

Leonardo Spagnuolo Belluzzo

# **Quantum Machine Learning Aplicado à Detecção de Esquizofrenia**

São Paulo, SP

2024



Leonardo Spagnuolo Belluzzo

# **Quantum Machine Learning Aplicado à Detecção de Esquizofrenia**

Trabalho de conclusão de curso apresentado ao Departamento de Engenharia de Computação e Sistemas Digitais da Escola Politécnica da Universidade de São Paulo para obtenção do diploma de Bacharel em Engenharia.

Departamento de Engenharia de Computação e Sistemas Digitais (PCS)

Escola Politécnica

Universidade de São Paulo – USP

Orientadora: Profa. Dra. Regina Melo Silveira

Coorientador: Dr. Luiz Gustavo Esmenard Arruda

São Paulo, SP

2024

Gerar a ficha catalográfica em <https://www.poli.usp.br/bibliotecas/servicos/catalogacao-na-publicacao>  
Salvar o pdf e incluir na monografia





# Agradecimentos

Agradeço, em primeiro lugar, a Deus pela minha vida e pela oportunidade de escrever este trabalho.

À minha família e aos meus amigos pelo apoio constante e irrefreável, pelos exemplos de resiliência e amor e pela oportunidade de partilhar sonhos e aventuras com eles.

Aos meus orientadores pela paciência e perspicácia com as quais me conduziram neste projeto. Sem eles, este trabalho não seria possível.

A todos os docentes, pesquisadores, funcionários e alunos que fizeram parte de minha jornada de cinco longos anos na Escola Politécnica da USP.

Por fim, gostaria de saudar minha sobrinha que nasceu no mesmo dia da entrega do meu trabalho de conclusão de curso. Seja bem-vinda, Maria Eliza!



# Resumo

A esquizofrenia assola, aproximadamente, 1% da população mundial e está entre as 20 maiores causas de invalidez conhecidas (décima quarta posição) de acordo com a Organização Mundial da Saúde. O diagnóstico precoce da esquizofrenia permite a prevenção de crises e melhora a qualidade de vida do paciente a longo prazo. Técnicas de aprendizado de máquina são utilizadas no diagnóstico de diversos transtornos e doenças pela capacidade de incorporar padrões estatísticos no processo de classificação de dados. Por outro lado, a computação quântica tem características únicas como a superposição e o emaranhamento de estados. Tais características são essenciais para a construção de algoritmos quânticos os quais, em tese, são capazes de solucionar problemas computacionais mais rapidamente que os computadores clássicos. Logo, métodos de Aprendizado de Máquina Quântico podem ser utilizados para identificar indivíduos com esquizofrenia e o presente trabalho analisa as vantagens e desvantagens da utilização de técnicas de Aprendizado de Máquina Quântico no diagnóstico de esquizofrenia.

**Palavras-chave:** Computação Quântica. Aprendizado de Máquina. Aprendizado de Máquina Quântico. Esquizofrenia.



# Abstract

Schizophrenia affects nearly 1% of global population and is in the top twenty cause of disability (fourteenth place) according to the World Health Organization. The early diagnosis prevent crisis and improve the prognosis of patients long term. Machine Learning models are able to identify different disorders and illnesses by learning statistical patterns within the dataset. Quantum Computing has unique attributes like superposition and entanglement. Those attributes play a essential role in quantum algorithms that are able to surpass classical computers in terms of computational time. Therefore, quantum machine learning models are used to classify patient with schizophrenia. In this work, we analyze the advantages and disadvantages present in applying quantum machine learning algorithms in schizophrenia diagnosis.

**Keywords:** Quantum Computing, Machine Learning, Quantum Machine Learning, Schizophrenia.



# Lista de ilustrações

Figura 1 – Exemplo de árvore de decisão com um nível. . . . .	37
Figura 2 – Exemplo de árvore de decisão com dois níveis. . . . .	39
Figura 3 – Exemplo simplificado. . . . .	42
Figura 4 – Fronteira de decisão inicial do algoritmo SVM. . . . .	43
Figura 5 – Fronteira de decisão do algoritmo SVM. . . . .	44
Figura 6 – Exemplo de circuito quântico à esquerda e o circuito gerador do estado de Bell à direita. . . . .	51
Figura 7 – Representação da equação 4.17 no circuito quântico. . . . .	61
Figura 8 – Circuito quântico utilizado como <i>kernel</i> quântico. . . . .	62
Figura 9 – Circuito quântico implementando a técnica de <i>kernel</i> quântico. . . . .	63
Figura 10 – Exemplo de circuito variacional. . . . .	64
Figura 11 – Exemplo de aplicação de PCA com <i>kernel</i> quântico. . . . .	70
Figura 12 – Divisão dos dados utilizados utilizando o algoritmo SVM com <i>kernel</i> linear. . . . .	72
Figura 13 – Fronteira de Decisão do algoritmo K-NN. . . . .	87
Figura 14 – Fronteira de Decisão do algoritmo Naive Bayes. . . . .	88
Figura 15 – Fronteira de Decisão do algoritmo Árvore de Decisão. . . . .	89
Figura 16 – Fronteira de Decisão do algoritmo Floresta Aleatória. . . . .	90
Figura 17 – Fronteira de Decisão do algoritmo SVM com <i>kernel</i> linear. . . . .	91
Figura 18 – Fronteira de Decisão do algoritmo SVM com <i>kernel</i> polinomial. . . . .	92
Figura 19 – Fronteira de Decisão do algoritmo SVM com <i>kernel</i> RBF. . . . .	93
Figura 20 – Fronteira de Decisão do algoritmo SVM com <i>kernel</i> sigmoid. . . . .	94
Figura 21 – Fronteira de Decisão do algoritmo SVM com <i>kernel</i> $Z$ Feature. . . . .	95
Figura 22 – Fronteira de Decisão do algoritmo SVM com <i>kernel</i> $ZZ$ Feature. . . . .	96



# Lista de tabelas

Tabela 1 – Exemplo de dados de meteorologia adaptado de Witten et al. (2017). . . . .	28
Tabela 2 – Dado a ser classificado do exemplo de meteorologia adaptado de Witten et al. (2017). . . . .	29
Tabela 3 – Exemplo de dados com atributos numéricos de meteorologia adaptado de Witten et al. (2017). . . . .	33
Tabela 4 – Distância entre o novo dado a ser classificado e o conjunto de dados com atributos numéricos de meteorologia adaptado de Witten et al. (2017). . . . .	34
Tabela 5 – Exemplo simplificado de meteorologia adaptado de Witten et al. (2017). . . . .	41
Tabela 6 – Matriz de confusão. . . . .	45
Tabela 7 – Representação dos operadores no circuito quântico. . . . .	52
Tabela 8 – Métricas de desempenho para cada um dos métodos utilizados. . . . .	71
Tabela 9 – Matriz de confusão para o algoritmo K-NN com K igual a três. . . . .	97
Tabela 10 – Matriz de confusão para o algoritmo Naive Bayes. . . . .	97
Tabela 11 – Matriz de confusão para o algoritmo Árvore de Classificação. . . . .	98
Tabela 12 – Matriz de confusão para o algoritmo Floresta Aleatória. . . . .	98
Tabela 13 – Matriz de confusão para o algoritmo SVM com <i>Kernel</i> linear. . . . .	98
Tabela 14 – Matriz de confusão para o algoritmo SVM com <i>Kernel</i> ZFeature. . . . .	98
Tabela 15 – Matriz de confusão para o algoritmo SVM com <i>Kernel</i> ZZFeature. . . . .	98
Tabela 16 – Matriz de confusão invertida para o algoritmo SVM com <i>Kernel</i> ZZFeature. . . . .	98
Tabela 17 – Resultado de Aksoy et al. (2024) utilizando cinco atributos para o canal Cz. . . . .	99
Tabela 18 – Resultado de Aksoy et al. (2024) utilizando cinco atributos para o canal T3. . . . .	100



# Lista de abreviaturas e siglas

IRM	Imagem de Ressonância Magnética
IRMF	Imagem de Ressonância Magnética Funcional
EEG	Eletroencefalograma
FNC	Functional Network Connectivity
SBM	Source-Based Morphometry
IEEE	Institute of Electrical and Electronics Engineers
MRN	Mind Research Networ
K-NN	K-Nearest Neighbors
SVM	Support Vector Machine
QSVM	Quantum Support Vector Machine
KRR-Opt-RVFL	Kernel Ridge Regression-based Optimized Random Vector Functional Link
EB	Ensemble Bagging
ELM	Extreme Learning Machine
IC	Independent Components
GICA	Group Independent Component Analysis
PC	Principal Components
PCA	Principal Component Analysis
HHL	Harrow, Hassidim e Lloyd
FFT	Fast Fourier Transform
DWT	Discrete Wavelet Transform
RLNDiP	Relaxed Local Neighbour Difference Pattern
NISQ	Noisy Intermediate-Scale Quantum
QKE	Quantum Kernel Estimation

RBF	Radial Basis Function
COBRE	Centers of Biomedical Research Excellence
NIH	National Institutes of Health
NAND	Not And
NOR	Not Or
CNOT	Controlled Not

# Lista de símbolos

$T$	<i>Dataset</i>
$X_i$	Conjunto de atributos de um dado
$x_{i,k}$	Atributo $k$ do dado $i$
$y_i$	Rótulo do dado $i$
$\hat{y}_i$	Rótulo predito de um dado $i$
$f_b$	Classificador bayesiano
$f_{nb}$	Classificador Naive Bayes
$\mu$	Média
$\sigma$	Desvio Padrão
$d_k$	Distância entre o $k$ -ésimo dado e a amostra a ser classificada



# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>21</b>
1.1	Motivação	22
1.2	Objetivos	24
1.3	Justificativa	25
1.4	Organização do Trabalho	26
<b>2</b>	<b>MACHINE LEARNING</b>	<b>27</b>
2.1	Naive Bayes	28
2.2	K-ésimo Vizinho Mais Próximo	31
2.3	Árvore de Decisão e Floresta Aleatória	34
2.4	Support Vector Machine	39
2.5	Matriz de Confusão	45
2.6	Machine Learning Aplicado a Diagnósticos	46
<b>3</b>	<b>COMPUTAÇÃO QUÂNTICA</b>	<b>49</b>
3.1	Postulados da Mecânica Quântica	52
3.2	Computação Universal	54
3.3	Ruído na Computação Quântica	55
<b>4</b>	<b>QUANTUM MACHINE LEARNING</b>	<b>57</b>
4.1	Quantum Encoding	59
4.2	Kernel Quântico	61
4.3	Circuito Variacional Quântico	62
4.4	Trabalhos Correlatos	64
<b>5</b>	<b>DESENVOLVIMENTO DO PROJETO</b>	<b>67</b>
5.1	Procedimento Metodológico	67
5.2	Ferramentas Utilizadas	67
<b>6</b>	<b>RESULTADOS OBTIDOS E ANÁLISES</b>	<b>69</b>
<b>7</b>	<b>CONCLUSÕES E CONSIDERAÇÕES FINAIS</b>	<b>75</b>
7.1	Limitações do Trabalho	75
7.2	Perspectivas de Continuidade	76
7.3	Reconhecimento	76

<b>REFERÊNCIAS</b> . . . . .	<b>77</b>
<b>APÊNDICE A – FRONTEIRAS DE DECISÃO DOS ALGORIT- MOS ANALISADOS.</b> . . . . .	<b>87</b>
<b>APÊNDICE B – MATRIZES DE CONFUSÃO.</b> . . . . .	<b>97</b>
<b>APÊNDICE C – RESULTADOS DO TRABALHO DE AKSOY ET AL. (2024).</b> . . . . .	<b>99</b>

# 1 Introdução

A esquizofrenia afeta, aproximadamente, 1% da população mundial e está entre as 20 maiores causas de invalidez conhecidas (décima quarta posição) de acordo com a Organização Mundial da Saúde. Alguns dos sintomas da esquizofrenia são delírios, alucinações, desorganização da fala e do comportamento, apatia e deterioração do funcionamento social. O diagnóstico de esquizofrenia é dado pelo aparecimento de características como delírios, alucinações, fala desorganizada, comportamento desorganizado ou catatônico e sintomas negativos como baixa motivação e expressividade reduzida. O indivíduo deve apresentar dois dos cinco sintomas (sendo um deles um dos três primeiros da enumeração) pelo período de um mês de acordo com o manual de diagnóstico de transtornos psiquiátricos (MARDER; CANNON, 2019; WHO, 2008). Além disso, aproximadamente, 10% dos pacientes diagnosticados com esquizofrenia comentem suicídio. A esquizofrenia apresenta consequências devastadoras tanto econômicas quanto sociais e, especialmente, para os familiares do paciente (MARDER; CANNON, 2019).

Um artigo publicado em 2007 aponta que o diagnóstico precoce da esquizofrenia acompanhado de tratamento melhora seu prognóstico (LOUZÁ, 2007), ademais, o tratamento pode ser ministrado antes da ocorrência de uma crise. Há trabalhos que evidenciam o diagnóstico por meio de imagens de ressonância magnética (IRM) e Eletroencefalograma (EEG) (GOGTAY; THOMPSON, 2010; VARAPRASAD et al., 2024; CHYZHYK; SAVIO; GRAÑA, 2015). É válido ressaltar que os sintomas de esquizofrenia aparecem também em outros distúrbios psiquiátricos e, portanto, a esquizofrenia é, em muitos casos, diagnosticada por eliminação. Assim, é de suma importância desenvolver técnicas capazes de diagnosticá-las precocemente, buscando melhorar a qualidade de vida de pacientes e reduzir os impactos na sociedade (GOGTAY; THOMPSON, 2010).

Em relação às causas do distúrbio, estima-se que fatores hereditários explicam 80% do risco de esquizofrenia em uma população, parte considerável desse risco está relacionada com a expressão de genes ligados ao sistema imunológico, a plasticidade sináptica, entre outros. Sabe-se que pessoas com esquizofrenia possuem uma quantidade menor de massa cinzenta. Como é apontado por Bennett (2011), há genes candidatos que podem indicar susceptibilidade a esquizofrenia, tendo sua ação principalmente em receptores tirosina-proteína quinase erbB-4 da neuregulina (NRG1) codificados pelo gene ERBB4.

Os exames de IRM apresentam como resultados dois tipos de atributos: *Functional Network Connectivity* (FNC) e *Source-Based Morphometry* (SBM). O FNC representa as conexões entre redes independentes no cérebro ao longo do tempo. O SBM representa a

concentração de massa cinzenta em regiões do cérebro; a massa cinzenta está relacionada com a capacidade de processamento de determinada região do cérebro. Os dados advindo dos exames de IRM foram compartilhados para a realização de uma competição do *Institute of Electrical and Electronics Engineers* (IEEE) em machine learning aplicada no processamento de sinais (GOGTAY; THOMPSON, 2010; ALLEN et al., 2014).

Os atributos FNC são relações criadas entre os componentes independentes (Independent Components – IC) como forma de relacionar a atividade cerebral em determinadas regiões com outras mostrando como é a conectividade entre IC (ALLEN et al., 2014; ALLEN et al., 2011). Busca-se utilizar algoritmos de classificação com atributos FNC para classificar pacientes como saudáveis ou como portadores de esquizofrenia. Os detalhes específicos do domínio não serão abordados neste trabalho, pois fogem do seu escopo. Os atributos SBM representam a concentração de massa cinzenta em diversas regiões do cérebro compondo uma análise estrutural. O artigo de Segall et al. (2012) utiliza a densidade de massa cinzenta ao longo do cérebro como meio para relacionar e compreender características funcionais e estruturais do mesmo.

Os dados foram coletados pela rede de pesquisa Mind Research Network (MRN) que realizou exames de ressonância magnética funcional para a realização do artigo Çetin et al. (2014) que investiga as diferenças da conectividade de certas regiões no cérebro em repouso e ao engajá-los em atividades. Por outro lado, há indícios que as características estruturais do cérebro podem auxiliar no diagnóstico de esquizofrenia, algo que se mostrou verdadeiro ao realizar a seleção de características, pois dois dos três atributos selecionados no presente trabalho são estruturais (BENNETT, 2011; GOGTAY; THOMPSON, 2010).

Por fim, os exames de ressonância magnética foram realizados por um aparelho de ressonância magnética 3-Tesla Siemens Trio Scanner de doze canais. Os dados foram pré-processados utilizando o SPM-5 e o INRIalign para que os vídeos sejam alinhados e depois determina-se o *frame* médio para representar o resultado de exames de imagem de ressonância magnética funcional (IRMF). Em seguida, realiza-se o *Group Independent Component Analysis* (GICA). Para realizar o GICA, utilizou-se a *GIFT Toolbox* e o algoritmo *infomax*. Inicialmente, reduz-se o número de componentes principais (*principal components* - PC) para 100, em seguida, após a redução e concatenação dos dados, o número final de componentes principais foi de 75. Posteriormente, utiliza-se o *independent component analysis* (ICA). Uma explicação mais completa dos componentes independentes pode ser encontrada nas referências Çetin et al. (2014) e Segall et al. (2012).

## 1.1 Motivação

Considerando o contexto supracitado, a classificação de indivíduos com esquizofrenia realizada por meio de algoritmos aplicados a dados de ressonância magnética mostra-se

como um candidato para uma técnica de auxílio diagnóstico não invasivo ([AKSOY et al., 2024](#)). Essa assistência no diagnóstico é capaz de auxiliar na atenuação dos sintomas e prevenção de crises em seus pacientes, melhorando a sua condição e qualidade de vida a longo prazo. Por outro lado, essas soluções demandam grande poder computacional ou necessitam de diferentes abordagens para obterem melhores métricas de desempenho. Sendo assim, avalia-se a aplicação de computação quântica nesse cenário.

A computação quântica, por sua vez, permite o processamento de informação de forma distinta da computação clássica, em alguns casos, há algoritmos quânticos que possuem uma complexidade menor em comparação com o melhor algoritmo clássico conhecido para resolvê-lo. Tendo em vista as características particulares da computação quântica, procura-se analisar os benefícios da utilização do *kernel* quântico. Contudo, a Vantagem Quântica ainda não foi comprovada na prática, pois a computação quântica encontra-se no início de seu desenvolvimento dado que possui taxas de erro muito altas e um número ainda insuficiente de qubits ([HARROW; MONTANARO, 2017](#); [PERAL-GARCÍA; CRUZ-BENITO; GARCÍA-PEÑALVO, 2024](#); [JADHAV; RASOOL; GYANCHANDANI, 2023](#); [PRESKILL, 2018](#)).

Pode-se dividir as abordagens existentes para aprendizado de máquina em quatro tipos utilizando-se dois eixos: sistema utilizado para a coleta de dados e o sistema no qual ocorre o processamento. Logo, processa-se dados clássicos em computadores clássicos, esse conjunto é a abordagem padrão para aprendizado de máquina.

A medição de estados quânticos gera um conjunto de informações utilizadas para representar modelos de mecânica quântica na física. A simulação de sistemas quânticos em computadores clássicos permite um maior entendimento da natureza dos fenômenos. Portanto, nesse caso, o processamento dados quânticos é feito por computadores clássicos.

A utilização de computadores quânticos para processar dados clássicos é onde as aplicações de *Quantum Machine Learning* se concentram. Ou seja, os dados clássicos são representados em estados quânticos por meio de codificações que permitem seu processamento por computadores quânticos.

Por último, o processamento de dados quânticos por hardware quântico permite que em tais dispositivos de mesma natureza tenha-se maior eficiência computacional. Tal prática é chamada de simulação quântica e tem como objetivo simular sistemas quânticos complexos sobre problemas de física e química, tendo grandes aplicações na simulação de moléculas e suas interações ([SCHULD; PETRUCCIONE, 2021](#)).

Em relação a trabalhos relacionados à computação quântica, tem-se o trabalho de [Aksoy et al. \(2024\)](#) que mostra como os dados de um eletroencefalograma são utilizados para realizar o diagnóstico de esquizofrenia por um algoritmo quântico. O trabalho supracitado relatou obter uma acurácia de 100% com um kernel quântico, ademais, o resultado é

comparado com outros algoritmos de *Machine Learning* como *K-Nearest Neighbors* (K-NN), Árvores de Decisão, Floresta Aleatória, *Naive Bayes* e *Support Vector Machine* (SVM).

Uma diferença crucial entre o artigo supracitado e o presente trabalho é a fonte dos dados. Os dados presentes em [Aksoy et al. \(2024\)](#) advêm de um estudo da faculdade estadual de Moscou com eletroencefalogramas enquanto este trabalho utiliza dados coletados pelo MRN por meio de ressonância magnética. Além do mais, os dados usados no presente trabalho foram pré-processados de acordo com o artigo [Çetin et al. \(2014\)](#) e disponibilizados em uma competição do IEEE no site Kaggle <sup>1</sup>. Infelizmente, não foi possível encontrar um conjunto com todos os dados devidamente classificados para a utilização de técnicas de aprendizado supervisionado.

Em relação a trabalhos que abordam o problema de diagnóstico utilizando redes neurais, é válido ressaltar o trabalho de [Varaprasad et al. \(2024\)](#) que apresenta uma aplicação da ResNet50 na seleção de características de imagens de ressonância magnética em outra tarefa de diagnóstico de esquizofrenia. A ResNet50 é uma rede convolucional residual cuja arquitetura é descrita no trabalho de [He et al. \(2016\)](#). O artigo compara a aplicação da *Kernel Ridge Regression-based Optimized Random Vector Functional Link* (KRR-Opt-RVFL) com outras técnicas como SVM, K-NN, Árvore de Decisão, Floresta Aleatória, *Ensemble Bagging* (EB), Extreme Learning Machine (ELM) ([CHYZHYK; SAVIO; GRAÑA, 2015](#)). Contudo, o presente trabalho não utiliza nenhuma técnica de rede neural para o problema de classificação. A técnica de redes neurais não é implementada, pois os dados do *dataset*, em sua maioria, são linearmente separáveis o que indica o uso de técnicas como Árvore de Decisão e Floresta Aleatória.

Por fim, tem-se que, em sua maioria, algoritmos baseados em árvores de decisão como o XGBoost, descrito no trabalho de [Chen e Guestrin \(2016\)](#), continuam sendo o estado da arte em problemas de classificação para dados tabulares ([GRINSZTAJN; OYALLON; VAROQUAUX, 2022](#)). Portanto, é necessário avaliar, principalmente, métodos que são derivados de árvores de decisão.

## 1.2 Objetivos

O objetivo principal deste trabalho de conclusão de curso é investigar o desempenho de técnicas de aprendizado de máquina quânticas (como o kernel quântico utilizado no SVM) e se são capazes de melhorar a capacidade preditiva do modelo em comparação com técnicas clássicas. Para isso, tem-se tanto os atributos funcionais quanto os estruturais para classificar utilizando técnicas de aprendizado de máquina e computação quântica se um paciente tem esquizofrenia ou não. Ademais, os resultados serão comparados com outros trabalhos recentes ([AKSOY et al., 2024](#)) que apresentaram bom desempenho utilizando

<sup>1</sup> Kaggle, disponível em: <https://www.kaggle.com/>. Acesso em: 10 dez. 2024.

*kernel* quântico resolvendo um problema semelhante, porém, com um conjunto de dados diferentes.

Apesar da inovação em realizar diagnósticos utilizando computação quântica, alguns autores mostram que algoritmos quânticos derivados de algoritmos clássicos apresentam uma acurácia menor que em sua versão clássica e que, raramente, possuem o mesmo desempenho (PERAL-GARCÍA; CRUZ-BENITO; GARCÍA-PEÑALVO, 2024). Por outro lado, há trabalhos que descrevem indícios de ganho de desempenho e de diminuição da complexidade computacional com o uso de computadores quânticos (JADHAV; RASOOL; GYANCHANDANI, 2023; ENAD; MOHAMMED et al., 2023; ULLAH; GARCIA-ZAPIRAIN, 2024). Assim, esse trabalho busca entender se o método de SVM com os *kernels* quânticos é capaz de obter um desempenho melhor que algoritmos clássicos dado que foi citado como um dos algoritmos de maior acurácia no trabalho de Aksoy et al. (2024). É válido ressaltar que alguns trabalhos divergem sobre a utilização de computadores quânticos em aprendizado de máquina, pois, no melhor dos casos, os algoritmos quânticos ou híbridos apresentam uma acurácia igual aos algoritmos clássicos (PERAL-GARCÍA; CRUZ-BENITO; GARCÍA-PEÑALVO, 2024).

### 1.3 Justificativa

Por meio da superposição coerente de estados quânticos e emaranhamento, em tese, é possível obter um melhor desempenho em computadores quânticos para determinados problemas. Não apenas o desempenho é uma característica relevante ao comparar-se um computador clássico e um computador quântico, mas também a fidelidade presente na simulação de fenômenos quânticos, pois o trabalho de FEYNMAN (1982) discute os motivos pelos quais sistemas quânticos não poderiam ser simulados eficientemente em computadores clássicos. Logo, busca-se encontrar algoritmos que sejam capazes tanto de resolver o problema proposto quanto de se beneficiar das características especiais do domínio da computação quântica para obter melhor desempenho.

É válido ressaltar que, ainda que não se possa provar por completo que a computação quântica não pode ser simulada por um computador clássico, é possível discutir a complexidade computacional dos algoritmos utilizados para resolver um determinado problema. Sendo assim, a teoria da Complexidade Computacional passa a ser uma das principais ferramentas para comparar o desempenho entre um computador clássico e um computador quântico (HARROW; MONTANARO, 2017).

Além disso, a efetividade de algoritmos quânticos na resolução de problemas de Machine Learning é corroborada pelos trabalhos de Aksoy et al. (2024) e Havlicek et al. (2019) os quais demonstram que a técnica Quantum Support Vector Machine (QSVM) obtém uma acurácia de 100% nos problemas apresentados. Destarte, os algoritmos quânticos

podem apresentar diversas vantagens em comparação com algoritmos clássicos o que reforça a importância desse trabalho.

Seguindo a motivação deste trabalho, o surgimento de um método não invasivo para diagnóstico por meio de IRM e EEG contribuirá para a detecção e o tratamento de indivíduos que tenham pré-disposição à esquizofrenia. Logo, avalia-se também a aplicabilidade de técnicas de Quantum Machine Learning na identificação de indivíduos com esquizofrenia.

