CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS

ENGENHARIA DE AUTOMAÇÃO INDUSTRIAL

TRABALHO DE CONCLUSÃO DE CURSO

APLICAÇÃO DE INTELIGÊNCIA ARTIFICIAL NA CLASSIFICAÇÃO DE IMAGENS DE ULTRASSONOGRAFIA MAMÁRIA PARA DETECÇÃO DE CÂNCER DE MAMA

MARIA PAULA PEREIRA DA TRINDADE

Junho de 2026

CAMPUS ARAXÁ

MARIA PAULA PEREIRA DA TRINDADE

APLICAÇÃO DE INTELIGÊNCIA ARTIFICIAL NA CLASSIFICAÇÃO DE IMAGENS DE ULTRASSONOGRAFIA MAMÁRIA PARA DETECÇÃO DE CÂNCER DE MAMA

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Automação Industrial, do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito para obtenção do grau de Bacharel em Engenharia de Automação Industrial.

Orientadora: Dra. Fabiana Alves Pereira

Araxá

2026

Trindade, Maria Paula Pereira da.
T833a Aplicação de inteligência artificial na classificação de imagens de ultrassonografia mamária para detecção de câncer de mama / Maria Paula Pereira da Trindade. – 2026.
79 f. : il.

Orientadora: Profa. Dra. Fabiana Alves Pereira.

Trabalho de conclusão de curso (Graduação) – Centro Federal de Educação Tecnológica de Minas Gerais. *Campus* Araxá. Engenharia de Automação Industrial, Departamento de Eletromecânica, Araxá, 2026.
Bibliografia.

1. Inteligência artificial - Saúde. 2. Computação evolutiva. 3. Exame - Imagem. 4. Redes neurais. I. Pereira, Fabiana Alves. II. Centro Federal de Educação Tecnológica de Minas Gerais. III. Título.

CDU 004.89:613

Ficha elaborada pela Biblioteca – *Campus* Araxá – CEFET-MG
Bibliotecária: Gleisa Mara Alves CRB-6/1713

MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
DEPARTAMENTO DE ELETROMECÂNICA - AX

FOLHA DE APROVAÇÃO DE TCC Nº 3 / 2026 - DELMAX (11.57.05)

Nº do Protocolo: 23062.034964/2026-70

Araxá-MG, 07 de julho de 2026.

MARIA PAULA PEREIRA DA TRINDADE

APLICAÇÃO DE INTELIGÊNCIA ARTIFICIAL NA CLASSIFICAÇÃO DE IMAGENS DE ULTRASSONOGRAFIAS MAMÁRIAS PARA DETECÇÃO DE CÂNCER DE MAMA

Trabalho de Conclusão de Curso apresentado ao Curso de **Engenharia de Automação Industrial** do Centro Federal de Educação Tecnológica de Minas Gerais, do Campus **Araxá**, como parte do requisito para obtenção do grau de Bacharel em **Engenharia de Automação Industrial**.

Aprovado em **19 de junho de 2026.**

Banca Examinadora:

Profa. Dra. Fabiana Alves Pereira - Orientadora
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Luis Paulo Fagundes
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Frederico Duarte Fagundes
Centro Federal de Educação Tecnológica de Minas Gerais

Araxá
2026

(Assinado digitalmente em 07/07/2026 14:11)
FABIANA ALVES PEREIRA
PROFESSOR ENS BASICO TECN TECNOLOGICO
DELMAX (11.57.05)
Matrícula: 1281564

(Assinado digitalmente em 13/07/2026 09:30)
FREDERICO DUARTE FAGUNDES
PROFESSOR ENS BASICO TECN TECNOLOGICO
DELMAX (11.57.05)
Matrícula: 1107165

(Assinado digitalmente em 08/07/2026 09:03)
LUIS PAULO FAGUNDES
PROFESSOR ENS BASICO TECN TECNOLOGICO
DELMAX (11.57.05)
Matrícula: 2993114

(Assinado digitalmente em 07/07/2026 09:45)
MARIA PAULA PEREIRA DA TRINDADE
DISCENTE
Matrícula: 20203009910

Visualize o documento original em **https://sig.cefetmg.br/public/documentos/index.jsp** informando seu número: **3**, ano: **2026**, tipo: **FOLHA DE APROVAÇÃO DE TCC**, data de emissão: **07/07/2026** e o código de verificação: **c78ef0fca7**

DEDICATÓRIA

Dedico este trabalho, com profunda admiração e saudade, ao professor Thiago Gomes Cardoso (in memoriam), que foi uma grande inspiração na minha trajetória acadêmica.

Muito mais do que transmitir conhecimento, ele acreditou no meu potencial, me incentivou nos momentos em que eu mais precisei e contribuiu de forma única para a minha formação. Seu exemplo, dedicação e paixão pela área sempre serão lembrados e levados comigo.

Este trabalho também é resultado de tudo o que aprendi com você.

AGRADECIMENTO

Agradeço, em primeiro lugar, à minha orientadora, Fabiana Alves Pereira, pela dedicação, paciência e orientação ao longo deste trabalho. Sua contribuição foi essencial para o desenvolvimento deste projeto, e seus ensinamentos tiveram grande impacto na minha formação acadêmica.

Agradeço, também, ao meu noivo, Gabriel Henrique Soares de Faria, por ter sido meu suporte durante toda a minha trajetória acadêmica. Seu apoio, incentivo e dedicação foram fundamentais para que eu seguisse em frente, sempre buscando o meu melhor. Obrigada por caminhar ao meu lado e contribuir para que este trabalho fosse realizado da melhor forma possível.

À minha mãe, Leide Mara Pereira, expresso minha profunda gratidão por todos os esforços realizados para que eu pudesse continuar meus estudos. Seu incentivo constante, sua fé em meu potencial e seu amor foram essenciais ao longo dessa jornada.

Aos meus avós, Baltazar Antonio Pereira e Maria Cantalina Pereira agradeço por todo o apoio, por sempre torcerem por mim e por me proporcionarem a oportunidade de me dedicar integralmente à minha formação. Esse gesto foi fundamental para que eu chegasse até aqui.

Aos meus grandes amigos Germano Gualberto Bittencourt e Pedro Henrique Rodrigues, minha sincera gratidão por compartilharem comigo essa caminhada. Obrigada por cada trabalho em grupo, pelas horas de estudo, pelas listas compartilhadas e, principalmente, pelo companheirismo e carinho. Sem vocês, essa trajetória não teria sido a mesma. Levo essa amizade para a vida.

Aos meus amigos Camila Eduarda Fernandes Pires e André Junji Morita, agradeço por estarem ao meu lado mesmo nos bastidores dessa jornada, sempre me incentivando e reforçando o quanto eu era capaz.

Por fim, agradeço a Deus pela saúde, pelas bênçãos recebidas e por me permitir concluir essa etapa. Agradeço também a mim mesma, por não ter desistido, mesmo diante das dificuldades. Ainda que este não tenha sido o curso que sempre sonhei, dediquei-me ao

máximo, com persistência e resiliência, aproveitando todas as oportunidades que me foram concedidas.

RESUMO

Este trabalho aborda a aplicação da inteligência artificial (IA) no diagnóstico do câncer de mama, destacando o desafio de aprimorar a precisão e a agilidade na detecção precoce da doença, especialmente diante das limitações dos métodos tradicionais e da necessidade de reduzir erros interpretativos em exames de imagem; tal contexto evidencia uma lacuna relacionada à integração efetiva de algoritmos avançados na prática clínica. O objetivo geral é analisar de que maneira técnicas de aprendizado de máquina, podem otimizar a identificação de padrões suspeitos em ultrassonografia mamária, contribuindo para diagnósticos mais assertivos; como objetivos específicos, busca-se comparar o desempenho de diferentes modelos, avaliar métricas de acurácia e propor uma configuração metodológica aplicável a cenários reais. O trabalho utiliza bancos de dados públicos com imagens de ultrassonografias mamárias, pré-processamento para padronização, treinamento supervisionado de modelos e validação cruzada para análise de desempenho. Os resultados incluem aumento da acurácia e sensibilidade dos modelos em relação a classificadores convencionais, além da identificação de combinações de parâmetros que favoreçam a robustez da predição. Como contribuição, o trabalho pretende fornecer evidências sobre a eficácia do uso de IA no apoio ao diagnóstico radiológico, ampliar a compreensão sobre o potencial dessas tecnologias no contexto da saúde e favorecer o desenvolvimento de soluções que possam auxiliar profissionais na tomada de decisão clínica, promovendo impacto social e tecnológico significativo.

Palavras-chave: Câncer de Mama, Diagnóstico, Inteligência Artificial, Aprendizado de Máquina, Redes Neurais Convolucionais.

ABSTRACT

This work addresses the application of artificial intelligence (AI) in breast cancer diagnosis, highlighting the challenge of improving accuracy and agility in the early detection of the disease, especially given the limitations of traditional methods and the need to reduce interpretative errors in imaging exams. This context reveals a gap related to the effective integration of advanced algorithms into clinical practice. The general objective is to analyze how machine learning techniques, can optimize the identification of suspicious patterns in breast ultrasound images, contributing to more assertive diagnoses. The specific objectives include comparing the performance of different models, evaluating accuracy metrics, and proposing a methodological configuration applicable to real-world scenarios. The study has a quantitative and experimental approach, using public databases containing breast ultrasound images, preprocessing for standardization, supervised model training, and cross-validation for performance analysis. Expected results include increased accuracy and sensitivity compared to conventional classifiers, as well as the identification of parameter combinations that enhance prediction robustness. As a contribution, the study aims to provide evidence on the effectiveness of AI in supporting radiological diagnosis, expand understanding of the potential of these technologies in the healthcare context, and foster the development of solutions that can assist professionals in clinical decision-making, promoting significant social and technological impact.

Keywords: Breast Cancer, Diagnosis, Artificial Intelligence, Machine Learning, Convolutional Neural Network.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE ABREVIATURAS

AUC – Area Under the Curve

BI-RADS – Breast Imaging Reporting and Data System

CAD – Computer-Aided Detection

CNN – Convolutional Neural Network

CPU – Central Processing Unit

DBT – Digital Breast Tomosynthesis

DBT/3D – Digital Breast Tomosynthesis 3D (Mamografia 3D)

DL – Deep Learning

GPU – Graphics Processing Unit

HD – Hard Disk Drive

IA – Inteligência Artificial

INCA – Instituto Nacional de Câncer

kNN – k-Nearest Neighbors

MCC – Matthews Correlation Coefficient

ML – Machine Learning

RAM – Random Access Memory

RBF – Radial Basis Function

ROC – Receiver Operating Characteristic

SSD – Solid State Drive

SUS – Sistema Único de Saúde

SVM – Support Vector Machine

SUMÁRIO

1. INTRODUÇÃO E JUSTIFICATIVA

1.1. Tema

Desde a Antiguidade, o câncer de mama já era reconhecido por diversas civilizações, como egípcios, gregos e romanos, que relatavam a presença de tumores nas mamas. No entanto, foi apenas ao longo do século XX, com o avanço das tecnologias de imagem e do conhecimento médico, que o diagnóstico e o tratamento da doença passaram por transformações significativas (American Cancer Society, 2025). No cenário global, estimou-se que em 2020 tenham ocorrido cerca de 2,3 milhões de novos casos de câncer de mama, tornando-o o tipo mais incidente entre mulheres no mundo (Porto; Teixeira; Silva, 2013). Projeções internacionais indicam que esse número continuará crescendo em todo o mundo nas próximas décadas, com maior intensidade em países de médio e baixo desenvolvimento, impulsionado pelo envelhecimento populacional, pelas mudanças no estilo de vida e pela ampliação dos métodos de rastreamento. (INCA, 2023).

1.2. Problema do trabalho

No Brasil, o câncer de mama representa um grande desafio para o sistema público de saúde. Segundo estimativas do Instituto Nacional de Câncer (INCA), em 2023, foram esperados 73.610 novos casos entre 2023 e 2025, com taxa ajustada de incidência de 41,89 casos a cada 100 mil mulheres (Santos et al., 2023). No que diz respeito à mortalidade por câncer, trata-se da principal causa de morte entre mulheres brasileiras, com taxa ajustada de 11,71 óbitos em 100 mil mulheres em 2021, já em 2023, foram registrados mais de 20 mil óbitos no Brasil (INCA, 2025). Mesmo com a recomendação de rastreamento para mulheres entre 50 e 74 anos, observa-se alta incidência em mulheres fora dessa faixa etária, como demonstram os 4,4 milhões de exames realizados pelo Sistema Único de Saúde (SUS) em 2024, muitos deles em grupos não prioritários (Gonçalves et al., 2016).

Neste contexto, a introdução da mamografia modificou profundamente o diagnóstico do câncer de mama ao permitir a detecção de tumores em estágios iniciais e, assim, a redução da mortalidade por meio do rastreamento. Com o passar dos anos, essa tecnologia evoluiu para a mamografia digital, a ultrassonografia mamária (3D) e os sistemas de detecção auxiliado por

computador (CAD), que trouxeram melhorias importantes. Ainda assim, atualmente, a interpretação visual realizada por um profissional da área da saúde permanece essencial para a identificação da patologia. No entanto este método de avaliação apresenta limitações significativas, como variabilidade entre os laudos profissionais, ocorrência de falsos positivos e falsos negativos e dificuldades específicas na análise de mamas densas, onde a sobreposição de tecidos pode comprometer a identificação de lesões (Hassan; Hamad; Mahar, 2022).

1.3. Hipótese do trabalho

Em contrapartida, o surgimento da Inteligência Artificial (IA) e a sua incorporação na etapa diagnóstica representa um novo marco nesse processo. Pesquisas recentes indicam que algoritmos baseados em aprendizagem profunda (do inglês: *Deep Learning*), têm demonstrado bom desempenho na análise de ultrassonografias mamárias e outras imagens, identificando padrões sutis muitas vezes imperceptíveis ao olhar humano. Tais estudos apontam avanços importantes no uso de *deep learning* para imagens mamográficas, com ganhos em sensibilidade, redução de falsos positivos e suporte ao trabalho dos radiologistas (Omega Boro; Nandi, 2023). Revisões sistemáticas também reforçam que técnicas baseadas em IA contribuem para diagnósticos mais rápidos, precisos e acessíveis, favorecendo a detecção precoce do câncer de mama. Assim, sua aplicação na classificação de imagens de ultrassonografia mamária surge como uma tecnologia promissora para reduzir a mortalidade, ampliar o acesso a diagnósticos de qualidade e fortalecer o sistema de saúde brasileiro (Wang, 2024).

Em um contexto de crescente incidência da doença, sobrecarga dos serviços públicos e demanda cada vez maior por eficiência nos processos diagnósticos, a IA apresenta potencial para agilizar atendimentos, reduzir erros nos diagnósticos e apoiar decisões clínicas. Pesquisas internacionais, como as desenvolvidas na Universidade de Stanford e sistemas como o *OncoSeek*, que é um teste de detecção precoce de múltiplos tipos de câncer baseado em um exame de sangue, como um tipo de detecção multicâncer precoce, reforçam esse potencial ao obter altos índices de acurácia na identificação de diferentes tipos de câncer (McKinney et al., 2020). Diante desse cenário mundial e nacional, marcado pela falta de especialistas, desigualdades regionais e limitações dos métodos tradicionais, torna-se necessário buscar soluções que padronizem análises e ampliem o acesso a interpretações confiáveis. Nesse sentido, o ponto central deste estudo consiste em compreender como a IA pode aprimorar a

classificação de imagens de ultrassonografia mamária, aumentando a precisão e a eficiência frente às limitações da interpretação humana.

1.4. Justificativa

A modernização do diagnóstico do câncer de mama tem se intensificado nas últimas décadas, com a adoção de tecnologias de imagem mais avançadas e o uso de ferramentas computacionais para apoiar a análise radiológica. A migração da mamografia convencional para modalidades mais modernas, como a *Digital Breast Tomosynthesis* (DBT, ou mamografia 3D), representa um avanço importante. Estudos mostram que a DBT eleva a detecção de câncer e reduz falsos positivos em comparação à mamografia digital 2D: em uma análise envolvendo milhares de exames, a DBT apresentou maior taxa de detecção, especialmente em mamas densas, e menor necessidade de convocação para exames complementares (RSNA, 2023).

Paralelamente, a incorporação de algoritmos de IA, especialmente via técnicas de *deep learning*, tem mostrado resultados promissores para superar limitações tradicionais da mamografia. Uma revisão recente indicou que a combinação de IA com ultrassonografia mamária aumenta a detecção precoce de câncer de mama, oferecendo mais precisão mesmo em contextos de alta demanda (Alves et al., 2025).

Em um estudo comparativo entre um algoritmo baseado em IA aplicado a imagens 3D, e um sistema tradicional CAD aplicado a imagens sintéticas 2D de mamografias, o algoritmo de IA teve desempenho significativamente superior: maior área sob a curva ROC (AUC), maior sensibilidade (94,3% vs. 72,6%) e maior especificidade (Bahl et al., 2025). Isso demonstra que a IA pode aumentar de forma considerável a acurácia diagnóstica, reduzindo falsos negativos e falsos positivos, o que é crucial para detecção precoce e confiável.

Além da precisão, a IA também contribui para otimizar o fluxo de trabalho e reduzir a sobrecarga dos radiologistas: em um estudo com leitura assistida por IA, o tempo médio de interpretação diminuiu, enquanto a acurácia melhorou e o grau de concordância entre leitores radiológicos aumentou (Park et al., 2024). Isso torna o processo de rastreamento mais eficiente, o que é especialmente relevante em sistemas de saúde públicos ou com grande demanda.

A utilização de algoritmos de Inteligência Artificial aplicados à análise de imagens mamográficas apresenta desempenho superior aos sistemas tradicionais de detecção assistida por computador (CAD), aumentando a acurácia diagnóstica e reduzindo o tempo de

interpretação radiológica. Assim, espera-se que a IA contribua simultaneamente para a melhoria da sensibilidade e especificidade na detecção do câncer de mama e para a otimização do fluxo de trabalho clínico, especialmente em ambientes com alta demanda.

Contudo, embora os avanços sejam promissores, há desafios para a implementação ampla dessas tecnologias. A adaptação de infraestrutura, aquisição de equipamentos modernos, treinamento adequado de profissionais e garantia de qualidade na captura e leitura de imagens são fundamentais para que a modernização seja efetiva. Além disso, a variabilidade anatômica entre paciente, como a densidade mamária, continua sendo um fator que pode comprometer a sensibilidade do exame, mesmo com tecnologias avançadas. Estratégias como o uso de IA multimodal, combinando diferentes modalidades de imagem conforme a densidade mamária, têm sido propostas para contornar tais limitações (Kakileti et al., 2025).

Em síntese, a modernização do diagnóstico de câncer de mama, por meio da adoção de tomossíntese, ultrassonografia mamária e inteligência artificial, representa uma evolução significativa no rastreamento e na detecção precoce da doença. Essas inovações oferecem maior sensibilidade e especificidade, reduzem erros interpretativos, aumentam a eficiência dos serviços e ampliam a possibilidade de diagnóstico em contextos diversos. Ao mesmo tempo, exigem investimento em tecnologia, capacitação e organização de sistemas de saúde, para que todo o potencial dessas ferramentas seja plenamente realizado (Park et al., 2024).

Assim, tendo em vista o exposto, nota-se que, apesar dos avanços tecnológicos apresentados, ainda existe uma carência na integração prática de algoritmos avançados nos ambientes médicos. Nesse contexto, a contribuição deste Trabalho de Conclusão de Curso consiste em expandir a análise originalmente proposta por (Dallagassa; Oliveira, 2024), que se concentrava na comparação entre redes neurais e o método *kNN*, incorporando também a avaliação de outros classificadores amplamente utilizados, como *Random Forest* e *SVM*. Além de ampliar o escopo comparativo, este estudo aprofunda a discussão sobre limitações dos modelos, riscos de generalização, sensibilidade a bases reduzidas e desafios de aplicação em cenários reais, aspectos fundamentais para compreender a viabilidade do uso dessas técnicas no diagnóstico por imagem.

Adicionalmente, o trabalho enfatiza o caráter educacional da implementação realizada no software *Orange Data Mining*, demonstrando como a ferramenta pode ser utilizada para fins de ensino, experimentação e validação inicial de modelos de IA aplicados à saúde. Ao

abordar tanto o potencial quanto as restrições desses algoritmos, busca-se contribuir para a incorporação mais consciente e fundamentada de soluções baseadas em inteligência artificial na rotina clínica, minimizando falhas interpretativas e promovendo maior segurança no diagnóstico precoce do câncer de mama.

1.5. OBJETIVOS

Nesta seção, são apresentados os objetivos que norteiam o desenvolvimento deste projeto, divididos entre o objetivo geral e os objetivos específicos, que juntos estruturam as etapas necessárias para a realização do estudo. Esses objetivos compõem a base que orienta a construção metodológica e a condução das atividades propostas, assegurando clareza sobre o que se pretende alcançar ao investigar a aplicação da automação e da IA no diagnóstico do câncer de mama.

1.5.1. Objetivo Geral

Classificar ultrassonografias mamárias em benignas e malignas por meio de técnicas de inteligência artificial.

