
APRENDIZADO QUÂNTICO-CLÁSSICO: APLICADO A CLASSIFICAÇÃO DE CÂNCER DE MAMA

Lucas Leonardo Fazioni
Instituto de Informática
Universidade Federal do Rio Grande do Sul
Porto Alegre, Rio Grande do Sul - Brasil
lucas.fazioni@inf.ufrgs.br

RESUMO

Este trabalho apresenta uma abordagem baseada em Aprendizado de Máquina utilizando uma arquitetura híbrida Clássico-Quântica para a tarefa de classificação de câncer de mama. O estudo compara o desempenho entre dois modelos: o *Support Vector Classifier* (SVC), tradicionalmente utilizado em problemas supervisionados, e o *Quantum Support Vector Classifier* (QSVC), uma versão quântica construída sobre kernels quânticos simulados. A pesquisa propõe uma pipeline completa de pré-processamento de dados de expressão gênica, seleção de atributos via técnicas embarcadas e aplicação de normalização para codificação eficiente em qubits. Os experimentos revelam que, apesar do maior tempo de execução, o QSVC apresentou resultados superiores em termos de *recall* e *F1-score*, evidenciando sua capacidade de identificar corretamente casos positivos em um contexto clínico. A análise das métricas também demonstra que o modelo quântico oferece uma alternativa viável e promissora frente aos métodos clássicos, principalmente em cenários com alta dimensionalidade e restrições computacionais. Esses resultados reforçam o potencial do Aprendizado Quântico como uma extensão estratégica para modelos híbridos em aplicações reais de classificação.

Palavra-chave Aprendizado Quântico · Clássico-Quântico · SVC · QSVC · Aprendizado de Máquina

1 Introdução

O avanço recente da computação quântica tem proporcionado novas oportunidades para o desenvolvimento de modelos híbridos no campo do Aprendizado de Máquina (ML), combinando abordagens clássicas com circuitos quânticos. Esse paradigma, conhecido como *Aprendizado Clássico-Quântico*, é especialmente promissor na era dos dispositivos quânticos [1]. Modelos como o *Quantum Support Vector Classifier* (QSVC) surgem como extensões quânticas de algoritmos validados, como o *Support Vector Classifier* (SVC), explorando a codificação de dados em qubits e o uso de kernels quânticos [2].

No contexto da saúde, especificamente na detecção de câncer de mama, modelos de aprendizado quântico oferecem novas perspectivas para diagnósticos mais precisos e eficientes, como discutido por Mohammed Sufiyan [3]. A aplicação de QSVC, combinada com pipelines clássicas de engenharia de atributos.

1.1 Bit versus Qubit

No contexto da computação clássica, a unidade fundamental de informação é o **bit**, que pode assumir apenas dois estados discretos: 0 ou 1. A notação formal de um bit é representada como:

$$Bit \in \{0, 1\} \quad (1)$$

Na computação quântica, por outro lado, a unidade básica de informação é o **qubit** (*quantum bit*). Diferente do bit clássico, o qubit pode estar em uma superposição linear dos estados 0 e 1 simultaneamente, de acordo com os princípios da mecânica quântica. Sua notação é dada por:

$$\psi = \alpha 0 + \beta 1 \quad (2)$$

Essa propriedade permite que o qubit represente múltiplas combinações de estados ao mesmo tempo, proporcionando uma vantagem em termos de paralelismo computacional. Principalmente, para tarefas de aprendizado de máquina, onde grandes espaços de busca e manipulação de dados em alta dimensionalidade são comuns [4].

1.2 Objetivos

O objetivo deste trabalho é investigar o desempenho de modelos clássico-quânticos aplicados à uma arquitetura para classificação de câncer de mama. Mais especificamente, busca-se:

- Desenvolver e avaliar uma arquitetura híbrida baseada em SVC e QSVC;
- Explorar a aplicabilidade de técnicas de seleção de atributos na preparação de dados para algoritmos quânticos;
- Comparar métricas de desempenho como *acurácia*, *recall*, *precision* e *F1-score*;
- Analisar a viabilidade computacional do modelo quântico frente ao modelo clássico;
- Contribuir com evidências experimentais sobre o potencial do *Aprendizado Quântico* no domínio médico.

