# Cancer prognosis prediction models based on deep learning

Chengyu Sun *

College of Life Sciences and Technology, Huazhong University of Science and Technology, Wuhan, China

* Corresponding author: chengyu_sun@hust.edu.cn

**Abstract.** Cancer is a hot issue that more and more people are concerned about at present. In recent years, the incidence of cancer in the world has been rising, but the treatment methods for cancer have been lacking in specificity. Some researchers have proposed the use of deep learning algorithms to assist cancer treatment, which can be analyzed according to the characteristics of patients, and has made progress in the prediction and treatment of a diverse of different cancers. However, there are few articles summarizing the application of deep learning in cancer prognosis. Therefore, this paper summarizes some advances in deep learning in the field of cancer prognosis in recent years, and discusses the limitations of deep learning and possible future development directions. The research believes that the current limitations of deep learning mainly come from the characteristics of data sets and algorithms themselves, and proposes their own solutions to the problems.

**Keywords:** cancer prognosis, deep learning, cancer treatment, artificial intelligence, cancer prediction.

## 1. Introduction

Globally, cancer is one of the most difficult problems in the field of medicine for a long time. Common types of cancer include lung cancer, stomach cancer, colorectal cancer, and breast cancer. The existence of cancer cells will destroy the normal metabolism of the human body, lead to the decline of immunity, destroy the physiological function of specific organs, may lead to local bleeding and other symptoms of specific organs, patients will have loss of appetite, physical fatigue and other Mandarin, and eventually will die due to multiple organ failure. In recent years, cancer incidence and mortality rates have been increasing rapidly worldwide.

So far, the most common cancer treatments are surgery, chemotherapy and radiation. Conventional treatment has had obvious effects in cancer treatment, but it is accompanied by large side effects, high recurrence rate of cancer after treatment, the effect of treatment varies from person to person and other limitations in the treatment process. Cancer prognosis refers to predicting the likely course and outcome of a cancer patient. Accurate and reasonable prediction of the possible course and outcome of cancer patients can assist doctors to choose appropriate medical programs and evaluate the effectiveness of medical means, thus improving the survival rate of patients. The pathogenesis of cancer is complex, influenced by many factors such as age, gender, genetics and so on. People gradually realize that every cancer is different and has its particularity. The prognosis of cancer in different patients, even in different periods of one same patient is different, and different medical measures are needed. Therefore, the analysis and prediction of specific patients' characteristics have become more and more people's demands.

Artificial intelligence (AI) is developing rapidly in recent years and gradually enters the medical field, assisting recovery of more and more patients. At present, there are many research achievements about artificial intelligence in cancer. In the field of lung cancer, the artificial intelligence clinical decision-support system WFO can already participate in cancer treatment, with an accuracy rate of 92.4%, according to research [1]. In the field of gastric cancer, artificial intelligence also aids cancer diagnosis through the recognition of endoscopic images, and the diagnostic accuracy rate of AI using ME-NBI still images and videos is higher than that of endoscopists [2]. The application of AI in the field of cancer has broad prospects. Deep learning (DL) is the latest field of machine learning (ML) and the frontier of artificial intelligence research. So far, some machine learning methods, such as

support vector machines (SCM), semi-supervised learning and decision tree models, have been applied to the prediction and treatment of cancer, and have achieved good results [3]. According to the survey, ML algorithms significantly improved the accuracy of cancer recurrence rates and mortality predictions [3]. At present, DL is mainly used for cancer image recognition and diagnosis, cancer classification, cancer prediction and prognosis. The DL approach can make cancer treatment more personalized and differentiated, and has important implications for exploring specific treatments for different patients.

DL has important implications for cancer prognosis prediction, but there are few articles in this field. This paper mainly summarizes some advances in DL in the field of cancer prognosis, and discusses the limitations and future directions of DL.

## 2. Application of DL for the treatment of cancer

In the study of Liu et al. [4], they invented an algorithm named ProgCAE to predict cancer prognosis based on multi-omics data, and got good results. For the first time, the algorithm uses Convolutional autoencoders (CAEs) for the analysis of multi-omics molecular data, and introduces feature reordering to improve CNN training, as shown in Fig. 1. This allows ProgCAE to extract and exploit nonlinear relationships that are hidden in multi-omics data and represent important biological processes. The results show that the use of survival information enables ProgCAE algorithm to accurately predict the prognosis of cancer patients. All 12 of the cancer subtypes predicted by ProgCAE had significant survival differences compared to other models, and 10 of the 12 subtypes had higher hold-out C-index, significantly better than traditional prediction methods. If this algorithm is applied to clinical diagnosis, it can assist doctors to make medical decisions and improve the survival probability of patients.

