



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA BIOMÉDICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA

ANDERSON FELIX DA SILVA

**SISTEMA INTELIGENTE PARA O APOIO AO DIAGNÓSTICO DO CÂNCER DE MAMA
USANDO IMAGENS TERMOGRÁFICAS E REDES NEURAIS ARTIFICIAIS
PROFUNDAS**

Recife

2022

ANDERSON FELIX DA SILVA

**SISTEMA INTELIGENTE PARA O APOIO AO DIAGNÓSTICO DO CÂNCER DE MAMA
USANDO IMAGENS TERMOGRÁFICAS E REDES NEURAIS ARTIFICIAIS
PROFUNDAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para a obtenção do título de Mestre em Engenharia Biomédica. Área de concentração: Computação Biomédica.

Orientador: Prof. Dr. Wellington Pinheiro dos Santos

Recife

2022

ANDERSON FELIX DA SILVA

**SISTEMA INTELIGENTE PARA O APOIO AO DIAGNÓSTICO DO CÂNCER DE
MAMA USANDO IMAGENS TERMOGRÁFICAS E REDES NEURAIS ARTIFICIAIS
PROFUNDAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para a obtenção do título de Mestre em Engenharia Biomédica. Área de concentração: Computação Biomédica.

Aprovada em: 24 / 02 / 2022.

BANCA EXAMINADORA

Prof. Dr. Wellington Pinheiro dos Santos (Orientador)
Universidade Federal de Pernambuco

Prof. Dr. Ricardo Emmanuel de Souza (Examinador Interno)
Universidade Federal de Pernambuco

Prof^a. Dr^a. Rita de Cássia Fernandes de Lima (Examinadora Externa)
Universidade Federal de Pernambuco

Dedico esse trabalho à minha família, que sempre foi meu apoio e minha fortaleza. Aos meus pais, que sempre buscaram me dar as melhores condições para que eu estudasse. Aos meus avós, em especial a minha avó Josefa Correia (in memoriam), que teve uma contribuição importante para a minha formação. Dedico à minha companheira de vida, Camila Bernardo, que sempre me incentivou e contribuiu para que eu fosse uma pessoa melhor. Dedico também a todos aqueles que lutaram e lutam por uma educação pública de qualidades para todos.

AGRADECIMENTOS

Gostaria de agradecer ao meu orientador, o Prof. Dr. Wellington Pinheiro dos Santos, pela confiança, por sua atenção e generosidade, sou grato pela oportunidade de trabalhar e aprender com esse ser humano tão admirável.

Aos componentes da banca examinadora, Prof. Dr. Ricardo Emmanuel de Souza e à Prof^a. Dr^a. Rita de Cássia Fernandes de Lima, que aceitaram o convite em participar da banca.

Aos professores e servidores do Departamento de Engenharia Biomédica, por todo conhecimento compartilhado, por todo apoio e disponibilidade.

Aos meus pais, Angela e Adenilson, que sempre buscaram as melhores condições para que eu conseguisse estudar. Agradeço por terem me incentivado e me apoiado nos momentos difíceis.

À minha irmã, Andresa, pela companhia e apoio durante essa jornada.

Agradeço a minha companheira de vida, Camila Bernardo, por seu apoio, incentivo e orientação em diversos momentos de minha vida. Camila foi uma pessoa fundamental para eu viesse a ser quem sou, como profissional e ser humano.

Aos meus amigos de trabalho, da STI/UFPE, por todo apoio, incentivo e compreensão.

Aos amigos do grupo de pesquisa, que me ajudaram diretamente ou indiretamente nesse trabalho, Valter, Rian, Thifany e Pedro.

Agradeço a todos aqueles que lutam ou um dia lutaram para que o acesso à educação pública de qualidade fosse para todos. O acesso ao nosso bem mais precioso, o conhecimento, é um fator importantíssimo na luta para diminuir a desigualdade social.

RESUMO

O câncer de mama é a forma mais mortal de câncer entre mulheres, tanto em países desenvolvidos quanto em países subdesenvolvidos e em desenvolvimento. A mortalidade do câncer de mama está ligada diretamente a estratégias de prevenção da doença, como campanhas educativas e tecnologias de apoio ao diagnóstico precoce do câncer. A técnica mais utilizada no apoio ao diagnóstico do câncer de mama por imagem é a mamografia por Raios-X. No entanto, a mamografia tem suas desvantagens, como o custo, o uso de raios ionizantes (que podem estar relacionados a fatores causadores de câncer), além do desconforto na obtenção da imagem por meio da compressão da mama. Uma técnica complementar à mamografia, e que vem sendo testada, é a termografia de mama, a qual baseia-se nas mudanças metabólicas resultantes do surgimento de células alteradas no tecido mamário, que resultam em modificações da distribuição de temperatura. A termografia vem sendo proposta como técnica complementar à mamografia, sendo mais vantajosa quando comparada ao toque da mama e que pode servir como sistema de triagem, permitindo a detecção precoce de lesões da mama e diminuindo a mortalidade. O presente trabalho tem como objetivo analisar o uso de redes neurais profundas juntamente com diferentes técnicas de classificação, para o reconhecimento de lesões em imagens termográficas utilizando o *software* de aprendizado de máquina Weka. Pretende também, fundamentar um modelo que possa ser explorado em aplicações de apoio ao diagnóstico do câncer de mama para classificação de lesões em imagens de termografia. Inicialmente, na etapa de extração de atributos, foram utilizadas diferentes redes profundas da biblioteca DeepLearning4j do Weka: *LeNet*, *ResNet50*, *NASNetMobile*, *SqueezeNet* e *Inception v3*. Em seguida, foi realizada a seleção dos melhores atributos utilizando PSO. Em seguida, na etapa de classificação e treinamento, foram gerados 30 experimentos para os seguintes classificadores: *Naive Bayes*, *Bayes Net*, *Random Tree*, Árvore de Decisão J48, *Random Forest*, Máquina de Vetor de Suporte *SVM* e Rede Perceptron Multicamadas *MLP*. Os resultados foram comparados utilizando gráficos *boxplots* e tabelas para as métricas de Acurácia, Índice Kappa, Sensibilidade, Especificidade, Área Sob a Curva *ROC* e Tempo de Treinamento (ms). Por fim, foram analisados os melhores desempenhos entre as redes neurais profundas e os classificadores utilizados. Além disso, também foi analisado o desempenho antes e após a seleção de atributos, com o objetivo de determinar o modelo mais eficiente a ser utilizado. Como resultados, a rede profunda *Inception V3* combinada com o classificador *SVM* com kernel polinomial de 3º teve a maior taxa de acurácia para a

abordagem sem a seleção de atributos, obtendo 79,92%. Porém, utilizando a seleção de atributos, a CNN *Inception V3* combinada com o classificador SVM com kernel polinomial de 4º, obteve uma acurácia de 78,55% com um tempo de treinamento duas vezes menor. Portanto, a seleção de atributos pode ser realizada sem perder muito em acurácia quando comparada ao conjunto completo de atributos, colaborando assim, para aplicações que visem otimizar o tempo de classificação e o custo computacional. Diante disso, o uso de modelos de redes neurais profundas e classificadores para auxiliar no diagnóstico do câncer de mama usando imagens termográficas, mostrou-se como uma ferramenta interessante a ser explorada em futuras aplicações e que pode contribuir para o diagnóstico precoce dessa doença.

Palavras-chaves: diagnóstico do câncer de mama; extração de atributos; seleção de atributos; termografia da mama; redes neurais profundas.

ABSTRACT

Breast cancer is the deadliest form of cancer among women, both in developed and in underdeveloped and developing countries. The mortality of breast cancer is directly linked to strategies for prevention of disease, such as educational campaigns and technologies to support early diagnosis of cancer. The most used technique to support the diagnosis of breast cancer by imaging is X-ray mammography. However, mammography has its disadvantages, such as cost, the use of ionizing radiation (which may be related to factors that cause cancer), in addition to the discomfort in obtaining the image through breast compression. A complementary technique to mammography that has been tested is breast thermography, which is based on metabolic changes resulting from the appearance of altered cells in the breast tissue, resulting in changes in temperature distribution. Thermography has established itself as a complementary technique to mammography, being more advantageous when compared to breast touch, and can serve as a screening system, allowing early detection of breast lesions and reducing mortality. The present work aims to analyze the use of deep neural networks together with different classification techniques, for the recognition of lesions in thermographic images using the machine learning software Weka. It also aims to develop a model that can be explored in applications to support the diagnosis of breast cancer for the classification of lesions in thermographic images. Initially, in the attribute extraction stage, different deep networks from the Weka DeepLearning4j library were used: LeNet, ResNet50, NASNetMobile, SqueezeNet and Inception V3. Soon after, the selection of the best attributes was performed using PSO. Then, in the classification and training stage, 30 experiments were generated for the following classifiers: Naive Bayes, Bayes Net, Random Tree, J48 Decision Tree, Random Forest, Support Vector Machine (SVM) and Multilayer Perceptron (MLP). The results were compared using boxplots and tables for the metrics of Accuracy, Kappa Index, Sensitivity, Specificity, Area Under the ROC Curve and Training Time (ms). Finally, the best performances between the deep neural networks and the classifiers used were analyzed. In addition, the performance before and after attribute selection was also analyzed, in order to determine the most efficient model to use. As results, the deep network *Inception V3* combined with the SVM classifier with a polynomial kernel of 3^o had the highest accuracy rate for the approach without attribute selection, obtaining 79.92%. However, using attribute selection, the CNN *Inception V3* combined with the SVM classifier with a 4^o polynomial kernel obtained an accuracy of 78.55% with twice as little training time. Therefore, the attribute selection can be performed without losing much in accuracy

when compared to the complete set of attributes, thus contributing to applications that aim to optimize classification time and computational cost. Given this, the use of deep neural network models and classifiers to aid in the diagnosis of breast cancer using thermographic images, has been shown to be an interesting tool that can be explored in future applications and that can contribute to the early diagnosis of this disease.

Keywords: diagnosis of breast cancer; attribute extraction; attribute selection; thermography; deep neural networks.

LISTA DE FIGURAS

Figura 1 – Como fazer o autoexame das mamas.	32
Figura 2 – Exame de mamografia.	33
Figura 3 – Ultrassonografia da mama.	34
Figura 4 – Exemplo de imagens de termografia da mama para as classes: Cisto, Lesão Benigna, Lesão Maligna e Sem Lesão.	37
Figura 5 – Exemplo do modelo de neurônio artificial proposto por McCulloch-Pits. . .	39
Figura 6 – Arquitetura de Rede Neural Artificial de Camada Única.	40
Figura 7 – Arquitetura de Rede Neural Artificial de Camada Múltipla.	40
Figura 8 – Arquitetura de Redes Recorrentes.	40
Figura 9 – Exemplo de classificação de imagem usando rede neural profunda. . . .	41
Figura 10 – Arquitetura da Rede Convolucional LeNet.	42
Figura 11 – Bloco de construção da aprendizagem residual.	43
Figura 12 – Organização dos filtros de convolução do <i>Fire Module</i> da <i>CNN SqueezeNet</i> . .	45
Figura 13 – Fluxograma relacionado a execução do PSO.	47
Figura 14 – Exemplo de MLP com duas camadas escondidas.	52
Figura 15 – Exemplo de fluxograma com as etapas seguidas para construção de sistemas auxiliados por computador.	53
Figura 16 – Diagrama de bloco do método proposto.	55
Figura 17 – Exemplos de imagens termográficas obtidas em diferentes posições. . .	57
Figura 18 – Parte do arquivo ".arff" carregado na entrada dos dados, onde é indicado as classes existentes e os caminhos para cada imagem.	58
Figura 19 – Validação cruzada para <i>10-fold</i> . A base de dados é aleatoriamente dividida em dez conjuntos diferentes, cada um com aproximadamente 10% dos dados. O modelo é treinado sobre o conjunto de treinamento e em seguida ele é aplicado ao conjunto de validação (BERRAR, 2018) . . .	63
Figura 20 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a <i>CNN LeNet</i>	67
Figura 21 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a <i>CNN LeNet</i>	68
Figura 22 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a <i>CNN NASNetMobile</i>	70

Figura 23 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN <i>NASNetMobile</i>	71
Figura 24 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN <i>ResNet50</i>	73
Figura 25 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN <i>ResNet50</i>	73
Figura 26 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN <i>SqueezeNet</i>	74
Figura 27 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN <i>SqueezeNet</i>	75
Figura 28 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN <i>Inception v3</i>	77
Figura 29 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN <i>Inception v3</i>	78
Figura 30 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN <i>LeNet</i> e selecionados com PSO	81
Figura 31 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN <i>LeNet</i> e selecionados com PSO	82
Figura 32 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN <i>NASNetMobile</i> e selecionados usando PSO.	83
Figura 33 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN <i>NASNetMobile</i> e selecionados usando PSO.	84
Figura 34 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN <i>ResNet50</i> e selecionados usando PSO.	85
Figura 35 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN <i>ResNet50</i> e selecionados usando a técnica PSO.	86
Figura 36 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN <i>SqueezeNet</i> e selecionados usando PSO.	88
Figura 37 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN <i>SqueezeNet</i> e selecionados usando PSO.	88
Figura 38 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN <i>Inception v3</i> e selecionados usando PSO.	90

Figura 39 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *Inception v3* e selecionados usando PSO. 91

LISTA DE TABELAS

Tabela 1 – Distribuição das amostras entre as classes definidas.	57
Tabela 2 – Modelos de redes neurais profundas utilizados e o número de atributos extraídos por cada rede.	60
Tabela 3 – Números de atributos antes e depois da seleção de atributos por cada modelo de rede neural profunda.	61
Tabela 4 – Intervalos de valores para a métrica de índice kappa e sua classificação de desempenho correspondente.	66
Tabela 5 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>LeNet</i> . São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.	68
Tabela 6 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>LeNet</i> . São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.	69
Tabela 7 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>NASNetMobile</i> . São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.	70
Tabela 8 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>NASNetMobile</i> . São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.	71
Tabela 9 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>ResNet50</i> . São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.	72
Tabela 10 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>ResNet50</i> . São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.	74

Tabela 11 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>SqueezeNet</i> . São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.	75
Tabela 12 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>SqueezeNet</i> . São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.	76
Tabela 13 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional InceptionV3. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.	77
Tabela 14 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>Inception v3</i> . São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.	78
Tabela 15 – Comparação entre a taxa de acurácia dos melhores classificadores para cada Rede Neural Convolutacional.	79
Tabela 16 – Comparação entre a taxa de índice kappa dos melhores classificadores para cada Rede Neural Convolutacional.	79
Tabela 17 – Comparação entre a taxa de sensibilidade dos melhores classificadores para cada Rede Neural Convolutacional.	79
Tabela 18 – Comparação entre a taxa de especificidade dos melhores classificadores para cada Rede Neural Convolutacional.	80
Tabela 19 – Comparação entre a taxa de área sob curva ROC dos melhores classificadores para cada Rede Neural Convolutacional.	80
Tabela 20 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>LeNet</i> e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade. . .	81

Tabela 21 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional LeNet e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas Área Sob Curva ROC e Tempo de Treinamento.	83
Tabela 22 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>NASNetMobile</i> e selecionados com PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.	84
Tabela 23 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>NASNetMobile</i> e selecionados com PSO. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.	85
Tabela 24 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>ResNet50</i> e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.	86
Tabela 25 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>ResNet50</i> e selecionados usando PSO. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.	87
Tabela 26 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>SqueezeNet</i> e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.	89
Tabela 27 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional <i>SqueezeNet</i> e selecionados usando PSO. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.	90

Tabela 28 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional InceptionV3 e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.	91
Tabela 29 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional InceptionV3 e selecionados usando PSO. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.	92
Tabela 30 – Comparação entre a taxa de acurácia dos melhores classificadores para cada Rede Neural Convolutacional com PSO.	93
Tabela 31 – Comparação entre a taxa de índice kappa dos melhores classificadores para cada Rede Neural Convolutacional com PSO.	93
Tabela 32 – Comparação entre a taxa de sensibilidade dos melhores classificadores para cada Rede Neural Convolutacional.	93
Tabela 33 – Comparação entre a taxa de especificidade dos melhores classificadores para cada Rede Neural Convolutacional com PSO.	93
Tabela 34 – Comparação entre a taxa de área sob curva ROC dos melhores classificadores para cada Rede Neural Convolutacional com PSO.	94
Tabela 35 – Comparação entre a Acurácia antes e depois da seleção de atributos usando PSO.	94
Tabela 36 – Comparação entre o índice kappa antes e depois da seleção de atributos usando PSO.	94
Tabela 37 – Comparação entre a Sensibilidade antes e depois da seleção de atributos usando PSO.	95
Tabela 38 – Comparação entre a Especificidade antes e depois da seleção de atributos usando PSO.	95
Tabela 39 – Comparação entre a Área Sob Curva ROC antes e depois da seleção de atributos usando PSO.	95

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Motivação e justificativa	19
1.2	Objetivos	21
1.3	Organização do Trabalho	21
2	TRABALHOS RELACIONADOS	23
3	FUNDAMENTAÇÃO TEÓRICA	29
3.1	Câncer de mama	29
3.1.1	Fatores de risco	30
3.1.2	Diagnóstico do câncer de mama	31
3.1.3	Diagnóstico precoce	35
3.2	Termografia de mama	36
3.3	Redes neurais artificiais	38
3.4	Aprendizado profundo	41
3.4.1	LeNet	42
3.4.2	ResNet50	42
3.4.3	NASNet mobile	43
3.4.4	SqueezeNet	43
3.4.5	Inception v3	45
3.5	Seleção de atributos	46
3.6	Classificação	48
3.6.1	Bayesiano ingênuo	48
3.6.2	Rede bayesiana	49
3.6.3	Random tree	49
3.6.4	Árvore de decisão J48	49
3.6.5	Random forest	50
3.6.6	Máquina de vetores de suporte (SVM)	50
3.6.7	Rede perceptron multicamadas (MLP)	51
3.7	Sistema de apoio ao diagnóstico	52
3.8	Considerações	53

4	MATERIAIS E MÉTODOS	55
4.1	Base de dados	56
4.2	Pré-processamento	57
4.3	Transferência de aprendizagem	59
4.4	Extração de atributos	59
4.5	Seleção de atributos	60
4.6	Balanceamento das classes	61
4.7	Treinamento e classificação	62
4.8	<i>Softwares</i> utilizados	64
4.9	Avaliação de desempenho	64
5	RESULTADOS E DISCUSSÃO	67
5.1	Análise da classificação	67
5.2	Análise da classificação após seleção de atributos com PSO	80
5.3	Comparação dos resultados antes e após a seleção de atributos	94
6	CONCLUSÃO	96
6.1	Dificuldades apresentadas	96
6.2	Contribuições e trabalhos futuros	97
	REFERÊNCIAS	98

1 INTRODUÇÃO

Este capítulo de introdução apresenta a motivação e os objetivos, gerais e específicos, desse trabalho. Como também a estrutura na qual o trabalho está organizado.

1.1 Motivação e justificativa

O câncer é uma doença que se origina da divisão celular desordenada resultante de mutação/lesões nos genes que coordenam o processo da tal divisão, ocasionando células “defeituosas” que se dividem descontroladamente (neoplasia), capazes de migrar para outra região do corpo, como também, não obedecem à morte programada da célula (apoptose) (BERNARDES et al., 2019). Por definição, câncer é o nome dado a mais de cem doenças que têm em comum o crescimento desordenado (maligno) de células que invadem os tecidos e órgãos, podendo espalhar-se (metástase) para outras regiões do corpo (RODRIGUES; CRUZ; PAIXÃO, 2015). Dentre os tipos de câncer, o câncer de mama se destaca, tanto no Brasil como no Mundo, como o segundo tipo de câncer mais comum entre as mulheres. Por outro lado, em homens a doença é rara, correspondendo a menos de 1% do total de casos desta doença. O câncer de mama corresponde a cerca de 28% dos novos casos de câncer em mulheres e no Brasil é o tipo de câncer que mais mata mulheres (SILVA; RIUL, 2011). Dados do Instituto Nacional do Câncer (INCA) mostram que o número de novos casos de câncer de mama em mulheres foi de aproximadamente 29,7% no ano de 2020 (INSTITUTO NACIONAL DO CÂNCER, 2020).

Para o câncer de mama, assim como todos os tipos de câncer, quanto mais cedo for diagnosticado mais efetivo será o tratamento, ou seja, o momento da detecção do câncer tem um grande impacto na sua letalidade. A detecção precoce reduz em 25% a mortalidade da doença. O padrão na prática clínica para apoio ao diagnóstico do câncer de mama por imagem ainda é a mamografia desde 1960. No entanto, a mamografia apresenta algumas desvantagens, como o alto custo e o desconforto na obtenção da imagem por meio da compressão da mama. Além disso os raios ionizantes, fundamentais para o exame, podem causar lesões no tecido mamário principalmente em pessoas acima de 70 anos, tornando o exame perigoso. Para pacientes jovens, o exame pode não ser eficaz na detecção devido à alta densidade do tecido mamário, o que dificulta a detecção de nódulos (HUSAINI et al., 2020).

Um método de exame de imagem em crescente investigação é a termografia. A termografia apresenta como vantagens não emitir radiação ionizante, ser não invasiva e indolor (HUSAINI et al., 2020). Uma câmera termográfica é capaz de mensurar a temperatura do tecido fotografado por meio da captura da radiação infravermelha, naturalmente emitida pelo corpo humano. A termografia baseia-se no princípio que a atividade metabólica e a circulação sanguínea nas células cancerígenas e tecidos adjacentes são maiores, consequentemente liberando mais calor que as células mamárias sadias (BOQUETE et al., 2012). A termografia vem sendo proposta como técnica complementar à mamografia, podendo ser mais eficiente comparada ao toque da mama e servindo como sistema de triagem, ajudando na detecção precoce de lesões de mama e consequentemente diminuindo a mortalidade.

Nos últimos anos, o Brasil sofreu e continua sofrendo devido às medidas austeras que apresentam consequências danosas aos direitos fundamentais, sobretudo aos direitos sociais (DIREITO; POLI, 2017). Dentre esses direitos, o direito à saúde pública e de qualidade é um dos mais afetados. Dito isto, trabalhar em pesquisas que possam auxiliar subáreas da saúde pública, de uma forma relativamente barata, é essencial diante do atual cenário como também o é incentivar o desenvolvimento de novas tecnologias de cunho nacional.

Com o crescente avanço tecnológico, o aprendizado de máquina tem sido cada vez mais utilizado e aplicado nas diversas áreas do saber humano. Dessa forma, a aplicação da aprendizagem de máquina como ferramenta auxiliar em práticas clínicas, tem mostrado ser uma poderosa ferramenta. Métodos como redes neurais, máquina de vetor de suporte, lógica *fuzzy*, árvores de decisão e outros são utilizados como classificadores para reconhecimento de possíveis doenças por meio de exames de imagens (BOQUETE et al., 2012). Em busca de agilizar o processo de diagnóstico do câncer de mama, algumas técnicas de aprendizado de máquina estão sendo utilizadas para o mapeamento de imagens da mama através do reconhecimento de padrões.

Este trabalho é orientado pela seguinte pergunta: como apoiar o diagnóstico do câncer de mama utilizando aprendizagem de máquina?

Sendo assim, o objetivo desse trabalho foi analisar o desempenho de diferentes técnicas de aprendizagem profunda para extração de atributos (LeNet, ResNet50, NAS-NetMobile, SqueezeNet e Inception v3), como também analisar o uso de algumas técnicas de classificação (*Naive Bayes*, *Bayes Net*, *Random Tree*, Árvore de Decisão J48, *Random Forest*, SVM e Rede Neural MLP) para o reconhecimento de lesões em imagens

termográficas utilizando o *software* de aprendizado de máquina Weka. A base de dados de imagens termográficas usada neste trabalho, foi fornecida pelo grupo de pesquisa de Engenharia Térmica do Departamento de Engenharia Mecânica da Universidade Federal de Pernambuco. Essa base continha 1052 imagens, divididas em quatro classes: lesão benígna, lesão málgna, cisto e sem lesão.

1.2 Objetivos

Este trabalho tem como objetivo geral propor um modelo de máquina de aprendizado para apoio ao diagnóstico do câncer de mama para classificação de lesões em imagens termográficas, utilizando classificadores tradicionais e redes neurais artificiais profundas como extratores de atributos. A ideia é investigar diferentes redes neurais profundas combinadas com diferentes classificadores, além de analisar o comportamento com e sem redução de atributos, de modo a proporcionar um modelo que sirva de base para utilização em aplicações que auxiliem o diagnóstico do câncer de mama usando imagens termográficas.

