

Deep learning: shaping the medicine of tomorrow

Konstantinos Vougas, Spyridon Almpanis & Vassilis Gorgoulis

To cite this article: Konstantinos Vougas, Spyridon Almpanis & Vassilis Gorgoulis (2020): Deep learning: shaping the medicine of tomorrow, Molecular & Cellular Oncology, DOI: [10.1080/23723556.2020.1723462](https://doi.org/10.1080/23723556.2020.1723462)

To link to this article: <https://doi.org/10.1080/23723556.2020.1723462>



Published online: 19 Feb 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

COMMENTARY



Deep learning: shaping the medicine of tomorrow

Konstantinos Vougas ^{a,b}, Spyridon Almpanis ^a, and Vassilis Gorgoulis ^{a,b,c,d}

^aMolecular Carcinogenesis Group, Department of Histology and Embryology, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece; ^bBiomedical Research Foundation of the Academy of Athens, Athens, Greece; ^cDivision of Cancer Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, Manchester Cancer Research Centre, NIHR Manchester Biomedical Research Centre, University of Manchester, Manchester, UK; ^dCentre for New Biotechnologies and Precision Medicine, Medical School, National and Kapodistrian University of Athens, Athens, Greece

ABSTRACT

Predicting response to therapy is a major challenge in medicine. Machine learning algorithms are promising tools for assisting this aim. Amongst them, Deep Neural Networks are emerging as the most capable of interrogating across multiple data types. Their further development will lead to sophisticated knowledge extraction, shaping the medicine of tomorrow.

ARTICLE HISTORY

Received 14 January 2020
Revised 23 January 2020
Accepted 24 January 2020

KEYWORDS

Personalized medicine;
cancer treatment; drug
response prediction;
metastasis detection;
machine learning; deep
neural networks (DNN)

The rapid development of high-throughput technologies such as Next Generation Sequencing, as well as the digitization of already existing diagnostic procedures like digital-radiology and digital-pathology, along with the global adoption of Electronic Health Records (EHR), provide a big-data landscape for the utilization of machine learning (ML) frameworks for accelerating the advent of personalized medicine. This landscape is however far from ideal, partly due to lack of large patient cohorts with detailed follow-up and full molecular profiling and partly due to lack of preclinical models assembled in large panels able to adequately recapitulate disease complexity in the clinic. Although such panels (e.g. Genomics of Drug Sensitivity in Cancer (GDSC) cancer cell-line pharmacogenomics database) have already been made available to the community¹ it has been demonstrated that they need to be further expanded, combined and processed by multiple data-mining frameworks to maximize knowledge extraction and clinical relevance.²

On a different note, it is widely accepted throughout the scientific community that the driving force of carcinogenesis is genomic instability operating on a patient's most unique characteristic, his/her genome,^{3,4} making cancer the most 'personal' among diseases. As such, a major hurdle in oncology is our current inability to predict the individual patient's response to the selected therapeutic strategy, since the latter is defined based on crude clinical characteristics along with a limited number of molecular biomarkers, which however cannot recapitulate the aforementioned complexity. As a result, patients do not receive always efficient treatments, translating in decreased overall survival and quality of life. In a proof-of-concept study Geeleher et al.,⁵ demonstrated that ML frameworks could be utilized for predicting clinical

drug response from baseline gene expression, while trained on *in-vitro* cell-line response data. However, Deep Neural Networks (DNNs), which have been shown to deliver "state-of-the-art" (SOTA) performance in a wide variety of tasks, were not part of this study. Hence, their value in that particular task was yet to be demonstrated.

The study published by our group bridged that void by utilizing the GDSC cell-line pharmacogenomic database to train DNNs to predict drug response from baseline gene expression and validated their predictive performance on an extended set of publically available clinical cohorts as well as on previously unpublished ones.⁶ Although the number of training cases (1001 cell-lines) was very low for an ML framework such as DNNs, which are known to require very large training sets for their full potential to be unleashed, and despite the dimensionality curse⁷ present in this training set, not favoring such complex models that tend to overfit on small training sets, they were surprisingly shown to generalize better in clinical cohorts than other SOTA ML frameworks, thus delivering superior predictive performance. A significant DNNs drawback, apart from the fact that they require large training sets and large computational resources to be trained, is their 'black-box' nature in the sense that the intrinsic multi-dimensional, non-linear relationships learned by the model across their hidden layers are extremely hard to interpret.⁸ Toward that end, our study presented a knowledge-extraction strategy for mining learned gene-importance toward drug-response prediction. Consequent pathway enrichment analysis of the extracted genes provided insight on the molecular mechanisms that drive drug-response, confirming already existing knowledge as well as revealing novel pathways as key-players in the response-to-therapy mechanisms.