2 Métodos

2.1 Conjunto de Dados

O conjunto de dados utilizado foi obtido da base GSE70947, curada e disponibilizada por meio do CuMiDa — *Curated Microarray Database* para cânceres humanos [5]. Essa base é composta por **289 amostras** com mais de **35.000 atributos**, representando perfis de expressão gênica associados ao câncer de mama, com variável de classe balanceada entre "breast_adenocarcinoma" e "normal".

2.2 Modelo Clássico

2.2.1 SVC (Support Vector Classifier)

O modelo clássico utilizado foi o *Support Vector Classifier* (SVC), uma implementação baseada no algoritmo de vetores de suporte com kernel RBF. O SVC busca maximizar a margem entre as classes ao encontrar o hiperplano ótimo de separação. Neste trabalho, os hiperparâmetros do modelo, como *C* e *gamma*, foram otimizados utilizando validação cruzada com *GridSearchCV*.

2.3 Modelo Quântico

2.3.1 QSVC (Quantum Support Vector Classifier)

O QSVC é a versão quântica do classificador SVC, implementado utilizando o framework Qiskit [6]. Ele utiliza um kernel quântico baseado em circuitos parametrizados que avaliam a similaridade entre amostras em um espaço. O modelo foi treinado utilizando um simulador quântico local [1, 7, 8].

2.4 Arquitetura Híbrida

A arquitetura híbrida combina componentes clássicos e quânticos em uma única pipeline de aprendizado. O objetivo é explorar o poder de representação quântica com o apoio da robustez de técnicas clássicas.