Como objetivos específicos, têm-se:

- Estudar e realizar experimentos com diferentes arquiteturas de redes neurais artificiais profundas para extrair atributos;
- Estudar e realizar experimentos com diferentes classificadores para os atributos que foram extraídos;
- Construir um modelo de rede neural artificial profunda para classificação de imagens termográficas de mama, com habilidade de detectar e classificar lesões de mama;
- Habilitar a termografia como uma técnica complementar ao diagnóstico do câncer de mama.

1.3 Organização do Trabalho

Este trabalho foi organizado em seis capítulos. O primeiro capítulo, a introdução, é realizada a contextualização do problema de pesquisa como também a explanação dos objetivos.

O segundo capítulo mostra os trabalhos relacionados à detecção e classificação de lesões mamárias em imagens termográficas. O terceiro capítulo apresenta os conceitos fundamentais para o entendimento desse trabalho.

No quarto capítulo é explicada a metodologia utilizada: são apresentadas as informações sobre a base de dados de imagens termográficas, os detalhes das etapas de pré-processamento (extração e seleção de atributos), a classificação e análise dos resultados.

O quinto capítulo apresenta os resultados e as discussões sobre os experimentos que foram realizados. São mostrados os resultados para a abordagem sem selecionar os atributos (sem PSO) e os resultados para a abordagem com a seleção de atributos (com PSO).

No último capítulo, é apresentada uma conclusão a partir do que foi experimentado e apresentado, como também algumas sugestões para trabalhos e contribuições futuras.

2 TRABALHOS RELACIONADOS

Existem diversos trabalhos na literatura que utilizam imagens termográficas de mama, onde o principal objetivo é identificar anormalidades mamárias.

Gonçalves et al. (GONÇALVES, 2017), apresentou um trabalho utilizando uma base de dados composta por imagens termográficas de 70 pacientes distribuída da seguinte forma: 21 pacientes sem alterações na mama (normal), 23 com alterações benignas e 26 com alterações malignas. Na etapa de extração de atributos, foram extraídos 17 atributos. As características consideradas foram de medidas estatísticas e dimensão fractal. Na etapa de classificação, foram utilizadas Máquina de Vetores de Suporte (*SVM*) e redes neurais artificiais (*RNA*). O resultado com maior precisão foi obtido com o classificador *SVM*, usando a função de kernel cúbica, com uma taxa de acurácia de 80,95%. Esta pesquisa contribuiu para fortificar a ideia de que as imagens termográficas podem ser utilizadas no diagnóstico precoce do câncer de mama.

No trabalho de Madhu et al. (MADHU et al., 2016) foram propostas novas abordagens para extrair atributos que possuíam mais significado médico e seriam facilmente interpretáveis pelo clínico, além de aumentar a especificidade na análise de imagens termográficas. Foram utilizadas imagens de 265 pacientes, divididas da seguinte forma: 120 imagens de pacientes normais, 78 de pacientes com anomalias malignas, 67 distribuídas entre pacientes com anomalias benignas, casos não-malignos com tecidos sensíveis a hormônios e pacientes lactantes. O objetivo desse trabalho foi distinguir tumores malignos de outras lesões não-malignas sob diferentes condições, incluindo tecidos sensíveis a hormônios, condições de lactação e certas condições benignas, que têm uma resposta térmica aumentada. Como atributos, foram utilizadas a temperatura relativa das mamas, a comparação das temperaturas das mamas contralaterais e a integridade das bordas da mama. O estudo apresentou um resultado de especificidade de até 98,90%. Esse estudo foi importante porque aumentou a especificidade, algo que era um problema conhecido no rastreamento termográfico. O aumento foi possível devido às propriedades que foram utilizadas na classificação para diferenciar entre lesão maligna e lesão benigna: as bordas são mais irregulares nas lesões malignas e mais simétricas no tamanho e número de pontos quentes para as não-malignas.

Mambou et al. (MAMBOU et al., 2018) realizaram um estudo comparativo de vá-

rias técnicas de detecção de câncer de mama usando técnicas de visão computacional e modelos de aprendizagem profunda. Os objetivos desse trabalho, foram deixar mais compreensível o processo de detecção de câncer de mama usando imagens termográficas e propor um modelo inteligente para ajudar na detecção precoce do câncer de mama. O modelo proposto baseou-se no conceito de transferência de aprendizagem, onde foi inicialmente utilizado um modelo pré-treinado chamado de Inception V3, que apresenta uma excelente eficiência na extração de atributos, aumentando a precisão e reduzindo a entropia do modelo. A rede neural profunda utilizada possui 10 neurônios na primeira camada, onde cada um está conectado a 20 neurônios na próxima camada e, da mesma forma, cada um está conectado a 10 neurônios na terceira camada. Além disso, a camada final classificada foi retreinada para que pudesse distinguir entre ter câncer e não ter câncer com uma taxa de confiança maior que 0,6. Caso, após a saída da camada final, a confiança tivesse seu valor entre 0,6 e 0,5, a matriz de atributos seria enviada para um SVM para realização da previsão da saída. Foram treinadas 64 amostras de mama, incluindo 32 imagens de mamas saudáveis e 32 imagens de mamas não saudáveis, onde cada mama tinha 20 imagens sequenciais. Na etapa da classificação adicional, foi utilizada uma variação embutida do SVM que foi o LinearSVC. O LinearSVC tem mais flexibilidade, suporta entrada densa e esparsa e é adequado para classificação multiclasse. O modelo apresentou uma confiança de 0,78 para imagens de mama não saudáveis e uma confiança de 0,94 para imagens de mamas saudáveis.

O trabalho de Ekici e Jawzal (EKICI; JAWZAL, 2020) teve como objetivo desenvolver um *software* para detecção automática do câncer de mama que utilizasse técnicas de processamento de imagem e algoritmos para analisar imagens térmicas da mama, buscando detectar os sinais da doença nessas imagens e contribuir para a detecção precoce do câncer de mama. Foi proposto um novo algoritmo para extração de características das imagens de termografia da mama com base em bio-dados, análise de imagens e estatísticas de imagens. A metodologia do trabalho foi dividida em cinco partes: aquisição dos dados, pré-processamento, segmentação, extração de características, classificação e decisão. Na etapa de extração de características, os atributos que foram extraídos das imagens, foram usados na etapa de classificação para classificar as imagens como normais ou suspeitas usando redes neurais convolucionais (CNNs) otimizadas pelo algoritmo de Bayes. O algoritmo proposto obteve uma taxa de acurácia de 98,95% para um conjunto de dados

de imagens térmicas de 140 indivíduos.

O artigo de Roslidar et al. (ROSLIDAR et al., 2019) discutiu sobre experimentos iniciais usando *fine-tuning* com alguns modelos de redes neurais convolucionais (CNNs) para detecção de câncer de mama: *ResNet101*, *DenseNet*, *MobileNet2* e *ShuffleNetV2*. O *fine-tuning* consiste em treinar a camada completamente conectada da CNN, enquanto os parâmetros das outras camadas ficam congelados. A base de dados de imagens térmicas da mama utilizada foi do *Database for Mastology Research* (DMR). Foram definidas duas classes de classificação: com câncer e saudável. Na etapa de treinamento da rede, o conjunto de dados foi dividido em 80% para treinamento e 20% para validação. Além disso, antes do treinamento as imagens foram redimensionadas para $224 \times 244 \times 3$. Como resultado, o *DenseNet201* obteve uma classificação com 100% de acurácia. Porém, a *MobileNetV2* destacou-se pelo seu desempenho, com baixo tempo de treinamento e com a menor perda, além de ter obtido uma excelente acurácia de 99,6 %.

Chaves et al. (2020) propõem um sistema inteligente de detecção de câncer de mama, baseado na análise de imagens termográficas e algoritmos de aprendizagem profunda. Para isso, foram utilizadas redes neurais convolucionais (CNN) aplicadas a um banco de dados de 440 imagens por infravermelho de 88 pacientes, cada imagem possuía 640×480 pixels. A base de dados foi coletada pelo Departamento de Ciências da Computação da Universidade Federal Fluminense, utilizando a câmera FLIR SC620. As classes de classificação foram normal (mama sem o tumor) e patologia (mama com tumor). Para o treinamento, foi utilizada a abordagem de transferência de aprendizagem com as seguintes arquiteturas de redes neurais convolucionais: *AlexNet*, *GoogLeNet*, *ResNet-18*, *VGG-16* e *VGG-19*. Os experimentos foram realizados variando o número de épocas do treinamento: 10, 20 e 30 épocas. Os melhores resultados foram obtidos com 20 épocas para as CNNs VGG-16 e VGG-19. A VGG-16 obteve 77,5% de acurácia, 85% de sensibilidade e 70% de especificidade. Já a VGG-19 obteve 77,5% de acurácia, 90% de sensibilidade e 65% de especificidade. Portanto, os resultados mostraram um grande potencial de uso das técnicas de aprendizagem profunda combinadas com imagens por infravermelho no auxílio ao diagnóstico do câncer de mama.

No artigo de Chebbah et al. (CHEBBAH; OUSLIM; RYAD, 2018) foi proposta uma abordagem baseada em técnicas de processamento de imagem, aprendizagem de máquina e termografia para detecção precoce de anormalidades mamárias. Foi utilizada uma base

de dados pública do PROENG, envolvendo 80 imagens termográficas, sendo 50 de mamas saudáveis e 30 de mamas com algum tumor. Na metodologia desse trabalho, inicialmente as imagens termográficas foram convertidas de coloridas para escala de cinza, em seguida foram usadas técnicas de processamento de imagem para segmentação das regiões de interesse (ROIs), que estão relacionadas com as áreas de separação da imagem: isolar as mamas esquerda e direita; logo após, foram usadas uma GLCM (*Gray Level Co-occurrence Matrix*) e operações morfológicas na imagem segmentada para extrair características estatísticas e textuais. Por fim, foi utilizado um classificador SVM para detecção automática de anormalidade mamária. Os resultados mostraram-se interessantes com 93,3% de sensibilidade, 90% de especificidade e 91,25% de acurácia. Portanto, foi possível concluir que a termografia por infravermelho aliada ao uso de um algoritmo de classificação adequado, pode ser uma ferramenta valiosa e complementar para detecção do câncer de mama, podendo contribuir para redução da mortalidade.

Com o objetivo de contribuir na detecção correta de doenças da mama, o trabalho de Chebbah et al. (CHEBBAH; OUSLIM; BENABID, 2022) propôs um novo sistema baseado em inteligência artificial e termografia da mama. Foram usadas 170 imagens de termografia da mama, coletadas de um banco de dados *open-source*. Inicialmente, na etapa de segmentação, foi utilizado um modelo de *deep learning* específico para imagens biomédicas, chamado de U-net. Esse modelo é uma rede totalmente convolucional baseada em uma arquitetura de ponta a ponta e composta de várias camadas de rede neural convolucional e deconvolucional (CNN e DCNN). A CCN reduz a dimensão espacial dos dados de entrada, enquanto que a DCNN aplica uma transformação reversa na CNN para restaurar o tamanho original das imagens. Para extração de características, foi usado o método GLCM, que descreve a textura da imagem calculando a frequência na qual as combinações de pixels com valores de cinza definidos ocorrem em uma posição espacial específica em uma imagem. Como o crescimento dos vasos sanguíneos representa o primeiro sinal da existência de células cancerosas, isso ocasionará a produção de redes vasculares ao redor do tumor, aumentando assim a temperatura da mama. Diante disso, o presente trabalho buscou extrair essas redes vasculares usando combinação de operações de limiarização e morfológicas. Por fim, com as características extraídas usando GLCM e da análise de rede vascular, foi realizada a classificação usando quatro classificadores: KNN (*K-Nearest Neighbour*) com K igual a 7, *AdaBoost* usando árvore de decisão como classificador padrão

e um número de estimador igual a 50, *Random Forest* com 100 árvores e SVM (*Support Vector Machine*) com kernel linear e valor de C igual a 1. O modelo U-net de aprendizado profundo alcançou um excelente desempenho, exibindo 89,03% de interseção sobre união. Já a extração de características usando análise de textura e rede vascular teve melhores resultados experimentais usando o SVM como classificador, resultando em 94,4% de acurácia, precisão de 96,2%, *recall* de 86,7%, *F1-score* de 91,2% e uma taxa de verdadeiro negativo de 98,3%. Esse trabalho obteve excelentes resultados de classificação, além de mostrar um exemplo da aplicação por meio de uma interface gráfica, podendo assim ser melhor explorada em trabalhos futuros e levar o acesso a esse exame de diagnóstico do câncer de mama para regiões mais afastadas dos grandes centros urbanos.

Tiwari et al. (TIWARI et al., 2021) teve como objetivo fornecer um método preciso para diagnosticar o câncer de mama usando imagens termográficas e comparar os resultados de métodos de aprendizado de máquina com técnicas de aprendizado profundo. As imagens térmicas foram coletadas de várias bases de dados, como *Ann Arbor Thermography* e *DMR-IR-database*. Foi realizado um pré-processamento nas imagens, onde primeiramente as imagens termográficas foram convertidas em imagens do modelo de cores RGB. Em seguida, as imagens foram redimensionadas para um tamanho padrão de 128×128 . Para a abordagem usando métodos de aprendizagem de máquinas, foram realizadas as seguintes etapas: a segmentação foi baseada na região de interesse da imagem, histograma e métodos estatísticos; a extração de característica, foi realizada usando *Hu Moments*; a classificação foi testada usando os classificadores *Random Forest*, SVM e *Naive Bayes* para classificar em três classes: estágio inicial, mediano e avançado do câncer de mama. Já na abordagem usando aprendizado profundo, foi utilizada segmentação baseada em DCNN, e em seguida a classificação usando as CNNs Inception V3 e ResNet50 para classificar as imagens termográficas em três classes também: estágio inicial, mediano e avançado do câncer de mama. Como resultado, a rede residual, ResNet50, obteve a maior taxa de acurácia na classificação com 96%.

Mishra et al. (MISHRA et al., 2020), utilizou uma base de imagens dinâmicas disponível no conjunto DMR. Foram utilizadas 521 imagens classificadas como saudáveis e 160 imagens classificadas como anormais (imagens com regiões de crescimento de câncer). Inicialmente, houve a conversão das imagens em escala de cinza e o redimensionamento a fim de garantir o mesmo tamanho para todas as imagens. Para balancear o conjunto de

dados foi usada a técnica de *data augmentation* em tempo real usando rotação aleatória, translação no eixo X e translação no eixo Y. Por fim, as imagens foram segmentadas e classificadas, obtendo excelentes resultados com um modelo de CNN: 95,8% de acurácia.

O trabalho de Rodrigues et al. (RODRIGUES et al., 2019) teve como objetivo propor uma abordagem de seleção de características usando algoritmos genéticos (AG) e otimização por enxame de partículas (PSO) em imagens termográficas com lesões mamárias. Essa abordagem teve a finalidade de otimizar a identificação e classificação das lesões mamárias. Foram utilizados diversos classificadores para avaliar o desempenho dos subconjuntos com características selecionadas, onde o SVM obteve o melhor desempenho nesses experimentos. Como resultados, com o AG foi possível reduzir de 169 para 57 atributos com acurácia de 91,12%. Já com o PSO, foi possível reduzir de 169 para 60 atributos com acurácia de 86,16%. Portanto, o trabalho concluiu que a redução de atributos pode ser realizada sem perder muito em acurácia comparado ao conjunto completo com todos os atributos.

Os trabalhos abordados acima são estudos recentes que fazem uso de técnicas de aprendizagem de máquina ou de aprendizagem profunda em imagens termográficas, com a finalidade de contribuir na detecção de lesões mamárias. Os trabalhos apresentados possuem resultados excelentes na detecção do câncer de mama em imagens termográficas. O presente trabalho, possui uma metodologia semelhante aos trabalhos citados, porém, buscou-se realizar experimentos de extração de características com diferentes redes convolucionais, além de selecionar os melhores atributos com PSO e, por fim, analisar cada conjunto das características extraídas e selecionadas com diferentes classificadores. Portanto, neste trabalho a ideia foi fazer uma análise geral sobre fatores importantes no contexto de classificação de imagens: análise sobre os melhores extratores de atributos, os melhores classificadores e o comportamento a partir da seleção dos melhores atributos.

3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo apresentar a fundamentação teórica relacionada ao contexto do câncer de mama, como também os conceitos básicos sobre redes neurais artificiais, aprendizagem profunda e sistemas de apoio ao diagnóstico.

3.1 Câncer de mama

O câncer é originado a partir de uma mutação genética, ou seja, de uma alteração no DNA da célula, essa célula recebe instruções erradas para realizações de suas atividades. Essas alterações no DNA, ocorrem em genes especiais: os proto-oncogenes. Os proto-oncogenes, que são os genes que ajudam as células a crescer, saem do estado inativo para o estado ativo, tornando-se oncogenes, responsáveis por transformar as células normais em células cancerosas (MINISTÉRIO... , 2011). As células normais responsáveis pela formação dos tecidos do corpo humano são capazes de se multiplicar e isso é um processo contínuo e natural. De modo geral, as células normais crescem, multiplicam-se e morrem de maneira ordenada. O crescimento das células cancerosas é diferente do crescimento das células normais. As células cancerosas, em vez de morrerem, continuam crescendo de forma incontrolável, formando outras novas células anormais (MINISTÉRIO... , 2011).

O câncer de mama é uma doença causada por conta da multiplicação desordenada de células anormais da mama, que resulta na formação de um tumor com potencial de invadir outros órgãos (INSTITUTO NACIONAL DO CÂNCER, 2022d).

O processo de desenvolvimento do câncer é chamado de carcinogênese. A carcinogênese acontece de forma lenta, podendo levar vários anos para que uma célula cancerosa se prolifere e dê origem a um tumor visível ou palpável. Em média, 100 dias é o tempo para que ocorra a duplicação celular no câncer de mama, ou seja, o processo de carcinogênese. O tumor pode ser palpável quando atinge 1 centímetro de diâmetro. Nessa esfera de 1 centímetro, contém uma estimativa de 1 bilhão de células, consistindo no resultado de 30 duplicações celulares em um período de 10 anos. Portanto, em aproximadamente 10 anos, uma célula maligna torna-se um tumor de 1 centímetro (INSTITUTO NACIONAL DO CÂNCER, 2002).

3.1.1 Fatores de risco

O câncer de mama é uma doença associada a vários fatores: idade; fatores endócrinos/história reprodutiva; fatores comportamentais/ambientais e fatores hereditários. O principal fator associado ao câncer de mama é a idade, pois à medida em que se envelhece há um acúmulo de exposições, como também ocorrem mudanças biológicas e hormonais ao longo dos anos (VASCONCELOS; SANTOS; LIMA, 2018; SANTANA et al., 2020; PEREIRA et al., 2020b; PEREIRA et al., 2020a; SANTOS; SILVA; SANTANA, 2020; SANTANA et al., 2021; PEREIRA et al., 2021; PEREIRA et al., 2021; BARBOSA et al., 2020; SILVA et al., 2021; SANTANA et al., 2021; MACEDO et al., 2021; PEREIRA et al., 2021).

Quanto maior for a exposição a certos fatores endócrinos/história reprodutiva - que estão relacionados principalmente ao estímulo estrogênico, seja endógeno ou exógeno - maior será o risco. Dentre esses fatores, estão: menarca precoce (antes dos 12 anos), menopausa tardia (após 55 anos), primeira gravidez após os 30 anos, uso de contraceptivos orais (estrogênio-progesterona) e terapia de reposição hormonal pós-menopausa (estrogênio-progesterona) (INSTITUTO NACIONAL DO CÂNCER, 2022c).

Os fatores comportamentais e ambientais incluem: ingestão de bebida alcoólica, sobrepeso e obesidade, falta de atividade física e exposição à radiação ionizante (INSTITUTO NACIONAL DO CÂNCER, 2022c; VASCONCELOS; SANTOS; LIMA, 2018; SANTANA et al., 2020; PEREIRA et al., 2020b; PEREIRA et al., 2020a; SANTOS; SILVA; SANTANA, 2020; SANTANA et al., 2021; PEREIRA et al., 2021; PEREIRA et al., 2021; BARBOSA et al., 2020; SILVA et al., 2021; SANTANA et al., 2021; MACEDO et al., 2021; PEREIRA et al., 2021). Os fatores genéticos/hereditários foram relacionados à presença de mutações em determinados genes. Essas mutações são encontradas principalmente nos genes BRCA1 e BRCA2. O câncer de mama hereditário corresponde de 5 % a 10 % do total de casos (INSTITUTO NACIONAL DO CÂNCER, 2022c; VASCONCELOS; SANTOS; LIMA, 2018; SANTANA et al., 2020; PEREIRA et al., 2020b; PEREIRA et al., 2020a; SANTOS; SILVA; SANTANA, 2020; SANTANA et al., 2021; PEREIRA et al., 2021; PEREIRA et al., 2021; BARBOSA et al., 2020; SILVA et al., 2021; SANTANA et al., 2021; MACEDO et al., 2021; PEREIRA et al., 2021).

3.1.2 Diagnóstico do câncer de mama

O câncer de mama é o tipo de câncer que mais acomete as mulheres em todo o mundo, tanto em países desenvolvidos quanto em países em desenvolvimento. Os diagnósticos e os tratamentos dos variados tipos de câncer, passaram por importantes e expressivos avanços. Os métodos ficaram mais modernos e específicos, contribuindo assim para um diagnóstico mais apurado, um melhor acompanhamento e uma melhor avaliação de prognósticos dos pacientes. Os avanços nos diagnósticos e nos tratamentos têm possibilitado sobrevida progressivamente maior em casos considerados incuráveis (Do Nascimento; Da Rocha Pitta; De Melo Rêgo, 2015).

O diagnóstico do câncer de mama deve estar ancorado em um tripé: exame clínico, exame de imagem e análise histopatológica (análise microscópica dos tecidos para a detecção de possíveis lesões). Além disso, é fundamental que seja realizada uma boa anamnese (histórico dos sintomas informados pelo paciente), exame físico e o complemento por meios de exames de imagens (mamografia, ultrassonografia ou ressonância magnética) para avaliar a necessidade de realizar a biópsia da lesão (INSTITUTO NACIONAL DO CÂNCER, 2022b; SILVA et al., 2021; CORDEIRO; BEZERRA; SANTOS, 2017; CORDEIRO et al., 2012; CORDEIRO; SANTOS; SILVA-FILHO, 2016a; CORDEIRO; SANTOS; SILVA-FILHO, 2016b; CORDEIRO; SANTOS; SILVA-FILHO, 2017; LIMA; SILVA-FILHO; SANTOS, 2016). O diagnóstico final é confirmado por meio da biópsia, que é uma técnica que consiste na retirada de um fragmento do nódulo ou da lesão suspeita por meio de punções (extração por agulha) ou de uma pequena cirurgia. O material que foi retirado é analisado pelo patologista para a definição do diagnóstico (INSTITUTO NACIONAL DO CÂNCER, 2022a).

- Autoexame

O autoexame das mamas (AEM) é um dos métodos conhecidos para detecção de lesões mamárias. Porém, como faltam evidências com relação à redução de mortalidade por câncer de mama, essa técnica não faz parte de programas de rastreamento de tal malignidade (CASTRO, 2021). O autoexame, juntamente com o exame clínico das mamas, são métodos menos eficazes, porém são mais simples e mais acessíveis. O AEM é uma alternativa que pode ser utilizada em locais com poucos recursos, onde o exame clínico de rotina e a mamografia podem não ser viáveis (CASTRO, 2021). A Figura 1 mostra o passo

a passo sobre como realizar o autoexame.

Figura 1 – Como fazer o autoexame das mamas.

Uma das formas mais simples de se fazer a prevenção do câncer de mama é o autoexame.

É importante você saber que:

- Ao detectar qualquer anormalidade, não entre em pânico e procure sempre um médico.
- O autoexame das mamas não substitui a consulta de rotina que deve ser feita ao mastologista.
- Toda mulher deve realizar uma mamografia de base aos 35 anos e anualmente após os 40 anos.
- Não tenha medo de examinar-se e descobrir algo, faça com que o autoexame seja uma rotina na sua vida.
- Lembre-se o câncer de mama detectado precocemente tem altíssima probabilidade de cura.

Prevenir é um ato de amor à vida.

Toque-se !

