

Análise de Redes Neurais para CRISPR: Uma Abordagem com Computação Quântica *

Marcus Navarro Gabrich, Henrique Cota de Freitas, Matheus Alcântara Souza

Departamento de Ciência da Computação, Pontifícia Universidade Católica de Minas Gerais
Belo Horizonte, MG, Brazil

1276826@sga.pucminas.br, {cota, matheusalcantara}@pucminas.br

Resumo. O sistema CRISPR transformou a biotecnologia ao possibilitar a manipulação precisa do DNA e avançar a compreensão de distúrbios genéticos. No entanto, as complexas interações entre o DNA e o CRISPR continuam a representar desafios significativos, exigindo processamento de alta performance. Este estudo propõe uma abordagem que integra tecnologia quântica, algoritmos de aprendizado de máquina e CRISPR para lidar com esses desafios. A análise foca na relação entre a dependência do gene CRISPR e o número de cópias de genes repetidos para prever a dependência genética e avaliar o impacto dos experimentos CRISPR. Embora alguns modelos tenham alcançado valores de R^2 negativos, esses resultados destacam a complexidade do problema e as oportunidades para melhorias futuras. Este estudo também avalia a viabilidade de redes neurais híbridas quânticas para problemas de regressão, mostrando a computação quântica como uma solução para superar os desafios de HPC em biotecnologia e análise de dados em larga escala.

1. Introdução

Biotecnologia e computação quântica são duas áreas científicas em ascensão. O sistema de edição genética *Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR), que possibilita a modificação precisa e direcionada do DNA, revolucionou a biologia molecular [Bortesi and Fischer 2014]. No entanto, interpretar os resultados desses experimentos continua sendo um desafio significativo devido à complexidade das interações entre o DNA alvo e os componentes do sistema CRISPR. A computação quântica, por sua vez, oferece novas perspectivas para lidar com problemas complexos, como a análise de grandes volumes de dados genéticos e moleculares, prometendo ganhos em desempenho computacional [Preskill 2023, Mangini et al. 2021, Mauranyapin et al. 2022]. Embora essa tecnologia tenha avançado em diversas áreas, seu uso em problemas de regressão ainda apresenta desafios.

Apesar da capacidade do CRISPR de revolucionar a edição do genoma, há dificuldades na previsão precisa dos efeitos genéticos, particularmente em problemas de regressão. A análise da relação entre a dependência do gene CRISPR e o número de cópias de genes repetidos ainda carece de modelos preditivos eficazes que possam ser aplicados a dados reais. Nesse contexto, a aplicação de técnicas de redes neurais quânticas surge como uma abordagem promissora.

*Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), código 001; ao Conselho Nacional de Desenvolvimento Científico (CNPq), código 311697/2022-4; à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG); e à PUC Minas; pelo suporte no desenvolvimento deste trabalho.

O objetivo deste trabalho é combinar a tecnologia quântica com o CRISPR, não apenas para prever resultados, mas também para melhorar a compreensão das interações moleculares, proporcionando uma análise mais qualitativa e quantitativa. A hipótese central é que modelos quânticos podem aumentar o desempenho na análise e previsão dos efeitos das edições genéticas, explorando os dados de maneira mais rápida e eficiente [Kak 1995, Gyongyosi and Imre 2019].

Este estudo investiga se a aplicação de técnicas quânticas em problemas de regressão pode trazer melhorias significativas na análise dos resultados, apesar das dificuldades iniciais. Durante o desenvolvimento, foram observados resultados desafiadores, como valores negativos de R^2 , indicando a complexidade do problema. No entanto, tais resultados também ressaltam a importância de aprimorar essas técnicas para resolver problemas reais e concretos.

Para a análise, foram desenvolvidos algoritmos de aprendizado de máquina, tanto quânticos quanto clássicos, para interpretar as interações genéticas. Utilizamos os conjuntos de dados do projeto DepMap, que incluem informações sobre dependência genética e o número de cópias de genes repetidos. O estudo explora como essas assinaturas genéticas podem servir como indicadores confiáveis do impacto das edições CRISPR no comportamento celular, com o valor de dependência genética CRISPR sendo a principal métrica de referência.

O restante deste trabalho está organizado da seguinte forma: na Seção 2, são apresentados fundamentos teóricos e trabalhos relacionados ao tema. A Seção 3 descreve os procedimentos metodológicos adotados. Na Seção 4, são apresentadas as análises realizadas, e, por fim, na Seção 6, as considerações finais e direções futuras são discutidas.

2. Referencial Teórico

Esta seção apresenta os fundamentos teóricos envolvidos no contexto interdisciplinar entre computação quântica, aprendizado de máquina e biologia/genética.

