

# Classification of Histopathological Images of Breast Cancer Using Convolutional Neural Networks

Bernardo Teixeira de Miranda\*, Phelipe Romano Magalhães Rosa\*, \* Vinicius do Carmo Gomes

\*Instituto de Ciências Exatas e Tecnológicas - Universidade Federal de Viçosa - Campus Rio Paranaíba (UFV-CRP)

Rodovia MG230, Km7, Caixa Postal 22 - 38.810-000 - Rio Paranaíba - MG - Brasil

Emails: {bernardo.t.miranda, phelipe.rosa, vinicius.c.gomes}@ufv.br

**Abstract**—Convolutional neural networks (CNNs) play a crucial role in early diagnosis detection, aiding healthcare professionals in decision-making. This study utilizes different CNN architectures (AlexNet, ResNet-50, and EfficientNet) to classify breast cancer histopathological images as benign or malignant, using the BreakHis dataset. The models were trained and evaluated with various magnifications and epochs, measuring the performance of each model based on metrics such as accuracy, recall, and specificity in image classification. The results showed that EfficientNet achieved an average of 98.15%, ResNet-50 reached 98.18%, and AlexNet obtained 95.47%.

**Index Terms**—Convolutional Neural Networks (CNNs), Breast Cancer Detection, Histopathological Image Classification, AlexNet, ResNet-50, EfficientNet, BreakHis, Image Classification.

## I. INTRODUÇÃO

O câncer de mama é uma doença caracterizada pela multiplicação de células anormais na mama, resultando na formação de um tumor com potencial para invadir outros órgãos. Existem várias variações deste tipo de câncer, algumas das quais podem se desenvolver mais rapidamente do que outras. Embora raro, pode afetar homens, representando 1% dos casos [1].

Segundo os dados do INCA [2], foram estimados aproximadamente 73.610 novos casos de câncer de mama em 2023, com uma margem de risco de aproximadamente 66,54 casos para cada 100 mil. O câncer de mama, sendo a principal causa de mortalidade por câncer entre as mulheres no Brasil, apresentou uma taxa de mortalidade ajustada por idade de 11,71 por 100 mil habitantes, resultando em 18.139 óbitos em 2021. As regiões Sul e Sudeste do país registram as maiores taxas de incidência e mortalidade dessa doença.

As estratégias utilizadas para a detecção precoce do câncer de mama incluem diagnóstico e rastreamento. O diagnóstico envolve abordagem a pessoas com sinais/sintomas, enquanto rastreamento visa identificar alterações em populações sem sintomas. No Brasil, as Diretrizes para a Detecção Precoce do câncer de mama, são baseadas em revisão sistemática que orientam profissionais de saúde [3].

A aplicação da tecnologia na medicina tem revelado avanços significativos, especialmente no campo do câncer de mama. Novas ferramentas e técnicas estão sendo constantemente desenvolvidas. Na área da visão computacional, surgem oportunidades para explorar estudos, análises e processamento de características em imagens [4]. Como exemplificado por

Tan et al. [5], que apresenta um estudo de Redes Neurais Convolucionais (CNNs) utilizando Aprendizado Federado (FL) para a classificação de imagens de câncer de mama, e também pelo trabalho de trabalho de Macaço e Neagoe [6], que propõe um novo modelo para a detecção de câncer de mama por meio da termografia e Redes Neurais Convolucionais (CNNs).

Este trabalho propõe a identificação e classificação do câncer de mama utilizando imagens histopatológicas, com o apoio de redes neurais convolucionais. O objetivo principal é Desenvolver um projeto completo de classificação de imagens em um comparativo entre três modelos de CNNs. O projeto incluirá o treinamento dos modelos e a validação do desempenho, utilizando métricas apropriadas e uma matriz de confusão para análise dos resultados.