1.5.2. Objetivos Específicos

Tendo em vista o objetivo geral descrito anteriormente e o desenvolvimento do trabalho, os seguintes objetivos específicos ficam em destaque:

- Compreender de que maneira a inteligência artificial e o processamento de dados podem otimizar a precisão diagnóstica do câncer de mama;
- Realizar a classificação das imagens de ultrassonografias mamárias em benignas ou malignas utilizando o *Orange Data Mining*;
- Examinar o desempenho da rede neural empregada no desenvolvimento da IA voltada ao diagnóstico por imagem;
- Analisar a performance de outras arquiteturas de IA empregadas no desenvolvimento de modelos para diagnóstico por imagem.

2. REFERENCIAL TEÓRICO

Nesta seção serão apresentados os conceitos necessários para o desenvolvimento deste projeto. Definir-se-á o câncer de mama e inteligência artificial para poder permitir a compreensão desse trabalho.

2.1. Câncer de Mama

O câncer de mama é caracterizado pela proliferação anormal de células presentes no tecido mamário, resultando na formação de tumores com potencial de invasão local e disseminação para outros órgãos. A literatura aponta que existem diversos subtipos dessa neoplasia, os quais diferem quanto ao comportamento biológico, agressividade e resposta terapêutica. Alguns tipos apresentam crescimento acelerado, enquanto outros evoluem de forma mais lenta e gradual. Estudos também reforçam que, apesar da heterogeneidade clínica, grande parte dos casos apresenta boa resposta aos tratamentos disponíveis, sobretudo quando o diagnóstico é realizado precocemente, condição que aumenta significativamente as chances de sucesso terapêutico (Rosa et al., 2021).

O diagnóstico do câncer de mama envolve um conjunto de procedimentos clínicos e de imagem que visam identificar alterações suspeitas no tecido mamário. O exame clínico das mamas constitui a etapa inicial, permitindo ao profissional avaliar características como nódulos, assimetrias ou alterações cutâneas. Em seguida, utilizam-se exames de imagem, como mamografia, ultrassonografia e ressonância magnética, que auxiliam na visualização interna das estruturas mamárias. Entre esses métodos, a mamografia se destaca como a principal ferramenta de rastreamento populacional, por ser capaz de detectar lesões em estágios iniciais, frequentemente antes do aparecimento de sintomas, o que aumenta significativamente as chances de tratamento bem-sucedido. Em casos nos quais as imagens identificam alterações suspeitas, a biópsia é indicada para confirmação histopatológica do diagnóstico (Oncoguia, 2025).

2.1.1. Análise da Mamografia

A análise da mamografia para diagnóstico segue um processo estruturado que utiliza o sistema *BI-RADS* (*Breast Imaging Reporting and Data System*), uma padronização desenvolvida para uniformizar a interpretação das imagens e garantir maior consistência nos

laudos radiológicos. Durante a avaliação, o radiologista examina características específicas da mama, como densidade mamária, presença de nódulos, calcificações, distorções arquiteturais e assimetrias. Cada achado é descrito de acordo com critérios estabelecidos pelo *BI-RADS*, que orienta a classificação da imagem em categorias que variam de 0 a 6, como mostra o Quadro 2.1. A categoria *BI-RADS* 0 indica que a avaliação está incompleta e que são necessários exames complementares para esclarecimento. As categorias 1 e 2 correspondem a achados benignos: a categoria 1 significa exame normal, enquanto a categoria 2 descreve achados claramente benignos, como cistos simples ou calcificações típicas. A categoria 3 engloba achados provavelmente benignos, com baixo risco de malignidade, recomendando-se acompanhamento para monitoramento. Já a categoria 4 reúne lesões suspeitas, subdivididas em 4A, 4B e 4C conforme o grau de suspeição, sendo indicada biópsia para elucidação. A categoria 5 representa alta suspeita de malignidade, com forte recomendação de intervenção diagnóstica e terapêutica. Por fim, a categoria 6 é utilizada apenas quando já existe diagnóstico histopatológico confirmando câncer. Essa estrutura sistematizada melhora a comunicação entre profissionais, orienta decisões clínicas e reduz a variabilidade na interpretação das imagens, contribuindo para maior precisão no diagnóstico (Johns Hopkins Medicine, 2025).

Quadro 2.1 – Classificação BI-RADS

CLASSIFICAÇÃO BI-RADS	
Categorias	Definição
0	Avaliação incompleta
1	Exame normal
2	Achado benigno
3	Provavelmente benigno
4	Achado suspeito
5	Altamente sugestivo de malignidade
6	Malignidade comprovada

Fonte: Adaptado de (Johns Hopkins Medicine, 2025)

2.1.2. Ultrassonografia mamária

A ultrassonografia mamária é um método amplamente utilizado na avaliação complementar das mamas, especialmente em mulheres com tecido mamário denso, nas quais a mamografia apresenta menor sensibilidade. Por utilizar ondas sonoras de alta frequência, trata-se de um exame seguro, não invasivo e isento de radiação ionizante, permitindo a visualização detalhada de estruturas internas e a diferenciação entre lesões sólidas e císticas.

Sua aplicação é particularmente relevante no rastreamento suplementar e na investigação de achados suspeitos (Pisano et al., 2008).

Além do diagnóstico por imagem, a ultrassonografia desempenha papel essencial na condução de procedimentos intervencionistas, como biópsias percutâneas e punções aspirativas. A possibilidade de guiar instrumentos em tempo real aumenta a precisão da coleta de material e reduz complicações, tornando o método preferido para lesões visíveis ao ultrassom. A técnica também permite avaliar características morfológicas, margens, ecogenicidade e vascularização das lesões, auxiliando na categorização segundo o sistema *BI-RADS* e contribuindo para decisões clínicas mais assertivas (Caseiras et al., 2010).

2.1.3. Problemas técnicos e de imagem

A qualidade do exame mamográfico pode ser comprometida por falhas no controle de qualidade, equipamentos desatualizados ou mal calibrados, posicionamento inadequado da mama e compressão incorreta, fatores que reduzem a nitidez da imagem e dificultam a detecção de lesões sutis. Em contextos com deficiência na manutenção de equipamentos e ausência de protocolos padronizados, a sensibilidade diagnóstica da ultrassonografia mamária pode cair significativamente (Morais et al., 2025).

Além disso, a densidade mamária elevada representa uma limitação intrínseca da mamografia, uma vez que, em mamas densas, o tecido fibroglandular pode “ocultar” tumores, fenômeno conhecido como efeito máscara. Esse efeito reduz a sensibilidade do exame e aumenta o risco de falsos negativos (Azevedo; Catarino; Ribeiro, 2021).

A ultrassonografia mamária, embora amplamente utilizada como método complementar no rastreamento e diagnóstico do câncer de mama, apresenta limitações importantes que precisam ser reconhecidas. Revisões recentes destacam que, em mamas extremamente densas, a mamografia permanece a modalidade primária de rastreamento, enquanto o ultrassom atua apenas como exame suplementar, em parte devido ao *masking effect* causado pela sobreposição do tecido fibroglandular. No caso do ultrassom manual, estudos apontam desvantagens como maior demanda de tempo do radiologista, dependência significativa do operador, maiores taxas de reconvocação e menor valor preditivo positivo, fatores que motivaram o desenvolvimento do ultrassom automatizado para melhorar documentação, orientação e reprodutibilidade (Allajbeu et al., 2021). A ultrassonografia apresenta elevada variabilidade entre observadores, uma vez que a detecção, a caracterização

e a interpretação de lesões podem diferir significativamente mesmo entre médicos experientes, o que reforça sua marcada dependência do operador (Berg et al., 2006). Além disso, a qualidade da imagem é influenciada por múltiplos ajustes manuais, como seleção do transdutor, foco, profundidade, ganho, compensação de ganho no tempo, harmônica, composição espacial e Doppler, que, quando inadequadamente configurados, podem comprometer a nitidez e a confiabilidade diagnóstica (Lee et al., 2024). Soma-se a isso a presença de ruído *speckle*, um artefato inerente às imagens ultrassonográficas que degrada a qualidade visual, reduz contraste e pode gerar estruturas artificiais que dificultam a interpretação (Ayana et al., 2022). Revisões tecnológicas também reforçam que o diagnóstico por ultrassonografia mamária permanece altamente dependente do operador e sujeito a grande variação entre observadores, o que limita sua padronização e consistência clínica (Guo et al., 2018). Em conjunto, essas limitações evidenciam que, apesar de seu valor clínico, a ultrassonografia mamária enfrenta desafios técnicos e operacionais que podem impactar a acurácia diagnóstica e reforçam a necessidade de ferramentas complementares, como sistemas de apoio baseados em IA.

2.1.4. Falsos positivos, falsos negativos e sobrediagnósticos

A ultrassonografia mamária, mesmo em programas de rastreamento, não é infalível. A sensibilidade costuma variar, e há uma proporção relevante de resultados falsos positivos, exames que indicam uma anormalidade suspeita, mas que após exames complementares ou biópsia se revelam benignos (IOM/NRC–CEDBC, 2005). Por outro lado, falsos negativos também ocorrem, por exemplo: pode haver tumores presentes que não são detectados pela ultrassonografia mamária, especialmente em casos de densidade mamária alta, posicionamento inadequado ou tumores pequenos/subtis (Comparetto; Borruto, 2025).

Além disso, o rastreamento por ultrassonografia mamária carrega o risco de sobrediagnóstico, ou seja, a detecção de lesões que nunca se tornariam clinicamente relevantes ao longo da vida da paciente. Esse fenômeno, embora estimado como relativamente raro, representa um dos principais danos potenciais associados ao rastreamento em massa (Nehmat; Nehmat, 2017).

2.1.5. Variabilidade humana e interpretação

Variabilidade significa que dois radiologistas diferentes podem interpretar a mesma ultrassonografia mamária de modos distintos, afetando a consistência dos diagnósticos e

aumentando a probabilidade de erro (Elmore, 2002). É importante destacar, que mesmo com imagens de boa qualidade, a interpretação radiológica depende fortemente da experiência e da formação do radiologista, o que introduz variabilidade nos laudos. Inclusive, já existem estudos publicados que mostram ampla variação entre profissionais quanto às taxas de falsos positivos e à sensibilidade da detecção de câncer (Elmore, 2002).

2.1.6. Limitações Estruturais e Operacionais

Em sistemas de saúde com alta demanda, a sobrecarga de exames para radiologistas pode levar a leituras rápidas ou menos cuidadosas, comprometendo a acurácia. A falta de protocolos de qualidade, controle rigoroso dos equipamentos e treinamento continuado também contribuem para a diminuição da confiabilidade dos resultados (Morais et al., 2025). Adicionalmente, em localidades com recursos limitados, pode haver restrição no acesso a métodos adicionais de imagem (ultrassonografia, ressonância magnética, tomossíntese), o que limita a detecção em casos de mamas densas ou imagens inconclusivas (Rosa Júnior et al., 2025).

Essas limitações combinadas comprometem o potencial da ultrassonografia mamária de garantir um diagnóstico precoce confiável e universal. A presença de falsos negativos pode atrasar o início do tratamento; falsos positivos e sobrediagnósticos podem causar ansiedade, exames desnecessários e sobrecarga no sistema de saúde; e desigualdades no acesso e na qualidade dos exames agravam, até mesmo, as disparidades regionais. A variabilidade entre profissionais e a dependência de fatores técnicos evidenciam a necessidade de aprimorar os processos de imagem, capacitação e controle de qualidade (Bleyer; Welch, 2012).

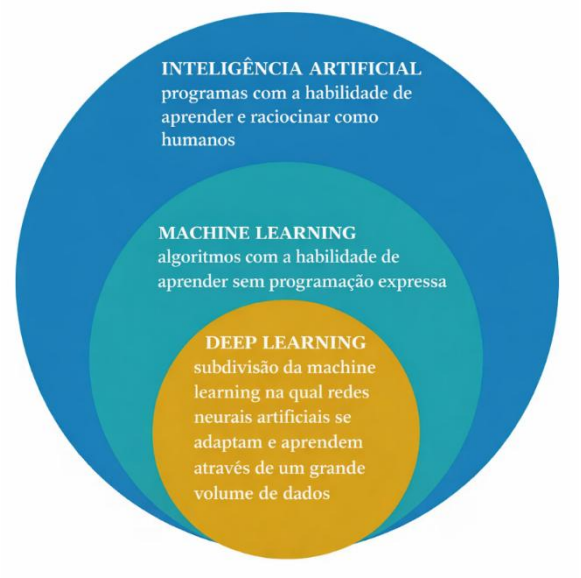
2.2. Inteligência Artificial (IA)

A IA é um campo da ciência da computação dedicado ao desenvolvimento de sistemas capazes de realizar tarefas que normalmente requerem inteligência humana, como reconhecer padrões, tomar decisões, resolver problemas, aprender com experiências e interpretar dados complexos. Seu funcionamento se baseia em algoritmos que permitem ao computador analisar informações, identificar relações e ajustar seu desempenho conforme novos dados são apresentados (Jordan; Mitchell, 2015).

Uma das abordagens mais importantes dentro da IA é o aprendizado de máquina (em inglês: *machine learning*), no qual modelos são treinados com grandes conjuntos de dados para

aprender a realizar previsões ou classificações. Técnicas mais avançadas, como o aprendizado profundo, utilizam redes neurais artificiais compostas por diversas camadas, capazes de transformar progressivamente os dados até gerar uma resposta, como reconhecer imagens, processar linguagem ou detectar anomalias. Assim, a inteligência artificial funciona combinando algoritmos matemáticos, estruturas computacionais e grandes volumes de dados para simular aspectos do raciocínio humano (Jordan; Mitchell, 2015). A Figura 2.1 ilustra como as áreas se organizam de forma abrangente, intermediária e especializada no campo dos algoritmos e modelos computacionais.

Figura 2.1 – Relação hierárquica entre Inteligência Artificial, *Machine Learning* e *Deep Learning*



Fonte: Adaptado de (Chagas, 2019)

2.2.1. *Machine Learning*

O machine learning é uma área da inteligência artificial dedicada ao desenvolvimento de algoritmos capazes de aprender a partir de dados e melhorar seu desempenho ao longo do tempo sem serem explicitamente programados para cada tarefa. Seu princípio central é permitir que sistemas computacionais identifiquem padrões, façam previsões ou tomem decisões com base em exemplos fornecidos durante o treinamento. Entre as abordagens mais conhecidas estão o aprendizado supervisionado, em que o modelo aprende a partir de pares de entrada e saída; o aprendizado não supervisionado, voltado à descoberta de estruturas ou agrupamentos nos dados; e o aprendizado por reforço, no qual um agente aprende por meio de interações com o ambiente e recompensas recebidas. Os algoritmos de *machine learning* utilizam técnicas como regressão, árvores de decisão, máquinas de vetores de suporte e redes neurais, variando

em complexidade conforme o tipo de problema. A eficácia desses métodos depende da qualidade dos dados, da seleção das características relevantes e dos processos de validação e ajuste de parâmetros. Dessa forma, o *machine learning* se consolidou como uma ferramenta essencial em aplicações modernas, como reconhecimento de padrões, análise preditiva, sistemas de recomendação e automação inteligente (Alpaydın, 2020).

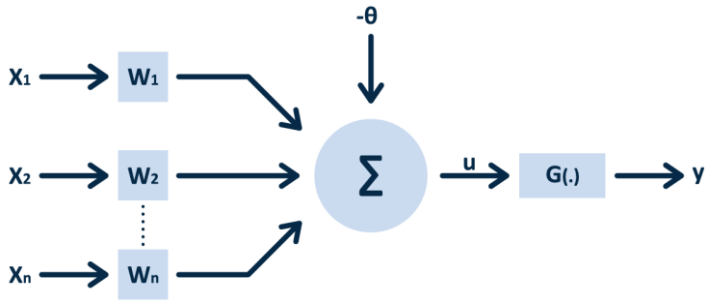
2.2.2. *Deep Learning*

O *deep learning* é um subcampo do aprendizado de máquina que utiliza redes neurais artificiais com múltiplas camadas para extrair automaticamente padrões complexos a partir de grandes volumes de dados. Essas redes profundas são capazes de transformar representações simples em níveis crescentes de abstração, permitindo que o modelo aprenda características relevantes sem necessidade de intervenção humana direta. O funcionamento do *deep learning* se baseia em operações matemáticas sucessivas realizadas por várias camadas ocultas, que ajustam seus parâmetros internos durante o treinamento por meio de algoritmos como o *backpropagation*. Essa estrutura possibilita que o sistema aprenda relações altamente não lineares, o que faz do *deep learning* uma técnica especialmente eficaz em tarefas como reconhecimento de imagens, processamento de linguagem natural, detecção de anomalias e análise de sinais. Esse avanço só se tornou possível devido ao aumento da capacidade computacional, à disponibilidade de grandes bases de dados e à evolução das arquiteturas de redes neurais, consolidando o *deep learning* como uma das tecnologias centrais da inteligência artificial moderna (Alpaydın, 2020).

2.2.3. Redes Neurais Artificiais

As redes neurais artificiais são modelos computacionais inspirados no funcionamento do cérebro humano, estruturados em unidades chamadas neurônios artificiais, que recebem sinais, processam informações e geram respostas. Essas redes operam por meio de camadas organizadas em arquitetura de entrada, camadas intermediárias (ou ocultas) e uma camada de saída, permitindo que transformem dados brutos em representações cada vez mais abstratas. A Figura 2.2 apresenta a estrutura de um neurônio artificial, responsável por coletar os sinais em suas entradas e produzir uma resposta em sua saída (Silva, 2016).

Figura 2.2 – Esquema funcional de um neurônio artificial utilizado em redes neurais.



Fonte: Adaptado de (Silva, 2016)

Os elementos presentes na Figura 2.2 são:

- $\{X1,X2,\cdots,Xn\}$ → sinais de entrada vindos do meio externo;
- $\{W1,W2,\cdots,Wn\}$ → pesos sinápticos usados para ponderar cada uma das variáveis de entrada;
- $\sum$ → combinador linear responsável por realizar o somatório dos sinais ponderados;
- θ → limiar de ativação define o nível adequado para que o resultado gerado pelo combinador linear possa ativar a saída do neurônio;
- u → potencial de ativação (resultado da diferença entre o combinador linear e o limiar de ativação);
- G → função de ativação, que limita a saída dentro de um intervalo de valores e
- y → sinal de saída produzido pelo neurônio.

Durante o processo de treinamento, as conexões entre os neurônios, representadas por pesos numéricos, são ajustadas para minimizar erros e melhorar o desempenho do modelo, tornando possível aprender padrões complexos e realizar tarefas como classificação, regressão, reconhecimento de imagens e análise de séries temporais. Esse processo de aprendizagem é fundamentado em algoritmos como o *backpropagation*, que calcula os gradientes de erro e ajusta os pesos de forma iterativa até que o modelo atinja uma performance satisfatória. Assim, redes neurais artificiais são ferramentas fundamentais dentro da inteligência artificial moderna,

especialmente no contexto do aprendizado profundo, que utiliza múltiplas camadas para lidar com grandes volumes de dados e padrões altamente complexos (Russell; Norvig, 2021).

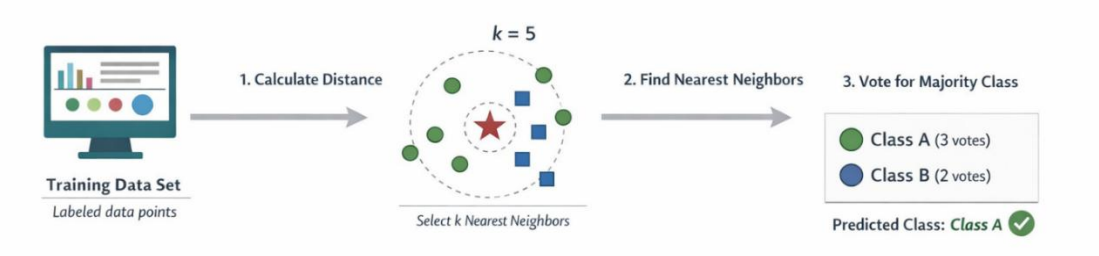
2.2.4. Redes Neurais Convolucionais

As redes neurais convolucionais, também conhecidas como *CNN* (*Convolutional neural network*) são modelos de aprendizado profundo desenvolvidos para processar dados visuais por meio de operações matemáticas capazes de extrair automaticamente padrões presentes em imagens, como bordas, texturas e formas, permitindo um reconhecimento progressivamente mais complexo conforme as camadas da rede são percorridas. Segundo (LeCun; Bengio; Hinton, 2015), sua arquitetura baseia-se em filtros convolucionais que detectam características locais, camadas de *pooling* que reduzem a dimensionalidade preservando informações essenciais e camadas totalmente conectadas responsáveis pela etapa final de classificação. Esse arranjo possibilita que as *CNNs* aprendam representações hierárquicas dos dados, o que as torna altamente eficientes em tarefas de visão computacional, incluindo diagnóstico por imagem, reconhecimento facial e detecção de padrões complexos, como descrito por (Goodfellow; Bengio; Courville, 2016). Na área da saúde, a literatura demonstra que esse tipo de rede apresenta desempenho superior em análises de imagens médicas devido à sua capacidade de identificar sinais sutis não percebidos facilmente pelo observador humano, conforme apontado por (Litjens et al., 2017), consolidando-se como uma das principais abordagens para classificação e detecção em radiologia e outras modalidades.

2.2.5. *K-Nearest Neighbors*

O algoritmo *K-Nearest Neighbors* (*kNN*) é um método de aprendizado supervisionado baseado na ideia de que instâncias próximas no espaço de atributos tendem a pertencer à mesma classe. Ele funciona identificando os *k* vizinhos mais próximos de uma nova amostra e classificando-a de acordo com a maioria das classes presentes entre esses vizinhos, conforme mostra a Figura 2.3. Por ser um método simples, intuitivo e não paramétrico, o *kNN* é amplamente utilizado em tarefas de classificação e regressão, especialmente quando a relação entre as variáveis não é linear (Cover; Hart, 1967).