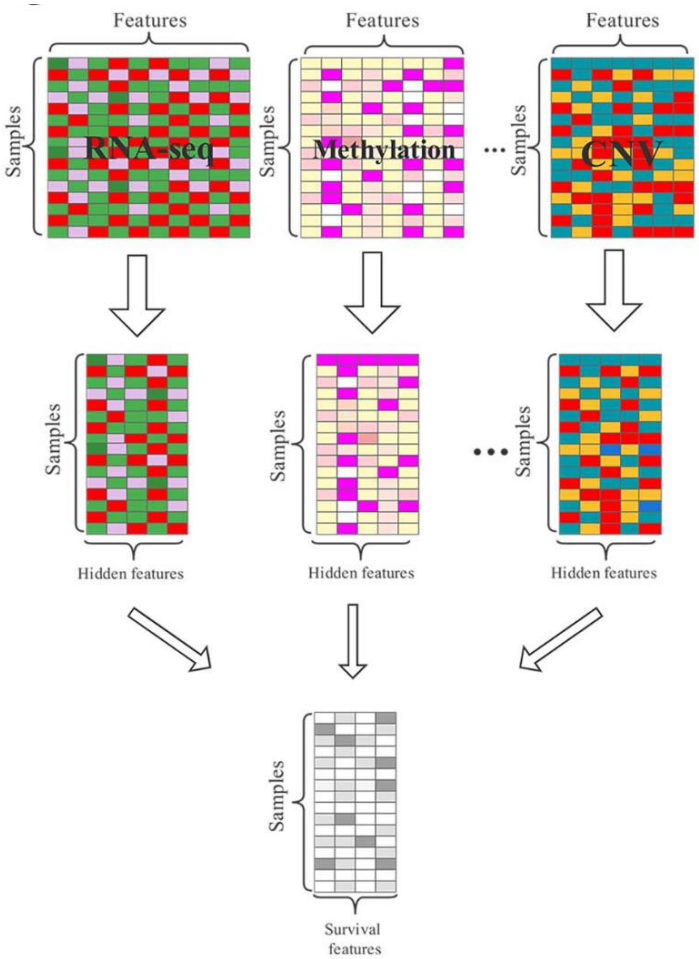



**Figure 1.** Process from the original to the final survival matrix [4].

According to Zhi et al. [5], they used three different DL-based cancer prognosis models to predict survival of patients with 12 types of TCGA cancer. The results show that the deep learning algorithm has better performance than the traditional ML model. In addition, the authors also found a significant negative correlation between tumor mutation burden (TMB) and the overall survival time of patients. TMB can be used as an important factor to predict cancer prognosis, providing a direction for future research on cancer prognosis. The author introduced the concordance index as an indicator to measure the effect of the model, and the results showed that the consistency index varied greatly among different cancers, and the model had the best predictive effect on the prognosis of renal papillary cell carcinoma (KIRP) and the worst predictive effect on lung squamous cell carcinoma (LUSC).

In the study of Pham et al. on rectal cancer [6], they proposed a new prognostic method for rectal cancer. They combine DL and wavelet scattering for feature extraction of IHC image data. The method captures the expression of RhoB protein in cancer tissue, which is closely related to the sensitivity of cancer cells to radiation therapy, to determine the degree of resistance of rectal cancer to radiation therapy. After experiments, the accuracy of the model where RhoB protein was included in the prognosis of rectal cancer was as high as 94%, much higher than that of the DL model with only 67% accuracy after pretreatment alone. This provides a broader idea and a more accurate method for the prognosis prediction of rectal cancer after radiotherapy.

Cheng et al. used DL models for the first time to predict the prognosis of patients with osteosarcoma [7], which until now had limited prognostic accuracy. They used the DeepSurv DL model, which can flexibly handle nonlinear data, taking into account age, primary location, tumor size, surgical method and other indicators, and predicted the results of the C-index of 0.800, higher than the traditional method of 0.774. This study, as the first application of DL models in the field of osteosarcoma, provides a more accurate and personalized prognosis prediction method for patients with this type of cancer, improving the probability of survival of patients.

Huang et al. introduced multi-omics data into the prognosis analysis of breast cancer [8]. They used a multi-omics analysis tool called SALMON to predict the survival of breast cancer patients, as shown in Fig. 2. Instead of using the original gene expression data, the model aggregated gene expression data and cancer biomarkers, and used the characteristic gene modules in the gene co-expression network analysis results as the model input, which improved the prediction performance. The results show that more omics data can improve the accuracy of breast cancer prognosis prediction, which provides a broader new idea for cancer prognosis prediction and integrated analysis based on DL. At the same time, the researchers also found that the main characteristics that control cancer prognosis change with age, which provides a more accurate and personalized prognosis.