## II. TRABALHOS RELACIONADOS

No estudo de Parvin e Hasan [7], foi realizada uma análise comparativa de diferentes arquiteturas de redes neurais convolucionais (CNN), incluindo LeNet-5, AlexNet, VGG-16, ResNet-50 e Inception-v1, aplicadas à classificação de imagens histopatológicas para a detecção de câncer de mama. A pesquisa utilizou o conjunto de dados BreakHis, que contém imagens de biópsias microscópicas de tumores benignos e malignos em várias ampliações. O desempenho dos modelos foi avaliado com base em métricas como precisão, AUC, *recall* e F1-score. Todas as imagens foram redimensionadas para 224x224 *pixels*, com um tamanho de lote de 32 e um *momentum* de 0,9. O treinamento foi realizado por 300 épocas, com uma taxa de aprendizado inicial de 0,01 e o uso de estratégias de aumento de dados para prevenir overfitting. O conjunto de dados foi dividido em 80% para treinamento e 20% para teste. Os resultados mostraram que a arquitetura Inception-v1 obteve o melhor desempenho, alcançando as maiores acurácias de 89%, 92%, 94% e 90%, respectivamente, para os diferentes fatores de ampliação (40X, 100X, 200X e 400X).

Melekoodappattu et al. [8] realizaram uma modificação na arquitetura e nas especificações da rede, para obter um melhor desempenho no diagnóstico de malignidade no tecido mamário. Aplicando o modelo CNN de nove camadas para o diagnóstico de câncer de mama, obteve uma precisão de 97,8% e 98% utilizando o *dataset* MIAS e uma precisão de 98,3% e 97,9% utilizando o *dataset* DDSM. O estudo

previamente mencionado busca aprimorar uma arquitetura e especificações proposta no trabalho, o trabalho atual tem como o foco principal em realizar uma avaliação comparativa das CNNs existentes, com um objetivo de identificar quais delas possuem um desempenho melhor entre as demais.

Com base nos estudos de Barbosa. et al. [9], foram implementadas seis redes neurais convolucionais (CNNs) com o uso de aumento de dados nos conjuntos de imagens de câncer de mama (BreakHis e Biglycan). O objetivo foi medir o impacto individual e combinado da otimização e da taxa de aprendizado na precisão de treinamento dos modelos de classificação de imagens de câncer de mama. Foram conduzidos experimentos para avaliar o impacto desses parâmetros, tanto individualmente quanto em seus níveis específicos (0,001 e 0,0001), na precisão de validação. No conjunto de dados BrakHis, o modelo EfficientNet alcançou uma acurácia de 98,86%, enquanto o modelo ShuffleNet obteve uma acurácia de 97,06% no conjunto de imagens do Biglycan, ambos utilizando o otimizador Adam. Por meio dessa abordagem, foi possível analisar o impacto dos fatores de otimização e do *learning rate* na acurácia dos modelos de previsão, com o auxílio de um método parcial.

De acordo com a pesquisa de Tan et al. [5], foi demonstrada a integração das Redes Neurais Convolucionais (CNNs) com Aprendizado Federado (FL) para abordar a classificação de imagens de câncer de mama, envolvendo a extração de características de múltiplos ambientes participantes em vez de um centro de aprendizagem centralizado. O trabalho tem como foco principal na análise do desempenho de *recall* do que na *accuracy* devido aos riscos associados aos falsos negativos. A base de dados utilizada no estudo foi a Digital Database for Screening Mammography (DDSM) e o Curated Breast Imaging Subset of DDSM (CBIS-DDSM), os dados foram pré-processados e convertidos em imagens de 299 x 299 pixels, com Regiões de Interesse (ROIs) extraídas das imagens. Foram armazenados como um arquivo tfrecords para TensorFlow, contendo 55.890 exemplos de treinamento, sendo a maioria negativos e uma minoria positivos. Essa abordagem resultou em uma melhoria significativa, alcançando uma classificação multiclasse com um *recall* de 100% e uma pontuação de 99.804% na Área Sob a Curva (AUC). Estes resultados foram obtidos ao utilizar um modelo MobileNet pré-treinado para a extração de características.

### III. METODOLOGIA

Neste capítulo, será apresentado a metodologia para a realização deste estudo, utilizando dados do *DataSet* BreakHis [10]. Serão aplicadas três Redes Neurais Convolucionais distintas (CNNs), sendo elas a AlexNet, ResNet-50 e a EfficientNet utilizando como ferramenta a biblioteca *PyTorch* para a implementação e execução dessas redes. A metodologia seguirá o fluxo do diagrama geral ilustrado na Figura 1.