Sobre o exame de toque:

- Durante a palpitação da mama, observe o surgimento de nódulos, note sua consistência, mobilidade, limites, localização e tamanho.
- A dor também deve ser considerada, embora não seja frequente em casos de câncer de mama.
- Nódulos que surgem antes da menstruação e desaparecem após a mesma, geralmente não têm valor clínico importante.
- O uso de anticoncepcionais, terapia de reposição hormonal, período pré-menstrual, amamentação, gestação e aborto recente podem influenciar e dificultar a identificação de nódulos.

Como fazer o autoexame das Mamas



NO ESPELHO

Coloque as mãos na cintura e observe o formato, tamanho e contorno de suas mamas (observe se estão iguais, simétricas). Veja se existem pregas, depressões ou alterações na pele da mama, aréola e mamilo.

NO ESPELHO

Coloque as mãos para o alto e fique atenta aos mesmos sinais de igualdade da pele já descritos. É uma oportunidade única de observar sua simetria mamária. Observe se as mamas estão alinhadas, assim como as aréolas e mamilos. Eles deverão sempre estar na mesma linha. Observe se o sutiã faz marca só em uma mama, e qual é o local para verificar se há edema (inchaço).



NO CHUVEIRO

Ensaíe bem as mamas e axilas, deste modo suas mãos percorrerão melhor e fará o exame. Coloque o braço atrás da nuca, coluna reta e com a ponta dos dedos percorra todas as áreas da mama em movimentos circulares de fora para dentro, procurando por espessamentos ou caroços. Use a mão direita para examinar a mama esquerda e vice-versa. Você também poderá realizar esse exame fora do chuveiro, com maior atenção.



MOVIMENTOS CIRCULARES

Devem ocorrer da parte externa superior da mama (perto da axila) em direção ao mamilo, de modo firme, mas delicado.

MAMILOS E ARÉOLAS

Pressione suavemente os mamilos e observe com cuidado a presença de secreções e a presença de lesões na sua pele sensível, assim como nas aréolas.



AXILAS

Elas também fazem parte do tecido mamário, e devem ser examinadas do mesmo modo que as mamas. Realize movimentos circulares da mama para a axila, observando a presença de caroços na região.



DEITADA

Coloque um apoio sob o ombro direito (pode ser um travesseiro ou toalha) e a mão direita atrás da cabeça. Examine sua mama direita com a mão esquerda, repetindo os movimentos circulares. Repita esse passo com o lado esquerdo.



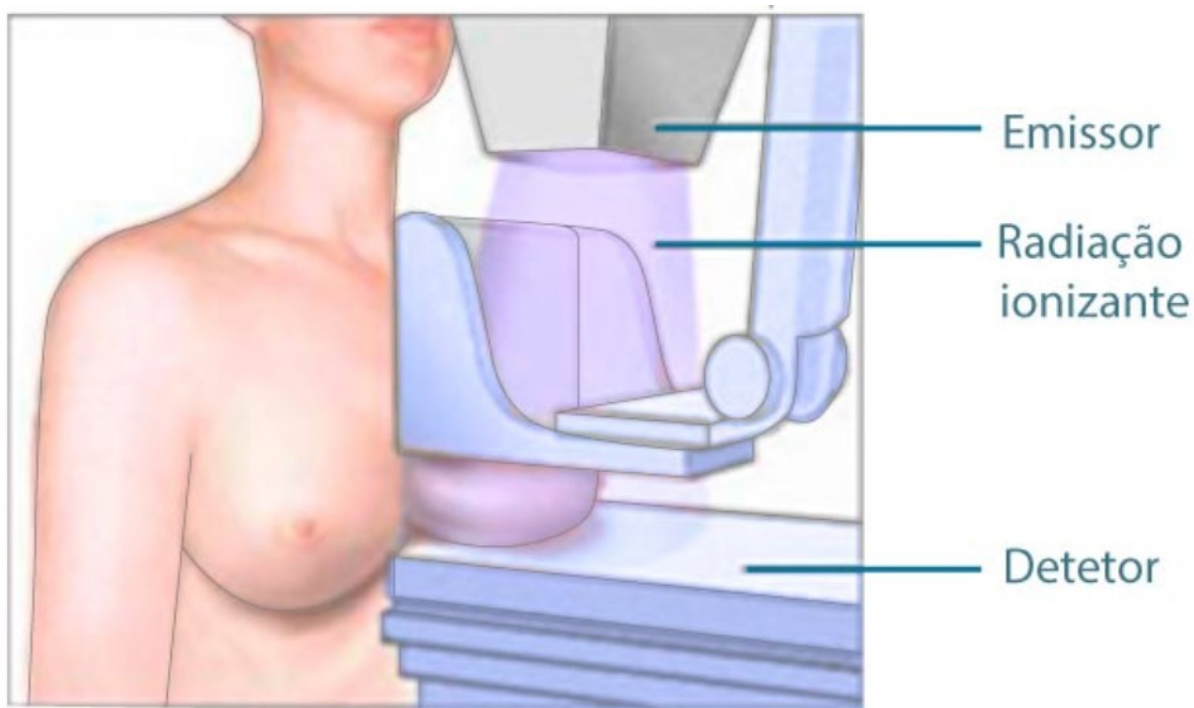
Fonte: Caldeira (2015).

• Mamografia

A mamografia é um exame de imagem que fornece algumas informações sobre a morfologia, anatomia e patologias das mamas. O objetivo do exame de mamografia é registrar imagens da mama para diagnosticar possíveis presenças de estruturas que indiquem uma patologia. Para isso, ele utiliza uma abordagem semelhante ao procedimento de Raios X, porém, com níveis baixos de radiação entre intervalos específicos, sendo assim, um tipo específico de radiografia (CAMPOS A. P. DE; CAMARGO, 2015; SANTANA et al., 2021).

A Figura 2 mostra um exemplo de como é realizada a mamografia por meio do uso do mamógrafo. Por meio desse exame, será gerada uma imagem da mama, como consequência do uso de baixas dosagem de irradiação de Raios-X (feixes ionizantes) sobre a mama da paciente. Atualmente, a mamografia é o procedimento padrão para o diagnóstico de câncer de mama (KANDLIKAR et al., 2017).

Figura 2 – Exame de mamografia.



Fonte: Saúde bem estar (2020).

Porém, a mamografia apresenta algumas desvantagens: altas taxas de falso positivo; diagnóstico falso negativo para mulheres jovens, já que suas mamas tendem a ter um tecido mais denso (glandular), resultando em uma imagem mamográfica da mesma cor de uma lesão; e a exposição à radiação ionizante cumulativa, podendo aumentar o risco de câncer (PEREIRA, 2020).

- Ultrassonografia de mama

A ultrassonografia da mama é um exame de imagem que auxilia detecção de algumas alterações mamárias. Esse exame é realizado por meio de um transdutor que é encostado ao corpo da paciente, conforme pode-se ver na Figura 3. Durante a realização do exame, a paciente deve permanecer na posição dorsal com os braços atrás da cabeça, enquanto o profissional passa o transdutor sobre a mama em busca de detectar algum nódulo (MORAIS, 2015).

Apesar de ser capaz de diagnosticar casos de câncer de mama assintomáticos e não identificados por meio do exame físico ou mamografia, a ultrassonografia possui algumas limitações: altos custos globais, devido a realização do exame como também por conta do número relativamente alto de exames falsos positivos, ou seja, mulheres

são submetidas à investigação diagnóstica com punções, biópsias ou cirurgia sem que se chegue ao diagnóstico do câncer (NASTRI; MARTINS; LENHARTE, 2011).

Figura 3 – Ultrassonografia da mama.



Fonte: Moraes (2015).

- Ressonância magnética

Esse exame consiste na absorção de energia ressonante emitida por átomos de hidrogênio, alinhados a partir de um forte campo magnético. Sistemas computadorizados formam imagens tridimensionais. Desta forma, é possível analisar características morfológicas detalhadas do tecido, como também a impregnação de contraste que está relacionada à vascularização e à permeabilidade vascular aumentada na região com a lesão (MORAIS, 2015).

Contudo, apesar de a ressonância magnética ser útil como exame de imagem para diagnosticar lesões, principalmente para definir com precisão o tamanho do tumor, a sua adição ao exame de mamografia eleva os custos. Outrossim, não é indicada como um exame de rastreamento de forma isolada, pois pode deixar passar alguns tipos de câncer que poderia ser diagnosticado com a mamografia (PEREIRA, 2020). Com isso, reforça-se

a necessidade de estudar, desenvolver e promover outras metodologias que sejam um suporte de baixo custo ao diagnóstico do câncer de mama.

3.1.3 Diagnóstico precoce

As principais estratégias para a detecção precoce do câncer de mama são por meio do diagnóstico precoce e do rastreamento (INSTITUTO NACIONAL DO CÂNCER, 2022b).

O diagnóstico precoce contribui para o início do tratamento adequado de forma mais breve, evitando assim que o câncer desenvolva-se para um estado mais crítico. Nesse contexto, destaca-se a importância da educação da mulher e dos profissionais de saúde para o reconhecimento dos sinais e sintomas suspeitos do câncer de mama, como também do acesso rápido e facilitado aos serviços de saúde tanto na atenção primária quanto nos serviços de referência para investigação diagnóstica (INSTITUTO NACIONAL DO CÂNCER, 2022b).

O rastreamento consiste na aplicação de teste ou exame numa população sem sinais e sintomas sugestivos de câncer de mama, objetivando-se identificar alterações sugestivas de câncer e encaminhar as mulheres com resultados anormais para investigação diagnóstica. O rastreamento pode ser oportunístico ou organizado. O primeiro caso é ofertado às mulheres que chegam oportunamente às unidades de saúde, enquanto o modelo organizado convida formalmente as mulheres na faixa etária alvo para realizar os exames periódicos. Países que implantaram programas efetivos de rastreamento, com cobertura da população alvo, qualidade dos exames e, sobretudo, tratamento adequado e oportuno, a mortalidade por câncer de mama vem diminuindo (INSTITUTO NACIONAL DO CÂNCER, 2022b).

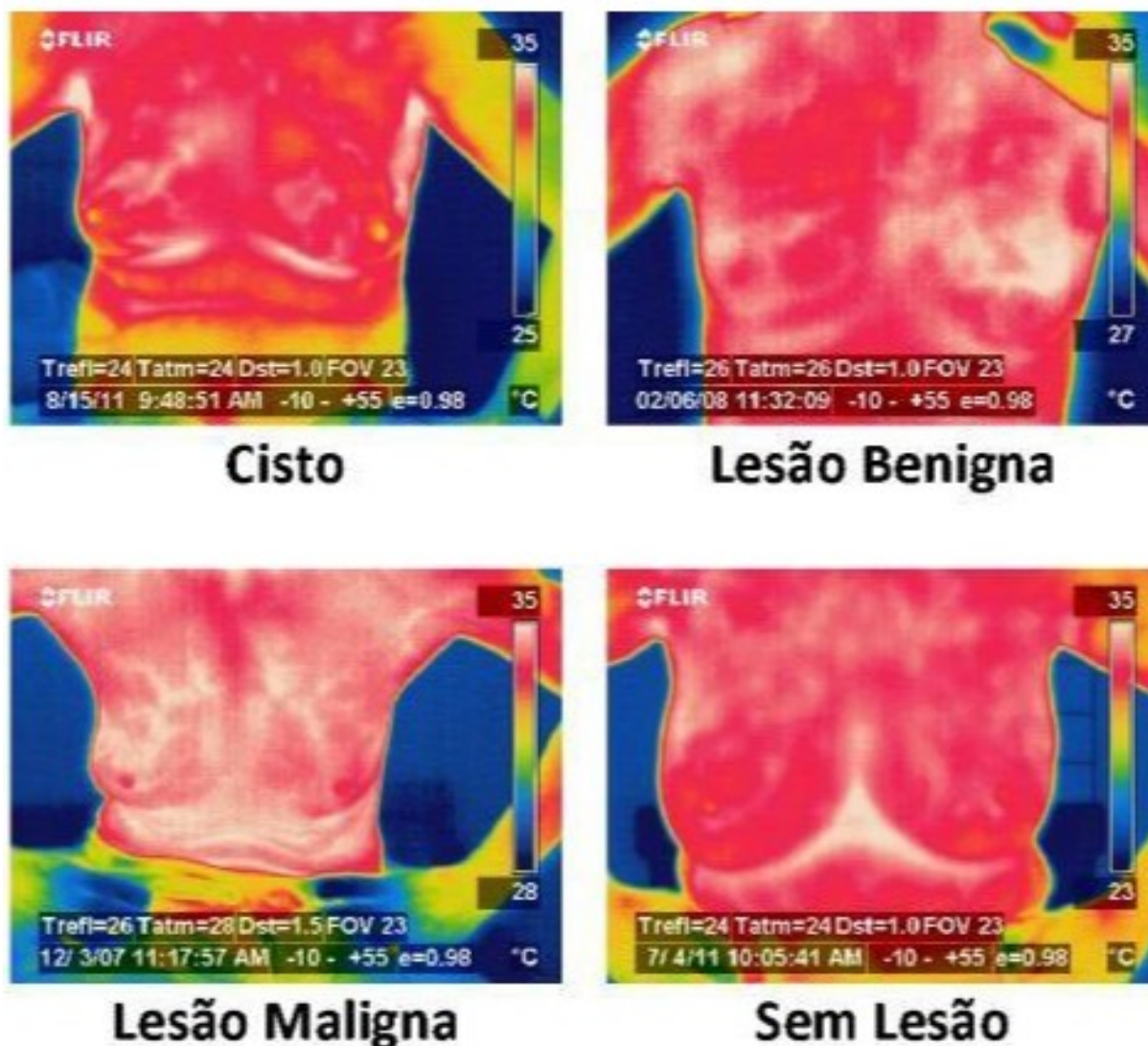
Ademais, o câncer de mama pode ser detectado em fases iniciais, em grande parte dos casos, aumentando assim a possibilidade de tratamento menos agressivo e com taxas de sucesso satisfatórias (INSTITUTO NACIONAL DO CÂNCER, 2022a). Tudo isso reforça a importância em investir em ferramentas aliadas ao diagnóstico precoce na luta contra o câncer de mama.

3.2 Termografia de mama

A FDA (*Food and Drug Administration*), em 29 de janeiro de 1982, publicou a confirmação da termografia como procedimento complementar de diagnóstico por imagem para detecção do câncer de mama (ETEHADTAVAKOL; NG, 2013). Assim, a termografia é uma técnica auxiliar no processo de diagnóstico do câncer de mama, sendo uma técnica não invasiva, barata e rápida. Essa técnica tem por base a utilização de imagens registradas por uma câmera de infravermelho, onde busca-se registrar a radiação térmica emitida pela superfície da pele da paciente, apresentando na região uma distribuição de temperatura (SILVA et al., 2018).

A termografia mamária é realizada totalmente sem contato, sem incidência de radiação alguma e sem compressão das mamas da paciente (ETEHADTAVAKOL; NG, 2013). Esse exame é útil para todas as mulheres de todas as idades, mulheres grávidas e lactantes, mulheres com seios densos ou fibrocísticos, mulheres em terapia de reposição hormonal e mulheres na pré ou pós-menopausa. A termografia é um teste fisiológico, assim, ela é especialmente útil durante os estágios iniciais do crescimento do tumor (ETEHADTAVAKOL; NG, 2013).

Figura 4 – Exemplo de imagens de termografia da mama para as classes: Cisto, Lesão Benigna, Lesão Maligna e Sem Lesão.



Fonte: Silva et al. (2018).

O procedimento de termografia da mama ocorre com a radiação infravermelha emitida pela superfície a ser analisada e capturada por uma câmera, apresentando como vantagens ser não invasivo e não expor o pacientes a níveis de radiação ionizantes. Essa técnica se baseia nas mudanças metabólicas resultantes do surgimento de células alteradas no tecido mamário, que por sua vez resultam em modificações de distribuição de calor na mama, conhecido como o processo de angiogênese. Nesse processo, o tumor estimula a criação de novos vasos sanguíneos para sua alimentação. Com mais vasos alimentando o tumor, a temperatura da região se mostra superior, possibilitando assim uma melhor visualização das alterações fisiológicas ocasionadas pela doença (ANDRADE, 2017).

Assim, a termografia é considerada um método capaz de identificar atividade vas-

cular anormal precursora do câncer de mama, sem exposição à radiação ionizante e configurando-se como uma atraente técnica de rastreamento e que pode servir à detecção precoce do câncer de mama [13]. Mas a interpretação visual de imagens mamográficas por vezes é difícil, principalmente nos casos onde as lesões estão longe da superfície da pele, implicando em temperaturas difusas e não pontuais.

3.3 Redes neurais artificiais

As redes neurais ou redes neurais artificiais (RNAs) são modelos matemáticos que se desenvolveram a partir do princípio do conceito básico das redes neurais biológicas. Assim, esses modelos são inspirados no comportamento do cérebro humano, buscando realizar uma simulação do funcionamento do cérebro (CRUZ, 2017).

Além disso, RNAs podem ser descritas como sistemas paralelos que são constituídos por unidades de processamento simples, no qual são capazes de calcular determinadas funções matemáticas. Essas unidades estão organizadas em uma ou mais camadas, interligadas por conexões, sendo geralmente unidirecionais. Na maioria dos modelos, essas conexões são associadas a pesos sinápticos, onde são armazenados o conhecimento que é representado no modelo e, ponderando a entrada recebida por cada neurônio da rede (MOREIRA et al., 2021).

As RNAs são constituídas por pequenas unidades, chamadas de neurônios artificiais, que são fundamentais para a operação da rede (HAYKIN, 1998). Um neurônio artificial é formado pelos seguintes elementos: um conjunto de *sinapses* ou *conectores*, caracterizadas por um peso w , onde um sinal x_i na entrada da sinapse j conectada ao neurônio k é multiplicado pelo peso sináptico $w_{k,j}$; um *somador* para adicionar os sinais de entrada, ponderados pelas respectivas sinapses do neurônio (equação 3.1); e também por uma função de ativação para limitar a saída de um neurônio (equação 3.2) (HAYKIN, 1998). Pode-se visualizar um exemplo do modelo de um neurônio artificial na Figura 5.

Um neurônio pode ser representado matematicamente da seguinte forma:

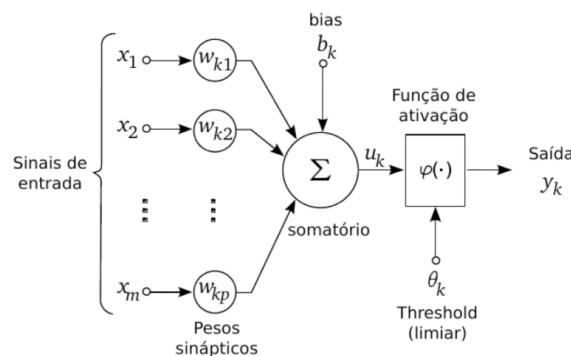
$$u_k = \sum_{j=1}^m w_{k,j} x_i \quad (3.1)$$

e pela equação

$$y_k = \varphi(u_k + b_k) \quad (3.2)$$

onde $x_1, x_2, \dots, x_n$ representam os sinais de entrada; $w_0, w_1, w_2, \dots, w_n$ representam os pesos sinápticos dos sinais de entrada; b_k , é o termo conhecido como *bias* e φ é função de ativação do neurônio.

Figura 5 – Exemplo do modelo de neurônio artificial proposto por McCulloch-Pits.

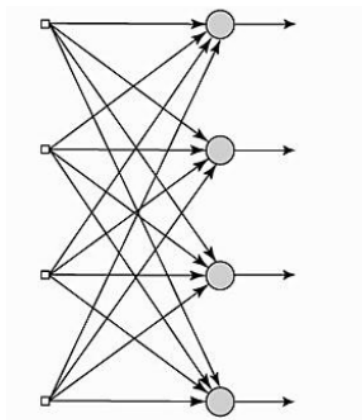


Fonte: Haykin (1998).

As redes neurais são classificadas de acordo com as seguintes arquiteturas:

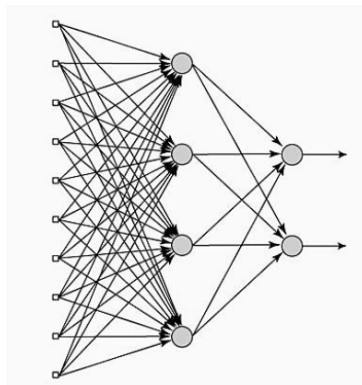
- A Rede Neural de Camada Única, é uma arquitetura com apenas uma camada de entrada e vários neurônios, onde esses últimos projetam o sinal para uma camada de saída (Figura 6).
- A Rede Neural Multicamadas, é uma configuração onde a rede pode ter mais de uma camada, as camadas intermediárias são chamadas de camadas escondidas e os neurônios têm como entrada os valores da saída da camada anterior (Figura 7).
- A Rede Recorrente, é um modelo de arquitetura onde cada neurônio é alimentado pelo sinal de saída de cada neurônio da rede (Figura 8).

Figura 6 – Arquitetura de Rede Neural Artificial de Camada Única.



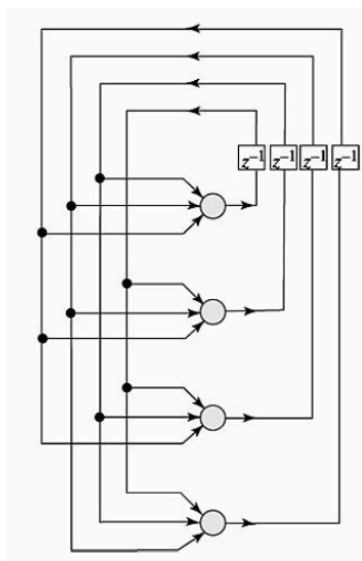
Fonte: Haykin (1998).

Figura 7 – Arquitetura de Rede Neural Artificial de Camada Múltipla.



Fonte: Haykin (1998).

Figura 8 – Arquitetura de Redes Recorrentes.

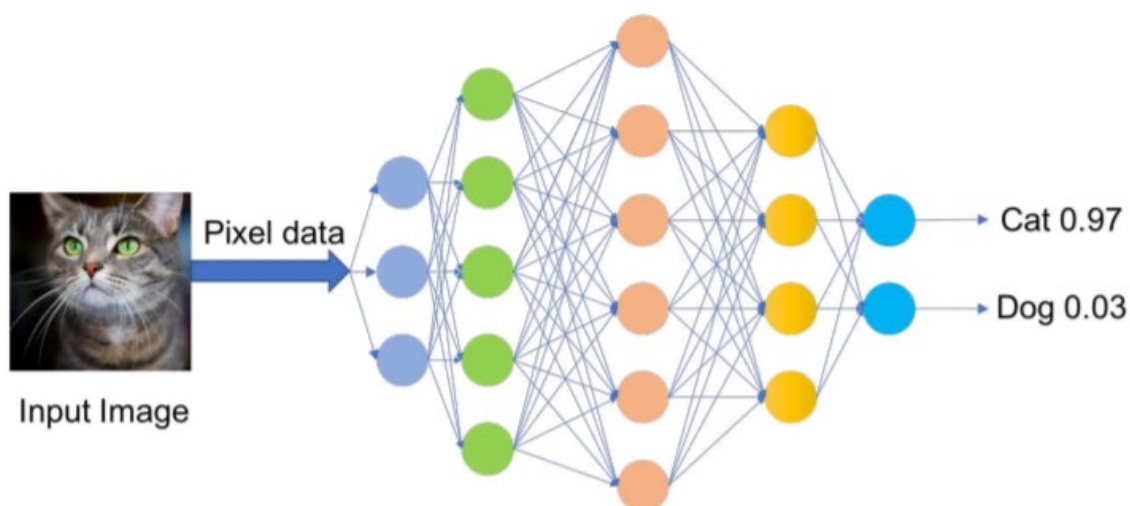


Fonte: Haykin (1998).

3.4 Aprendizado profundo

O Aprendizado Profundo (AP - *Deep Learning*), consiste em uma técnica de aprendizagem de máquina baseada no modelo de Rede Neural Artificial com várias camadas (Figura 9) (CHEN; RAN, 2019). No Aprendizado Profundo os dados de entrada passam pelas camadas em sequência e cada camada executa várias multiplicações de matrizes sobre os dados. A saída de uma camada é geralmente a entrada para a camada seguinte. Após os dados serem processados pela última camada, a saída resultante será uma característica que foi extraída do conjunto de dados ou o resultado final da classificação. Quando o modelo possui muitas camadas em sequência, a rede é conhecida como Rede Neural Profunda (*Deep Neural Network* - DNN) (CHEN; RAN, 2019).

Figura 9 – Exemplo de classificação de imagem usando rede neural profunda.