2.1. Biotecnologia, edição genética e Tecnologia CRISPR

O sistema CRISPR-Cas9, abreviação de *Clustered Regularly Interspaced Short Palindromic Repeats* e *CRISPR-associated proteins*, é uma técnica de edição genética presente em procariotos. Este sistema funciona como um mecanismo de defesa contra a invasão de vírus e outros patógenos, e introduz um método revolucionário na biologia molecular [Bortesi and Fischer 2014]. Composto por dois elementos principais, a região CRISPR contém sequências de DNA palindrômicas curtas intercaladas por *spacers*, fragmentos de DNA viral capturados durante infecções anteriores, e as *Proteínas Cas*, especialmente a *Cas9*, que atuam como “tesouras moleculares”, cortando o DNA em locais específicos. A utilização de uma nucleáse guiada por RNA *Cas9*, possibilita a modificação precisa do DNA, incluindo adições (Knock-in ou KI), remoções (Knock-out ou KO), ou substituições de bases. Esse processo tem três etapas: reconhecimento, corte e reparo, onde o DNA é quebrado e pode ser reparado por vias como o reparo de DNA não homólogo (NHEJ) ou a recombinação homóloga dirigida (HDR) [Xu and Li 2020].

As aplicações do CRISPR abrangem diversas áreas. Por exemplo, na medicina, destaca-se o tratamento de doenças genéticas, terapia do câncer e desenvolvimento de novos medicamentos. Na agricultura, a edição genética pode melhorar a produtividade

das plantações e conferir resistência a pragas e condições climáticas adversas. Na biotecnologia, contribui para a produção de biocombustíveis e biomateriais com propriedades específicas, além do desenvolvimento de novas variedades de plantações.

2.2. Computação quântica

A mecânica quântica descreve o comportamento da matéria e da energia em escalas atômicas e subatômicas. Alguns conceitos fundamentais para a computação quântica incluem a superposição, emaranhamento e interferência. A superposição permite que um sistema quântico esteja em múltiplos estados simultaneamente. O emaranhamento conecta os estados de dois sistemas quânticos, enquanto a interferência influencia a probabilidade de eventos quânticos [Preskill 2018].

Os modelos de computação quântica utilizam qubits como unidades básicas de informação, que diferem dos bits tradicionais por sua capacidade de estar em superposição. Os circuitos quânticos consistem em operações realizadas em qubits, semelhantes às portas lógicas dos computadores clássicos, sendo uma sequência de operações que são realizadas em qubits [Gyongyosi and Imre 2019].

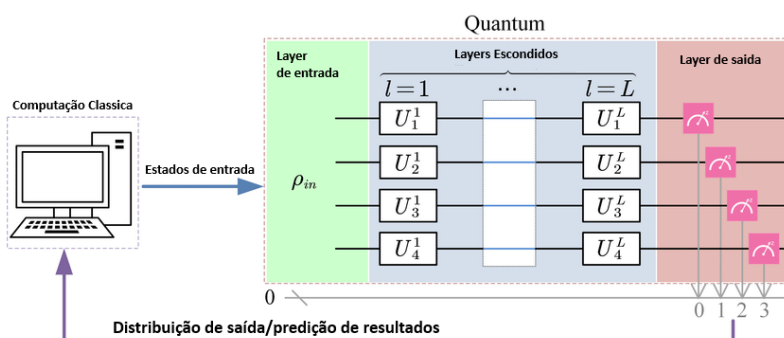


Figura 1. Rede neural Quântica Híbrida

As Redes Neurais Quânticas, do inglês *Quantum Neural Networks* (QNNs), são uma abordagem emergente na interseção da computação quântica e do aprendizado de máquina [Maria Schuld and Petruccione 2015]. Elas buscam combinar a capacidade de processamento paralelo das redes neurais clássicas com a superposição e entrelaçamento da mecânica quântica. As Redes Neurais Híbridas-Quânticas (*Hybrid QNNs*) representam uma tentativa de combinar os pontos fortes das redes neurais clássicas e quânticas. Em uma HQNN, uma rede neural clássica é usada para pré-processar os dados de entrada e/ou pós-processar os dados de saída, enquanto uma rede neural quântica é usada para realizar computações complexas, como indicado na Figura 1. As HQNNs têm o potencial de aproveitar a eficiência computacional e a capacidade de processamento paralelo das redes neurais quânticas, ao mesmo tempo que mantêm a facilidade de treinamento e a flexibilidade das redes neurais clássicas. No entanto, a pesquisa em QNNs ainda está em seus estágios iniciais, e muitos desafios técnicos e teóricos permanecem [Biamonte et al. 2017].

2.3. Trabalhos Relacionados

Alguns estudos da literatura são correlatos ao presente trabalho. Como exemplo, em relação a técnicas e abordagens CRISPR, um trabalho apresentou uma visão

abrangente sobre a edição do genoma, abordando a modificação do DNA genômico em locais específicos em diversos tipos celulares e organismos [Xu and Li 2020]. Os autores destacaram a capacidade de inserção (Knock-In), remoção (Knock-Out), e substituição de DNA, resultando na inativação de genes-alvo, aquisição de características genéticas inovadoras e correção de mutações genéticas patogênicas. Ainda em relação a CRISPR, uma visão detalhada do sistema CRISPR/Cas e sua adoção em diferentes organismos, especialmente plantas, para várias aplicações foi apresentada [Bortesi and Fischer 2014]. As investigações revelaram as implicações do sistema CRISPR/Cas como uma ferramenta promissora para modificações genéticas direcionadas em plantas.

