


Challenges of Deep Learning in Cancers

Yudong Zhang, PhD¹  and Jin Hong, PhD²

Received: March 17, 2023; Accepted: April 7, 2023.

Technology in Cancer Research & Treatment
Volume 22: 1-2
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15330338231173495
journals.sagepub.com/home/tct


Cancer is a significant crisis worldwide. Firstly, cancer is one of the leading sources of death worldwide. According to World Health Organization, cancer caused estimated 9.6 million deaths in 2018. Secondly, its treatment is costly,¹ and the corresponding care expenses continue to rise. This puts a weighty financial burden on individuals and families, as well as on healthcare systems and governments. Thirdly, cancer and its treatment drastically impact patients' quality of life. Treatments such as chemotherapy, radiation therapy,² and surgery cause physical and emotional side effects that are debilitating and long-lasting. Fourthly, not all cancer types are preventable. Many cases are connected to lifestyle factors, for example, smoking, alcohol consumption,³ poor diet, and so on. Finally, cancer research is a chief focus of public health efforts around the world. Advances in cancer diagnosis, drug treatment, and disease prevention lead to better survival rates, but there is much research to be studied about its pathology and treatment.

Cancer screening,⁴ diagnosis, prediction, survival rate estimation,⁵ treatment, and control measures are still the foremost challenges in the recent decade. With the development of biomedical imaging, inspection, and health management technologies, medical big data (MBD), such as biomedical images, omics,⁶ and clinical electronic medical records, are accumulating rapidly. How can we use this MBD to build better health records and more accurate prediction models to help disease diagnose and treat cancer better? In the past decade, the rapid development of machine learning (ML) methods has provided many successful cases to answer this fundamental question.

Commonly used ML methods for cancer analysis include linear regression, logistic regression,⁷ decision trees, random forests, support vector machines,⁸ neural networks, k-means clustering, principal component analysis,⁹ naïve Bayes, gradient boosting,¹⁰ and so on. However, ML methods suffer several shortcomings. Traditional ML models cannot predict outcomes with a high level of accuracy due to a lack of complexity in their underlying algorithms. Scalability is another shortcoming. The number of input variables in traditional ML

methods is often limited, making it difficult to build systems that are robust and scalable.

Recently, deep learning (DL)¹¹ has been the hottest ML method for cancer analysis nowadays. The reasons are 5 aspects. (i) DL can handle large and complex datasets with many variables, which is essential for massive amounts of genomic and clinical data. (ii) DL can detect patterns and identify correlations in cancer data that might not be visible to the human eye. (iii) DL can help predict how individual patients will respond to different treatments based on their genetic profile, medical history,¹² and other factors. (iv) DL models can be trained on large cancer datasets, such as mammograms¹³ or 3-dimensional computed tomography scans, to detect early signs of cancer that are invisible to the human eye. (v) DL can predict the effectiveness of new cancer drugs¹⁴ and identify potential drug targets based on genetic and other data.

Some top DL algorithms include convolutional neural networks, recurrent neural networks, autoencoders, generative adversarial networks,¹⁵ long short-term memory networks,¹⁶ capsule networks, and so on. There are many other variants of DL algorithms and combinations of these algorithms for various applications of cancer analysis.¹⁷

Although ML and DL have shown great potential in helping cancer screening, diagnosis, prognosis, and treatment, many significant data-related problems still hinder the application of ML and DL in cancer analysis. These problems include excessive noise, labeled data deficiency, heterogeneous data, unbalanced data, multisource domain data,¹⁸ and data isolation.

¹ School of Computing and Mathematical Sciences, University of Leicester, Leicester, UK

² Brain Information and Human Factors Engineering Laboratory, Zhongshan Institute of Changchun University of Science and Technology, Zhongshan, China

Corresponding Author:

Yudong Zhang, School of Computing and Mathematical Sciences, University of Leicester, Leicester, LE1 7RH, UK.

Email: yudongzhang@ieee.org



Interpretability and validation are 2 other issues. Most ML and DL models are often referred to as black boxes because it is difficult to interpret how they arrive at their predictions. In cancer analysis, interpretability is crucial to understanding how the ML and DL models arrive at their decisions and gaining insights into the disease's biological mechanism. Meanwhile, ML and DL models can suffer from overfitting, where the model learns to perform well on the training data but performs poorly on new, unseen data. Valuing ML and DL models in cancer analysis is challenging when the data is limited or biased.

To better apply ML and DL models to cancer analysis, it is important to develop novel data preprocessing¹⁹ methods, data representation learning methods,²⁰ and novel efficient ML and DL models.

Ethics Statement

Not applicable. Our study did not require ethical board approval because it is an editorial.


Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This paper is partially supported by MRC, UK (MC_PC_17171); Royal Society, UK (RP202G0230); BHF, UK (AA/18/3/34220); Hope Foundation for Cancer Research, UK (RM60G0680); GCRF, UK (P202PF11); Sino-UK Industrial Fund, UK (RP202G0289); LIAS, UK (P202ED10, P202RE969); Data Science Enhancement Fund, UK (P202RE237); Fight for Sight, UK (24NN201); Sino-UK Education Fund, UK (OP202006); BBSRC, UK (RM32G0178B8).

ORCID iD

Yudong Zhang  <https://orcid.org/0000-0002-4870-1493>

References

- Singh KR, Agarwal G. Prioritizing the expenses of breast cancer treatment makes sense—not just in developing countries, but across the globe. *World J Surg*. 2014;38(8):2187-2188.
- Cappon DJ, Fang S, Berry K, et al. Clinical best practices for radiation safety during lutetium-177 therapy. *Health Phys*. 2023;124(2):139-146.
- Gaspari M, Agrela-Romero B. Migrations, ethnified communities and alcohol consumption: from “ecuadorianized” drinking in Genoa. *Migraciones*. 2022;56:1-22.
- Michaels E, Worthington RO, Rusiecki J. Breast cancer risk assessment, screening, and primary prevention. *Med Clin N Am*. 2023;107(2):271-284.
- Iwai N. Estimation of growth rate, survival rate, and longevity of the endangered Otton frog (*Babina subaspera*) using the capture-mark-recapture method. *Herpetol Conserv Biol*. 2022;17(3):548-552.
- Patkulkar PA, Subbalakshmi AR, Jolly MK, et al. Mapping spatio-temporal heterogeneity in tumor profiles by integrating high-throughput imaging and omics analysis. *ACS Omega*. 2023;8(7):6126-6138.
- Avila FR, Spaulding AC, Rinker BD, et al. Demographic characteristics influence treatment costs of invasive melanoma in Florida. *Ann Plast Surg*. 2023;90(3):248-254.
- Anupong W, Muda I, AbdulAmeer SA, et al. Energy consumption and carbon dioxide production optimization in an educational building using the supported vector machine and ant colony system. *Sustainability*. 2023;15(4):3118. doi:10.3390/su15043118
- Zagkos L, Dib MJ, Pinto R, et al. Associations of genetically predicted fatty acid levels across the phenome: a Mendelian randomisation study. *PLoS Med*. 2022;19(12):e1004141. doi:10.1371/journal.pmed.1004141
- Hagiwara Y, Shiroya T, Taira N, et al. Gradient boosted tree approaches for mapping European organization for research and treatment of cancer quality of life questionnaire core 30 onto 5-level version of eq-5d index for patients with cancer. *Value Health*. 2023;26(2):269-279.
- Hong J. Brain age prediction of children using routine brain MR images via deep learning. *Front Neurol*. 2020;11:584682.
- Al-Husban A, Abdulridha MM, Mohamad AAH, et al. Biocomposite's multiple uses for a new approach in the diagnosis of Parkinson's disease using a machine learning algorithm. *Adsorpt Sci Technol*. 2022;2022:Article ID 615932. doi:10.1155/2022/615932
- Tiryaki VM, Kaplanoglu V. Deep learning-based multi-label tissue segmentation and density assessment from mammograms. *IRBM*. 2022;43(6):538-548.
- Matthews JM, Schuster B, Kashaf SS, et al. OrganoID: a versatile deep learning platform for tracking and analysis of single-organoid dynamics. *PLoS Comput Biol*. 2022;18(11):e1010584. doi:10.1371/journal.pcbi.1010584
- Al-Rasheed A, Ksibi A, Ayadi M, et al. An ensemble of transfer learning models for the prediction of skin cancers with conditional generative adversarial networks. *Diagnostics*. 2022;12(12):3145. doi:10.3390/diagnostics12123145
- Raguvaran S, Anandamurugan S, Rahman A. Harnessing lstm classifier to suggest nutrition diet for cancer patients. *Intell Autom Soft Comput*. 2022;35(2):2171-2187. doi:10.32604/iasc.2023.028605
- Hong J, Yu SCH, Chen WT. Unsupervised domain adaptation for cross-modality liver segmentation via joint adversarial learning and self-learning. *Appl Soft Comput*. 2022;121:108729. doi:10.1016/j.asoc.2022.108729
- Hong J, Zhang YD, Chen WT. Source-free unsupervised domain adaptation for cross-modality abdominal multi-organ segmentation. *Knowl Based Syst*. 2022;250:109155. doi:10.1016/j.knsys.2022.109155
- Wang S. Advances in data preprocessing for biomedical data fusion: an overview of the methods, challenges, and prospects. *Inf Fusion*. 2021;76:376-421.
- Zhang Y-D, Dong Z-C. Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation. *Inf Fusion*. 2020;64:149-187.