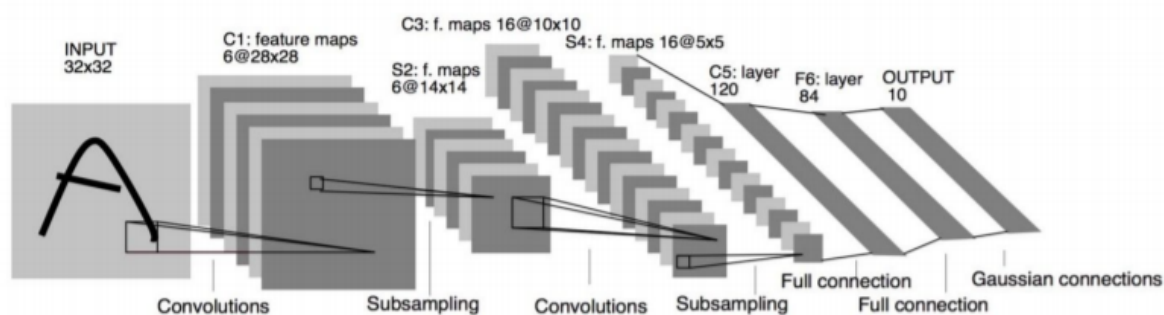
Fonte: Chen e Ran (2019).

Há diversos tipos de Redes Neurais Profundas para diferentes tipos de aplicações. Um caso especial de DNNs é quando as multiplicações de matrizes incluem operações utilizando filtro convolucional. Esse modelo é conhecido como Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs), e é usada para análise de imagem e vídeo. Um outro modelo de DNN bastante usado são as Redes Neurais Recorrentes (*Recurrent Neural Networks* - RNNs), que são utilizadas em problemas envolvendo previsão de séries temporais. As RNNs são formadas por *loops* em suas conexões de camada, com a finalidade de manter o estado e permitir previsões em entradas sequenciais (CHEN; RAN, 2019).

3.4.1 LeNet

LeNet é uma arquitetura de rede convolucional criada por Lecun et al. (Lecun et al., 1998). Essa CNN apresenta uma arquitetura formada por 7 camadas e possui parâmetros ou pesos treináveis, conforme mostrado na Figura 10. Na LeNet as camadas convolucionais são rotuladas Cx, já as camadas de subamostragem são rotuladas Sx e as camadas totalmente conectadas são rotuladas Fx, onde x é o índice da camada. A primeira camada convolucional, C1, é organizada em seis planos, cada plano é um mapa de característica de tamanho 28x28. A camada S2 é a primeira camada de subamostragem, ela possui seis mapas de características de tamanho 14x14. A camada C3 é a segunda camada convolucional, ela contém 16 mapas de características de tamanho 10x10. A segunda camada de subamostragem, S4, possui 16 mapas de características de tamanho 5x5. A camada C5 é a terceira camada convolucional, ela possui 120 mapas de características de tamanho 1x1. A camada F6 contém 84 mapas de características e é totalmente conectada à camada C5. Por fim, a camada de saída é formada por unidades *RBH* Euclidianas, uma para cada classe, cada uma com 84 entradas. Cada unidade *RBH* de saída, vai calcular a distância Euclidiana entre o vetor de entrada e seu vetor de parâmetro. Quanto mais distante a entrada estiver do seu vetor de parâmetro, maior será o valor de saída do *RBH*.

Figura 10 – Arquitetura da Rede Convolucional LeNet.



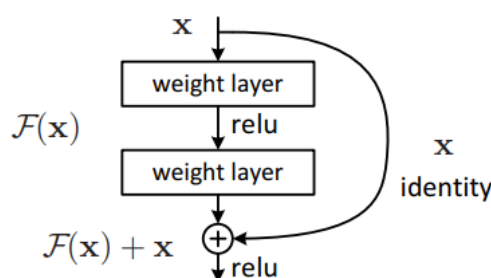
Fonte: Lecun et al. (1998).

3.4.2 ResNet50

ResNet é uma rede residual que tem como objetivo reduzir o problema do *Vanish Gradient*, para isso a ResNet faz uso de conexões de atalho que realizam o que é chamado de mapeamento de identidade, onde suas saídas são adicionadas às saídas das camadas empilhadas, conforme pode-se visualizar na Figura 11 (HE et al., 2016). O mapeamento

de identidade permite que o modelo ignore uma camada CNN se a camada não for mais necessária, ajudando a evitar o problema de *overfitting* ao conjunto de treinamento (THECKEDATH; SEDAMKAR, 2020). A ResNet50 é uma forma mais compacta da rede residual, possuindo 50 camadas. O uso de conexões de atalho apresenta como vantagem o não aumento do número parâmetros e também o não aumento da complexidade computacional para utilizar essa rede (HE et al., 2016).

Figura 11 – Bloco de construção da aprendizagem residual.



Fonte: He et al. (2016).

3.4.3 NASNet mobile

NASNet (*Neural Architecture Search Network*) é uma arquitetura CNN escalável formada por blocos de construção básicos, também chamados de células, que são otimizados utilizando aprendizagem por reforço. Por sua vez, uma célula consiste em apenas algumas operações (como várias convoluções separáveis e *pooling*), onde essas operações são repetidas por várias vezes de acordo com a capacidade necessária da rede. A NASNet Mobile, versão móvel da NASNet, possui 12 células com 5,3 milhões de parâmetros e 564 milhões de operações *multiply-accumulates* (MACs) (SAXEN et al., 2019).

3.4.4 SqueezeNet

A rede neural convolucional SqueezeNet foi desenvolvida com o objetivo de gerar uma rede com boa precisão usando poucos parâmetros. A ideia por trás do uso de poucos parâmetros, seria para facilitar o seu funcionamento em sistemas com pouca memória e também para facilitar a transmissão de informações pela rede.

Segundo Iandola et al. (IANDOLA et al., 2016), uma arquitetura CNN com poucos parâmetros possui três grandes vantagens: promover um treinamento distribuído mais

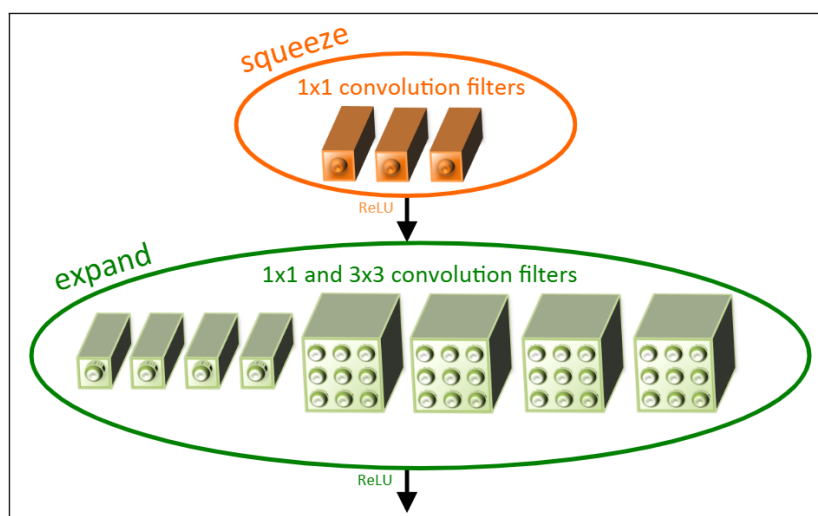
eficiente, pois a diminuição na quantidade de parâmetros usados diminui a sobrecarga na comunicação de rede, assim, pode-se ter mais eficiência ao realizar um treinamento distribuído entre vários servidores; menor sobrecarga (*overhead*) ao exportar novos modelos para o lado do cliente, pois modelos menores irão ocupar menos banda de rede em atualizações a serem realizadas nos clientes, viabilizando assim o uso desse modelo de aprendizagem de máquinas em sistemas embarcados, por exemplo, em carros autônomos; e por fim, viabilizar a implantação em placas FPGA, onde geralmente possuem pouca memória (menos de 10MB) em seu chip, limitando o uso de certos modelos preditivos. Dessa forma, a SqueezeNet busca aliar o uso de uma arquitetura CNN de poucos parâmetros com uma acurácia aceitável comparada à de outros modelos.

Para construir essa arquitetura CNN com poucos parâmetros, foram utilizadas algumas estratégias. A primeira estratégia foi substituir filtros convolucionais 3x3 por filtros 1x1, pois os filtros convolucionais 1x1 possuem 9x menos parâmetros comparados a um filtro 3x3. A segunda estratégia foi diminuir o número de canais de entrada para filtros 3x3, pois a quantidade de parâmetros também é proporcional a este número. A terceira estratégia foi adiar a redução de amostragem (*downsampling*) na rede para que as camadas de convolução tivessem grandes mapas de ativação. A ideia é que grandes mapas de ativação - devido ao atraso na redução da resolução - podem levar a uma maior precisão de classificação (IANDOLA et al., 2016).

Dessa forma, as duas primeiras estratégias tratam de diminuir a quantidade de parâmetros em uma CNN enquanto tentam preservar a precisão. A terceira estratégia trata de maximizar a precisão com uma quantidade limitada de parâmetros. Baseado nessas estratégias, foi descrito um módulo, chamado de *Fire Module*, que representa o bloco de construção para arquitetura CNN.

O *Fire Module* é composto por uma cada convolucional compressora (*squeeze layer*) (que possui apenas filtros 1x1), alimentando uma camada de expansão (*expand layer*) que possui uma mistura de filtros 1x1 e 3x3, conforme pode-se visualizar na Figura 12.

Figura 12 – Organização dos filtros de convolução do *Fire Module* da *CNN Squeezenet*.



Fonte: Iandola et al. (2016).

3.4.5 Inception v3

A arquitetura Inception v3 (Szegedy et al., 2015), foi desenvolvida tendo como base a arquitetura de *inceptions* da GoogLeNet (Szegedy et al., 2014). A GoogLeNet original possui poucos parâmetros, mostrando que o uso de *inceptions* de fato pode reduzir o custo computacional, viabilizando a sua aplicação em cenários de big data, como também permitindo aplicações em contextos com restrição de *hardware*.

Os ganhos da GoogLeNet original são justificados em grande parte ao uso de redução dimensional. Analizando a estrutura computacional das convoluções, é possível substituir filtros espaciais maiores por estruturas de rede com múltiplas camadas, possuindo o mesmo número de parâmetros de entrada, com o objetivo de reduzir o custo computacional.

No modelo Inception v3, foram sugeridas várias técnicas para otimizar a rede com o objetivo de diminuir as restrições e facilitar a adaptação do modelo. Dentre as técnicas sugeridas, estão: convoluções fatoradas, regularização, redução de dimensão e cálculos paralelizados.

3.5 Seleção de atributos

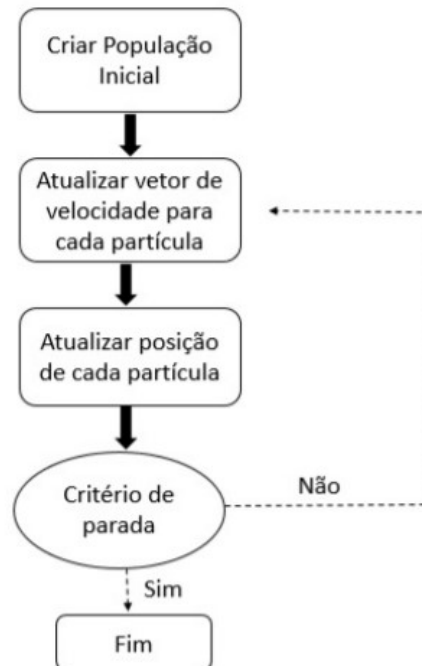
A seleção de atributos tem uma grande importância na fase de pré-processamento. Ela permite a ordenação de atributos segundo algum critério de importância (*ranking*), a redução da dimensionalidade do espaço de busca, a remoção de dados que contenham ruídos entre outros. A seleção de atributos é utilizada principalmente em casos onde a medição de certos atributos é custosa, viabilizando assim o uso de um subconjunto representativo e menor que o conjunto original (LEE, 2005). Como resultado da seleção de atributos, a qualidade dos dados pode ser melhorada e os modelos que serão construídos podem ser mais compreensíveis (LEE, 2005). Além disso, como consequência da melhoria na qualidade dos dados, é possível que a performance dos algoritmos de aprendizado seja melhorada, contribuindo assim para um processo de treinamento mais rápido.

Há diversas técnicas existentes na literatura para selecionar os atributos mais relevantes em um conjunto de dados. No trabalho de Silva et al. (SILVA et al., 2018) foram utilizados dois algoritmos para seleção de atributos: Algoritmo Genético e o PSO. Os dois algoritmos mostraram bons resultados na redução de atributos e no desempenho obtido. Portanto, neste trabalho, baseado nos resultados positivos obtidos em Silva et al. (SILVA et al., 2018) e Neto et al. (NETO et al., 2016), foi utilizada a técnica de seleção de atributos de Otimização por Enxame de Partículas (PSO - *Particle Swarm Optimization*).

- PSO - *Particle Swarm Optimization*

O PSO (*Particle Swarm Optimization*) foi apresentado em 1995 por James Kennedy e Russel Eberhart (KENNEDY; EBERHART, 1995). Essa técnica é baseada no movimento coletivo de um grupo de partículas: o enxame de partículas. Cada partícula deste enxame é movimentado por meio do espaço de busca do problema por duas forças: uma força os atrai com uma magnitude aleatória para a melhor localização já encontrada por ele próprio (*pbest*); e a outra força atrai para a melhor localização encontrada entre alguns ou todos os membros do enxame (*gbest*) (SILVA et al., 2018). A cada nova iteração, a posição e a velocidade de cada partícula são atualizadas até todo o enxame convergir para o melhor resultado (SILVA et al., 2018). A Figura 13 apresenta o fluxograma de execução do PSO.

Figura 13 – Fluxograma relacionado a execução do PSO.



Fonte: Menezes (2007).

Inicialmente, o algoritmo vai gerar N partículas que formarão "enxame"(ou *swarm*) com suas respectivas posições. Neste momento, também podem ser atribuídas velocidades aleatórias para cada partícula (MENEZES, 2007).

Em seguida, o algoritmo vai trabalhar atualizando os vetores e a posição ciclicamente de cada partícula, até atingir algum critério de parada: o número máximo de iterações ou quando encontrar uma partícula com a aptidão desejada (MENEZES, 2007).

O vetor velocidade de cada partícula deve ser atualizado conforme a seguinte equação:

$$v_{k+1}^i = w \cdot v_k^i + c_1 \cdot r_1 \cdot (p^i - x_k^i) + c_2 \cdot r_2 \cdot (p_k^S - x_k^i) \quad (3.3)$$

O parâmetro v_k^i é a velocidade atual da partícula, p^i é a melhor posição encontrada pela partícula i , p_k^S é a melhor posição dentre todas as partículas na iteração k . Já w é um parâmetro que representa a inércia da partícula e controla a sua capacidade de exploração do espaço de soluções, de forma que um alto valor vai determinar uma busca global enquanto um baixo valor vai determinar uma busca local, geralmente esses valores oscilam entre 0,4 e 1,4. Os parâmetros c_1 e c_2 são os parâmetros de confiança e definem o

quanto uma partícula confia em si (c_1) ou no bando (c_2), comumente ambos assumem o valor igual a dois. Por fim, os parâmetros r_1 e r_2 são os números aleatórios compreendidos entre zero e um (SILVA et al., 2018).

Para o cálculo da posição futura de cada partícula no algoritmo é utilizada a seguinte equação:

$$x_{K+1}^i = x_{k+1}^i + v_{k+1}^i \quad (3.4)$$

Onde,

x_{K+1}^i é a posição de cada partícula i na iteração $k + 1$;

v_{k+1}^i é o vetor velocidade da partícula.

3.6 Classificação

A classificação é o processo de realizar atribuição de classes baseados em suas características. A classificação pode ser por meio de uma aprendizagem supervisionada ou por meio de uma aprendizagem não-supervisionada. A aprendizagem supervisionada ocorre quando o classificador utiliza uma base de dados com classes pré-definidas e uma fase de treinamento é realizada antes da classificação, para que os parâmetros que representam cada classe sejam aprendidos. Na aprendizagem não-supervisionada, não há o conhecimento prévio na aplicação do algoritmo de classificação (PEREIRA, 2020).

Existem vários classificadores, alguns dos mais tradicionais são: Bayesiano Ingênuo (*Naive Bayes*), Rede Bayesiana (*Bayes Net*), *Random Tree*, Árvore de Decisão J48, *Random Forest*, Máquinas de Vetores de Suporte (SVM - *Support Vector Machine*) e Rede Perceptron Multicamadas (MLP - *Multilayer Perceptron*).

3.6.1 Bayesiano ingênuo

O classificador Bayesiano Ingênuo (*Naive Bayes*) é um classificador baseado no Teorema de Bayes. Ele é um dos métodos de aprendizagem de máquina mais simples e rápidos. Ele é bastante usado na classificação de documentos de textos e também no diagnóstico clínico. Sua aplicação é indicada para conjuntos de treinamento médios ou grandes. Esse classificador é a forma mais simples de rede bayesiana, na qual todos

os atributos são independentes dado o valor da variável de classe. Isso é chamado de independência condicional (ZHANG, 2004).

3.6.2 Rede bayesiana

Redes Bayesianas (*Bayes Net*), apresentam um modelo baseado em teoria dos grafos para representar uma situação e suas variáveis, e diante disso realizar inferências. Como baseia-se na teoria de grafos, essa técnica usa um conjunto de variáveis e um conjunto de arcos ligando as variáveis, formando um grafo dirigido e sem ciclos. Porém, cada variável possui um número limitado de estados mutualmente exclusivos. Assim, Redes Bayesianas constituem uma forma natural para representação de informações condicionalmente independentes, sendo uma boa solução para trabalhar em domínios onde não há informação suficiente.

3.6.3 Random tree

Random Tree é um classificador baseado em árvore de decisão. Uma árvore de decisão é uma estrutura de dados definida de forma recursiva, onde possui um nó folha correspondente a uma classe ou possui um nó de decisão que contém um teste sobre algum atributo. Para cada resultado do teste há uma aresta para uma subárvore. Cada subárvore tem a mesma estrutura que a árvore (MONARD; BARANAUSKAS, 2003).

3.6.4 Árvore de decisão J48

A árvore de decisão J48 é um algoritmo de código aberto implementado para a ferramenta Weka. Ele é baseado no algoritmo C4.5 *release 8* que gera uma árvore de decisão. O J48 constroi um modelo de árvore de decisão baseado num conjunto de dados de treinamento, sendo esse modelo usado para classificar as instâncias do conjunto de teste. É importante prestar atenção em alguns parâmetros dessa técnica, que podem ser modificados para obter melhores resultados como, por exemplo, o uso de podas na árvore, o número mínimo de instâncias por folha e a construção de árvore binária. Essa árvore é boa para problemas de classificação, pois a árvore de decisão é gerada a partir de um conjunto de dados de treinamento (MACIEL; DE, 1997).

3.6.5 Random forest

Random Forest (ou Floresta Aleatória), é uma técnica que trabalha com a criação de várias árvores de decisão, de maneira aleatória, formando o que podemos enxergar como uma floresta, onde cada árvore será utilizada na escolha do resultado final. Inicialmente, o algoritmo *Random Forest* vai selecionar aleatoriamente algumas amostras dos dados de treino, e não a sua totalidade. Nessa mesma etapa, será usada a técnica de *bootstrap*, que é um método de reamostragem onde as amostras que forem selecionadas podem ser repetidas na seleção. A partir dessa primeira seleção de amostras será construída a primeira árvore de decisão.

Para definição do primeiro nó da árvore, nó raiz, ele utiliza o algoritmo de entropia ou o índice Gini, para escolher a melhor variável para compor o nó raiz. Assim, para definir a variável a ser usada no primeiro nó, o *Random Forest* escolhe de maneira aleatória duas ou mais variáveis, e em seguida realiza os cálculos com base nas amostras selecionadas. A escolha da variável do próximo nó será feita excluindo-se as variáveis já selecionadas anteriormente, e o mesmo processo para a escolha da nova variável se repetirá, até formar a árvore construída até o último nó.

Nas construções das próximas árvores, os processos anteriores serão repetidos. Novas árvores tendem a ser diferentes das árvores já criadas, pois tanto a seleção das amostras, quanto a seleção das variáveis, acontecem de forma aleatória. Quanto mais árvores forem criadas, melhores serão os resultados do modelo, porém, o custo computacional será maior, e também haverá um limiar onde o desempenho não será tão significativo.

Após a criação do modelo com *Random Forest*, podem-se apresentar os novos dados e obter o resultado da previsão. Cada árvore criada vai apresentar um resultado, e em problemas de classificação o resultado que mais vezes foi apresentado será o escolhido (Didádita Tech, 1997).

3.6.6 Máquina de vetores de suporte (SVM)

Uma das técnicas mais utilizadas em problemas envolvendo classificação e reconhecimento de padrões é a técnica de Máquina de Vetor de Suporte (SVM. *Support Vector Machine*). A técnica de SVM foi proposta por Vladimir Vapnik em 1979. A ideia por trás de uma Máquina de Vetor de Suporte é construir um hiperplano com superfície de decisão

de forma a maximizar a separação entre exemplos positivos e negativos. O SVM tem seu funcionamento baseado no mapeamento linear no conjunto de dados em um espaço de alta dimensão, chamado de espaço de características. Nesse espaço, uma superfície linear de decisão, denominada de hiperplano, é construída de modo a separar classes distintas (CORTES; VAPNIK, 1995). Assim, esse hiperplano atua como uma superfície de decisão de forma a maximizar a fronteira de separação entre as classes (HAYKIN, 1998).

O SVM utiliza a função kernel para realizar o mapeamento. Dentre os tipos de SVM, as funções mais utilizadas são: Máquina de Vetor de Suporte com Kernel Polinomial (dada pela Equação 3.5 e a Máquina de Vetor de Suporte com Kernel RBF (função de base radial, dada pela Equação 3.6). Os termos $\mathbf{x}$ e $\mathbf{y}$ são instâncias do conjunto de treinamento. Os parâmetros p e γ deverão ser definidos pelo usuário, onde p define o grau do kernel polinomial, por exemplo, quando $p = 1$ dizemos kernel linear, e γ define a largura do kernel RBF (HAYKIN, 1998).

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^p \quad (3.5)$$

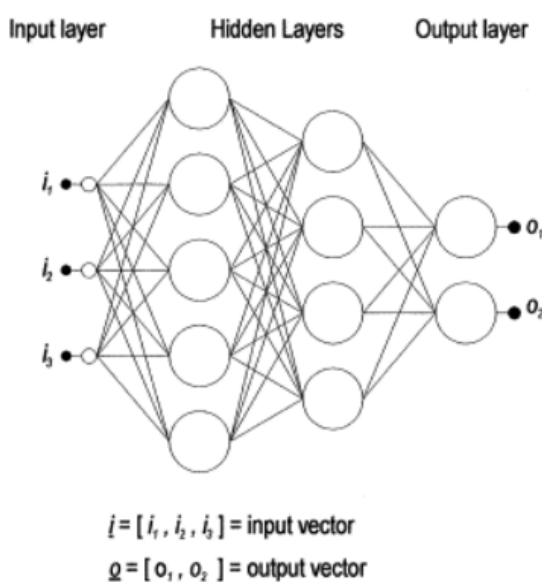
$$K(\mathbf{x}, \mathbf{y}) = \exp(-\gamma(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})) \quad (3.6)$$

3.6.7 Rede perceptron multicamadas (MLP)

Uma rede Perceptron Multicamadas (MLP ou *Multilayer Perceptron*) é um variante do modelo Perceptron original. O MLP possui uma ou mais camadas ocultas entre suas camadas de entrada e saída (Figura 14), os neurônios são organizados em camadas e as conexões são sempre direcionadas das camadas inferiores para as camadas superiores. Os nós são conectados por pesos e sinais de saída que são uma função da soma das entradas para o nó modificado por uma simples função de transferência não linear ou função de ativação. Conforme ilustrado na Figura 14, o Perceptron Multicamadas (MLP) consiste em um sistema de neurônios simples interconectados, também chamado de nós, que é um modelo que representa um mapeamento não-linear entre um vetor de entrada e um vetor de saída (GARDNER; DORLING, 1998). Os nós são conectados por pesos e sinais de saída, esses últimos são representados por uma função de soma das entradas para o nó, modificadas por uma simples transferência não-linear, ou função de

ativação. É a superposição de muitas funções de transferência não-lineares simples que permite ao perceptron multicamadas aproximar funções extremamente não-lineares. Se a função de transferência fosse linear, o perceptron multicamadas só seria capaz de modelar funções lineares. A saída de um nó é escalonada pelo peso de conexão e alimentada para ser uma entrada para os nós na próxima camada do rede. Isso implica uma direção de processamento de informação, portanto, o perceptron multicamadas é conhecido como uma rede neural *feed-forward* (GARDNER; DORLING, 1998). A arquitetura de uma MLP é variável, mas em geral consiste em várias camadas de neurônios. A camada de entrada não desempenha nenhuma função computacional, mas serve apenas para passar o vetor de entrada para a rede. Os termos vetores de entrada e saída referem-se às entradas e saídas da MLP e podem ser representados como vetores únicos, conforme mostrado na Figura 14. Um perceptron multicamadas pode ter uma ou mais camadas ocultas, terminando com uma camada de saída. Perceptrons multicamadas são descritos como totalmente conectados (*fully connected*), com cada nó conectado a todos os nós da camada seguinte e anterior (GARDNER; DORLING, 1998).