No contexto de computação quântica, foi feita uma análise que se baseia nos fundamentos da mecânica quântica, como a superposição quântica, o emaranhamento quântico e o teorema da não clonagem [Gyongyosi and Imre 2019]. Dado que esses fenômenos não têm um análogo clássico, resultados semelhantes não podem ser alcançados dentro do contexto da computação tradicional. Outro estudo destacou como tensores químicos quânticos são utilizados para interpretar e prever a eficiência do sgRNA CRISPR-Cas9 [Noshay et al. 2023]. Essa pesquisa aprimora a compreensão dos processos biológicos envolvidos na edição genética, introduzindo novas abordagens para interpretação dos experimentos.

Na área de interesse de aprendizado de máquina integrado a computação quântica, um artigo ofereceu uma visão geral do campo do aprendizado de máquina quântico [Maria Schuld and Petruccione 2015]. Os autores apresentaram as abordagens e detalhes técnicos, discutindo o potencial de uma futura teoria de aprendizado quântico. Foi explicado nesse trabalho, como os algoritmos de aprendizado de máquina aprendem a relação entre entrada e saída a partir de exemplos, usando a tecnologia quântica.

Ainda nessa linha, Maria Schuld et al. apresentaram um estudo fundamental sobre classificadores centrados em circuitos quânticos, aplicáveis em diversos contextos de aprendizado de máquina [Schuld et al. 2020]. Esse trabalho demonstra como técnicas de circuitos quânticos podem ser adaptadas para redes neurais, destacando o potencial das redes neurais híbridas quânticas para lidar com tarefas complexas de análise de dados, fazendo uso da biblioteca Pennylane, uma plataforma open-source para a implementação de algoritmos quânticos e aprendizado de máquina quântico.

Outro artigo explorou como os computadores quânticos podem superar os computadores clássicos em tarefas de aprendizado de máquina [Biamonte et al. 2017]. Foi investigado como desenvolver e implementar software quântico concreto que ofereça tais vantagens. No mesmo sentido, de que algoritmos quânticos podem resolver problemas complexos de química quântica, outro estudo mostrou como a computação quântica resolveria problemas intratáveis de química, com implicações para a medicina personalizada e a biotecnologia [Aspuru-Guzik 2020].

Na área de redes generativas quânticas um trabalho introduziu o conceito de redes adversárias generativas quânticas (QuGANs), que são uma ferramenta poderosa para o aprendizado de máquina clássico [Lloyd and Weedbrook 2018]. Essas QuGANs combinam a mecânica quântica com o aprendizado de máquina clássico, permitindo melhorias significativas na eficiência do treinamento e na qualidade dos resultados. Além disso, outro artigo trouxe um novo método de computação evolutiva baseado em

princípios da mecânica quântica [Narayanan and Moore 1996]. Os autores exploraram como a mecânica quântica pode informar e inspirar métodos de computação evolutiva mais eficientes. Eles compararam um algoritmo genético clássico com um método inspirado na quântica para resolver o problema do caixeiro-viajante, mostrando informalmente que o algoritmo inspirado na quântica supera seu equivalente clássico em um pequeno domínio. Por fim, no contexto de aprendizado de máquina com ênfase em Redes Neurais Quânticas, foram explorados temas de modelos computacionais baseados nos princípios da mecânica quântica [Kak 1995]. Estas redes combinam modelos de redes neurais artificiais clássicas com os benefícios da informação quântica, visando desenvolver algoritmos mais eficientes. Essas redes têm potencial para analisar grandes conjuntos de dados de forma mais eficiente do que os métodos tradicionais, o que as torna promissoras para aplicações no contexto da edição genética CRISPR.

Os trabalhos relacionados apontam uma lacuna na área, em que geralmente são realizadas tarefas de classificação. O presente trabalho preenche essa lacuna com modelos de aprendizado de máquina com redes neurais híbridas quânticas para análise de problemas de regressão, visando prever os valores objetivo e avaliando o desempenho.

3. Metodologia

A partir do estudo do referencial teórico, o processo de implementação da solução proposta para o problema de previsão da dependência genética CRISPR foi iniciado. A linguagem de programação escolhida foi o Python, devido à sua versatilidade e à disponibilidade de bibliotecas para aprendizado de máquina e computação quântica. O ambiente de desenvolvimento utilizado foi o Python Notebook. Para a geração de circuitos e técnicas quânticas, foram empregadas as plataformas *IBM Quantum*, *Pennylane* e *Qiskit*. A *Pennylane* desempenhou um papel fundamental neste projeto, oferecendo a capacidade de integrar camadas quânticas com redes neurais clássicas, possibilitando a implementação de redes híbridas quânticas. A técnica de classificadores centrados em circuitos quânticos, descrita por Schuld et al. [Maria Schuld and Petruccione 2015], foi utilizada para desenvolver os modelos híbridos. Essa integração foi essencial para explorar o potencial dos circuitos quânticos em problemas de regressão, utilizando dados biológicos reais.