#### A. Conjuntos de Imagens

1) *BreakHis*: O *dataset* BreakHis [10] é composto por 9.109 imagens microscópicas de tecido tumoral de mama

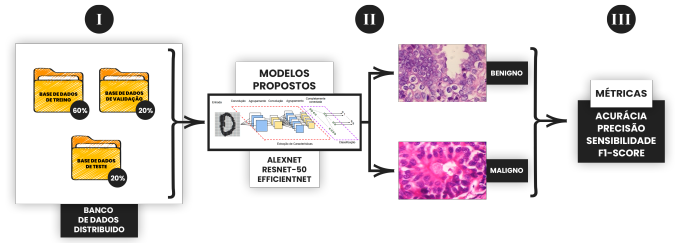


Fig. 1. Diagrama Geral da metodologia adaptada de [10] [11]

coletadas de 82 pacientes usando diferentes fatores de ampliação (40X, 100X, 200X e 400X). Nelas contém 2.480 amostras benignas e 5.429 malignas com uma resolução de 700x460 *pixels* com um RGB de 3 canais, profundidade de 8 bits em cada canal, do formato PNG. A Figura 2 mostra um exemplo de lâmina de tumor maligno de mama nas diferentes ampliações.

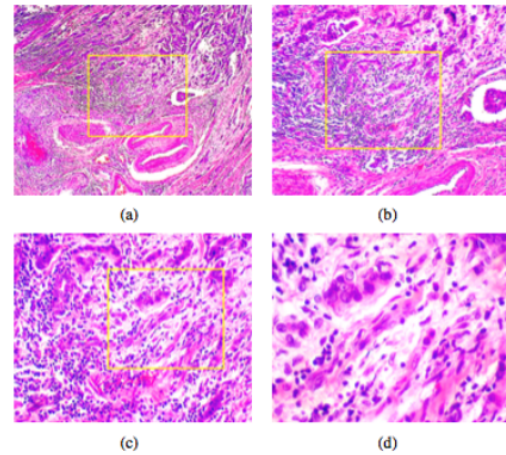


Fig. 2. Exemplo de Lâmina de tumor maligno de mama em diferentes tipos de ampliação disponíveis: (a) 40x, (b) 100x, (c) 200x e (d) 400x [10].

#### B. Coleta e Preparação de Dados

Conforme a Figura 1 (I), esta etapa envolve a coleta e organização dos dados da base de dados BreakHis [10], que são essenciais para o treinamento e validação das Redes Neurais Convolucionais (CNNs). Primeiramente, foi feita a coleta das imagens, que foram cuidadosamente organizadas em duas categorias: Benignas e Malignas. Em seguida, foi realizada uma pré-distribuição das imagens, garantindo a mesma quantidade de amostras para ambas as categorias. Dessa forma, foi assegurado que cada categoria possuísse uma distribuição equilibrada de imagens.

Após essa etapa de organização e pré-distribuição, os dados foram divididos para compor os conjuntos de treinamento, teste e validação, sendo 60% do total das imagens destinado ao treinamento, 20% ao teste e 20% à validação.

Nesta etapa, conforme detalhado na Figura 1 (II), será implementada três Redes Neurais Convolucionais distintas (CNNs), sendo elas a AlexNet, ResNet-50 e a EfficientNet com a utilização da biblioteca *PyTorch*. Cada CNN será carregada com pesos pré-treinados no dataset ImageNet, o que oferece uma base inicial de aprendizado para as redes. Em seguida, serão treinadas e avaliadas com foco na capacidade do modelo de classificar as imagens como benignas ou malignas.

Para a execução dos treinamentos, utilizou-se um computador com as seguintes especificações técnicas: uma GPU Nvidia GeForce RTX 4060 Ti com 4GB de memória dedicada, um processador Intel Core i5-4430, e 32 GB de memória RAM DDR3 com frequência de 800 MHz. Assim foram definidos parâmetros básicos no código, incluindo o tamanho do lote (*batch size*) de 32 treinamento e 10 para validação e teste, uma *SEED* de 42, a taxa de aprendizado (*learning rate*) de 0,001 e o momento (*momentum*) de 0,9 em 100 épocas. Além disso, o critério de perda escolhido foi a função *CrossEntropyLoss*, e o otimizador utilizado foi o *SGD* da biblioteca *PyTorch*.