Figura 2.3 – Algoritmo do método *kNN*



Fonte: Adaptado de (Haykin, 2001)

Uma característica fundamental do *kNN* é que ele é considerado um algoritmo *lazy learning*, ou seja, não realiza uma fase de treinamento explícita. Em vez disso, o modelo armazena todo o conjunto de dados e realiza os cálculos apenas no momento da predição, quando determina as distâncias entre a nova instância e os dados existentes. Essa abordagem reduz o custo computacional na etapa de treinamento, mas aumenta significativamente o custo na etapa de classificação, especialmente em bases de dados grandes. (Altman, 1992).

O desempenho do *kNN* depende diretamente da escolha do valor de K e da métrica de distância utilizada. Valores muito baixos de K tornam o modelo sensível a ruídos e outliers, enquanto valores muito altos podem levar ao sub-ajuste, diluindo padrões importantes. Além disso, a distância Euclidiana é a métrica mais utilizada, mas outras métricas, como *Manhattan* ou *Minkowski*, podem ser mais adequadas dependendo da natureza dos dados. (Haykin, 2001).

A distância Euclidiana é amplamente utilizada no kNN por representar a medida geométrica mais direta de proximidade entre duas instâncias, sendo calculada como a raiz quadrada da soma dos quadrados das diferenças entre cada atributo, conforme Equação (1)**Erro! Fonte de referência não encontrada.**; essa métrica funciona bem em espaços de baixa dimensionalidade e quando os atributos estão devidamente normalizados, pois evita que variáveis com escalas maiores dominem o cálculo, conforme discutido por (Haykin, 2001).

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

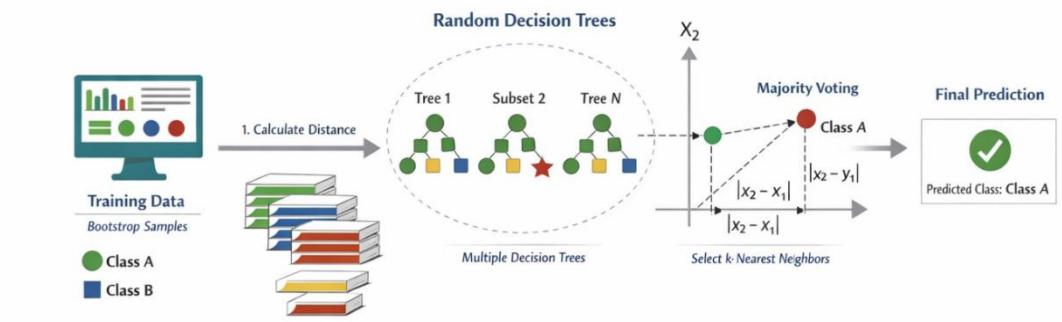
O *kNN* tem sido amplamente aplicado em áreas como diagnóstico médico, mineração de dados e sistemas de recomendação, devido à sua robustez e capacidade de adaptação a

diferentes tipos de dados. Em aplicações biomédicas, por exemplo, o *kNN* demonstra desempenho competitivo na detecção de padrões complexos, mesmo quando comparado a modelos mais sofisticados. (Nogueira; Penatti; Dos Santos, 2017).

2.2.6. *Random Forest*

O *Random Forest* é um algoritmo de aprendizado supervisionado baseado em conjuntos, que combina múltiplas árvores de decisão para melhorar a precisão e reduzir o risco de sobre-ajuste. Ele foi introduzido por (Breiman, 2001) como uma extensão dos métodos de *bagging*, utilizando amostras aleatórias dos dados e subconjuntos aleatórios de atributos em cada divisão das árvores, onde essa aleatoriedade garante diversidade entre as árvores e, consequentemente, maior robustez do modelo. A Figura 2.4 representa o algoritmo aplicado no método *Random Forest*.

Figura 2.4 – Algoritmo do método *Random Forest*



Fonte: Adaptado de (Breiman, 2001)

Cada árvore individual é construída a partir de uma amostra *bootstrap* do conjunto de treinamento, conforme mostrado na Figura 2.4, e as previsões finais são obtidas por votação majoritária (para classificação) ou média (para regressão). Essa abordagem reduz a variância em relação a uma única árvore de decisão, mantendo o viés relativamente baixo (Pohlert et al., 2007).

Uma das principais vantagens do *Random Forest* é sua capacidade de estimar a importância das variáveis. O algoritmo calcula o impacto de cada atributo na redução da impureza das árvores, permitindo identificar quais variáveis são mais relevantes para o processo de decisão. Essa funcionalidade é especialmente útil em contextos de análise exploratória e seleção de características, como demonstrado por (Díaz-Uriarte; Alvarez De

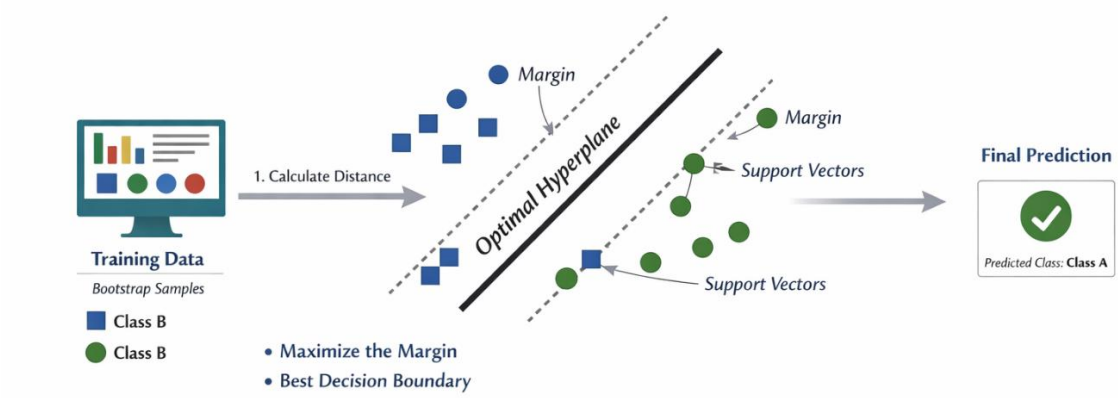
Andrés, 2006), que aplicaram o método em estudos genômicos para identificar genes associados a doenças.

Além disso, o método apresenta alta resistência ao *overfitting*, graças à combinação de múltiplas árvores independentes e ao uso de amostragem aleatória. Mesmo quando o número de árvores é elevado, o modelo tende a manter estabilidade e generalização, desde que os parâmetros, como o número de árvores e o número de atributos considerados em cada divisão, sejam ajustados adequadamente. Essa característica foi explorada por (Liaw; Wiener, 2002), que demonstraram a eficiência do algoritmo em diferentes cenários de classificação.

2.2.7. *Support Vector Machine*

O *Support Vector Machine* (*SVM*) é um algoritmo de aprendizado supervisionado amplamente utilizado para tarefas de classificação e regressão, baseado na ideia de encontrar um hiperplano ótimo que separa as classes de forma maximamente distante. O objetivo do *SVM* é maximizar a margem entre os pontos de dados das diferentes classes, garantindo maior capacidade de generalização. Essa abordagem foi formalmente introduzida por (Cortes; Vapnik, 1995), que demonstraram que o uso de vetores de suporte é suficiente para definir a fronteira de decisão, tornando o modelo eficiente e matematicamente elegante. A Figura 2.5 representa o algoritmo aplicado no método *SVM*.

Figura 2.5 – Algoritmo do método *SVM*



Fonte: Adaptado de (Schölkopf; Smola, 2001)

Uma das principais vantagens do *SVM* é sua capacidade de lidar com dados não linearmente separáveis por meio do uso de funções *kernel*, que transformam o espaço original dos dados em um espaço de maior dimensão, onde a separação linear se torna possível. Entre

os *kernels* mais utilizados estão o linear, o polinomial e o *radial basis function* (*RBF*). Essa técnica permite que o *SVM* capture relações complexas entre as variáveis, mantendo um bom equilíbrio entre viés e variância. Estudos como o de Schölkopf e Smola (2001) discutem o papel dos *kernels* na expansão da aplicabilidade do *SVM* em problemas de alta complexidade.

As funções *kernel* são mecanismos matemáticos usados no *SVM* para transformar os dados de entrada em um espaço de maior dimensão, permitindo que o algoritmo encontre um hiperplano que separe as classes de forma linear mesmo quando isso não é possível no espaço original. Em outras palavras, o *kernel* calcula o produto interno entre vetores em um espaço transformado sem precisar realizar explicitamente essa transformação (Schölkopf; Smola, 2001).

O *Kernel* Linear é usado quando os dados são linearmente separáveis e define o hiperplano diretamente no espaço original dos atributos. A Equação (2) mostra a representação matemática do *kernel* linear.

$$K(x, x') = x^T x' \tag{2}$$

Onde,

- x é um vetor de características (um exemplo do conjunto de dados).
- x' é outro vetor de características, geralmente um segundo exemplo usado para medir similaridade.
- Ambos pertencem ao espaço original dos dados:

$$x, x' \in \mathbb{R}^n$$

Onde,

- n é o número de atributos.

O *Kernel* Polinomial introduz não linearidades ao elevar o produto interno a uma potência, permitindo capturar relações mais complexas entre as variáveis. A Equação (3) mostra a representação matemática do *kernel* polinomial.

$$K(x, x') = (x^\top x' + c)^d \quad (3)$$

Onde,

- $c \geq 0$ é um termo de ajuste,
- d é o grau do polinômio.

O *Kernel RBF* mede a similaridade entre pontos com base na distância Euclidiana, criando fronteiras de decisão altamente flexíveis. A Equação (4) mostra a representação matemática do *kernel RBF*.

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right) \quad (4)$$

Onde,

- $\gamma > 0$ controla o alcance da influência dos pontos.

O SVM se destaca pela sua robustez em conjuntos de dados com alta dimensionalidade, como os encontrados em bioinformática e reconhecimento de padrões. Por utilizar apenas os vetores de suporte para definir o hiperplano, o algoritmo é menos sensível a ruídos e redundâncias nos dados. Essa característica foi explorada por (Guyon et al., 2002) em estudos sobre seleção de características, mostrando que o *SVM* pode identificar variáveis relevantes e melhorar a interpretabilidade dos modelos em contextos científicos e médicos.

2.3. *Software Orange Data Mining*

O *Orange Data Mining* é uma plataforma de código aberto desenvolvida pela Universidade de Ljubljana (Eslovênia), voltada para análise exploratória de dados, mineração de dados e aprendizado de máquina. Seu principal diferencial é a interface visual intuitiva, que permite criar fluxos de trabalho por meio de blocos conectáveis (*widgets*), cada um responsável por uma tarefa específica, como importar dados, processá-los, aplicar algoritmos ou visualizar resultados (Viegas, 2025).

O software funciona por meio de uma interface visual baseada em fluxos de trabalho, permitindo que o usuário construa análises conectando módulos chamados *widgets*. Cada widget executa uma função específica, como importar dados, realizar pré-processamento, treinar modelos de aprendizado de máquina ou visualizar resultados. O processo ocorre de forma sequencial: os dados são carregados, transformados conforme necessário e encaminhados para algoritmos de classificação, regressão ou agrupamento. Em seguida, os resultados podem ser avaliados por meio de métricas como acurácia, sensibilidade, matriz de confusão e curva *ROC*. Essa abordagem modular facilita a experimentação, pois o usuário pode alterar parâmetros, substituir algoritmos ou adicionar novas etapas apenas rearranjando os widgets, sem necessidade de programação. Assim, o Orange se destaca como uma ferramenta acessível para explorar dados, testar modelos e compreender o funcionamento de técnicas de machine learning de maneira intuitiva.

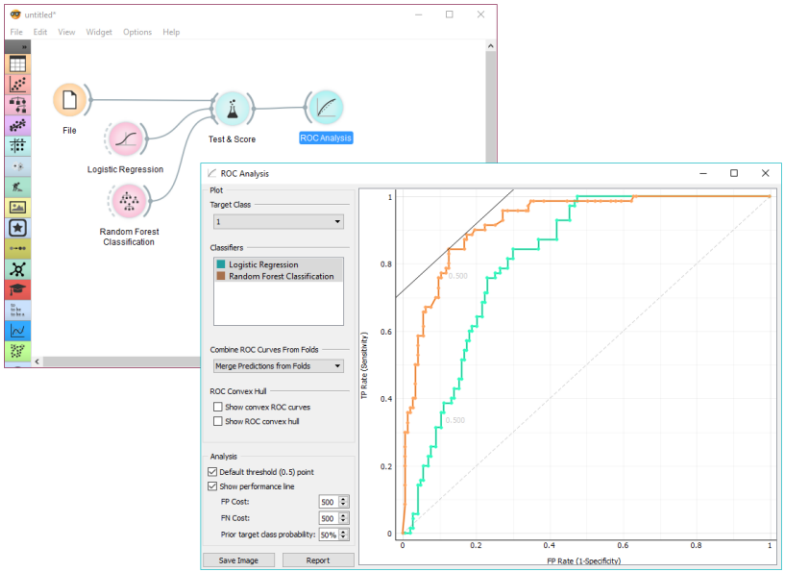
Acurácia e sensibilidade são métricas fundamentais para avaliar o desempenho de modelos de classificação em saúde, especialmente em tarefas como a detecção de câncer de mama. A acurácia representa a proporção total de previsões corretas feitas pelo modelo, considerando simultaneamente acertos de ambas as classes, benigno e maligno, em relação ao número total de amostras avaliadas. Já a sensibilidade (ou taxa de verdadeiros positivos) mede especificamente a capacidade do modelo de identificar corretamente os casos positivos, ou seja, quantas lesões malignas foram detectadas entre todas as realmente malignas. Enquanto a acurácia fornece uma visão global do desempenho, a sensibilidade é crucial em contextos clínicos, pois reflete o quanto o modelo consegue evitar falsos negativos, que são especialmente perigosos por representarem casos de câncer não detectados (Powers, 2020).

A matriz de confusão é uma ferramenta essencial para avaliar o desempenho de modelos de classificação, pois permite visualizar de forma detalhada como o algoritmo se comporta ao distinguir entre as classes previstas e as reais. Ela é composta por quatro elementos principais: verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. Essa estrutura possibilita identificar não apenas quantas amostras foram corretamente classificadas, mas também os tipos de erros cometidos pelo modelo, como confundir um nódulo maligno com benigno ou vice-versa (Grandini; Bagli; Visani, 2020).

A curva *ROC* (*Receiver Operating Characteristic*) é uma representação gráfica utilizada para avaliar o desempenho de modelos de classificação binária, relacionando a taxa de verdadeiros positivos (*True Positive Rate*) com a taxa de falsos positivos (*False Positive*

Rate) em diferentes limiares de decisão. Essa curva permite visualizar a capacidade discriminativa do classificador, indicando o quanto ele consegue separar corretamente as classes. O desempenho global costuma ser medido pela AUC (*Area Under the Curve*), cuja variação vai de 0 a 1, sendo valores próximos de 1 indicativos de melhor desempenho. A curva *ROC* é amplamente aplicada em áreas como medicina, detecção de sinais e aprendizado de máquina, pois fornece uma avaliação robusta e independente da distribuição das classes (Fawcett, 2006). É possível ver um exemplo da curva *ROC* na Figura 2.6.

Figura 2.6 – Exemplo de curva *ROC* no *software Orange Data Mining*



Fonte: (University of Ljubljana, 2025)

Entre seus principais recursos, o *Orange* oferece uma ampla coleção de *widgets* para análise estatística, visualização gráfica e modelagem preditiva, permitindo desde tarefas simples, como inspeção de dados, até experimentos complexos envolvendo aprendizado supervisionado e não supervisionado. A plataforma é gratuita, de código aberto e compatível com *Windows*, *macOS* e *Linux*, o que facilita sua adoção em ambientes acadêmicos e profissionais. Outro diferencial é a disponibilidade de extensões temáticas, como módulos para bioinformática, mineração de texto e análise de imagens, que ampliam suas aplicações. Além disso, o software integra bibliotecas de machine learning, possibilitando a criação de modelos avançados sem exigir conhecimento em linguagens de programação. Por ser altamente visual e interativo, o *Orange* é amplamente utilizado no ensino de ciência de dados, na prototipagem rápida de soluções e em pesquisas que demandam experimentação ágil e comparações entre diferentes algoritmos.

3. METODOLOGIA

O uso da inteligência artificial no diagnóstico por imagem tem se tornado um campo amplamente estudado e promissor dentro da medicina moderna, resultando em diversas aplicações voltadas ao aprimoramento do diagnóstico e tratamento do câncer de mama. No entanto, ainda existem limitações no acesso e na implementação dessas tecnologias em larga escala, especialmente nos sistemas públicos de saúde. A integração de ferramentas baseadas em IA apresenta condições para otimizar os processos clínicos, permitindo diagnósticos mais precisos, tratamentos personalizados e maior eficiência no atendimento oncológico.

Neste capítulo, são apresentadas as etapas de desenvolvimento do projeto voltada à aplicação da IA no diagnóstico do câncer de mama. Essas etapas envolvem:

- Análise da eficácia das tecnologias na detecção precoce e no acompanhamento de pacientes com câncer de mama;
- Avaliação do impacto da IA na agilidade e precisão dos diagnósticos;
- Utilização da plataforma *Orange Data Mining* para implementação do primeiro modelo de classificação binária entre imagens benignas e malignas;
- Análise com base em métricas quantitativas dos desempenhos obtidos, a fim de reconhecer os desafios de implementação.

3.1. Análise da eficácia das tecnologias na detecção precoce e no acompanhamento de pacientes com câncer de mama

Após o levantamento das principais tecnologias foi realizada a análise da eficácia das tecnologias empregadas na detecção precoce e no acompanhamento de pacientes com câncer de mama. Para isso, foram examinados os resultados apresentados nas pesquisas, considerando métricas como acurácia, sensibilidade, especificidade e capacidade de identificar lesões em estágios iniciais. Essa avaliação permitiu compreender em que medida as ferramentas tecnológicas, especialmente as baseadas em inteligência artificial, têm contribuído para aumentar a precisão diagnóstica e reduzir a ocorrência de falsos positivos e falsos negativos.

Também é importante destacar que foi investigado como essas tecnologias têm sido aplicadas ao monitoramento da evolução clínica dos pacientes, incluindo o acompanhamento de respostas terapêuticas e a identificação de possíveis recorrências da doença. A análise dos estudos possibilitou observar se os modelos de IA apresentam vantagens em relação aos métodos tradicionais, tanto em termos de agilidade quanto de consistência nas interpretações.

3.2. Avaliação do impacto da IA na agilidade e precisão dos diagnósticos

Esta etapa do estudo consistiu na avaliação do impacto da inteligência artificial na precisão e na agilidade dos diagnósticos relacionados ao câncer de mama. Para isso, foram analisados os resultados reportados no referencial teórico quanto ao tempo necessário para processamento das imagens, emissão de laudos assistidos por algoritmos e detecção automatizada de achados suspeitos. Essa análise permitiu identificar se as ferramentas de IA contribuem para reduzir o tempo entre o exame e a interpretação clínica, favorecendo um fluxo mais eficiente na rotina diagnóstica.

Paralelamente, foi examinada a precisão alcançada pelos sistemas de IA em comparação aos métodos tradicionais, observando métricas como acurácia, sensibilidade, especificidade e taxa de detecção de lesões em estágios iniciais. A partir dessa avaliação, foi possível compreender em que medida a IA melhora a assertividade dos diagnósticos, diminui variações entre avaliadores e auxilia na identificação de padrões sutis que podem passar despercebidos em análises exclusivamente humanas.

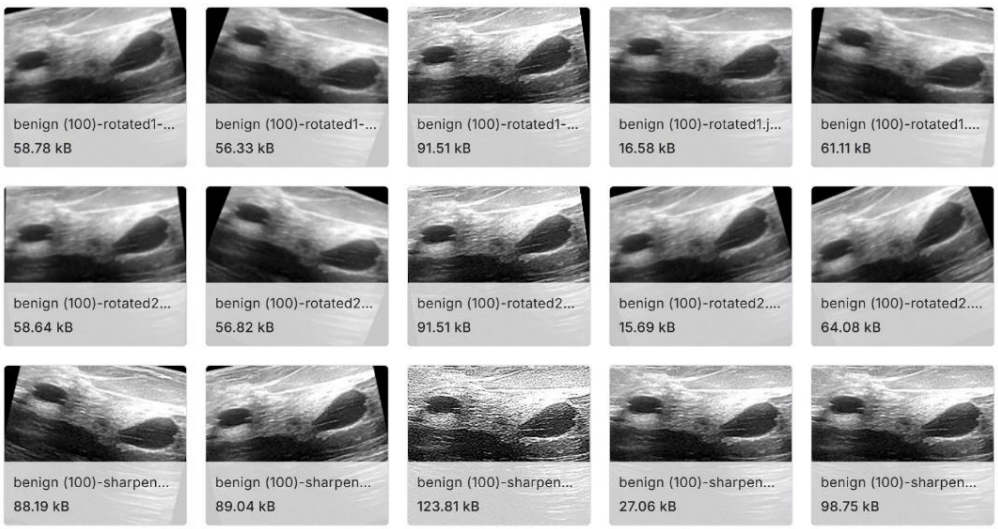
3.3. Utilização da plataforma *Orange Data Mining* para implementação do primeiro modelo de classificação binaria entre imagens benignas e malignas

O estudo envolveu a utilização da plataforma *Orange Data Mining* para a implementação do primeiro modelo de classificação binária entre imagens benignas e malignas, baseada e adaptada do estudo de (Dallagassa; Oliveira, 2024). Inicialmente, foi realizada a preparação do conjunto de dados, incluindo a importação das imagens, a organização das classes e a verificação da consistência dos arquivos. Em seguida, os dados foram submetidos a procedimentos de pré-processamento adequados, como redimensionamento, normalização e extração automática de características, de acordo com os recursos disponíveis na plataforma.