Future plans for carrying this work forward include (a) increasing the cell-line training-set size and dimensionality as well as utilizing more ML frameworks for comparison (linear models, Decision Trees, Support Vector Machines, and Bayesian Classifiers) in order to unambiguously establish DNN superiority, (b) training DNNs to model combinatorial therapeutic schemes, on actual clinical cohorts and finally (c) integrate on the latter, information from alternate data sources such as histopathological whole slide images (WSIs). Regarding the first goal, since DNNs have demonstrated the ability to learn meaningful biological information from cell-lines that generalizes to the clinic, we need to push this boundary further by increasing the number of available cell lines as well as the information we feed the networks. The DNNs of our study used only gene-expression as input. Including mutation and methylation status, as well as proteomic and metabolomic information will boost toward this direction. Additionally, the *in-vitro* drug response data that we utilized originated from simple 2D-cultures, which are sub-optimal for recapitulating the tumor microenvironment that plays a pivotal role to therapeutic response.⁹ Hence, it is obvious that the response data have to be enhanced through organotypic cultures or *in-vivo* dose–response experiments through xenografts in mice. As to the second goal, large clinical cohorts with complete follow-up and baseline tumor molecular profiling, having received specific therapeutic combinatorial schemes, need to be identified and used as training-sets to train combination therapy-specific models. In respect to the last goal, our group has developed know-how on applying specialized image processing DNNs, namely Convolutional-Neural-Network (CNNs), Convolutional-Autoencoders (CAEs) and Generative-Adversarial-Networks (GANs) on histopathology WSIs. Specifically, our submitted solution in the Camelyon-17 contest achieved the fifth best score at the time of submission (Cohen's k of 0.9052).¹⁰ Camelyon-17 is a global contest for evaluating the performance of Artificial Intelligence-powered algorithms for automated detection of metastasis through analysis of digitized histopathological lymph nodes sections from breast cancer patients and consequent patient-level N-staging classification. Through the application of such networks, feature embeddings corresponding to tissue architecture will be extracted and integrated along with features describing the tumor molecular profile. Such an integration will allow us to develop even more powerful models for delivering personalized therapy to the clinic.

Abbreviations

AI	Artificial Intelligence
CAE	Convolutional Autoencoders
CNN	Convolutional Neural Network (networks used for image processing)
CNV	Copy Number Variation
DNN	Deep Neural Network
HER	Electronic Health Records
GAN	Generative Adversarial Networks

LN	Lymph Nodes
ML	Machine Learning
SOTA	State of the art
WSI	Whole Slide Images

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Funding

Financial support was from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grants agreement No. 722729 (SYNTRAIN); the Welfare Foundation for Social & Cultural Sciences (KIKPE), Greece; Pentagon Biotechnology Ltd, UK; DeepMed IO Ltd, UK, NKUA-SARG grants No 70/3/9816, 70/3/12128.

ORCID

Konstantinos Vougas  <http://orcid.org/0000-0002-8873-4740>
 Spyridon Almpanis  <http://orcid.org/0000-0001-7003-735X>
 Vassilis Gorgoulis  <http://orcid.org/0000-0001-9001-4112>

References

1. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, Aben N, Gonçalves E, Barthorpe S, Lightfoot H, et al. A landscape of pharmacogenomic interactions in cancer. *Cell*. 2016;166(3):740–754. doi:10.1016/j.cell.2016.06.017.
2. Vougas K, Sakellaropoulos T, Kotsinas A, Foukas G-RP, Ntargaras A, Koinis F, Polyzos A, Myrianthopoulos V, Zhou H, Narang S, et al. Machine learning and data mining frameworks for predicting drug response in cancer: an overview and a novel in silico screening process based on association rule mining. *Pharmacol Ther*. 2019;203:107395. doi:10.1016/j.pharmthera.2019.107395.
3. Halazonetis TD, Gorgoulis VG, Bartek J. An oncogene-induced DNA damage model for cancer development. *Science*. 2008;319(5868):1352–1355. doi:10.1126/science.1140735.
4. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability—an evolving hallmark of cancer. *Nat Rev Mol Cell Biol*. 2010;11(3):220–228. doi:10.1038/nrm2858.
5. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol*. 2014;15(3):R47. doi:10.1186/gb-2014-15-3-r47.
6. Sakellaropoulos T, Vougas K, Narang S, Koinis F, Kotsinas A, Polyzos A, Moss T, Piha-Paul S, Zhou H, Kardala E, et al. A deep learning framework for predicting response to therapy in cancer. *Cell Rep*. 2019;29(11):3367–3373.e4. doi:10.1016/j.celrep.2019.11.017.
7. Bellman RE. *Dynamic programming*. princeton landmarks in mathematics. Princeton (NJ): Princeton University Press; 1957.
8. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*. 2019;51:93. doi:10.1145/3236009.
9. Hirata E, Sahai E. Tumor microenvironment and differential responses to therapy. *Cold Spring Harb Perspect Med*. 2017;7(7):pii: a026781. doi:10.1101/cshperspect.a026781.
10. Vasileiou I, Tagkalakis F, Diamandis A, Mitsianis E, Mavropoulos E, Mallios D, Agrogiannis G, Pateras I, Evangelou K, Joseph L, et al. Breast cancer pN staging with deep learning. CAMELYON-17 submission. <https://camelyon17.grand-challenge.org/evaluation/results/>