Figura 14 – Exemplo de MLP com duas camadas escondidas.



Fonte: Gardner e Dorling (1998).

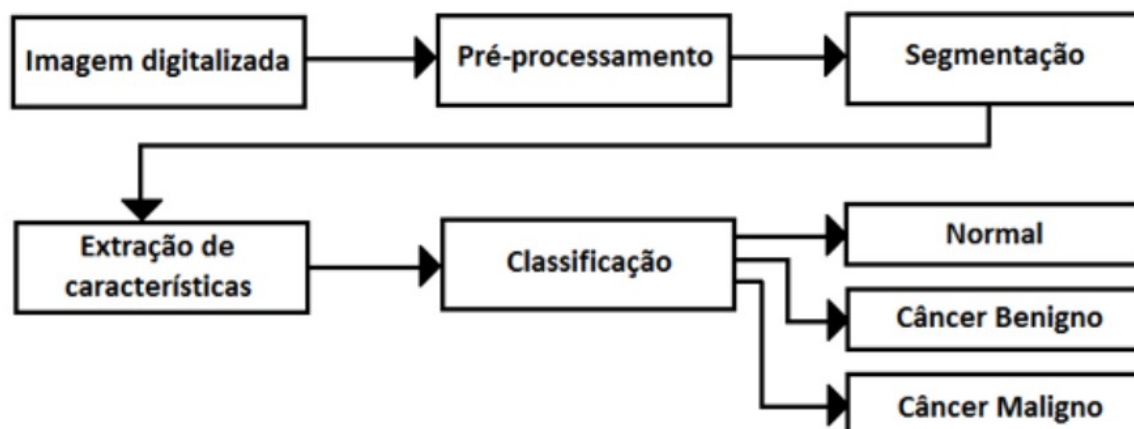
3.7 Sistema de apoio ao diagnóstico

Sistemas de apoio ao diagnóstico, ou diagnóstico auxiliado por computador (CAD), têm sido utilizados como ferramentas de apoio a decisão clínica, mostrando um alto potencial

no auxílio aos especialistas para classificação e detecção de possíveis anormalidades (CRUZ, 2017). Nesse contexto, há duas aplicações básicas para o sistema de apoio ao diagnóstico: auxiliar detecção de lesões por meio da varredura da imagem a partir do local com padrão anormal; e auxiliar no diagnóstico usando quantificação das características presentes na imagem, como também utilizar sua classificação como condizente com padrões normais ou anormais (CRUZ, 2017).

O CAD é bastante utilizado para auxiliar na resolução de diversos problemas na área da saúde (DALVI, 2018). Na literatura, existem vários trabalhos que buscam encontrar soluções para o problema de detecção e classificação de câncer a partir de imagens de mamografia. Esses trabalhos, geralmente, utilizam uma metodologia baseada nas seguintes etapas: pré-processamento, segmentação, extração de características e classificação. Cada etapa dessa consiste na utilização de técnicas de Processamento Digital de Imagens (PDI) e reconhecimento de padrões (DALVI, 2018).

Figura 15 – Exemplo de fluxograma com as etapas seguidas para construção de sistemas auxiliados por computador.



Fonte: Dalvi (2018).

3.8 Considerações

Neste capítulo, foi apresentada a fundamentação teórica deste trabalho. Foram mostrados alguns conceitos relacionados ao câncer de mama, como os principais métodos de diagnósticos e a importância do diagnóstico precoce. Também foi apresentada a técnica de termografia de mama como uma alternativa interessante a ser utilizada no diagnóstico do câncer de mama. Além disso, foram apresentadas, brevemente, as técnicas de Redes

Neurais Artificiais e Aprendizado Profundo. Por fim, foi falado sobre a importância do Sistema de Apoio ao Diagnóstico por Imagem.

4 MATERIAIS E MÉTODOS

Neste trabalho, foram propostas técnicas para extração e seleção de características em imagens de termografia de mama usando redes neurais profundas, como também o uso dessas características para treinamento usando técnicas de aprendizagem de máquina tradicionais. Neste capítulo foi apresentada a metodologia utilizada para o desenvolvimento deste trabalho, assim como a explanação de como foram realizados os experimentos: inicialmente a etapa de extração de atributos, usando técnicas de redes neurais profundas; passando pela etapa de seleção dos melhores atributos; e por fim a etapa de classificação, usando diferentes classificadores. A metodologia foi dividida nas seguintes etapas: descrição da base de dados de imagens termográficas da mama, pré-processamento das imagens (seleção das imagens, extração de atributos, seleção de atributos e balanceamento das classes), classificação das lesões da mama e avaliação de desempenho dos classificadores utilizados. A Figura 16 mostra de forma simplificada o diagrama de bloco do método proposto. A base de dados é uma base de imagens termográficas da mama fornecida pelo Departamento de Engenharia Mecânica da UFPE. No pré-processamento as imagens foram selecionadas para as classes Cisto, Lesão Benigna, Lesão Maligna e Sem Lesão. Em seguida, foram aplicadas técnicas de extração de atributos sobre cada imagem para cada classe. Na etapa seguinte, foram selecionados os melhores atributos usando a técnica de Otimização por Enxame de Partículas (PSO - *Particle Swarm Optimization*). Logo após, foi realizado o balanceamento das classes usando SMOTE. A etapa de classificação foi realizada com diversos classificadores: *Random Tree*, *Random Forest*, *J48*, *Naive Bayes Multinomial*, *Bayes Net*, *SVM* e *MLP*. Por fim, os resultados foram avaliados usando as métricas de Acurácia, Índice Kappa, Sensibilidade, Especificidade, Área Sob Curva ROC e Tempo de Treinamento.

Figura 16 – Diagrama de bloco do método proposto.



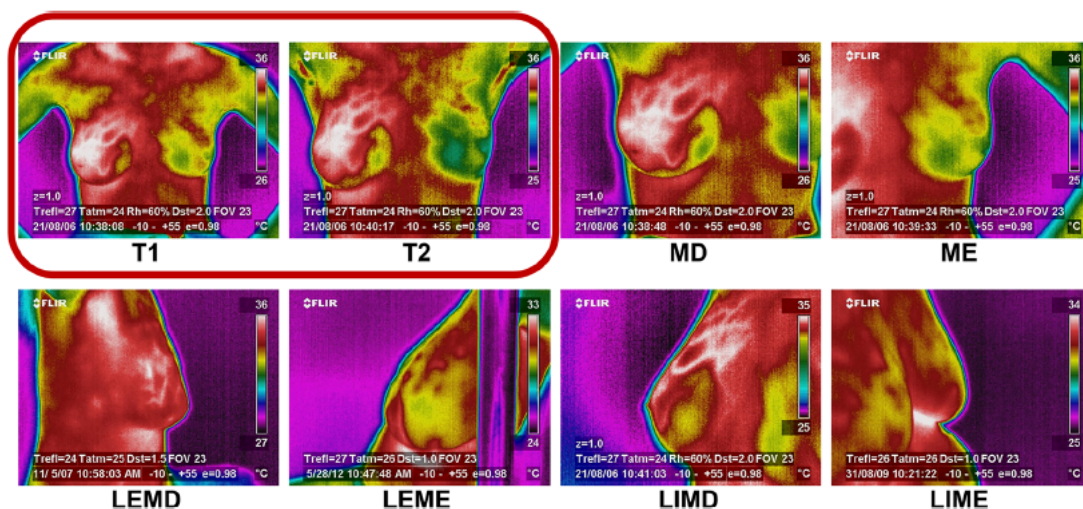
Fonte: o autor (2022).

4.1 Base de dados

A base de dados com imagens de termografia da mama, foi fornecida pelo Grupo de Pesquisa de Engenharia Térmica do Departamento de Engenharia Mecânica (DEMEC) da Universidade Federal de Pernambuco, Brasil. Essas imagens termográficas foram adquiridas no Hospital das Clínicas(HC)-UFPE, usando uma câmera infravermelha modelo FLIR S45 (PEREIRA et al., 2017) (SANTANA et al., 2018). O método adotado no processo de aquisição das imagens foi o estático, no qual a imagem é obtida após a aclimação térmica da paciente à sala de exame. O procedimento de aclimação térmica da paciente teve a duração aproximada de 10 minutos. Durante o exame, a paciente ficou sentada em um aparato mecânico desenvolvido pelo DEMEC/UFPE e instalado no HC/UFPE (SANTANA, 2020). Assim, para cada paciente foram obtidas imagens em oito posições distintas (Figura 17). Foram realizadas duas imagens frontais de cada mama, T1 (frontal com as mãos na cintura) e T2 (frontal com as mãos levantadas segurando uma barra localizada acima da cabeça), bem como três imagens de cada mama em diferentes ângulos: MD (imagem frontal da mama direita), ME (imagem frontal da mama esquerda), LMD (lateral interna da mama direita), LIME (lateral interna da mama esquerda), LEMD (lateral externa da mama direita) e LEME (lateral externa da mama esquerda) (PEREIRA et al., 2017) (SANTANA et al., 2018). Os exemplos de cada imagem sob diferentes ângulos podem ser vistos na Figura 17. Após adquiridas, as imagens foram organizadas de acordo com o tipo de diagnóstico do paciente. Cada paciente foi classificada de acordo com um dos quatro possíveis diagnósticos: lesão cística, lesão benigna, lesão maligna e sem lesão (RODRIGUES et al., 2019). Todos os diagnósticos foram estabelecidos após uma investigação prévia utilizando exames específicos para cada situação: os casos de cistos, o diagnóstico foi confirmado por punção aspirativa com agulha fina (PAAF) ou ultrassonografia; para lesões malignas e benignas foram confirmadas por meio de biópsias; os casos sem lesão foram verificados por meio de exames de mamografia e ultrassonografia, classificando com BI-RADS 1, ou seja, sem achados (SILVA, 2015).

Nesse trabalho, foram usadas imagens de todos os ângulos, divididas entre as quatro classes: lesão benígna (371 imagens), lesão málgna (235 imagens), cisto (219 imagens) e sem lesão (227 imagens). A Tabela 1 mostra a divisão das classes, bem como a atribuição de cada classe a uma *label* numérica: Lesão Benigna = Classe 0,

Figura 17 – Exemplos de imagens termográficas obtidas em diferentes posições.



Fonte: Rodrigues et al. (2019).

Lesão Maligna = Classe 1, Cisto = Classe 2 e Sem Lesão = Classe 3. Conforme pode-se ver na Tabela 1, a quantidade de imagens por classe é desbalanceada. Isso pode gerar resultados tendenciosos durante a etapa de treinamento, onde algumas imagens poderiam ser classificadas mais frequentemente para a classe com mais amostras, desprezando o padrão das outras classes.

Tabela 1 – Distribuição das amostras entre as classes definidas.

Classes	Amostras	Total
0 Lesão Benigna	371	1052
1 Lesão Maligna	235	
2 Cisto	219	
3 Sem Lesão	227	

Fonte: o autor (2022).

4.2 Pré-processamento

Os experimentos foram realizados utilizando todas as imagens de todos os ângulos: T1, T2, MD, ME, LEMD, LEME, LIMD e LIME. Ao todo foram selecionadas 1052 imagens, divididas em quatro classes (Tabela 1). Para referenciar cada imagem com a sua devida classe e sua localização, foi necessário criar um arquivo ".arff" indicando o nome, caminho e classe (0, 1, 2 e 3) ao qual cada imagem pertence. A Figura 18 mostra uma parte deste arquivo que foi criado. Tanto o arquivo ".arff" quanto o diretório onde estão as imagens,

foram selecionados na ferramenta utilizada para realização dos experimentos: o *software* de aprendizagem de máquina Weka.

Figura 18 – Parte do arquivo ".arff" carregado na entrada dos dados, onde é indicado as classes existentes e os caminhos para cada imagem.

```

1 @relation dados_termografia_cancer_de_mama
2 @attribute filename string
3 @attribute class {0,1,2,3}
4 @data
5 Lesao_Benigna\Lesao_Benigna_LEMD_1.jpg,0
6 Lesao_Benigna\Lesao_Benigna_LEMD_2.jpg,0
7 Lesao_Benigna\Lesao_Benigna_LEMD_3.jpg,0
8 Lesao_Benigna\Lesao_Benigna_LEMD_4.jpg,0
9 Lesao_Benigna\Lesao_Benigna_LEMD_5.jpg,0
10 Lesao_Benigna\Lesao_Benigna_LEMD_6.jpg,0
11 Lesao_Benigna\Lesao_Benigna_LEMD_7.jpg,0
12 Lesao_Benigna\Lesao_Benigna_LEMD_8.jpg,0
13 Lesao_Benigna\Lesao_Benigna_LEMD_9.jpg,0
14 Lesao_Benigna\Lesao_Benigna_LEMD_10.jpg,0
15 Lesao_Benigna\Lesao_Benigna_LEMD_11.jpg,0
16 Lesao_Benigna\Lesao_Benigna_LEMD_12.jpg,0
17 Lesao_Benigna\Lesao_Benigna_LEMD_13.jpg,0
18 Lesao_Benigna\Lesao_Benigna_LEMD_14.jpg,0
19 Lesao_Benigna\Lesao_Benigna_LEMD_15.jpg,0
20 Lesao_Benigna\Lesao_Benigna_LEMD_16.jpg,0
21 Lesao_Benigna\Lesao_Benigna_LEMD_17.jpg,0
22 Lesao_Benigna\Lesao_Benigna_LEMD_18.jpg,0
23 Lesao_Benigna\Lesao_Benigna_LEMD_19.jpg,0
24 Lesao_Benigna\Lesao_Benigna_LEMD_20.jpg,0
25 Lesao_Benigna\Lesao_Benigna_LEMD_21.jpg,0
26 Lesao_Benigna\Lesao_Benigna_LEMD_22.jpg,0
27 Lesao_Benigna\Lesao_Benigna_LEMD_23.jpg,0
28 Lesao_Benigna\Lesao_Benigna_LEMD_24.jpg,0
29 Lesao_Benigna\Lesao_Benigna_LEMD_25.jpg,0
30 Lesao_Benigna\Lesao_Benigna_LEMD_26.jpg,0
31 Lesao_Benigna\Lesao_Benigna_LEMD_27.jpg,0
32 Lesao_Benigna\Lesao_Benigna_LEMD_28.jpg,0
33 Lesao_Benigna\Lesao_Benigna_LEMD_29.jpg,0
34 Lesao_Benigna\Lesao_Benigna_LEMD_30.jpg,0
35 Lesao_Benigna\Lesao_Benigna_LEMD_31.jpg,0
36 Lesao_Benigna\Lesao_Benigna_LEMD_32.jpg,0
37 Lesao_Benigna\Lesao_Benigna_LEMD_33.jpg,0
38 Lesao_Benigna\Lesao_Benigna_LEMD_34.jpg,0
39 Lesao_Benigna\Lesao_Benigna_LEMD_35.jpg,0
40 Lesao_Benigna\Lesao_Benigna_LEMD_36.jpg,0
41 Lesao_Benigna\Lesao_Benigna_LEMD_37.jpg,0
42 Lesao_Benigna\Lesao_Benigna_LEMD_38.jpg,0
43 Lesao_Benigna\Lesao_Benigna_LEMD_39.jpg,0
44 Lesao_Benigna\Lesao_Benigna_LEMD_40.jpg,0

```

Fonte: o autor (2022).

Devido ao desbalanceamento entre as classes, foi necessário utilizar a técnica SMOTE (*Synthetic Minority Over-sampling Technique*) para balanceá-las. Entretanto, a aplicação dessa técnica foi realizada na etapa logo após a extração de atributos, porque o desbalanceamento não afeta a extração de atributos. Na verdade, para obter um melhor desempenho na extração de atributos é importante utilizar o conjunto de dados originais.

4.3 Transferência de aprendizagem

Por meio da transferência de aprendizado, é possível aplicar uma Rede Neural Artificial para um problema diferente para o qual ela foi treinada. Isto é, utilizar uma rede pré-treinada em um conjunto de dados para resolver um problema distinto (Rosebrock, A., 2019), seja de uma forma mais rápida, mais eficiente ou que exija um conjunto de treinamento menor (PAN; YANG, 2010).

No contexto da aprendizagem profunda, o uso da transferência de aprendizado é muito importante, principalmente para problemas com conjunto de dados pequeno. Além disso, redes profundas são complexas e possuem muitos parâmetros a serem ajustados durante o seu processo de treinamento, o que exige conjuntos de dados com um grande número de exemplos rotulados para um bom treinamento supervisionado. Entretanto, construir um grande conjunto de dados pode ser custoso e demandar muito tempo, acima de tudo, quando se tratando de aplicações médicas. Ou seja, conseguir um conjunto de dados para um problema médico que seja grande o suficiente para treinar uma rede profunda pode ser inviável.

Basicamente, no contexto da aprendizagem profunda há dois tipos de transferência de aprendizado: a transferência através da extração de atributos e a transferência via ajuste fino (Rosebrock, A., 2019). Na modalidade de extração de atributos, a rede pré-treinada é utilizada como um extrator de atributos arbitrário. Assim, a instância de entrada é propagada pela rede até uma preestabelecida camada. A saída desta camada é obtida como atributos que poderão ser usados para treinar um classificador qualquer (Rosebrock, A., 2019).

Já modalidade de ajuste fino, apenas as camadas totalmente conectadas do modelo são treinadas do zero com o conjunto de dados alvo. As outras camadas são mantidas fixas (ou "congeladas") com os parâmetros pré-treinados (Rosebrock, A., 2019).

Neste trabalho, foi utilizada a modalidade de extração de atributos da transferência de aprendizagem.

4.4 Extração de atributos

Ainda na fase de pré-processamento, na etapa de Extração de Atributos, foram utilizadas técnicas de aprendizagem profunda da biblioteca *Deeplearning4j* do Weka (LANG

et al., 2019). A biblioteca *Deeplearning4j* possui diversas arquiteturas de Redes Neurais Convolucionais (CNN) utilizadas na literatura. Neste trabalho, foram usadas as seguintes técnicas da biblioteca *Deeplearning4j* do Weka: LeNet, ResNet50, NASNet Mobile, SqueezeNet e Inception v3.

Inicialmente, foi escolhida uma das Redes Neurais Convolucionais pré-treinadas disponíveis no Weka *Deeplearning4j*. Para cada rede selecionada, foi indicado o tipo da base pré-treinada a ser utilizada: *MNIST* ou *IMAGENET*. Neste momento de pré-extração de atributos, também foi necessário carregar o arquivo ".arff" que indicava quais as classes existentes - quatro no total (0, 1, 2 e 3) - e os caminhos para cada imagem pertencente a cada classe.

Após a escolha e configuração do modelo, foi executada a extração de atributos. Para cada modelo, foram extraídos quantidades diferentes de atributos, conforme pode-se ver na Tabela 2.

Tabela 2 – Modelos de redes neurais profundas utilizados e o número de atributos extraídos por cada rede.

Modelo	Base de Pré-Treinamento	Nº de Atributos Extraídos
LeNet	MNIST	500
ResNet50	IMAGENET	2048
NASNetMobile	IMAGENET	1056
SqueezeNet	IMAGENET	1000
Inception V3	IMAGENET	2048

Fonte: o autor (2022).

4.5 Seleção de atributos

Nessa etapa, foi realizada a seleção dos atributos mais revelantes em cada conjunto de atributos extraídos com as redes profundas. Foi utilizada a técnica de Otimização por Enxame de Partículas (*Particle Swarm Optimization*, PSO) para realização dessa seleção. O objetivo em realizar a seleção de atributos, foi fazer uma comparação sobre o desempenho dos modelos antes e após a seleção de atributos. Neste caso, foram avaliadas as mesmas métricas: tempo de treinamento, acurácia, índice kappa, sensibilidade, especificidade e área sob curva ROC. A ideia é que essa avaliação, sobre utilizar a técnica de seleção de atributos com PSO, sirva de base para a tomada de decisão sobre qual o melhor modelo a ser utilizado. Na Tabela 3 pode-se visualizar a comparação entre a quantidade de atributos por rede neural profunda, antes e depois de ser aplicado o PSO.

Tabela 3 – Números de atributos antes e depois da seleção de atributos por cada modelo de rede neural profunda.

Modelo	Nº de Atributos Extraídos	Nº de Atributos Seleccionados
LeNet	500	117
ResNet50	2048	481
NASNetMobile	1056	15
SqueezeNet	1000	159
Inception V3	2048	616

Fonte: o autor (2022).

4.6 Balanceamento das classes

Conforme pode-se visualizar na Tabela 1, há um desbalanceamento entre as classes de imagens termográficas da mama. A Classe 0 (Lesão Benigna) possui 371 imagens, a Classe 1 (Lesão Maligna) possui 235 imagens, a Classe 2 (Cisto) possui 219 imagens e a Classe 3 (Sem Lesão) possui 227 imagens. Antes da etapa de treinamento, é importante que as classes estejam balanceadas para que o classificador tenha um melhor aprendizado diante da equivalência entre o número de amostras por classe. Caso esse balanceamento não fosse realizado, o treinamento ficaria tendencioso para a classe com o maior número de amostras. Para este trabalho, foi utilizada a técnica SMOTE para balanceamento entre as classes de imagens termográficas.

O *Synthetic Minority Oversampling TEchnique* (SMOTE) é um método de geração de instâncias sintéticas para diminuir o efeito do desequilíbrio de classes em bases desbalanceadas (BLAGUS; LUSA, 2013). A técnica SMOTE foi primeiramente proposta por Chawla et al. (2002) (KOVÁCS et al., 2020), onde foi mostrado que o SMOTE usa a combinação linear de duas instâncias da classe minoritária para construir uma nova instância sintética. Por meio dessa técnica, considerando x uma instância da classe minoritária, uma nova instância sintética s é obtida por meio da seguinte equação:

$$s = x + u \cdot (x_R - x), \quad (4.1)$$

onde u é um número aleatório gerado dentro do intervalo $[0, 1]$; e x_R é escolhido aleatoriamente entre as k instâncias da classe minoritária vizinhas de x . Geralmente, é utilizado $k = 5$ (BLAGUS; LUSA, 2013; KOVÁCS et al., 2020). Desta maneira, a instância s , dada pela Equação 4.1, se encontrará em algum ponto aleatório ao longo do segmento de linha

entre x e x_R , isso faz com que a região de decisão da classe minoritária se torne mais geral (KOVÁCS et al., 2020).

4.7 Treinamento e classificação

Após o processo de extração de atributos, seleção de atributos e balanceamento das classes, foi iniciado o processo de treinamento e classificação com o objetivo de observar o desempenho de diferentes técnicas de aprendizagem de máquina.

Neste trabalho foram realizadas duas abordagens: uma utilizando o conjunto de atributos completo que foram extraídos; e outra utilizando o conjunto de atributos reduzidos que foram selecionados usando PSO. Assim, o processo de treinamento e classificação usou como entrada, para treinar os modelos, os atributos resultantes das duas abordagens (com PSO e sem PSO). Primeiro foi utilizada a abordagem sem PSO e depois a abordagem com PSO. Por conseguinte, foram utilizados algoritmos de classificação tradicionais da área de aprendizagem de máquina com o propósito de comparar a classificação e detecção de lesões na mama usando imagens termográficas. Foram realizados experimentos com dezenove configurações de diferentes classificadores presentes no Weka: *Naive Bayes*, *Bayes Net*, *Random Tree*, *J48*, *Random Forest* com configuração de 10, 20, 50 e 100 árvores, SVM (Máquina de Vetor de Suporte) com Kernel Polinomial com graus 1, 2, 3, 4 e 5, SVM com Kernel Gaussiano (RBF) com o parâmetro γ (gamma) igual a 0,01; 0,25; e 0,5, e MLP com 20, 50 e 100 neurônios.

