"Interpretable Deep Learning for Cancer Research"

NCI Division of Cancer Biology, September 20-21st

Executive Summary of the Workshop

Workshop Motivation and Organization

The previous five years have seen a paradigm-shifting ascendance of artificial neural network-based methods in computer vision, imaging, and natural language processing – collectively and popularly called "Deep Learning." A key methodology within the broader fields of artificial intelligence and machine learning, deep learning approaches have been rapidly adopted for use in health records mining and in cancer imaging, however their use by the broader cancer research community has been less marked. This can be attributed, at least in part, to a limitation frequently referred to as the "black box," where deep learning models can make correct predictions but without direct association to underlying mechanisms, hindering biological interpretation. This workshop brought together deep learning researchers, cancer systems biologists, and computational biologists to discuss challenges and opportunities to developing interpretable deep learning methods that can be applied to cancer biology investigations in a manner enabling biological interpretation and knowledge generation.

For this workshop, the NCI and co-chairs, Dr. Trey Ideker of UCSD and Dr. Claire Tomlin of UC Berkeley, identified a range of participants (attendee list and agenda attached). Prior to the workshop, attendees responded to a series of questions designed to scope the challenges and opportunities for using deep learning more broadly in cancer research. Based on their input, the two-day workshop was structured into five discussion sessions focused on topics that address the following questions:

1. How do we deal with the insufficient quantity and/or quality of data that is paramount for the success of deep learning model development in computer vision?
2. How do we transfer successful deep learning methods developed in other fields into cancer research? For example, AlphaGo's reinforcement learning is entirely rule-based without training using existing data. Can we use this approach to model biological problems?
3. How do we transfer biological domain knowledge into deep learning method development?
4. What can we learn from successful deep learning models? Why is each successful?
5. What are the requirements for model validation and model comparison?

Across two days of discussions and presentations, some common themes emerged regarding future directions and needs in deep learning development to make the method more useful in cancer research, summarized below in two general categories:

Interpretable Deep Learning – Methods Development Needs

- Making deep learning methods interpretable is key to making them more broadly useful in cancer research:
  - Deep learning model quality must be improved and validated for cancer research questions.
  - There is then a need to determine *why* a validated model happens to be successful in making correct associations and predictions. This will require interpretable models.

- To make the models more impactful, stronger connections to real meaningful biological/clinical problems in cancer research need to be made, so that models can be used to directly testing hypothesis and making new discoveries.

- Creative ways of integrating existing domain knowledge into model development are needed. While it has been accomplished in other fields, advancements in the cancer research space are limited. Some examples were discussed at the workshop including: 1) the design of a model topology that is biological network informed, 2) insertion of explicit mathematical models or other methods as components of deep learning models, 3) associating specific deep learning parameters with biological features/measurements. Additionally, more research is needed in integrating causal modeling with modern machine learning techniques.

- More research is needed for testing new deep learning methods that are successful in other fields (e.g. reinforcement learning, transfer learning, adversarial networks) to determine if they are transferable to modeling biological problems.

Interpretable Deep Learning – Expertise and Resource Needs

- NCI needs to develop domain-specific expertise in deep learning and machine learning, instead of outsourcing to the larger computer science community.
- NCI should establish a clear definition of "interpretable deep learning" and share that with the community.
- Well curated, well labelled, and benchmarked data sets with sufficient quantity are critically needed in all areas of cancer research. These resources need to be continually updated and expanded. Currently, public release of data in usable/curated form is under-incentivized.
- NCI should enhance model/code sharing and encourage open source development in machine learning methods development. Preprint servers such as bioRxiv should be adopted more broadly.
- Industry partnerships will remain important:
    - NIH may be well-poised to incentivize commercialization of biomedical deep learning models through existing avenues, such as SBIR grants and technology transfer efforts.
    - Industry remains data-limited, so the biomedical community could incentivize industry participation by making data available.
- Community organized challenges such as the DREAM Challenge have piloted existing and robust frameworks for sharing both data and software. These efforts need to be encouraged.