Com o *dataset* devidamente estruturado, foi montado um fluxo de trabalho no *Orange* que permitiu testar diferentes algoritmos de aprendizagem de máquina aplicados ao problema de classificação. Durante essa etapa, foram ajustados hiper parâmetros, que podem ser considerados parâmetros de configuração, e aplicadas estratégias de validação cruzada a fim de assegurar maior confiabilidade nos resultados. O desempenho do modelo foi analisado com base em métricas como acurácia, sensibilidade, especificidade e matriz de confusão, possibilitando comparar a capacidade de cada algoritmo em distinguir imagens benignas de malignas.

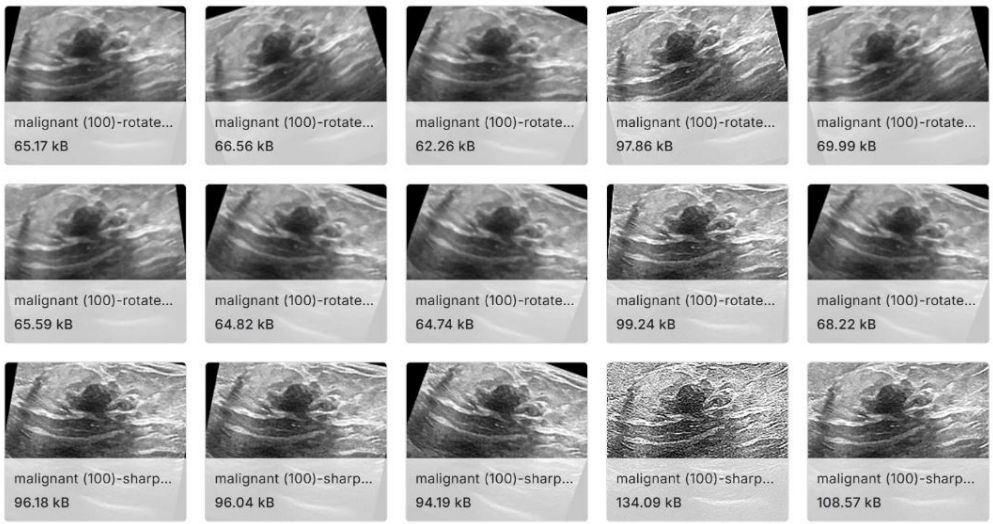
Após a realização de revisões sobre o estado da arte da aplicação de IA em exames de imagem para câncer de mama, as quais permitiram a construção do embasamento teórico apresentado neste projeto, foram realizados testes de classificação binária utilizando o software Orange Data Mining, um software de código aberto utilizado em processos de análise e modelagem de dados, em que sua principal característica é a facilidade de uso. Para tal, foi selecionado um banco de imagens, que possui 9.016 imagens, disponibilizado gratuitamente na plataforma *Kaggle*, que é uma plataforma para testes de IA. Esse *dataset* é composto exclusivamente por ultrassonografias das mamas, referentes a exames nos quais foram identificados nódulos. Cada imagem incluída no conjunto já apresentava a indicação diagnóstica correspondente, informando se o nódulo era classificado como benigno (Figura 3.1) ou maligno (Figura 3.2).

Figura 3.1 –Imagens de ultrassonografia de nódulos benignos utilizadas no treinamento do modelo.



Fonte: (“Ultrasound Breast Images for Breast Cancer”, 2022)

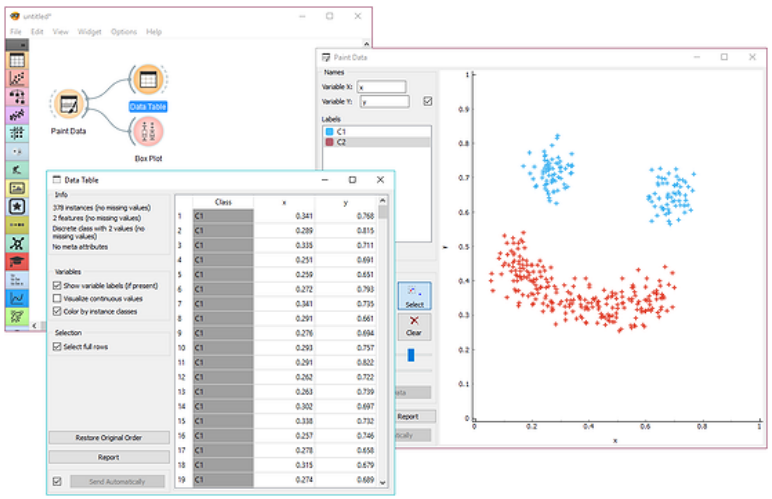
Figura 3.2 – Imagens de ultrassonografia de nódulos malignos utilizadas no treinamento do modelo.



Fonte: (“Ultrasound Breast Images for Breast Cancer”, 2022)

O *Orange* permitiu realizar tarefas de processamento de linguagem natural, mineração de texto, extração de regras de associação e análises baseadas em redes. Além disso, disponibiliza diversos recursos estatísticos e visuais, como gráficos, árvores de decisão, mapas de calor e projeções lineares, podendo ainda ser ampliado por meio da instalação de bibliotecas adicionais que ampliam suas funcionalidades (Dallagassa; Oliveira, 2024). A Figura 3.3 representa algumas dessas funcionalidades do *Orange*.

Figura 3.3 – Interface e funcionalidades do *Orange Data Mining* para análise de imagens.



Fonte: (University of Ljubljana, 2025)

Para a utilização do software *Orange Data Mining*, é indispensável contar com um computador que ofereça desempenho adequado para lidar com o grande volume de imagens processadas ao longo do estudo. Como o projeto envolve a extração de *embeddings*, que é uma

técnica utilizada em inteligência artificial para transformar características em imagens ou outros tipos de dados em vetores numéricos que representam significado, o treinamento de diferentes modelos de inteligência artificial e a validação cruzada dos resultados, o *hardware* precisa ser capaz de executar operações matemáticas intensivas sem comprometer a estabilidade do sistema.

Neste trabalho, utilizou-se uma máquina equipada com um processador Intel i5-2400, placa-mãe Afox Ib75-Ma5 e 8 GB de memória RAM (*Random Access Memory*) e uma placa de vídeo GTX 1650, responsável por acelerar cálculos matriciais e operações gráficas essenciais para o processamento das imagens. Além disso, o sistema contou com um *SSD* (*Solid State Drive*) de 1 TB para maior velocidade de leitura e escrita, um *HD* (*Hard Disk Drive*) de 1 TB para armazenamento complementar, uma fonte de 500 W para garantir estabilidade energética e um gabinete adequado para acomodar todos os componentes. Os componentes citados e seus modelos ou especificações estão sintetizadas na Tabela 3.1.

Tabela 3.1 – Componentes e preço do computador utilizado

Componente	Modelo/Especificação	Custo
Processador (CPU)	*Intel i5-2400*	R$ 100,00
Placa mãe	*Afox Ib75-Ma5 (soquete 1155, DDR3)*	R$ 388,00
Memória RAM	*8 GB DDR3*	R$ 63,00
Placa de vídeo (GPU)	*GTX 1650 PNY*	R$ 1.335,00
Armazenamento	*SSD* 1 TB	R$ 327,77
Armazenamento	*HD* 1TB	R$ 277,77
Fonte de alimentação	*500 W*	R$ 182,00
Gabinete	-	R$ 149,00

Fonte: Elaborado pela autora

Além disso, é importante ressaltar que a execução completa das etapas de pré-processamento, classificação e análise de desempenho exigem que o computador suporte operações repetitivas e de alto custo computacional, especialmente quando se trabalha com modelos mais complexos, como redes neurais profundas. A capacidade gráfica também pode influenciar o desempenho, uma vez que alguns módulos do *Orange* fazem uso de aceleração por *GPU* (*Graphics Processing Unit*) para otimizar cálculos matriciais, função desempenhada pela GTX 1650 neste estudo. Dessa forma, a escolha de um bom equipamento não apenas garante maior fluidez durante o desenvolvimento do projeto, mas também contribui para a obtenção de resultados mais consistentes, evitando travamentos, lentidão e limitações que poderiam comprometer a qualidade das análises realizadas.

Ao analisar os custos dos componentes utilizados, é possível organizar os valores em categorias funcionais, facilitando a compreensão do investimento total. No grupo de processamento, que reúne tanto o processador quanto a placa de vídeo por serem responsáveis pela maior parte do desempenho computacional, encontram-se o Intel i5-2400 e a GTX 1650 PNY, totalizando R$ 1.435,00. A categoria de placa-mãe e memória inclui o modelo Afox Ib75-Ma5 e a especificação 8 GB de *RAM* DDR3, com custo de R$ 451,00. No grupo de armazenamento, somam-se o *SSD* (R$ 327,77) e o *HD* (R$ 277,77), resultando em R$ 605,54. Por fim, os componentes complementares, que englobam componentes essenciais ao funcionamento físico do sistema, incluem a fonte de 500 W (R$ 182,00) e o gabinete (R$ 149,00), totalizando R$ 331,00.

Essa organização evidencia a distribuição dos custos e destaca que a maior parte do investimento está concentrada no processamento, etapa crítica para o desempenho das tarefas de inteligência artificial. Com isso, temos que o custo total do *hardware* utilizado no projeto é de R$ 2.822,54.

3.4. Análise com base em métricas quantitativas dos desempenhos obtidos, a fim de reconhecer os desafios de implementação

O estudo analisou de forma quantitativa os desempenhos obtidos pelos modelos de IA analisados, a fim de reconhecer os principais desafios relacionados à sua implementação. Inicialmente, foram consolidadas as métricas nas fases anteriores, permitindo uma avaliação comparativa entre os diferentes métodos testados. Essa análise possibilitou identificar tendências de desempenho, pontos fortes de cada abordagem e eventuais limitações na capacidade de generalização dos modelos.

Em continuidade, foram examinados fatores que possam ter influenciado os resultados, como heterogeneidade do conjunto de dados, desbalanceamento entre classes, variações na qualidade das imagens, ausência de padronização nos rótulos e limitações dos algoritmos utilizados. A interpretação dessas métricas, associada ao contexto técnico do estudo, permitiu reconhecer gargalos que dificultam a adoção prática dos modelos, bem como desafios relacionados ao fluxo de pré-processamento, à robustez das previsões e à necessidade de validações adicionais.

4. RESULTADOS E DISCUSSÕES

No *Orange*, as análises são desenvolvidas por meio de fluxos estruturados compostos por *widgets* interligados, organizados sequencialmente para formar o processo de treinamento e teste dos modelos. Essa estrutura permite aplicar e comparar diferentes métodos de classificação, como Rede Neural, *kNN*, *Random Forest* e *SVM*, dentro de um mesmo ambiente experimental.

4.1. Treinamento Rede Neural

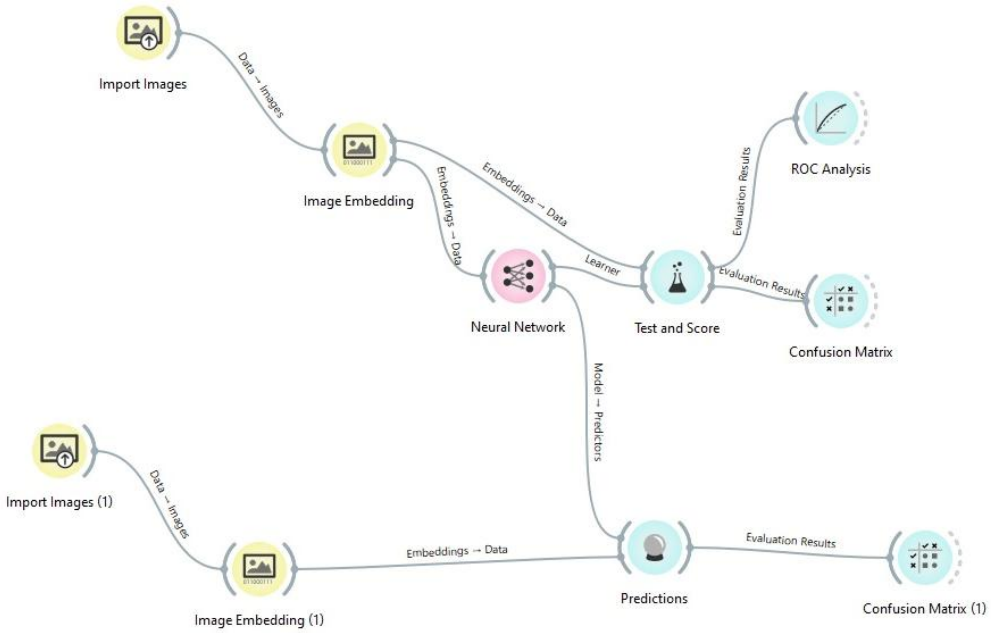
Nesta fase do projeto, foram elaborados dois fluxos distintos: o primeiro destinado ao treinamento do modelo e o segundo responsável pela etapa de testes. Ambos utilizaram o mesmo tipo de dado, composto por imagens de ultrassonografias contendo nódulos classificados como benignos ou malignos. Os dois fluxos estão representados na Figura 4.1.

No estágio de validação do modelo, foi incorporado ao fluxo o *widget Predictions*, cuja função é aplicar os classificadores já treinados a um novo conjunto de dados e retornar os resultados inferidos. Esse recurso permite observar, de forma organizada, tanto as probabilidades atribuídas a cada classe quanto a classificação final gerada pelos modelos, oferecendo uma visão clara do comportamento preditivo durante os testes.

Tanto o *Predictions* quanto o *ROC Analysis* pertencem ao mesmo módulo que também inclui o Test and Score. O avaliador *Predictions* foi conectado ao modelo preditivo Rede Neural. O *widget Predictions* fez a ligação entre o modelo treinado no fluxo de treinamento e o fluxo de teste, sendo fundamental para medir o desempenho do experimento. O esquema geral elaborado no *workflow* do *Orange* para a análise das imagens de ultrassonografia pode ser visto na Figura 4.1.

O fluxo de treinamento teve como finalidade “ensinar” o sistema a diferenciar corretamente as imagens, identificando quais representavam nódulos malignos e quais eram benignos; já o fluxo de testes teve a função de aplicar o conhecimento adquirido na etapa anterior, permitindo verificar o desempenho do modelo em dados não vistos anteriormente.

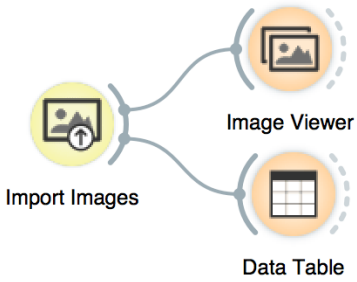
Figura 4.1 – Fluxograma do processo de classificação binária de nódulos mamários (benigno x maligno) utilizando o módulo Rede Neural



Fonte: Elaborado pela autora

Após a seleção do *dataset* destinado ao treinamento e teste, foi necessário preparar o ambiente de trabalho para recebê-los. Nessa etapa, procedeu-se à instalação da biblioteca *Image Analytics*, que não fazia parte dos *widgets* nativos do Orange, mas cuja presença é essencial para permitir o upload e o tratamento de imagens dentro da plataforma, esse *widget* é representado pela Figura 4.2. A utilização dessa biblioteca é indispensável para alcançar os objetivos propostos, uma vez que o estudo trabalha especificamente com dados visuais.

Figura 4.2 – Interface do *Orange* para importação e transformação de imagens em *embeddings, via Image Analytics*.

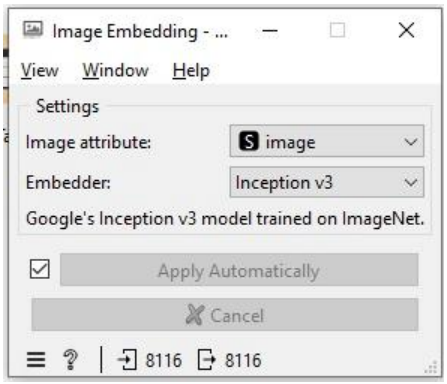


Fonte: (University of Ljubljana, 2025)

O processo iniciou-se com o módulo *Import Images*, responsável pela importação do conjunto de imagens de mama, previamente organizadas de acordo com suas classes diagnósticas. Esse módulo gerou a estrutura inicial dos dados, permitindo que cada imagem

fosse reconhecida pelo sistema juntamente com seu respectivo rótulo. Em seguida, as imagens importadas foram enviadas ao módulo *Image Embedding*, no qual foram extraídas representações numéricas a partir de modelos pré-treinados de *deep learning*. Essas representações, denominadas *embeddings*, sintetizam características visuais relevantes das imagens como textura do tecido mamário, variação de brilho e contraste, presença de bordas e contornos, formato geral das estruturas internas, densidade dos padrões radiológicos, além de características de alto nível aprendidas pela rede, como indícios de massas, texturas ou assimetrias e servem como entrada adequada para algoritmos de aprendizagem de máquina, substituindo a necessidade de trabalhar diretamente com imagens brutas, que possuem alta dimensionalidade. A Figura 4.3 mostra a configuração do *embendding* no *software*.

Figura 4.3 – Configuração do módulo de *embedding* de imagens utilizando *Inception v3* no *Orange*.



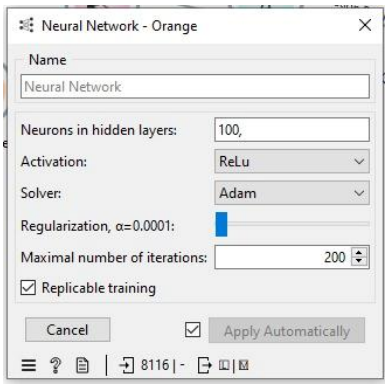
Fonte: (University of Ljubljana, 2025)

No total, foram processadas 8.116 imagens na fase de treinamento, o que significa 90% do banco de imagens utilizado, sendo 4.074 referentes a nódulos benignos e 4.042 a nódulos malignos. Na fase de testes, foram utilizadas 900 ultrassonografias, que representam os outros 10% do mesmo banco de dados, das quais 500 continham nódulos benignos e 400 apresentavam nódulos malignos.

Entre os componentes essenciais do fluxo analítico destaca-se o uso dos modelos de predição, que no *Orange* integram um conjunto específico de *widgets* denominado *Models*. Esses *widgets* são responsáveis por realizar o processamento e a classificação dos dados. Na primeira versão desenvolvida para o fluxo de treinamento, foi testado o modelo de aprendizado Rede Neural, que tem seus parâmetros definidos para teste representados na Figura 4.4, onde foi definida a quantidade de camadas adequada ao tipo de banco de dados utilizado, sendo que 100 camadas se mostraram suficientes para garantir o bom desempenho do modelo; os demais

parâmetros permaneceram com as configurações padrão do próprio software. A finalidade dessa etapa foi identificar qual o desempenho na análise das imagens o modelo apresentaria.

Figura 4.4 – Configuração para treinamento e teste da Rede Neural



Fonte: Elaborado pela Autora

A partir das análises realizadas no *widget Test and Score*, constatou-se que o modelo de Rede Neural apresentou desempenho consistente, destacando-se principalmente pelos valores de acurácia e precisão observados, conforme apresentado na Tabela 4.1. Esse widget, integrante do módulo *Evaluate* do *Orange*, é responsável por aplicar diferentes algoritmos de aprendizagem ao conjunto de dados e comparar seus resultados. Ele gera uma tabela com diversas métricas de avaliação como acurácia, precisão e área sob a curva ROC e permite encaminhar essas informações para outros widgets, possibilitando uma investigação mais detalhada do comportamento dos classificadores.

As métricas apresentadas em modelos de classificação, como na Tabela 4.1, têm como objetivo avaliar diferentes dimensões do desempenho do algoritmo. A *AUC* (*Area Under the Curve*) mede a capacidade geral do modelo de separar corretamente as classes ao longo de todos os limiares possíveis, sendo útil para entender o comportamento global do classificador. A *CA* (*Classification Accuracy*) representa a proporção total de acertos, considerando tanto verdadeiros positivos quanto verdadeiros negativos. O *F1-score* combina precisão e *recall* em uma única métrica harmônica, sendo especialmente relevante quando há desequilíbrio entre classes. A Precisão indica a proporção de previsões positivas que realmente pertencem à classe positiva, enquanto o *Recall* (Sensibilidade) mostra a capacidade do modelo de identificar corretamente todos os elementos da classe positiva. Por fim, o *MCC* (*Matthews Correlation Coefficient*) é uma métrica mais robusta que considera todos os quadrantes da matriz de confusão, oferecendo uma visão equilibrada do desempenho mesmo em bases desbalanceadas.

Juntas, essas métricas permitem uma avaliação completa e confiável da qualidade do modelo de classificação.