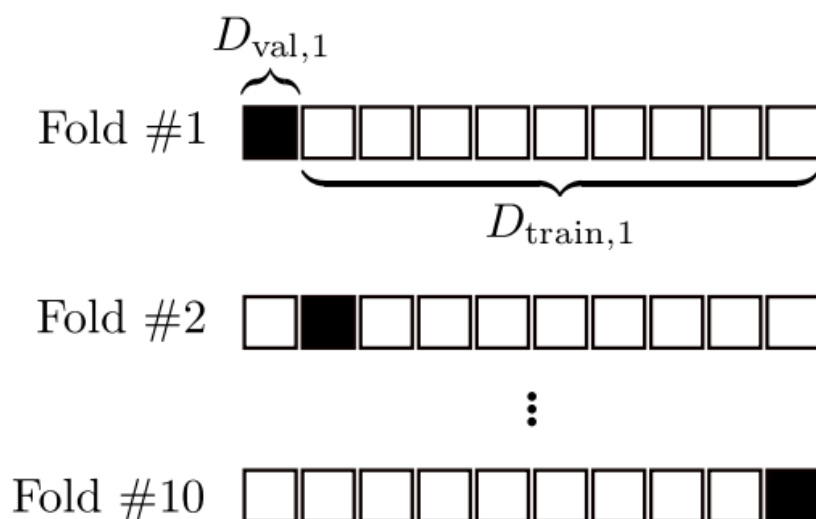
- Tipo do experimento e número de iterações

O Tipo do Experimento refere-se à forma como a amostra foi explorada ao longo do treinamento. Foi escolhida a técnica de validação cruzada K *folds* com o número de 10 *folds*. O K *folds* é uma técnica computacional intensiva onde são utilizadas todas as amostras disponíveis como amostras de treinamento e teste.

Na validação cruzada k -*fold*, a base de dados é aleatoriamente dividida em k subconjuntos disjuntos de tamanho aproximadamente iguais (BERRAR, 2018). Os *folds* referem-se ao número de subconjuntos resultantes. O modelo é treinado usando $k - 1$ subconjuntos, deixando o subconjunto restante, chamado de conjunto de validação, para avaliar o desempenho do modelo. Em cada repetição, um subconjunto diferente é escolhido como conjunto

de validação. Esse procedimento é repetido até que cada um dos subconjuntos k tenha servido como conjunto de validação. A média entre todas as medições de desempenho em todos os k conjuntos de validação é o desempenho da validação cruzada (BERRAR, 2018). Na Figura 19, pode-se ver o processo de validação cruzada para *10-fold*, com $k = 10$. No primeiro *fold*, o subconjunto de validação é representado por $D_{val,1}$, enquanto o restante dos nove subconjuntos, subconjuntos de treinamento, são representados por $D_{train,1}$. No segundo *fold*, o subconjunto de validação será o segundo bloco e os restantes serão os de treinamento. E assim, ao decorrer dos próximos *folds*, o subconjunto de validação vai sucessivamente mudando de bloco até que todos os subconjuntos seja usado como um subconjunto de validação.

Figura 19 – Validação cruzada para *10-fold*. A base de dados é aleatoriamente dividida em dez conjuntos diferentes, cada um com aproximadamente 10% dos dados. O modelo é treinado sobre o conjunto de treinamento e em seguida ele é aplicado ao conjunto de validação (BERRAR, 2018)



Fonte: o autor (2022).

O número de iterações, refere-se ao número de repetições de treinamento para cada classificador. Nesse caso, com a finalidade em obter resultados estatisticamente significativos, o número de repetições utilizado foi de 30. Assim, cada classificador foi executado 30 vezes, e em cada iteração dessa foi aplicada a validação cruzada para 10 *folds*, totalizando 300 classificações.

4.8 Softwares utilizados

Neste trabalho, foram utilizados os *softwares* Weka, LibreOffice e SciDAVis. O Weka (*Waikato Enviroment for Knowledge Analysis*), desenvolvido pela Universidade de Waikato na Nova Zelândia, é um *software* que apresenta uma coleção de algoritmos de aprendizagem de máquina e ferramentas para pré-processamento de dados. É uma ferramenta desenvolvida em Java e, por ser um *software* de código aberto, está sob os termos da GNU General Plubic License. Além disso, o Weka pode ser executado nos sistemas operacionais Windows, Linux e Mac OS (WITTEN; FRANK, 2005). Neste trabalho, foi utilizada a versão 3.8 do Weka para extrair os atributos das imagens termográficas, como também para realizar o treinamento usando diferentes classificadores. Na etapa de extração de atributos, foram utilizadas diferentes técnicas de redes neurais profundas por meio da biblioteca WekaDeepLearning4j (LANG et al., 2019).

O SciDAVis (*Scientific Data Analysis and Visualization*) é um *software* de código aberto e multiplataforma para plotagem interativa de gráficos científicos e análise de dados. Ele apresenta uma interface gráfica fácil e intuitiva com diversos modos de construção de gráficos e cálculos estatísticos (SciDAVis, 2021). Nesse trabalho, foram utilizados os recursos do SciDAVis de: Análise Estatística por Coluna e Gráficos Boxplot. A Análise Estatística por Coluna, mostrou diversas métricas estatísticas, como a Média e Desvio Padrão. Já os *Box Plots* foram gerados para comparar os valores de Acurácia e Índice Kappa.

Para fazer a leitura das planilhas geradas com os resultados dos experimentos de classificação e treinamento, foi utilizado o *software* LibreOffice Calc do conjunto de aplicativos de escritório LibreOffice (LibreOffice, 2021).

4.9 Avaliação de desempenho

Nesta etapa de avaliação de desempenho, foi escolhido realizar uma avaliação multidimensional dos modelos utilizados. Foram escolhidas diversas métricas para avaliar o desempenho dos resultados das classificações realizadas no *software* Weka: a acurácia, o índice kappa, a sensibilidade, a especificidade, a área sob a curva ROC e o tempo de treinamento (ms). Essas métricas são representadas pelas seguintes equações, onde VP,

VN, FN e FP representam respectivamente: Verdadeiros Positivos, Verdadeiros Negativos, Falsos Negativos e Falsos Positivos.

A sensibilidade é a capacidade de um exame identificar a doença em pacientes que realmente estão doentes. Assim, a sensibilidade é a probabilidade de uma paciente ser corretamente diagnosticada quando ela estiver doente (MOGHBEL; MASHOHOR, 2013; COMMOWICK et al., 2018; BARBOSA et al., 2021b; BARBOSA et al., 2021a; GOMES et al., 2020; LIMA; SILVA-FILHO; SANTOS, 2014; FREITAS et al., 2022; SOUZA et al., 2021; SILVA et al., 2020; SANTANA et al., 2022; GOMES et al., 2021). A seguinte equação representa a sensibilidade:

$$Sensibilidade = \frac{VP}{VP + FN}. \quad (4.2)$$

Já a especificidade, é a probabilidade de um exame determinar que a pessoa está sem a doença quando ela realmente está sadia (MOGHBEL; MASHOHOR, 2013; COMMOWICK et al., 2018; BARBOSA et al., 2021b; BARBOSA et al., 2021a; GOMES et al., 2020; LIMA; SILVA-FILHO; SANTOS, 2014; FREITAS et al., 2022; SOUZA et al., 2021; SILVA et al., 2020; SANTANA et al., 2022; GOMES et al., 2021). A especificidade é definida pela seguinte equação:

$$Especificidade = \frac{VN}{VN + FP}. \quad (4.3)$$

A acurácia avalia a capacidade de um exame diagnosticar como doentes, as pessoas que realmente possuem a doença e, diagnosticar como sadias as pessoas que realmente estão livres da doença (MOGHBEL; MASHOHOR, 2013; COMMOWICK et al., 2018; BARBOSA et al., 2021b; BARBOSA et al., 2021a; GOMES et al., 2020; LIMA; SILVA-FILHO; SANTOS, 2014; FREITAS et al., 2022; SOUZA et al., 2021; SILVA et al., 2020; SANTANA et al., 2022; GOMES et al., 2021). A acurácia é dada pela seguinte equação:

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}. \quad (4.4)$$

O índice kappa é um método estatístico para avaliar o nível de concordância ou reprodutibilidade entre dois conjuntos de dados (BARBOSA et al., 2021b; BARBOSA et al., 2021a; GOMES et al., 2020; LIMA; SILVA-FILHO; SANTOS, 2014; FREITAS et al., 2022;

SOUZA et al., 2021; SILVA et al., 2020; SANTANA et al., 2022; GOMES et al., 2021). Ele assume valores dentro do intervalo $[0, 1]$, conforme pode ser visto na Tabela 4 onde é mostrado cada subintervalo e sua respectiva classificação de desempenho.

Tabela 4 – Intervalos de valores para a métrica de índice kappa e sua classificação de desempenho correspondente.

Kappa	Desempenho
$\kappa < 0,0$	Muito Ruim
$0,0 < \kappa \leq 0,2$	Ruim
$0,2 < \kappa \leq 0,4$	Razoável
$0,4 < \kappa \leq 0,6$	Bom
$0,6 < \kappa \leq 0,8$	Muito Bom
$0,8 < \kappa \leq 1,0$	Excelente

Fonte: Azevedo, Oliveira e Vasconcelos (2015), Santana et al. (2018).

O kappa pode ser representado pela seguinte equação:

$$\kappa = \frac{\rho_o - \rho_e}{1 - \rho_e}, \quad (4.5)$$

O ρ_o é a concordância observada, que é a mesma da acurácia (Eq. 4.4), e ρ_e é a concordância esperada, definida da seguinte forma:

$$\rho_e = \frac{(VP + FP)(VP + FN) + (FN + VN)(FP + VN)}{(VP + FP + FN + VN)^2}. \quad (4.6)$$

A curva ROC é construída em um plano cartesiano, onde no eixo Y encontra-se a sensibilidade e, no eixo X encontra-se 1 menos a especificidade (1-E). A curva ROC pode descrever quantitativamente o desempenho de um teste de diagnóstico, onde para cada ponto de corte utilizado pelo teste, são calculadas a sensibilidade e a especificidade, em seguida este ponto é inserido no gráfico. A união desses pontos forma a curva ROC (LOPES et al., 2014). A área sob a curva ROC representa a acurácia ou performance global de teste, pois ela leva em consideração todos os valores de sensibilidade e especificidade para cada valor da variável do teste. Quanto maior o poder do teste em discriminar os indivíduos doentes e não doentes, mais a curva se aproximará do canto superior esquerdo, no ponto que representa a sensibilidade e 1-especificidade do melhor valor de corte. Assim, quanto melhor for o teste, mais a área sob a curva ROC se aproximará de 1 (LOPES et al., 2014).

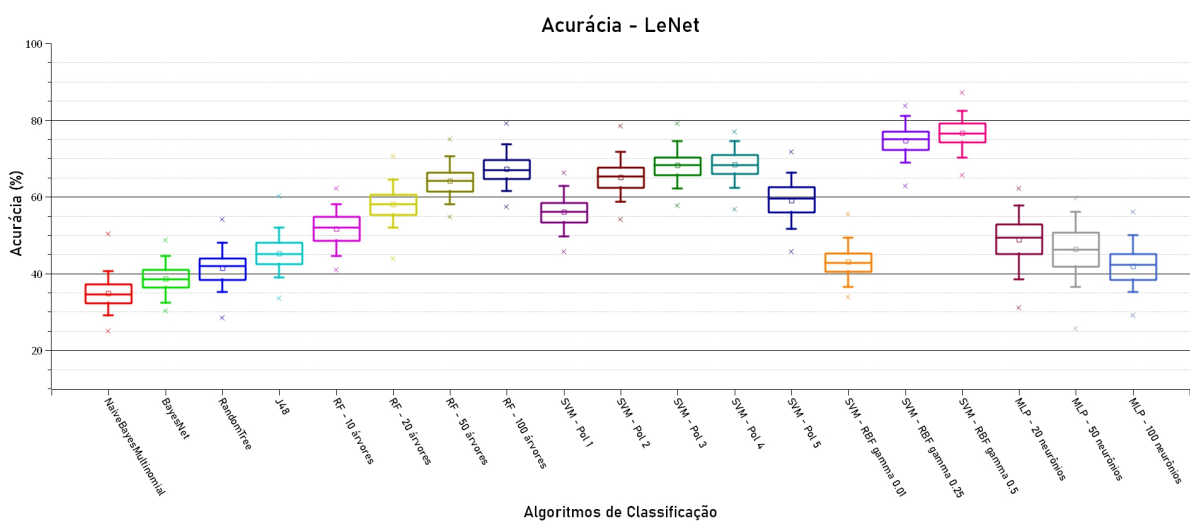
5 RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados e discutidos os resultados dos experimentos realizados. Esse capítulo foi dividido em três sessões principais: análise da classificação, análise da classificação após a seleção de atributos com PSO e, por fim, comparação dos resultados antes e após a seleção de atributos. A ideia foi analisar os desempenhos das redes convolucionais combinadas com diferentes classificadores, observando o comportamento antes e após a utilização da técnica de PSO para selecionar os atributos mais relevantes.

5.1 Análise da classificação

Cada classificador foi avaliado em 30 rodadas utilizando a validação cruzada com 10 *folds*. A Tabela 5 e a Tabela 6 mostram os resultados dos classificadores para a Rede Convolucional *LeNet*. De acordo com a Tabela 5, para a *LeNet*, foram obtidos os seguintes resultados: o classificador SVM com Kernel RBF $\gamma = 0,5$ obteve o melhor desempenho para Acurácia (Figura 20) com 76,65% e desvio padrão de 3,80, Índice Kappa (Figura 21) com 0,69 e desvio padrão de 0,05, e para Sensibilidade de 84,97% com desvio padrão de 6,70; para a Especificidade, o melhor classificador foi o *Bayes Net* com 91,50% e 2,90 de desvio padrão.

Figura 20 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN *LeNet*.



Fonte: o autor (2022).

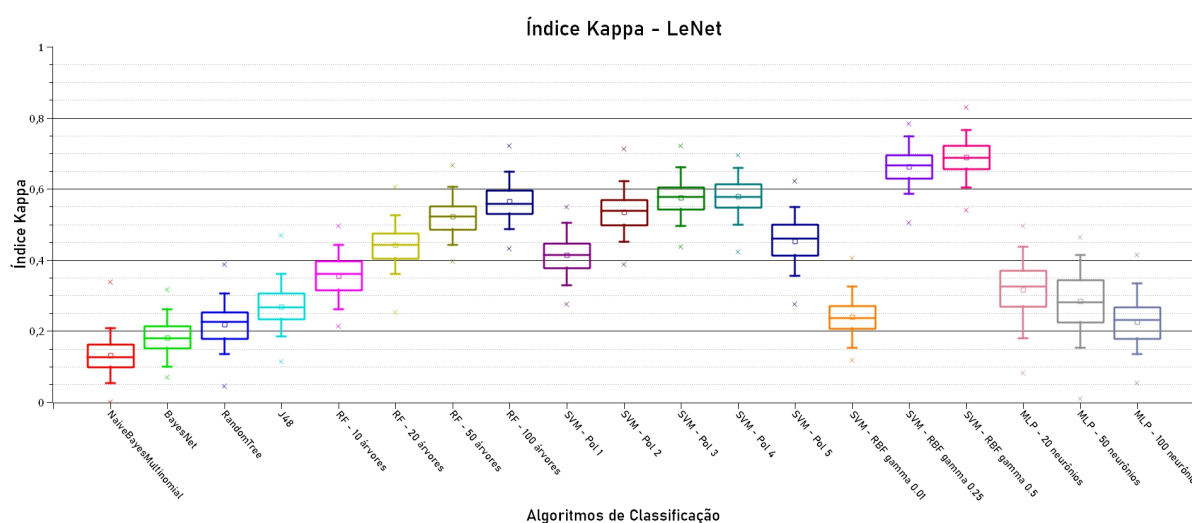
Na Tabela 6 são mostrados os resultados, ainda para a *LeNet*, da Área Sob Curva ROC e Tempo de Treinamento (em milissegundos). O classificador com melhor desempenho

Tabela 5 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *LeNet*. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

Classificador	Rede Neural Convolutacional LeNet			
	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	34,85 ± 3,56	0,13 ± 0,05	36,27 ± 6,90	74,32 ± 4,46
Bayes Net	38,56 ± 3,54	0,18 ± 0,05	11,40 ± 5,19	91,50 ± 2,90
Random Tree	41,42 ± 4,09	0,22 ± 0,05	32,23 ± 7,47	78,85 ± 3,92
J48	45,18 ± 4,03	0,27 ± 0,05	34,26 ± 7,83	79,10 ± 4,14
RF - 10 árvores	51,68 ± 4,19	0,36 ± 0,06	42,99 ± 7,85	77,65 ± 4,35
RF - 20 árvores	58,11 ± 3,99	0,44 ± 0,05	42,05 ± 8,18	83,18 ± 3,73
RF - 50 árvores	64,21 ± 3,73	0,52 ± 0,05	42,22 ± 7,90	87,74 ± 3,46
RF - 100 árvores	67,31 ± 3,70	0,56 ± 0,05	42,26 ± 7,83	90,10 ± 2,90
SVM - Polinômio 1º	56,02 ± 3,63	0,41 ± 0,05	44,45 ± 8,03	80,32 ± 3,95
SVM - Polinômio 2º	65,13 ± 3,79	0,54 ± 0,05	53,00 ± 7,91	83,71 ± 3,56
SVM - Polinômio 3º	68,21 ± 3,70	0,58 ± 0,05	52,33 ± 7,87	86,24 ± 3,37
SVM - Polinômio 4º	68,39 ± 3,65	0,58 ± 0,05	50,58 ± 7,99	87,02 ± 3,71
SVM - Polinômio 5º	59,01 ± 4,50	0,45 ± 0,06	34,63 ± 13,28	86,79 ± 7,08
SVM - RBF $\gamma = 0,01$	43,01 ± 3,70	0,24 ± 0,05	20,67 ± 6,27	88,69 ± 3,47
SVM - RBF $\gamma = 0,25$	74,73 ± 3,69	0,66 ± 0,05	66,76 ± 7,36	83,90 ± 3,78
SVM - RBF $\gamma = 0,5$	76,65 ± 3,80	0,69 ± 0,05	84,97 ± 6,70	76,23 ± 4,57
MLP - 20 neurônios	48,78 ± 5,89	0,32 ± 0,08	30,04 ± 16,91	84,00 ± 14,34
MLP - 50 neurônios	46,37 ± 6,01	0,28 ± 0,08	29,42 ± 16,55	85,27 ± 14,98
MLP - 100 neurônios	41,96 ± 4,97	0,23 ± 0,06	30,81 ± 17,72	82,31 ± 17,24

Fonte: o autor (2022).

Figura 21 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *LeNet*.



Fonte: o autor (2022).

para Área Sob Curva ROC foi o *Random Forest* de 100 árvores, com 0,78 e desvio padrão de 0,04. Já para o Tempo de Treinamento, o *Random Tree* obteve o menor tempo com 25,72 ms e desvio padrão de 7,90.

Tabela 6 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *LeNet*. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.

Classificador	Rede Neural Convolutacional <i>LeNet</i>	
	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,58 ± 0,05	4,64 ± 5,68
Bayes Net	0,50 ± 0,05	103,82 ± 68,48
Random Tree	0,56 ± 0,04	25,72 ± 7,90
J48	0,57 ± 0,05	1858,11 ± 618,67
RF - 10 árvores	0,65 ± 0,05	141,56 ± 43,90
RF - 20 árvores	0,69 ± 0,04	311,40 ± 9,14
RF - 50 árvores	0,75 ± 0,04	795,67 ± 180,33
RF - 100 árvores	0,78 ± 0,04	1423,31 ± 586,27
SVM - Polinômio 1º	0,66 ± 0,05	2716,79 ± 975,02
SVM - Polinômio 2º	0,71 ± 0,04	7256,84 ± 2859,18
SVM - Polinômio 3º	0,72 ± 0,05	6632,34 ± 2535,88
SVM - Polinômio 4º	0,72 ± 0,04	5998,14 ± 2165,32
SVM - Polinômio 5º	0,67 ± 0,05	1601,20 ± 382,21
SVM - RBF $\gamma = 0,01$	0,62 ± 0,05	1812,77 ± 431,88
SVM - RBF $\gamma = 0,25$	0,77 ± 0,04	1888,26 ± 653,92
SVM - RBF $\gamma = 0,5$	0,81 ± 0,04	1700,04 ± 568,86
MLP - 20 neurônios	0,61 ± 0,06	140908,07 ± 41624,42
MLP - 50 neurônios	0,61 ± 0,06	353052,28 ± 107009,53
MLP - 100 neurônios	0,61 ± 0,05	814353,27 ± 323036,35

Fonte: o autor (2022).

As Tabelas 7 e 8 mostram os resultados para a CNN *NASNetMobile*. O SVM polinomial de grau 4 obteve o melhor desempenho para a Acurácia (Figura 22), com 73,81% e desvio padrão de 3,36. Para o Índice Kappa (Figura 23), os melhores resultados foram para SVM polinomial de grau 4 e o SVM com kernel RBF $\gamma = 0,25$, ambos classificadores com taxa de 0,65 e desvio padrão de 0,05. O classificador SVM com kernel RBF $\gamma = 0,5$ apresentou o melhor desempenho para a Sensibilidade, com taxa de 98,16% e desvio padrão de 2,20. Para Especificidade, o Bayes Net teve a melhor taxa com 93,93% e desvio padrão de 14,99.