Durante a pesquisa, foi realizado vários testes para ajustar o tamanho do lote e a taxa de aprendizado, até encontrar os valores mais adequados para alcançar um melhor desempenho. Esses parâmetros foram aplicados da mesma maneira em todos os modelos durante o treinamento, teste e validação, garantindo uma comparação justa dos resultados.

#### D. Métricas

Na Figura 1 (III), foram utilizados métricas importantes, como acurácia, precisão, sensibilidade e F1-score, permitindo a identificação do modelo mais eficiente na tarefa de classificação das imagens em benignas e malignas, assim garantindo a precisão na detecção do câncer de mama.

Obtemos a acurácia (ACC), precisão (PP), sensibilidade (SE) e F1-Score (F1), que são obtidas pelas Equações 1 a 4 conforme os livros de BALAS et al. [12], Wani et al. [13] e do trabalho de Mariano e Diego [14]. Essas métricas são necessárias para avaliar a eficácia dos modelos na classificação de imagens em categorias específicas, como benignas e malignas. Para calcular essas métricas, utilizamos as seguintes definições: VP representa o número de verdadeiros positivos, FP o número de falsos positivos, VN o número de verdadeiros negativos e FN o número de falsos negativos.

$$ACC = \frac{(VP + VN)}{(VP + VN + FP + FN)} \quad (1)$$

$$PP = \frac{VP}{(VP + FP)} \quad (2)$$

$$SE = \frac{VP}{(VP + FN)} \quad (3)$$

$$F1 = 2 \cdot \frac{(PP \cdot SE)}{(PP + SE)} \quad (4)$$

## IV. RESULTADOS E DISCUSSÃO

Inicialmente, utilizamos a base de dados padrão já separada, sem a aplicação de filtros citadas anteriormente na Seção 4.3. Foram realizados testes em 100 épocas utilizando os modelos AlexNet, Resnet-50 e EfficientNet, em diferentes ampliações (40x, 100x, 200x e 400x) que foram analisadas de forma conjunta gerando os resultados apresentados na I.

TABLE I  
RESULTADOS DOS TESTES OBTIDOS POR TODOS OS MODELOS EM 100 ÉPOCAS.

Modelo	Classe	ACC	Precisão	SE	F1-Score
AlexNet	Benigna	93,75%	93,23%	94,35%	93,79%
	Maligna		94,29%	93,15%	93,71%
ResNet-50	Benigna	95,36%	96,30%	94,35%	95,32%
	Maligna		94,47%	96,37%	95,41%
EfficientNet	Benigna	94,96%	96,46%	93,35%	94,88%
	Maligna		93,55%	96,57%	95,04%

Analisando os resultados apresentados na Tabela I, nota-se que o modelo ResNet-50 alcançou um melhor desempenho geral, destacando-se com 95,36% de acurácia para a classe benigna e 94,47% para a classe maligna. Esse desempenho foi corroborado pela matriz de confusão (Figura 5), que evidencia uma quantidade reduzida de falsos positivos e falsos negativos, reforçando sua capacidade de generalização. Além disso, as curvas de erro e acurácia do modelo (Figura 6) demonstraram uma boa estabilidade, indicando uma rápida convergência e baixo sobreajuste.

Por outro lado, o modelo AlexNet obteve desempenho inferior em relação aos outros modelos, com 93,75% de acurácia para a classe benigna e 94,29% para a classe maligna. A matriz de confusão do AlexNet (Figura 3) indica uma maior ocorrência de falsos negativos, especialmente na classe maligna, o que impacta sua sensibilidade. Ademais, as curvas de erro e acurácia (Figura 4) revelam uma maior discrepância entre os dados de treinamento e validação, indicando uma capacidade limitada de generalização.

O modelo EfficientNet também apresentou resultados competitivos, com uma acurácia de 94,96% para a classe benigna e 93,55% para a classe maligna. Embora sua matriz de confusão (Figura 7) revele uma quantidade de falsos negativos ainda considerada boa, suas curvas de erro e acurácia (Figura 8) mostram uma ligeira instabilidade durante o treinamento, sugerindo a necessidade de ajustes adicionais nos hiperparâmetros.