Os testes foram realizados localmente utilizando simulações de hardware clássico para emular o comportamento de circuitos quânticos. Embora as simulações de hardware quântico real apresentem limitações de tempo, elas permitiram uma análise preliminar do desempenho dos modelos híbridos. A *API da Pennylane* foi utilizada para gerenciar essas simulações e gerar circuitos quânticos complexos. Os conjuntos de dados utilizados no estudo foram extraídos do projeto *DepMap*, que inclui dados de dependência genética (*CRISPRGeneDependency*) e de número de cópias de genes (*OmicsCNGene*). Esses conjuntos de dados, que representam dependências genéticas de centenas de linhagens celulares cancerígenas, são essenciais para os experimentos. A análise dos dados foi realizada de maneira a garantir que não houvesse registros redundantes, e os conjuntos de treinamento e teste foram ajustados para permitir experimentos em diferentes escalas.

A técnica empregada para o desenvolvimento do algoritmo incluiu análise estatística e aprendizado de máquina. Para a manipulação e pré-processamento dos dados, foi utilizada a biblioteca *Pandas*, enquanto as bibliotecas *Matplotlib* e *Seaborn*

foram empregadas para criar visualizações, como gráficos de dispersão, que ilustram a correlação entre o Knock-out de CRISPR e os níveis de dependência gênica em diferentes linhagens celulares. A *engenharia de características* foi aplicada para selecionar as variáveis mais relevantes, reduzindo a dimensionalidade do conjunto de dados através de técnicas de filtragem baseadas em correlação.

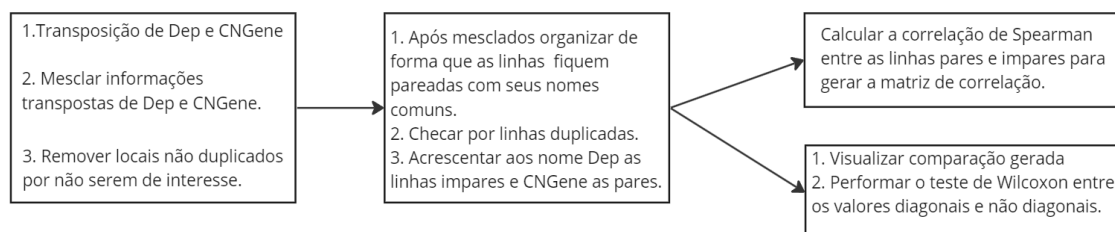


Figura 2. Pipeline do processo de triagem sistemática da relação entre dependência de gene CRISPR (Dep) e contagem de cópias de genes (CNGene) no projeto CCLE (Cancer Cell Line Encyclopedia).

Para análises estatísticas adicionais, as bibliotecas *NumPy* e *SciPy* foram utilizadas para realizar validações estatísticas, como a geração de matrizes de correlação e o teste de Wilcoxon. Esse processo de triagem sistemática, conforme ilustrado na Figura 2, permitiu identificar relações significativas entre a dependência do gene CRISPR e o número de cópias de genes, fornecendo insights importantes para o ajuste dos modelos.

O uso da plataforma *IBM Quantum* possibilitou a produção e teste dos circuitos quânticos integrados às redes neurais. Foi possível realizar experimentos com camadas quânticas internas nas redes neurais híbridas, utilizando circuitos desenvolvidos na *Pennylane* para analisar o comportamento dessas redes em problemas de regressão.

Diversos modelos de aprendizado de máquina foram desenvolvidos utilizando as bibliotecas *Tensorflow* e *Keras*, com a integração da *Pennylane* para a construção de redes híbridas com camadas intermediárias quânticas e completamente quânticas. Esses modelos foram treinados para prever a dependência do gene CRISPR com base nos padrões de número de cópias de genes presentes nas linhas celulares.

Para avaliar o desempenho dos modelos, foram utilizadas métricas específicas como: tempo de execução, Acurácia, *Loss* (perda), Erro Médio Absoluto (MAE), Erro Quadrático Médio (MSE), Erro Percentual Absoluto Médio (MAPE), Erro Logarítmico Quadrático Médio (MSLE), Similaridade de Cosseno (*Cosine Similarity*) e *Log-cosh*. Também foram realizadas análises de ganho computacional e eficiência de escalabilidade, comparando o desempenho operacional dos modelos em termos de consumo de recursos computacionais.

Essa abordagem metodológica permitiu uma análise robusta do desempenho dos modelos híbridos quânticos, demonstrando a viabilidade dessas técnicas para resolver problemas complexos de regressão em biotecnologia, especialmente na análise de dados de dependência genética CRISPR.

