Plant Breeding Biomolecular Classification in Quantum Bayesianism (QBism) Physics-Informed Neural Network Architecture

Barbara S. Keary[1] and Karriem A. J. Perry[2]*

[1]*Emory University, USA*

[2]*Capitol Technology University, USA*

*****Corresponding Author**
Karriem A. J. Perry, Capitol Technology University, USA.

Disclosures: The research, findings, and opinions expressed in this manuscript are solely those of the authors.

Plant science, in particular plant breeding, with its emphasis on gene manipulation thereby containing extraordinarily large datasets, can benefit from greater penetration of Machine-Learning (ML) and Deep Learning (DL) tools for solutions to problems from multi-classification (M-class) and optimization to anomaly detection, to time-series analysis and forecasting [1,2]. Deep learning (DL) contributes to plant science by simulating biochemical behaviors and assisting with the spatial analysis of plant evolution. A majority of the improvements made in machine learning (ML) and deep learning (DL) are not fully explored by disciplines outside of computer science, the physical sciences, engineering, and mathematics [3]. Plant science will advance more rapidly if it makes better use of related advances in ML, DL, and other constructs to solve the growing complexity of challenges in food security [4,5]. Of particular utility is physics-informed ML; physics-informed, or scientific ML, analyzes and interprets scientific datasets using novel methods drawn from both ML and the advanced mathematics field known as scientific computing. These techniques are critical for the next set of advances in data-heavy scientific disciplines, including plant breeding.

This research communication specifically posits that gene editing for plant breeding has yet to begin exploration into physics-informed ML. The intent of this short communication is to promote further interests in the utility ML modelling can offer to plant sciences. As ML modelling performance improves and other aspects of computation become available, including as quantum computing continues to develop, the probability of computational intractability will decrease, as will the likelihood of adverse computational environmental effects. If these factors reach equilibrium, the potential for more effective plant breeding experiments using ML modelling will significantly benefit both fields of research.

Leveraging the full capacity of AI techniques and connecting somewhat disparate forms of mathematics to accelerate scientific discovery in discipline-dependent domains requires a comprehensive and incisive understanding of AI constraints and capabilities, such as the vastly different approaches used in model selection and tuning effectiveness of robotics as compared to natural language processing model development and physical artificial intelligence. ML model selection is critically important in scientific computing; when it is done incorrectly, computations will not properly resolve, making what was a problem of intractability likely an impossibility. For example, the uniformed loss function of Sum of Squared Errors (SSE) results in computation and environmental energy waste and algorithm non-convergence [6]. Failure to identify anomalous environments and adapt model parameters in the simplest design, a feed-forward neural network, can result in a chain-reaction of conditions which render modeling vulnerable to irregularities common as Type-I (false positive) or Type-II (false negatives) in confusion matrices; they also can include the more challenging-to-detect stationarity issues associated with time-series or streaming datatypes [7].

Classical DL, defined as inference calculated on a classical information system, has been used in plant science already to help draw conclusions about yield. However, quantum deep learning informed by quantum Bayesianism (QBism), has not been fully exploited by the plant sciences [8]. QBism is a philosophical interpretation of quantum mechanics wherein the perspective of the observer is considered quantum deep learning is an approach used in machine learning model architecture [9,10]. Quantum deep learning informed by QBism permits the user to perform calculations on a quantum system from the Bayesian perspective. This helps overcome the computational intractability of large datasets, including those found in plant science.

No two ML models behave the same and this exclusivity is amplified when considering the application of relational

quantum ML methods, a derivation of scientific ML, directed at biomolecular processes. Developing biology-centric scientific functions in statistical and multi-purpose of programming languages, R and Python respectively [11,12]. This is because the datasets are so large and the data in question are so complex, the task of initial exploratory data analysis' time intensity outweighs the benefit of model development. Future research will analyze a derivation of a physics-informed M-class intelligent agent designated as a Relational Quantum-Informed Neural Network or RQINN.

