Combing Multiple Immunotherapy Studies and Real-World Data Improves Prediction of IO Treatment Efficacy and Highlights Key Driving Features

Gustavo Arango*, Elly Kipkogei, EtaI Jacob. Data Science and AI, Oncology R&D, Astrazeneca, Waltham, MA

Background Despite the clinical success of immune checkpoint blockade therapies, many patients do not respond to treatments or become resistant. Previous attempts to predict treatment efficacy suffered from limited accuracy and deficiency to uncover determinants of response. Here, we introduce an AI framework which addresses these issues.

Methods We present a new explainable deep learning framework based on transformer architecture\(^1\)\(^2\) which combines data with different feature sets (including sparse date) and clinical endpoints for survival prediction or classification. This framework includes: (1) A new loss function based on a sigmoid approximation of Harrell’s concordance-index.\(^3\) (2) Explainability module providing feature importance and similarity score between features based on mutual contribution to predictions. (3) Transfer learning strategy to enable leveraging diverse clinical datasets in the public or private domain.

Results We utilized seven data sets comprising of more than 140,000 patients from IO, targeted and Chemotherapy treatments to benchmark our prediction models (table 1) in addition to 10 train/test splits performance evaluations.

Consistently, our framework outperformed other methods previously described in the literature, including CoxPH\(^4\) and random survival forest.\(^5\)\(^6\) For example, using the concordance index, our framework achieved 0.60 (0.04) vs. 0.60 (0.04) of the second-best method (Random survival forest in all cases) on MYSTIC IO arms clinical data. This improvement was a result of including transfer learning in the training process (table 2) which also achieved better performance in less training steps (figure 1).

Utilizing our Explainability module, we identified key features driving response prediction consistent with previous publications. For example, in Choswell et al. dataset, we identified Albumin, NLR, Chemo-before-IO-treatment and TMB as the most important features (figure 2). We also identified in sparse mutation calls of 469 genes from Samstein et al. dataset, functional modules of several genes only, each with a strong predictive power. For example, the functional module comprising of the genes: AKT2, BTK, CDC73, HLA-B, IKBKE, INPP1L1, RFWD2, TRAF2 and WHSC1, related to adaptive immunity, stratified patients to two groups with a Hazard Ratio of 0.58 for the Samstein dataset and 0.42 for the validation dataset (figure 3).

Conclusions We propose a new framework with state-of-the-art performance in survival prediction and potential to uncover biological and clinical insights related to patient response and resistance. Importantly, our framework simplifies the process of translating complex AI models to clinical practice and may accelerate the benefit immunotherapy can bring to patients.

REFERENCES


Abstract 1273 Table 1 Description of the datasets used to train and evaluate the clinical transformer

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Cancer Type</th>
<th>Description</th>
<th>Task</th>
<th>Samples</th>
<th>Arm</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mystic (1)</td>
<td>NSCLC</td>
<td>Stage IV NSCLC treatment naïve anti-PDL1 and CTLA-4 combo. Only patients with available features to predict TMB and HLA (spatial).</td>
<td>Survival</td>
<td>150</td>
<td>IO</td>
<td>15</td>
</tr>
<tr>
<td>OAK (5)</td>
<td>NSCLC</td>
<td>Stage IV NSCLC patients treated with anti-PDL1 after failing chemotherapy</td>
<td>Survival</td>
<td>396</td>
<td>IO</td>
<td>418</td>
</tr>
<tr>
<td>Choswell et al. (2016)[6]</td>
<td>Pan Cancer</td>
<td>Patients treated with anti-PD1/PDL1 and/or cetuximab in multiple cancer types. This dataset consists of 17 features (e.g., TMB, HLA, IDH1, IDH2, ALB, TNFA)</td>
<td>Survival</td>
<td>1,479</td>
<td>IO</td>
<td>17</td>
</tr>
<tr>
<td>Samstein et al. (2019)[9]</td>
<td>Pan Cancer</td>
<td>Response to IO (anti-PD1/PDL1 and/or cetuximab) in a Pan cancer setting. This dataset includes molecular data (mutations and copy numbers) and somatic features for 134 patients</td>
<td>Survival</td>
<td>1,610</td>
<td>IO</td>
<td>474</td>
</tr>
<tr>
<td>Thomson et al. (2016)[10]</td>
<td>Pan Cancer</td>
<td>The Spanish landscape of cancer. A study on the UCIA data that defines features associated with the TMB</td>
<td>Survival</td>
<td>6,012</td>
<td>SOC</td>
<td>49</td>
</tr>
<tr>
<td>Mine et al. (2018)[12]</td>
<td>Pan Cancer</td>
<td>Whole-exome sequencing (WES) of 299 tumors from patients with clinically actionable features associated with immunotherapy</td>
<td>Validation dataset</td>
<td>249</td>
<td>IO</td>
<td>&gt;10,800</td>
</tr>
</tbody>
</table>