4. Avaliação de Resultados

4.1. Pre-processamento

Durante o desenvolvimento do projeto, foram realizadas uma série de testes para avaliar a relação entre a dependência genética do CRISPR e a contagem do número de cópias de genes nas linhas genéticas, como gráficos de dispersão, matrizes de correlação e teste de Wilcoxon. Em sequência, utilizando técnicas de aprendizado de máquina de Redes Neurais, foram desenvolvidos modelos preditivos para essa relação.

Os gráficos de dispersão mostram as correlações entre os efeitos de Knock-out do CRISPR e os níveis de cópias de genes em diferentes linhagens celulares. Conforme Figura 3, visualizações ajudam a identificar associações e dependências potenciais entre as perturbações genéticas do CRISPR e os a contagem do número de cópias de genes. Por meio de direção da Associação e padrões Visuais em diferentes genes, mostra-se um reforço que existe relação e uma associação positiva entre as variáveis.

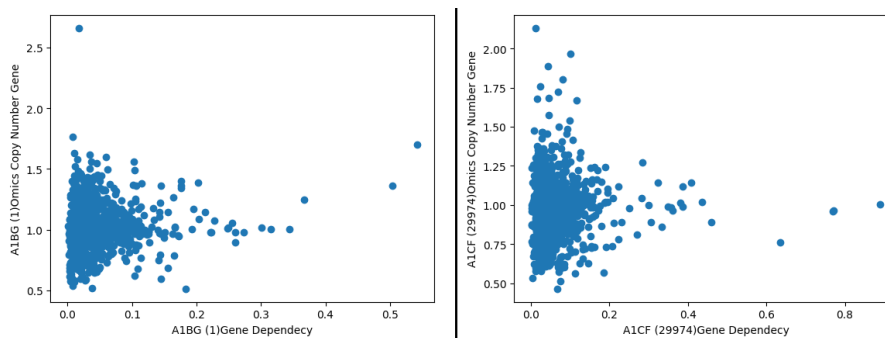


Figura 3. Gráficos de dispersão(scatterplot) para os genes A1BG (1)', 'A1CF (29974)'

A matriz de correlação exibe a análise de Spearman entre a dependência genética do CRISPR e a contagem de cópias de genes. Conforme mostrado na Figura 4, a correlação de Spearman é útil para dados não normais ou relações não lineares. Os valores diagonais indicam auto-correlações da dependência genética com o número de cópias do gene, enquanto os valores fora da diagonal destacam correlações significativas entre genes e a dependência do CRISPR. Essa análise fornece uma visão inicial sobre como o número de cópias de genes se relaciona com a dependência genética do CRISPR, permitindo a comparação de valores diagonais e não diagonais para *insights* adicionais na construção dos modelos.

A fim de realizar uma exploração adicional a relação entre a dependência genética do CRISPR e o número de cópias de genes, foi realizado um teste de Wilcoxon. Esse teste mede a força da associação entre essas variáveis. Um valor maior da estatística do teste indica uma associação mais forte. O teste compara as correlações dentro da matriz de correlação para verificar se são estatisticamente diferentes. No estudo, a estatística média do teste de Wilcoxon foi 112.246,404, indicando uma forte associação entre as variáveis. O valor-p médio foi extremamente baixo ($2,64e-43$), mostrando uma associação altamente significativa. Em resumo, o teste de Wilcoxon revela uma associação forte e significativa entre a dependência genética do CRISPR e o número de cópias de genes, sugerindo que essa relação merece mais investigação.

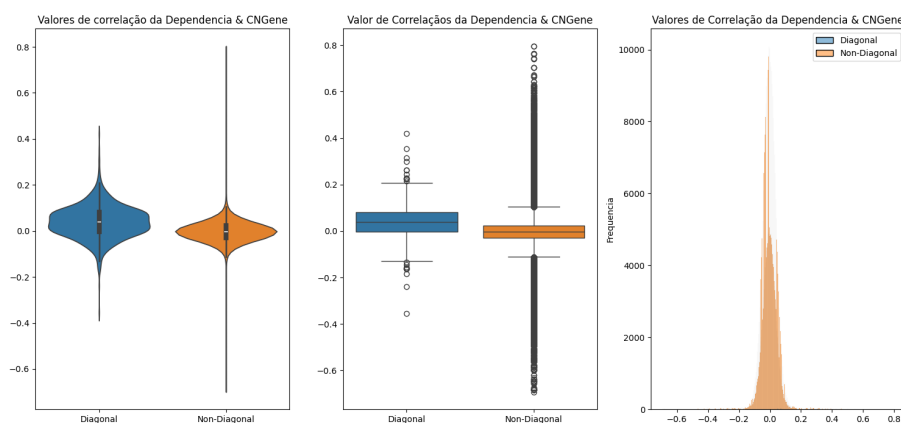


Figura 4. Valores de correlação da Dependência e CNGene para ViolinPlot, Boxplot e Histograma de valores diagonais e não diagonais

4.2. Modelos de aprendizado e abordagens quânticas

Foram utilizadas Redes Neurais, sendo desenvolvidos modelos preditivos para a relação entre a dependência genética do CRISPR com base em padrões na contagem do números de cópias de genes. Foram criados como base 3 modelos de redes neurais direcionados para os *datasets* processados, sendo um modelo de rede neural clássica, um modelo de rede neural com uma camada intermediária quântica, e as outras clássicas e um modelo de rede neural com camadas intermediárias totalmente quânticas. A Tabela 1 apresenta um comparativo das arquiteturas de redes neurais elaboradas.