Classical neural networks may encounter datatypes causing computational intractability or conditions comparable to those seen in currently insoluble problems from relatively common anomalies such as noisy datasets, non-static time series, missing values in streaming datasets and biomolecular processes containing robust molecules composed of (> 15) non-hydrogen atoms. Therefore, physics-informed ML provides tools unavailable in current ML techniques. Considering plant breeding as the end-goal, the initial challenge may include deciding the most utilitarian physical approach to employ, for instance chemical versus biological and establishing the applicable partial differential equation (PDE), which is part of physics-informed ML.

These potential RQINN exploits will positively impact the trajectory of a number of crucial areas in molecular plant science; as described by these areas include, but are not limited to, genome editing in plant breeding; plant-based edible vaccines, RNA interference biotechnology, sustainable agriculture supporting food security, and many other areas. These advances provide enhanced granularity by describing some of the emerging technologies, or new plant breeding technology, associated with the plant sciences.

The technology associated with the physical and biological sciences has progressed concurrently with classical ML evolutions but not quantum ML [13]. A vital method in evaluating and understanding physical traits associated with a given species is the use of the breeder's equation [14]. As with a majority of PDEs, solving for an unknown function, for instance $ff \ iiii \ f(xx)$, is dependent on myriad factors not initially identifiable [15]. Therein lies the importance in knowledge of the structure of those data and how relations among metadata, example include nodes and the composition of entities, objects, elements [14]. Moreover, due to the complexity of biological systems, these domains often provide favorable test-beds for often intractable M-class from a classical optic. Considering the QBism perspective, we can to update our understanding of the breeder's equation as more data are made available [16]. This update of data does not interfere with the initial data quality and potentially enhances the classifier at convergence or the understanding of the observer.

It is questionable that conducting biological computation renders the operation inconsequential due to computational cost. This postulation is non-linear and should be considered subjective without the presence of a comprehensive problem statement and successive hypotheses development, a necessary exercise in preparation for discretization of experimental data. Moreover,

this inherent complexity accompanying biological data effects both binary and M-class problems. Let us briefly consider a utilitarian aspect of the breeder's equation, where ZZ represents the central limit of an unidentified trait of a species population. From a PDE point of observation, this is equivalent to some unknown function $f(xx)$. Moreover, $\Delta ZZ = h2SS$ denotes the breeder's equation rate of change over time and should indicate to plant scientists or ML engineers that time series analysis and forecasting may produce a useful modeling alternative or exists as an additional tool to aid in the understanding and manipulation of biological data. Time series analysis can help to identify trends and patterns in the data, which can inform breeding decisions and improve prediction accuracy. Additionally, machine learning techniques such as neural networks and decision trees can be applied to biological data to improve classification accuracy and aid in the discovery of new knowledge. However, it is important to note that the application of these techniques requires careful consideration of the unique characteristics and complexities of biological data, as well as appropriate data preprocessing and feature selection technique.

One of the main challenges in plant science and plant breeding is the large and complex datasets that need to be analyzed. Machine learning (ML) and deep learning (DL) techniques have shown promise in handling such data and have the potential to significantly advance plant science research. These techniques can assist in solving complex problems, such as multi-classification and optimization, anomaly detection, time-series analysis, and forecasting. Furthermore, physics-informed ML is a useful tool that can provide additional capabilities to handle data-heavy scientific disciplines, including plant breeding.

A critical aspects of ML is model selection, which plays a crucial role in scientific computing.

Selecting the wrong model can result in computations that do not resolve, rendering the problem intractable or even impossible to solve. For example, the uniform loss function of Sum of Squared Errors (SSE) can result in computation and environmental energy waste and algorithm nonconvergence.

Additionally, failure to identify anomalous environments and adapt model parameters in a feedforward neural network can result in confusion matrices vulnerable to irregularities such as Type-I (false positive) or Type-II (false negatives) or the more challenging-to-detect stationarity issues associated with time-series or streaming datatypes.