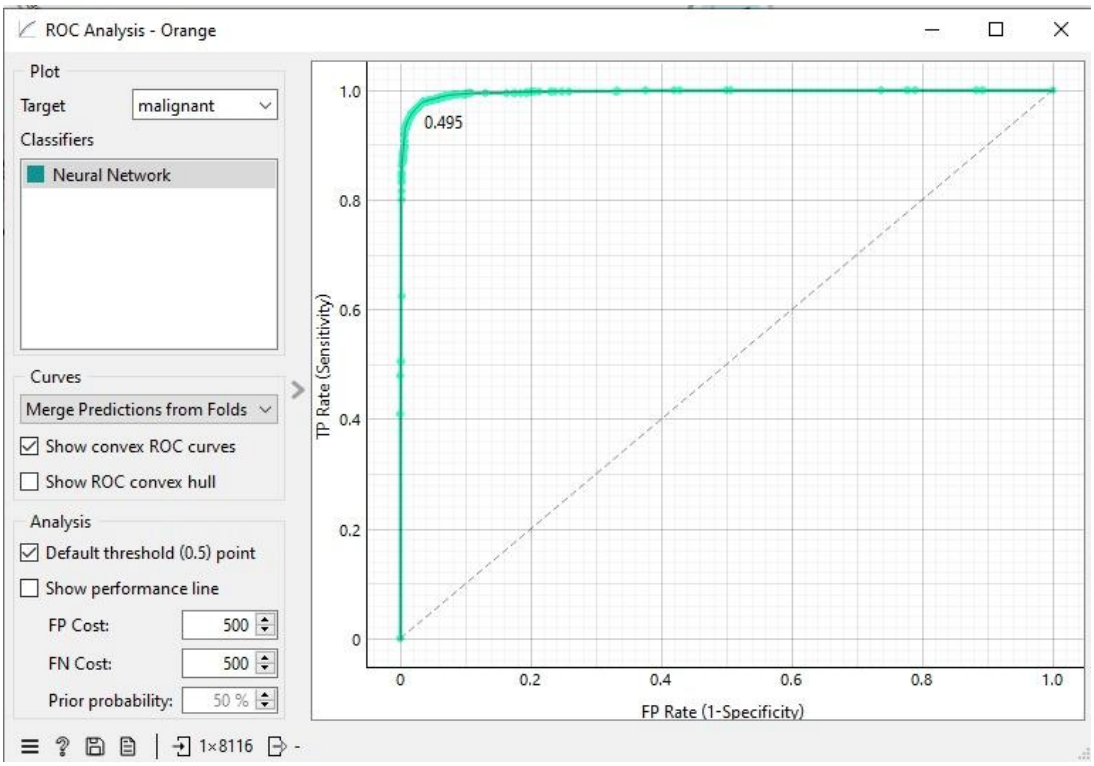
Tabela 4.1 – Indicadores de performance do classificador de imagens de ultrassonografias mamárias, na etapa de treinamento da Rede Neural

Model	AUC	CA	F1	Prec	Recall	MCC
Rede Neural	0,996	0,972	0,972	0,972	0,972	0,944

Fonte: Elaborado pela autora

No fluxo de treinamento, foi incorporado o widget *ROC Analysis* com o objetivo de comparar graficamente o desempenho do classificador. A curva *ROC* permite avaliar a precisão do modelo, sendo que a curva mais próxima da região superior esquerda indica classificações mais eficientes. A Figura 4.5 apresenta o resultado satisfatório no formato de curva *ROC* quando se trata dos testes da rede neural identificando nódulos malignos.

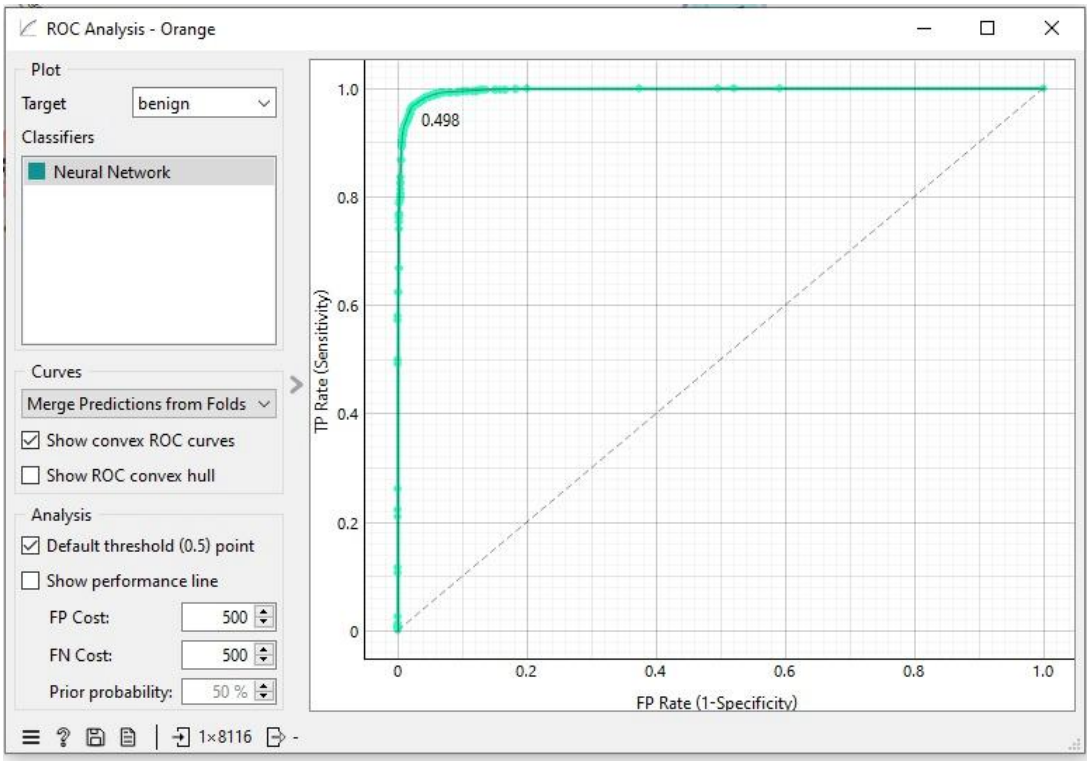
Figura 4.5 – Curva *ROC* para avaliação da sensibilidade e especificidade do modelo de classificação para nódulos malignos utilizando a Rede Neural



Fonte: Elaborado pela autora

Da mesma forma, a Figura 4.6 demonstra a curva *ROC* com resultados satisfatórios da rede neural quando está classificando nódulos benignos.

Figura 4.6 – Curva *ROC* para avaliação da sensibilidade e especificidade do modelo de classificação para nódulos benignos utilizando a Rede Neural



Fonte: Elaborado pela autora

O resultado do treinamento do modelo foi analisado por meio do *widget Confusion Matrix* (em português: Matriz de Confusão), ferramenta normalmente utilizada para visualizar o desempenho obtido através do *Test and Score*. A Matriz de Confusão apresenta a quantidade e a proporção de acertos e erros em cada classe, permitindo avaliar a eficiência do modelo.

No treinamento, a Neural Network classificou corretamente 3.963 das 4.074 imagens de nódulos benignos, cometendo 111 erros, o que corresponde a uma taxa de acerto de aproximadamente 97,3% para essa classe. Para os nódulos malignos, o modelo classificou corretamente 3.925 das 4.042 amostras, errando 117, o que representa taxa de acerto de aproximadamente 97,1% para a classe maligna. Os 117 erros para a classe maligna é o mais perigoso de todos, pois se há a presença de um nódulo maligno classificado como benigno, geralmente, para-se de fazer mais testes e só é descoberto o câncer quando já é tarde demais. A acurácia global, apresentada na Tabela 4.1 pela métrica CA, considera simultaneamente os acertos das duas classes em relação ao total de amostras avaliadas. Já o F1-score não representa a taxa de acerto de uma classe específica, mas uma métrica que combina precisão e recall. O resultado visual desse desempenho pode ser consultado na Figura 4.7.

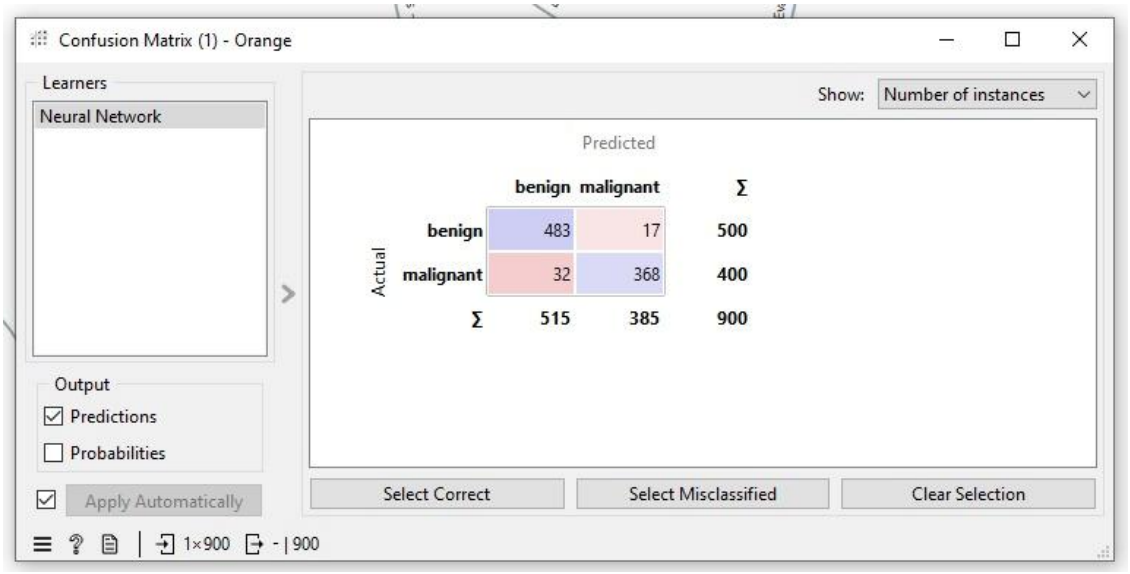
Figura 4.7 – Matriz de confusão do modelo Rede Neural aplicada ao conjunto de treinamento, demonstrando os acertos e erros na classificação de nódulos benignos e malignos

Fonte: Elaborado pela autora

4.2. Teste Rede Neural

Os resultados da etapa de treinamento mostram que a Rede Neural conseguiu aprender adequadamente os padrões presentes nas imagens utilizadas. Ainda assim, em modelos de inteligência artificial, um bom desempenho no treinamento, por si só, não comprova a eficácia do sistema, pois o modelo pode ter se ajustado a características específicas desse conjunto de dados. Por isso, a análise mais importante deve ser feita com o conjunto de teste, formado por imagens que não participaram do treinamento. Essa etapa permite avaliar com maior precisão a capacidade de generalização do modelo, ou seja, sua aptidão para classificar corretamente novas 500 imagens, em que a Rede Neural apresentou um comportamento consistente, classificando corretamente 483 das 500 imagens de nódulos benignos e cometendo 17 erros, o que corresponde a um aproveitamento de 96,6%. Para as imagens com nódulos malignos, o modelo identificou corretamente 368 das 400 amostras, com 32 erros, resultando em uma taxa de acerto de 92,0%. Em termos gerais, os resultados obtidos no fluxo de testes foram similares aos observados no treinamento, demonstrando estabilidade na assertividade do modelo. A Figura 4.8 apresenta o desempenho registrado pela rede neural.

Figura 4.8 – Matriz de confusão do modelo Rede Neural aplicada ao conjunto de teste, demonstrando os acertos e erros na classificação de nódulos benignos e malignos



Fonte: Elaborado pela autora

Considerando o total de 900 imagens avaliadas, o modelo obteve 851 classificações corretas, resultando em acurácia global de aproximadamente 94,6%, valor compatível com a CA apresentada na Tabela 4.2.

Tabela 4.2 – Indicadores de performance do classificador de imagens de ultrassonografias mamárias, na etapa de teste da Rede Neural

Model	AUC	CA	F1	Prec	Recall	MCC
Rede Neural	0,987	0,946	0,945	0,946	0,946	0,890

Fonte: Elaborado pela autora

A tabela resume o desempenho de uma Rede Neural, e os valores indicam um modelo com performance muito alta, sendo que valores acima de 0,8 já são bem vistos para a aplicação do trabalho e consistente em todas as métricas avaliadas.

O valor de $AUC = 0{,}987$ mostra que o modelo possui excelente capacidade de distinguir entre as classes, mantendo alta sensibilidade e especificidade ao longo de diferentes limiares. Já a $CA = 0{,}946$ indica que o modelo acerta cerca de 94,6% de todas as previsões, um resultado bastante elevado para tarefas de classificação.

As métricas $F1 = 0{,}945$, $Precisão = 0{,}946$ e $Recall = 0{,}946$ reforçam esse desempenho equilibrado, em que o modelo não apenas identifica corretamente a maioria dos casos positivos, como também mantém baixo o número de falsos positivos e falsos negativos.

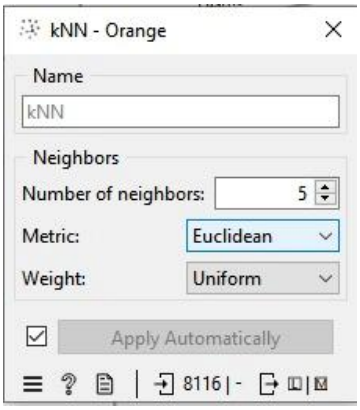
Isso demonstra que a Rede Neural é eficiente tanto em detectar corretamente quanto em evitar classificações equivocadas.

Por fim, o $MCC = 0{,}89$ confirma a robustez do modelo. Essa métrica, mais rigorosa e sensível a desbalanceamentos, indica forte correlação entre as previsões e os valores reais, evidenciando que o modelo é confiável e generaliza bem.

4.3. Treinamento *kNN*

Após os testes da Rede Neural foi feito o treinamento e testes com outros modelos de IA. O primeiro deles é o *kNN* (*K-Nearest Neighbors*) que consiste em classificar um novo dado de acordo com os "k" pontos mais próximos dele no conjunto de treinamento. A configuração para o treinamento e teste do *kNN* mostra que foram utilizados 5 vizinhos e métrica euclidiana, conforme a Figura 4.9.

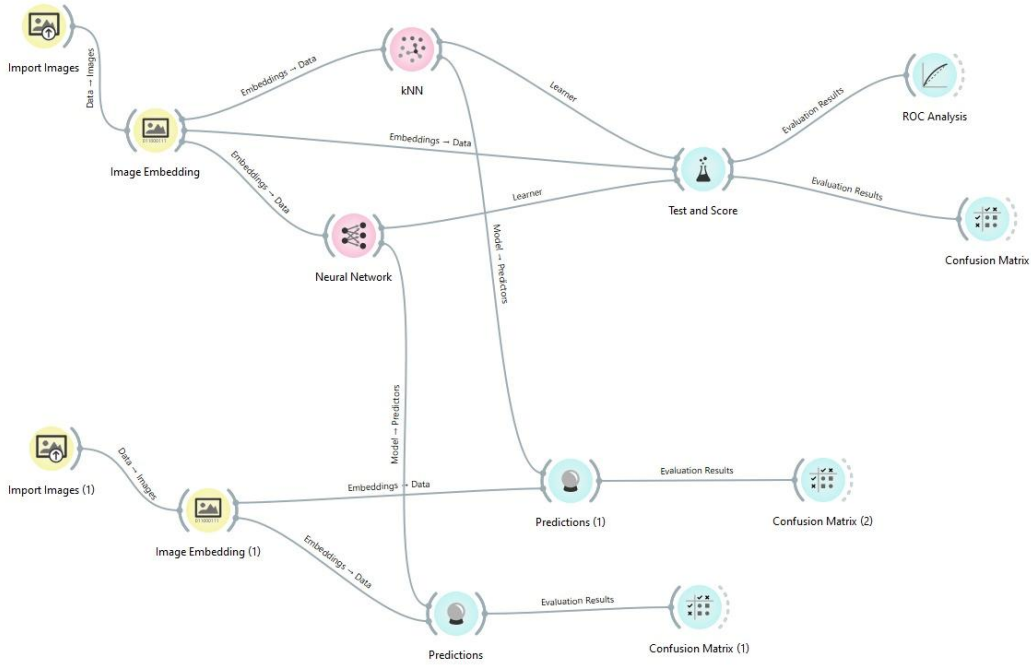
Figura 4.9 – Configuração para treinamento e teste *kNN*



Fonte: Elaborado pela autora

Para isso, foram estruturados quatro fluxos: dois destinados ao treinamento e ao teste da Rede Neural, já apresentados na Seção 4.1, e outros dois voltados ao treinamento e ao teste do *kNN*, conforme ilustrado na Figura 4.10. Os fluxos originais da Rede Neural foram mantidos para permitir, posteriormente, a comparação entre os dois modelos.

Figura 4.10 – Fluxograma do processo de classificação binária de nódulos mamários (benigno x maligno) utilizando os módulos Rede Neural e *kNN*



Fonte: Elaborado pela autora

Com os treinamentos há cinco resultados extraídos, três do conjunto de treinamento e dois do conjunto de teste. O primeiro deles é *Predictions*, conforme a Tabela 4.3, que apresenta um conjunto de métricas de desempenho, que permitem comparar a performance dos modelos Rede Neural e *kNN*.

Tabela 4.3 – Indicadores de performance do classificador de imagens de ultrassonografias mamárias, na etapa de treinamento, comparando a Rede Neural e *kNN*

Model	*AUC*	*CA*	*F1*	*Prec*	*Recall*	*MCC*
Rede Neural	0,996	0,972	0,972	0,972	0,972	0,944
kNN	0,990	0,951	0,951	0,952	0,951	0,903

Fonte: Elaborado pela autora

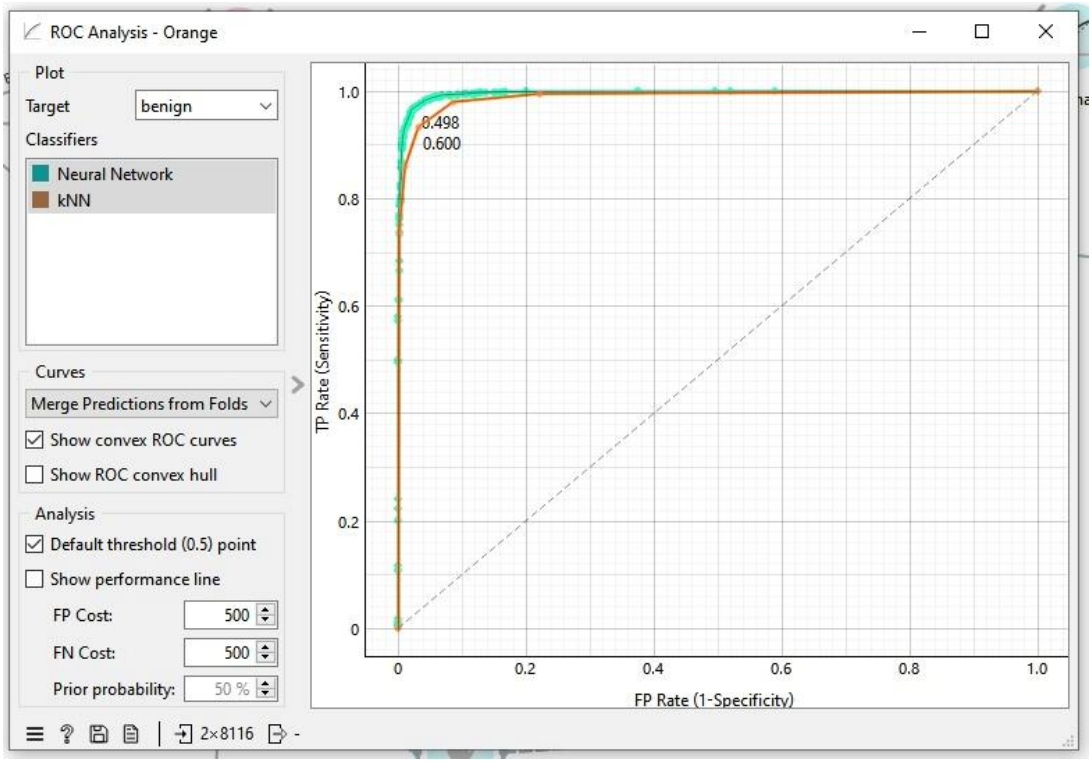
A métrica AUC evidencia a superioridade da Rede Neural que apresenta valor de 0,996, enquanto o *kNN* alcança 0,990. Embora ambos os valores sejam excelentes, o desempenho da rede neural se aproxima mais do ideal, reforçando sua maior habilidade em separar corretamente as classes. A acurácia (CA) segue o mesmo padrão, com 0,972 para a Rede Neural e 0,951 para o kNN, mostrando que o primeiro modelo acerta uma proporção maior de instâncias no conjunto de teste.

As métricas F1, Precisão e *Recall* também confirmam essa tendência. A Rede Neural apresenta valores iguais de 0,972 nessas três métricas, indicando um equilíbrio consistente entre acertos de positivos e controle de falsos positivos e falsos negativos. Já o *kNN*, embora apresente desempenho sólido, mantém valores ligeiramente inferiores $F1 = 0{,}951$; $Precisão = 0{,}952$; $Recall = 0{,}951$, o que demonstra que o modelo é um pouco menos eficiente tanto em identificar corretamente as instâncias positivas quanto em evitar erros de classificação.