Portanto, busca-se comparar o desempenho entre algoritmos clássicos e o QSVM com o intuito de verificar as métricas de desempenho de modelos que utilizam computação quântica.

## 1.4 Organização do Trabalho

No capítulo 2, define-se os algoritmos de *machine learning Naive Bayes*, K-NN, árvore de decisão, floresta aleatória e *support vector machine*. O capítulo 3 apresenta os postulados da mecânica quântica, as operações realizadas e exemplos de circuitos quânticos. Relacionado com as aplicações de computação quântica e *machine learning*, o capítulo 4 apresenta as formas de codificação de atributos. Em especial, tem-se o *kernel* quântico, parte essencial do algoritmo QSVM. O capítulo 5 trata do método proposto no trabalho, dos requisitos dos algoritmos e, por último, das ferramentas utilizadas. O capítulo 6 discorrerá sobre os resultados obtidos e análises relacionadas às fronteiras de decisão. Por último, o capítulo de 7 tecerá considerações sobre o desempenho e o uso de algoritmos de *Quantum Machine Learning* para diagnóstico de transtornos psiquiátricos como a esquizofrenia.

## 2 Machine Learning

Machine Learning é uma subárea de Inteligência Artificial definida pelo aprendizado inerente aos algoritmos. Há quatro categorias de aprendizado de máquina: aprendizado não-supervisionado; aprendizado por reforço; aprendizado supervisionado e aprendizado semissupervisionado (NORVIG; RUSSELL, 2014),(CHAPELLE; SCHOLKOPF; ZIEN, 2006).

O aprendizado não-supervisionado é definido por um conjunto de exemplos não classificados previamente. Busca-se agrupar os exemplos em “clusters” levando-se em consideração a semelhança entre seus atributos. A semelhança é inversamente proporcional a medida de distância. Algumas distâncias presentes na literatura são: euclidiana, Manhattan e Mahalanobis (NORVIG; RUSSELL, 2014).

O aprendizado por reforço é dado no contexto de recompensas e punições, por exemplo, suponha que um agente é avaliado de acordo com uma função chamada de função de utilidade onde o objetivo do agente é maximizá-la. Para cada ação que o faz chegar mais próximo do objetivo final, o agente recebe uma recompensa, ou seja, o valor da função de utilidade aumenta. Caso o agente realize ações que o impedem de completar seu objetivo, tal agente recebe uma punição, ou seja, o valor da função de utilidade diminui. Portanto, os reforços tanto negativos quanto positivos alteram as ações tomadas pelo agente para que se alcance o maior valor da função de utilidade (NORVIG; RUSSELL, 2014).

No aprendizado supervisionado, o modelo observa os pares entrada e saída de um determinado fenômeno e precisa aproximar a função que liga o domínio de entrada com o contradomínio da saída (NORVIG; RUSSELL, 2014). Por último, o aprendizado semi-supervisionado é utilizado quando há um grande volume de dados dos quais somente parte estão classificados. Sendo assim, cria-se um modelo não-supervisionado cuja função é classificar os dados que não possuem uma classe para que um outro modelo realize o aprendizado supervisionado (CHAPELLE; SCHOLKOPF; ZIEN, 2006).

Considerando o exemplo da tabela 1 em relação ao aprendizado supervisionado, os atributos do problema são “visual”, “temperatura”, “umidade” e “ventando” e o atributo meta, ou seja, o atributo que se quer descobrir em novos dados é “brincar”. Observa-se que algoritmos de Machine Learning utilizados em problemas de aprendizado supervisionado buscam relacionar seus atributos com seu rótulo. A aplicação dos algoritmos mostrados será demonstrado nas seções posteriores.

Tabela 1 – Exemplo de dados de meteorologia adaptado de [Witten et al. \(2017\)](#).

Visual	Temperatura	Umidade	Ventando	Brincar
Ensolarado	Quente	Alta	Falso	Não
Ensolarado	Quente	Alta	Verdadeiro	Não
Nublado	Quente	Alta	Falso	Sim
Chuvoso	Amena	Alta	Falso	Sim
Chuvoso	Fria	Normal	Falso	Sim
Chuvoso	Fria	Normal	Verdadeiro	Não
Nublado	Fria	Normal	Verdadeiro	Sim
Ensolarado	Amena	Alta	Falso	Não
Ensolarado	Fria	Normal	Falso	Sim
Chuvoso	Amena	Normal	Falso	Sim
Ensolarado	Amena	Normal	Verdadeiro	Sim
Nublado	Amena	Alta	Verdadeiro	Sim
Nublado	Quente	Normal	Falso	Sim
Chuvoso	Amena	Alta	Verdadeiro	Não

Fonte: [Witten et al. \(2017\)](#).

## 2.1 Naive Bayes

O nome Naive Bayes refere-se a um conjunto de classificadores os quais utilizam a teoria de Bayes ([NORVIG; RUSSELL, 2014](#)). O termo “naive” pressupõe a independência condicional entre cada característica e sua classe. Inicialmente, considere dois eventos  $A$  e  $B$  independentes entre si, tem-se que a probabilidade da ocorrência de ambos os eventos dado um terceiro evento  $C$  é:

$$p(A \cap B | C) = p(A | C) p(B | C) \quad (2.1)$$

Supondo que  $X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,k}\}$  e sua classe seja dada por  $y_i$ , define-se o *dataset* como  $T = \{(X_i, y_i)\}_{i=1}^m$  sendo  $m$  o número de exemplos presentes no conjunto de dados. Sendo assim, supondo que cada característica  $x_{i,j}$  seja independente em relação a cada característica do conjunto  $\{x_{i,1}, x_{i,2}, \dots, x_{i,j-1}, x_{i,j+1}, \dots, x_{i,k}\}$ , tem-se:

$$p(X_i | y) = \prod_{j=1}^k p(x_{i,j} | y) \quad (2.2)$$

A regra de Bayes define a probabilidade de uma determinada amostra  $X_i$  ser classificado como pertencente a classe  $y = 1$ . A probabilidade é dada por:

$$p(y_i | X_i) = \frac{p(X_i | y) p(y)}{p(X_i)} \quad (2.3)$$

Portanto, um classificador bayesiano  $f_b$  tem como classe predita  $\hat{y}$  para um conjunto de características  $X_i$  de acordo com:

Tabela 2 – Dado a ser classificado do exemplo de meteorologia adaptado de [Witten et al. \(2017\)](#).

Visual	Temperatura	Umidade	Ventando	Brincar
Ensolarado	Quente	Alta	Falso	Não
Ensolarado	Quente	Alta	Verdadeiro	Não
Nublado	Quente	Alta	Falso	Sim
Chuvoso	Amena	Alta	Falso	Sim
Chuvoso	Fria	Normal	Falso	Sim
Chuvoso	Fria	Normal	Verdadeiro	Não
Nublado	Fria	Normal	Verdadeiro	Sim
Ensolarado	Amena	Alta	Falso	Não
Ensolarado	Fria	Normal	Falso	Sim
Chuvoso	Amena	Normal	Falso	Sim
Ensolarado	Amena	Normal	Verdadeiro	Sim
Nublado	Amena	Alta	Verdadeiro	Sim
Nublado	Quente	Normal	Falso	Sim
Chuvoso	Amena	Alta	Verdadeiro	Não
Chuvoso	Fria	Alta	Verdadeiro	?

Fonte: [Witten et al. \(2017\)](#).

$$\hat{y} = f_b(X_i) = \begin{cases} 1, & \text{se } \frac{p(y=1|X_i)}{p(y=-1|X_i)} > 1, \\ -1, & \text{caso contrário.} \end{cases} \quad (2.4)$$

Logo, o classificador Naive Bayes  $f_{nb}$  é dado por:

$$\hat{y} = f_{nb}(X_i) = \begin{cases} 1, & \text{se } \frac{p(y=1)}{p(y=-1)} \prod_{j=1}^k \frac{p(x_{ij}|y=1)}{p(x_{ij}|y=-1)} > 1, \\ -1, & \text{caso contrário.} \end{cases} \quad (2.5)$$

Por fim, como as características  $x_{i,j}$  são numéricas, utiliza-se uma distribuição gaussiana para avaliar a probabilidade  $p(x_{ij}|y)$ :

$$p(x_{i,j}|y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_j-\mu)^2}, \quad (2.6)$$

sendo que  $\mu$  representa a média e  $\sigma$  o desvio padrão da distribuição. Os conceitos desta seção advêm dos trabalhos de [Zhang \(2004\)](#), [Norvig e Russell \(2014\)](#), [Barber \(2012\)](#).

A aplicação do classificador Naive Bayes é dado definindo as probabilidades  $p(y)$ ,  $p(x)$  e  $p(x_{i,j}|y)$ , considerando uma nova amostra presente na tabela 2, tem-se as seguinte probabilidades:

$$p(y=1) = \frac{9}{14} \text{ e } p(y=-1) = \frac{5}{14}, \quad (2.7)$$

considerando:

$$\text{Brincar} = \begin{cases} \text{Não} \rightarrow -1, \\ \text{Sim} \rightarrow 1. \end{cases} \quad (2.8)$$

A probabilidade de  $y$  dado o atributo visual é:

$$p(\text{Visual}|y = 1) = \begin{cases} \text{se Ensolarado, } p(\text{Ensolarado}|y = 1) = \frac{2}{9}, \\ \text{se Nublado, } p(\text{Nublado}|y = 1) = \frac{4}{9}, \\ \text{se Chuvoso, } p(\text{Chuvoso}|y = 1) = \frac{3}{9}. \end{cases} \quad (2.9)$$

$$p(\text{Visual}|y = -1) = \begin{cases} \text{se Ensolarado, } p(\text{Ensolarado}|y = -1) = \frac{3}{5}, \\ \text{se Nublado, } p(\text{Nublado}|y = -1) = \frac{0}{5}, \\ \text{se Chuvoso, } p(\text{Chuvoso}|y = -1) = \frac{2}{5}. \end{cases} \quad (2.10)$$

Em relação ao atributo temperatura, tem-se:

$$p(\text{Temperatura}|y = 1) = \begin{cases} \text{se Fria, } p(\text{Fria}|y = 1) = \frac{3}{9}, \\ \text{se Amena, } p(\text{Amena}|y = 1) = \frac{4}{9}, \\ \text{se Quente, } p(\text{Quente}|y = 1) = \frac{2}{9}. \end{cases} \quad (2.11)$$

$$p(\text{Temperatura}|y = -1) = \begin{cases} \text{se Fria, } p(\text{Fria}|y = -1) = \frac{1}{5}, \\ \text{se Amena, } p(\text{Amena}|y = -1) = \frac{2}{5}, \\ \text{se Quente, } p(\text{Quente}|y = -1) = \frac{2}{5}. \end{cases} \quad (2.12)$$

O mesmo se faz para o atributo umidade:

$$p(\text{Umidade}|y = 1) = \begin{cases} \text{se Normal, } p(\text{Normal}|y = 1) = \frac{6}{9}, \\ \text{se Alta, } p(\text{Alta}|y = 1) = \frac{3}{9}. \end{cases} \quad (2.13)$$

$$p(\text{Umidade}|y = -1) = \begin{cases} \text{se Normal, } p(\text{Normal}|y = -1) = \frac{1}{5}, \\ \text{se Alta, } p(\text{Alta}|y = -1) = \frac{4}{5}. \end{cases} \quad (2.14)$$

Para o atributo ventando:

$$p(\text{Ventando}|y = 1) = \begin{cases} \text{se Verdadeiro, } p(\text{Verdadeiro}|y = 1) = \frac{3}{9}, \\ \text{se Falso, } p(\text{Falso}|y = 1) = \frac{6}{9}, \end{cases} \quad (2.15)$$

$$p(\text{Ventando}|y = -1) = \begin{cases} \text{se Verdadeiro, } p(\text{Verdadeiro}|y = -1) = \frac{3}{5}, \\ \text{se Falso, } p(\text{Falso}|y = -1) = \frac{2}{5}, \end{cases} \quad (2.16)$$

por último, a probabilidade  $p(X_i)$  será dada pela multiplicação das probabilidades de cada atributo independente de seu rótulo, como mostrado a seguir:

$$p(X_i) = p(\text{Chuvoso}) p(\text{Fria}) p(\text{Alta}) p(\text{Verdadeiro}). \quad (2.17)$$

Portanto, pode-se classificar o novo dado da tabela 2. A probabilidade  $p(y = 1|X_i)$  é dada por:

$$\frac{p(\text{Chuvoso}|y = 1) p(\text{Fria}|y = 1) p(\text{Alta}|y = 1) p(\text{Verdadeiro}|y = 1) p(y = 1)}{p(\text{Chuvoso}) p(\text{Fria}) p(\text{Alta}) p(\text{Verdadeiro})}, \quad (2.18)$$

Além disso, a probabilidade  $p(y = -1|X_i)$  é dada por:

$$\frac{p(\text{Chuvoso}|y = -1) p(\text{Fria}|y = -1) p(\text{Alta}|y = -1) p(\text{Verdadeiro}|y = -1) p(y = -1)}{p(\text{Chuvoso}) p(\text{Fria}) p(\text{Alta}) p(\text{Verdadeiro})}. \quad (2.19)$$

Sendo assim, pode-se simplificar a expressão  $\frac{p(y=1|X_i)}{p(y=-1|X_i)}$  para

$$\frac{p(\text{Chuvoso}|y = 1) p(\text{Fria}|y = 1) p(\text{Alta}|y = 1) p(\text{Verdadeiro}|y = 1) p(y = 1)}{p(\text{Chuvoso}|y = -1) p(\text{Fria}|y = -1) p(\text{Alta}|y = -1) p(\text{Verdadeiro}|y = -1) p(y = -1)} \quad (2.20)$$

cujo valor é

$$\frac{p(y = 1|X_i)}{p(y = -1|X_i)} = \frac{\frac{3}{9} \frac{3}{9} \frac{6}{9} \frac{3}{9} \frac{9}{14}}{\frac{2}{5} \frac{1}{5} \frac{4}{5} \frac{3}{5} \frac{5}{14}} \approx 1,1574. \quad (2.21)$$

Como o valor  $\frac{p(y=1|X_i)}{p(y=-1|X_i)} > 1$ , o rótulo atribuído ao dado é 1 ou “sim”.

## 2.2 K-ésimo Vizinho Mais Próximo

O K-NN é um método não parametrizado de Machine Learning o qual utiliza dados com características semelhantes para classificar um novo dado. No caso, a semelhança é similar àquela utilizada no aprendizado não-supervisionado, no qual mede-se a diferença numérica entre os atributos. Como o algoritmo K-NN utiliza tanto as características quanto os rótulos já conhecidos para classificar um novo dado, seu aprendizado é supervisionado. Por fim, a sigla K representa o número de vizinhos mais próximos que serão considerados

na classificação do novo dado. Tal classificação é feita por meio de maioria simples. Caso haja empate, decide-se de forma aleatória a classe do exemplo em questão (FIX, 1985; COVER; HART, 1967).

Considere um conjunto  $A$ , define-se distância como uma função tal que  $d : A \times A \rightarrow \mathbb{R}$ , para todo  $b, c \in A$ , a medida de distância deve satisfazer três propriedades (DEZA et al., 2009):

- $d(b, c) \geq 0$  (não-negatividade);
- $d(b, c) = d(c, b)$  (simetria);
- $d(b, b) = 0$  (reflexividade).

Considere a notação do dado  $X_i$  apresentada na seção anterior. No presente trabalho, usa-se a distância euclidiana entre o dado e o restante do *dataset* para cada exemplo  $l$  como:

$$d_l = \sqrt{\sum_{j=1}^k (x_{i,j} - x_{l,j})^2}. \quad (2.22)$$

Sendo assim, após calcular a distância  $d_l$  para cada um dos dados do conjunto de treinamento, seleciona-se os rótulos das  $K$  menores distâncias encontradas e atribui-se ao dado  $i$  o rótulo em maior quantidade (COVER; HART, 1967). Distâncias diferentes como Manhattan e Mahalanobis também podem ser utilizadas, contudo, não são apresentadas no presente trabalho.

Considerando o conjunto de dados da tabela 1, suponha que um novo exemplo seja classificado utilizando o algoritmo K-NN. O exemplo em questão tem como atributos chuvoso, frio, alta umidade e ventando. Primeiramente, o algoritmo K-NN utiliza atributos numéricos para calcular a distância, assim, é necessário traduzir os atributos categóricos em atributos numéricos. Atributos categóricos que não são sequenciais podem ser substituídos do seguinte modo:

$$\text{Ventando} = \begin{cases} \text{Falso} \rightarrow 0, \\ \text{Verdadeiro} \rightarrow 1. \end{cases} \quad (2.23)$$

Porém, parte dos atributos categóricos apresentam uma ordem entre os valores, por exemplo:

Tabela 3 – Exemplo de dados com atributos numéricos de meteorologia adaptado de Witten et al. (2017).

Visual	Temperatura	Umidade	Ventando	Brincar
1	3	2	0	Não
1	3	2	1	Não
2	3	2	0	Sim
3	2	2	0	Sim
3	1	1	0	Sim
3	1	1	1	Não
2	1	1	1	Sim
1	2	2	0	Não
1	1	1	0	Sim
3	2	1	0	Sim
1	2	1	1	Sim
2	2	2	1	Sim
2	3	1	0	Sim
3	2	2	1	Não
3	1	2	1	?

Fonte: Witten et al. (2017).

$$\text{Visual} = \begin{cases} \text{Ensolarado} \rightarrow 1, \\ \text{Nublado} \rightarrow 2, \\ \text{Chuvoso} \rightarrow 3. \end{cases} \quad (2.24)$$

Sendo assim, tem-se a seguinte transformação nos atributos restantes:

$$\text{Temperatura} = \begin{cases} \text{Fria} \rightarrow 1, \\ \text{Amena} \rightarrow 2, \\ \text{Quente} \rightarrow 3. \end{cases} \quad (2.25)$$

$$\text{Umidade} = \begin{cases} \text{Normal} \rightarrow 1, \\ \text{Alta} \rightarrow 2. \end{cases} \quad (2.26)$$

Portanto, é possível calcular a distância que o dado a ser classificado tem para cada item da tabela. Um exemplo sobre o cálculo da distância na equação 2.27, tem-se o cálculo da distância entre a primeira amostra da tabela e o dado a ser classificado  $x_i$ :

$$d_1 = \sqrt{\sum_{j=1}^k (x_{i,j} - x_{1,j})^2} = \sqrt{(3-1)^2 + (1-3)^2 + (2-2)^2 + (1-0)^2} = 3. \quad (2.27)$$

Considera-se os três vizinhos mais próximos para a classificação da nova amostra, ou seja, os dois vizinhos de distância 1 e um dos vizinhos de distância 1,41. Percebe-se que

Tabela 4 – Distância entre o novo dado a ser classificado e o conjunto de dados com atributos numéricos de meteorologia adaptado de [Witten et al. \(2017\)](#).

Visual	Temperatura	Umidade	Ventando	Distância	Brincar
1	3	2	0	3,00	Não
1	3	2	1	2,83	Não
2	3	2	0	2,45	Sim
3	2	2	0	1,41	Sim
3	1	1	0	1,41	Sim
3	1	1	1	1,00	Não
2	1	1	1	1,41	Sim
1	2	2	0	2,45	Não
1	1	1	0	2,45	Sim
3	2	1	0	1,73	Sim
1	2	1	1	2,45	Sim
2	2	2	1	1,41	Sim
2	3	1	0	2,65	Sim
3	2	2	1	1,00	Não
3	1	2	1	0,00	?

Fonte: [Witten et al. \(2017\)](#).

há quatro amostras que possuem uma distância de 1,41 para o caso a ser classificado, contudo, os dois vizinhos mais próximos com distância 1 formam maioria simples para uma votação de três vizinhos. Sendo assim, temos como resposta o rótulo “não” para o atributo meta “brincar”. Contudo, é válido ressaltar que o número de vizinhos mais próximos é um hiperparâmetro do algoritmo e, no caso de dois ou mais vizinhos terem a mesma distância, sorteia-se o vizinho a ser considerado com probabilidades equiprováveis entre si.

## 2.3 Árvore de Decisão e Floresta Aleatória

O algoritmo de árvores de decisão procura diminuir a desordem dos dados. Neste caso, utiliza-se a medida de desordem conhecida como índice de Gini. O índice de Gini pode ser calculado pela seguinte fórmula:

$$\text{Gini}(x) = 1 - \sum_{i=1}^k p_i^2. \quad (2.28)$$

Sendo  $p_i$  a proporção de elementos da classe  $i$  e  $k$  o número de classes. Um dos pontos negativos da árvore de classificação é o viés advindo do sobreajuste do modelo ao conjunto de dados, portanto, a quantidade de níveis da árvore é um hiperparâmetro que altera a capacidade preditiva do modelo. Tal alteração acontece porque árvores mais profundas, ou seja, com mais níveis, tendem a diminuir sua capacidade de generalização dos dados quanto maior sua profundidade.

Uma técnica utilizada para melhorar a capacidade preditiva da árvore de classificação chama-se Bagging (*Bootstrap Aggregating*). Nessa técnica, sorteia-se, com reposição, amostras do conjunto total de dados criando subconjuntos, ajusta-se uma árvore de classificação para cada subconjunto e, para a classificação, considera-se o resultado de cada árvore como um voto para a classe resultante, a classe com a maior quantidade de votos é dada como a resposta do modelo (NORVIG; RUSSELL, 2014; BREIMAN, 2001).

Além da técnica de Bagging, realiza-se a estimativa do erro de generalização por meio de um conjunto de modelos. Suponha um dado  $(X_i, y_i)$  do *dataset* e seja  $T$  um subconjunto do *dataset* original. Para avaliar o dado  $(X_i, y_i)$ , utiliza-se modelos que foram treinados em subconjuntos  $T$  tal que  $(X_i, y_i) \notin T$ . Sendo assim, o erro de generalização é dado pelo número de elementos classificados corretamente (BREIMAN, 2001).

Novamente, considere o exemplo estabelecido na tabela 2. Para cada atributo, procura-se aquele que apresenta uma menor taxa de erro quando utilizado para classificar o conjunto de dados. Para cada atributo, calcula-se a probabilidade de erro de classificação. Ao observar a probabilidade  $p(y|Atributo)$ , escolhe-se a maior probabilidade como a classe a ser atribuída a um novo dado.

$$p(y = 1|Visual) = \begin{cases} \text{se Ensolarado, } \frac{2}{5}, \\ \text{se Nublado, } \frac{4}{4}, \\ \text{se Chuvoso, } \frac{3}{5}. \end{cases} \quad (2.29)$$

$$p(y = -1|Visual) = \begin{cases} \text{se Ensolarado, } \frac{3}{5}, \\ \text{se Nublado, } \frac{0}{4}, \\ \text{se Chuvoso, } \frac{2}{5}. \end{cases} \quad (2.30)$$

Dado a fórmula 2.28, o cálculo do índice de Gini para o conjunto de dados é

$$\text{Gini}(\{(X_i, y_i)\}_{i=1}^m) = 1 - \left(\frac{5}{14}\right)^2 - \left(\frac{9}{14}\right)^2 = \frac{90}{196}. \quad (2.31)$$

Para o atributo “visual”, é necessário calcular o índice de Gini para cada valor desse atributo:

$$\text{Gini(Ensolarado)} = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 = \frac{12}{25}; \quad (2.32)$$

$$\text{Gini(Nublado)} = 1 - \left(\frac{0}{4}\right)^2 - \left(\frac{4}{4}\right)^2 = 0; \quad (2.33)$$

$$\text{Gini(Chuvoso)} = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 = \frac{12}{25}. \quad (2.34)$$

Por fim, o ganho de informação do atributo “visual”:

$$\text{Ganho de Informação (GI)} = \frac{90}{196} - \frac{12}{25} \frac{5}{14} - 0 \frac{4}{14} - \frac{12}{25} \frac{5}{14} \approx 0,1163. \quad (2.35)$$

Portanto, é necessário realizar o mesmo cálculo para cada atributo e escolher o atributo que possui o maior GI. Para o atributo “temperatura”, tem-se:

$$\text{Gini(Fria)} = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = \frac{6}{16}; \quad (2.36)$$

$$\text{Gini(Amena)} = 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 = \frac{16}{36}; \quad (2.37)$$

$$\text{Gini(Quente)} = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = \frac{8}{16}. \quad (2.38)$$

O ganho de informação é

$$\text{Ganho de Informação (GI)} = \frac{90}{196} - \frac{6}{16} \frac{4}{14} - \frac{16}{36} \frac{6}{14} - \frac{8}{16} \frac{4}{14} \approx 0,0187. \quad (2.39)$$

Para o atributo “Umidade”, tem-se:

$$\text{Gini(Normal)} = 1 - \left(\frac{6}{7}\right)^2 - \left(\frac{1}{7}\right)^2 = \frac{12}{49}; \quad (2.40)$$

$$\text{Gini(Alta)} = 1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2 = \frac{24}{49}. \quad (2.41)$$

O ganho de informação é:

$$\text{Ganho de Informação (GI)} = \frac{90}{196} - \frac{12}{49} \frac{7}{14} - \frac{24}{49} \frac{7}{14} \approx 0,0918. \quad (2.42)$$

Por último, para o atributo “Ventando”:

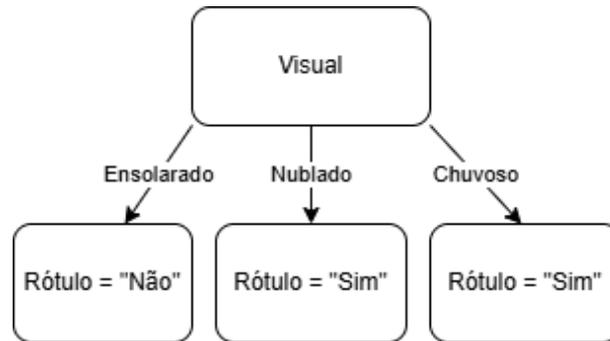
$$\text{Gini(Falso)} = 1 - \left(\frac{6}{8}\right)^2 - \left(\frac{2}{8}\right)^2 = \frac{24}{64}; \quad (2.43)$$

$$\text{Gini(Verdadeiro)} = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = \frac{18}{36}. \quad (2.44)$$

O ganho de informação é:

$$\text{Ganho de Informação (GI)} = \frac{90}{196} - \frac{24}{64} \frac{8}{14} - \frac{18}{36} \frac{6}{14} \approx 0,0306. \quad (2.45)$$

Figura 1 – Exemplo de árvore de decisão com um nível.