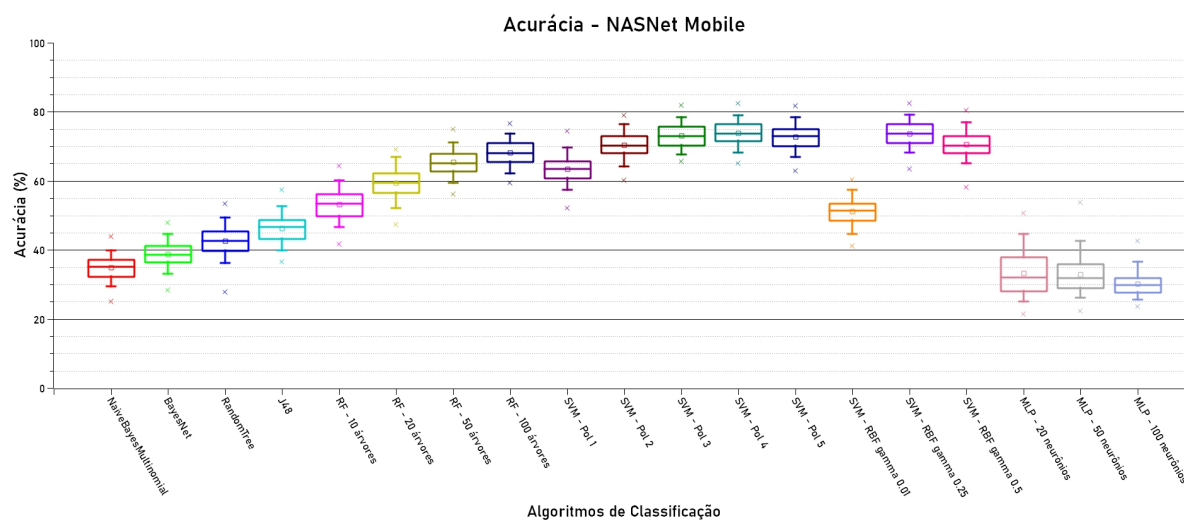
Na Tabela 8, o melhor classificador foi SVM com Kernel RBF $\gamma = 0,5$ que obteve o desempenho de 0,80 e desvio padrão de 0,03 para a Área Sob Curva ROC. Já para o Tempo de Treinamento em milissegundo, o *Random Tree* obteve o menor tempo com 0,04

Tabela 7 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *NASNetMobile*. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

Classificador	Rede Neural Convolutacional <i>NASNetMobile</i>			
	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	34,92 ± 3,34	0,13 ± 0,04	17,06 ± 4,71	89,80 ± 4,94
Bayes Net	38,74 ± 3,63	0,18 ± 0,05	17,66 ± 5,66	93,93 ± 2,61
Random Tree	42,52 ± 4,04	0,23 ± 0,05	32,56 ± 7,87	79,00 ± 3,75
J48	46,22 ± 3,83	0,28 ± 0,05	35,73 ± 7,97	79,33 ± 4,08
RF - 10 árvores	53,17 ± 4,16	0,38 ± 0,06	43,13 ± 8,24	79,21 ± 4,12
RF - 20 árvores	59,37 ± 4,15	0,46 ± 0,06	41,79 ± 8,26	84,47 ± 3,65
RF - 50 árvores	65,36 ± 3,63	0,54 ± 0,05	40,98 ± 8,35	89,06 ± 2,94
RF - 100 árvores	68,19 ± 3,42	0,58 ± 0,05	40,47 ± 7,96	91,27 ± 2,67
SVM - Polinômio 1º	63,35 ± 3,65	0,51 ± 0,05	45,98 ± 8,02	82,67 ± 3,55
SVM - Polinômio 2º	70,33 ± 3,57	0,60 ± 0,05	50,45 ± 8,27	86,65 ± 3,22
SVM - Polinômio 3º	73,04 ± 3,35	0,64 ± 0,04	51,33 ± 8,49	88,71 ± 2,80
SVM - Polinômio 4º	73,81 ± 3,36	0,65 ± 0,05	51,04 ± 8,86	89,76 ± 2,93
SVM - Polinômio 5º	72,77 ± 3,43	0,64 ± 0,05	49,71 ± 9,91	89,73 ± 3,41
SVM - RBF $\gamma = 0,01$	51,11 ± 3,66	0,35 ± 0,05	24,12 ± 6,64	91,86 ± 2,82
SVM - RBF $\gamma = 0,25$	73,63 ± 3,49	0,65 ± 0,05	80,12 ± 6,84	74,42 ± 4,37
SVM - RBF $\gamma = 0,5$	70,60 ± 3,57	0,61 ± 0,05	98,16 ± 2,20	61,64 ± 4,71
MLP - 20 neurônios	33,28 ± 6,30	0,11 ± 0,08	23,28 ± 33,01	82,14 ± 33,65
MLP - 50 neurônios	32,88 ± 5,14	0,11 ± 0,07	24,62 ± 31,43	82,04 ± 32,87
MLP - 100 neurônios	30,24 ± 3,29	0,07 ± 0,04	27,44 ± 34,48	79,05 ± 36,29

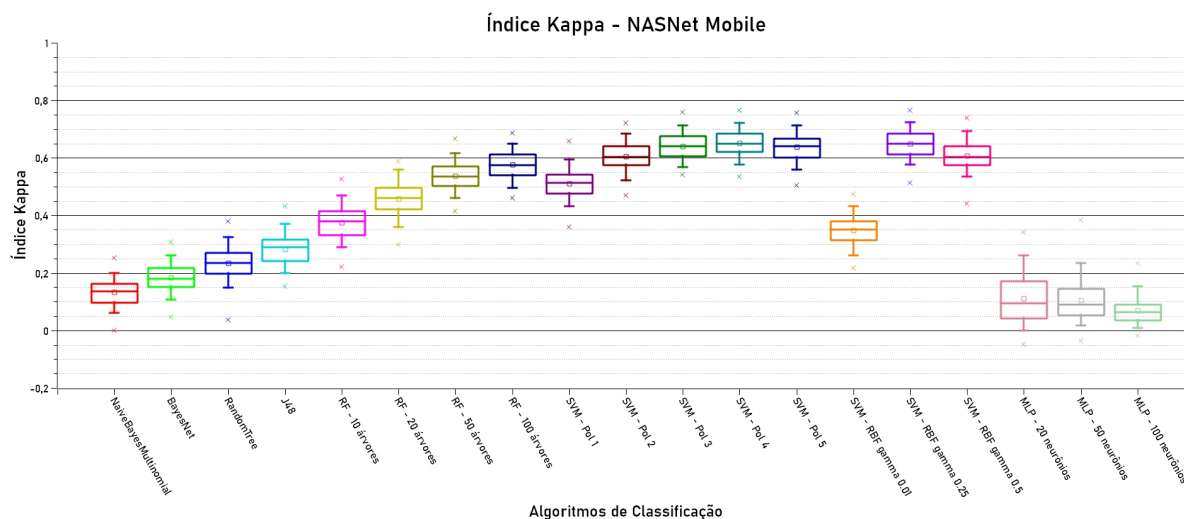
Fonte: o autor (2022).

Figura 22 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN *NASNetMobile*.



Fonte: o autor (2022).

Figura 23 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *NASNetMobile*.



Fonte: o autor (2022).

ms e 0,01 de desvio padrão.

Tabela 8 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *NASNetMobile*. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.

Rede Neural Convolutacional <i>NASNetMobile</i>		
Classificador	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,53 ± 0,05	6,89 ± 1,49
Bayes Net	0,62 ± 0,05	0,42 ± 0,03
Random Tree	0,56 ± 0,04	0,04 ± 0,01
J48	0,58 ± 0,05	5233,21 ± 1778,82
RF - 10 árvores	0,66 ± 0,05	0,24 ± 0,01
RF - 20 árvores	0,70 ± 0,05	0,48 ± 0,01
RF - 50 árvores	0,75 ± 0,05	1190,50 ± 9,43
RF - 100 árvores	0,78 ± 0,04	2162,94 ± 694,45
SVM - Polinômio 1º	0,65 ± 0,05	4563,53 ± 1691,84
SVM - Polinômio 2º	0,69 ± 0,05	4965,15 ± 1689,15
SVM - Polinômio 3º	0,71 ± 0,05	4608,70 ± 1419,62
SVM - Polinômio 4º	0,72 ± 0,05	2929,27 ± 1112,06
SVM - Polinômio 5º	0,72 ± 0,05	2718,24 ± 530,97
SVM - RBF $\gamma = 0,01$	0,64 ± 0,05	2960,01 ± 521,55
SVM - RBF $\gamma = 0,25$	0,78 ± 0,04	2861,87 ± 806,57
SVM - RBF $\gamma = 0,5$	0,80 ± 0,03	2579,20 ± 1012,44
MLP - 20 neurônios	0,55 ± 0,06	287326,85 ± 104687,12
MLP - 50 neurônios	0,56 ± 0,06	847132,06 ± 300240,91
MLP - 100 neurônios	0,55 ± 0,06	1754819,15 ± 860312,74

Fonte: o autor (2022).

Os resultados para a *ResNet50* são mostrados nas Tabelas 9 e 10. Na Tabela 9, o SVM com kernel RBF $\gamma = 0,25$ apresentou o melhor desempenho para Acurácia (Figura 24)

e Índice Kappa (Figura 25), com 75,72% e desvio padrão de 3,21 para Acurácia, e Kappa de 0,68 com 0,04 de desvio padrão. Para Sensibilidade, o SVM com kernel RBF $\gamma = 0,5$ teve o melhor resultado, obtendo 83,90% e 6,27 de desvio padrão. Para Especificidade, o *Bayes Net* apresentou o melhor desempenho com 99,03% e 0,99 de desvio padrão.

Tabela 9 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *ResNet50*. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

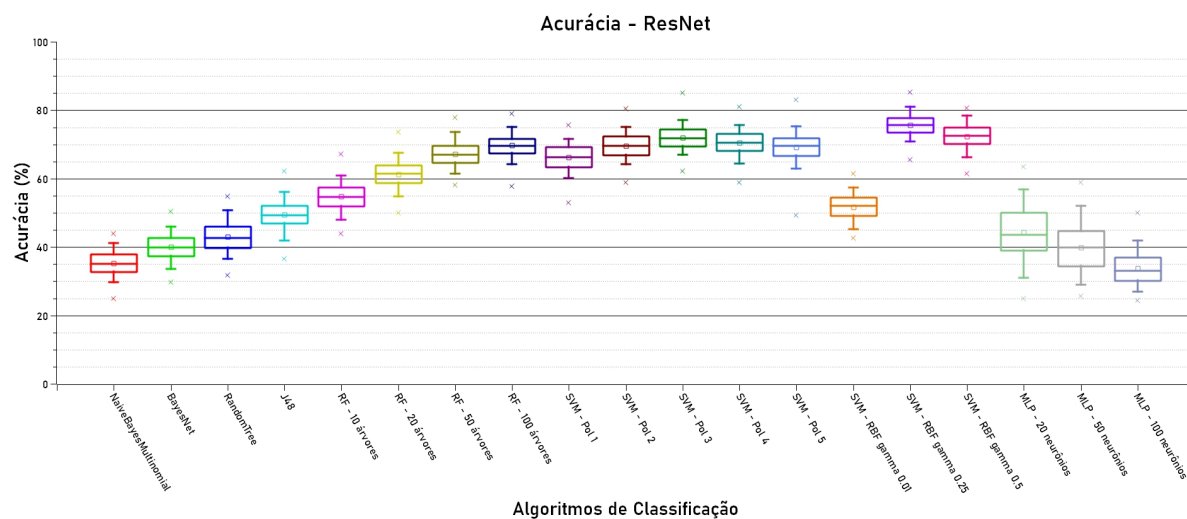
Classificador	Rede Neural Convolutacional ResNet50			
	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	35,30 ± 3,50	0,14 ± 0,05	1,78 ± 2,08	97,75 ± 1,68
Bayes Net	39,96 ± 3,78	0,20 ± 0,05	3,67 ± 3,01	99,03 ± 0,99
Random Tree	42,92 ± 4,21	0,24 ± 0,06	34,07 ± 8,08	79,32 ± 3,99
J48	49,46 ± 4,30	0,33 ± 0,06	39,51 ± 8,17	80,04 ± 4,15
RF - 10 árvores	54,72 ± 3,96	0,40 ± 0,05	44,04 ± 7,95	80,70 ± 3,62
RF - 20 árvores	61,30 ± 3,84	0,48 ± 0,05	42,89 ± 8,24	86,51 ± 3,28
RF - 50 árvores	67,21 ± 3,73	0,56 ± 0,05	40,49 ± 8,66	91,79 ± 2,56
RF - 100 árvores	69,65 ± 3,40	0,60 ± 0,05	39,12 ± 8,30	94,11 ± 2,17
SVM - Polinômio 1º	66,17 ± 3,71	0,55 ± 0,05	53,11 ± 8,35	84,75 ± 3,59
SVM - Polinômio 2º	69,60 ± 3,51	0,59 ± 0,05	56,44 ± 8,09	85,57 ± 3,34
SVM - Polinômio 3º	71,99 ± 3,39	0,63 ± 0,05	57,11 ± 7,87	87,74 ± 3,20
SVM - Polinômio 4º	70,42 ± 3,62	0,61 ± 0,05	54,42 ± 9,37	88,00 ± 4,51
SVM - Polinômio 5º	69,24 ± 4,13	0,59 ± 0,06	54,36 ± 10,58	86,53 ± 5,80
SVM - RBF $\gamma = 0,01$	51,68 ± 3,74	0,36 ± 0,05	18,08 ± 6,51	94,73 ± 2,19
SVM - RBF $\gamma = 0,25$	75,72 ± 3,21	0,68 ± 0,04	69,80 ± 7,77	83,42 ± 3,38
SVM - RBF $\gamma = 0,5$	72,36 ± 3,61	0,63 ± 0,05	83,90 ± 6,27	71,53 ± 4,61
MLP - 20 neurônios	44,30 ± 7,69	0,26 ± 0,10	24,82 ± 26,80	84,70 ± 24,98
MLP - 50 neurônios	39,90 ± 6,84	0,20 ± 0,09	25,68 ± 27,73	83,05 ± 27,59
MLP - 100 neurônios	33,83 ± 4,70	0,12 ± 0,06	24,94 ± 30,21	82,19 ± 31,33

Fonte: o autor (2022).

Na Tabela 10 estão os resultados de Área Sob Curva ROC e Tempo de Treinamento (em milissegundos) para a *ResNet50*. O *Random Forest* para 100 árvores obteve o melhor resultado para Área Sob Curva ROC, com 0,80 e 0,04 de desvio padrão. O Tempo de Treinamento em milissegundos foi menor para o *Random Tree*, com 60,58 e 15,54 de desvio padrão.

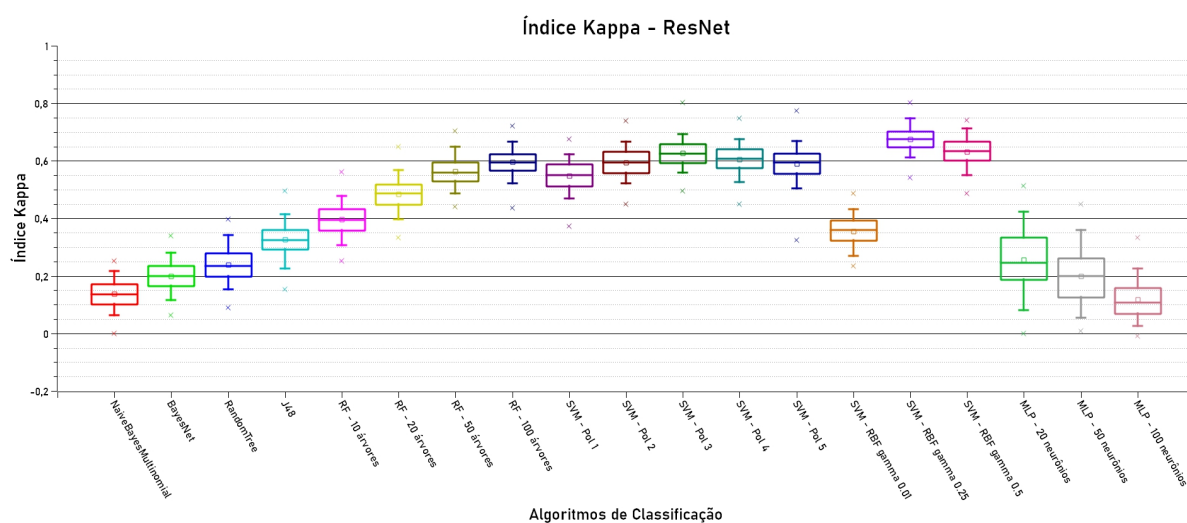
Nas Tabelas 11 e 12, são apresentados os resultados de classificação para a arquitetura CNN *SqueezeNet*. Pela Tabela 11, para a taxa de Acurácia (Figura 26) e Índice Kappa (Figura 27), o classificador *Random Forest* com 100 árvores apresentou os melhores resultados, com Acurácia de 67,17% e desvio padrão de 3,51, e Kappa de 0,56 e desvio padrão de 0,05. A taxa de Sensibilidade teve o melhor desempenho para o SVM com kernel RBF $\gamma = 0,5$, obtendo 90,64% e desvio padrão de 6,25. Para Especificidade, o *Bayes Net*

Figura 24 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN *ResNet50*.



Fonte: o autor (2022).

Figura 25 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *ResNet50*.



Fonte: o autor (2022).

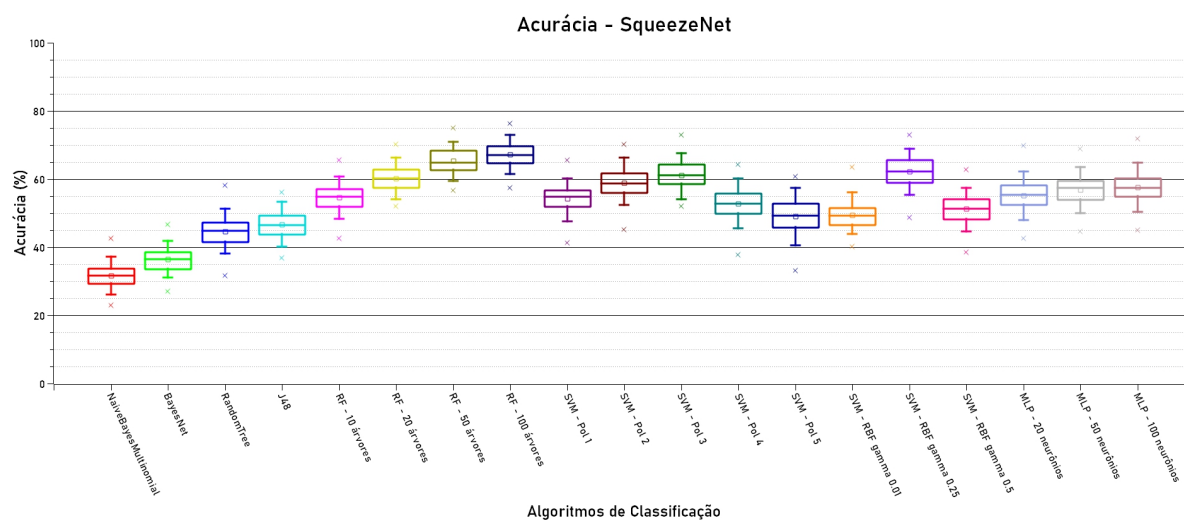
Tabela 10 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *ResNet50*. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.

Rede Neural Convolutacional ResNet50		
Classificador	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,53 ± 0,05	13,81 ± 5,11
Bayes Net	0,59 ± 0,05	692,69 ± 236,09
Random Tree	0,57 ± 0,04	60,58 ± 15,54
J48	0,60 ± 0,05	6000,93 ± 2322,99
RF - 10 árvores	0,67 ± 0,05	333,92 ± 124,22
RF - 20 árvores	0,71 ± 0,05	677,91 ± 200,90
RF - 50 árvores	0,77 ± 0,04	1644,71 ± 559,49
RF - 100 árvores	0,80 ± 0,04	3482,88 ± 858,13
SVM - Polinômio 1º	0,72 ± 0,04	7394,09 ± 2472,50
SVM - Polinômio 2º	0,73 ± 0,04	9611,75 ± 3889,66
SVM - Polinômio 3º	0,75 ± 0,04	9638,83 ± 3171,32
SVM - Polinômio 4º	0,74 ± 0,04	5130,55 ± 1783,93
SVM - Polinômio 5º	0,73 ± 0,05	4725,02 ± 1827,82
SVM - RBF $\gamma = 0,01$	0,62 ± 0,05	5070,49 ± 1961,88
SVM - RBF $\gamma = 0,25$	0,78 ± 0,04	5509,30 ± 2100,34
SVM - RBF $\gamma = 0,5$	0,78 ± 0,04	5237,15 ± 1811,35
MLP - 20 neurônios	0,59 ± 0,06	865997,70 ± 436068,46
MLP - 50 neurônios	0,58 ± 0,06	2430502,50 ± 818942,36
MLP - 100 neurônios	0,55 ± 0,06	4780833,09 ± 1903411,03

Fonte: o autor (2022).

apresentou o melhor resultado com 95,31% e 2,28 de desvio padrão.

Figura 26 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN *SqueezeNet*.



Fonte: o autor (2022).

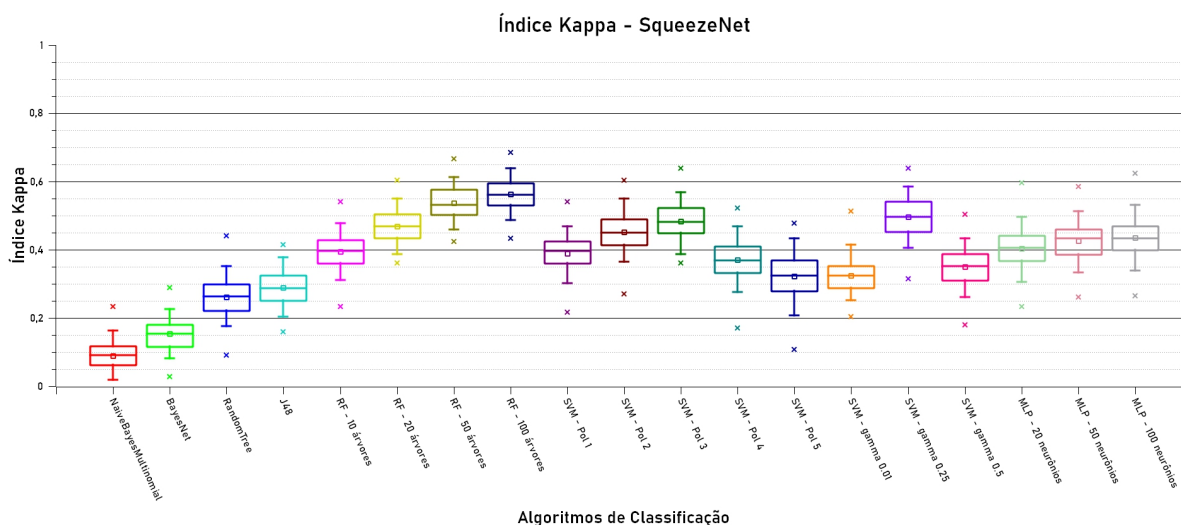
Para *SqueezeNet*, Tabela 12, a métrica de Área Sob Curva ROC teve 0,80 e desvio padrão de 0,04 como seu melhor resultado, obtido com o classificador *Random Forest* com

Tabela 11 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *SqueezeNet*. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

Classificador	Rede Neural Convolutacional SqueezeNet			
	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	31,59 ± 3,36	0,09 ± 0,04	13,92 ± 10,60	87,97 ± 8,23
Bayes Net	36,46 ± 3,35	0,15 ± 0,04	16,89 ± 5,16	95,31 ± 2,28
Random Tree	44,58 ± 4,34	0,26 ± 0,06	16,89 ± 5,16	80,53 ± 4,17
J48	46,60 ± 4,00	0,29 ± 0,05	42,37 ± 7,94	80,68 ± 3,84
RF - 10 árvores	54,62 ± 3,82	0,39 ± 0,05	48,00 ± 8,40	81,15 ± 3,71
RF - 20 árvores	60,18 ± 3,72	0,47 ± 0,05	48,12 ± 8,28	85,81 ± 3,36
RF - 50 árvores	65,29 ± 3,64	0,54 ± 0,05	48,40 ± 8,14	89,63 ± 2,91
RF - 100 árvores	67,17 ± 3,51	0,56 ± 0,05	48,25 ± 8,01	90,89 ± 2,75
SVM - Polinômio 1º	54,22 ± 3,94	0,39 ± 0,05	50,71 ± 8,53	78,33 ± 3,85
SVM - Polinômio 2º	58,83 ± 4,16	0,45 ± 0,06	54,62 ± 8,44	79,00 ± 4,01
SVM - Polinômio 3º	61,19 ± 4,05	0,48 ± 0,05	54,87 ± 8,38	80,84 ± 3,91
SVM - Polinômio 4º	52,76 ± 4,58	0,37 ± 0,06	48,51 ± 15,10	78,94 ± 11,48
SVM - Polinômio 5º	49,10 ± 4,96	0,32 ± 0,07	45,38 ± 17,65	79,15 ± 14,07
SVM - RBF $\gamma = 0,01$	49,36 ± 3,77	0,36 ± 0,05	35,76 ± 7,43	90,18 ± 2,92
SVM - RBF $\gamma = 0,25$	62,23 ± 4,18	0,50 ± 0,06	66,19 ± 8,59	71,78 ± 5,98
SVM - RBF $\gamma = 0,5$	51,33 ± 4,24	0,35 ± 0,06	90,64 ± 6,25	41,40 ± 6,73
MLP - 20 neurônios	55,21 ± 4,37	0,40 ± 0,06	47,17 ± 9,44	82,43 ± 4,79
MLP - 50 neurônios	56,89 ± 4,15	0,43 ± 0,06	47,54 ± 8,68	84,18 ± 4,27
MLP - 100 neurônios	57,58 ± 4,34	0,43 ± 0,06	47,88 ± 8,99	84,50 ± 4,09

Fonte: o autor (2022).

Figura 27 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *SqueezeNet*.



Fonte: o autor (2022).

100 árvores. Já o Tempo de Treinamento (em milissegundos) foi menor para o *Random Tree* com 0,07 ms e desvio padrão de 0,01.

Tabela 12 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolucional *SqueezeNet*. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.

Rede Neural Convolucional SqueezeNet		
Classificador	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,52 ± 0,05	8,01 ± 6,20
Bayes Net	0,59 ± 0,06	0,44 ± 0,02
Random Tree	0,59 ± 0,05	0,07 ± 0,01
J48	0,62 ± 0,05	4990,28 ± 1891,31
RF - 10 árvores	0,70 ± 0,05	0,41 ± 0,02
RF - 20 árvores	0,74 ± 0,04	0,79 ± 0,04
RF - 50 árvores	0,78 ± 0,04	1802,51 ± 537,34
RF - 100 árvores	0,80 ± 0,04	3468,45 ± 1185,11
SVM - Polinômio 1º	0,69 ± 0,05	14039,28 ± 5241,25
SVM - Polinômio 2º	0,71 ± 0,05	32861,62 ± 11439,96
SVM - Polinômio 3º	0,71 ± 0,05	29019,05 ± 8940,70
SVM - Polinômio 4º	0,69 ± 0,05	3867,13 ± 1227,51
SVM - Polinômio 5º	0,67 ± 0,05	3489,30 ± 1334,88
SVM - RBF $\gamma = 0,01$	0,66 ± 0,05	3894,31 ± 1372,83
SVM - RBF $\gamma = 0,25$	0,69 ± 0,05	3673,81 ± 1295,96
SVM - RBF $\gamma = 0,5$	0,66 ± 0,03	3710,77 ± 1308,37
MLP - 20 neurônios	0,71 ± 0,05	269281,31 ± 84059,79
MLP - 50 neurônios	0,72 ± 0,05	657240,48 ± 240910,49
MLP - 100 neurônios	0,73 ± 0,05	1349616,81 ± 425069,16

Fonte: o autor (2022).