Já o MCC, uma métrica que considera todos os elementos da matriz de confusão, reforça a superioridade da Rede Neural neste estudo, com valor de 0,944, contra 0,903 do *kNN*. Como o MCC é especialmente sensível ao equilíbrio entre classes, esse resultado indica que a rede neural mantém uma correlação mais forte entre predições e valores reais, mesmo diante de possíveis assimetrias nos dados.

O segundo resultado extraído é a Curva *ROC*, que mostra em verde a avaliação da sensibilidade da Rede Neural e em amarelo a avaliação da sensibilidade do modelo *kNN*, conforme a Figura 4.11.

Figura 4.11 – Curva *ROC* para avaliação da sensibilidade e especificidade do modelo de classificação para nódulos benignos comparando a Rede Neural e a *kNN*

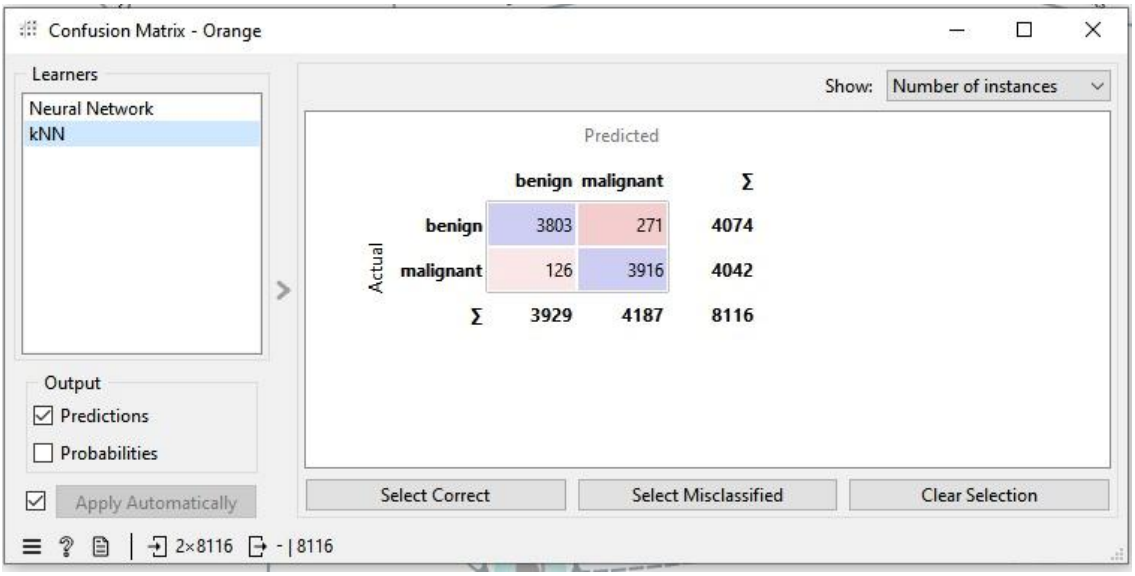


Fonte: Elaborado pela autora

A curva *ROC* evidencia que ambos os modelos apresentam desempenho superior ao acaso, uma vez que suas curvas permanecem acima da linha diagonal, embora a Rede Neural demonstre desempenho superior ao do *kNN* por se aproximar mais do canto superior esquerdo do gráfico, indicando maior sensibilidade com menor taxa de falsos positivos. Dessa forma, a comparação direta entre as curvas mostra que a Rede Neural domina o *kNN* em toda a extensão do gráfico, apresentando maior capacidade discriminativa independentemente do limiar adotado.

O terceiro é a *Confusion Matrix* do modelo *kNN*, que permite avaliar de forma detalhada o desempenho do classificador na distinção entre as classes *benign* (benigno) e *malignant* (maligno), conforme Figura 4.12.

Figura 4.12 – Matriz de confusão do modelo *kNN* aplicada ao conjunto de treinamento, demonstrando os acertos e erros na classificação de nódulos benignos e malignos



Fonte: Elaborado pela autora

Observa-se que o modelo classificou corretamente 3.803 instâncias benignas, enquanto 271 foram incorretamente rotuladas como malignas. Esse resultado indica que o *kNN* apresenta boa capacidade de identificar casos benignos, com um número relativamente baixo de falsos negativos, o que demonstra uma sensibilidade satisfatória para essa classe. No caso da classe dos malignos, o modelo também apresentou bom desempenho, acertando 3.916 instâncias e cometendo apenas 126 erros ao classificá-las como benignas.

A distribuição dos acertos e erros revela que o modelo mantém um equilíbrio adequado entre as duas classes, sem tendência significativa a favorecer uma delas. O total de acertos

$3.803 + 3.916 = 7.719$ é substancialmente superior ao total de erros $126 + 271 = 397$, o que evidencia uma acurácia global elevada. Além disso, o baixo volume de classificações incorretas reforça a capacidade discriminativa do modelo, em consonância com o bom desempenho observado anteriormente na análise da curva *ROC*.

4.4. Teste *kNN*

Seguindo para o conjunto de testes, temos mais dois resultados extraídos, sendo o primeiro deles o *Predictions*, que mostra os indicadores de performance para o classificador de imagens de ultrassonografias mamárias do teste do modelo *kNN*, conforme mostra a Tabela 4.4.

Tabela 4.4 – Indicadores de performance do classificador de imagens de ultrassonografias mamárias, na etapa de teste, comparando a Rede Neural e *kNN*

Model	*AUC*	*CA*	*F1*	*Prec*	*Recall*	*MCC*
Rede Neural	0,987	0,946	0,945	0,946	0,946	0,890
kNN	0,931	0,858	0,858	0,861	0,858	0,716

Fonte: Elaborado pela autora

O desempenho do *kNN*, com base nos valores apresentados na tabela, mostra que o modelo apresenta resultados consistentes, porém inferiores aos obtidos pela Rede Neural. O $AUC = 0{,}931$ indica que o *kNN* possui uma boa capacidade de discriminar entre as classes, embora não alcance níveis de excelência. Esse valor demonstra que o modelo consegue separar razoavelmente bem os casos positivos e negativos ao longo de diferentes limiares de decisão.

As métricas CA, F1, Precisão e *Recall*, todas em torno de 0,858 e 0,861, revelam um desempenho equilibrado, em que o modelo acerta uma proporção semelhante de casos positivos e negativos, mantendo coerência entre sensibilidade e precisão. Isso significa que o *kNN* não tende a favorecer excessivamente uma classe em detrimento da outra.

O $MCC = 0{,}716$ reforça a interpretação de que o *kNN* apresenta um desempenho sólido, ainda que moderado. O MCC é uma métrica mais rigorosa e sensível a desbalanceamentos, e um valor acima de 0,70 indica que o modelo mantém uma boa correlação entre as previsões e os valores reais.

Por fim, também obtivemos os resultados do teste por meio da *Confusion Matrix*, que mostra de forma quantitativa a capacidade de classificação do modelo *kNN*, conforme apresentado na Figura 4.13.

Figura 4.13 – Matriz de confusão do modelo *kNN* aplicada ao conjunto de teste, demonstrando os acertos e erros na classificação de nódulos benignos e malignos

Fonte: Elaborado pela autora

A matriz de confusão apresentada resume o desempenho do teste do modelo *kNN* na classificação das amostras entre as classes *benign* e *malignant*, permitindo avaliar os acertos e os erros do modelo. Considerando a classe "maligno" como positiva, quando um caso maligno é classificado como benigno ocorre um falso negativo, que representa o erro clinicamente mais crítico. Entre as 500 amostras benignas, o modelo classificou corretamente 419 delas, alcançando 83,8% de acerto, enquanto 81 foram incorretamente previstas como malignas, representando 16,2% de erro. Esses casos correspondem aos falsos positivos, que, embora indesejáveis, costumam ser menos críticos em diagnósticos médicos, já que levam a exames complementares, não à perda de detecção da doença.

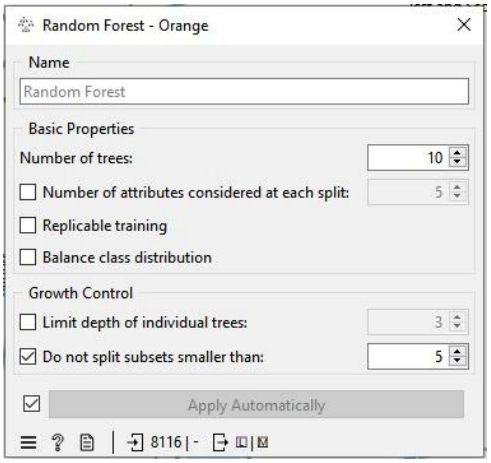
Nos casos malignos, o modelo apresentou desempenho um pouco melhor: das 400 amostras, 353 foram corretamente identificadas como malignas (88,3% de acerto), enquanto 47 foram classificadas como benignas (11,7% de erro). Esses falsos negativos são os erros mais sensíveis, pois representam situações em que o modelo falha em identificar a presença da doença, podendo atrasar o diagnóstico e o tratamento.

Considerando o conjunto total de 900 amostras, o *kNN* obteve 772 acertos, o que corresponde a 85,8% de acurácia, enquanto 128 previsões foram incorretas (14,2% de erro). Esses resultados mostram que o modelo apresenta um desempenho sólido e relativamente equilibrado entre as classes, embora ainda exista espaço para melhorias, especialmente na redução dos falsos negativos, que são os erros mais críticos em aplicações de saúde.

4.5. Treinamento *Random Forest*

O segundo método para treinamento e teste é o *Random Forest*, que é um método de aprendizado de máquina que combina várias árvores de decisão para melhorar a precisão e reduzir *overfitting*. Cada árvore é treinada com amostras aleatórias dos dados e, ao final, todas "votam" no resultado. A configuração para o treinamento e teste do *Random Forest* mostra que foram utilizadas 10 árvores de decisão, conforme mostra a Figura 4.14.

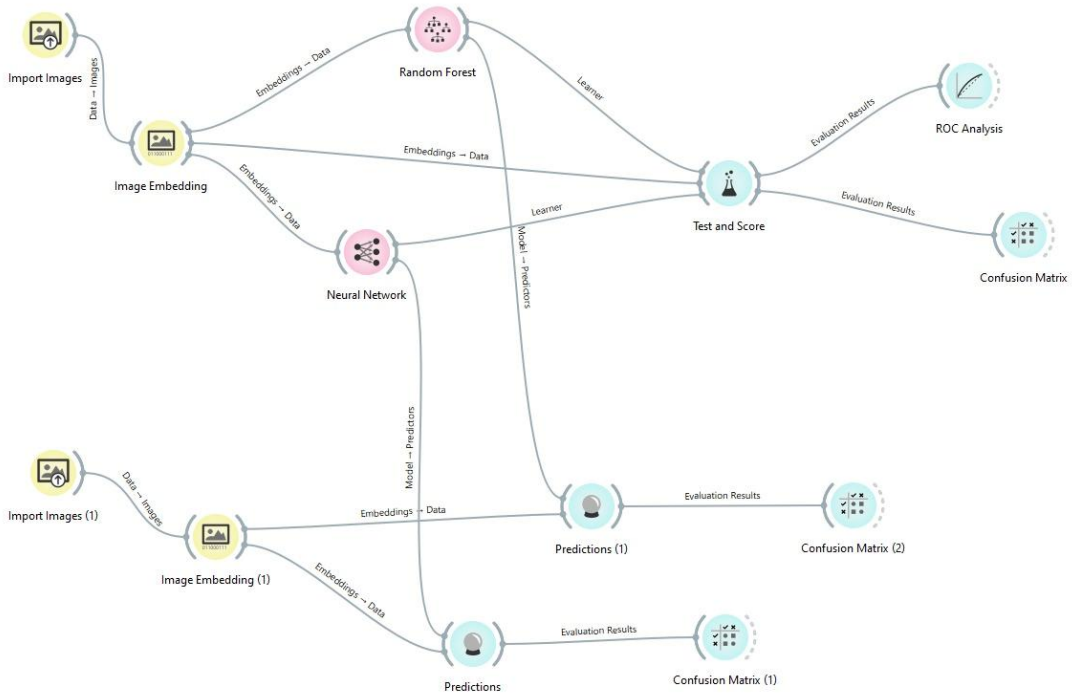
Figura 4.14 – Configuração para treinamento e teste *Random Forest*



Fonte: Elaborado pela autora

Da mesma forma feita com o método *kNN* foram montados quatro fluxos: dois para treinamento e teste da Rede Neural e outros dois para treinamento e teste da *Random Forest*, conforme a Figura 4.15 mostra. E, da mesma forma, os fluxos da Rede Neural foram mantidos para ser possível para a comparação entre os dois modelos.

Figura 4.15 – Fluxograma do processo de classificação binária de nódulos mamários (benigno x maligno) utilizando os módulos Rede Neural e *Random Forest*



Fonte: *Orange Data Mining*

Após a configuração e a montagem dos fluxos são extraídos cinco resultados, assim como no modelo *kNN*, três do conjunto de treinamento e dois do conjunto de teste. Sendo o primeiro deles o *Test and Score*, conforme exposto na Tabela 4.5, que reúne métricas de desempenho essenciais para a comparação da performance entre os modelos Rede Neural e *Random Forest*.

Tabela 4.5 – Indicadores de performance do classificador de imagens de ultrassonografias mamárias, na etapa de treinamento, comparando a Rede Neural e *Random Forest*

Model	*AUC*	*CA*	*F1*	*Prec*	*Recall*	*MCC*
Rede Neural	0,996	0,972	0,972	0,972	0,972	0,944
Random Forest	0,934	0,861	0,861	0,861	0,861	0,722

Fonte: Elaborado pela autora

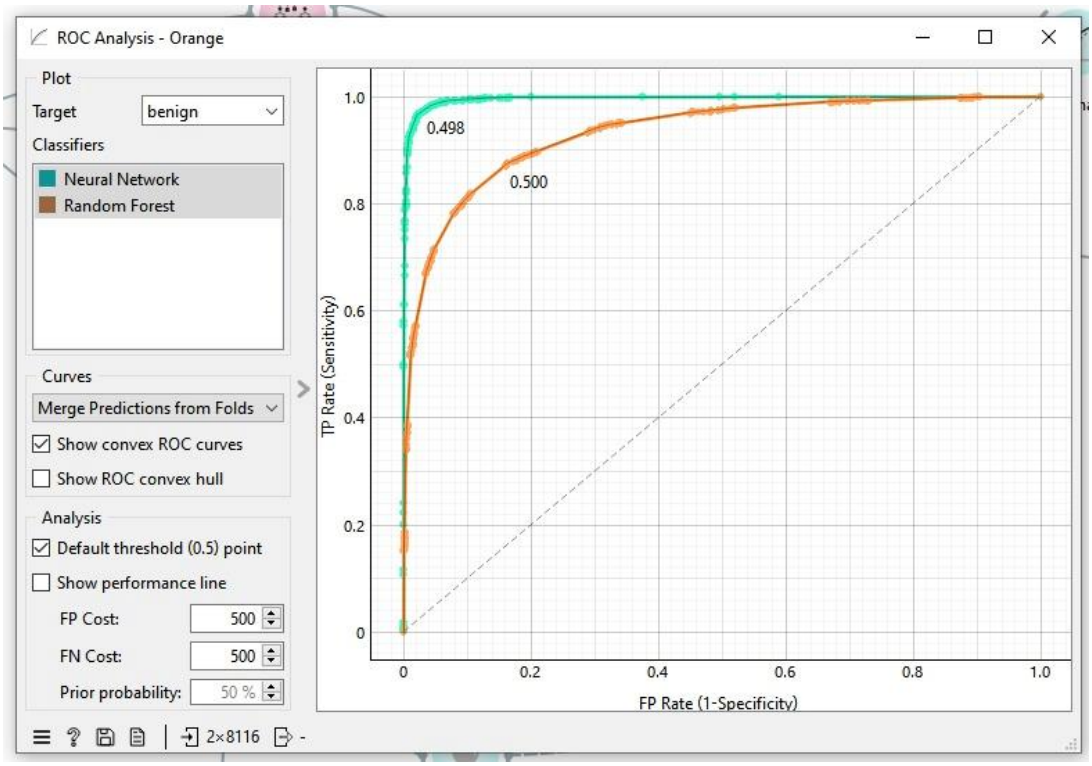
O desempenho do *Random Forest* na tabela mostra um modelo funcional, porém claramente inferior à Rede Neural em todos os indicadores avaliados. O valor de $AUC = 0{,}934$ mostra que o *Random Forest* possui uma boa capacidade de discriminar entre as classes, mas ainda distante da performance quase perfeita da Rede Neural. Já as métricas CA, F1, Precisão e *Recall*, todas em 0,861, indicam que o modelo mantém um equilíbrio razoável entre

acertos gerais, capacidade de identificar corretamente casos positivos e controle de falsos positivos, mas com uma taxa de acerto consideravelmente menor que a do modelo concorrente.

O $MCC = 0{,}722$, embora positivo e indicativo de correlação moderada entre previsões e valores reais, reforça que o *Random Forest* não alcança o mesmo nível de robustez e confiabilidade observado na Rede Neural, esse valor evidencia que o modelo ainda comete uma quantidade relevante de erros.

O segundo resultado extraído é a curva *ROC*, que mostra as curvas de acordo com a legenda à esquerda da Figura 4.16.

Figura 4.16 – Curva *ROC* para avaliação da sensibilidade e especificidade do modelo de classificação para nódulos benignos comparando a Rede Neural e a *Random Forest*



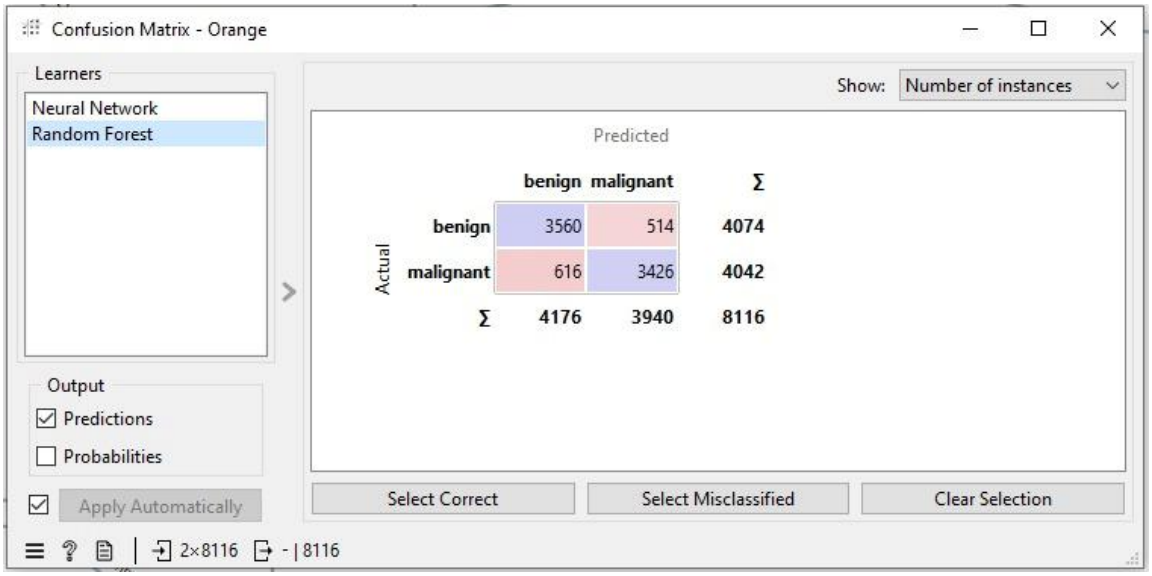
Fonte: Elaborado pela autora

A curva *ROC* apresentada permite comparar o desempenho da Rede Neural e do *Random Forest* na tarefa de distinguir corretamente entre amostras benignas e malignas. Observa-se que a curva da Rede Neural permanece mais próxima do canto superior esquerdo do gráfico, indicando maior sensibilidade para os mesmos níveis de taxa de falsos positivos. Por outro lado, a curva do *Random Forest* se mantém mais baixa e mais próxima da diagonal

aleatória, o que revela menor capacidade discriminativa. Esse comportamento indica que o modelo apresenta desempenho inferior ao da Rede Neural.

Por último, no treinamento, temos a *Confusion Matrix* da *Random Forest,* que resume o desempenho do modelo ao classificar amostras como benignas ou malignas, conforme a Figura 4.17.

Figura 4.17 – Matriz de confusão do modelo *Random Forest* aplicada ao conjunto de treinamento, demonstrando os acertos e erros na classificação de nódulos benignos e malignos



Fonte: Elaborado pela autora

A matriz de confusão apresentada permite avaliar o desempenho do *Random Forest* de forma detalhada, revelando como o modelo se comporta ao classificar amostras entre as classes *benign* e *malignant*.

Entre as 4.074 amostras benignas, o modelo classificou corretamente 3.560, o que corresponde a 87,4% de acerto, enquanto 514 foram incorretamente previstas como malignas, representando 12,6% de erro. Já entre as 4.042 amostras malignas, o modelo identificou corretamente 3.426 delas (84,8% de acerto), mas classificou 616 como benignas (15,2% de erro).

No total de 8.116 amostras, o *Random Forest* acertou 6.986 previsões, o que equivale a 86,1% de acurácia, enquanto 1.130 amostras foram classificadas incorretamente (13,9% de erro). Esses resultados mostram que o modelo apresenta um desempenho sólido, com boa taxa de acertos em ambas as classes, mas ainda com uma quantidade relevante de falsos negativos,

indicando que, embora eficiente, o modelo pode não ser o mais adequado quando a prioridade é minimizar ao máximo a chance de deixar casos malignos passarem despercebidos.