Tabela 1. Comparação das Arquiteturas das Redes Neurais

Modelo	Camada (Tipo)	Forma de Saída	Parâmetros
NNW Clássica	Dense	(None, 8)	14.544
	Dense_1	(None, 8)	72
	Dense_2	(None, 8)	72
	Dense_3	(None, 8)	72
	Dense_4	(None, 6)	54
NNW Parcial Quântica	Dense	(None, 8)	14.544
	KerasLayer	(None, 8)	48
	Dense_1	(None, 8)	72
	Dense_2	(None, 8)	72
	Dense_3	(None, 6)	54
NNW Totalmente Quântica	Dense	(None, 8)	14.544
	KerasLayer	(None, 8)	72
	KerasLayer_1	(None, 8)	72
	KerasLayer_2	(None, 8)	72
	Dense_1	(None, 6)	54
Total de parâmetros (Clássica, Totalmente Quântica):			14.814
Total de parâmetros (Parcial Quântica):			14.790

4.2.1. Rede Neural Clássica

A rede clássica é um modelo sequencial composto por várias camadas densamente conectadas (*fully connected*). Inicialmente, há uma camada de entrada com 8 neurônios,

onde cada neurônio está conectado a todos os neurônios da camada seguinte. Essa camada de entrada utiliza a função de ativação ReLU (*Rectified Linear Unit*), uma função comumente empregada em redes neurais para introduzir não-linearidade.

Após a camada de entrada, são adicionadas três camadas ocultas, todas compostas por 8 neurônios cada. Novamente, a função de ativação utilizada em todas essas camadas é a ReLU. Essas camadas ocultas são responsáveis por aprender e representar características mais complexas nos dados. Por fim, há uma camada de saída densamente conectada, com 6 neurônios, correspondendo ao número de saídas desejadas.

4.2.2. Rede neural com camadas intermediárias parcialmente quânticas

A Rede neural com camadas intermediárias parcialmente quânticas é um modelo híbrido que mescla camadas convencionais com camadas quânticas. Na configuração proposta, há uma camada de entrada e uma camada de saída clássicas, duas camadas intermediárias clássicas e uma camada intermediária quântica. A camada quântica sendo construída utilizando o módulo PennyLane, permitindo a incorporação de circuitos quânticos como camadas em modelos de redes neurais clássicas. O circuito quântico é definido através da função *qnode* para indicar sua natureza quântica. Este circuito é composto por uma camada de encaixe de ângulo (*AngleEmbedding*) seguida por camadas de entrelaçamento forte (*StronglyEntanglingLayers*), padrões comuns em circuitos quânticos usados para aprendizado de máquina quântica.

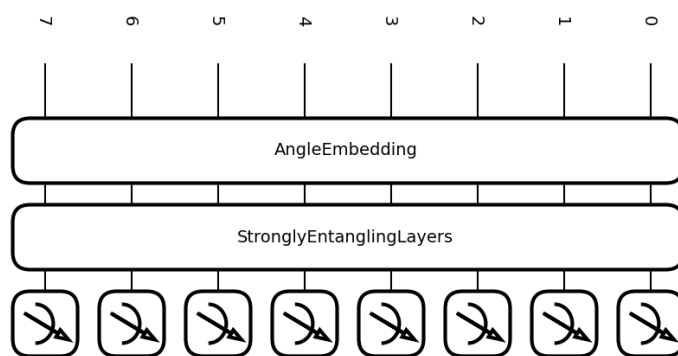


Figura 5. Representação básica do circuito quântico utilizado para montagem das camadas nas redes neurais

Na Figura 5, é possível visualizar a definição da camada presente na rede neural quântica. A implementação prática da camada da rede neural híbrida envolve entradas correspondentes aos qubits *RxEntrada*, enquanto o *AngleEmbedding* refere-se às três colunas iniciais (*rz*, *ry* e *rz theta*), com *theta* inicial gerado aleatoriamente. Os *StronglyEntanglingLayers* representam os valores intermediários (*Rz*, *Ry* e *Rz theta*), associados aos pesos da camada quântica, definidos pelo formato desses pesos. O valor final é determinado pela medida geral de todos os qubits, utilizando a expectativa do operador Pauli-Z para decodificação.