DL has already been applied in plant science to simulate biochemical behaviors and assist with the spatial analysis of plant evolution. However, quantum deep learning, informed by quantum Bayesianism (QBism), has not been fully exploited in plant science. QBism is a philosophical interpretation of quantum mechanics that considers the perspective of the observer, while quantum deep learning is an approach used in machine learning model architecture. Using quantum deep learning informed by QBism can help overcome the computational intractability of large datasets, including those found in plant science.

Relational quantum ML methods, a derivation of scientific ML,

are directed at biomolecular processes. Developing biology-centric scientific functions in statistical and multi-purpose programming languages such as R and Python, respectively, is critical because the datasets are large and the data in question are complex

The potential of physics-informed M-class intelligent agents, such as the Relational Quantum-Informed Neural Network or RQINN, can positively impact molecular plant science in areas such as genome editing in plant breeding, plant-based edible vaccines, RNA interference biotechnology, and sustainable agriculture supporting food security.

In conclusion, plant science is a complex and data-intensive field that can benefit greatly from the application of machine learning and deep learning techniques. However, the complexity of biological systems and the large-scale datasets involved present significant challenges for traditional machine learning methods. Physics-informed machine learning techniques offer a promising solution to these challenges by incorporating known physical properties of biological systems into the model architecture. By leveraging these physical properties, physics-informed machine learning can improve the accuracy of predictions, reduce the amount of training data required, and enable faster and more efficient modeling of complex biological systems.

Competing Interests
The authors declare no competing financial interests in relation to the work described in this manuscript.

References

1. Sammut, C., & Webb, G. I. (2017). *Encyclopedia of machine learning and data mining.* Springer Publishing Company, Incorporated.
2. Yoosefzadeh-Najafabadi, M., Earl, H. J., Tulpan, D., Sulik, J., & Eskandari, M. (2021). Application of machine learning algorithms in plant breeding: predicting yield from hyperspectral reflectance in soybean. *Frontiers in plant science, 11*, 624273.
3. Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., & Kohli, P. (2021). Advancing mathematics by guiding human intuition with AI. *Nature, 600*(7887), 70-74.
4. Bharadwaj, D. N. (Ed.). (2018). *Advanced molecular plant breeding: meeting the challenge of food security.* CRC Press.
5. Du, K. L., & Swamy, M. N. S. (2019). Neural Networks and Statistical Learning.
6. Berrar, D. (2018). *Bayes' theorem and naive Bayes classifier. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, 403.*
7. Sugiyama, M., & Kawanabe, M. (2012). *Machine learning in non-stationary environments: Introduction to covariate shift adaptation.* MIT press.
8. Rovelli, C. (2022). Relational Quantum Mechanics. *Bulletin of the American Physical Society.*
9. Von Baeyer, H. C. (2016). QBism: The Future of Quantum Physics. *QBism: The Future of Quantum Physics.*
10. Agarwal, C. C. (2018). *Neural Networks and Deep Learning.*
11. Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical statistics for data scientists: 50+ essential concepts using R and Python.* O'Reilly Media.
12. Geng, H. (Ed.). (2016). *Internet of Things and Data Analytics Handbook.* John Wiley & Sons.
13. Sajjan, M., Li, J., Selvarajan, R., Sureshbabu, S. H., Kale, S. S., Gupta, R., & Kais, S. (2021). Quantum computing enhanced machine learning for physico-chemical applications.
14. Kelly, J. K. (2011) The breeder's equation. Nature Education Knowledge 4(5):5.
15. Farlow, S. J. (2012). Partial differential equations for scientists and engineers. Courier Corporation.
16. Koller, D., Friedman, N., Džeroski, S., Sutton, C., McCallum, A., Pfeffer, A., & Yih, W. T. (2007). *Introduction to statistical relational learning.* MIT press.