Fonte: De autoria própria.

Logo, o maior ganho de informação é obtido utilizando o atributo “visual” para separar os dados como pode ser visto na figura 1. Para o segundo nível da árvore de decisão, é necessário fazer o mesmo processo para as novas folhas. O índice de Gini para o atributo “temperatura” é:

$$\text{Gini(Fria)} = 1 - \left(\frac{1}{1}\right)^2 - \left(\frac{0}{1}\right)^2 = 0; \quad (2.46)$$

$$\text{Gini(Amena)} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{1}{2}; \quad (2.47)$$

$$\text{Gini(Quente)} = 1 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2 = 0. \quad (2.48)$$

O ganho de informação é:

$$\text{Ganho de Informação (GI)} = \frac{12}{25} - \frac{1}{2} \frac{2}{5} = 0,2800. \quad (2.49)$$

Para o atributo “Umidade”:

$$\text{Gini(Normal)} = 1 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2 = 0; \quad (2.50)$$

$$\text{Gini(Alta)} = 1 - \left(\frac{0}{3}\right)^2 - \left(\frac{3}{3}\right)^2 = 0. \quad (2.51)$$

O ganho de informação é:

$$\text{Ganho de Informação (GI)} = \frac{12}{25} = 0,4800. \quad (2.52)$$

Para o atributo “Ventando”:

$$\text{Gini(Falso)} = 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2 = \frac{4}{9}; \quad (2.53)$$

$$\text{Gini(Verdadeiro)} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{1}{2}. \quad (2.54)$$

O ganho de informação é:

$$\text{Ganho de Informação (GI)} = \frac{12}{25} - \frac{4}{9} \frac{3}{5} - \frac{1}{2} \frac{2}{5} \approx 0,0133. \quad (2.55)$$

Portanto, utiliza-se o atributo “Umidade” na folha com o atributo “visual” de valor “Ensolarado”. Como o índice de Gini para o valor “Nublado” é zero, não é necessário ter outro nó para dividir novamente o conjunto de dados. Por último, para o nó de valor “Chuvoso”:

$$\text{Gini(Fria)} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{1}{2}; \quad (2.56)$$

$$\text{Gini(Amena)} = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 = \frac{4}{9}; \quad (2.57)$$

$$\text{Gini(Quente)} = 1. \quad (2.58)$$

O ganho de informação é:

$$\text{Ganho de Informação (GI)} = \frac{12}{25} - \frac{1}{2} \frac{2}{5} - \frac{4}{9} \frac{3}{5} \approx 0,0133. \quad (2.59)$$

Para o atributo “Umidade”, tem-se:

$$\text{Gini(Normal)} = 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2 = \frac{4}{9}; \quad (2.60)$$

$$\text{Gini(Alta)} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{1}{2}. \quad (2.61)$$

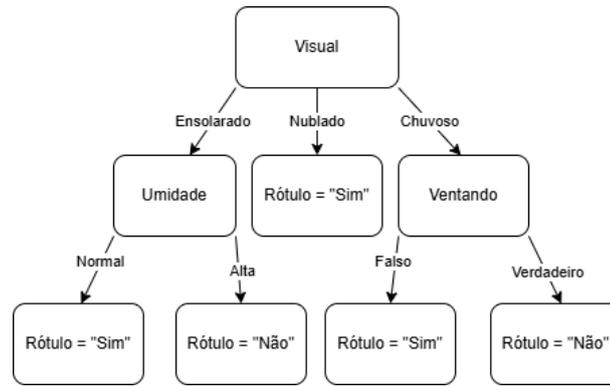
O ganho de informação é:

$$\text{Ganho de Informação (GI)} = \frac{12}{25} - \frac{4}{9} \frac{3}{5} - \frac{1}{2} \frac{2}{5} \approx 0,01333. \quad (2.62)$$

Para o atributo “Ventando”, tem-se:

$$\text{Gini(Falso)} = 1 - \left(\frac{3}{3}\right)^2 - \left(\frac{0}{3}\right)^2 = 0; \quad (2.63)$$

Figura 2 – Exemplo de árvore de decisão com dois níveis.



Fonte: De autoria própria.

$$\text{Gini(Verdadeiro)} = 1 - \left(\frac{0}{2}\right)^2 - \left(\frac{2}{2}\right)^2 = 0. \quad (2.64)$$

O ganho de informação é:

$$\text{Ganho de Informação (GI)} = \frac{12}{25} = 0,4800. \quad (2.65)$$

Por fim, o segundo nível da árvore de decisão para o valor “Chuvoso” será do atributo “Ventando”. A árvore de decisão com dois níveis é dada na figura 2. Caso o novo dado da tabela 2 fosse classificado por essa árvore, o rótulo de classificação seria “Não”.

## 2.4 Support Vector Machine

SVM é um algoritmo de machine learning no âmbito do aprendizado supervisionado. É válido ressaltar a técnica SVM, pois sua versão quântica apresenta resultados promissores no trabalho de Aksoy et al. (2024). Define-se um par de vetores de atributos  $X_i \in \mathbb{R}^N$  (*Features*) e classes  $y_i \in \{-1, 1\}$  (*labels*). Considerando que os pontos estão distribuídos em um espaço vetorial, define-se margem como o espaço interno a dois hiperplanos paralelos os quais são separados pela distância  $\frac{2}{\|w\|}$ . Não há pontos dentro da margem e define-se  $\vec{w}$  como o vetor normal ao hiperplano que divide os pontos de dados  $(X_i, y_i)$  de acordo com sua classe  $y_i$  e  $b$  é o termo linear do hiperplano. Ademais, os parâmetros  $w$  e  $b$  são treináveis. Os parâmetros serão otimizados para maximizar a margem a qual divide os pontos de diferentes classes no espaço, no caso, minimiza-se:

$$\frac{\|w\|^2}{2} \quad (2.66)$$

Com a restrição para todo elemento  $i$ :

$$y_i(w^T \cdot X_i + b) \geq 1. \quad (2.67)$$

Lembrando que maximizar  $\frac{2}{\|w\|}$  é equivalente a minimizar  $\frac{\|w\|^2}{2}$ .

Destarte, tem-se a definição dual do problema solucionado pela técnica SVM:

$$L(\vec{\alpha}) = \sum_{j=1}^M y_j \alpha_j - \frac{1}{2} \sum_{j,k=1}^M \alpha_j K_{jk} \alpha_k. \quad (2.68)$$

Tal mudança é feita para aplicar o “Kernel Trick” (WITTEN et al., 2017; HOFMANN, 2006). A matriz kernel  $K_{jk}$  pode ser definida como uma função:

$$K_{jk} = k(X_j, X_k) = X_j^T \cdot X_k. \quad (2.69)$$

O problema dual mostrado acima tem complexidade temporal de  $O(\log(1/\epsilon)M^2(N+M))$  onde  $\epsilon$  representa a acurácia do algoritmo,  $M$  representa o número de dados no conjunto e  $N$  representa a dimensão do vetor de características  $X_j$ . As definições foram tiradas de Witten et al. (2017), Rebertrost, Mohseni e Lloyd (2014), Hofmann (2006) e Cortes (1995). Por fim, tem-se a equação:

$$\hat{y}(X) = \sum_{i=1}^M \alpha_i \cdot k(X_i, X) + b, \quad (2.70)$$

Por simplicidade, não será utilizado um kernel para o exemplo, logo, a equação da classe predita  $\hat{y}$  é:

$$\hat{y}_i(X_i) = (w^T \cdot X_i + b). \quad (2.71)$$

Define-se a função de *loss* chamada “Hinge Loss”:

$$\mathcal{L}(T, w, b) = \sum_i \max(0, 1 - y_i \cdot \hat{y}_i) = \sum_i \max(0, 1 - y_i \cdot (w^T \cdot X_i + b)) \quad (2.72)$$

A função de custo inclui também a margem:

$$C(T, w, b) = \mathcal{L}(T, w, b) + \lambda \frac{\|w\|^2}{2}. \quad (2.73)$$

O parâmetro  $\lambda$  é um hiperparâmetro relacionado a regularização. Cada parâmetro treinável,  $w$  e  $b$ , será atualizado neste exemplo com a técnica de gradiente descendente, sendo assim, tal técnica consiste em alterar os parâmetros treináveis de um modelo no sentido oposto ao seu gradiente. A próxima iteração de um parâmetro  $w_k$  é dada abaixo:

Tabela 5 – Exemplo simplificado de meteorologia adaptado de [Witten et al. \(2017\)](#).

Índice	Temperatura	Umidade	Brincar
1	3	2	-1
2	2	2	1
3	1	1	-1
4	2	1	-1
5	3	1	-1

Fonte: [Witten et al. \(2017\)](#).

$$w_{k+1} = w_k - \eta \cdot \frac{\partial C(T, w, b)}{\partial w}. \quad (2.74)$$

A derivada parcial da função custo em relação ao parâmetro  $w$  é dado por:

$$\frac{\partial C(T, w, b)}{\partial w} = \frac{\partial \mathcal{L}(T, w, b)}{\partial w} + \frac{\lambda}{2} \frac{\partial \|w\|^2}{\partial w}. \quad (2.75)$$

A derivada parcial da função “Hinge Loss” em relação ao parâmetro  $w$  é:

$$\frac{\partial \mathcal{L}(T, w, b)}{\partial w} = \begin{cases} 0, & \text{se } y_i \cdot (w^T \cdot X_i + b) \geq 1, \\ -X_i \cdot y_i, & \text{se } y_i \cdot (w^T \cdot X_i + b) < 1. \end{cases} \quad (2.76)$$

A derivada parcial do termo de regularização em relação ao parâmetro  $w$  é:

$$\frac{\partial \|w\|^2}{\partial w_i} = \frac{\partial ((\sum_j w_j^2)^{\frac{1}{2}})^2}{\partial w_i} = \frac{\partial \sum_j w_j^2}{\partial w_i} = \sum_j \frac{\partial w_j^2}{\partial w_i} = \sum_j 2w_j \delta_{i,j} = 2w_i. \quad (2.77)$$

Sendo  $\delta_{i,j}$  o delta de Kronecker:

$$\delta_{i,j} = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j. \end{cases} \quad (2.78)$$

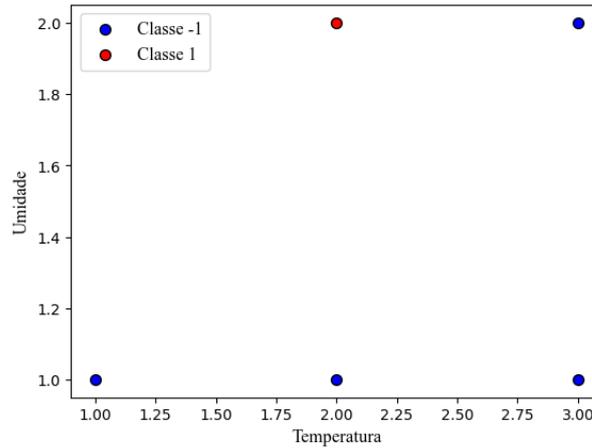
Por fim, a derivada parcial da função de custo em relação ao parâmetro  $w$  é dado por:

$$\frac{\partial C(X_i, y_i, w, b)}{\partial w} = \begin{cases} \lambda w, & \text{se } y_i \cdot (w^T \cdot X_i + b) \geq 1, \\ \lambda w - X_i \cdot y_i, & \text{se } y_i \cdot (w^T \cdot X_i + b) < 1. \end{cases} \quad (2.79)$$

O parâmetro  $b$  é treinável e é atualizado de acordo com a seguinte expressão:

$$b_{k+1} = b_k - \eta \cdot \frac{\partial C(T, w, b)}{\partial b}. \quad (2.80)$$

Figura 3 – Exemplo simplificado.



Fonte: De autoria própria.

O valor do parâmetro é modificado de acordo com a derivada parcial da função de custo em relação ao parâmetro  $b$  a qual é dada por:

$$\frac{\partial C(T, w, b)}{\partial b} = \frac{\partial \mathcal{L}(T, w, b)}{\partial b} + \frac{\lambda}{2} \frac{\partial \|w\|^2}{\partial b}. \quad (2.81)$$

A derivada parcial do termo de regularização em relação ao parâmetro linear  $b$  é:

$$\frac{\partial \|w\|^2}{\partial b} = 0. \quad (2.82)$$

Logo, a derivada parcial da função de custo pelo parâmetro  $b$  é:

$$\frac{\partial C(X_i, y_i, w, b)}{\partial b} = \frac{\partial \mathcal{L}(T, w, b)}{\partial b} = \begin{cases} 0, & \text{se } y_i \cdot (w^T \cdot X_i + b) \geq 1, \\ -y_i, & \text{se } y_i \cdot (w^T \cdot X_i + b) < 1. \end{cases} \quad (2.83)$$

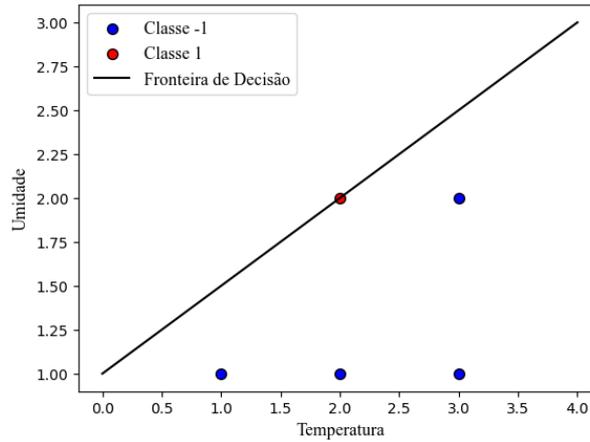
Utiliza-se o exemplo da tabela 5 e os atributos “Temperatura” e “Umidade” para realizar a classificação de um novo dado. O gráfico referente ao problema está na figura 3. Os valores iniciais de  $w$  e  $b$  para o exemplo formam a fronteira de decisão mostrada na figura 4. Os parâmetros iniciais são:

$$w_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \text{ e } b_1 = -2. \quad (2.84)$$

Para a realização do exemplo, utiliza-se os parâmetros  $\eta = 0,05$  e  $\lambda = 0$  com o intuito de simplificar o exemplo mostrado.

Inicialmente, é necessário verificar se os dados presentes no *dataset* são classificados de forma correta, ou seja, se  $y_i \cdot (w^T \cdot X_i + b) \geq 1$  para dado  $i$  considerado. Caso um dado seja classificado de forma correta, este não impacta nos erros propagados para os parâmetros treináveis. Para o dado  $(X_1, y_1)$ :

Figura 4 – Fronteira de decisão inicial do algoritmo SVM.



Fonte: De autoria própria.

$$y_1 \cdot (w_1^T \cdot X_1 + b_1) = -1 \cdot \left( (-1 \ 2) \cdot \begin{pmatrix} 3 \\ 2 \end{pmatrix} - 2 \right) = 1, \quad (2.85)$$

assim, tem-se:

$$\frac{\partial C(X_1, y_1, w, b)}{\partial w} = 0 \quad (2.86)$$

e

$$\frac{\partial C(X_1, y_1, w, b)}{\partial b} = 0. \quad (2.87)$$

Para  $(X_2, y_2)$ :

$$y_2 \cdot (w_1^T \cdot X_2 + b_1) = 1 \cdot \left( (-1 \ 2) \cdot \begin{pmatrix} 2 \\ 2 \end{pmatrix} - 2 \right) = 0, \quad (2.88)$$

logo,

$$\frac{\partial C(X_2, y_2, w, b)}{\partial w} = -X_2 \cdot y_2 = - \begin{pmatrix} 2 \\ 2 \end{pmatrix} \cdot 1 = \begin{pmatrix} -2 \\ -2 \end{pmatrix} \quad (2.89)$$

e

$$\frac{\partial C(X_2, y_2, w, b)}{\partial b} = -1. \quad (2.90)$$

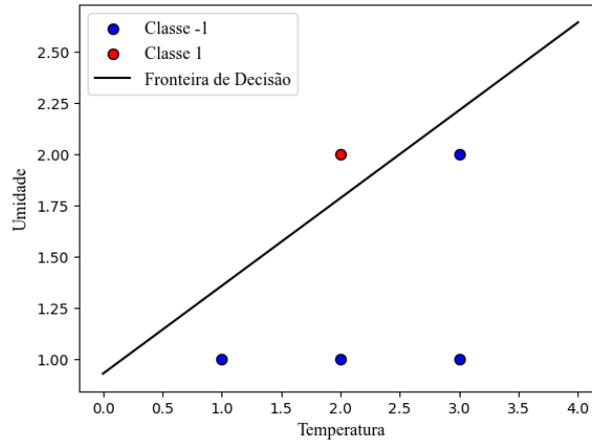
Para  $(X_3, y_3)$ :

$$y_3 \cdot (w_1^T \cdot X_3 + b_1) = -1 \cdot \left( (-1 \ 2) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} + -2 \right) = 1, \quad (2.91)$$

portanto,

$$\frac{\partial C(X_3, y_3, w, b)}{\partial w} = 0 \quad (2.92)$$

Figura 5 – Fronteira de decisão do algoritmo SVM.



Fonte: De autoria própria.

e

$$\frac{\partial C(X_3, y_3, w, b)}{\partial b} = 0. \quad (2.93)$$

Para  $(X_4, y_4)$ :

$$y_4 \cdot (w_1^T \cdot X_4 + b_1) = -1 \cdot \left( (-1 \ 2) \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} - 2 \right) = 2, \quad (2.94)$$

destarte,

$$\frac{\partial C(X_4, y_4, w, b)}{\partial w} = 0 \quad (2.95)$$

e

$$\frac{\partial C(X_4, y_4, w, b)}{\partial b} = 0 \quad (2.96)$$

Para  $(X_5, y_5)$ :

$$y_5 \cdot (w_1^T \cdot X_5 + b_1) = -1 \cdot \left( (-1 \ 2) \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix} - 2 \right) = 3, \quad (2.97)$$

dessarte,

$$\frac{\partial C(X_5, y_5, w, b)}{\partial w} = 0 \quad (2.98)$$

e

$$\frac{\partial C(X_5, y_5, w, b)}{\partial b} = 0. \quad (2.99)$$

Os parâmetros são atualizados de acordo com as equações abaixo:

Tabela 6 – Matriz de confusão.

Matriz de confusão		Classe Verdadeira	
		Verdadeiro	Falso
Classe Verdadeira	Verdadeiro	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Predita	Falso	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: De autoria própria.

$$\begin{aligned}
 w_2 &= w_1 - \eta \cdot \left( \sum_i -\frac{\partial C(X_i, y_i, w, b)}{\partial w} \right) = \begin{pmatrix} -1 \\ 2 \end{pmatrix} - 0,05 \cdot \left( \begin{pmatrix} -2 \\ -2 \end{pmatrix} \cdot 1 \right) \\
 &= \begin{pmatrix} -0,9 \\ 2,1 \end{pmatrix}, \tag{2.100}
 \end{aligned}$$

$$b_2 = b_1 - \eta \cdot \left( \sum_i \frac{\partial C(X_i, y_i, w, b)}{\partial b} \right) = -2 - 0,05 \cdot (-1) = -1,95. \tag{2.101}$$

Com os valores obtidos, tem-se que a fronteira de decisão divide os pontos da classe  $-1$  dos pontos da classe  $1$  como pode ser visto na figura 5.

## 2.5 Matriz de Confusão

A matriz de confusão permite avaliar de forma mais ampla o desempenho das técnicas de classificação. Tem-se quatro elementos na matriz de confusão: Verdadeiro Positivo (VP), ou seja, o exemplo é da classe de interesse e a classe predita pelo modelo concorda com a classe verdadeira; Falso Negativo (FN), o exemplo é da classe de interesse e a classe predita pelo modelo discorda da classe verdadeira; Falso Positivo (FP), o exemplo não é da classe de interesse e o modelo discorda da classe verdadeira; Verdadeiro Negativo (VN), o exemplo não é da classe de interesse e o modelo concorda com a classe verdadeira.

Entre as relações medidas, tem-se a acurácia a qual mede a taxa de acerto do modelo:

$$\text{Acurácia} = \frac{VP + VN}{Total}. \tag{2.102}$$

A revocação mede a chance de se identificar a classe considerando que a classe do dado seja, de fato, positivo. Em esquizofrenia, pode-se dizer que a revocação mede a chance de um indivíduo que tenha esquizofrenia seja classificado corretamente pelo modelo. A revocação também é conhecida como sensibilidade no contexto de pesquisas relacionadas à área da saúde:

$$\text{Revocação} = \frac{VP}{VP + FN}. \quad (2.103)$$

A precisão indica a proporção de exemplos que são classificados na classe de interesse corretamente.

$$\text{Precisão} = \frac{VP}{VP + FP}. \quad (2.104)$$

Por fim, em problemas relacionados à medicina, utiliza-se uma métrica chamada especificidade a qual é dada por:

$$\text{Especificidade} = \frac{VN}{FP + VN}. \quad (2.105)$$

Neste capítulo, apresenta-se as técnicas de machine learning clássicas utilizadas e a matriz de confusão usada para explicar os resultados obtidos. Ademais, exemplifica-se o funcionamento de cada algoritmo tendo em vista que tais características alteram sua fronteira de decisão e seu desempenho.

## 2.6 Machine Learning Aplicado a Diagnósticos

Os trabalhos apresentados nesta seção não constituem uma busca exaustiva, logo, constituem apenas uma fração dos artigos presentes na literatura. Porém, os trabalhos permitem um maior entendimento do cenário da aplicação de *machine learning* em diagnóstico de doenças e transtornos.

Técnicas de *machine learning* são usadas extensivamente para auxiliar no diagnóstico de diversos transtornos e doenças. Tais técnicas são utilizadas no diagnóstico de doenças renais (QEZELBASH-CHAMAK et al., 2022; WALSE et al., 2021; NITHYA et al., 2020; IMRAN; AMIN; JOHORA, 2018; NAVANEETH; SUCHETHA, 2020; BRUNETTI et al., 2019), de Parkinson (MEI; DESROSIERS; FRASNELLI, 2021; NURROHMAN; ABDULLAH; MURFI, 2020; ANNABEL; SREENIDHI; VISHALI, 2021; ASMAE et al., 2020; GÜRÜLER, 2017; SHETTY; RAO, 2016), de Alzheimer (TANVEER et al., 2020; HUANG; SUN; GUO, 2021; KUANG et al., 2021; MANZAK; ÇETINEL; MANZAK, 2019; CHAPLOT; PATNAIK; JAGANNATHAN, 2006; ZHANG et al., 2011), de câncer de mama (YUE et al., 2018; RAJENDRAN; JAYABALAN; THIRUCHELVAM, 2020; ABDEL-NASSER et al., 2015; SHARMA; KHANNA, 2015; MOON et al., 2017; LO et al., 2016; VENKATESH et al., 2015), de diabetes (FITRIYANI et al., 2019; FERNÁNDEZ-EDREIRA; LIÑARES-BLANCO; FERNANDEZ-LOZANO, 2021; ALI et al., 2020; TSAO; CHAN; SU, 2018; QTEAT; AWAD, 2021), de COVID-19 (LI et al., 2020; CHEN et al.,

2020; SONG et al., 2021; JIN et al., 2020; AHSAN et al., 2020), de cardiopatias (ENAD; MOHAMMED et al., 2023), entre outras doenças. Em relação às técnicas aplicadas, tem-se que as redes neurais convolucionais apresentam os melhores resultados em problemas relacionados a análise de imagens. Outras técnicas que aparecem com grande frequência no diagnóstico de doenças com *machine learning* são o SVM, Floresta Aleatória e Árvore de Decisão (AHSAN; LUNA; SIDDIQUE, 2022). Uma das lacunas relatada no trabalho de Ahsan, Luna e Siddique (2022) é a falta de interpretabilidade de modelos de *machine learning*. Sendo assim, o presente trabalho busca não a avaliação dos resultados, mas também a análise dos modelos obtidos (ENAD; MOHAMMED et al., 2023; ULLAH; GARCIA-ZAPIRAIN, 2024).

Tendo em vista modelos de *machine learning* aplicados a detecção de transtornos mentais utilizando IRM, o trabalho de Lazli, Boukadoum e Mohamed (2020) apresenta uma revisão sistemática da literatura para problemas de segmentação de regiões do cérebro e classificação de transtornos mentais. Nesse trabalho, tem-se que as técnicas de aprendizado supervisionado são preferidas em relação às técnicas de aprendizado não supervisionado. Por fim, o artigo apresenta que a combinação de imagens cerebrais estruturais com funcionais permitem que os modelos de classificação alcancem desempenho superior do que modelos treinados apenas em um dos dois tipos de imagem (LAZLI; BOUKADOUM; MOHAMED, 2020).



### 3 Computação Quântica

Computação quântica é um método que utiliza dos princípios e das características de mecânica quântica para processar dados (HIDARY, 2019). Entre as principais características quânticas que são úteis para a computação estão a superposição e o emaranhamento. Considerando um sistema quântico de dois níveis  $|0\rangle$  e  $|1\rangle$  e amplitudes  $\alpha, \beta \in \mathbb{C}$ , tem-se a superposição desses estados em um estado quântico  $|\psi\rangle$  tal que:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \quad (3.1)$$

Os estados  $|0\rangle$  e  $|1\rangle$  são equivalentes a:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ e } |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3.2)$$

Os estados quânticos também podem ser descrito pela sua representação dual:  $\langle 0|$  e  $\langle 1|$ . Como os estados quântico são representados por matrizes coluna, os *bras* que são sua representação dual, são o conjugado transposto e são representados por matrizes linha. Em mecânica quântica, os estados quânticos são vetores que pertencem a um espaço vetorial complexo chamado de espaço de Hilbert. Vale ressaltar que este espaço vetorial complexo é munido de um produto interno. Definimos o produto interno entre dois vetores do espaço de Hilbert como se segue:

$$\langle a|b\rangle = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \quad (3.3)$$

Sendo que os estados  $|a\rangle$  e  $|b\rangle$  são dados da seguinte forma  $|a\rangle = a_1|0\rangle + a_2|1\rangle$  e  $|b\rangle = b_1|0\rangle + b_2|1\rangle$ . As amplitudes  $\alpha$  e  $\beta$  que aparecem Eq. 3.1 são chamadas de amplitudes de probabilidade e seguem a lei de Born:

$$|\alpha|^2 + |\beta|^2 = 1, \quad (3.4)$$

onde o módulo quadrado das amplitudes de probabilidade representam as probabilidades de medição dos estados  $|0\rangle$  e  $|1\rangle$ .

