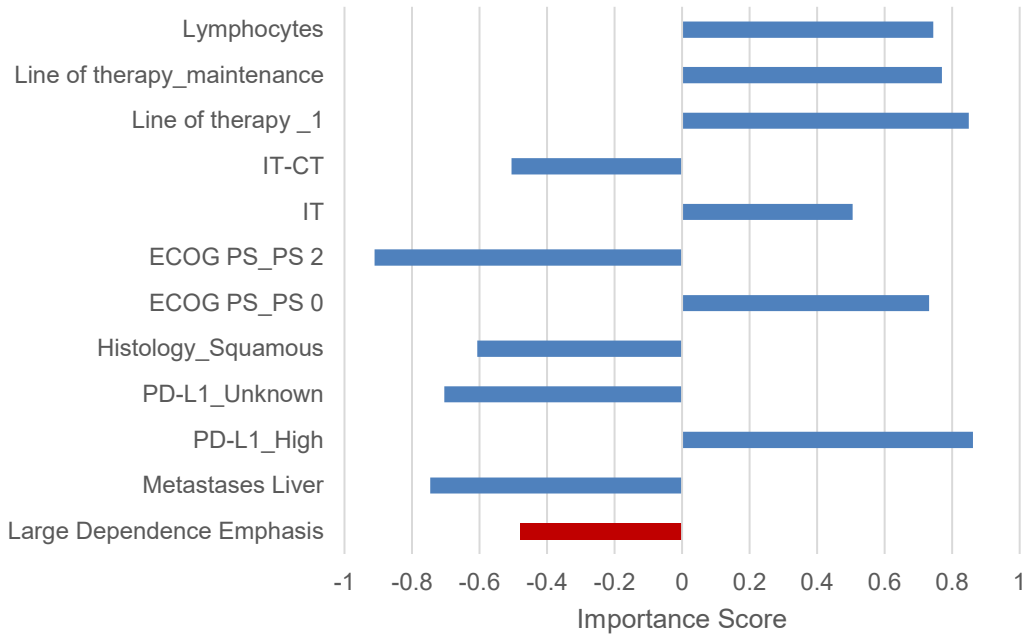


Experimental Results: Overall Survival

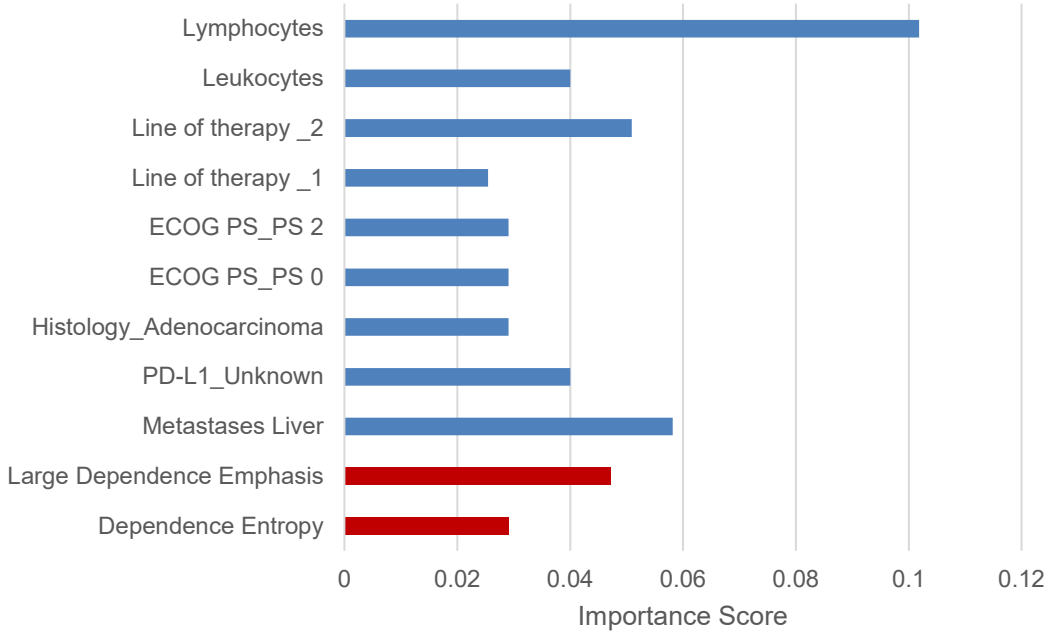
Experimental Results: Feature Importance Analysis for Disease Control Rate

The feature importance analysis is presented for the best results obtained on the **DCR** outcome for the **LR** and **SVM** trained on the combination set (**RW** + **RD**).

Logistic Regression



Support Vector Machine



Conclusions

- Overall, ML algorithms showed to better perform using Real World Evidence features. However, the results obtained from the combination set suggest that the integration of different data set (RW data + radiomics) need to be further exploit.
- Radiomic features seem have a greater importance when they are implied to predict response to immunotherapy rather then survival outcomes.

Future Developments:

Include more patients' data

Set a clear image acquisition protocol

Assure a higher variability in terms of radiomic features

Thanks for your time and attention