# Evaluation of deep learning techniques in RNA sequencing data for the prediction of patient response to immune checkpoint inhibitor treatment.

Cancer is considered one of the biggest public health problems worldwide. Although research of Cancer disease has improved and increased in the last decades, not all types of cancer have been benefited to the same extent by these investigations. For instance, kidney cancer. This work focuses on the most common type of Renal Cell Carcinoma, clear cell Renal Carcinoma. Concretely, the metastasis phase of this disease also known as metastasis Renal Cell Carcinoma (mRCC).

In recent years, immune checkpoint treatments have demonstrated to improve Progression Free Survival (PFS) and overall survival for mRCC patients in several phase III clinical trials. Consequently, current research lines are looking for genes signatures to identify predictors of response to treatments using genomic and transcriptomic data. However, the healthcare domain is characterized by heterogeneous and often high-dimensional data sets. Therefore, the application of Deep Learning (DL) approaches appears to be and appropriated solution.

The main objective of this work is studying the performance of different DL models (autoencoder and CNN) intended to predict the PFS of patients treated with NIVOLUMAB, an immune checkpoint inhibitor treatment. For this purpose, we have followed several research lines including the creation of combined dataset with clinic and RNA sequencing data from 181 mRCC patients and comparing the DL models results with the performance of traditional Machine Learning (ML) models. Finally, we came up with an interpretability analysis of those black-box models using LIME and SHAP values.

Results achieved during the research confirmed we are able to model the response for NIVOLUMAB using RNA sequencing data. However, DL models have not demonstrated to be significantly better than traditional ML methods, when predicting the response of patients. Although, promising results were showed by deep autoencoder which achieved 68.9\% accuracy, most accurate model was logistic regression classifier, a classic ML algorithm which achieved 86.4\% of hit rate.

Finally, interpretability results revealed that most relevant genes for the decision making were related with the immune response and the regulation of kinase which meets the biological explanation. However, interpretability results revealed that clinic features seemed to have more relevance when making the decision.

|  |  |  |
| --- | --- | --- |
| Model | Data set | Accuracy |
| Logistic regression | 30 most reliable genes and MSKCC variable. | 0.864 |
| Deep autoencoder + decision tree | 30 most reliable genes and 12 clinic variables. | 0.689 |
| CNN | 400 most reliable genes | 0.541 |

Cambios Jose:

# Evaluation of deep learning techniques in RNA sequencing data for the prediction of patient response to immune checkpoint inhibitor treatment.

Cancer is considered one of the biggest public health problems worldwide. Although research of Cancer disease has improved and increased in the last decades, not all types of cancer have been benefited to the same extent by these investigations. For instance, kidney cancer. This work focuses on the most common type of Renal Cell Carcinoma, clear cell Renal Carcinoma. Concretely, the metastasis phase of this disease also known as metastasis Renal Cell Carcinoma (mRCC).

In recent years, immune checkpoint treatments have demonstrated to improve Progression Free Survival (PFS) and overall survival for mRCC patients in several phase III clinical trials. Consequently, current research lines are looking for gene signatures to identify predictors of response to treatments using genomic and transcriptomic data. However, the healthcare domain is characterized by heterogeneous and often high-dimensional data sets. In this sense, the application of Deep Learning (DL) approaches appears as a suitable approach.

The main objective of this work is studying the performance of different DL models (autoencoder and CNN) intended to predict the PFS of patients treated with NIVOLUMAB, an immune checkpoint inhibitor treatment. For this purpose, we have followed several research lines including the creation of combined datasets with clinic and RNA sequencing data from 181 mRCC patients, and comparing the results of the DL models against the performance of traditional Machine Learning (ML) models. Finally, we came up with an interpretability analysis of those black-box models using LIME and SHAP values.

