Response to Reviewer 1 Comments

The authors are grateful for your reading and suggestion. We worked to ameliorate all the indicated items that must be improved: justifying the research choices, describing the methods adequately, and clarifying the presentation of the results.

Point 1: The entire study was based on data mining, therefore the lack of independent validation with wet lab experiments using cell lines or clinical specimens weakens this study. Please validate your results with functional experiments.

Response 1: This work does not require wet-lab experiments, since our main goal was to compare distinct gene signatures from the literature and to generate new gene signatures using feature selection methods. Our work applied the validations and statistical analysis mostly used in machine learning studies. Using two independent datasets, our results had statistical significance (Adj p-values and FDR) and best-performing metrics (AUC Uno). Our methodology contains a series of justified steps:

- A differential gene expression analysis comparing normal against tumor tissues, and non-metastatic against metastatic samples of ccRCC;
- An internal validation using RNAseq and clinical data from TCGA-KIRC data;
- An external validation using RNAseq and clinical data from ICGC-RECA;
- An enrichment analysis of gene-disease associations (GDAs) with expert-curated databases;
- A statistical comparison of cancer stages with protein expression of each gene in a third dataset of ccRCC from Clinical Proteomic Tumor Analysis Consortium (CPTAC);

The wet-lab experiments are not in our scope of work. Here, we present a study on ccRCC which combines bioinformatics analysis and machine learning. The feature selection and machine learning methods are computational methodologies that can be executed before experimental validations. Further studies can be executed to elucidate the renal carcinogenic processes using wet-lab experiments, with a different focus and goal.

We believe that our Machine Learning approach is appropriate for publication in the Special Issue of *Artificial Intelligence and Machine Learning in Cancer Research*, given our keywords, methodology, and results.

In order to elucidate this point, we elaborated the explanations needed in the penultimate paragraph of the introduction section:

> Nowadays, the scientific community is still searching for new biomarkers for ccRCC, and feature selection methods using survival analysis provide a robust exploratory methodology before experimental validations. Survival analysis is a field of statistics that predicts the time until an event of interest happens in many domains [15]. The most commonly used method for survival analysis is the Cox Regression model [16]. The Cox model is semi-parametric, that is, the distribution of the event of interest is unknown. In addition, Cox models are widely used for censored data, i.e., when the event is not observed during the study period due to loss to follow-up, study termination, or the patient's death by other causes. Regularized Cox models provide suitable predictions for high-dimensional data using penalty functions with the main regularizers Lasso-Cox, Ridge-Cox, and Elastic net-Cox [15]. Ensemble learning methods are committees of machine learning models, in other words, they combine the majority of the votes for each model in an ensemble or they adjust the weighted vote of each model. Moreover, this

approach results in a more robust, efficient, and stable model compared to singular models. In this work, we applied Cox models and ensemble methods using gene expression to predict the overall survival (OS) after diagnosis of ccRCC.

Lasso-Cox regression generated most of the reviewed gene signatures for ccRCC [9,10,13,17–19]. All the studies reviewed in this work use TCGA-KIRC dataset to train and validate the results. Fewer studies validated their results with other datasets such as GEO database [10,13,2], ICGC-RECA [2,11], and data from Fudan University Shanghai Cancer Center (FUSCC). The most common methodologies used to discover and validate gene signatures were differentially expression analysis (DEA), and gene set enrichment analysis (GSEA). Only one study compared its methodology to 3 other biomarker signatures from our literature selection [9]. In addition, there was a lack of comparisons between the gene signatures. As far we know, our study presents a most comprehensive comparison between gene signatures, including ensemble methods, machine learning, and feature selection.

We also have modified the conclusion by adding the following clarifications:

Our main goal was to compare distinct gene signatures from literature and generate new gene signatures using feature selection methods. We contributed by providing a list of new genes, some of them not previously reported as biomarkers for ccRCC. The gene signature created by the mRMR method achieved a score of 0.82 with AUC, being the best performer. We identified two clusters of genes with high expression (SAA1, OTX1, ZIC2, LINC01732, GNB3 and IL4) and low expression (AL353637.1, AR, HHLA2, LIMCH1, SEMA3G, DPP6, and FOXJ1) which are correlated with poor prognosis. We validated our 13-gene signature for ccRCC and confirmed our results with the literature, and by comparing each cancer stage of ccRCC with CPTAC and the survival effects of gene expression of individual genes in TCGA. We believe that further studies on the involvement of these genes in renal carcinogenic processes can improve the understanding of cancer biology. After experimental validations, new possible applications in clinical practices can benefit from the biomarker found with machine learning and feature selection.