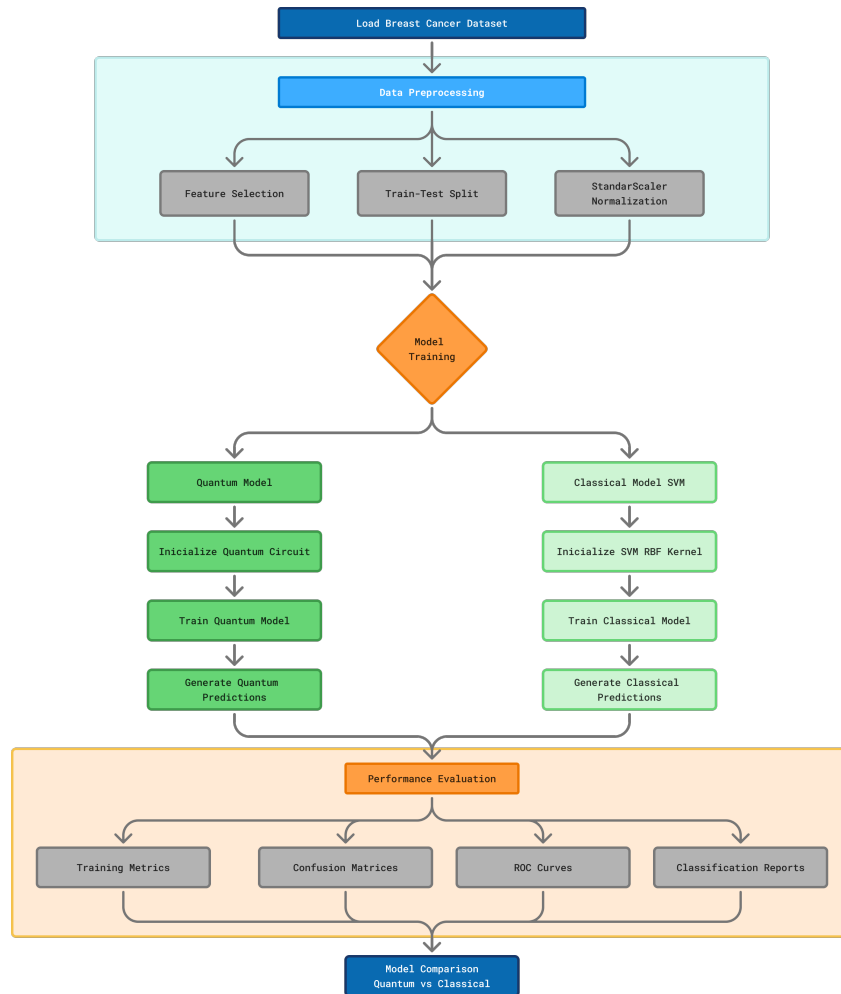


Figura 1: Pipeline desenvolvida para o processamento e classificação de cancer de mama.

2.4.1 Feature Selection

Foi empregada uma abordagem mista para seleção de atributos. O que possibilitou compreender a distribuição das classes, identificar atributos redundantes e visualizar correlações relevantes. Técnicas como análise de variância, histogramas e matrizes de correlação foram utilizadas para avaliar a importância relativa das *features*. O método empregado para seleção de características, foi *embedded* (como LGBMClassifier) com filtros estatísticos (correlação, VIF). Essa etapa reduziu o número de atributos de **35.983** para **84**, possibilitando a codificação eficiente dos dados em circuitos quânticos [7].

2.4.2 Modelagem

A modelagem foi realizada paralelamente com os classificadores SVC e QSVC. Ambos os modelos foram treinados com os mesmos dados e atributos, utilizando divisão de treino e teste de 70/30. O QSVC foi implementado com kernel quântico customizado, enquanto o SVC utilizou kernel RBF otimizado.

2.4.3 Métricas de Avaliação

A comparação entre os modelos foi realizada com base na Tabela 1, além da análise da matriz de confusão, com especial atenção ao *recall*, fundamental em contextos médicos.

Tabela 1: Descrição das métricas de avaliação utilizadas nos modelos

Métrica	Descrição
Acurácia	Proporção de classificações corretas em relação ao total de amostras avaliadas. Mede a performance global do modelo.
Precision	Taxa entre os verdadeiros positivos e todas as previsões positivas. Mede a exatidão entre os casos classificados como positivos.
Recall (Sensibilidade)	Proporção de verdadeiros positivos entre todos os exemplos que realmente pertencem à classe positiva. Avalia a capacidade de detectar todos os casos positivos.
F1-Score	Média harmônica entre <i>precision</i> e <i>recall</i> , equilibrando as duas métricas em um único indicador robusto.
AUC-ROC	Área sob a curva ROC. Representa a capacidade do classificador em separar corretamente as classes ao longo de diferentes limiares de decisão.

3 Trabalhos Relacionados

Diversos estudos recentes têm investigado o uso de algoritmos de aprendizado quântico para tarefas de classificação supervisionada. Rebentrost et al. [2] propuseram um modelo de *Quantum Support Vector Machine* (QSVM) com complexidade logarítmica, baseado em inversão de matrizes quânticas, demonstrando potencial para acelerar algoritmos clássicos de grande escala. Chen et al. [1] avançaram nesta linha ao propor uma arquitetura híbrida treinável de ponta a ponta, combinando redes tensoriais com circuitos variacionais quânticos.

No contexto de dados biomédicos, Sufiyan et al. [3] demonstraram que circuitos quânticos podem ser eficazes na extração de características de dados médicos de alta dimensionalidade, como nos casos de câncer de mama.

4 Resultados

Nesta seção, apresentamos os resultados obtidos a partir da aplicação de métodos clássicos-quânticos, como: SVC e QSVC. A pipeline desenvolvida para este estudo consistiu em quatro etapas principais: (1) pré-processamento dos dados (CuMiDa), (2) Seleção de Atributos (*Feature Selection*), (3) arquitetura híbrida, e (4) avaliação de desempenho e performance.

