



OPEN ACCESS

EDITED BY

Yanqing Liu,
Columbia University, United States

REVIEWED BY

Sijia Yue,
Columbia University, United States
Sisi Chen,
University of Pennsylvania, United States
Jun Xia,
Massachusetts General Hospital and Harvard
Medical School, United States

*CORRESPONDENCE

Fangshi Xu
✉ xfs8113@163.com

RECEIVED 10 June 2024

ACCEPTED 05 September 2024

PUBLISHED 19 September 2024

CITATION

Xu F and Lai J (2024) Commentary: Immune cell infiltration and prognostic index in cervical cancer: insights from metabolism-related differential genes. *Front. Immunol.* 15:1446741. doi: 10.3389/fimmu.2024.1446741

COPYRIGHT

© 2024 Xu and Lai. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Commentary: Immune cell infiltration and prognostic index in cervical cancer: insights from metabolism-related differential genes

Fangshi Xu^{1*} and Jiawei Lai²

¹Department of Vascular Surgery, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China, ²Department of Urology, Shaanxi Provincial People's Hospital, Xi'an, Shaanxi, China

KEYWORDS

bioinformatic technology, machine learning algorithms, prognostic models, cancer, clinical cohort

A Commentary on:

Immune cell infiltration and prognostic index in cervical cancer: insights from metabolism-related differential genes

By Ma B, Ren C, Yin Y, Zhao S, Li J and Yang H (2024). *Front Immunol.* 2024 May 22;15:1411132. doi: 10.3389/fimmu.2024.1411132.

Introduction

With the ongoing advancement of bioinformatic technology, utilizing public transcriptome data to predict the clinical outcomes of cancer patients has become standard. A recent study by Ma B et al., titled 'Immune cell infiltration and prognostic index in cervical cancer: insights from metabolism-related differential genes,' served as an inspiration for us. This study introduced a novel metabolism-related (MR) model designed to enhance the prognostic and therapeutic evaluations of patients with cervical cancer (CC). The model demonstrated significant prognostic value and was closely linked to the levels of immune cell infiltration and the response to immunotherapy. Although the study was well-designed and analyzed, it warrants further enhancements in the selection of machine learning algorithms, comprehensive bioinformatic evaluations, and validation through extensive clinical cohorts.

Selection of machine learning algorithms

Currently, lasso regression is the most frequently used machine learning approach for clinical modeling (2). Similarly, the MR model was also reliant on lasso regression for construction (1). In 2024 alone, more than 800 studies have utilized this method to establish clinical assessment models in various cancers. The reasons for its wide application

are predominantly due to that lasso regression effectively handles the dilemma of variable selection and model overfitting through creating a penalty function λ (3).

At present, other machine learning algorithms in exception with lasso regression have also been widely applied in the clinical modeling process owing to their particular advantages. For instance, the support vector machine (SVM) is a classical supervised learning algorithm, which has expertise in addressing the issues referring to multi-classification and sample disequilibrium (4). Yu Y et al. have utilized this method to characterize heart failure (5). As for random forest, it has excelled when dealing with non-linear data and complex relational data (6), by which Zhang Y et al. constructed a prediction model for deep vein thrombosis in patients with digestive system tumors (7). Clearly, lasso regression is not the only option in clinical modeling. Notably, the original intention of modeling is to accurately predict survival outcome or therapeutic response of cancer patients, which can be evaluated by C-index or area under curve (AUC) value (8). This elicits a question: Does the models constructed by lasso regression have the best predictive performance, or is lasso regression the optimal solution for modeling?

To address this point, Liu Z et al. demonstrated a remarkable research strategy, termed comparison of multiple machine learning algorithms (9). In their study, 101 kinds of prediction models were fitted based on various machine learning algorithms and their combinations, such as support vector machine-recursive feature elimination (SVM-RFE), random survival forest (RSF), and Ridge method. They then calculated the C-index of each model in all validation cohorts and found the optimal solution for immune lncRNA signature in colorectal cancer (CRC), which was the combination of lasso regression and step cox regression. Therefore, there may be better approaches for fitting the MR model, which awaits verification by comparing multiple machine learning algorithms.