Uma das diferenças principais entre computação quântica e clássica são os operadores utilizados. Os operadores quânticos, em sua maioria, são hermitianos e unitários. Para operadores hermitianos, dado um operador  $A$  tal que  $A \in \mathbb{C}^{2^n \times 2^n}$  sendo  $n$  o número de qubits utilizado no circuito, tem-se:

$$A = A^\dagger. \quad (3.5)$$

Um operador é unitário se  $AA^\dagger = A^\dagger A = I$ .

Por último, tem-se três operações elementares realizadas na computação quântica: produto interno; produto externo; produto tensorial. Define-se o produto interno, Considere os estados  $|0\rangle$  e  $|1\rangle$ , mostra-se que os dois estados são ortogonais entre si

$$\langle a|b\rangle = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = a_1 b_1 + a_2 b_2. \quad (3.6)$$

A base  $\{|0\rangle |1\rangle\}$  é o que chamamos de base computacional e ela forma uma base ortonormal, pois  $\langle i|j\rangle = \delta_{i,j}$ , onde  $\delta_{i,j}$  é a delta de Kronecker definida na equação 2.78.

O produto externo tem como resultado uma matriz e é representado no exemplo abaixo:

$$|0\rangle\langle 0| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (3.7)$$

Por fim, tem-se o produto tensorial o qual representa a expansão de um estado quântico:

$$|0\rangle \otimes |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ 0 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 \\ 1 \cdot 0 \\ 0 \cdot 1 \\ 0 \cdot 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = |00\rangle. \quad (3.8)$$

Existem três operadores de suma importância em computação quântica que são chamados de operadores de Pauli, X, Y e Z. Esses operadores são representados matricialmente da seguinte forma:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (3.9)$$

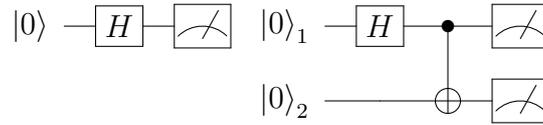
$$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad (3.10)$$

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.11)$$

Tais operadores são os geradores de rotações nos eixos x, y e z. Essas rotações são dadas de forma geral:

$$R_{\hat{n} \cdot \vec{\sigma}}(\theta) = e^{-i \frac{\theta \hat{n} \cdot \vec{\sigma}}{2}}, \quad (3.12)$$

Figura 6 – Exemplo de circuito quântico à esquerda e o circuito gerador do estado de Bell à direita.



Fonte: De autoria própria.

onde  $\vec{\sigma} = X\hat{i} + Y\hat{j} + Z\hat{k}$ , onde  $\hat{n}$  é um versor que estipula o eixo e  $\theta$  o ângulo da rotação.

Existem infinitas bases para se representar um qubit. Além da base computacional formada pelos vetores  $|0\rangle$  e  $|1\rangle$ , também podemos citar as bases associadas aos operadores de Pauli X e Y:  $\{|+\rangle, |-\rangle\}$  e  $\{|+i\rangle, |-i\rangle\}$  respectivamente. Os vetores  $|+\rangle$  e  $|-\rangle$  são os autovetores do operador de Pauli X, enquanto os estados  $|+i\rangle$  e  $|-i\rangle$  são os autovetores do operador de Pauli Y. Estes vetores são escritos em termos da base computacional da seguinte forma (HIDARY, 2019):

$$|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}, \quad (3.13)$$

$$|\pm i\rangle = \frac{|0\rangle \pm i|1\rangle}{\sqrt{2}}. \quad (3.14)$$

Os algoritmos quânticos podem ser implementados via circuitos quânticos. As linhas representam os qubits, as caixas representam os operadores e os estados iniciais são apresentados no início de cada circuito. No fim do circuito, percebe-se a operação de medição o qual provoca o colapso do estado quântico. O circuito da figura 6 apresenta como resultado o seguinte estado quântico antes da medida:

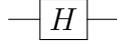
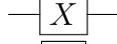
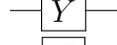
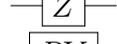
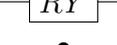
$$H|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = |+\rangle. \quad (3.15)$$

Por último, discorre-se sobre o conceito de emaranhamento. Emaranhamento é uma característica de sistemas quânticos que ocorre quando não é possível separar estados quânticos de múltiplos qubits em qubits individuais. Um dos estados quânticos utilizados para explicar o conceito de emaranhamento é o estado de Bell, o circuito que gera-o pode ser visto na figura 6. A representação dos operadores no circuito quântico pode ser vista na tabela 7.

O estado resultante é não-separável e é dado pela seguinte equação:

$$CNOT H_1|0\rangle_1 \otimes |0\rangle_2 = CNOT \frac{|0\rangle|0\rangle \otimes |1\rangle|0\rangle}{\sqrt{2}} = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \quad (3.16)$$

Tabela 7 – Representação dos operadores no circuito quântico.

Operador	Representação no Circuito
H	
X	
Y	
Z	
RY	
CNOT <sub>1,2</sub>	

Fonte: De autoria própria.

### 3.1 Postulados da Mecânica Quântica

Os postulados da mecânica quântica são um conjunto de princípios fundamentais que formam a base da teoria quântica. Eles descrevem como os sistemas físicos são representados matematicamente, como os estados evoluem no tempo, e como as medições afetam esses estados. Aqui estão os principais postulados da mecânica quântica:

O primeiro postulado da mecânica quântica afirma que, todo sistema físico isolado, pode ser descrito por um estado  $|\psi(t_0)\rangle$  em um determinado instante de tempo  $t_0$ . Tal estado pertence a um espaço de estados, ou seja, um espaço vetorial complexo munido de produto interno.

O segundo postulado da mecânica quântica diz que a evolução de um sistema quântico fechado pode ser descrito por uma transformação unitária. Considere dois estados quânticos  $|\Psi\rangle$  e  $|\Psi'\rangle$  associados a instantes de tempo  $t_1$  e  $t_2$  respectivamente. A evolução de um estado para outro está relacionada com o operador unitário  $U$  da seguinte forma:

$$|\Psi'\rangle = U|\Psi\rangle. \quad (3.17)$$

O terceiro postulado discorre sobre as medidas feitas em sistemas quânticos. Toda grandeza física mensurável é descrita por um operador que age em um espaço de estados. O operador é chamado de observável. As medidas são descritas como um conjunto de operadores de medida  $\{M_m\}$ . As medidas ocorrem no espaço de estados e sua probabilidade de medir o operador de índice  $m$  é dada por:

$$p(m) = \langle \psi | M_m^\dagger M_m | \psi \rangle. \quad (3.18)$$

Ademais, o estado quântico depois da medida é dado por:

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \quad (3.19)$$

Os operadores de medida precisam satisfazer a equação de completude:

$$\sum_m M_m^\dagger M_m = I, \quad (3.20)$$

tal equação é uma consequência de:

$$\sum_m p(m) = 1 \text{ e } \sum_m p(m) = \sum_m \langle\psi|M_m^\dagger M_m|\psi\rangle. \quad (3.21)$$

Um exemplo da utilização de medidas é a medida de um qubit na base computacional. Define-se os operadores hermitianos  $M_0 = |0\rangle\langle 0|$  e  $M_1 = |1\rangle\langle 1|$  e pode-se verificar a seguinte equação:

$$\sum_m M_m^\dagger M_m = I. \quad (3.22)$$

Logo,

$$\sum_m M_m^\dagger M_m = M_0^\dagger M_0 + M_1^\dagger M_1 = M_0^2 + M_1^2. \quad (3.23)$$

Lembrando que os operadores  $M_0$  e  $M_1$  são hermitianos, ou seja,  $M_0 = M_0^\dagger$ . Ademais, sabe-se que  $M_0^2$  e  $M_1^2$  são iguais a

$$M_0^2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (3.24)$$

e

$$M_1^2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad (3.25)$$

respectivamente. Portanto, tem-se:

$$\sum_m M_m^\dagger M_m = M_0 + M_1 = I, \quad (3.26)$$

Para explicar o conceito de valores esperados de operadores, é necessário explicar o que são medidas espectrais. O conceito refere-se a uma medida relacionada a um observável  $M$  o qual é um operador hermitiano no espaço de estados de um determinado sistema físico. O observável pode ser decomposto espectralmente:

$$M = \sum_m m P_m. \quad (3.27)$$

Onde  $P_m$  é um projetor do autoespaço de  $M$  com o autovalor  $m$ . Os autovalores  $m$  são os valores possíveis de serem obtidos ao medir-se um estado  $\psi$  com o operador  $M$ :

$$p(m) = \langle \psi | P_m | \psi \rangle. \quad (3.28)$$

O estado quântico após a medição é:

$$\frac{P_m | \psi \rangle}{\sqrt{p(m)}}. \quad (3.29)$$

Por fim, o valor esperado de uma medida  $M$  é dado por:

$$E(M) = \sum_m m p(m) = \sum_m m \langle \psi | P_m | \psi \rangle = \langle \psi | \sum_m m P_m | \psi \rangle = \langle \psi | M | \psi \rangle. \quad (3.30)$$

O valor esperado de um operador é dado pela notação:

$$\langle M \rangle = \langle \psi | M | \psi \rangle. \quad (3.31)$$

Por fim, o resultado da medida de uma grandeza física é, necessariamente, um dos autovalores do respectivo observável.

O quarto postulado afirma que o estado de espaços de um sistema físico composto por sistemas menores é igual ao produto tensorial dos espaços de estados de tais sistemas menores. Supondo que um sistema seja composto por  $n$  sistemas menores e cada sistema esteja em um estado  $|\psi_i\rangle$ , o estado de espaços do sistema composto por sistemas menores é  $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$ .

O último conceito a ser apresentado neste capítulo refere-se a matriz densidade. Suponha que  $|\psi_i\rangle$  seja um dos estado quântico que representa um sistema com probabilidade  $p_i$  e seja  $i$  um índice. A matriz densidade é definida pela equação:

$$\rho \equiv \sum_i p_i |\psi_i\rangle \langle \psi_i|. \quad (3.32)$$

Os postulados e suas definições foram tiradas de [Nielsen e Chuang \(2010\)](#).

## 3.2 Computação Universal

Um circuito eletrônico digital realiza operações computacionais por meio de portas lógicas e leva a informação a tais portas por meio de fios. Considerando as portas *not and* (NAND) e *not or* (NOR), é possível representar qualquer porta lógica. Tal conceito é chamado de computação universal ([NIELSEN; CHUANG, 2010](#)).

De forma análoga a computação clássica, diz-se que um conjunto de operadores é universal para à computação quântica se puder aproximar qualquer operador unitário dado uma determinada métrica. O grupo de Clifford composto pelas portas Hadamard, CNOT e S, contudo, o operador S pode ser construído pela porta T:

$$S = T^2 = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{2}} \end{pmatrix}. \quad (3.33)$$

O conjunto de Clifford não é universal, porém, com a adição da porta T, o conjunto de portas Hadamard, T e CNOT é universal (HIDARY, 2019).

Mais especificamente, o teorema de Solovay-Kitaev diz que para um dado conjunto de portas quânticas, pode-se aproximar qualquer porta quântica com precisão  $\delta$ . Considerando  $L$  como o número de portas quânticas do circuito quântico original e  $L'$  como o número de portas quânticas do circuito quântico aproximado, tem-se um limite superior para  $L'$  (HIDARY, 2019):

$$L' \geq O\left(L \log^4\left(\frac{L}{\delta}\right)\right). \quad (3.34)$$

Portanto, ao aproximar portas quânticas que não estão presente em um determinado conjunto, aumenta-se o número de operações computacionais necessárias no circuito quântico. O aumento no número de portas e as taxas de erros presentes nas portas quânticas são desafios para o uso de algoritmos complexos em computadores quânticos (PRESKILL, 2018).

### 3.3 Ruído na Computação Quântica

Os sistemas quânticos apresentados até esta seção são sistemas isolados e, por isso, não sofrem a ação de ruídos externos proveniente da interação dos sistemas com o ambiente. Contudo, em aplicações reais, as taxas de erro em *hardware* quântico limitam o uso de algoritmos quânticos mais complexos (PRESKILL, 2018). O conceito de decoerência representa o processo pelo qual sistemas quânticos abertos perdem suas características quânticas devido a interação com o ambiente e passam a se comportar como sistemas clássicos. Esse processo se dá pela ocorrência de erros coerentes e incoerentes (NIELSEN; CHUANG, 2010; MÁRTON; ASBÓTH, 2023). Define-se como ruído incoerente interações do sistema quântico com o ambiente. Os erros coerentes advém, principalmente, de imprecisões das portas quânticas aplicadas, inicialização de estados e no processo de medição (NIELSEN; CHUANG, 2010; MÁRTON; ASBÓTH, 2023). No caso, apresenta-se os três diferentes tipos de erros coerentes e incoerentes em sistemas quânticos, os erros podem ser divididos em *bit flip*, *phase flip* e *bit-phase flip*. Os erros do tipo *bit flip* representam a chance de um qubit mudar seu valor na base computacional  $Z$  e pode ser descrito como um operador  $U_{bit-flip}$  pela expressão (NIELSEN; CHUANG, 2010):

$$U_{bit-flip} = (1 - p_X)I + p_X X. \quad (3.35)$$

Sendo  $p_X$  a probabilidade de ocorrer um erro do tipo *bit flip*. O erro do tipo *phase flip* é descrito pelo operador  $U_{phase\ flip}$ :

$$U_{bit\ flip} = (1 - p_Z)I + p_Z Z. \quad (3.36)$$

Onde  $p_Z$  representa a probabilidade de ocorrer um erro do tipo *phase flip*. Por último, o erro do tipo *bit-phase flip* é descrito pelo operador  $U_{bit-phase\ flip}$ :

$$U_{bit-phase\ flip} = (1 - p_Y)I + p_Y Y. \quad (3.37)$$

Onde  $p_Y$  representa a probabilidade de ocorrer um erro do tipo *bit-phase flip* (NIELSEN; CHUANG, 2010).

Neste capítulo, discorreremos sobre os postulados da mecânica quântica, espaços de Hilbert, estados quânticos, operadores e circuitos quânticos. Tais conhecimentos são essenciais para justificar as vantagens e desvantagens da computação quântica em problemas de classificação.

## 4 Quantum Machine Learning

Conforme descrito em [Biamonte et al. \(2017\)](#), técnicas de machine learning obtiveram sucesso tanto em encontrar padrões estatísticos quanto produzir dados que contenham tais padrões. No caso da computação quântica, espera-se que os computadores quânticos sejam capazes não só de produzir dados com padrões estatísticos da mecânica quântica, mas também de identificar os mesmos padrões em dados clássicos.

Contudo, espera-se que, além da capacidade dos computadores quânticos de reconhecerem diferentes padrões, haja diminuição do tempo necessário para a resolução de diversos problemas, tal conceito é conhecido como *quantum speedup*. A complexidade de um algoritmo quântico pode ser apresentado de duas maneiras: número de execuções e número de portas quânticas.

*Quantum Machine Learning* apresenta avanços para diferentes problemas abordados em técnicas de *Machine Learning* clássicas. Um conjunto de problemas que será beneficiado pela aplicação de *Quantum Machine Learning* é o conjunto de problemas de álgebra linear. Problemas como encontrar a solução de um sistema linear, método dos mínimos quadrados e *Principal Component Analysis* (PCA) possuem menor complexidade temporal em computadores quânticos. No caso do algoritmo Harrow, Hassidim e Lloyd (HHL), é possível inverter uma matriz hermitiana para resolver o problema  $Ax = b$ . A principal aplicação é a resolução de sistemas lineares por meio da inversão da matriz  $A$  supracitada. Por fim, sabe-se que o algoritmo HHL é capaz de resolver sistemas lineares com uma complexidade temporal de  $O((\log N)^2)$ , enquanto o melhor algoritmo clássico conhecido precisa de  $O(N \log N)$  ([HARROW; HASSIDIM; LLOYD, 2009](#)).

Como citado anteriormente, o PCA é um algoritmo amplamente utilizado em *Machine Learning* e advém da álgebra linear, mais especificamente, é utilizado para representar informação de um conjunto de vetores em um espaço vetorial de dimensão  $n$  em um espaço vetorial de dimensão  $m$  tal que  $m < n$ , ademais, procura-se preservar relações presentes entre os atributos. No trabalho de [Lloyd, Mohseni e Rebentrost \(2014\)](#), apresenta-se um algoritmo quântico capaz de codificar os autoestados e autovalores de uma matriz densidade  $\rho$ . Considere que  $|\chi_i\rangle$  sejam os autoestados e que  $r_i$  sejam os autovalores correspondentes a matriz densidade  $\rho$ , tem-se que o estado quântico resultante do PCA quântico é:

$$\sum_i r_i |\chi_i\rangle\langle\chi_i| \otimes |\tilde{r}_i\rangle\langle\tilde{r}_i| \quad (4.1)$$

Por fim,  $\tilde{r}_i$  são aproximações dos autovalores  $r_i$ . É válido ressaltar que o algoritmo

HHL também é utilizado no PCA quântico.

Em relação a algoritmos promissores em aplicações reais, tem-se também a transformada de Fourier quântica a qual é a versão quântica da *Fast Fourier Transform* (FFT) e foi introduzida pelo trabalho de Shor (1997). Tem-se que o para um dado estado  $|a\rangle$ , o número  $a$  refere-se a interpretação dos qubits como dígitos binários, resultando em um número representado em base decimal. Tal número respeita a seguinte desigualdade  $0 \leq a < q$ . Ademais, assume-se que  $q = 2^n$  sendo  $n$  o número de qubits e o número  $a$  seja codificado como:

$$|a\rangle = \sum_{j=0}^{2^n-1} a_j |j\rangle. \quad (4.2)$$

Logo, o estado  $|a\rangle$  evoluirá para o  $|b\rangle$  de tal forma que:

$$b_c = \frac{1}{\sqrt{2^n}} \sum_{j=0}^{2^n-1} a_j \exp\left(\frac{2\pi i a_j c}{2^n}\right). \quad (4.3)$$

Sendo o estado  $b$ :

$$|b\rangle = \sum_{c=0}^{2^l-1} b_c |c\rangle. \quad (4.4)$$

Portanto, tem-se que a transformação de Fourier quântica é capaz de evoluir o estado  $|a\rangle$  para um estado  $|b\rangle$  (NIELSEN; CHUANG, 2010).

Contudo, tais algoritmos de *Quantum Machine Learning* apresentam problemas no desempenho da entrada de dados em circuitos quânticos, na quantidade de bits de saída, no número de portas quânticas e na comparação entre algoritmo quânticos e clássicos. Sobre os problemas de desempenho da entrada de dados em circuitos quânticos, ainda que diversos algoritmos quânticos apresentem vantagens no processamento de informação, raras são as vezes nas quais tais algoritmos também apresentam vantagem em mapear informações clássicas no ambiente quântico. Por fim, há casos em que o custo de gerar estados quânticos com as informações clássicas necessárias é dominante em relação ao tempo de processamento necessário. Por fim, sobre a largura dos circuitos quânticos, é válido ressaltar que algoritmos quânticos aplicados em problemas reais necessitam de um grande número de bits resultantes. Logo, algoritmos quânticos tornam-se muito complexos para aplicações práticas (BIAMONTE et al., 2017).

Sobre a comparação entre algoritmos clássicos e quânticos, muitas vezes é desafiador tecer comparações precisas entre algoritmos quânticos e clássicos, pois, para isso, seria necessário não somente comparações com soluções analíticas, mas também com diversos algoritmos heurísticos. Logo, torna-se difícil avaliar o ganho de desempenho de algoritmos quânticos em relação diversos métodos clássicos (BIAMONTE et al., 2017).

Contudo, computadores quânticos apresentam vantagens em comparação com a computação clássica. Uma das vantagens observadas é a dimensionalidade do espaço de

estados, pois o número de dimensões desse espaço escala exponencialmente em relação ao número de qubits utilizados (LIU; ARUNACHALAM; TEMME, 2021). Outra vantagem, são as características de superposição e emaranhamento, pois elas não são encontradas no paradigma da computação clássica. Essas características têm o potencial de solucionar problemas computacionais de forma mais rápida ou com maior eficácia tendo em vista os algoritmos de Grover e Shor (GROVER, 1996; SHOR, 1997; HIDARY, 2019; NIELSEN; CHUANG, 2010).

Destarte, percebe-se que a aplicação de computação quântica traz diversas vantagens na resolução de diversos problemas tanto ganho de desempenho quanto na capacidade de generalização de modelos de *machine learning*. Porém, há diversos ainda há diversos desafios na aplicação dessas técnicas em cenários reais. Como supracitado, o tempo necessário para mapear dados clássicos no ambiente quântico podem dominar a complexidade computacional do algoritmo (BIAMONTE et al., 2017).

Logo, tendo em vista os desafios relacionados a inserção de dados clássicos nos computadores quânticos, a aplicação de *Quantum Machine Learning* mostra-se promissora nas áreas relacionadas à química e biologia, pois o computador quântico é capaz de processar dados de origem quântica sem a necessidade de traduzi-los para o domínio clássico (BIAMONTE et al., 2017; CERESO et al., 2022).

## 4.1 Quantum Encoding

Como foi descrito na subseção anterior, um dos desafios em problemas práticos de Quantum Machine Learning é a capacidade de codificar informação no domínio quântico, sendo assim, é necessário apresentar os tipos de codificação presentes na literatura. A forma mais simples de codificar informação é utilizando o próprio estado quântico. Suponha que  $x$  seja um vetor e  $x \in \{0, 1\}^n$ , além disso, suponha um estado  $|\psi\rangle$  no espaço de Hilbert tal que  $|\psi\rangle \in \mathbb{C}^{2^n}$ :

$$|\psi\rangle = |x_1, x_2, \dots, x_n\rangle \quad (4.5)$$

Informação clássica também pode ser codificada por meio da amplitude dos estados quânticos. Para um vetor coluna  $c \in \mathbb{R}^N$  de  $2^n = N$  valores, sendo  $n$  o número de qubits, tem-se:

$$c = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_N \end{pmatrix} \quad (4.6)$$

Logo, a codificação de amplitude do vetor  $c$  é:

$$|\psi_c\rangle = \sum_{i=0}^N c_i |i\rangle \quad (4.7)$$

Sujeito a:

$$\sum_i^N |c_i|^2 = 1 \quad (4.8)$$

Também é possível codificar uma matriz  $A$ , ou seja,  $A \in \mathbb{R}^{2^n \times 2^m}$ :

$$|\psi_A\rangle = \sum_j^M \sum_i^N a_{ij} |i\rangle |j\rangle \quad (4.9)$$

Sujeito a:

$$\sum_j^M \sum_i^N |a_{ij}|^2 = 1 \quad (4.10)$$

Caso a matriz  $A$  seja quadrada, ou seja,  $A \in \mathbb{R}^{N \times N}$ . Pode-se codificar a matriz  $A$  em uma matriz de densidade  $\rho$ :

$$\rho_A = \sum_j^M \sum_i^N a_{ij} |i\rangle \langle j| \quad (4.11)$$

Contudo,  $A$  precisa ser uma matriz hermitiana e unitária, ou seja,  $A = A^\dagger$  e  $AA^\dagger = A^\dagger A = I$  respectivamente.

Dado um escalar  $x \in \mathbb{R}$ , associa-o com o tempo  $t$  a uma evolução unitária de um hamiltoniano  $H$ :

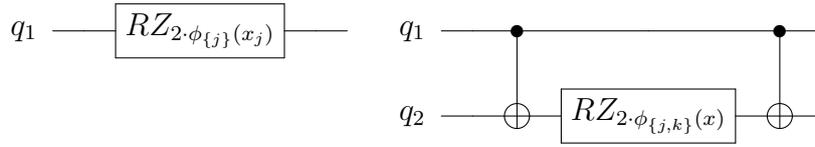
$$U(x) = e^{-ixH} \quad (4.12)$$

Dado um estado inicial  $|\psi_0\rangle$  e o operador  $U(x)$ , tem-se que o estado resultante da evolução unitária é:

$$|\psi(x)\rangle = U(x)|\psi_0\rangle \quad (4.13)$$

Por fim, tem-se a codificação por hamiltoniano. Caso uma matriz quadrada  $A$  seja hermitiana, pode-se utilizar  $A$  como um hamiltoniano. Caso contrário, é possível criar uma matriz hermitiana partindo-se de  $A$ :

Figura 7 – Representação da equação 4.17 no circuito quântico.



Fonte: Adaptado de Havlicek et al. (2019).

$$H_A = \begin{pmatrix} 0 & A \\ A^\dagger & 0 \end{pmatrix} \quad (4.14)$$

Portanto, apresenta-se as principais formas de se codificar informação em um computador quântico. É válido ressaltar que a técnica de *kernel* quântico aplicada refere-se a codificação por evolução temporal como é explicado na próxima seção (SCHULD; PETRUCCIONE, 2021).