4.1 Desempenho Comparativo dos Métodos Clássico-Quântico

A comparação entre os modelos SVC (clássico) e QSVC (quântico) foi realizada com base na mesma divisão dos dados e atributos selecionados. Ambos foram treinados com conjunto de dados balanceado, reduzido para comportar os limites do simulador quântico. A Tabela 2 apresenta os resultados obtidos:

Tabela 2: Desempenho comparativo entre modelos SVC e QSVC em arquitetura híbrida

Modelo	Tempo (s)	Accuracy	F1-Score	ROC-AUC	Precision	Recall
SVC	11.27	0.87	0.88	0.95	0.86	0.91
QSVC	385.55	0.90	0.91	0.93	0.88	0.93

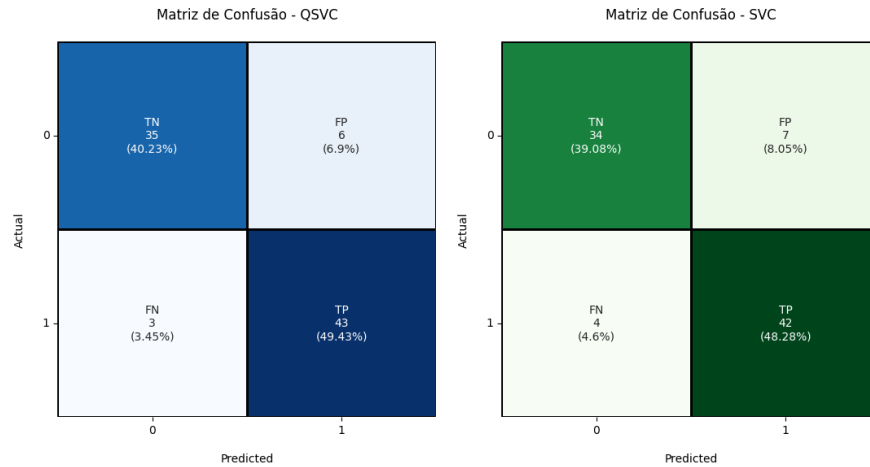


Figura 2: Matriz de confusão para os modelos SVC (direita) e QSVK (esquerda)

4.2 Avaliação das métricas de classificação e performance

Os resultados indicam que o modelo QSVK obteve desempenho superior nas métricas relacionadas à sensibilidade (*recall*) e equilíbrio entre precisão e revocação (*F1-score*). Isso é especialmente relevante para aplicações médicas, onde a capacidade de identificar corretamente os casos positivos é fundamental.

O SVC, por outro lado, apresentou um valor ligeiramente maior de AUC (0.95), conforme representação abaixo 3, evidenciando maior capacidade discriminativa em cenários binários com variação de limiar.

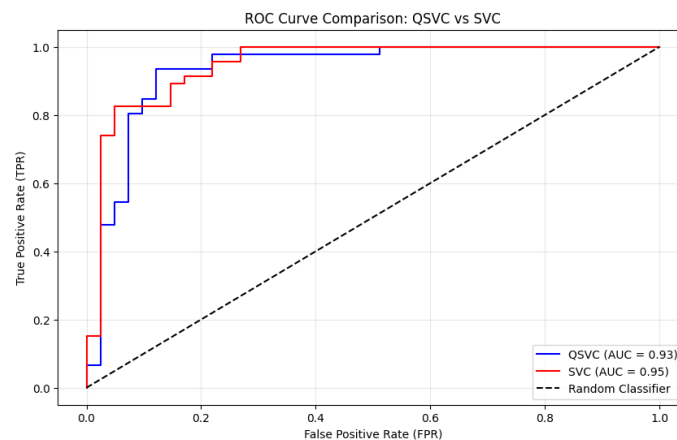


Figura 3: Área sob curva, comparando performance do modelo clássico-quântico

4.3 Limitações e Desafios

Nossos experimentos também revelaram limitações importantes:

4.3.1 SVC (Support Vector Classifier)

Apesar da eficiência computacional e fácil integração em pipelines clássicas, o modelo SVC enfrenta limitações em:

- Generalização em espaços de alta dimensionalidade;
- Dependência significativa da escolha do kernel e hiperparâmetros.

