Trustworthy Artificial Intelligence Models using Real World and liquid biopsy data for the prediction of Immunotherapy efficacy in Non-Small Cell Lung Cancer patients

Aim
Integration of RW (clinical, radiological, lab exams) and blood genomic (G) data at the baseline of IO treatment to develop an explainable ML model with improved accuracy to predict the IO response in patients with advanced NSCLC.

Methodology

Introduction
• PD-L1 is the only biomarker used to predict the response to Immunotherapy (IO) in advanced non-small-cell lung cancer (NSCLC), but its predictive performance is not satisfactory.
• Due to the complexity of the immune system there is a need to integrate different type of data (e.g. clinical and omics data) to achieve an accurate prediction.
• Liquid biopsy (LB) is a feasible tool able to detect genomic alterations in cancer at baseline.
• Machine learning (ML) allows the integration of different types of data (e.g. Real world (RW) and genomics (G)) which leads to increased predictive accuracy.
• Explainable Artificial Intelligence (XAI) is a valuable tool for understanding how the ML models generate the results.

Results

Statistical analysis:
DRC: 42.3% non responders
OS: 69.6% OS≥6monts, 81.4% OS<6months
ORR: 75.3% non responders, 24.7% responders

Machine learning analysis:
• CatBoost (CB) achieved the best results on the test dataset for all the outcomes.
• Adding G features increases test accuracy for 8.9%, 27.7%, and 13.8% for DCR, OS and ORR, respectively.

Conclusion
Our results suggest that adding genomic features to RW features increases the ML prediction accuracy for all tested outcomes (DCR, OS and ORR). The explainable AI algorithms showed a negative predictive role of TP53, KIT and BRAF mutation.

References:

Corresponding author
Arsela Prelaj, MD
Arsela.Prelaj@istitutotumori.mi.it
Arsela.Prelaj@polimi.it
Giuseppe.lorusso@istitutotumori.mi.it

Output | Accuracy
---|---
DCR | 0.56 | 0.61
OS | 0.65 | 0.83
ORR | 0.65 | 0.74

SHAP summary graphs:

<table>
<thead>
<tr>
<th>NLR</th>
<th>10/C10</th>
<th>Line of IO</th>
<th>Age at IO</th>
<th>LDH</th>
<th>PKR1</th>
<th>ECOG PS</th>
<th>BMI</th>
<th>ALC</th>
<th>pack/year</th>
<th>A1C</th>
<th>sum of other 17 features</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRC outcome, RW features</td>
<td>DRC outcome, RW+G features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>