De forma geral, o modelo Resnet-50 mostrou-se o mais robusto, equilibrando altas taxas de acurácia, precisão e sensibilidade, enquanto o EfficientNet demonstrou um bom potencial para aplicações que demandam especificidade elevada. Em contrapartida, o AlexNet, apesar de seu desempenho aceitável, apresenta limitações que o tornam mais adequado para cenários com menor complexidade nos dados.

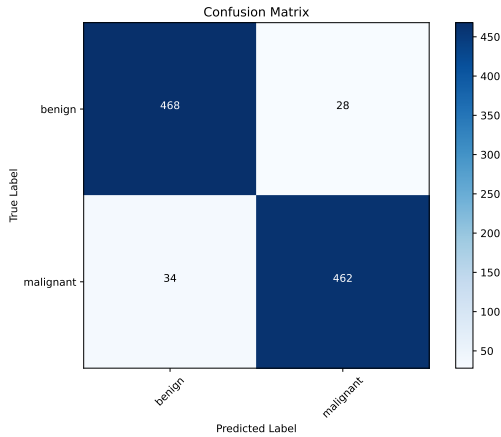


Fig. 3. Matriz de Confusão do modelo AlexNet

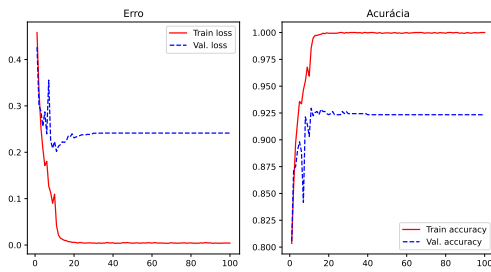


Fig. 4. Erro e Acurácia do modelo AlexNet

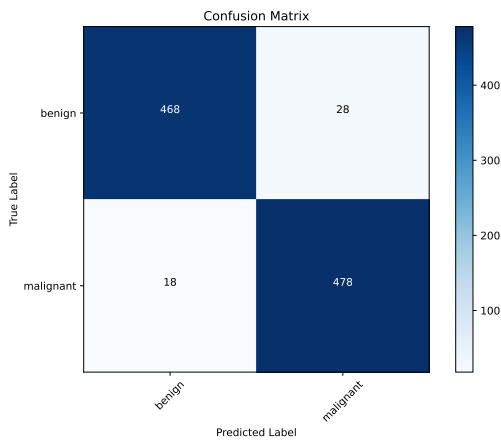


Fig. 5. Matriz de Confusão do modelo ResNet-50

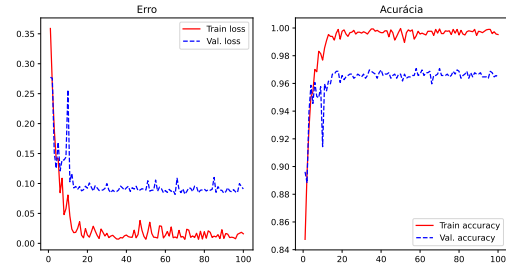


Fig. 6. Erro e Acurácia do modelo ResNet-50

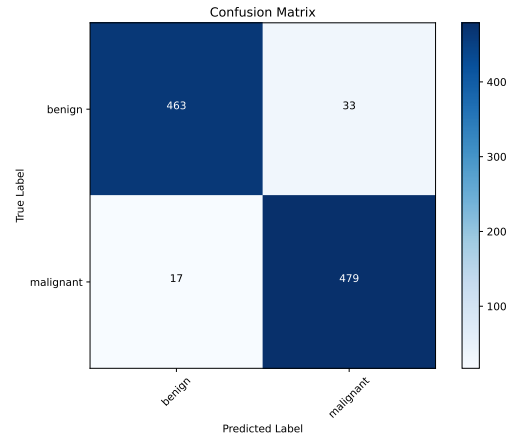


Fig. 7. Matriz de Confusão do modelo EfficientNet

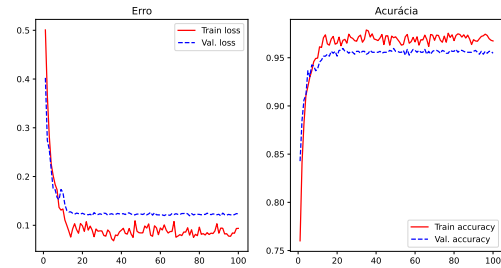


Fig. 8. Erro e Acurácia do modelo EfficientNet

## V. CONCLUSÃO

O modelo ResNet-50 destacou-se como a melhor opção, combinando alta acurácia e robustez para ambas as classes analisadas. O EfficientNet apresentou desempenho competitivo, com alta especificidade e F1-Score, sendo adequado para aplicações exigentes. Por outro lado, o AlexNet, apesar de aceitável, mostrou limitações em cenários mais complexos. Esses resultados ressaltam a importância de selecionar o modelo conforme os requisitos da aplicação, priorizando equilíbrio entre desempenho e generalização.