## 4.2 Kernel Quântico

Os *kernels* quânticos são circuitos quânticos nos quais dados são codificados, os circuitos podem ser fixos como nas técnicas *Z feature* e *ZZ feature* ou podem ser variacionais. Inicialmente, define-se o operador unitário  $U_{\phi(X_i)}$  (HAVLICEK et al., 2019):

$$U_{\phi(X_i)} = \exp(-i \sum_{S \subseteq [n]} \phi_S(X_i) \prod_{d \in S} Z_d) \quad (4.15)$$

sendo

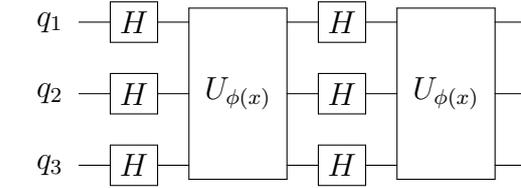
$$\phi_S(X_i) = \begin{cases} x_{i,j}, & \text{se } |S| = 1, \\ (\pi - x_{i,j})(\pi - x_{i,k}), & \text{se } |S| = 2. \end{cases} \quad (4.16)$$

Onde  $j, k \in S$ ,  $S$  é um subconjunto de  $[n]$  e  $n$  representa o número de qubits. Por fim, é válido ressaltar que o módulo do conjunto  $S$  é limitado tal que  $|S| \leq 2$ . Considerando a aplicação do *Z feature* e do *ZZ feature*, tem-se respectivamente:

$$U_{\phi_{\{j\}}(x_{i,j})} = \exp(-i\phi_{\{j\}}(x_{i,j})Z_j) \text{ e } U_{\phi_{\{j,k\}}(x_{i,j}, x_{i,k})} = \exp(-i\phi_{\{j,k\}}(x_{i,j}, x_{i,k})Z_j Z_k). \quad (4.17)$$

A representação em circuitos quântico para cada caso de  $\phi_S(x)$  pode ser vista na figura 7.

O circuito quântico utilizado no *ZZ Feature* utiliza uma sequência de portas Hadamard com um operador  $U_{\phi(X_i)}$  como pode ser visto na figura 8. O estado quântico resultante da aplicação do circuito da figura em questão é:

Figura 8 – Circuito quântico utilizado como *kernel* quântico.

Fonte: Adaptado de Havlicek et al. (2019).

$$|\Phi(X_i)\rangle = U_{\phi(X_i)} H^{\otimes n} U_{\phi(X_i)} H^{\otimes n} |0\rangle^{\otimes n} \quad (4.18)$$

Logo, o *kernel* obtém o valor do produto interno entre dois exemplos:

$$K(X_i, X_j) = |\langle \Phi(X_i) | \Phi(X_j) \rangle|^2. \quad (4.19)$$

Todo o desenvolvimento matemático e notação utilizados acima é descrito em Havlicek et al. (2019).

### 4.3 Circuito Variacional Quântico

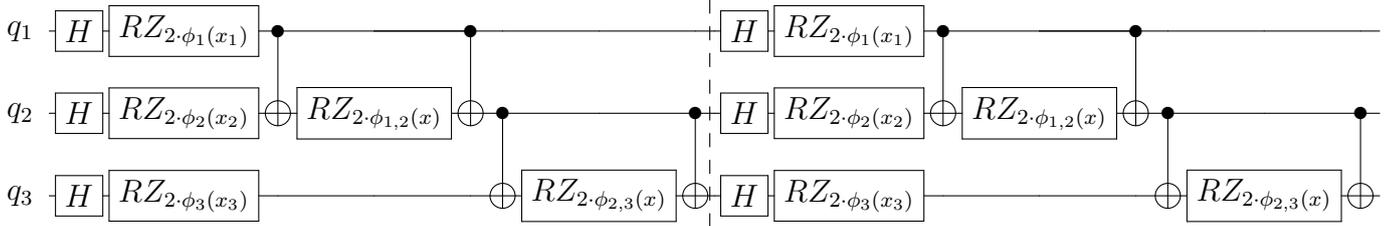
Os circuitos variacionais apresentam como característica principal as portas parametrizadas. Tais portas realizam uma operação cuja magnitude é controlada por um conjunto de parâmetros ajustáveis  $\theta$ . Para que os parâmetros sejam adequados a um determinado problema, os parâmetros do circuito são otimizados. O processo de otimização consiste em maximizar ou minimizar o valor esperado de um dado operador  $A$ . Considerando o estado inicial do circuito como  $|\Psi\rangle$ , o operador que representa o circuito variacional como  $U(\theta)$  e a função que representa o valor esperado como  $f(\theta)$ , tem-se:

$$\theta^* = \arg \min_{\theta} f(\theta) = \langle \Psi | U(\theta)^\dagger A U(\theta) | \Psi \rangle. \quad (4.20)$$

Onde  $\theta^*$  representa o conjunto de parâmetros que minimiza a função  $f(\theta)$ . Observa-se um exemplo de circuito variacional na figura 10. As portas quânticas  $RY_{\theta_1}$  e CNOT são definidas abaixo:

$$RY_{\theta_1} = \begin{pmatrix} \cos\left(\frac{\theta_1}{2}\right) & -\text{sen}\left(\frac{\theta_1}{2}\right) \\ \text{sen}\left(\frac{\theta_1}{2}\right) & \cos\left(\frac{\theta_1}{2}\right) \end{pmatrix} \text{ e } \text{CNOT}_{1,2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (4.21)$$

No exemplo considerado, o objetivo é minimizar o operador  $I_1 \otimes Z_2$  alterando o parâmetro  $\theta_1$ .

Figura 9 – Circuito quântico implementando a técnica de *kernel* quântico.

Fonte: Adaptado de Havlicek et al. (2019).

$$I_1 \otimes Z_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (4.22)$$

O estado  $U(\theta)|\Psi\rangle$  é dado por:

$$\begin{aligned} U(\theta)|\Psi\rangle &= \text{CNOT}_{1,2} RY_{\theta_1} |0\rangle_1 \otimes |0\rangle_2 = CX_{1,2} \begin{pmatrix} \cos\left(\frac{\theta_1}{2}\right) & -\text{sen}\left(\frac{\theta_1}{2}\right) \\ \text{sen}\left(\frac{\theta_1}{2}\right) & \cos\left(\frac{\theta_1}{2}\right) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes |0\rangle_2 \\ &= \text{CNOT}_{1,2} \begin{pmatrix} \cos\left(\frac{\theta_1}{2}\right) \\ \text{sen}\left(\frac{\theta_1}{2}\right) \end{pmatrix} \otimes |0\rangle_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \cos\left(\frac{\theta_1}{2}\right) \\ 0 \\ \text{sen}\left(\frac{\theta_1}{2}\right) \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\left(\frac{\theta_1}{2}\right) \\ 0 \\ 0 \\ \text{sen}\left(\frac{\theta_1}{2}\right) \end{pmatrix}. \end{aligned} \quad (4.23)$$

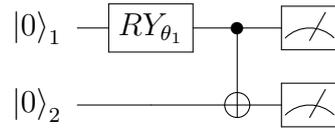
O valor esperado do operador  $I_1 \otimes Z_2$  é:

$$\begin{aligned} \langle \Psi | U(\theta)^\dagger I_1 \otimes Z_2 U(\theta) | \Psi \rangle &= \begin{pmatrix} \cos\left(\frac{\theta_1}{2}\right) & 0 & 0 & \text{sen}\left(\frac{\theta_1}{2}\right) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\left(\frac{\theta_1}{2}\right) \\ 0 \\ 0 \\ \text{sen}\left(\frac{\theta_1}{2}\right) \end{pmatrix} \\ &= \begin{pmatrix} \cos\left(\frac{\theta_1}{2}\right) & 0 & 0 & -\text{sen}\left(\frac{\theta_1}{2}\right) \end{pmatrix} \begin{pmatrix} \cos\left(\frac{\theta_1}{2}\right) \\ 0 \\ 0 \\ \text{sen}\left(\frac{\theta_1}{2}\right) \end{pmatrix} = \cos^2\left(\frac{\theta_1}{2}\right) - \text{sen}^2\left(\frac{\theta_1}{2}\right). \end{aligned} \quad (4.24)$$

O valor mínimo da expressão  $\cos^2\left(\frac{\theta_1}{2}\right) - \text{sen}^2\left(\frac{\theta_1}{2}\right)$  é  $-1$  e ocorre quando:

$$\frac{\theta_1^*}{2} = \frac{\pi}{2} \text{ ou } \frac{\theta_1^*}{2} = \frac{3\pi}{2}, \quad (4.25)$$

Figura 10 – Exemplo de circuito variacional.



Fonte: De autoria própria.

portanto, para minimizar o valor esperado do operador  $I_1 \otimes Z_2$ , e considerando que o valor de  $\theta_1^*$  varia de 0 a  $2\pi$ , tem-se:

$$\theta_1^* = \pi \quad (4.26)$$

## 4.4 Trabalhos Correlatos

A busca de trabalhos correlatos não é exaustiva e apresenta apenas uma fração dos artigos presentes na literatura. Contudo, os trabalhos apresentados auxiliam no entendimento do cenário de *quantum machine learning* aplicado a diagnósticos.

Em relação aos trabalhos correlatos, é válido ressaltar o artigo de [Havlicek et al. \(2019\)](#) o qual apresenta a técnica de *kernel* quântico chamada *ZZ Feature*. Nesse trabalho, os autores utilizam dados artificiais para aplicar as técnicas *ZZ Feature* e *Variational Quantum Classifier*. Os modelos *Variational Quantum Classifier* são circuitos variacionais cujos parâmetros são otimizados para classificar um conjunto de dados de modo a minimizar uma função de *loss*. Os dados artificiais foram gerados por meio dos circuitos quânticos do *kernel ZZ Feature* e um operador  $V$  de dois qubits. Para obter maior variabilidade nos dados, os autores amostraram três operadores aleatoriamente e, para cada operador  $V$ , gera-se um conjunto de dados:

$$y_i(X_i) = \begin{cases} 1, & \text{se } \langle \Phi(X_i) | V^\dagger Z_1 Z_2 V | \Phi(X_i) \rangle \geq 0,3 \text{ e} \\ -1, & \text{caso contrário.} \end{cases} \quad (4.27)$$

Sendo  $X_i$  os atributos de uma amostra  $i$ ,  $y_i$  a classe de uma amostra  $i$  e  $|\Phi(X_i)\rangle$  o estado quântico anterior à medição.

Outro artigo relevante no cenário de *kernels* quânticos é o trabalho de [Glick et al. \(2024\)](#) o qual investiga a aplicação de *kernels* quânticos associados a circuitos variacionais. O uso de circuitos variacionais associados aos circuitos de *kernel* quântico adiciona graus de liberdade a codificação de características, o que leva a ganho de acurácia em problemas de classificação de dados pertencentes a grupos distintos. Tal técnica é chamada de *Quantum Kernel Estimation* (QKE). A utilização de portas quânticas variacionais permite que os estados quânticos sejam reposicionados por meio de um processo de otimização dos

parâmetros dos operadores variacionais. Os autores reportam alta acurácia utilizando dados sintéticos que representam a classificação de dados advindos de distribuições estatísticas distintas (GLICK et al., 2024).

Algoritmos de *quantum machine learning* já foram aplicados no diagnóstico de esquizofrenia (AKSOY et al., 2024), de cardiopatias (ENAD; MOHAMMED et al., 2023; MAHESHWARI et al., 2023; ABDULSALAM; MESHOU; SHAIBA, 2023; HEIDARI; HELLSTERN, 2022; ULLAH et al., 2022a; YANO et al., 2021), de diabetes (PRAKASH et al., 2023; GUPTA et al., 2022; ISHWARYA; CHERUKURI, 2021; MAHESHWARI; GARCIA-ZAPIRAIN; SIERRA-SOSO, 2020), de câncer de mama (DÍAZ-SANTOS; ESCANEZ-EXPOSITO, 2023; YANO et al., 2021; SAINI et al., 2020; POMARICO et al., 2021), de câncer de fígado (KONAR et al., 2023), de COVID-19 (ULLAH et al., 2022b), entre outros. Os algoritmos quânticos utilizados os diagnósticos são QSVM, *Quantum Random Forests*, *Quantum Decision Trees*, *Quantum Neural Networks*, *Quantum Convolutional Neural Networks*, *Quantum Random Access Coding*, *Quantum Long Short Term Memory*, *Quantum K-Nearest Neighbor* e *Variational Quantum Classifier* (ULLAH; GARCIA-ZAPIRAIN, 2024).

A principal limitação relatada está na quantidade de qubits, no tempo de vida dos qubits e nas taxas de erro dos computadores quânticos atuais (ULLAH; GARCIA-ZAPIRAIN, 2024; ENAD; MOHAMMED et al., 2023). O desempenho de técnicas de *quantum machine learning* é sensível a codificação dos dados em estados quânticos, ademais, é necessário que testes com diferentes conjuntos de dados sejam feitos para avaliar a robustez e capacidade de generalização dos modelos. Uma técnica para aumentar robustez e a generalização dos modelos é o uso de conjuntos (*ensembles*) de classificadores quânticos. Por fim, modelos computação quântica possuem baixa interpretabilidade dado o elevado número de dimensões, portanto, são necessárias técnicas que ampliem a capacidade de interpretação dos resultados para sua aplicação na medicina (ULLAH; GARCIA-ZAPIRAIN, 2024).

Em relação aos benefícios do uso de *quantum machine learning*, tem-se o grande número de dimensões dos modelos, a adição de não-linearidades, a capacidade de processar grandes quantidades de informação simultaneamente graças a superposição e o uso de emaranhamento para processos de otimização e representação de dados. Ademais, os modelos quânticos apresentam maior acurácia que os modelos clássicos nos trabalhos de Prakash et al. (2023), Abdulsalam, Meshoul e Shaiba (2023), Díaz-Santos e Escanez-Exposito (2023), Ullah et al. (2022b), Pomarico et al. (2021), entre outros (ULLAH; GARCIA-ZAPIRAIN, 2024).

Neste capítulo, abordamos o algoritmo QSVM, formas de codificação em computadores quânticos e circuitos variacionais. *Quantum machine learning* apresenta vantagem na quantidade de informação codificada ao comparar-se bits e qubits. Ademais, o emaranhamento de estados leva a novas formas de representar dados e informação (ULLAH;

GARCIA-ZAPIRAIN, 2024; ENAD; MOHAMMED et al., 2023).

## 5 Desenvolvimento do Projeto

### 5.1 Procedimento Metodológico

O método do trabalho consiste em comparar as métricas de desempenho como acurácia, precisão e revocação obtidas em um problema de classificação de indivíduos com esquizofrenia com os resultados obtidos em [Aksoy et al. \(2024\)](#). O intuito das comparações está em investigar o desempenho de modelos de *quantum machine learning* em problemas reais. Além das métricas de desempenho, realiza-se análises relacionadas às características únicas de algoritmos quânticos observando suas fronteiras de decisão.

A técnica de *Kernel* Quântico escolhida é apresentada no trabalho de [Havlicek et al. \(2019\)](#) e foi utilizado no artigo de [Aksoy et al. \(2024\)](#), tendo desempenho comparável a algoritmos clássicos. É válido ressaltar que a computação quântica e suas aplicações ainda encontram-se em um nível de maturidade baixo ([FINEP, 2022](#); [\(IAS\), 2022](#)), pois os computadores quânticos atuais apresentam tanto o número de qubits quanto as taxas de erro insuficientes para a realização de computação quântica tolerante a faltas ([HARROW; MONTANARO, 2017](#); [PERAL-GARCÍA; CRUZ-BENITO; GARCÍA-PEÑALVO, 2024](#); [JADHAV; RASOOL; GYANCHANDANI, 2023](#); [PRESKILL, 2018](#)).

*Quantum machine learning* é descrito como uma técnica promissora em problemas relacionados a exames de eletroencefalogramas para previsão de esquizofrenia, capaz de apresentar uma acurácia igual ou melhor que métodos de *machine learning* clássico de acordo com [Aksoy et al. \(2024\)](#). Porém, tal afirmação não está alinhada com o que é apresentado nos trabalhos de [Peral-García, Cruz-Benito e García-Peñalvo \(2024\)](#) e [Jadhav, Rasool e Gyanchandani \(2023\)](#), nos quais descreve-se os desafios da aplicação de computação quântica.

É válido ressaltar que o presente trabalho não explora as especificidades sobre o uso de IRM e EEG no diagnóstico de esquizofrenia, contudo, explora-se as características de técnicas clássicas e quânticas dos algoritmos de *machine learning*.

### 5.2 Ferramentas Utilizadas

Os modelos de machine learning e os circuitos quânticos foram executados por meio da linguagem de programação Python 3.10.6 utilizando, principalmente, as bibliotecas Scikit-Learn <sup>1</sup> e Qiskit <sup>2</sup>. Para a execução dos kernels quânticos, utiliza-se a biblioteca

---

<sup>1</sup> [Pedregosa et al. \(2011\)](#).

<sup>2</sup> [Javadi-Abhari et al. \(2024\)](#).

Qiskit Machine Learning a qual contém o código da técnica *ZZ Feature* (HAVLICEK et al., 2019).<sup>3</sup> Tanto o código quanto o desenvolvimento da técnica de kernel quântico advém do trabalho Havlicek et al. (2019)<sup>4</sup>. As bibliotecas auxiliares utilizadas para organizar e processar os dados são o Numpy, Pandas, Matplotlib e Plotly.

O estudo será realizado sobre o conjunto de dados usado no artigo Çetin et al. (2014) e retirado do site do Kaggle. Inicialmente, são 86 exemplos, sendo 46 exemplos da classe 0 (não diagnosticado com esquizofrenia) e 40 exemplos da classe 1 (diagnosticado com esquizofrenia), cada exemplo possui 378 atributos para o IRM funcional e 32 atributos para o IRM estrutural, esses dados foram pré-processados utilizando-se as técnicas SPM-5, INRIalign, GICA, entre outras conforme descrito em Çetin et al. (2014). Sendo assim, realiza-se uma análise de variância nos atributos tanto funcionais quanto estruturais com o intuito de selecionar os atributos que apresentam maior expressividade para a classificação.

Para verificar o desempenho do modelo divide-se o conjunto de dados em duas partes, sendo uma das partes utilizada para treino dos modelos formada por 80% dos dados. Os 20% restantes são utilizados para testar os modelos. Entre as técnicas de *Machine Learning* a serem empregadas, tem-se K-NN, Árvores de Decisão, Floresta Aleatória, Naive Bayes e SVM de forma análoga ao que foi aplicado em Aksoy et al. (2024), entre os algoritmos quânticos, são utilizados dois *kernels* quânticos *Z feature* e o *ZZ Feature*, a plataforma utilizada será a IBM Quantum para simular e executar os circuitos quânticos na biblioteca Qiskit (JAVADI-ABHARI et al., 2024).

Os algoritmos precisam apresentar uma acurácia maior que 53% aproximadamente considerando a distribuição dos dados, um dos quesitos mais importantes no desempenho é a revocação ou sensibilidade, pois um modelo que tenha alta revocação pode auxiliar os especialistas a distinguir dentro de um conjunto de diversos transtornos mentais e de sintomas semelhantes à esquizofrenia.

---

<sup>3</sup> Qiskit Quantum Machine Learning, disponível em: <https://qiskit-community.github.io/qiskit-machine-learning/index.html>. Acesso em: 05 dez. 2024.

<sup>4</sup> Kernel Quântico, disponível em: [https://qiskit-community.github.io/qiskit-machine-learning/tutorials/03\\_quantum\\_kernel.html](https://qiskit-community.github.io/qiskit-machine-learning/tutorials/03_quantum_kernel.html). Acesso em: 05 dez. 2024.

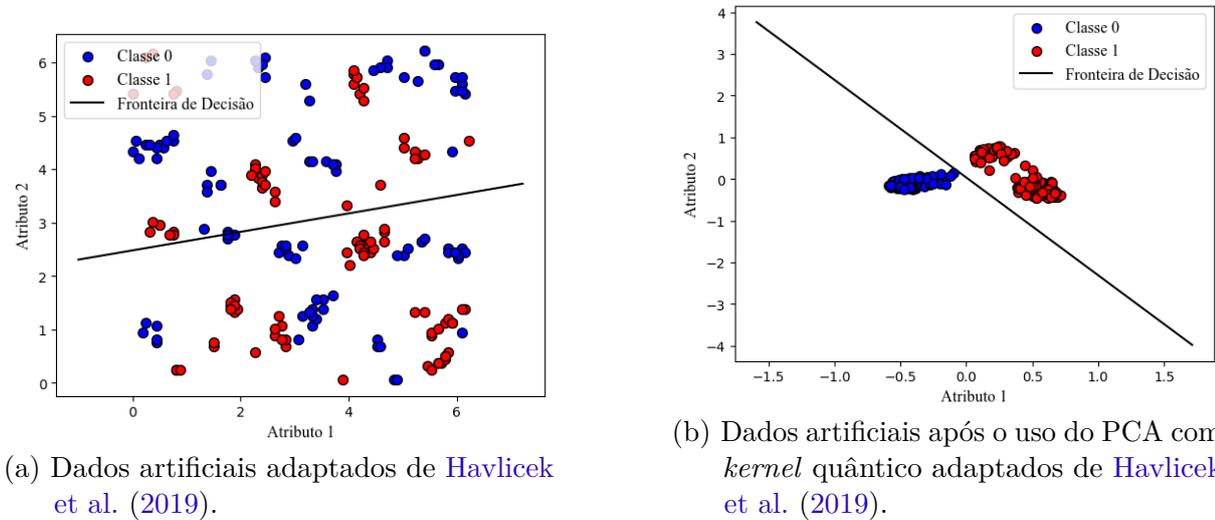
## 6 Resultados Obtidos e Análises

O trabalho de [Aksoy et al. \(2024\)](#) tem como objetivo investigar a eficácia da classificação de indivíduos com esquizofrenia utilizando algoritmos de aprendizado de máquina. Para isso, os autores utilizam um conjunto de exames de eletroencefalograma nos quais é feito um pré-processamento utilizando *Discrete Wavelet Transform* (DWT) nos sinais medidos. Além disso, realiza-se comparações com outros artigos os quais utilizam o mesmo conjunto de dados, porém, métodos de seleção de características e algoritmos de classificação diferentes. De acordo com artigo de [Aksoy et al. \(2024\)](#), os autores apresentam acurácia de 100% com algoritmos tanto clássicos (K-NN e regressão logística) quanto quânticos (QSVM).

É válido ressaltar que o artigo de [Sairamya, Subathra e Thomas George \(2022\)](#) também obteve 100% de acurácia utilizando o mesmo conjunto de dados com técnicas de pré-processamento semelhantes àquelas presentes em [Aksoy et al. \(2024\)](#). No artigo de [Sairamya, Subathra e Thomas George \(2022\)](#), os autores obtêm 100% de acurácia com a técnica de redes neurais e os algoritmos de pré-processamento *Relaxed Local Neighbour Difference Pattern* (RLNDiP) e DWT. Logo, é possível que parte da alta acurácia obtida em ambos os trabalhos esteja relacionada com o pré-processamento do algoritmo DWT. Outros trabalhos relacionados a detecção de IRM anormais ([CHAPLOT; PATNAIK; JAGANNATHAN, 2006](#); [ZHANG et al., 2011](#)).

O trabalho de [Aksoy et al. \(2024\)](#) não especifica o kernel utilizado na técnica de SVM clássica utilizada em seu trabalho, contudo, o trabalho de [Sharma, Tripathi e Sekhon \(2023\)](#) é citado como justificativa para a utilização do método SVM. No trabalho de [Sharma, Tripathi e Sekhon \(2023\)](#), o kernel utilizado é o *Radial Basis Function* (RBF).

A distribuição dos dados considerando os três atributos, SBM61, SBM67 e FNC183, é dada na figura [12a](#). O FNC 183 relaciona as regiões *visual network* (29) e *auditory network* (71). O SBM 61 corresponde a *GM Visual cortex V1 BA17* ([YSTAD et al., 2010](#)). Sobre o atributo SBM 67, não foi encontrado nenhuma informação no artigo original ou em suas referências. Os pontos amarelos indicam os indivíduos os quais foram diagnosticados com esquizofrenia e os pontos azuis indicam indivíduos que não foram diagnosticados com esquizofrenia. Percebe-se que os dados se concentram em diferentes regiões do espaço, contudo, há pontos de ambas as classes que aparecem em regiões dominadas por pontos da outra classe. No caso, ao analisar a figura [12b](#), percebe-se concentrações dos pontos de ambas as classes e a superfície de decisão é capaz de separá-las. Contudo, há dados de uma classe dentro da concentração de pontos de outra classe. Logo, ainda que o problema apresentado neste trabalho não seja linearmente separável, percebe-se que o problema é,

Figura 11 – Exemplo de aplicação de PCA com *kernel* quântico.

Fonte: De autoria própria.

em parte, linearmente separável.

A aplicação do *kernel* quântico *ZZ Feature* melhora as métricas de desempenho no conjunto de dados artificiais criados em Havlicek et al. (2019). Na figura 11a, apresenta-se um *dataset* que não é linearmente separável no espaço dos atributos 1 e 2. Contudo, ao aplicar-se a técnica PCA com o *kernel ZZ Feature*, o problema se torna linearmente separável como apresentado na figura 11b. Sendo assim, a efetividade do *kernel* depende da distribuição dos dados nos quais as transformações não lineares são aplicadas.

A etapa de seleção de características revelou que apenas três dos atributos presentes no conjunto de dados eram necessários para atingir uma acurácia de, aproximadamente, 89% utilizando o algoritmo *Random Forest*. A figura 12a mostra a disposição dos dados em relação aos atributos FNC 183, SBM map 67, SBM map 61, sendo que os pontos amarelos indicam o indivíduo foi diagnosticado com esquizofrenia e os pontos azuis indicam indivíduos que não foram diagnosticados com esquizofrenia.

Para o auxílio de diagnóstico em esquizofrenia, o *kernel* quântico apresenta mais desvantagens do que vantagens, pois o conjunto de dados em questão apresenta um comportamento próximo de um conjunto de dados linearmente separáveis. As vantagens da utilização de *kernels* quânticos são o número de dimensões do espaço de Hilbert dos circuitos quânticos e as características específicas de sistemas quânticos como superposição e emaranhamento.

Todavia, a utilização de algoritmos quânticos apresentam desvantagens advindas, principalmente, do baixo nível de maturidade tecnológica (FINEP, 2022; (IAS), 2022). A computação quântica encontra-se em um período conhecido como *Noisy intermediate-scale quantum* (NISQ), no qual as taxas de erro causadas pelos ruídos são maiores do que o

Tabela 8 – Métricas de desempenho para cada um dos métodos utilizados.

Método	Acurácia	Revocação	Precisão
K-NN	67%	58%	88%
Naive Bayes	72%	67%	89%
Árvore de Classificação	67%	50%	100%
Floresta Aleatória	89%	83%	100%
SVM (linear)	72%	58%	100%
QSVM - ZFeature	50%	50%	100%
QSVM - ZZFeature	44% (56%)	17% (83%)	100% (63%)

Fonte: De autoria própria.