Inicialmente, há uma camada de entrada com 8 neurônios, onde cada neurônio está conectado a todos os neurônios da camada seguinte. Essa camada de entrada utiliza

a função de ativação ReLU (Rectified Linear Unit), após a camada de entrada, a camada quântica, definida anteriormente, é adicionada como uma camada intermediária na rede e então são adicionadas duas camadas intermediárias clássicas, ocultas, todas compostas por 8 neurônios cada. Por fim, há uma camada de saída densamente conectada, com 6 neurônios, correspondendo ao número de saídas desejadas.

4.2.3. Rede neural com camadas intermediárias totalmente quântica

Esta rede neural híbrida é caracterizada por uma camada intermediária composta exclusivamente por camadas quânticas. Inicialmente, há uma camada de entrada com 8 neurônios, onde cada neurônio está conectado a todos os neurônios da camada seguinte. Essa camada de entrada utiliza a função de ativação ReLU (Rectified Linear Unit), após a camada de entrada, a camada quântica, definida anteriormente, é adicionada como três camadas intermediárias na rede, ocultas, todas compostas por 8 neurônios cada. Por fim, há uma camada de saída densamente conectada, com 6 neurônios, correspondendo ao número de saídas desejadas.

5. Discussão dos Resultados e Tabelas

5.1. Explicação das Métricas

A métrica de *Loss* refere-se à diferença entre os valores previstos pelo modelo e os valores reais. Uma *Loss* baixa indica que o modelo está mais próximo de prever corretamente os resultados. Por outro lado, uma *Loss* alta significa que as previsões estão distantes dos valores verdadeiros. A *Acurácia* mede a proporção de previsões corretas em relação ao total de previsões. Uma acurácia mais alta indica um melhor desempenho preditivo. O *Erro Médio Absoluto* (MAE) calcula a média das diferenças absolutas entre as previsões e os valores reais, sendo uma métrica interpretável que indica o quanto as previsões se desviam em média dos resultados verdadeiros. O *Erro Quadrático Médio* (MSE) é semelhante ao MAE, mas penaliza erros maiores, tornando-o uma métrica mais sensível a grandes desvios. Já o *Erro Percentual Absoluto Médio* (MAPE) fornece uma perspectiva percentual do erro, facilitando a comparação entre diferentes conjuntos de dados. A métrica de *Similaridade de Cosseno* avalia o alinhamento entre os vetores de previsão e os valores reais. Valores mais próximos de 1 indicam que os vetores estão bem alinhados, sugerindo boas previsões. Finalmente, o R^2 é uma métrica estatística que indica o quão bem os dados se ajustam ao modelo. Um R^2 positivo próximo de 1 significa que o modelo explica bem a variabilidade dos dados. Valores negativos indicam um desempenho abaixo do esperado.

5.2. Comparação dos Modelos

Ao analisar os resultados da Tabela 2, observa-se que a **Rede Neural Clássica** se destaca pela consistência entre a primeira e a vigésima época, apresentando uma *Loss* e *Acurácia* superiores em relação aos outros modelos. O tempo de CPU e *Wall Time* também foram os menores, demonstrando a eficiência deste modelo tanto em termos de treinamento quanto de melhoria ao longo do tempo. Isso sugere que o modelo é eficiente para tarefas que requerem desempenho consistente ao longo das épocas.

Tabela 2. Comparação dos Resultados de Redes Neurais

Métrica	NNW Clássica (1)	NNW Clássica (20)	H-NNW (1)	H-NNW (20)	HQNNW (1)	HQNNW (20)
Loss	0,0070	0,0059	0,1805	0,0191	0,5564	0,0200
Accuracy	0,9182	0,9352	0,0386	0,9750	0,0750	0,9750
MAE	0,0517	0,0483	0,2736	0,0798	0,4784	0,0853
MSE	0,0070	0,0059	0,1805	0,0191	0,5564	0,0200
MAPE	161,5174	156,8017	319,6315	183,6465	993,0101	200,2417
MSLE	0,0041	0,0036	0,0796	0,0097	0,1039	0,0101
Cosine Similarity	0,9867	0,9886	0,5031	0,9641	-0,2423	0,9631
Logcosh	0,0034	0,0029	0,0811	0,0093	0,2055	0,0097
MSE (final)	-	0,0444	-	0,0188	-	0,0196
MAE (final)	-	0,1054	-	0,0801	-	0,0847
R^2	-	-0,5858	-	-0,0303	-	-0,3112
CPU Time	-	1,12 s	-	20min 1s	-	1h 17min 22s
Wall Time	-	4,07 s	-	20min 51s	-	1h 18min 59s

A **Rede Intermediária Parcialmente Híbrida (H-NNW)**, embora tenha começado com um desempenho inicial mais fraco, apresentou melhorias significativas até a última época. O tempo de processamento foi intermediário, o que pode ser atribuído à complexidade adicional da camada de processamento quântico e da simulação envolvida. No entanto, isso sugere que o modelo tem potencial para evoluir com mais treinamento e ajustes. A teoria sugere que, com o uso de hardware quântico real, onde os valores dos qubits seriam mapeados diretamente para o hardware, o tempo de processamento seria similar ao da Rede Neural Clássica.