## REFERENCES

- [1] “Câncer de mama,” Instituto Nacional de Câncer - INCA, 2023, acessado em 15 de julho de 2024. [Online]. Available: <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/mama>
- [2] “Outubro rosa 2023,” Instituto Nacional de Câncer - INCA, 2023, acessado em 15 de julho de 2024. [Online]. Available: <https://www.gov.br/inca/pt-br/assuntos/campanhas/2023/outubro-rosa>
- [3] “Detecção precoce,” Instituto Nacional de Câncer - INCA, 2023, acessado em 15 de julho de 2024. [Online]. Available: <https://www.gov.br/inca/pt-br/assuntos/gestor-e-profissional-de-saude/controlado-cancer-de-mama/acoes/deteccao-precoce>
- [4] P. C. R. Boasquevisque, “Classificação do grau histológico de malignidade em câncer de mama utilizando programas de análise automatizada de imagens e aprendizado de máquina,” Ph.D. dissertation, brasil, 2020.
- [5] Y. N. Tan, V. P. Tinh, P. D. Lam, N. H. Nam, and T. A. Khoa, “A transfer learning approach to breast cancer classification in a federated learning framework,” *IEEE Access*, vol. 11, pp. 27 462–27 476, 2023.
- [6] I.-R. Macaço and V.-E. Neagoe, “Breast cancer detection using thermal infrared image analysis based on dempster-shafer decision fusion of cnn classifiers,” in *2023 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2023, pp. 01–06.
- [7] F. Parvin and M. A. M. Hasan, “A comparative study of different types of convolutional neural networks for breast cancer histopathological image classification,” in *2020 IEEE Region 10 Symposium (TENSYP)*. IEEE, 2020, pp. 945–948.
- [8] J. G. Melekoodappattu, A. S. Dhas, B. K. Kandathil, and K. Adarsh, “Breast cancer detection in mammogram: Combining modified cnn and texture feature based approach,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 9, pp. 11 397–11 406, 2023.
- [9] G. Barbosa, L. Moreira, P. Moises de Sousa, R. Moreira, and A. Backes, “Optimization and learning rate influence on breast cancer image classification,” in *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP, INSTICC*. SciTePress, 2024, pp. 792–799.
- [10] “Breast cancer histopathological database (brekhis),” BreakHis, 2023, acesso em 23 de outubro de 2023. [Online]. Available: <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/>
- [11] A. C. G. Vargas, A. Paes, and C. N. Vasconcelos, “Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres,” in *Proceedings of the xxix conference on graphics, patterns and images*, vol. 1, no. 4. sn, 2016.
- [12] V. E. Balas, S. S. Roy, D. Sharma, and P. Samui, Eds., *Handbook of Deep Learning Applications*, 1st ed., ser. Smart Innovation, Systems and Technologies. Springer Cham, 2019. [Online]. Available: <https://doi.org/10.1007/978-3-030-11479-4>
- [13] M. A. Wani, F. A. Bhat, S. Afzal, and A. I. Khan, *Advances in Deep Learning*, 1st ed., ser. Studies in Big Data. Springer Singapore, 2020. [Online]. Available: <https://doi.org/10.1007/978-981-13-6794-6>
- [14] D. Mariano, “Métricas de avaliação em machine learning: acurácia, sensibilidade, precisão, especificidade e f-score,” *Alfahelix*, 2021. [Online]. Available: [https://www.academia.edu/50221775/Mtricas\\_de\\_avaliao\\_em\\_machine\\_learning\\_acurcia\\_sensibilidade\\_preciso\\_especificidade\\_e\\_F\\_score](https://www.academia.edu/50221775/Mtricas_de_avaliao_em_machine_learning_acurcia_sensibilidade_preciso_especificidade_e_F_score)