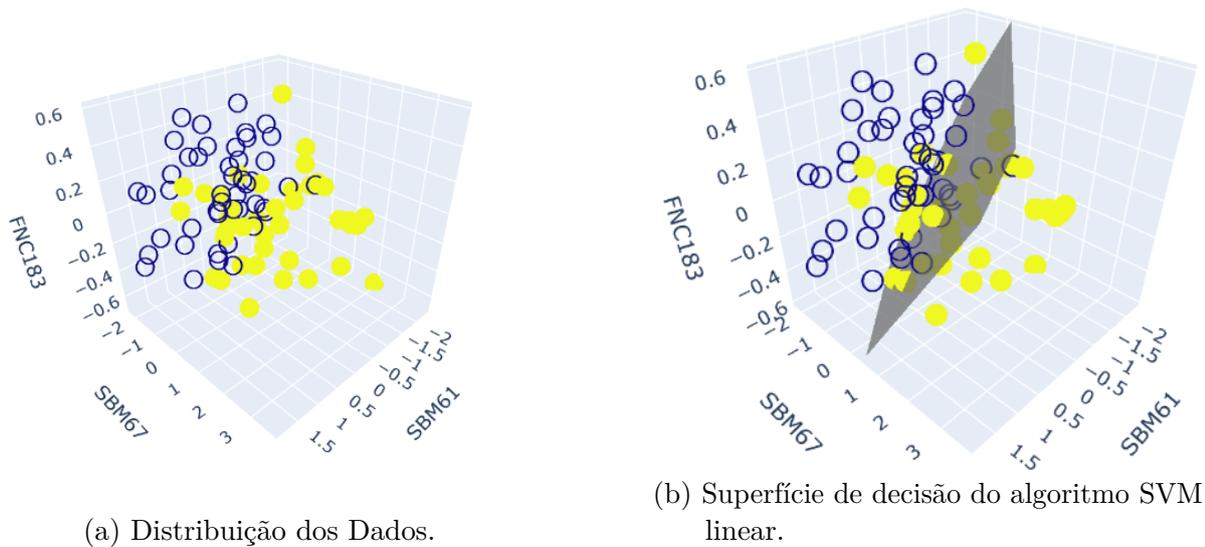
aceitável em um cenário de computação quântica de grande escala (PRESKILL, 2018). Ademais, alterações e atualizações para os algoritmos quânticos já existentes estão em desenvolvimento como o trabalho de Glick et al. (2024). Como explicado anteriormente, esse artigo apresenta o conceito de QKE que consiste na aplicação de portas quânticas variacionais antes da aplicação do circuito do *ZZ Feature*. Outra desvantagem presente em modelos de *machine learning* tanto clássicos quanto quânticos é a falta de interpretabilidade de modelos complexos, tal característica é essencial em aplicações relacionadas à medicina (ENAD; MOHAMMED et al., 2023).

A técnica QKE não influencia o resultado obtido de forma significativa, obtém-se 56% de acurácia aproximadamente. Ou seja, obtém-se o mesmo resultado com a aplicação do *ZZ Feature* sem as portas variacionais. Isso ocorre, pois tem-se o mesmo parâmetro para realizar três rotações em torno do eixo Y, logo, há apenas um grau de liberdade no processo de otimização. Mais detalhes sobre o algoritmo QKE pode ser encontrado no site da comunidade de *Quantum Machine Learning* do Qiskit (GLICK et al., 2024).

Em comparação com o artigo de Aksoy et al. (2024), o presente trabalho apresenta métricas de desempenho diferentes em relação a ordem dos algoritmos com melhor desempenho. Como descrito anteriormente, os algoritmos com maior métrica de desempenho são o K-NN, o QSVM e a regressão logística em Aksoy et al. (2024), enquanto o algoritmo de floresta aleatória possui a maior acurácia entre as técnicas utilizadas neste trabalho. De acordo com Grinsztajn, Oyallon e Varoquaux (2022), métodos derivados das técnicas árvore de decisão e floresta aleatória têm o melhor desempenho em problemas de dados tabulares.

Com relação aos resultados obtidos no presente trabalho, as fronteiras de decisão são apresentadas nas figuras no apêndice, tais figuras representam a fronteira de decisão para cada algoritmo considerando uma combinação de duas características das três características selecionadas. Portanto, cada figura possui três gráficos os quais representam cada uma das possíveis combinações. As fronteiras de decisão dos algoritmos são diferentes entre si, ou seja, cada algoritmo divide o espaço de modo a refletir as características de seu

Figura 12 – Divisão dos dados utilizados utilizando o algoritmo SVM com *kernel* linear.



Fonte: De autoria própria.

funcionamento. Na figura 13, percebe-se que o algoritmo K-NN não possui um formato definido em sua fronteira de decisão entre os três gráficos, pois o algoritmo leva em consideração a distância entre a amostra a ser classificada e o conjunto de dados de treino. Ademais, a distribuição dos dados é diferente para cada combinação de atributos (FIX, 1985; COVER; HART, 1967).

Outro exemplo de como as características de um algoritmo impactam sua fronteira de decisão é a figura 15. Nessa figura, percebe-se os intervalos presentes nos nós da árvore de decisão. Além disso, são notáveis as semelhanças entre as fronteiras de decisão do algoritmo de floresta aleatória e árvores de decisão.

O algoritmo SVM utilizando um *kernel* linear apresenta uma reta como fronteira de decisão conforme apresentado no capítulo 2. A utilização de diferentes *kernels* permite que a fronteira de decisão separe o conjunto de dados com superfícies diferentes as quais impactam diretamente nas métricas de desempenho.

Os *kernels* quânticos *Z Feature* e *ZZ Feature* apresentam fronteiras de decisão distintas de todos os outros algoritmos vistos. Observa-se que as figuras 21 e 22 apresentam repetições de padrões. Tais repetições devem-se ao fato de que as portas quânticas de rotação são periódicas, pois a variação dos atributos na fórmula  $(\pi - x_1)(\pi - x_2)$  ao longo dos eixos leva a repetição de padrões.

Além disso, as matrizes de confusão apresentam o número de verdadeiros positivos, falsos positivos, falsos negativos e verdadeiros negativos para cada algoritmo utilizado. Para a elaboração das matrizes de confusão, divide-se o *dataset* em 80% das amostras utilizadas para treino e 20% das amostras utilizadas para teste. Por fim, é válido ressaltar

que os resultados de circuitos quânticos advêm de simulações dentro da biblioteca Qiskit.

Portanto, percebe-se que as técnicas de *Quantum Machine Learning* implementadas possuem características únicas quando comparadas com técnicas clássicas. No caso de [Havlicek et al. \(2019\)](#) e [Aksoy et al. \(2024\)](#), as técnicas quânticas apresentam acurácia melhor ou igual às técnicas clássicas. Porém, neste trabalho, não se observa ganho de acurácia ao utilizar os *kernels* *Z Feature* e *ZZ Feature*.



## 7 Conclusões e Considerações Finais

Computação quântica representa uma nova forma de se realizar computação criada à partir das leis da mecânica quântica. Algoritmos quânticos podem trazer ganho de desempenho por possuir fenômenos como a superposição e o emaranhamento (HARROW; MONTANARO, 2017). Ademais, diagnósticos realizados por meio de modelos de *machine learning* possuem desempenho comparável com o diagnóstico feito por médicos (ZHANG; SHI; WANG, 2023). O diagnóstico de esquizofrenia traz benefícios para os pacientes permitindo a prevenção de crises psicóticas e melhorando o prognóstico do distúrbio mental. Portanto, o trabalho de Aksoy et al. (2024) avalia a eficácia de métodos de *machine learning* no diagnóstico de esquizofrenia.

A comparação dos resultados obtidos com Aksoy et al. (2024) apresentam diferenças no desempenho dos modelos. Em especial, algoritmos derivados de árvores de decisão e floresta aleatória como o XGBoost (CHEN; GUESTRIN, 2016) são o estado da arte em problemas de classificação de dados tabulares (GRINSZTAJN; OYALLON; VAROQUAUX, 2022). Contudo, os modelos que apresentam a maior acurácia são K-NN, regressão logística e QSVM.

No presente trabalho, os algoritmos floresta aleatória, SVM com *kernel* linear e Naive Bayes apresentam as maiores acurácias obtidas para uma divisão de 80% dos dados para treino e 20% para teste. Considerando possíveis variações das métricas de desempenho em relação a diferentes divisões entre treino e teste, as técnicas de *quantum machine learning* apresentam acurácia inferior às técnicas clássicas, todavia, possuem características únicas devido aos postulados da mecânica quântica. Tal diferença é notável ao comparar-se as diferentes fronteiras de decisão contidas no apêndice. Destarte, técnicas quânticas são promissoras em problemas de aprendizado de máquina tendo em vista a atual maturidade tecnológica da computação quântica (HARROW; MONTANARO, 2017; PERAL-GARCÍA; CRUZ-BENITO; GARCÍA-PEÑALVO, 2024; JADHAV; RASOOL; GYANCHANDANI, 2023; (IAS), 2022).

### 7.1 Limitações do Trabalho

Neste trabalho, simulamos os algoritmos quânticos levando a resultados que não consideram os ruídos presentes na computação quântica. Como dito anteriormente, a computação quântica está na era NISQ, assim, a taxa de erros presente nos computadores quânticos é substancial e influencia no uso dos algoritmos quânticos. Sendo assim, os circuitos quânticos são simulados para analisar a eficácia dos algoritmos de quantum machine learning desconsiderando erros presentes na tecnologia atual.

Ademais, utiliza-se um número reduzido de características por dados (3 características), pois a simulação de circuitos quânticos também é custosa em computadores clássicos. Logo, a simulação de circuitos quânticos com um número maior de qubits torna-se impraticável no computador utilizado.

Por fim, as figuras elaboradas carecem de informações mais precisas sobre o impacto das tecnologias apresentadas na divisão de *datasets* por fronteiras de decisão. A representação da classificação dos dados por regiões no espaço passa a ser desafiadora ao utilizar-se três ou mais parâmetros (LI et al., 2018).

## 7.2 Perspectivas de Continuidade

Para a continuidade do projeto, tem-se a aplicação de técnicas de QSVM conforme descritas em Li et al. (2015). Ademais, como o conjunto de dados disponíveis é de 86 amostras, propõe-se como continuidade do projeto a utilização de outro conjunto mais abundante de dados ou a adição de um outro conjunto de dados ao *dataset* utilizado.

## 7.3 Reconhecimento

A coleta de dados foi feita pelo *Mind Research Network* (descrito parcialmente em Çetin et al. (2014)) e foi financiado pelo *Centers of Biomedical Research Excellence* (COBRE) com código 5P20RR021938/P20GM103472 do *National Institutes of Health* (NIH) pelo Dr. Vince Calhoun.

# Referências

- ABDEL-NASSER, M. et al. Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern. *Expert Systems with Applications*, Elsevier, v. 42, n. 24, p. 9499–9511, 2015. Citado na página 46.
- ABDULSALAM, G.; MESHOUL, S.; SHAIWA, H. Explainable heart disease prediction using ensemble-quantum machine learning approach. *Intell. Autom. Soft Comput*, v. 36, n. 1, p. 761–779, 2023. Citado na página 65.
- AHSAN, M. M. et al. Deep mlp-cnn model using mixed-data to distinguish between covid-19 and non-covid-19 patients. *Symmetry*, MDPI, v. 12, n. 9, p. 1526, 2020. Citado 2 vezes nas páginas 46 e 47.
- AHSAN, M. M.; LUNA, S. A.; SIDDIQUE, Z. Machine-learning-based disease diagnosis: A comprehensive review. *Healthcare*, v. 10, n. 3, 2022. ISSN 2227-9032. Disponível em: <<https://www.mdpi.com/2227-9032/10/3/541>>. Citado na página 47.
- AKSOY, G. et al. Quantum machine-based decision support system for the detection of schizophrenia from eeg records. *Journal of Medical Systems*, v. 48, 2024. ISSN 1573-689X. Disponível em: <<https://doi.org/10.1007/s10916-024-02048-0>>. Citado 14 vezes nas páginas 13, 23, 24, 25, 39, 65, 67, 68, 69, 71, 73, 75, 99 e 100.
- ALI, A. et al. Diabetes diagnosis based on knn. *Iium engineering journal*, v. 21, n. 1, p. 175–181, 2020. Citado na página 46.
- ALLEN, E. et al. Tracking whole-brain connectivity dynamics in the resting state. *Cereb Cortex*, 2014. Citado na página 22.
- ALLEN, E. et al. A baseline for the multivariate comparison of resting-state networks. *Front Syst Neurosci*, 2011. Citado na página 22.
- ANNABEL, L. S. P.; SREENIDHI, S.; VISHALI, N. A novel diagnosis system for parkinson’s disease using k-means clustering and decision tree. In: SPRINGER. *Communication and Intelligent Systems: Proceedings of ICCIS 2020*. [S.l.], 2021. p. 607–615. Citado na página 46.
- ASMAE, O. et al. Parkinson’s disease identification using knn and ann algorithms based on voice disorder. In: IEEE. *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*. [S.l.], 2020. p. 1–6. Citado na página 46.
- BARBER, D. *Bayesian reasoning and machine learning*. [S.l.]: Cambridge University Press, 2012. Citado na página 29.
- BENNETT, M. Schizophrenia: susceptibility genes, dendritic-spine pathology and gray matter loss. *Progress in Neurobiology*, v. 95, n. 3, p. 275–300, 2011. ISSN 0301-0082. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0301008211001481>>. Citado 2 vezes nas páginas 21 e 22.

- BIAMONTE, J. et al. Quantum machine learning. *Nature*, Springer Science and Business Media LLC, v. 549, n. 7671, p. 195–202, set. 2017. ISSN 1476-4687. Disponível em: <<http://dx.doi.org/10.1038/nature23474>>. Citado 3 vezes nas páginas 57, 58 e 59.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, p. 5–32, 2001. Citado na página 35.
- BRUNETTI, A. et al. Detection and segmentation of kidneys from magnetic resonance images in patients with autosomal dominant polycystic kidney disease. In: SPRINGER. *Intelligent Computing Theories and Application: 15th International Conference, ICIC 2019, Nanchang, China, August 3–6, 2019, Proceedings, Part II 15*. [S.l.], 2019. p. 639–650. Citado na página 46.
- CEREZO, M. et al. Challenges and opportunities in quantum machine learning. *Nature Computational Science*, v. 2, p. 567 – 576, 2022. Disponível em: <<https://api.semanticscholar.org/CorpusID:252323115>>. Citado na página 59.
- CHAPELLE, O.; SCHOLKOPF, B.; ZIEN, A. Semi-supervised learning. 2006. *Cambridge, Massachusetts: The MIT Press View Article*, v. 2, p. 1, 2006. Citado na página 27.
- CHAPLOT, S.; PATNAIK, L. M.; JAGANNATHAN, N. R. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical signal processing and control*, Elsevier, v. 1, n. 1, p. 86–92, 2006. Citado 2 vezes nas páginas 46 e 69.
- CHEN, J. et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. *Scientific reports*, Nature Publishing Group UK London, v. 10, n. 1, p. 19196, 2020. Citado 2 vezes nas páginas 46 e 47.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 785–794. ISBN 9781450342322. Disponível em: <<https://doi.org/10.1145/2939672.2939785>>. Citado 2 vezes nas páginas 24 e 75.
- CHYZHYK, D.; SAVIO, A.; GRAÑA, M. Computer aided diagnosis of schizophrenia on resting state fmri data by ensembles of elm. *Neural Networks*, v. 68, p. 23–33, 2015. ISSN 0893-6080. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0893608015000805>>. Citado 2 vezes nas páginas 21 e 24.
- CORTES, C. Support-vector networks. *Machine Learning*, 1995. Citado na página 40.
- COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21–27, 1967. Citado 2 vezes nas páginas 32 e 72.
- DEZA, E. et al. *Encyclopedia of distances*. [S.l.]: Springer, 2009. Citado na página 32.
- DÍAZ-SANTOS, S.; ESCANEZ-EXPOSITO, D. Classical vs. quantum machine learning for breast cancer detection. In: IEEE. *2023 19th International Conference on the Design of Reliable Communication Networks (DRCN)*. [S.l.], 2023. p. 1–5. Citado na página 65.

ENAD, H. G.; MOHAMMED, M. A. et al. A review on artificial intelligence and quantum machine learning for heart disease diagnosis: Current techniques, challenges and issues, recent developments, and future directions. *Fusion: Pract Appl (FPA)*, v. 11, n. 1, p. 08–25, 2023. Citado 5 vezes nas páginas 25, 47, 65, 66 e 71.

FERNÁNDEZ-EDREIRA, D.; LIÑARES-BLANCO, J.; FERNANDEZ-LOZANO, C. Machine learning analysis of the human infant gut microbiome identifies influential species in type 1 diabetes. *Expert Systems with Applications*, v. 185, p. 115648, 2021. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417421010393>>. Citado na página 46.

FEYNMAN, R. P. Simulating physics with computers. v. 21, p. 467–488, June 1982. ISSN 1572-9575. Disponível em: <<https://doi.org/10.1007/BF02650179>>. Citado na página 25.

FINEP. *Definição de Nível de Maturidade Tecnológica*. 2022. <[http://www.finep.gov.br/images/chamadas-publicas/2022/15\\_07\\_2022\\_EL\\_Anexo\\_7-Definicao\\_do\\_Nivel\\_de\\_Maturidade\\_Tecnologica.pdf/](http://www.finep.gov.br/images/chamadas-publicas/2022/15_07_2022_EL_Anexo_7-Definicao_do_Nivel_de_Maturidade_Tecnologica.pdf/)>. [Acesso em: 10 dez. 2024]. Citado 2 vezes nas páginas 67 e 70.

FITRIYANI, N. L. et al. Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. *IEEE Access*, v. 7, p. 144777–144789, 2019. Citado na página 46.

FIX, E. *Discriminatory analysis: nonparametric discrimination, consistency properties*. [S.l.]: USAF school of Aviation Medicine, 1985. v. 1. Citado 2 vezes nas páginas 32 e 72.

GLICK, J. R. et al. Covariant quantum kernels for data with group structure. *Nature Physics*, Nature Publishing Group UK London, v. 20, n. 3, p. 479–483, 2024. Citado 3 vezes nas páginas 64, 65 e 71.

GOGTAY, N.; THOMPSON, P. M. Mapping gray matter development: Implications for typical development and vulnerability to psychopathology. *Brain and Cognition*, v. 72, n. 1, p. 6–15, 2010. ISSN 0278-2626. Adolescent Brain Development: Current Themes and Future Directions. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0278262609001626>>. Citado 2 vezes nas páginas 21 e 22.

GRINSZTAJN, L.; OYALLON, E.; VAROQUAUX, G. Why do tree-based models still outperform deep learning on typical tabular data? In: KOYEJO, S. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2022. v. 35, p. 507–520. Disponível em: <[https://proceedings.neurips.cc/paper\\_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets\\_and\\_Benchmarks.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf)>. Citado 3 vezes nas páginas 24, 71 e 75.

GROVER, L. K. A fast quantum mechanical algorithm for database search. In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. [S.l.: s.n.], 1996. p. 212–219. Citado na página 59.

GUPTA, H. et al. Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. *Complex & Intelligent Systems*, Springer, v. 8, n. 4, p. 3073–3087, 2022. Citado na página 65.

- GÜRÜLER, H. A novel diagnosis system for parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. *Neural Computing and Applications*, Springer, v. 28, p. 1657–1666, 2017. Citado na página 46.
- HARROW, A. W.; HASSIDIM, A.; LLOYD, S. Quantum algorithm for linear systems of equations. *Physical Review Letters*, American Physical Society (APS), v. 103, n. 15, out. 2009. ISSN 1079-7114. Disponível em: <<http://dx.doi.org/10.1103/PhysRevLett.103.150502>>. Citado na página 57.
- HARROW, A. W.; MONTANARO, A. Quantum computational supremacy. *Nature*, v. 549, p. 203–209, 2017. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/nature23458>>. Citado 4 vezes nas páginas 23, 25, 67 e 75.
- HAVLICEK, V. et al. Supervised learning with quantum-enhanced feature spaces. *Nature*, v. 567, p. 209–212, 2019. ISSN 7747. Disponível em: <<https://doi.org/10.1038/s41586-019-0980-2>>. Citado 9 vezes nas páginas 25, 61, 62, 63, 64, 67, 68, 70 e 73.
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 24.
- HEIDARI, H.; HELLSTERN, G. Early heart disease prediction using hybrid quantum classification. *arXiv preprint arXiv:2208.08882*, 2022. Citado na página 65.
- HIDARY, J. D. *Quantum computing: an applied approach*. [S.l.]: Springer, 2019. v. 1. Citado 4 vezes nas páginas 49, 51, 55 e 59.
- HOFMANN, M. Support vector machines — kernels and the kernel trick an elaboration for the hauptseminar “ reading club : Support vector machines ”. In: . [s.n.], 2006. Disponível em: <<https://api.semanticscholar.org/CorpusID:35071766>>. Citado na página 40.
- HUANG, Z.; SUN, M.; GUO, C. Automatic diagnosis of alzheimer's disease and mild cognitive impairment based on cnn+ svm networks with end-to-end training. *Computational Intelligence and Neuroscience*, Wiley Online Library, v. 2021, n. 1, p. 9121770, 2021. Citado na página 46.
- (IAS), I. for A. S. *Technology Readiness Level of Quantum Computing Technology (QTRL)*. 2022. <<https://www.fz-juelich.de/en/ias/jsc/about-us/structure/research-groups/qip/technology-readiness-level-of-quantum-computing-technology-qtrl/>>. [Acesso em: 10 dez. 2024]. Citado 3 vezes nas páginas 67, 70 e 75.
- IMRAN, A. A.; AMIN, M. N.; JOHORA, F. T. Classification of chronic kidney disease using logistic regression, feedforward neural network and wide & deep learning. In: *IEEE. 2018 International Conference on Innovation in Engineering and Technology (ICIET)*. [S.l.], 2018. p. 1–6. Citado na página 46.
- ISHWARYA, M.; CHERUKURI, A. K. Quantum-inspired ensemble approach to multi-attributed and multi-agent decision-making. *Applied Soft Computing*, Elsevier, v. 106, p. 107283, 2021. Citado na página 65.

JADHAV, A.; RASOOL, A.; GYANCHANDANI, M. Quantum machine learning: Scope for real-world problems. *Procedia Computer Science*, v. 218, p. 2612–2625, 2023. ISSN 1877-0509. International Conference on Machine Learning and Data Engineering. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050923002351>>. Citado 4 vezes nas páginas 23, 25, 67 e 75.

JAVADI-ABHARI, A. et al. *Quantum computing with Qiskit*. 2024. Citado 2 vezes nas páginas 67 e 68.

JIN, C. et al. Development and evaluation of an artificial intelligence system for covid-19 diagnosis. *Nature communications*, Nature Publishing Group UK London, v. 11, n. 1, p. 5088, 2020. Citado 2 vezes nas páginas 46 e 47.

KONAR, D. et al. 3-d quantum-inspired self-supervised tensor network for volumetric segmentation of medical images. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, 2023. Citado na página 65.

KUANG, J. et al. Prediction of transition from mild cognitive impairment to alzheimer's disease based on a logistic regression–artificial neural network–decision tree model. *Geriatrics & Gerontology International*, Wiley Online Library, v. 21, n. 1, p. 43–47, 2021. Citado na página 46.

LAZLI, L.; BOUKADOUM, M.; MOHAMED, O. A. A survey on computer-aided diagnosis of brain disorders through mri based on machine learning and data mining methodologies with an emphasis on alzheimer disease diagnosis and the contribution of the multimodal fusion. *Applied Sciences*, v. 10, n. 5, 2020. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/10/5/1894>>. Citado na página 47.

LI, H. et al. Visualizing the loss landscape of neural nets. *Advances in neural information processing systems*, v. 31, 2018. Citado na página 76.

LI, L. et al. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology*, Radiological society of north america, 2020. Citado 2 vezes nas páginas 46 e 47.

LI, Z. et al. Experimental realization of a quantum support vector machine. *Phys. Rev. Lett.*, American Physical Society, v. 114, p. 140504, Apr 2015. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevLett.114.140504>>. Citado na página 76.

LIU, Y.; ARUNACHALAM, S.; TEMME, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, Nature Publishing Group UK London, v. 17, n. 9, p. 1013–1017, 2021. Citado na página 59.

LLOYD, S.; MOHSENI, M.; REBENTROST, P. Quantum principal component analysis. *Nature Physics*, Springer Science and Business Media LLC, v. 10, n. 9, p. 631–633, jul. 2014. ISSN 1745-2481. Disponível em: <<http://dx.doi.org/10.1038/nphys3029>>. Citado na página 57.

LO, C.-M. et al. Feasibility testing: Three-dimensional tumor mapping in different orientations of automated breast ultrasound. *Ultrasound in medicine & biology*, Elsevier, v. 42, n. 5, p. 1201–1210, 2016. Citado na página 46.