4.6. Teste *Random Forest*

Na fase de testes, foi considerada as previsões obtidas por meio do *widget Predictions*, conforme mostrado na Tabela 4.6. Os resultados permitiram identificar padrões significativos nos dados avaliados e serviram como base para comparações com outros métodos utilizados no projeto. A interpretação desses valores foi fundamental para determinar o desempenho do modelo aplicado.

Tabela 4.6 – Indicadores de performance do classificador de imagens de ultrassonografias mamárias, na etapa de teste, comparando a Rede Neural e *Random Forest*

Model	*AUC*	*CA*	*F1*	*Prec*	*Recall*	*MCC*
Rede Neural	0,987	0,946	0,945	0,946	0,946	0,890
Random Forest	0,923	0,833	0,830	0,840	0,833	0,666

Fonte: Elaborado pela autora

O desempenho do *Random Forest* na tabela reforça o que vimos no treinamento, um modelo competente, porém claramente inferior ao Rede Neural em todos os indicadores avaliados. O valor de $AUC = 0{,}923$ indica que o modelo focal dessa análise possui uma boa capacidade de discriminar entre as classes, mas ainda distante do desempenho mais robusto do modelo concorrente. Já as métricas CA, F1, Precisão e *Recall*, todas entre 0,830 e 0,840, mostram que o modelo mantém um equilíbrio razoável entre acertos gerais, capacidade de identificar corretamente os casos positivos e controle de falsos positivos. No entanto, esses valores evidenciam uma taxa de acerto mais modesta, sugerindo que o *Random Forest* comete uma quantidade maior de erros em comparação à Rede Neural.

O $MCC = 0{,}666$, reforça essa interpretação. Embora o valor seja positivo e indique correlação significativa entre previsões e valores reais, ele também mostra que o modelo não alcança o mesmo nível de confiabilidade e consistência observado no Rede Neural.

O segundo e último resultado dos testes consiste na *Confusion Matrix*, que apresenta quantitativamente como o modelo *Random Forest* realiza a classificação, conforme ilustrado na Figura 4.18.

Figura 4.18 – Matriz de confusão do modelo *Random Forest* aplicada ao conjunto de teste, demonstrando os acertos e erros na classificação de nódulos benignos e malignos

Fonte: Elaborado pela autora

A matriz de confusão apresentada permite avaliar o desempenho do modelo *Random Forest* tanto em números absolutos quanto em porcentagens.

O modelo conseguiu identificar corretamente 463 das 500 amostras benignas, atingindo um índice de acerto de 92,6%. As outras 37 foram classificadas erroneamente como malignas, o que corresponde a 7,4% de erro. Esses casos, conhecidos como falsos positivos, são considerados menos graves em contexto médico, já que geralmente resultam apenas na realização de exames adicionais em vez de falha na identificação da doença.

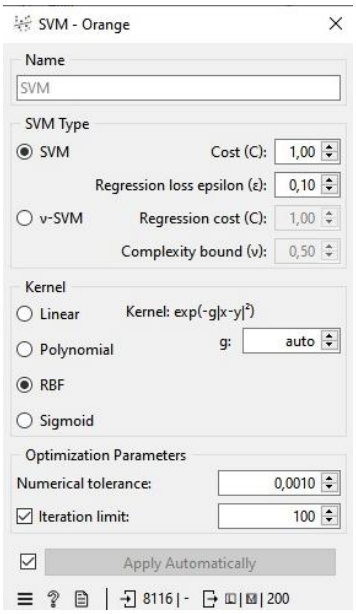
Das 400 amostras malignas avaliadas, o modelo obteve uma taxa de acerto de 71,8%, identificando corretamente 287 casos. No entanto, 113 amostras foram erroneamente classificadas como benignas, resultando em uma taxa de erro de 28,2%. Esses falsos negativos são particularmente relevantes, pois correspondem a situações em que a presença da doença não foi detectada pelo modelo, o que pode impactar negativamente o diagnóstico e o início do tratamento.

No total de 900 amostras, o Random Forest acertou 750 previsões, o que equivale a 83,3% de acurácia, enquanto 150 amostras foram classificadas incorretamente, que significa 16,7% de erro.

4.7. Treinamento *SVM*

O terceiro modelo de IA utilizado para comparação nessa pesquisa é o *SVM*, que conforme mostrado na Seção 2.2.7 é um método de aprendizado supervisionado que busca encontrar o hiperplano que melhor separa as classes, maximizando a margem entre elas. Essa maximização da distância entre os grupos garante maior capacidade de generalização e torna o modelo eficiente para tarefas de classificação e regressão. A Figura 4.19, mostra que foi utilizado o $C = 1$ e $\varepsilon = 0{,}1$, além do uso do *Kernel RBF*.

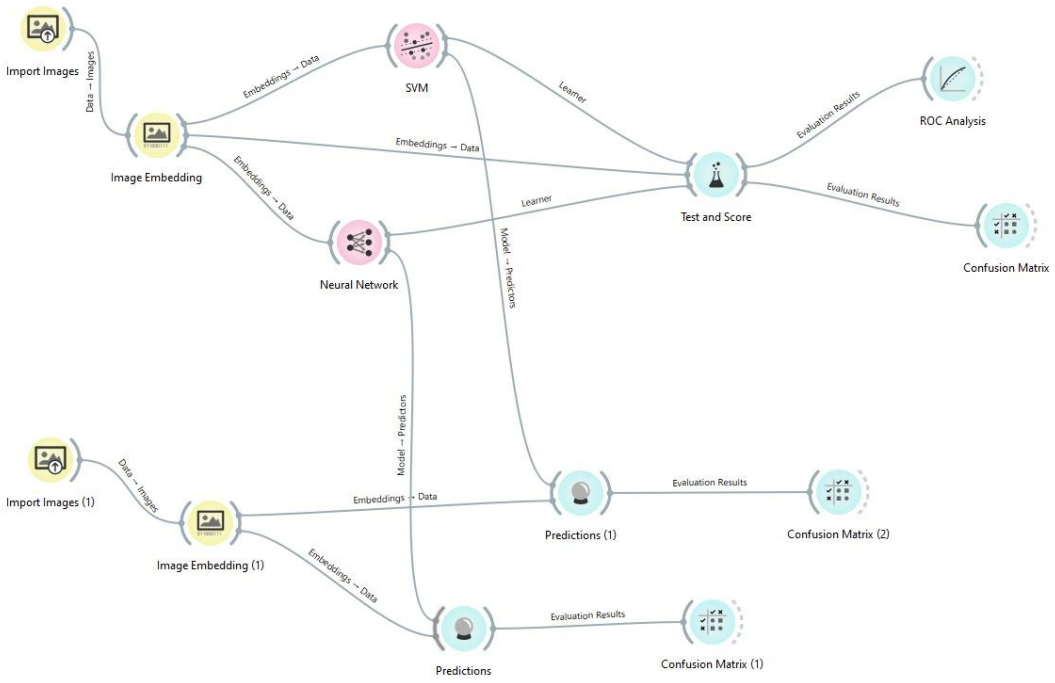
Figura 4.19 – Configuração para treinamento e teste *SVM*



Fonte: Elaborado pela autora

Como nos modelos anteriores, foram organizados os fluxos de treinamento e teste tanto para o *SVM* quanto para a Rede Neural, permitindo assim que os dois métodos fossem comparados. O fluxograma ilustrado na Figura 4.20 representa esse processo.

Figura 4.20 – Fluxograma do processo de classificação binária de nódulos mamários (benigno x maligno) utilizando os módulos Rede Neural e *SVM*



Fonte: Elaborado pela autora

Após montar e configurar os fluxos, são extraídos cinco resultados: três do conjunto de treinamento e dois do conjunto de teste. O primeiro, *Test and Score*, apresenta métricas essenciais para comparar o desempenho dos modelos Rede Neural e SVM, conforme mostra a Tabela 4.7.

Tabela 4.7 – Indicadores de performance do classificador de imagens mamográficas, na etapa de treinamento, comparando a Rede Neural e *SVM*

Model	*AUC*	*CA*	*F1*	*Prec*	*Recall*	*MCC*
Rede Neural	0,996	0,972	0,972	0,972	0,972	0,944
SVM	0,871	0,792	0,791	0,794	0,792	0,585

Fonte: Elaborado pela autora

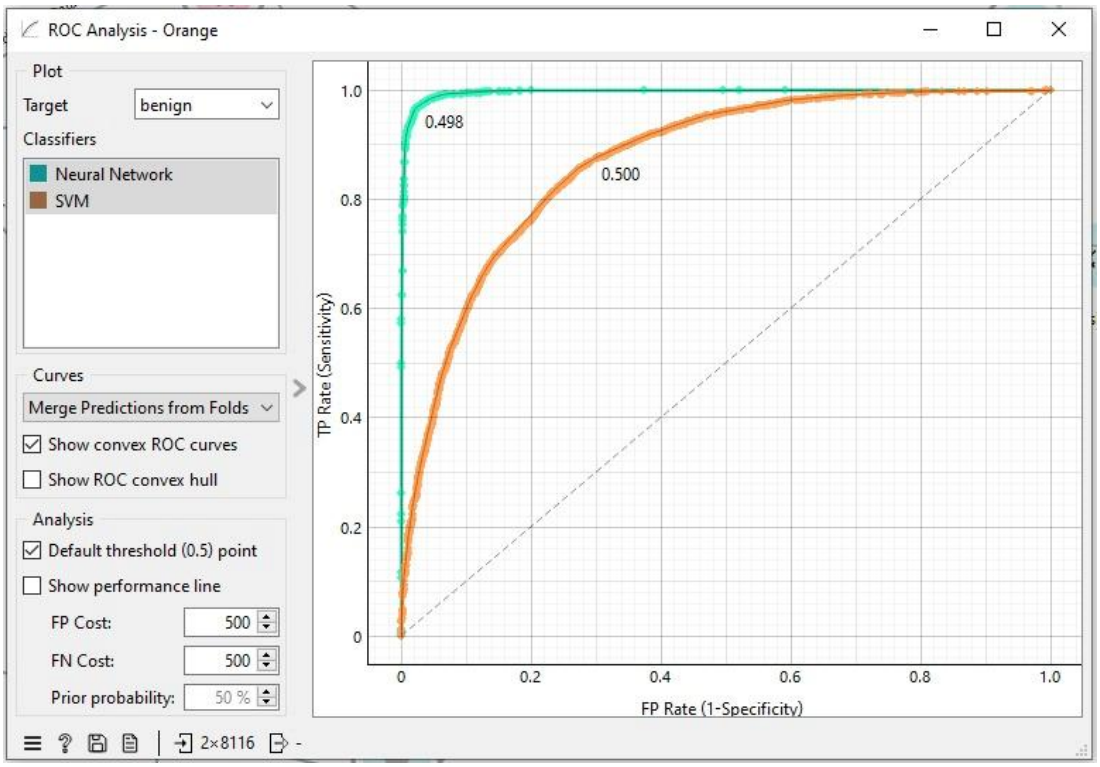
A tabela compara o desempenho do *SVM* com o Rede Neural, revelando uma diferença significativa entre os dois modelos em todas as métricas avaliadas. O *SVM* apresenta $AUC = 0{,}871$, indicando uma capacidade moderada de separação entre as classes, bem inferior ao valor de 0,996 obtido pela Rede Neural.

As métricas CA, F1, Precisão e *Recall*, todas entre de 0,791 e 0,794, mostram que o modelo acerta cerca de 79% das previsões e mantém um equilíbrio razoável entre sensibilidade e precisão. No entanto, esses valores evidenciam que o *SVM* comete mais erros tanto ao identificar casos positivos quanto negativos, refletindo uma performance mais limitada quando comparado a modelos mais robustos.

O $MCC = 0{,}585$ reforça essa interpretação. Como essa métrica é mais rigorosa e sensível a desbalanceamentos, um valor abaixo de 0,60 indica correlação apenas moderada entre as previsões e os valores reais, sugerindo que o *SVM* não apresenta alta confiabilidade para a tarefa analisada.

O segundo resultado obtido na fase de treinamento é a curva *ROC*, que mostra o comportamento do modelo *SVM*, em laranja, em comparação com a Rede Neural, em verde. A Figura 4.21 mostra esse comportamento.

Figura 4.21 – Curva *ROC* para avaliação da sensibilidade e especificidade do modelo de classificação para nódulos benignos comparando a Rede Neural e a *SVM*



Fonte: Elaborado pela autora

A curva *ROC* do *SVM* evidencia um desempenho apenas moderado do modelo na tarefa de classificação. A trajetória da curva permanece distante do canto superior esquerdo do

gráfico. Isso indica que o modelo tem dificuldade em manter simultaneamente uma boa capacidade de identificar corretamente os casos positivos e, ao mesmo tempo, evitar classificações incorretas de negativos como positivos.

Além disso, a inclinação mais suave da curva sugere que, ao longo dos diferentes limiares de decisão, o modelo não alcança níveis elevados de discriminação entre as classes. Ou seja, o *SVM* apresenta limitações tanto na sensibilidade quanto na especificidade. A presença de pontos de limiar próximos ao valor padrão reforça que ajustes no *threshold* não produzem ganhos expressivos de desempenho.

No conjunto, a análise visual da *ROC* mostra que o *SVM* funciona, mas sua capacidade de separação entre as classes é limitada, e o modelo demonstra menor robustez quando comparado a Rede Neural. Isso o torna menos adequado para cenários em que a precisão e a redução de erros críticos são essenciais, como em aplicações médicas.

Para concluir o treinamento, o resultado obtido é a *Confusion Matrix*, utilizada para avaliar o desempenho do modelo ao classificar amostras entre as classes *benign* e *malignant*. Esse resultado está ilustrado na Figura 4.22.

Figura 4.22 – Matriz de confusão do modelo *SVM* aplicada ao conjunto de treinamento, demonstrando os acertos e erros na classificação de nódulos benignos e malignos

Fonte: Elaborado pela autora

O *SVM* conseguiu identificar corretamente 3.415 das 4.074 amostras benignas, representando uma precisão de 83,9%, no entanto, classificou 659 dessas amostras como

malignas, ou seja, 16,1% incorretamente. Quanto às amostras malignas, das 4.042 analisadas, 3.009 foram reconhecidas de modo correto, o que representa 74,5% de acerto, enquanto 1.033 receberam classificação incorreta como benignas, correspondendo a 25,5% de erro.

Em um total de 8.116 amostras, o *SVM* obteve 6.424 previsões corretas, correspondendo a uma acurácia de 79,2%, enquanto 1.692 amostras foram classificadas de forma incorreta, resultando em uma taxa de erro de 20,8%. Os resultados indicam que o *SVM* proporciona uma capacidade moderada para identificar casos benignos, apresentando uma taxa significativa de falsos negativos.

4.8. Teste *SVM*

Ao avançar para a etapa de testes, temos dois resultados extraídos. O primeiro deles é a previsão, que pode ser medida pelos indicadores apresentados na Tabela 4.8.

Tabela 4.8 – Indicadores de performance do classificador de imagens mamográficas, na etapa de teste, comparando a Rede Neural e *SVM*

Model	*AUC*	*CA*	*F1*	*Prec*	*Recall*	*MCC*
Rede Neural	0,987	0,946	0,945	0,946	0,946	0,890
SVM	0,860	0,778	0,770	0,793	0,778	0,558

Fonte: Elaborado pela autora

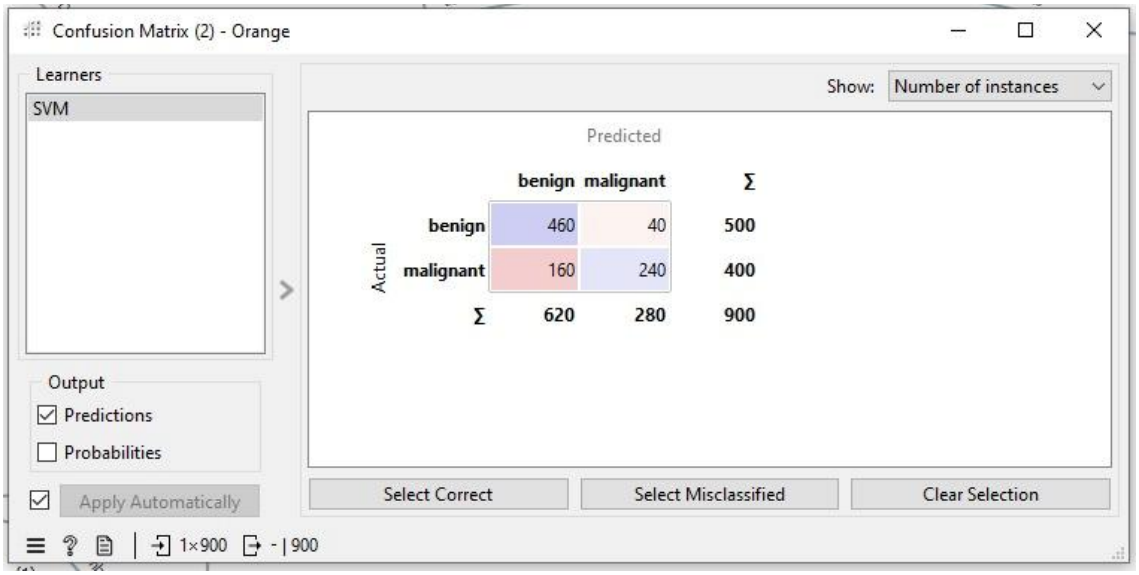
O desempenho do *SVM* apresentado na tabela revela um modelo com performance moderada, ficando claramente abaixo da Rede Neural em todos os indicadores avaliados. O valor de $AUC = 0{,}860$ indica que o *SVM* possui uma capacidade limitada de separar corretamente as classes ao longo dos diferentes limiares de decisão, sugerindo menor poder discriminativo.

As métricas $CA = 0{,}778$, $F1 = 0{,}770$, $Precisão = 0{,}793$ e $Recall = 0{,}778$ mostram que o modelo acertou cerca de 77,8% das previsões, apesar de apresentar um desempenho razoável entre identificar corretamente os casos positivos e evitar falsos positivos. No entanto, esses valores evidenciam que o *SVM* demonstrou limitação importante na detecção da classe maligna, com 160 falsos negativos.

O $MCC = 0{,}558$ reforça essa interpretação, onde um valor abaixo de 0,60 indica que o modelo apresenta apenas uma correlação moderada entre as previsões e os valores reais, revelando menor robustez e confiabilidade.

Para finalizar a etapa de testes, foi obtido o resultado da *Confusion Matrix* para o modelo *SVM*, conforme apresentado na Figura 4.23.

Figura 4.23 – Matriz de confusão do modelo *SVM* aplicada ao conjunto de teste, demonstrando os acertos e erros na classificação de nódulos benignos e malignos



Fonte: Elaborado pela autora

A matriz de confusão do *SVM* mostra um desempenho moderado do modelo ao separar as classes *benign* e *malignant*. Entre as 500 amostras benignas, o modelo classificou corretamente 460 delas, o que representa 92% de acerto, enquanto 40 foram rotuladas incorretamente como malignas, correspondendo a 8% de erro. Por outro lado, o desempenho do modelo na classe maligna é mais preocupante, onde em 400 amostras de nódulos malignos, apenas 240 foram corretamente identificadas, o que equivale a 60% de acerto, enquanto 160 foram classificadas como benignas, representando 40% de erro. Esses falsos negativos são os erros mais graves, pois indicam casos de doença que o modelo não conseguiu detectar.

No total de 900 amostras avaliadas, o *SVM* acertou 700 previsões, resultando em uma acurácia global de 77,8%, enquanto 200 amostras foram classificadas incorretamente, correspondendo a 22,2% de erro. Resumindo, o modelo consegue identificar bem os casos benignos, porém tem dificuldades expressivas na detecção dos casos malignos, o que compromete sua confiabilidade em situações críticas, principalmente em aplicações de saúde.

4.9. Resultados e Comparação entre os Modelos Avaliados

Ao consolidar os dados da etapa de treinamento, conforme a Tabela 4.9, permite observar um padrão consistente entre os modelos avaliados. De modo amplo, nota-se que as técnicas mais robustas, como a Rede Neural e, em menor grau, o *kNN*, apresentaram desempenho superior nas principais métricas, indicando maior capacidade de distinguir corretamente as classes do conjunto de dados. Já modelos como *Random Forest* e *SVM* mostraram resultados mais modestos, sugerindo menor adaptação às características específicas das imagens analisadas. No conjunto, as análises apontam para uma tendência clara: algoritmos capazes de capturar relações mais complexas nos dados tendem a oferecer classificações mais precisas e equilibradas, reforçando a importância de selecionar modelos adequados ao tipo de informação trabalhada e ao objetivo diagnóstico do estudo.