Nonetheless, the advantages of lasso regression over other algorithms should not be overlooked. First, the model fitted by lasso regression offers significant interpretability (10). In contrast, models such as random forest, SVM, XGBoost, and artificial neural network (ANN) often result in 'black box' models, which are challenging to interpret internally. This difficulty arises because the predictive results of black box models can typically only be explained by the correlations between input and output, without providing a specific interpreting or reasoning process (11). Furthermore, black box models are often nonlinear, making their decision boundaries more elusive than those of the linear models well-represented by lasso regression. Second, lasso regression is adept at addressing model overfitting and high model complexity. However, other algorithms may not always be competent. For example, when no multicollinearity exists between independent variables, Ridge regression may diminish the predictive performance of the constructed model (12). Similarly, the random forest model is susceptible to overfitting on certain noisy features, and its training time is relatively long due to its dependence on training multiple decision trees simultaneously. Therefore, although lasso regression may not always excel in diagnostic performance, it remains a stable solution with low fault tolerance for most clinical analyses.

Comprehensive bioinformatics evaluation

Prognostic analysis is a crucial component of personalized cancer medicine and represents a primary concern for patients. In their study, Ma B et al. evaluated the survival differences between patients with high and low MR risks and the accuracy of the MR model in predicting the overall survival rate (OSR) at 1, 3, and 5 years (1). However, other prognostic properties also merit further exploration. For instance, clinical subgroup analyses could test the predictive capacities of the MR model in cancer patients at different clinical stages (13). Additionally, while the AJCC and TNM systems are established foundations for cancer prognostic assessments (14), replacing them with the MR score is impractical. It could be clinically more significant to investigate whether the MR risk score could enhance the predictive accuracy or decision-making benefits of the AJCC or TNM systems using decision curve analysis (DCA) (15).

Regarding immunotherapy response, Ma B et al. explored the associations between MR risk score and Tumor Immune Dysfunction and Exclusion (TIDE) score, tumor mutation burden (TMB), and expressions of immune checkpoints (ICs). These are considered valuable biomarkers for predicting the efficacy of immune checkpoint inhibitors (ICIs) (16–18). Notably, some clinical cohorts could be used for therapeutic effect analysis due to their available transcriptome and clinical data.

Compared to the work of Ma B et al., Betancor YZ et al. demonstrated some advantages in predicting therapeutic response (19). They utilized real-world evidence from three related clinical trials (CheckMate-009, CheckMate-010, and CheckMate-025) to assess the predictive capacity of a three-gene model for anti-PD-1 blockade therapy, highlighting its effectiveness in actual clinical practice. Additionally, considering that nivolumab is the first-line treatment for advanced renal cell carcinoma (RCC), the topic chosen by Betancor YZ et al. may hold more appeal for researchers in this field. Clearly, the research of Ma B et al., particularly in prognostic and immunotherapy analyses, could be further refined with more comprehensive bioinformatic investigations.

Validation of extensive clinical cohort

Although prediction models are increasingly prevalent, only a minority have undergone validation using the authors' own clinical cohorts, significantly limiting their clinical applicability (20). Regrettably, the prognostic value of the MR model was validated only in an external cohort (GSE52904 dataset) and not in more comprehensive clinical cohorts.

There are two primary drawbacks to relying solely on public data. First, demographic differences among various clinical cohorts can introduce selection bias. For example, the age range of patients in the GSE52904 dataset spanned from 24 to 74 years (Mean=50.5 years) (21), whereas the average age in another cervical cancer cohort, GSE6791, was 43.9 years (22). Additionally, there were discrepancies among patients in terms of race, clinical stages, pathological types, and follow-up times across different cohorts.

Thus, using external clinical data could help minimize this selection bias to some extent. Second, the clinical validation within their own centers is a critical first step toward the clinical application of constructed models. If models developed by authors do not perform well in their own clinical cohorts, it is challenging to convince surgeons of their clinical potential and value.