- LOUZÃ, M. R. Detecção precoce: é possível prevenir a esquizofrenia? *Archives of Clinical Psychiatry (São Paulo)*, Faculdade de Medicina da Universidade de São Paulo, v. 34, p. 169–173, 2007. ISSN 0101-6083. Disponível em: <<https://doi.org/10.1590/S0101-60832007000800004>>. Citado na página 21.
- MAHESHWARI, D.; GARCIA-ZAPIRAIN, B.; SIERRA-SOSO, D. Machine learning applied to diabetes dataset using quantum versus classical computation. In: IEEE. *2020 IEEE international symposium on signal processing and information technology (ISSPIT)*. [S.l.], 2020. p. 1–6. Citado na página 65.
- MAHESHWARI, D. et al. Quantum machine learning applied to electronic healthcare records for ischemic heart disease classification. *Hum.-Cent. Comput. Inf. Sci*, v. 13, n. 06, 2023. Citado na página 65.
- MANZAK, D.; ÇETINEL, G.; MANZAK, A. Automated classification of alzheimer’s disease using deep neural network (dnn) by random forest feature elimination. In: IEEE. *2019 14th International Conference on Computer Science & Education (ICCSE)*. [S.l.], 2019. p. 1050–1053. Citado na página 46.
- MARDER, S. R.; CANNON, T. D. Schizophrenia. *New England Journal of Medicine*, v. 381, n. 18, p. 1753–1761, out. 2019. ISSN 0028-4793, 1533-4406. Disponível em: <<http://www.nejm.org/doi/10.1056/NEJMr1808803>>. Citado na página 21.
- MÁRTON, Á.; ASBÓTH, J. K. Coherent errors and readout errors in the surface code. *Quantum*, Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften, v. 7, p. 1116, set. 2023. ISSN 2521-327X. Disponível em: <<https://doi.org/10.22331/q-2023-09-21-1116>>. Citado na página 55.
- MEI, J.; DESROSIERS, C.; FRASNELLI, J. Machine learning for the diagnosis of parkinson’s disease: a review of literature. *Frontiers in aging neuroscience*, Frontiers Media SA, v. 13, p. 633752, 2021. Citado na página 46.
- MOON, W. K. et al. The adaptive computer-aided diagnosis system based on tumor sizes for the classification of breast tumors detected at screening ultrasound. *Ultrasonics*, Elsevier, v. 76, p. 70–77, 2017. Citado na página 46.
- NAVANEETH, B.; SUCHETHA, M. A dynamic pooling based convolutional neural network approach to detect chronic kidney disease. *Biomedical Signal Processing and Control*, Elsevier, v. 62, p. 102068, 2020. Citado na página 46.
- NIELSEN, M. A.; CHUANG, I. L. *Quantum computation and quantum information*. [S.l.]: Cambridge university press, 2010. Citado 5 vezes nas páginas 54, 55, 56, 58 e 59.
- NITHYA, A. et al. Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images. *Measurement*, Elsevier, v. 149, p. 106952, 2020. Citado na página 46.
- NORVIG, P.; RUSSELL, S. *Inteligência artificial: Tradução da 3a Edição*. Elsevier Brasil, 2014. ISBN 9788535251418. Disponível em: <<https://books.google.com.br/books?id=BsNeAwAAQBAJ>>. Citado 4 vezes nas páginas 27, 28, 29 e 35.

- NURROHMAN, A.; ABDULLAH, S.; MURFI, H. Parkinson's disease subtype classification: Application of decision tree, logistic regression and logit leaf model. In: AIP PUBLISHING. *AIP Conference Proceedings*. [S.l.], 2020. v. 2242, n. 1. Citado na página 46.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na página 67.
- PERAL-GARCÍA, D.; CRUZ-BENITO, J.; GARCÍA-PEÑALVO, F. J. Systematic literature review: Quantum machine learning and its applications. *Computer Science Review*, v. 51, p. 100619, 2024. ISSN 1574-0137. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1574013724000030>>. Citado 4 vezes nas páginas 23, 25, 67 e 75.
- POMARICO, D. et al. A proposal of quantum-inspired machine learning for medical purposes: An application case. *Mathematics*, MDPI, v. 9, n. 4, p. 410, 2021. Citado na página 65.
- PRAKASH, B. et al. Comparative performance analysis of quantum algorithm with machine learning algorithms on diabetes mellitus. In: IEEE. *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*. [S.l.], 2023. p. 1178–1183. Citado na página 65.
- PRESKILL, J. Quantum Computing in the NISQ era and beyond. *Quantum*, Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften, v. 2, p. 79, ago. 2018. ISSN 2521-327X. Disponível em: <<https://doi.org/10.22331/q-2018-08-06-79>>. Citado 4 vezes nas páginas 23, 55, 67 e 71.
- QEZELBASH-CHAMAK, J. et al. A survey of machine learning in kidney disease diagnosis. *Machine Learning with Applications*, v. 10, p. 100418, 2022. ISSN 2666-8270. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2666827022000937>>. Citado na página 46.
- QTEAT, H.; AWAD, M. Using hybrid model of particle swarm optimization and multi-layer perceptron neural networks for classification of diabetes. *International Journal of Intelligent Engineering & Systems*, v. 14, n. 3, 2021. Citado na página 46.
- RAJENDRAN, K.; JAYABALAN, M.; THIRUCHELVAM, V. Predicting breast cancer via supervised machine learning methods on class imbalanced data. *International Journal of Advanced Computer Science and Applications*, Science and Information (SAI) Organization Limited, v. 11, n. 8, 2020. Citado na página 46.
- REBENTROST, P.; MOHSENI, M.; LLOYD, S. Quantum support vector machine for big data classification. *Phys. Rev. Lett.*, American Physical Society, v. 113, p. 130503, Sep 2014. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevLett.113.130503>>. Citado na página 40.
- SAINI, S. et al. Quantum driven machine learning. *International Journal of Theoretical Physics*, Springer, v. 59, n. 12, p. 4013–4024, 2020. Citado na página 65.
- SAIRAMYA, N.; SUBATHRA, M.; Thomas George, S. Automatic identification of schizophrenia using eeg signals based on discrete wavelet transform and rln dip technique with ann. *Expert Systems with Applications*, v. 192, p. 116230, 2022. ISSN 0957-4174.

Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417421015438>>. Citado na página 69.

SCHULD, M.; PETRUCCIONE, F. *Machine Learning with Quantum Computers*. [S.l.: s.n.], 2021. ISBN 978-3-030-83097-7. Citado 2 vezes nas páginas 23 e 61.

SEGALL, J. et al. Correspondence between structure and function in the human brain at rest. *Front Neuroinform.*, 2012. Citado na página 22.

SHARMA, R.; TRIPATHI, S.; SEKHON, K. K. Detection of schizophrenia using machine learning. In: *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. [S.l.: s.n.], 2023. p. 505–512. Citado na página 69.

SHARMA, S.; KHANNA, P. Computer-aided diagnosis of malignant mammograms using zernike moments and svm. *Journal of digital imaging*, Springer, v. 28, p. 77–90, 2015. Citado na página 46.

SHETTY, S.; RAO, Y. Svm based machine learning approach to identify parkinson's disease using gait analysis. In: IEEE. *2016 International conference on inventive computation technologies (ICICT)*. [S.l.], 2016. v. 2, p. 1–5. Citado na página 46.

SHOR, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, v. 26, n. 5, p. 1484–1509, 1997. Disponível em: <<https://doi.org/10.1137/S0097539795293172>>. Citado 2 vezes nas páginas 58 e 59.

SONG, Y. et al. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *IEEE/ACM transactions on computational biology and bioinformatics*, IEEE, v. 18, n. 6, p. 2775–2780, 2021. Citado 2 vezes nas páginas 46 e 47.

TANVEER, M. et al. Machine learning techniques for the diagnosis of alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, ACM New York, NY, USA, v. 16, n. 1s, p. 1–35, 2020. Citado na página 46.

TSAO, H.-Y.; CHAN, P.-Y.; SU, E. C.-Y. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. *BMC bioinformatics*, Springer, v. 19, p. 111–121, 2018. Citado na página 46.

ULLAH, U.; GARCIA-ZAPIRAIN, B. Quantum machine learning revolution in healthcare: A systematic review of emerging perspectives and applications. *IEEE Access*, v. 12, p. 11423–11450, 2024. Citado 4 vezes nas páginas 25, 47, 65 e 66.

ULLAH, U. et al. A fully connected quantum convolutional neural network for classifying ischemic cardiopathy. *IEEE Access*, IEEE, v. 10, p. 134592–134605, 2022. Citado na página 65.

ULLAH, U. et al. Severity classification of covid-19 patients data using quantum machine learning approaches. In: IEEE. *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. [S.l.], 2022. p. 1–6. Citado na página 65.

- VARAPRASAD, S. et al. An effective diagnosis of schizophrenia using kernel ridge regression-based optimized rvfl classifier. *Applied Soft Computing*, v. 157, p. 111457, 2024. ISSN 1568-4946. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S156849462400231X>>. Citado 2 vezes nas páginas 21 e 24.
- VENKATESH, S. S. et al. Going beyond a first reader: a machine learning methodology for optimizing cost and performance in breast ultrasound diagnosis. *Ultrasound in medicine & biology*, Elsevier, v. 41, n. 12, p. 3148–3162, 2015. Citado na página 46.
- WALSE, R. S. et al. Effective use of naïve bayes, decision tree, and random forest techniques for analysis of chronic kidney disease. In: SPRINGER. *Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2020, Volume 1*. [S.l.], 2021. p. 237–245. Citado na página 46.
- WHO, W. H. O. *The global burden of disease : 2004 update*. 2008. <<https://www.who.int/publications/i/item/9789241563710>>. Acessado: 19/05/2024. Citado na página 21.
- WITTEN, I. H. et al. *Data Mining: Practical Machine Learning Tools and Techniques*. [S.l.]: Morgan Kaufmann, 2017. Citado 7 vezes nas páginas 13, 28, 29, 33, 34, 40 e 41.
- YANO, H. et al. Efficient discrete feature encoding for variational quantum classifier. *IEEE Transactions on Quantum Engineering*, IEEE, v. 2, p. 1–14, 2021. Citado na página 65.
- YSTAD, M. et al. Subcortical functional connectivity and verbal episodic memory in healthy elderly—a resting state fmri study. *NeuroImage*, v. 52, n. 1, p. 379–388, 2010. ISSN 1053-8119. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1053811910003496>>. Citado na página 69.
- YUE, W. et al. Machine learning with applications in breast cancer diagnosis and prognosis. *Designs*, MDPI, v. 2, n. 2, p. 13, 2018. Citado na página 46.
- ZHANG, B.; SHI, H.; WANG, H. Machine learning and ai in cancer prognosis, prediction, and treatment selection: a critical approach. *Journal of multidisciplinary healthcare*, Taylor & Francis, p. 1779–1791, 2023. Citado na página 75.
- ZHANG, H. The optimality of naive bayes. *Aa*, v. 1, n. 2, p. 3, 2004. Citado na página 29.
- ZHANG, Y. et al. A hybrid method for mri brain image classification. *Expert Systems with Applications*, Elsevier, v. 38, n. 8, p. 10049–10053, 2011. Citado 2 vezes nas páginas 46 e 69.
- ÇETIN, M. S. et al. Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in schizophrenia. *NeuroImage*, v. 97, p. 117–126, 2014. Disponível em: <<https://doi.org/10.1016/j.neuroimage.2014.04.009>>. Citado 4 vezes nas páginas 22, 24, 68 e 76.



# APÊNDICE A – Fronteiras de decisão dos algoritmos analisados.

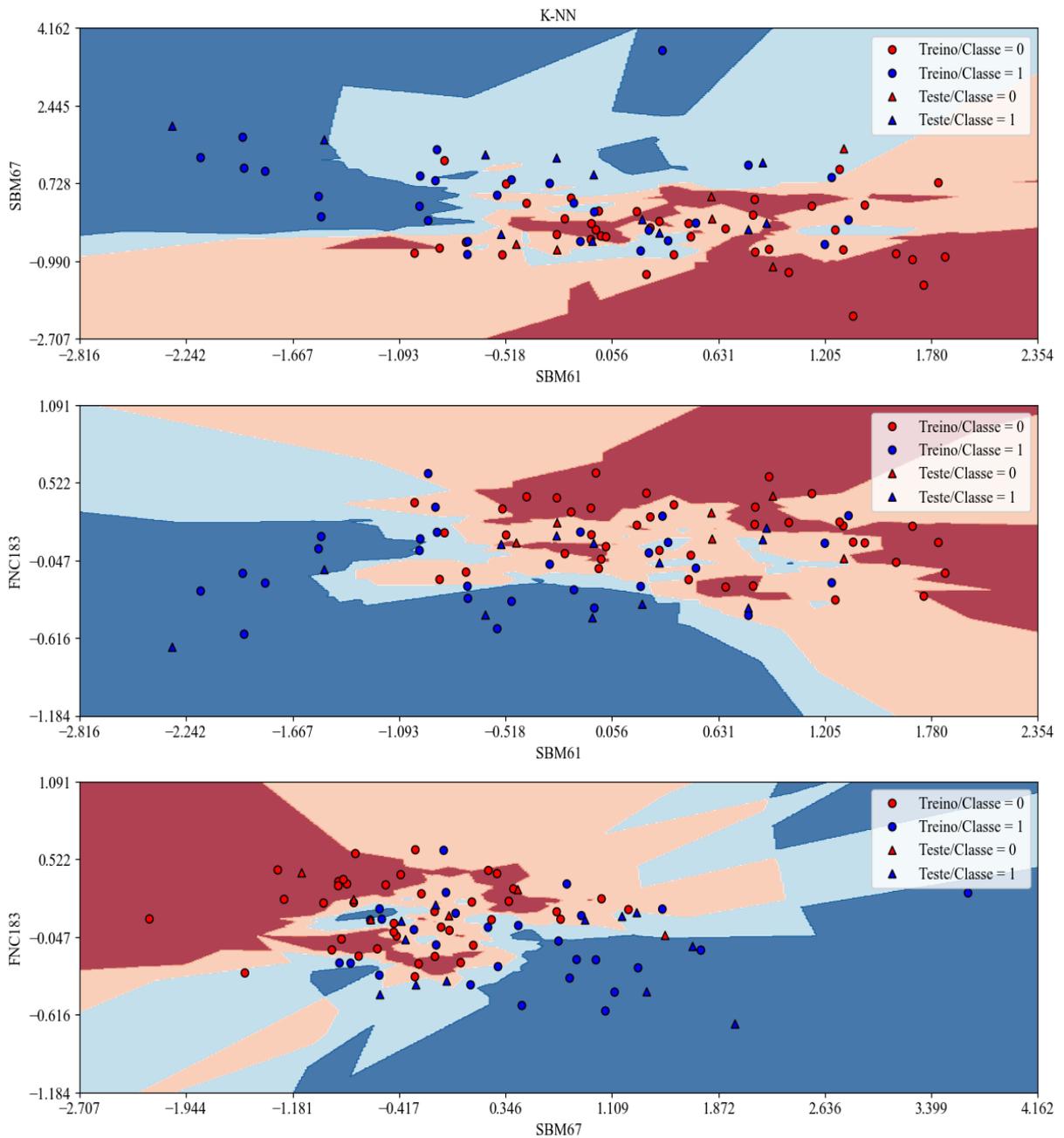


Figura 13 – Fronteira de Decisão do algoritmo K-NN.

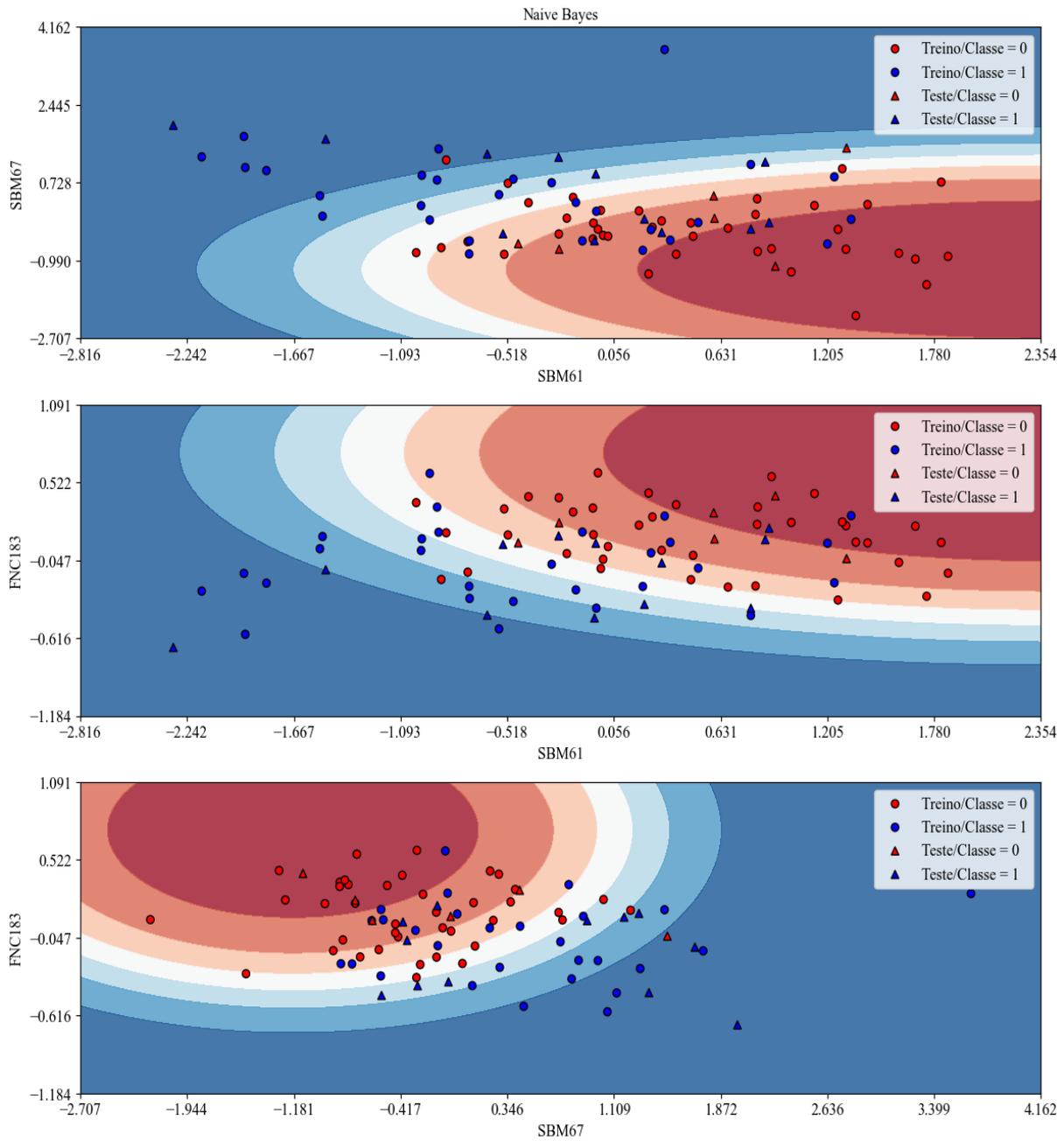


Figura 14 – Fronteira de Decisão do algoritmo Naive Bayes.

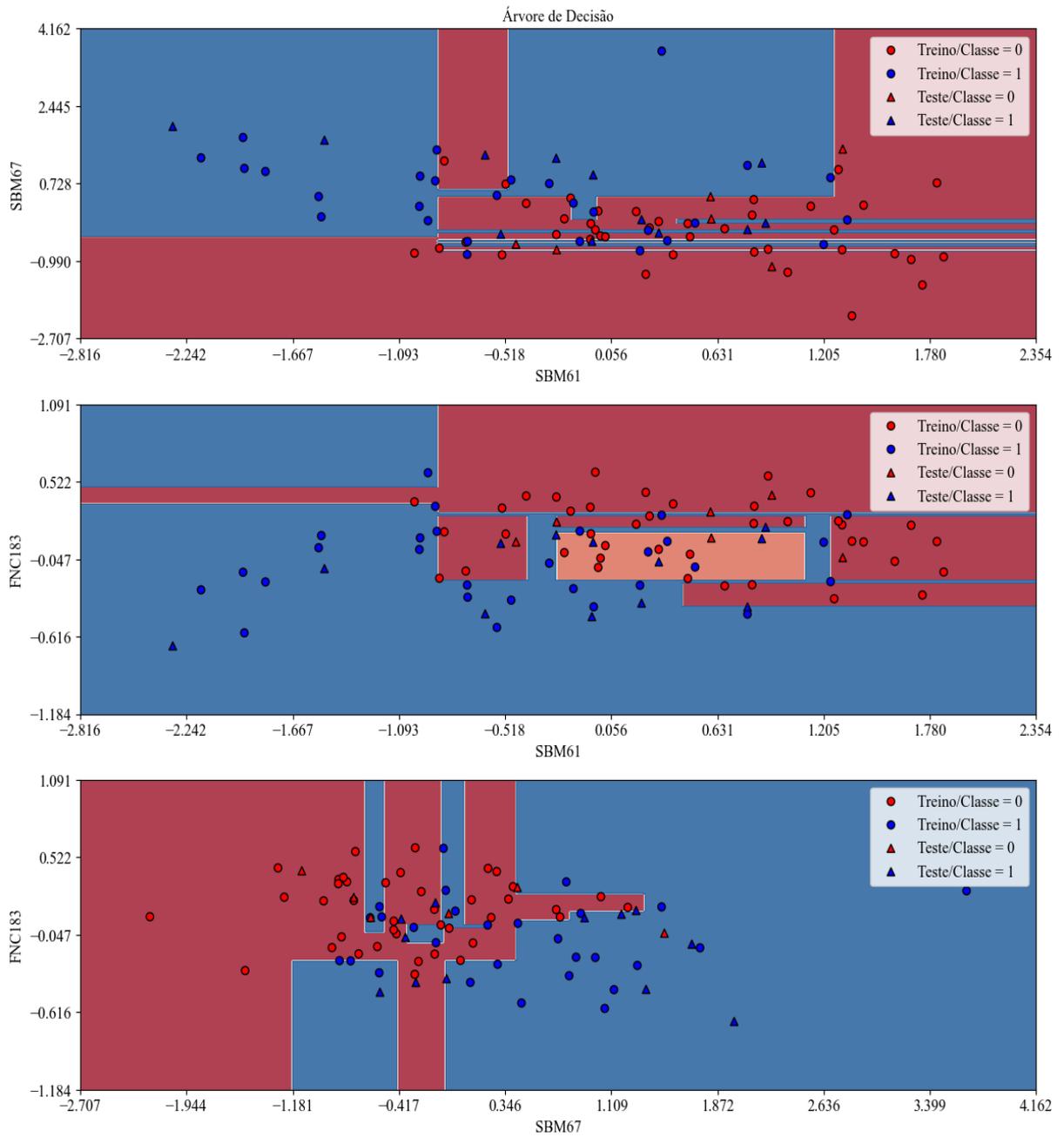


Figura 15 – Fronteira de Decisão do algoritmo Árvore de Decisão.

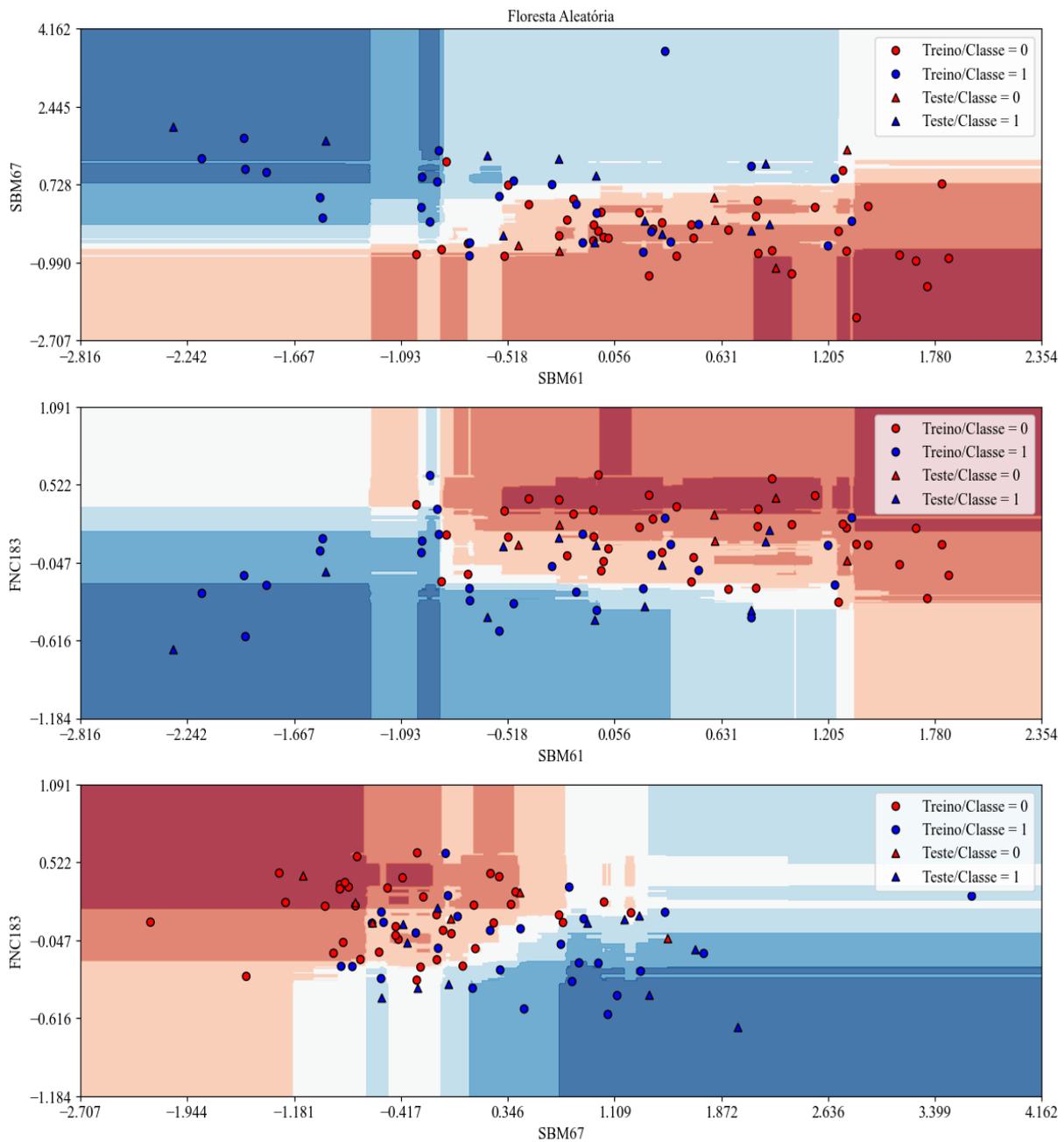


Figura 16 – Fronteira de Decisão do algoritmo Floresta Aleatória.

