

Preview

Shattering cancer with quantum machine learning: A preview

Joseph Geraci^{1,2,3,*}¹Department of Molecular Medicine, Queen's University, Kingston, ON, Canada²Centre for Biotechnology and Genomics Medicine, Medical College of Georgia, Augusta, GA, USA³NetraMark Corp, Toronto, ON, Canada*Correspondence: geracij@queensu.ca<https://doi.org/10.1016/j.patter.2021.100281>

Machine learning has become a standard tool for medical researchers attempting to model disease in various ways, including building models to predict response to medications, classifying disease subtypes, and discovering new therapies. In this preview, we review a paper that utilizes quantum computation in order to tackle a critical issue that exists with medical datasets: they are small, in that they contain few samples. The authors' work demonstrates the possibility that these quantum-based methods may provide an advantage for small datasets and thus have a real impact for medical researchers in the future.

Efforts to analyze genomic data utilizing machine learning have been underway for over a decade, and it is in oncology where these efforts have matured the furthest. The various faults of the molecular machinery present in cancer patients are exhausting to consider, but the use of algorithms, including machine learning, has helped the field begin to usher in precision medicine. In precision medicine the goal is, ultimately, to provide therapeutics based on data that are specific to an individual or class of individuals. Now, we are in a position to consider exciting alternative computing paradigms, specifically quantum computing, and the various ways it can help advance how we approach a deeper understanding of cancer and other complex diseases. Quantum computing is still at a nascent stage of development, but exciting advancements are being made. Several quantum-computing platforms are available now, and Python-based packages exist such as Qiskit (<https://qiskit.org/>), which allows programmers to create quantum circuits and run them on hardware or simulators. Circuit-based platforms such as the IBM Quantum Network, Rigetti, and IonQ run Qiskit. There are multiple approaches currently in play to create hardware to utilize quantum resources for computation. Xanadu is a quantum-computing platform that operates with light. Alternatively there is DWave, which utilizes what we refer to as quantum annealing, which is quite different from the aforementioned circuit model systems.

In this issue of *Patterns*, Li et al.¹ analyze cancer-based genetic data using several machine learning algorithms, including a novel approach on the DWave annealer.

Li et al. report on work they did comparing classical machine learning methods with some quantum-inspired methods that they refer to as Ising-type approaches. Ising models are a simple way to model complex interactive systems, like atoms connected to each other in a magnet.² The purpose of these models is to study phase transitions, and they can be used to model many phenomena. In Ising models, objects called spins are connected to each other with edges, and one can define the strength of the influence between spins as weights on the edges. The idea of tension plays a major role in these models, and one can think of moviegoers trying to find seats that are not only unoccupied but also not directly next to a stranger. The goal of minimizing tension by avoiding adjacent seat occupancy results in certain seating configurations to be more likely. In a similar way, Ising models encode a large number of configurations and their associated "tension" or "energy" values. Some configurations are relaxed, and others are "frustrated," or in other words "really tense," like being surrounded by moviegoers who speak during a film. Now, it turns out that one can encode information on these Ising models, and by minimizing different notions of tension or energy, these systems can be used to learn about data. Fortunately, DWave's

quantum-annealing hardware is a natural place to minimize these Ising energy functions. Thus, if we encode data and their labels onto an Ising model, then we can use this quantum system to learn the necessary weights to make predictions with this kind of model. In essence, we can use the quantum hardware to train an Ising-based model to make predictions.

These authors considered a collection of Ising-type models, one of which utilized the DWave quantum system. A popular Ising-type approach is referred to as a restricted Boltzmann machine (RBM), and the training for this kind of system can be done classically, which these authors include in their collection of methods. In Jain et al.,³ an RBM was trained by utilizing DWave's quantum annealer by taking advantage of the aforementioned annealing hardware, but in Li et al.,¹ they trained an RBM classically, in addition to other variations of Ising-type algorithms. However, they did utilize DWave for one Ising-type method and also included a popular set of classical methods including support vector machines, naive Bayes, random forest, and regression methods. They compared ten classifiers (five standard machine learning methods, one quantum Ising-type algorithm, and four classical Ising-type algorithms) with five binomial datasets and one six-cancer multi-class dataset, which included cancer data from the brain, breast, kidney, lung, liver, and colorectal. What they found has interesting implications for machine learning applied to small



data, which is currently a challenge. The standard set of machine learning algorithms outperformed the Ising-type models for the larger datasets, but there was superior reproducibility for models trained on smaller data when Ising-type models were applied. This is the punchline. Impressively, the Ising-type models were able to perform nearly as well as standard machine learning methods for various datasets, but when the number of samples are reduced, it turns out that these Ising-type algorithms are better able to avoid overfitting and thus performed better.

In medicine, the cost of obtaining a lot of data that integrate genetic, epigenetic, microbiome, and clinical data is high. However, the value of such an integrated dataset is obvious, because it will be through these kinds of data that a true understanding of patient populations will emerge. The current taxonomy of disease needs to be replaced by one that is driven by models that can elucidate the complex and varied ways that disease can manifest, despite looking similar from a clinical perspective. It is clear that the number of samples that can be brought together through an effort like this is expensive, which highlights the importance of the work done in Li et al.¹ In our opinion, the following characteristics will be critical

for the next generation of machine learning methods for medicine:

- the ability to create robust models from datasets with a relatively small number of samples despite having a large number of variables
- explainability so that the driving variables are clearly provided to medical experts
- unsupervised hypothesis generation surrounding patient population heterogeneity

As an example, in Qorri et al.,⁴ the authors used a novel methodology that encompasses these characteristics in order to extract the driving etiologies behind subpopulations of Alzheimer's patients. This, and the article previewed here, makes it clear that not only is it worth exploring novel machine learning approaches but that quantum computing can potentially provide major opportunities to process datasets like never before. Another example is provided in Abbas et al.,⁵ where the authors show that quantum neural networks may one day be superior to classical neural networks in terms of how quickly reliable models can be trained and how well these methods can *shatter* datasets, that is, how well they can classify data. It is now

clearer than ever that, as quantum computation matures, the opportunity to impact the medical sciences through superior analytical techniques is something worth pursuing, despite the inherent uncertainties.

DECLARATION OF INTERESTS

J.G. is a founder and major shareholder of NetraMark Corp, a company working with medical data for the purpose of optimizing drug discovery and development.

REFERENCES

1. Li, R.Y., Gujja, S., Bajaj, S.R., Gamel, O.E., Cilfone, N., Gulcher, J.R., Lidar, D.A., and Chittenden, T.W. (2021). Quantum processor-inspired machine learning in the biomedical sciences. *Patterns* 2. this issue, 100246-1–100246-13. <https://doi.org/10.1016/j.patter.2021.100246>.
2. Barry, A. (1987). Cibra, An Introduction to the Ising Model. *Am. Math. Mon.* 94, 937–959.
3. Jain, S., Ziauddin, J., Leonchik, P., Yenkanchi, S., and Geraci, J. (2020). Quantum and classical machine learning for the classification of non-small-cell lung cancer patients. *SN Applied Sciences* 2, 1088. <https://doi.org/10.1007/s42452-020-2847-4>.
4. Qorri, B., Tsay, M., Agrawal, A., Au, R., and Geraci, J. (2020). Using machine intelligence to uncover Alzheimer's disease progression heterogeneity. *Explor Med.* 1, 377–395. <https://doi.org/10.37349/emed.2020.00026>.
5. Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., and Woerner, S. (2020). The power of quantum neural networks. *arXiv*, 2011.00027 <https://arxiv.org/abs/2011.00027>.