Moreover, since the modeling process tends towards purely mathematical operations, the lack of clinical data correction can cause the constructed model to deviate from actual circumstances. For instance, a six-genes ZNF family model developed for clinical assessments of esophageal cancer (ESCS) (23) showed no significant differences in ZNF502 expressions between tumor and normal clinical samples, as determined by PCR. This undoubtedly diminishes the credibility of this model to some extent. Collectively, it is crucial to advocate for more extensive clinical validation.

Some suggestions on clinical data validation

Validating models or risk signatures on external clinical cohorts, such as those from a researcher's own center, significantly enhances their credibility, although this path is challenging yet worthwhile. From our perspective, there are two critical enabling factors. First, long-term and meticulous follow-up is essential. Prognostic assessment plays a pivotal role in individualized cancer therapy. Collecting detailed data on patient survival or therapeutic responses will greatly inform these issues. Second, the construction of a genome-wide library is crucial. With the rapid development of genomics, selecting appropriate patients and performing exon sequencing on their clinical samples can link molecular features to clinical phenotypes (24), enhancing our understanding of disease pathogenesis. However, these strategies are accompanied by considerable human and economic costs.

Conclusions

In this manuscript, we propose several approaches to refine the MR risk score, thereby enhancing its clinical application. First, optimizing the machine learning algorithm may improve its prediction performance. Second, the increased use of bioinformatics technologies, such as Decision Curve Analysis (DCA), is instrumental in assessing the clinical value of the MR risk score. Third, further validation using external clinical data will enhance the credibility of the MR risk score.

References

1. Ma B, Ren C, Yin Y, Zhao S, Li J, Yang H. Immune cell infiltration and prognostic index in cervical cancer: insights from metabolism-related differential genes. *Front Immunol.* (2024) 15:1411132. doi: 10.3389/fimmu.2024.1411132
2. Rafique R, Islam SMR, Kazi JU. Machine learning in the prediction of cancer therapy. *Comput Struct Biotechnol J.* (2021) 19:4003–17. doi: 10.1016/j.csbj.2021.07.003
3. Ternès N, Rotolo F, Michiels S. Empirical extensions of the lasso penalty to reduce the false discovery rate in high-dimensional Cox regression models. *Stat Med.* (2016) 35:2561–73. doi: 10.1002/sim.6927
4. Ding C, Bao TY, Huang HL. Quantum-inspired support vector machine. *IEEE Trans Neural Networks Learn systems.* (2022) 33:7210–22. doi: 10.1109/TNNLS.2021.3084467

The rapid expansion of bioinformatics technology has broadened the scope of cancer research. However, the majority of bioinformatics research encounters common issues, such as the selection of modeling algorithms and the scarcity of real-world data. These limitations significantly restrict the clinical application of novel prediction models. We highlight three improvement measures to address these deficiencies, which will contribute to a better understanding of the application of bioinformatics in the cancer field.

Author contributions

FX: Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing. JL: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by Natural Science Foundation of Shaanxi Province (2024JC-YBQN-0905).

Acknowledgments

All authors would like to thank Second Affiliated Hospital of Xi'an Jiaotong University for its support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