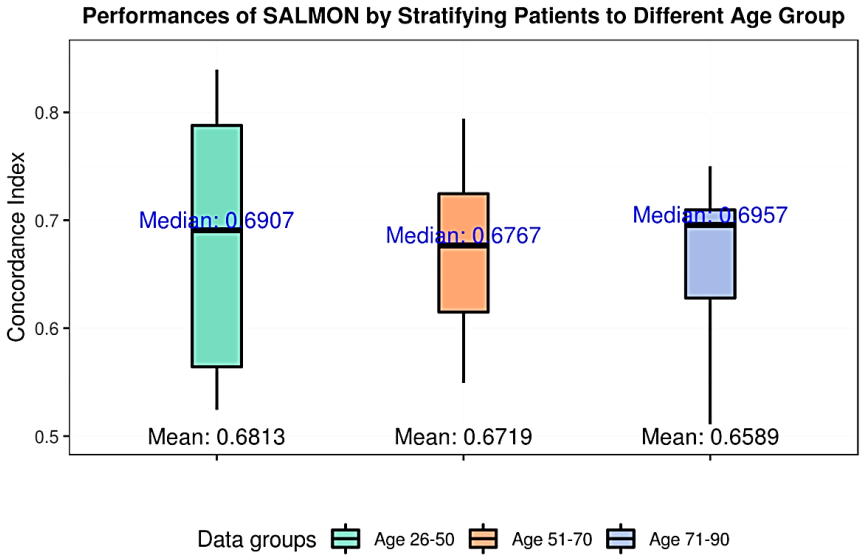



**Figure 2.** Performances of SALMON algorithm stratified by three age groups: 26–50 group; 51–70 group; 71–90 group with integrating all omics data (integration of mRNA, miRNA, CNB, TMB, diagnosis age, ER status, PR status) [8].

Xu et al. developed a weakly supervised learning dual-resolution DL model called WDRNet for prognostic prediction of colorectal cancer [9]. Most of the previous prediction methods only focus on local areas in histopathological images, but this method also takes global information into account, and the C-index of only local information is 5.9% lower than that of both global and local information, indicating that global information can improve the prediction performance. And compared to the other five prediction methods, WDRNet performed best on the two data sets used in the experiment. The researchers also stratified the patients according to the risk values predicted by the model, and the stratification results showed statistical significance in both univariate and multivariate analyses, which together demonstrated that the model performed better in predicting survival than a single high-resolution approach and could help patients choose a more appropriate individualized treatment plan.

## 3. Conclusion

DL plays an active role in cancer prognosis, mainly through recognition of images and omics data to make prognostic predictions. However, most of the current research is still at the theoretical level, and there is still a large gap from the clinical practical application, and many limitations hinder the further development of DL. First of all, the database of cancer patients is single and the amount of data is insufficient, so it is inevitable that some cases will be missed, leading to the deviation of prediction results. Moreover, there is no way to test the accuracy of results through multiple different data sets, and the extensibility of various models cannot be confirmed. Therefore, it is urgent to find more cancer cases to be included in the consideration of models. Create more databases for future generations to study. Image information may also be unavailable due to different lighting devices, different magnification sizes and coloring methods, so the integration of image data is also an important work in the future. Secondly, although the accuracy of DL is better than that of traditional methods, it is still not at the level of clinical use. The prediction accuracy of many models can be further improved. The prognosis is closely related to the patient's medical plan, so it is necessary to ensure the accuracy of model results, and more and more indicators should be included in the model's consideration. The accuracy of model prediction can be improved effectively by combining multiple omics data with each other, combining local information and global information, and combining image information and data. In addition, poor interpretability is also a reason for limiting the clinical application of deep learning. Interpretability is the basis of clinical application. In the future, the study of highly explanatory models is also an important task.

In the summary of this paper, DL has shown a good level of cancer prognosis, and it is also considered to have a good prospect in the future, and is expected to become an auxiliary tool for more and more doctors and patients in clinical treatment.

## References

[1] Zhang K, Chen K. Artificial intelligence: opportunities in lung cancer. Current Opinion in Oncology, 2022, 34 (1): 44 - 53.

[2] Horiuchi Y, Hirasawa T, Fujisaki J. Application of artificial intelligence for diagnosis of early gastric cancer based on magnifying endoscopy with narrow-band imaging. Clinical Endoscopy, 2024, 57 (1): 11.

[3] Bhatt M, Shende P. Advancement in machine learning: a strategic lookout from cancer identification to treatment. Archives of Computational Methods in Engineering, 2023, 30 (4): 2777 - 2792.

[4] Liu Q, Song K. ProgCAE: a deep learning-based method that integrates multi-omics data to predict cancer subtypes. Briefings in Bioinformatics, 2023: bbad196.

[5] Huang Z, Johnson T S, Han Z, et al. Deep learning-based cancer survival prognosis from RNA-seq data: approaches and evaluations. BMC medical genomics, 2020, 13: 1 - 12.

[6] Pham T D, Sun X F. Wavelet scattering of RhoB-expressed deep-learning features for rectal cancer prognosis. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). IEEE, 2023: 1 - 4.

[7] Cheng D, Liu D, Li X, et al. A deep learning model for accurately predicting cancer-specific survival in patients with primary bone sarcoma of the extremity: a population-based study. Clinical and Translational Oncology, 2023: 1 - 11.

[8] Huang Z, Zhan X, Xiang S, et al. SALMON: survival analysis learning with multi-omics neural networks on breast cancer. Frontiers in genetics, 2019, 10: 166.

[9] Xu Y, Jiang L, Chen W, et al. Computer-aided detection and prognosis of colorectal cancer on whole slide images using dual resolution deep learning. Journal of Cancer Research and Clinical Oncology, 2023, 149 (1): 91 - 101.