# CSCI 5481 Final Project
## Quantum machine learning in bioinformatics

Brian Cooper

coope824@umn.edu

University of Minnesota

December 4, 2019

# 1   Abstract

Quantum computing is an emergent field that provides promising potential for the field of bioinformatics. Among various problems, classification is a common task in bioinformatics machine learning tasks for identifying and predicting classes of objects, such as organisms or genes. Using a quantum variant of the classical support vector machine, a quantum model can be created in a similar manner to a classical support vector machine model for making predictions. Applied quantum computing is quite new, and it is not commonly used in modern bioinformatics experiments. Here, it is demonstrated that quantum computing frameworks, although in their infancy, can be used to design and test classification models, and are robust enough to prototype. The accuracy of the quantum support vector machine is demonstrated to compare competitively with its classical counterpart, despite claims that quantum computing is not yet mature enough for widespread use. This challenges the current stigma associated with quantum computing. These results demonstrate that quantum computing may be ready for testing in modern experiments extending beyond bioinformatics into other scientific domains.

# 2   Related Work

Quantum computing has a rich theoretical history, but not many have pursued it in an applied and experimental context. Biamonte, et al. [1] argue that the foundation for quantum machine learning algorithms is strong, yet considerably limited by hardware and software challenges. An important point in [2] is that the ability to simulate the actual, concrete learning process in quantum systems is unsatisfactorily documented, and thus requires more development. In contrast to these worldviews, it is argued in [1] that machine learning has a bright future when utilizing quantum processing power, and they mention that significant advancements have been made when it comes to hardware.

# 3 Methods

Qiskit provides two distinct methods for performing quantum computing experiments:

- Running tasks on real quantum computing hardware (connected with an API to IBM Q computers)

- A built-in quantum hardware simulator for performing experiments on a classical computer

The simulator was chosen, as I do not have any academic credit required to perform experiments on real quantum computing hardware. This necessarily rules out any accurate analysis of the time and space complexities, since quantum computing mathematics is calculated within a classical computer environment. Quantum computing algorithms are, of course, designed to be run on quantum hardware, so the computational time and space cannot be accurately gauged in this manner. Qiskit does provide GPU acceleration support through its QCGPU communication package. [?]

Multivariate data was selected from the UCI Machine Learning Repository [?]. Instances from three datasets were used:

- E. Coli [?]

- Yeast [?]

- Mice Protein Expression [?]

The full E. Coli (336 instances, 7 features) and Yeast (1484 instances, 8 features) datasets were used, but a subset of the mouse dataset was used (553 instances – after removing instances with missing attributes – and the first 9 features). This is because the original size of the mouse dataset, with 1080 instances and 82 features, would be intractable for a classical quantum simulation without distributing the task with GPU acceleration or a real quantum computer.

A support vector machine with a radial-basis kernel was chosen. Each of the datasets, having features and class labels for training and testing, are capable of building classification models.

The experiments were performed on an Intel processor at 4.60GHz running Linux.

# 4 Results

Table 1 shows the accuracy results of the classical and quantum support vector machine experiments on all three datasets.

|        | Classical SVM | Quantum SVM |
|--------|---------------|-------------|
| E. Coli | 90% | 70% |
| Yeast  | 45% | 49% |
| Mouse  | 43% | 22% |

Table 1: Model accuracy results.

The relatively lower accuracy on the mouse dataset may be improved by using more data from the mouse dataset (up to the entire set), which would be more practical with real quantum hardware due to its scale, as described previously in the *Methods* section. Although the removal of data results in lower accuracy, the primary focus of this research is to compare the performance of the two support vector machine approaches, rather than focus on optimizing the model performance itself.

# 5    Conclusion

Abc

# 6    Supplementary Material

All of the source code is publicly available at [URL].

# References

[1] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[2] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172–185, 2015.