4.3.2 QSVC (Quantum Support Vector Classifier)

O modelo QSVC mostrou resultados promissores, mas apresenta alguns desafios importantes:

- Alto tempo de simulação e dependência de infraestrutura especializada (hardware real);
- Limitação de escalabilidade devido ao número de qubits suportado por simuladores locais;
- Necessidade de normalização e compressão extrema dos dados para codificação eficiente [1, 7].

Apesar dessas dificuldades, os resultados obtidos validam o uso de abordagens híbridas como alternativa viável e complementar aos métodos clássicos, com potencial de ganho futuro conforme a maturação de hardwares e frameworks quânticos.

5 Conclusão

Este trabalho apresentou uma abordagem híbrida de aprendizado de máquina combinando modelos clássicos e quânticos (SVC e QSVC) para a classificação de câncer de mama, utilizando dados de expressão gênica do CuMiDa [5]. A partir da aplicação de técnicas de redução de dimensionalidade e normalização, foi feita a inicialização do circuito quântico e o kernel RFB, o foi induzido os classificadores a partir das suas arquiteturas.

Os resultados experimentais demonstraram que o modelo quântico (QSVC) alcançou desempenho superior em métricas cruciais como *F1-score* e *recall*, aspectos importantes especialmente em contextos para diagnósticos médicos, onde minimizar falsos negativos é vital. Por outro lado, o modelo clássico SVC se destacou em termos de *AUC* e eficiência computacional, sendo mais viável em aplicações em larga escala no cenário atual (simulador quântico local).

Apesar do custo computacional elevado e das limitações práticas dos simuladores quânticos, o QSVC demonstrou potencial significativo para tarefas de classificação, corroborando os achados em estudos anteriores [2, 1, 7].

Como trabalho futuro, recomenda-se a avaliação do modelo em hardware quântico real, bem como o uso de técnicas de mitigação de ruído e estratégias mais avançadas de seleção de atributos quânticos. A continuidade dessa linha de pesquisa pode abrir caminho para soluções diagnósticas mais precisas, personalizadas e eficientes.

Referências

- [1] Samuel Yen-Chi Chen, Chih-Min Huang, Chia-Wei Hsing, and Ying-Jer Kao. An end-to-end trainable hybrid classical-quantum classifier. *Machine Learning: Science and Technology*, 2(4):045021, 2021.
- [2] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Physical Review Letters*, 113(13):130503, 2014.
- [3] Mohammed Sufiyan, Anas Niaz, Saifuddin Syed, and Syed Riyan. Quantum-classical hybrid cancer classification system. *International Journal for Multidisciplinary Research*, 7(1):1–6, 2025.
- [4] Shivani Saini, P. K. Khosla, Manjit Kaur, and Gurmohan Singh. Quantum driven machine learning. *International Journal of Theoretical Physics*, 59:4013–4024, 2020.
- [5] Yuri A. L. Silva, Ana T. Winck, Roberto M. Cesar Jr., Juan C. Alcaraz, and Daniel C. Costa. A cancer gene expression dataset curated with microarray platform harmonization and batch effect removal. *Journal of Computational Biology*, 26(12):1244–1253, 2019.
- [6] IBM Quantum. Ibm quantum, 2022.
- [7] Michele Grossi, Noelle Ibrahim, Voica Radescu, Robert Loredo, Kirsten Voigt, Constantin von Altrock, and Andreas Rudnik. Mixed quantum–classical method for fraud detection with quantum feature selection. *IEEE Transactions on Quantum Engineering*, 3:1–12, 2022.
- [8] IBM Quantum. Ibm quantum learning. *learning. quantum. ibm. com*, 2023.