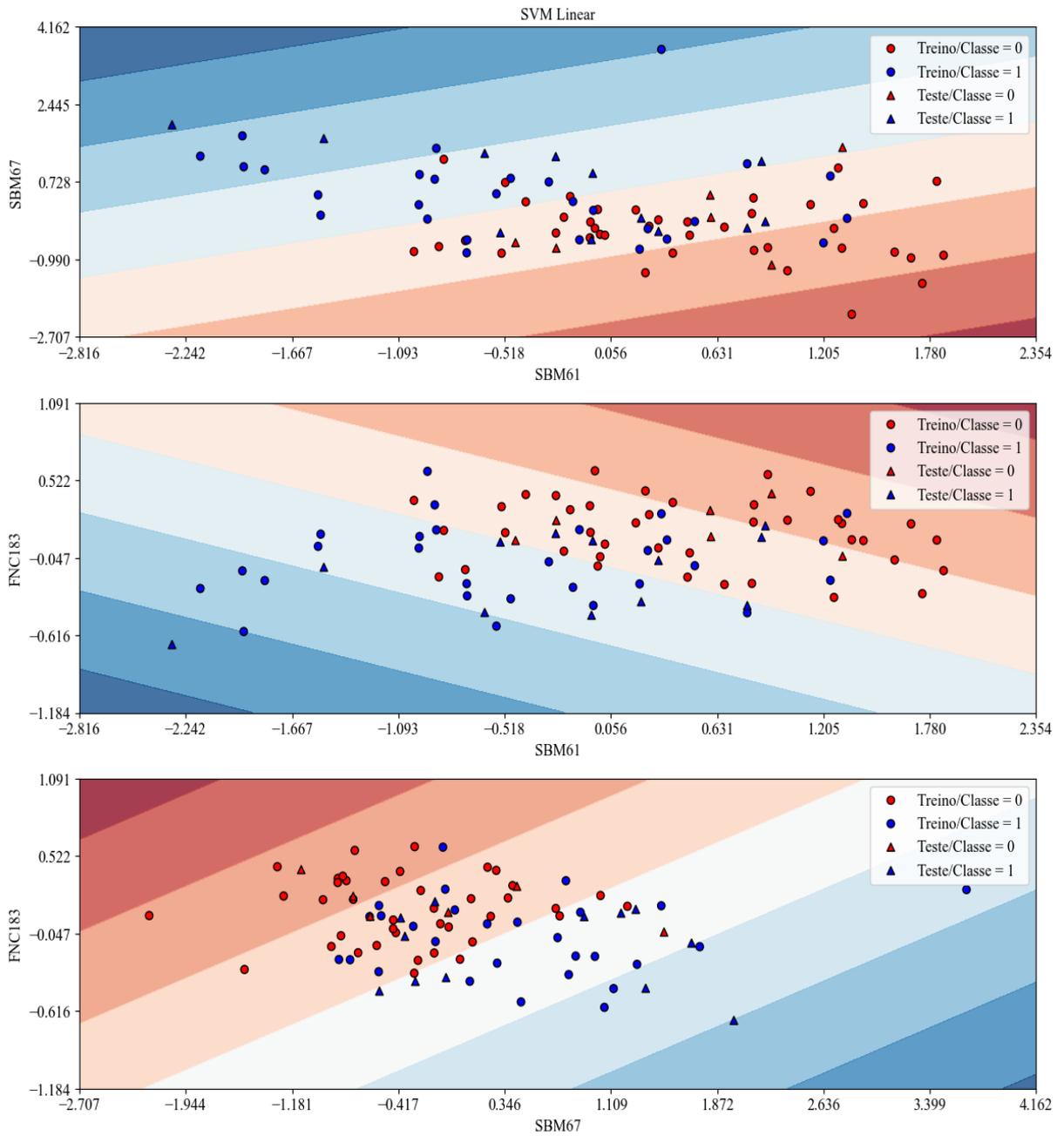
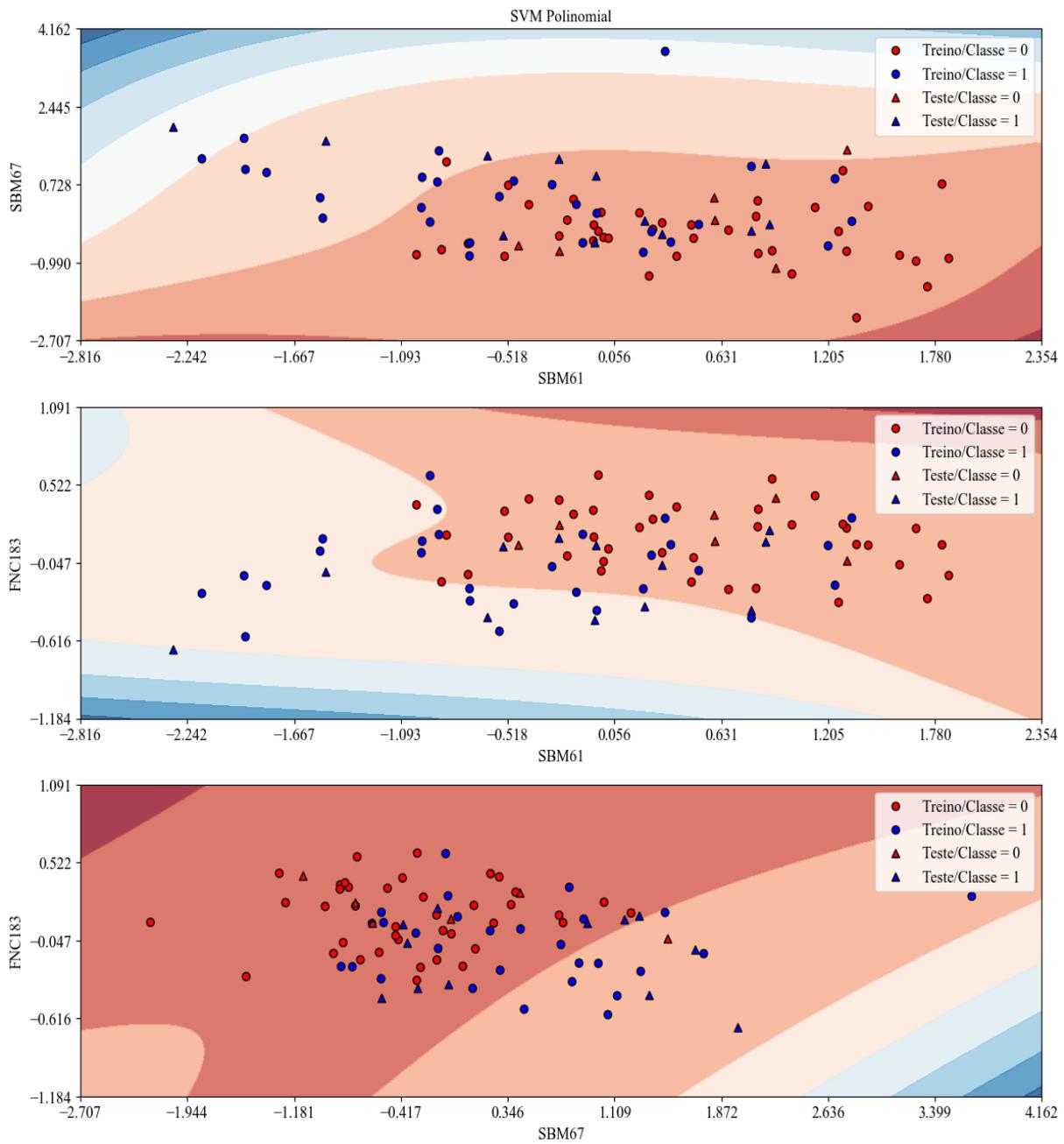


Figura 17 – Fronteira de Decisão do algoritmo SVM com *kernel* linear.

Figura 18 – Fronteira de Decisão do algoritmo SVM com *kernel* polinomial.

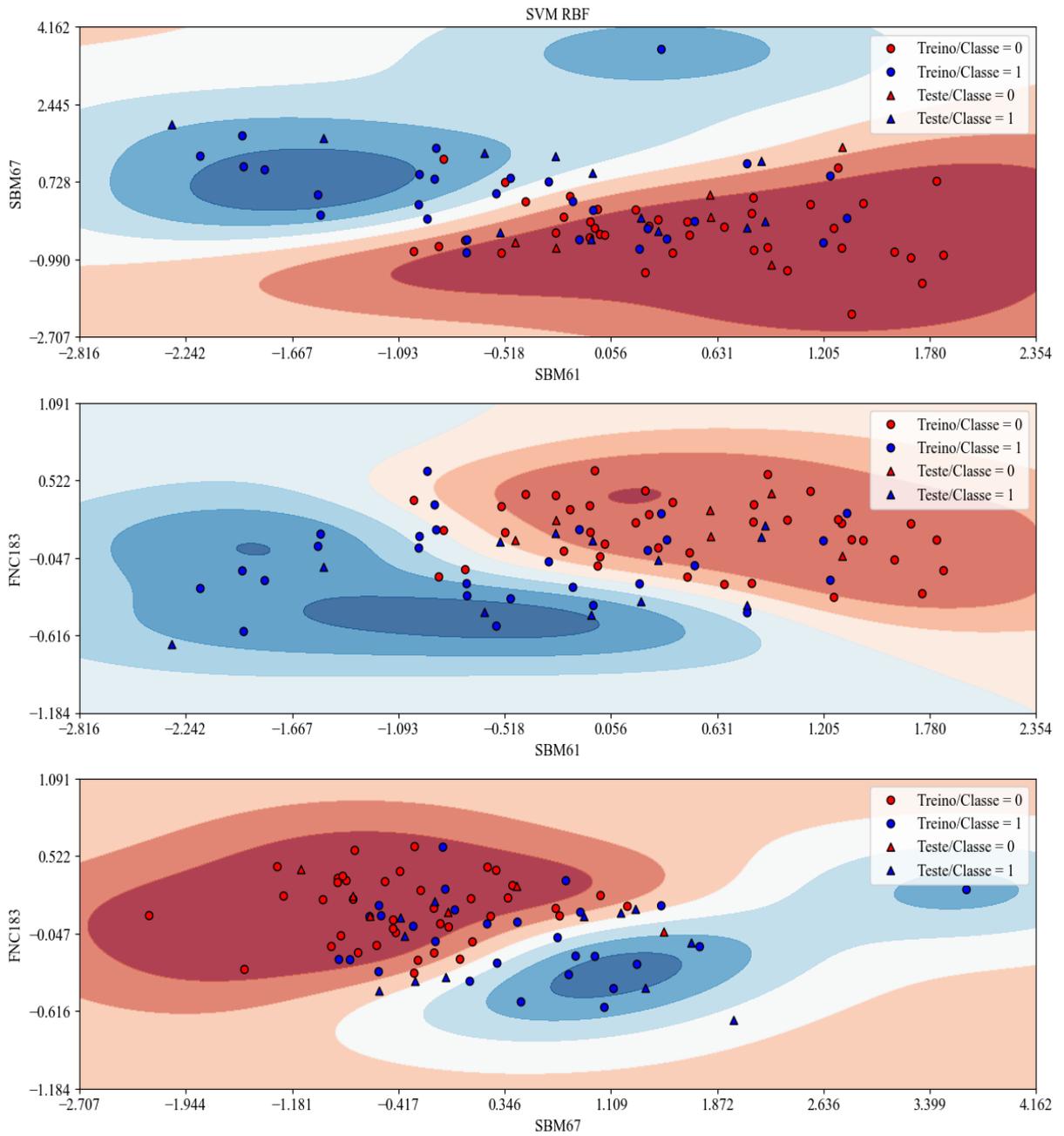
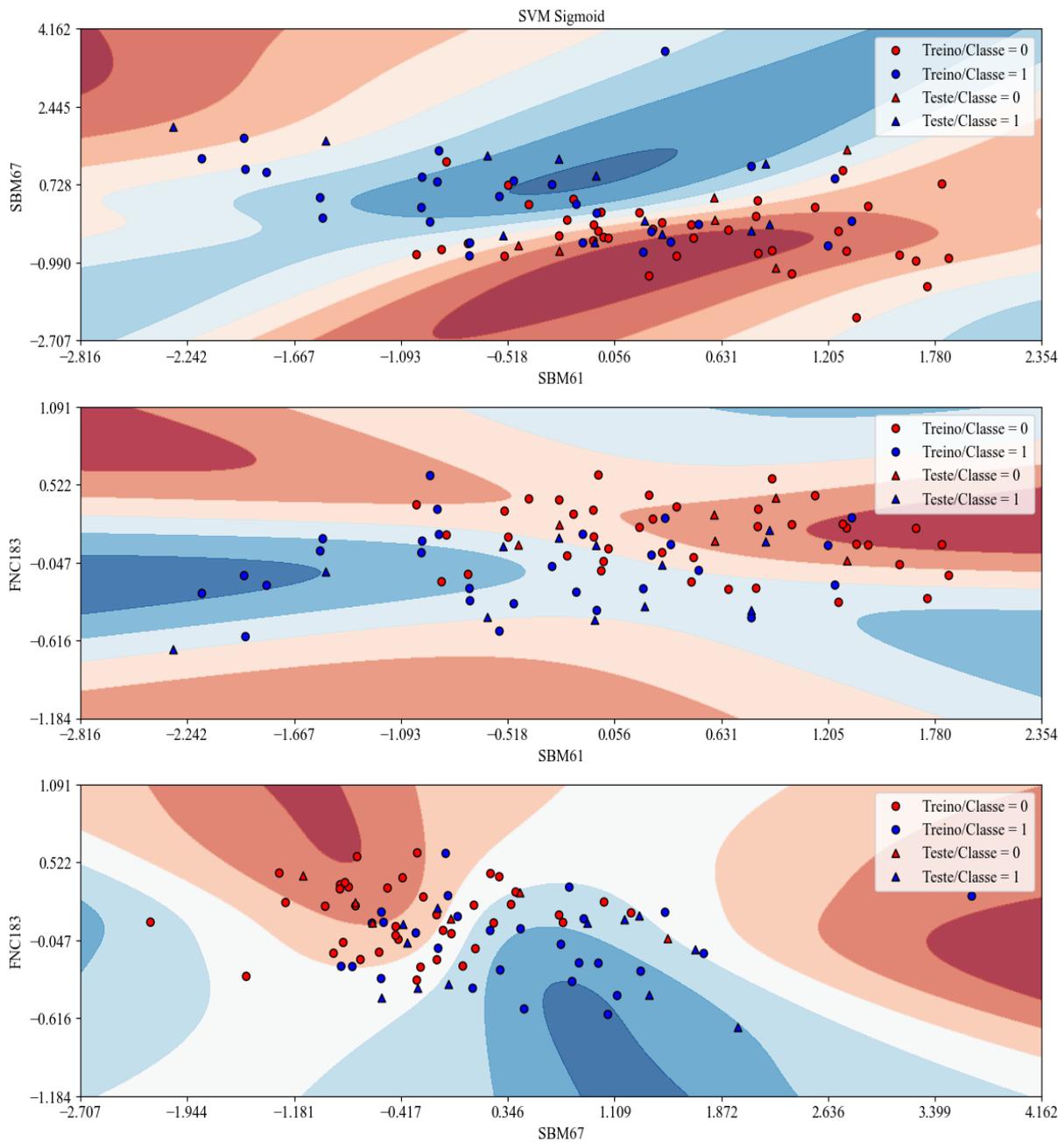


Figura 19 – Fronteira de Decisão do algoritmo SVM com *kernel* RBF.

Figura 20 – Fronteira de Decisão do algoritmo SVM com *kernel* sigmoid.

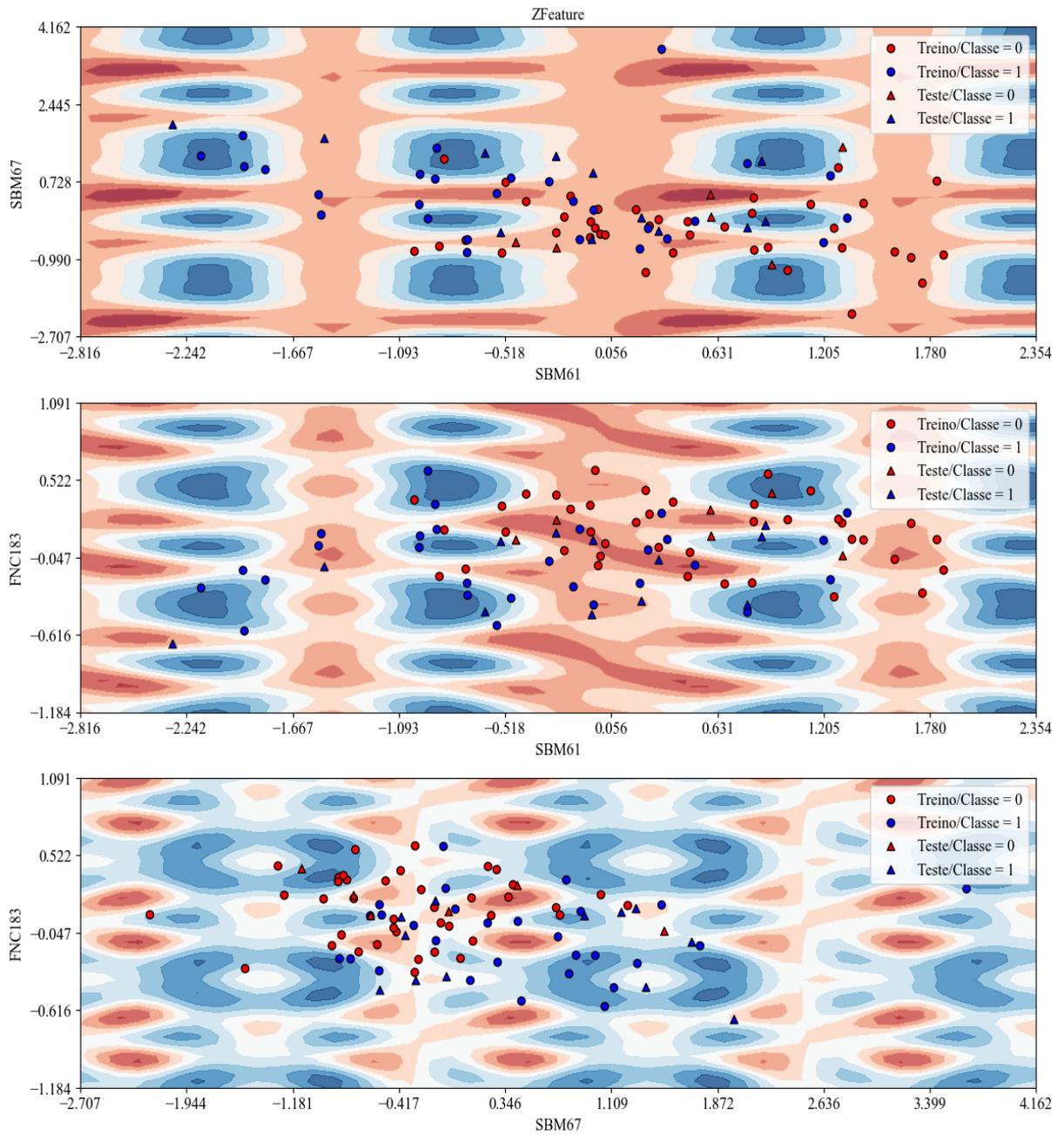


Figura 21 – Fronteira de Decisão do algoritmo SVM com *kernel Z Feature*.

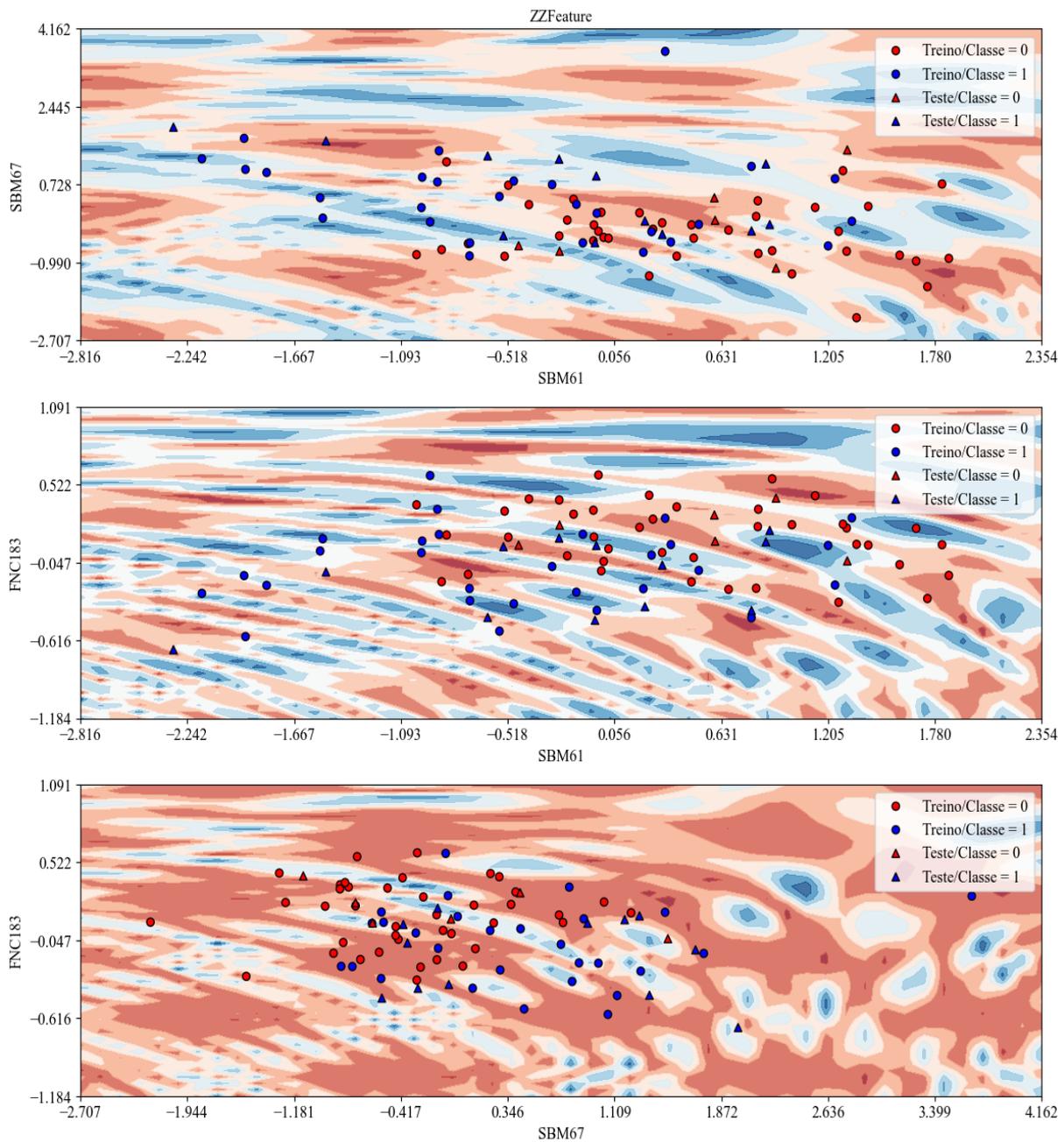


Figura 22 – Fronteira de Decisão do algoritmo SVM com *kernel ZZ Feature*.

## APÊNDICE B – Matrizes de confusão.

Tabela 9 – Matriz de confusão para o algoritmo K-NN com K igual a três.

3 - NN		Classe Verdadeira	
		Verdadeiro	Falso
Classe Predita	Verdadeiro	7	1
	Falso	5	5

Fonte: De autoria própria.

Tabela 10 – Matriz de confusão para o algoritmo Naive Bayes.

Naive Bayes		Classe Verdadeira	
		Verdadeiro	Falso
Classe Predita	Verdadeiro	8	1
	Falso	4	5

Fonte: De autoria própria.

Tabela 11 – Matriz de confusão para o algoritmo Árvore de Classificação.

Árvore de Classificação		Classe Verdadeira	
		Verdadeiro	Falso
Classe	Verdadeiro	6	0
Predita	Falso	6	6

Fonte: De autoria própria.

Tabela 12 – Matriz de confusão para o algoritmo Floresta Aleatória.

Floresta Aleatória		Classe Verdadeira	
		Verdadeiro	Falso
Classe	Verdadeiro	10	0
Predita	Falso	2	6

Fonte: De autoria própria.

Tabela 13 – Matriz de confusão para o algoritmo SVM com *Kernel* linear.

SVM Linear		Classe Verdadeira	
		Verdadeiro	Falso
Classe	Verdadeiro	7	0
Predita	Falso	5	6

Fonte: De autoria própria.

Tabela 14 – Matriz de confusão para o algoritmo SVM com *Kernel* ZFeature.

SVM ZFeature		Classe Verdadeira	
		Verdadeiro	Falso
Classe	Verdadeiro	3	0
Predita	Falso	9	6

Fonte: De autoria própria.

Tabela 15 – Matriz de confusão para o algoritmo SVM com *Kernel* ZZFeature.

SVM ZZFeature		Classe Verdadeira	
		Verdadeiro	Falso
Classe	Verdadeiro	2	0
Predita	Falso	10	6

Fonte: De autoria própria.

Tabela 16 – Matriz de confusão invertida para o algoritmo SVM com *Kernel* ZZFeature.

SVM ZZFeature		Classe Verdadeira	
		Verdadeiro	Falso
Classe	Verdadeiro	10	6
Predita	Falso	2	0

Fonte: De autoria própria.

# APÊNDICE C – Resultados do trabalho de Aksoy et al. (2024).

Tabela 17 – Resultado de Aksoy et al. (2024) utilizando cinco atributos para o canal Cz.

Método	Cz			
	Acurácia	Precisão	Especificidade	Sensibilidade
K-NN	92,31%	88%	81%	100%
Naive Bayes	96,15%	94%	90%	100%
Árvore de Classificação	96,15%	100%	100%	94%
Floresta Aleatória	96,15%	100%	100%	94%
SVM RBF	100%	100%	100%	100%
QSVM Z Feature	100%	100%	100%	100%
QSVM ZZ Feature	92,30%	88%	81%	100%

Fonte: Aksoy et al. (2024).

Tabela 18 – Resultado de Aksoy et al. (2024) utilizando cinco atributos para o canal T3.

Método	T3			
	Acurácia	Precisão	Especificidade	Sensibilidade
K-NN	100%	100%	100%	100%
Naive Bayes	92,30%	88%	81%	100%
Árvore de Classificação	96,15%	94%	90%	100%
Floresta Aleatória	96,15%	94%	90%	100%
SVM RBF	96,15%	94%	90%	100%
QSVM Z Feature	100%	100%	100%	100%
QSVM ZZ Feature	92,30%	88%	81%	100%

Fonte: Aksoy et al. (2024).