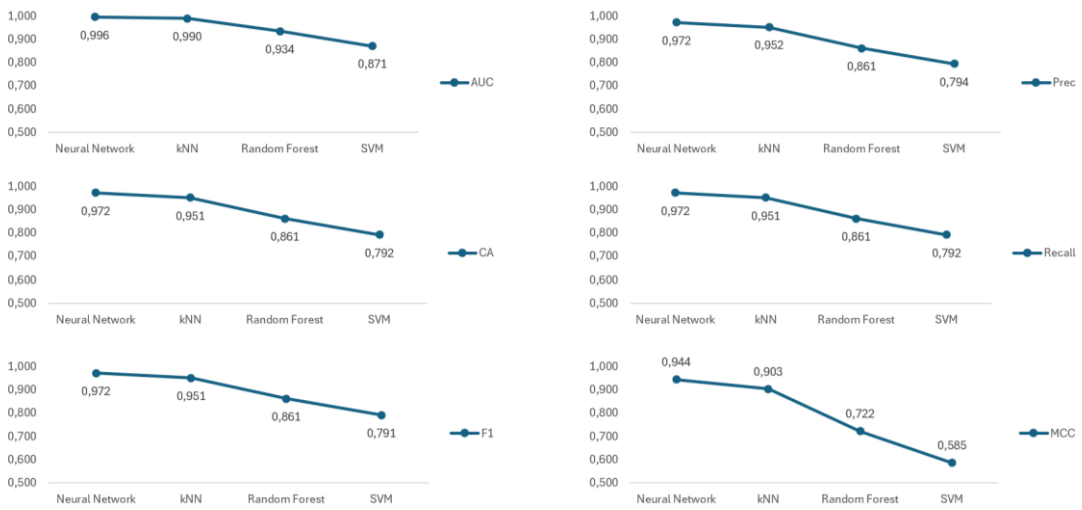
Tabela 4.9 – Indicadores de performance do classificador de imagens mamográficas, na etapa de treinamento, comparando a Rede Neural, *kNN*, *Random Forest* e *SVM*

Model	*AUC*	*CA*	*F1*	*Prec*	*Recall*	*MCC*
Rede Neural	0,996	0,972	0,972	0,972	0,972	0,944
kNN	0,990	0,951	0,951	0,952	0,951	0,903
Random Forest	0,934	0,861	0,861	0,861	0,861	0,722
SVM	0,871	0,792	0,791	0,794	0,792	0,585

Fonte: Elaborado pela autora

A Figura 4.24, reforça os pontos citados anteriormente de maneira gráfica.

Figura 4.24 – Indicadores de performance em forma gráfica, do classificador de imagens mamográficas, na etapa de treinamento, comparando a Rede Neural, *kNN*, *Random Forest* e *SVM*



Fonte: Elaborado pela autora

Da mesma forma da etapa de treinamento já apresentada, os resultados do teste, conforme apresentado na Tabela 4.10, reforçam o mesmo padrão observado anteriormente: modelos mais capazes de capturar relações complexas nos dados, como a Rede Neural, mantêm desempenho consistentemente superior em todas as métricas avaliadas. A diferença entre os algoritmos fica evidente quando se observa a queda gradual nos valores de *AUC*, acurácia, *F1-score* e *MCC* à medida que se avança para métodos menos sofisticados, como *Random Forest* e, sobretudo, *SVM*. Assim, a comparação consolidada dos modelos confirma que a escolha do algoritmo exerce impacto direto na qualidade do diagnóstico automatizado, destacando a importância de selecionar técnicas alinhadas às características do problema e ao nível de precisão esperado em aplicações clínicas.

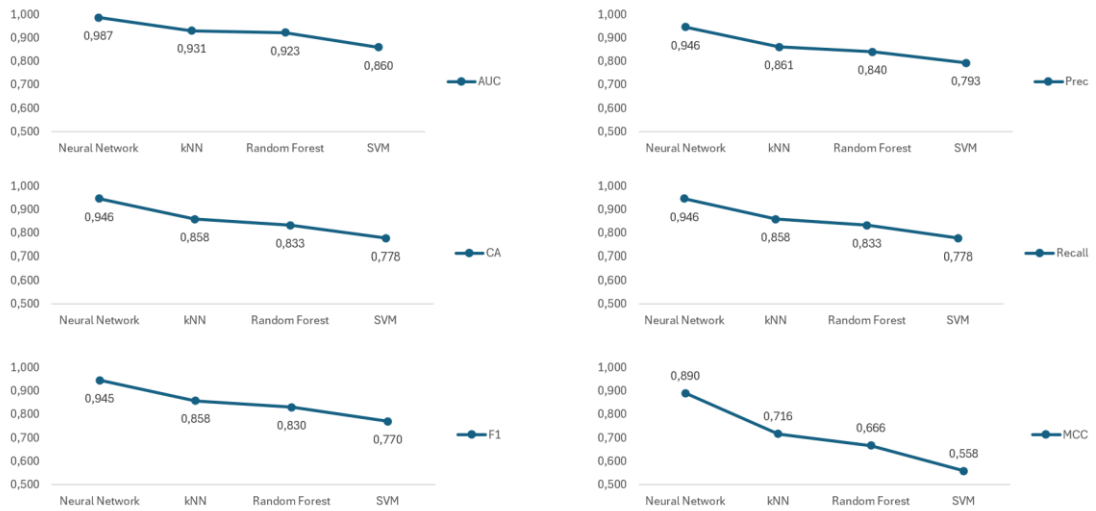
Tabela 4.10 – Indicadores de performance do classificador de imagens mamográficas, na etapa de teste, comparando a Rede Neural, *kNN*, *Random Forest* e *SVM*

Model	*AUC*	*CA*	*F1*	*Prec*	*Recall*	*MCC*
Rede Neural	0,987	0,946	0,945	0,946	0,946	0,890
kNN	0,931	0,858	0,858	0,861	0,858	0,716
Random Forest	0,923	0,833	0,830	0,840	0,833	0,666
SVM	0,860	0,778	0,770	0,793	0,778	0,558

Fonte: Elaborado pela autora

A Figura 4.25 reforça os pontos citados anteriormente de maneira gráfica.

Figura 4.25 – Indicadores de performance em forma gráfica, do classificador de imagens mamográficas, na etapa de teste, comparando a Rede Neural, *kNN*, *Random Forest* e *SVM*



Fonte: Elaborado pela autora

5. CONCLUSÃO

O presente Trabalho de Conclusão de Curso investigou a aplicação de técnicas de Inteligência Artificial (IA) na classificação de imagens de ultrassonografia mamária, com foco na distinção entre nódulos benignos e malignos. A partir do problema identificado, a persistência de limitações nos métodos tradicionais de diagnóstico, como variabilidade interpretativa, falsos positivos e falsos negativos, buscou-se compreender de que maneira algoritmos de aprendizado de máquina podem contribuir para diagnósticos médicos. A revisão bibliográfica evidenciou que, embora a IA já apresente resultados promissores em diferentes modalidades de imagem, sua integração prática ainda enfrenta desafios estruturais, operacionais e metodológicos, reforçando a relevância de estudos experimentais como este.

A contribuição central deste trabalho consistiu em expandir a análise originalmente proposta por (Dallagassa; Oliveira, 2024), incorporando a comparação entre quatro modelos de classificação: Rede Neural, *kNN*, *Random Forest* e *SVM*. Utilizando a plataforma *Orange Data Mining*, foram desenvolvidos fluxos completos de pré-processamento, extração de *embeddings*, treinamento supervisionado e validação cruzada, aplicados a um banco de 9.016 ultrassonografias mamárias. Os resultados demonstraram diferenças significativas entre os modelos: a Rede Neural apresentou o melhor desempenho, com métricas elevadas tanto no treinamento quanto no teste, seguida por *kNN*, *Random Forest* e *SVM*, que exibiram desempenhos progressivamente inferiores. A análise das matrizes de confusão evidenciou que os modelos variam principalmente na taxa de falsos negativos, que é um erro crítico em aplicações médicas, reforçando a necessidade de cautela na adoção de classificadores menos robustos.

Entretanto, apesar dos avanços observados, este estudo apresenta limitações importantes que precisam ser consideradas. A utilização de uma base pública de imagens, embora essencial para pesquisas acadêmicas, não reflete integralmente a variabilidade clínica real, podendo introduzir viés de seleção e limitar a capacidade de generalização dos modelos. Além disso, o escopo deste trabalho concentrou-se em uma classificação binária (benigno x maligno), enquanto, na prática clínica, a interpretação de ultrassonografias envolve múltiplas categorias, incertezas e gradações de risco. Outro ponto relevante é a ausência de validação clínica, uma vez que os modelos não foram testados em ambientes hospitalares, com imagens adquiridas em diferentes equipamentos, operadores e condições de exame. Esses fatores

reforçam que, embora promissora, a IA ainda enfrenta desafios significativos para sua adoção plena na rotina médica.

Por fim, destaca-se que a IA deve ser compreendida como uma ferramenta de apoio, e não como substituta do profissional de saúde. Algoritmos podem acelerar fluxos de trabalho, reduzir variabilidade interpretativa e auxiliar na identificação de padrões sutis, mas também podem gerar erros, vieses e interpretações equivocadas caso utilizados sem supervisão especializada. Assim, a contribuição deste trabalho reside não apenas na comparação ampliada de modelos, mas também na discussão crítica sobre riscos, limitações e possibilidades de uso educacional e prototípico da IA no *Orange Data Mining*. Espera-se que os resultados apresentados incentivem novas pesquisas, especialmente aquelas voltadas à validação clínica, integração multimodal e desenvolvimento de sistemas mais robustos e transparentes, capazes de apoiar de forma segura e eficaz o diagnóstico precoce do câncer de mama.

Como perspectivas para trabalhos futuros, recomenda-se a ampliação da base de dados utilizada, incorporando imagens provenientes de diferentes equipamentos, instituições e condições clínicas, de modo a reduzir vieses e aumentar a capacidade de generalização dos modelos. Também seria relevante explorar classificações multiclasses, aproximando o sistema de esquemas clínicos como o *BI-RADS*, além de investigar arquiteturas mais avançadas, como redes neurais profundas especializadas em ultrassonografia e modelos multimodais que integrem informações clínicas, textuais e de imagem. Outro caminho promissor envolve a realização de validações clínicas em ambiente hospitalar, permitindo avaliar o desempenho dos algoritmos em cenários reais e sob supervisão de especialistas. Por fim, estudos futuros podem desenvolver interfaces explicáveis e ferramentas de apoio à decisão que tornem a IA mais transparente e confiável para radiologistas, contribuindo para sua adoção segura e ética na prática médica.

REFERÊNCIAS

ALLAJBEU, Iris *et al.* Automated Breast Ultrasound: Technical Aspects, Impact on Breast Screening, and Future Perspectives. **Current Breast Cancer Reports**, v. 13, n. 3, p. 141–150, set. 2021.

ALPAYDIN, Ethem. **Introduction to Machine Learning**. 4. ed. Cambridge, MA, USA: MIT Press, 2020.

ALTMAN, N. S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. **The American Statistician**, v. 46, n. 3, p. 175–185, ago. 1992.

ALVES, Deborah *et al.* O Uso da inteligência artificial em exames de mamografia e diagnóstico do câncer de mama: uma revisão integrativa. **Revista da Faculdade de Ciências Médicas da Paraíba**, v. 3, n. 7, 5 maio 2025.

AMERICAN CANCER SOCIETY. **A Brief History of Cancer**. Disponível em: <https://www.cancer.org/cancer/understanding-cancer/history-of-cancer.html>. Acesso em: 2 dez. 2025.

AYANA, Gelan *et al.* De-Speckling Breast Cancer Ultrasound Images Using a Rotationally Invariant Block Matching Based Non-Local Means (RIBM-NLM) Method. **Diagnostics**, v. 12, n. 4, p. 862, 30 mar. 2022.

AZEVEDO, Rita; CATARINO, Patrícia; RIBEIRO, Maria Margarida. Tomossíntese e mamografia na avaliação de mulheres com elevada densidade mamária – Revisão sistemática. **ROENTGEN-Revista Científica das Técnicas Radiológicas**, v. 2, n. 1, p. 21–28, 21 jan. 2021.

BAHL, Manisha *et al.* Traditional versus modern approaches to screening mammography: a comparison of computer-assisted detection for synthetic 2D mammography versus an artificial intelligence algorithm for digital breast tomosynthesis. **Breast Cancer Research and Treatment**, v. 210, n. 3, p. 529–537, abr. 2025.

BERG, Wendie A. *et al.* Operator Dependence of Physician-performed Whole-Breast US: Lesion Detection and Characterization. **Radiology**, v. 241, n. 2, p. 355–365, nov. 2006.

BLEYER, Archie; WELCH, H. Gilbert. Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence. **New England Journal of Medicine**, v. 367, n. 21, p. 1998–2005, 22 nov. 2012.

BREIMAN, Leo. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, out. 2001.

CASEIRAS, Gisele B. *et al.* Relative cerebral blood volume measurements of low-grade gliomas predict patient outcome in a multi-institution setting. **European Journal of Radiology**, v. 73, n. 2, p. 215–220, fev. 2010.

CHAGAS, Edgar Thiago De Oliveira. Deep Learning e suas aplicações na atualidade. **Revista Científica Multidisciplinar Núcleo do Conhecimento**, v. 04, n. 05, p. 05–26, 8 maio 2019.

COMPARETTO, Ciro; BORRUTO, Franco. Cutting-edge Imaging Breakthroughs for Early Breast Cancer Detection. **Cancer Screening and Prevention**, v. 000, n. 000, p. 000–000, 30 mar. 2025.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine Learning**, v. 20, n. 3, p. 273–297, set. 1995.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, v. 13, n. 1, p. 21–27, jan. 1967.

DALLAGASSA, Marcelo Rosano; OLIVEIRA, Antonio Josenias Cordeiro de. Uso de machine learning no diagnóstico de câncer de mama através de ultrassonografia. **Journal of Health Informatics**, v. 16, n. Especial, 19 nov. 2024.

DÍAZ-URIARTE, Ramón; ALVAREZ DE ANDRÉS, Sara. Gene selection and classification of microarray data using random forest. **BMC Bioinformatics**, v. 7, n. 1, p. 3, dez. 2006.

ELMORE, J. G. Screening Mammograms by Community Radiologists: Variability in False-Positive Rates. **CancerSpectrum Knowledge Environment**, v. 94, n. 18, p. 1373–1380, 18 set. 2002.

FAWCETT, Tom. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, jun. 2006.

GONÇALVES, Juliana Garcia *et al.* Evolução histórica das políticas para o controle do câncer de mama no Brasil. 2016.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge, Mass: The MIT Press, 2016.

GRANDINI, Margherita; BAGLI, Enrico; VISANI, Giorgio. **Metrics for Multi-Class Classification: an Overview**. arXiv, , 2020. Disponível em: <https://arxiv.org/abs/2008.05756>. Acesso em: 25 jun. 2026

GUO, Rongrong *et al.* Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review. **Ultrasound in Medicine & Biology**, v. 44, n. 1, p. 37–70, jan. 2018.

GUYON, Isabelle *et al.* Gene Selection for Cancer Classification using Support Vector Machines. **Machine Learning**, v. 46, n. 1–3, p. 389–422, jan. 2002.

HASSAN, Nada M.; HAMAD, Safwat; MAHAR, Khaled. Mammogram breast cancer CAD systems for mass detection and classification: a review. **Multimedia Tools and Applications**, v. 81, n. 14, p. 20043–20075, jun. 2022.

HAYKIN, Simon (ORG.). **Kalman Filtering and Neural Networks**. 1. ed. *[S.l.]*: Wiley, 2001.

INCA. **Síntese de Resultados e Comentários**. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/estimativa/sintese-de-resultados-e-comentarios>. Acesso em: 2 dez. 2025.

INCA. **INCA lança publicação com panorama do câncer de mama no Brasil**. Disponível em: <https://www.gov.br/inca/pt-br/canais-de-atendimento/imprensa/releases/2025/inca-lanca-publicacao-com-panorama-do-cancer-de-mama-no-brasil>. Acesso em: 2 dez. 2025.

IOM/NRC–CEDBC. **Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis**. Washington (DC): National Academies Press (US), 2005.

JOHNS HOPKINS MEDICINE. **Understanding Mammogram Results | Johns Hopkins Medicine**. Disponível em: <https://www.hopkinsmedicine.org/health/conditions-and-diseases/breast-cancer/understanding-your-mammogram-report>. Acesso em: 2 dez. 2025.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255–260, 17 jul. 2015.

KAKILETI, Siva Teja *et al.* **A Density-Informed Multimodal Artificial Intelligence Framework for Improving Breast Cancer Detection Across All Breast Densities**. arXiv, , 16 out. 2025. Disponível em: <http://arxiv.org/abs/2510.14340>. Acesso em: 3 dez. 2025

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 28 maio 2015.

LEE, Christine U. *et al.* Breast ultrasound knobology and the knobology of twinkling for marker detection. **Translational Breast Cancer Research**, v. 5, p. 28–28, out. 2024.

LIAW, Andy; WIENER, Matthew. Classification and Regression by randomForest. dez. 2002.

LITJENS, Geert *et al.* A Survey on Deep Learning in Medical Image Analysis. **Medical Image Analysis**, v. 42, p. 60–88, dez. 2017.

MCKINNEY, Scott Mayer *et al.* International evaluation of an AI system for breast cancer screening. **Nature**, v. 577, n. 7788, p. 89–94, 2 jan. 2020.

MORAIS, Giovanna Ciolette *et al.* CONTROLE DE QUALIDADE DOS EXAMES DE MAMOGRAFIA: UMA REVISÃO DE LITERATURA. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 11, n. 10, p. 5285–5294, 29 out. 2025.

NEHMAT, Houssami; NEHMAT, Houssami. Overdiagnosis of breast cancer in population screening: does it make breast screening worthless? **Cancer Biology & Medicine**, v. 14, n. 1, p. 1–8, 2017.

NOGUEIRA, Keiller; PENATTI, Otávio A. B.; DOS SANTOS, Jefersson A. Towards better exploiting convolutional neural networks for remote sensing scene classification. **Pattern Recognition**, v. 61, p. 539–556, jan. 2017.

OMEGA BORO, Lal; NANDI, Gypsy. Diagnostic Performance of Deep Learning in Screened Mammogram:Systematic Review. **Current Medical Imaging Formerly Current Medical Imaging Reviews**, v. 20, p. e170423215871, 16 ago. 2023.

ONCOGUIA. **Mamografia das mamas**. Disponível em: <https://www.oncoguia.org.br/conteudo/mamografia-das-mamas/1393/264/>. Acesso em: 2 dez. 2025.

PARK, Eun Kyung *et al.* Impact of AI for Digital Breast Tomosynthesis on Breast Cancer Detection and Interpretation Time. **Radiology: Artificial Intelligence**, v. 6, n. 3, p. e230318, 1 maio 2024.

PISANO, Etta D. *et al.* Diagnostic Accuracy of Digital versus Film Mammography: Exploratory Analysis of Selected Population Subgroups in DMIST. **Radiology**, v. 246, n. 2, p. 376–383, fev. 2008.

POHLERT, T. *et al.* Integration of a detailed biogeochemical model into SWAT for improved nitrogen predictions—Model development, sensitivity, and GLUE analysis. **Ecological Modelling**, v. 203, n. 3–4, p. 215–228, maio 2007.

PORTO, Marco Antonio Teixeira; TEIXEIRA, Luiz Antonio; SILVA, Ronaldo Corrêa Ferreira da. Aspectos Históricos do Controle do Câncer de Mama no Brasil. **Revista Brasileira de Cancerologia**, v. 59, n. 3, p. 331–339, 30 set. 2013.

POWERS, David M. W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 2020.

ROSA JÚNIOR, Henrique Silveira *et al.* CORRELAÇÃO MAMOGRÁFICA E ULTRASSONOGRÁFICA NO DIAGNÓSTICO DO CÂNCER DE MAMA: UMA REVISÃO DE LITERATURA. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 11, n. 5, p. 7637–7646, 27 maio 2025.

ROSA, Kérem Gonçalves *et al.* A IMPORTÂNCIA DA CLASSIFICAÇÃO MOLECULAR NO PROGNÓSTICO DO CÂNCER DE MAMA: PERSPECTIVAS ATUAIS. **Revista Eletrônica da Faculdade de Ceres**, v. 10, n. 1, p. 46–70, 27 jul. 2021.

RSNA, RADIOLOGICAL SOCIETY OF NORTH AMERICA. **Huge Study Finds Tomosynthesis Better at Breast Cancer Detection**. Disponível em: <https://www.rsna.org/news/2023/march/study-finds-tomosynthesis-better-for-breast-cancer-detection>. Acesso em: 3 dez. 2025.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial intelligence: a modern approach**. Fourth Edition ed. Hoboken, NJ: Pearson, 2021.

SANTOS, Marceli de Oliveira *et al.* Estimativa de Incidência de Câncer no Brasil, 2023-2025. **Revista Brasileira de Cancerologia**, v. 69, n. 1, p. e-213700, 6 fev. 2023.

SCHÖLKOPF, Bernhard; SMOLA, Alexander J. **Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond**. *[S.l.]*: The MIT Press, 2001.

SILVA, Ivan Nunes da. **Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas. Fundamentos Teóricos e Aspectos Práticos**. *[S.l.]*: Artliber, 2016.

Ultrasound Breast Images for Breast Cancer. Disponível em: <https://www.kaggle.com/datasets/vuppalaadithyasairam/ultrasound-breast-images-for-breast-cancer>. Acesso em: 6 dez. 2025.

UNIVERSITY OF LJUBLJANA. **Orange Data Mining - Screenshots**. Disponível em: <https://orangedatamining.com>. Acesso em: 6 dez. 2025.

VIEGAS, Silvio Cesar. **Descubra o Orange: Análise de Dados e Machine Learning ao Alcance de Todos | LinkedIn**. Disponível em: <https://www.linkedin.com/pulse/descubra-o-orange-an%C3%A1lise-de-dados-e-machine-learning-viegas-fplof/>. Acesso em: 13 abr. 2026.

WANG, Lulu. Mammography with deep learning for breast cancer detection. **Frontiers in Oncology**, v. 14, p. 1281922, 12 fev. 2024.