5. Yu Y, Wang L, Hou W, Xue Y, Liu X, Li Y. Identification and validation of aging-related genes in heart failure based on multiple machine learning algorithms. *Front Immunol.* (2024) 15:1367235. doi: 10.3389/fimmu.2024.1367235
6. Ghosh D, Cabrera J. Enriched random forest for high dimensional genomic data. *IEEE/ACM Trans Comput Biol Bioinf.* (2022) 19:2817–28. doi: 10.1109/TCBB.2021.3089417
7. Zhang Y, Ma Y, Wang J, Guan Q, Yu B. Construction and validation of a clinical prediction model for deep vein thrombosis in patients with digestive system tumors based on a machine learning. *Am J Cancer Res.* (2024) 14:155–68. doi: 10.62347/LNDL8700
8. Sammon JD, Abdollah F, D'Amico A, Gettman M, Haese A, Suardi N, et al. Predicting life expectancy in men diagnosed with prostate cancer. *Eur urology.* (2015) 68:756–65. doi: 10.1016/j.eururo.2015.03.020
9. Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun.* (2022) 13:816. doi: 10.1038/s41467-022-28421-6
10. Zhao Y, Long Q. Multiple imputation in the presence of high-dimensional data. *Stat Methods Med Res.* (2016) 25:2021–35. doi: 10.1177/0962280213511027
11. Crețu AM, Guépin F, de Montjoye YA. Correlation inference attacks against machine learning models. *Sci Adv.* (2024) 10:eadj9260. doi: 10.1126/sciadv.adj9260
12. Yang R, He F, He M, Yang J, Huang X. Decentralized kernel ridge regression based on data-dependent random feature. *IEEE Trans Neural Networks Learn systems.* (2024) 12(7):1–10. doi: 10.1109/TNNLS.2024.3414325
13. Xu F, Cai D, Liu S, He K, Chen J, Qu L, et al. N7-methylguanosine regulatory genes well represented by METTL1 define vastly different prognostic, immune and therapy landscapes in adrenocortical carcinoma. *Am J Cancer Res.* (2023) 13:538–68.
14. Dicu-Andreescu IG, Marinca AM, Ungureanu VG, Ionescu SO, Prunoiu VM, Brătucu E, et al. Current therapeutic approaches in cervical cancer based on the stage of the disease: is there room for improvement. *Medicina (Kaunas Lithuania).* (2023) 59(7):1229. doi: 10.3390/medicina59071229
15. Tsatsanis A, Hozo I, Vickers A, Djulbegovic B. A regret theory approach to decision curve analysis: a novel method for eliciting decision makers' preferences and decision-making. *BMC Med Inf decision making.* (2010) 10:51. doi: 10.1186/1472-6947-10-51
16. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* (2018) 24:1550–8. doi: 10.1038/s41591-018-0136-1
17. Qin Y, Huo M, Liu X, Li SC. Biomarkers and computational models for predicting efficacy to tumor ICI immunotherapy. *Front Immunol.* (2024) 15:1368749. doi: 10.3389/fimmu.2024.1368749
18. Xu F, Guan Y, Zhang P, Xue L, Ma Y, Gao M, et al. Tumor mutational burden presents limiting effects on predicting the efficacy of immune checkpoint inhibitors and prognostic assessment in adrenocortical carcinoma. *BMC endocrine Disord.* (2022) 22:130. doi: 10.1186/s12902-022-01017-3
19. Betancor YZ, Ferreiro-Pantín M, Anido-Herranz U, Fuentes-Losada M, León-Mateos L, García-Acuña SM, et al. A three-gene expression score for predicting clinical benefit to anti-PD-1 blockade in advanced renal cell carcinoma. *Front Immunol.* (2024) 15:1374728. doi: 10.3389/fimmu.2024.1374728
20. Li J, Kong Z, Qi Y, Wang W, Su Q, Huang W, et al. Single-cell and bulk RNA-sequence identified fibroblasts signature and CD8+ T-cell - fibroblast subtype predicting prognosis and immune therapeutic response of bladder cancer, based on machine-learning bioinformatics retrospective study. *Int J Surg (London England).* (2024) 110(8):4911–31. doi: 10.1097/JS9.0000000000001516
21. Medina-Martinez I, Barrón V, Roman-Bassaure E, Juárez-Torres E, Guardado-Estrada M, Espinosa AM, et al. Impact of gene dosage on gene expression, biological processes and survival in cervical cancer: a genome-wide follow-up study. *PLoS One.* (2014) 9:e97842. doi: 10.1371/journal.pone.0097842
22. Pyeon D, Newton MA, Lambert PF, den Boon JA, Sengupta S, Marsit CJ, et al. Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Res.* (2007) 67:4605–19. doi: 10.1158/0008-5472.CAN-06-3619
23. Hong K, Yang Q, Yin H, Wei N, Wang W, Yu B. Comprehensive analysis of ZNF family genes in prognosis, immunity, and treatment of esophageal cancer. *BMC cancer.* (2023) 23:301. doi: 10.1186/s12885-023-10779-5
24. Dehghan A. Genome-wide association studies. *Methods Mol Biol (Clifton NJ).* (2018) 1793:37–49. doi: 10.1007/978-1-4939-7868-7_4