Além disso, o R^2 da (H-NNW) foi o mais próximo de ser positivo entre os modelos quânticos, sendo cerca de 15 vezes maior que o R^2 da Rede Clássica e 10 vezes maior que o da **Rede Híbrida Quântica Totalmente Híbrida (HQNNW)**. Isso sugere que uma abordagem com apenas uma camada quântica intermediária consegue capturar melhor as complexidades dos dados, resultando em um ajuste superior ao dos modelos clássicos. Por outro lado, a HQNNW, embora tenha começado com um desempenho inicial melhor que a H-NNW, não apresentou melhorias significativas até a última época. O desempenho inferior em termos de *Loss* e R^2 pode ser explicado pela maior complexidade associada à simulação de múltiplas camadas quânticas, o que contribuiu para o tempo de execução mais longo. Isso é esperado, dado que simular circuitos quânticos em hardware clássico é consideravelmente mais demorado. No entanto, espera-se que, com o uso de hardware quântico real, esse tempo de processamento seja significativamente reduzido.

De forma geral, as Redes HQNNW e H-NNW apresentaram uma acurácia final e um R^2 superiores à Rede Clássica, sugerindo que as técnicas híbridas oferecem vantagens em termos de previsão, especialmente em cenários de regressão complexos. Embora as métricas de *MAE* e *MSE* indiquem que, quando os modelos híbridos erram, eles tendem a se desviar mais dos valores reais, a taxa de erro geral foi menor, o que aponta para o potencial das técnicas quânticas em melhorar o desempenho preditivo.

6. Conclusões e Trabalhos Futuros

Este estudo integrou tecnologia quântica, algoritmos de aprendizado de máquina e CRISPR para investigar a relação entre a dependência genética e o número de cópias de genes repetidos, bem como a viabilidade de redes neurais híbridas quânticas para problemas de regressão. Embora os resultados indiquem que as redes neurais clássicas

são rápidas e eficientes, os modelos híbridos e quânticos demonstram um potencial significativo, especialmente em aplicações que exigem maior precisão preditiva.

Um dos desafios encontrados durante o estudo foi a obtenção de R^2 negativo em alguns dos modelos híbridos e quânticos, o que sugere a necessidade de melhorias na modelagem e no ajuste dos parâmetros. Embora haja desafios de precisão, o estudo demonstrou a viabilidade das redes quânticas em aplicações práticas, mostrando que, com a otimização adequada, essas redes podem superar os modelos clássicos.

Referências

- Aspuru-Guzik, A. (2020). Quantum algorithms for quantum chemistry. *Chemical Reviews*, 120(22):12685–12717.
- Biamonte, J. et al. (2017). Quantum machine learning. *Nature*, 549(7671):195–202.
- Bortesi, L. and Fischer, R. (2014). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology advances*.
- Gyongyosi, L. and Imre, S. (2019). A survey on quantum computing technology. *Computer Science Review*, 31:51–71.
- Kak, S. C. (1995). Quantum neural computing. In Hawkes, P. W., editor, *Advances in Imaging and Electron Physics*, volume 94, pages 259–313. Elsevier.
- Lloyd, S. and Weedbrook, C. (2018). Quantum generative adversarial learning. *Phys. Rev. Lett.*, 121:040502.
- Mangini, S., Tacchino, F., Gerace, D., Bajoni, D., and Macchiavello, C. (2021). Quantum computing models for artificial neural networks. *Europhysics Letters*, 134(1):10002.
- Maria Schuld, I. S. and Petruccione, F. (2015). An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172–185.
- Mauranyapin, N. P., Terrasson, A., and Bowen, W. P. (2022). Quantum biotechnology. *Advanced Quantum Technologies*, 5(9):2100139.
- Narayanan, A. and Moore, M. (1996). Quantum-inspired genetic algorithms. In *Proceedings of IEEE Int. Conf. on Evolutionary Computation*, pages 61–66.
- Noshay, J. M. et al. (2023). Quantum biological insights into CRISPR-Cas9 sgRNA efficiency from explainable-AI driven feature engineering. *Nucleic Acids Research*, 51(19):10147–10161.
- Preskill, J. (2018). Quantum computing in the NISQ era and beyond. *arXiv:1801.00862 [quant-ph]*. Focus.
- Preskill, J. (2023). Quantum computing 40 years later. In *Feynman Lectures on Computation*, pages 193–244. CRC Press.
- Schuld, M., Bocharov, A., Svore, K. M., and Wiebe, N. (2020). Circuit-centric quantum classifiers. *Phys. Rev. A*, 101:032308.
- Xu, Y. and Li, Z. (2020). CRISPR-Cas systems: Overview, innovations and applications in human disease research and gene therapy. *Comp. Struct. Biotech. J.*, 18:2401–2415.