Results achieved during the research confirmed that we are able to model the response for NIVOLUMAB using RNA sequencing data. However, DL models have not demonstrated to be significantly better than traditional ML methods, when predicting the response of patients. Although promising results were achieved by deep autoencoder which accomplished 68.9\% accuracy, the most accurate model was logistic regression classifier, a classic ML algorithm which achieved 86.4\% of hit rate.

Finally, interpretability results revealed that most relevant genes for the decision making were related with the immune response and the regulation of kinase which meets the biological explanation. However, interpretability results revealed that clinic features seemed to have more relevance when making the decision. Further research is intended in the field of the DL analysis, as data codification and data structure could bias the results.

|  |  |  |
| --- | --- | --- |
| Model | Data set | Accuracy |
| Logistic regression | 30 most reliable genes and MSKCC variable. | 0.864 |
| Deep autoencoder + decision tree | 30 most reliable genes and 12 clinic variables. | 0.689 |
| CNN | 400 most reliable genes | 0.541 |

Cambios oncólogos:

**INSTRUCCIONES**

**Abstract Body/Table:** The body of your abstract should describe the background, methods, results, and conclusions of your research. You may type your abstract directly into the text box, cut and paste it from an existing document, or upload a text file of your abstract. **Do not exceed 2,600 characters** (approximately 300 to 350 words) for the total of your abstract title, body including section titles, and table. The character count does not include spaces or author names or institutions. One data table is permitted per abstract. The composition process does not enable shading or the merging of cells with centered text. Limit your table to no more than 10 rows and eliminate the need for shading or merged cells with centered text. Illustrations and figures are not permitted.

**Title: Evaluation of deep learning techniques in RNA sequencing data**

**for the prediction of metastatic renal cancer patient response to immune checkpoint inhibitors.**

**Authors:**

**Background:**

Immune checkpoint inhibitors have become a cornerstone in the management of metastatic renal cell carcinoma (mRCC). However, to identify the most suitable patients for this treatment is an unmeet medical need.

We aimed to explore the utility of deep learning (DL) integrating clinical and molecular data to predict response to immunotherapy.

**Methods**: We conducted a retrospective analysis using publicly available data from patients treated with nivolumab in the clinical trials Chekmate 009, 010 and 025. The primary objective was to assess the performance of different DL models (autoencoder and CNN) predicting the PFS of these patients.

With that scope, we followed several research lines including the creation of combined datasets with clinic and RNA sequencing and comparing the results of the DL models against the performance of traditional Machine Learning (ML) models. Finally, we came up with an interpretability analysis of those black-box models using LIME and SHAP values.

**Results**: Clinical and transcriptomic data were available form from 181 nivolumab-treated patients. Outcomes achieved confirmed that we can model the response for NIVOLUMAB using RNA sequencing data. [Table 1] However, DL models have not demonstrated to be significantly better than traditional ML methods, when predicting the response of patients (p-value = 0.068) Although promising results were achieved by deep autoencoder [68.9% accuracy] the most accurate model was logistic regression classifier, which achieved 86.4% of hit rate. Interpretability results revealed that most relevant genes for the decision making were related with the immune response and the regulation of kinases. Regarding interpretability, best results were achieved integrating both transcriptomic and clinical data (70% of the DL and ML tested models achieved higher hit rates with combined data set. Concretely, logistic regression classifier improved its accuracy in 11%).

**Conclusions**: The integration of clinical and molecular data could lead to more accurate predictions of outcome than any dataset by its own. However, further research is intended in the field of the deep learning analysis, as data codification and data structure could bias the results. The ongoing study **ART** (**A**rtificial Intelligence in **R**enal **T**umors) will address this issue prospectively.

*Model Data set Accuracy*

*Logistic regression 30 most reliable genes and MSKCC variable. 0.864*

*Deep autoencoder + decision tree 30 most reliable genes and 12 clinic variables 0.689*

*CNN 400 most reliable genes 0.675*

*Table 1: Results in terms of accuracy for DL and ML best scored models.*