A CNN *Inception v3*, obteve os resultados mostrados por meio das Tabelas 13 e 14. Pela Tabela 13, o SVM com kernel polinomial de grau 3 teve os melhores resultados para Acurácia (Figura 28) e Kappa (Figura 29), com Acurácia de 79,92% e desvio padrão de 3,07, e Kappa de 0,73 e 0,04 de desvio padrão. Para a Sensibilidade, o melhor resultado foi o SVM com kernel RBF $\gamma = 0,25$, com taxa de 100,00%. Já para a Especificidade, o melhor resultado foi para o *Bayes Net* com 93,79% e 2,22 de desvio padrão.

Os resultados das métricas de Área Sob Curva ROC e Tempo de Treinamento (em milissegundos), mostrados na Tabela 14 para a *Inception v3*, foram: *Random Forest* com 100 árvores para Área Sob Curva ROC com taxa de 0,79 e desvio padrão de 0,04; e *Random Tree* com o menor Tempo de Treinamento, com 73,68 ms e desvio padrão de 27,47.

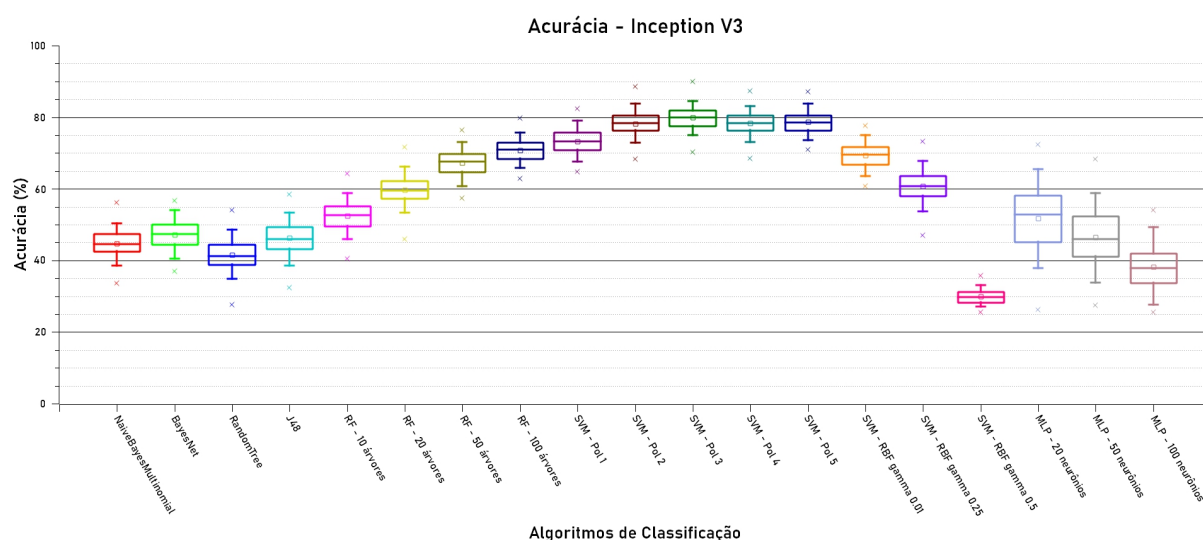
A Tabela 15 mostra a comparação entre os melhores classificadores por CNN. Entre todas as Redes Convolucionais, a *Inception v3* teve a melhor taxa de acurácia com o

Tabela 13 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutiva InceptionV3. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

Classificador	Rede Neural Convolutiva InceptionV3			
	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	44,73 ± 3,73	0,26 ± 0,05	16,51 ± 5,60	92,49 ± 2,67
Bayes Net	47,25 ± 3,93	0,30 ± 0,05	17,01 ± 5,83	93,79 ± 2,22
Random Tree	41,52 ± 4,19	0,22 ± 0,06	32,19 ± 7,77	78,88 ± 3,94
J48	46,28 ± 4,33	0,28 ± 0,06	42,37 ± 7,94	79,65 ± 3,88
RF - 10 árvores	52,49 ± 3,93	0,37 ± 0,05	42,49 ± 8,47	78,55 ± 3,97
RF - 20 árvores	59,72 ± 3,92	0,46 ± 0,05	41,33 ± 7,77	84,41 ± 3,70
RF - 50 árvores	67,21 ± 3,65	0,56 ± 0,05	40,51 ± 7,68	90,19 ± 2,77
RF - 100 árvores	70,75 ± 3,17	0,61 ± 0,04	40,11 ± 8,11	92,99 ± 2,23
SVM - Polinômio 1º	73,20 ± 3,39	0,64 ± 0,05	55,90 ± 7,78	87,89 ± 2,98
SVM - Polinômio 2º	78,16 ± 3,30	0,71 ± 0,04	57,48 ± 7,81	91,87 ± 2,38
SVM - Polinômio 3º	79,92 ± 3,07	0,73 ± 0,04	58,15 ± 7,73	93,19 ± 2,22
SVM - Polinômio 4º	78,33 ± 3,12	0,71 ± 0,04	54,63 ± 8,86	93,17 ± 2,65
SVM - Polinômio 5º	78,63 ± 3,05	0,72 ± 0,04	53,67 ± 9,13	93,59 ± 2,48
SVM - RBF $\gamma = 0,01$	69,33 ± 3,37	0,59 ± 0,04	44,82 ± 8,21	91,59 ± 2,44
SVM - RBF $\gamma = 0,25$	60,82 ± 4,34	0,48 ± 0,06	100,00 ± 0,00	47,74 ± 5,74
SVM - RBF $\gamma = 0,5$	29,88 ± 1,89	0,07 ± 0,02	90,00 ± 30,05	16,13 ± 28,10
MLP - 20 neurônios	51,83 ± 8,88	0,36 ± 0,12	29,42 ± 27,71	85,20 ± 21,14
MLP - 50 neurônios	46,55 ± 7,84	0,29 ± 0,10	29,79 ± 25,31	83,03 ± 22,41
MLP - 100 neurônios	38,19 ± 6,48	0,18 ± 0,09	27,42 ± 29,34	80,85 ± 28,76

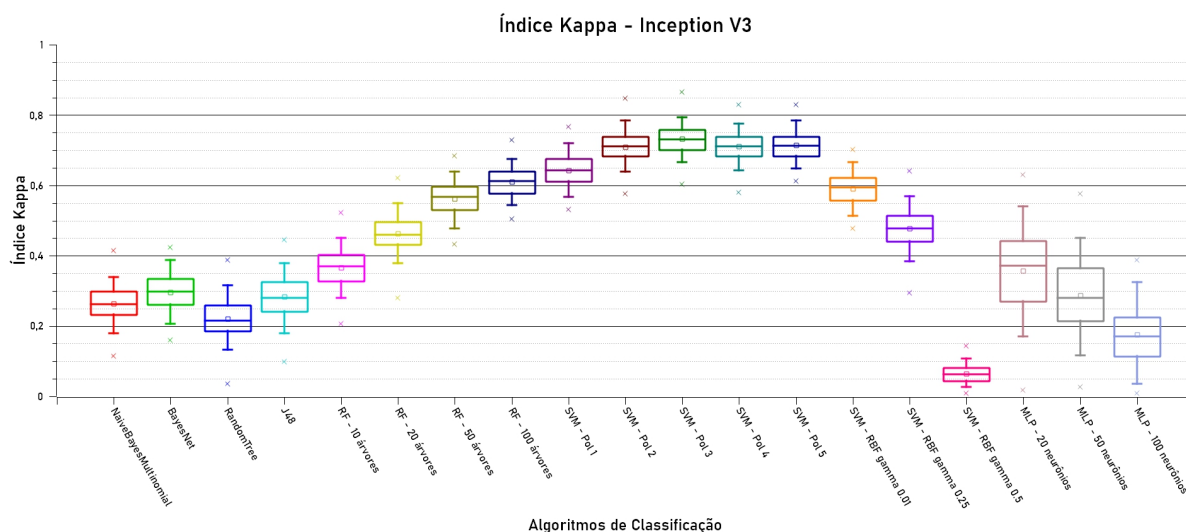
Fonte: o autor (2022).

Figura 28 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN Inception v3.



Fonte: o autor (2022).

Figura 29 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *Inception v3*.



Fonte: o autor (2022).

Tabela 14 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolucional *Inception v3*. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.

Rede Neural Convolucional InceptionV3		
Classificador	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,57 ± 0,05	13,45 ± 4,32
Bayes Net	0,66 ± 0,05	1107,75 ± 395,56
Random Tree	0,56 ± 0,05	73,68 ± 27,47
J48	0,58 ± 0,05	15344,45 ± 4753,87
RF - 10 árvores	0,64 ± 0,05	429,99 ± 165,42
RF - 20 árvores	0,68 ± 0,05	821,62 ± 308,43
RF - 50 árvores	0,75 ± 0,04	1917,00 ± 669,46
RF - 100 árvores	0,79 ± 0,04	3854,60 ± 1292,95
SVM - Polinômio 1º	0,73 ± 0,04	9212,31 ± 3217,03
SVM - Polinômio 2º	0,76 ± 0,04	8275,21 ± 2930,64
SVM - Polinômio 3º	0,77 ± 0,04	7744,16 ± 2350,95
SVM - Polinômio 4º	0,76 ± 0,04	5259,90 ± 1914,30
SVM - Polinômio 5º	0,76 ± 0,04	5371,11 ± 1657,20
SVM - RBF $\gamma = 0,01$	0,72 ± 0,04	5650,47 ± 1706,16
SVM - RBF $\gamma = 0,25$	0,74 ± 0,03	5644,23 ± 1777,81
SVM - RBF $\gamma = 0,5$	0,53 ± 0,01	5106,24 ± 2175,09
MLP - 20 neurônios	0,64 ± 0,07	518541,27 ± 192120,65
MLP - 50 neurônios	0,62 ± 0,06	1404933,52 ± 499019,12
MLP - 100 neurônios	0,55 ± 0,06	2983384,81 ± 1114101,18

Fonte: o autor (2022).

classificador SVM de kernel polinomial de grau 3, alcançando 79,92% e desvio padrão de 3,07 de acurácia. O Índice Kappa, de acordo com a Tabela 16, obteve melhor desempenho para a *Inception v3* para o SVM com kernel polinomial de grau 3, com taxa de 0,73 e desvio padrão de 0,04. Para Sensibilidade, Tabela 17, a *Inception v3* apresentou o melhor desempenho para o SVM com kernel RBF $\gamma = 0,25$ obtendo 100,00%. Para Especificidade, Tabela 18, a *ResNet50* teve destaque com o classificador *Bayes Net* com uma taxa de 99,03 e desvio padrão de 0,99. A *NASNetMobile* e a *SqueezeNet* apresentam as melhores taxas de Área Sob Curva ROC, com 0,80 e desvio padrão de 0,03, conforme Tabela 19.

Tabela 15 – Comparação entre a taxa de acurácia dos melhores classificadores para cada Rede Neural Convolucional.

Comparação da Taxa de Acurácia por CNN		
CNN	Classificador	Acurácia (%)
LeNet	SVM - RBF $\gamma = 0,5$	76,65 $\pm$ 3,80
NASNetMobile	SVM - Polinômio 4º	73,81 $\pm$ 3,36
ResNet50	SVM - RBF $\gamma = 0,25$	75,72 $\pm$ 3,21
SqueezeNet	RF - 100 árvores	67,17 $\pm$ 3,51
InceptionV3	SVM - Polinômio 3º	79,92 $\pm$ 3,07

Fonte: o autor (2022).

Tabela 16 – Comparação entre a taxa de índice kappa dos melhores classificadores para cada Rede Neural Convolucional.

Comparação do Índice Kappa por CNN		
CNN	Classificador	Índice Kappa
LeNet	SVM - RBF $\gamma = 0,5$	0,69 $\pm$ 0,05
NASNetMobile	SVM - Polinômio 4º e SVM - RBF $\gamma = 0,25$	0,65 $\pm$ 0,05
ResNet50	SVM - RBF $\gamma = 0,25$	0,68 $\pm$ 0,04
SqueezeNet	RF - 100 árvores	0,56 $\pm$ 0,05
InceptionV3	SVM - Polinômio 3º	0,73 $\pm$ 0,04

Fonte: o autor (2022).

Tabela 17 – Comparação entre a taxa de sensibilidade dos melhores classificadores para cada Rede Neural Convolucional.

Comparação da Taxa de Sensibilidade por CNN		
CNN	Classificador	Sensibilidade (%)
LeNet	SVM - RBF $\gamma = 0,5$	84,97 $\pm$ 6,70
NASNetMobile	SVM - RBF $\gamma = 0,5$	98,16 $\pm$ 2,20
ResNet50	SVM - RBF $\gamma = 0,5$	83,90 $\pm$ 6,27
SqueezeNet	SVM - RBF $\gamma = 0,5$	90,64 $\pm$ 6,25
InceptionV3	SVM - RBF $\gamma = 0,25$	100,00 $\pm$ 0,00

Fonte: o autor (2022).

Tabela 18 – Comparação entre a taxa de especificidade dos melhores classificadores para cada Rede Neural Convolucional.

Comparação da Taxa de Especificidade por CNN		
CNN	Classificador	Especificidade (%)
LeNet	Bayes Net	90,50 $\pm$ 2,90
NASNetMobile	Bayes Net	93,93 $\pm$ 2,61
ResNet50	Bayes Net	99,03 $\pm$ 0,99
SqueezeNet	Bayes Net	95,31 $\pm$ 2,28
InceptionV3	Bayes Net	93,79 $\pm$ 2,22

Fonte: o autor (2022).

Tabela 19 – Comparação entre a taxa de área sob curva ROC dos melhores classificadores para cada Rede Neural Convolucional.

Comparação da Área Sob Curva ROC por CNN		
CNN	Classificador	Área Sob Curva ROC
LeNet	RF - 100 árvores	0,78 $\pm$ 0,04
NASNetMobile	SVM - RBF $\gamma = 0,5$	0,80 $\pm$ 0,03
ResNet50	SVM - Polinômio 4º	0,74 $\pm$ 0,04
SqueezeNet	RF - 100 árvores	0,80 $\pm$ 0,04
InceptionV3	RF - 100 árvores	0,79 $\pm$ 0,04

Fonte: o autor (2022).

5.2 Análise da classificação após seleção de atributos com PSO

Neste trabalho, foram utilizadas Redes Neurais Convolucionais (CNN) para extrair atributos das imagens para cada classe definida. Em seguida, com o objetivo de selecionar os atributos mais relevantes para cada conjunto de atributos extraídos pelas CNNs, foi utilizada a técnica *PSO (Particle Swarm Optimization)* com 20 partículas e 50 gerações. Após a seleção dos atributos mais relevantes, foi realizada a etapa de classificação com diversos classificadores. Cada classificador foi avaliado em 30 rodadas utilizando a validação cruzada com 10 *folds*.

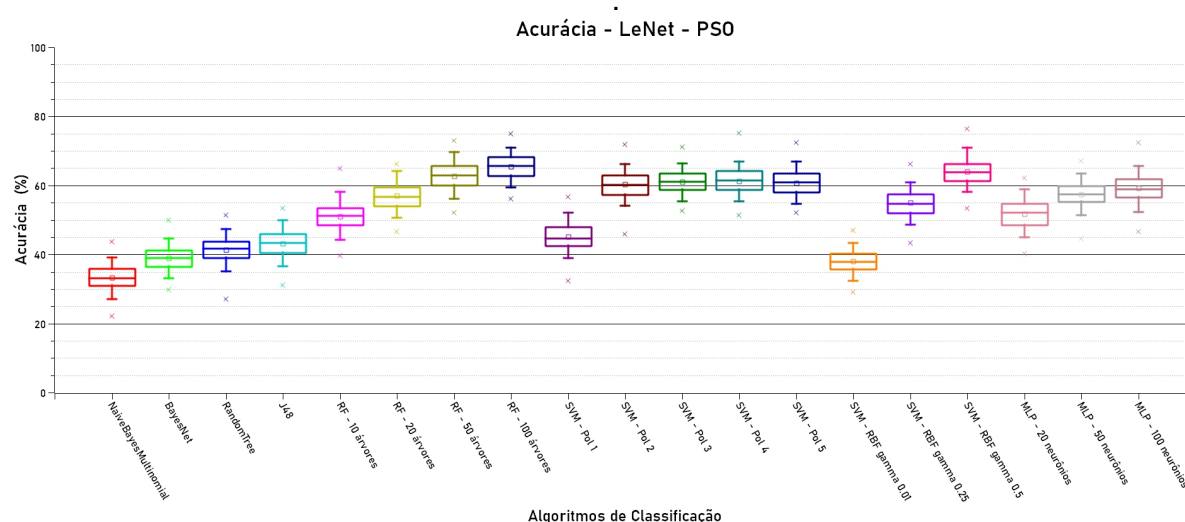
A Tabela 20 e a Tabela 21 mostram os resultados dos classificadores para a Rede Convolucional *LeNet* utilizando os atributos mais relevantes selecionados pela técnica PSO. De acordo com a Tabela 20, para a *LeNet*, foram obtidos os seguintes resultados: o classificador *Random Forest* com 100 árvores obteve o melhor desempenho para Acurácia (Figura 30) com 65,37% e desvio padrão de 3,62, como também para o Índice Kappa (Figura 31) com 0,54 e desvio padrão de 0,05; já para Sensibilidade, o SVM polinomial de 2º, obteve 47,16% com desvio padrão de 8,39; para a Especificidade, o melhor classificador foi o SVM com kernel RBF de $\gamma = 0,01$ com 94,73% e 2,35 de desvio padrão.

Tabela 20 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolucional *LeNet* e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

Classificador	Rede Neural Convolucional LeNet - PSO			
	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	33,31 ± 3,51	0,11 ± 0,05	38,12 ± 7,19	72,75 ± 4,78
Bayes Net	38,87 ± 3,43	0,19 ± 0,05	11,63 ± 5,29	92,34 ± 2,67
Random Tree	41,32 ± 3,75	0,22 ± 0,05	32,44 ± 7,63	78,64 ± 4,14
J48	43,11 ± 4,07	0,24 ± 0,05	34,11 ± 7,65	78,18 ± 4,07
RF - 10 árvores	51,02 ± 4,07	0,35 ± 0,05	41,44 ± 8,58	78,32 ± 4,05
RF - 20 árvores	56,97 ± 3,97	0,43 ± 0,05	40,67 ± 7,94	83,44 ± 3,54
RF - 50 árvores	62,68 ± 3,98	0,50 ± 0,05	39,54 ± 7,50	88,30 ± 3,30
RF - 100 árvores	65,37 ± 3,62	0,54 ± 0,05	39,44 ± 7,50	90,44 ± 3,03
SVM - Polinômio 1º	45,11 ± 3,95	0,27 ± 0,05	27,50 ± 6,96	80,25 ± 4,09
SVM - Polinômio 2º	60,25 ± 3,90	0,47 ± 0,05	47,16 ± 8,39	82,17 ± 3,91
SVM - Polinômio 3º	61,28 ± 3,70	0,48 ± 0,05	45,02 ± 7,84	83,15 ± 3,67
SVM - Polinômio 4º	61,28 ± 3,70	0,48 ± 0,05	45,02 ± 7,84	83,15 ± 3,67
SVM - Polinômio 5º	60,75 ± 3,82	0,48 ± 0,05	43,72 ± 7,78	83,44 ± 3,56
SVM - RBF $\gamma = 0,01$	37,98 ± 3,25	0,17 ± 0,04	11,02 ± 4,77	94,73 ± 2,35
SVM - RBF $\gamma = 0,25$	54,96 ± 3,95	0,40 ± 0,05	35,54 ± 7,51	85,13 ± 3,72
SVM - RBF $\gamma = 0,5$	63,95 ± 3,86	0,52 ± 0,05	43,85 ± 8,03	85,94 ± 3,56
MLP - 20 neurônios	51,76 ± 4,19	0,36 ± 0,06	38,72 ± 8,28	82,10 ± 4,06
MLP - 50 neurônios	57,44 ± 3,75	0,43 ± 0,05	41,63 ± 8,15	84,13 ± 3,61
MLP - 100 neurônios	59,16 ± 4,04	0,46 ± 0,05	41,67 ± 8,00	85,22 ± 3,51

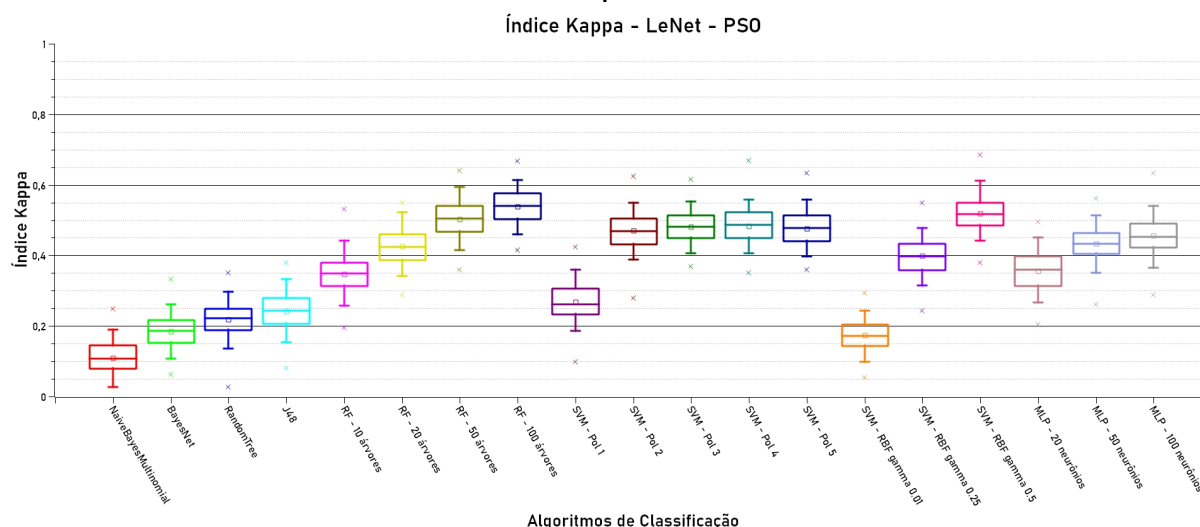
Fonte: o autor (2022).

Figura 30 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN *LeNet* e selecionados com PSO



Fonte: o autor (2022).

Figura 31 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *LeNet* e selecionados com PSO



Fonte: o autor (2022).

Na Tabela 21 são mostrados os resultados, para a *LeNet*, da Área Sob Curva ROC e Tempo de Treinamento (em milissegundos). O classificador com melhor desempenho para Área Sob Curva ROC foi o *Random Forest* de 100 árvores, com 0,77 e desvio padrão de 0,04. Já para o Tempo de Treinamento, o *Random Tree* obteve o menor tempo com 0,02 ms.

As Tabelas 22 e 23 mostram os resultados para a CNN *NASNetMobile*. O *Random Forest* de 100 árvores obteve o melhor desempenho para a Acurácia (Figura 32), com 55,97% e desvio padrão de 3,71, como também para o Índice Kappa (Figura 33) com taxa de 0,41 e desvio padrão de 0,05. O classificador SVM com kernel RBF $\gamma = 0,01$ apresentou o melhor desempenho para a Sensibilidade, com taxa de 42,57% e desvio padrão de 41,13. Para Especificidade, o SVM Polinomial de 5º teve a melhor taxa com 92,44% e desvio padrão de 2,54.

Na Tabela 23, o melhor classificador foi *Random Forest* com 100 árvores que obteve o desempenho de 0,67 e desvio padrão de 0,05 para a Área Sob Curva ROC. Já para o Tempo de Treinamento em milissegundo, o *Naive Bayes Multinomial* obteve o menor tempo com 0,24 ms.

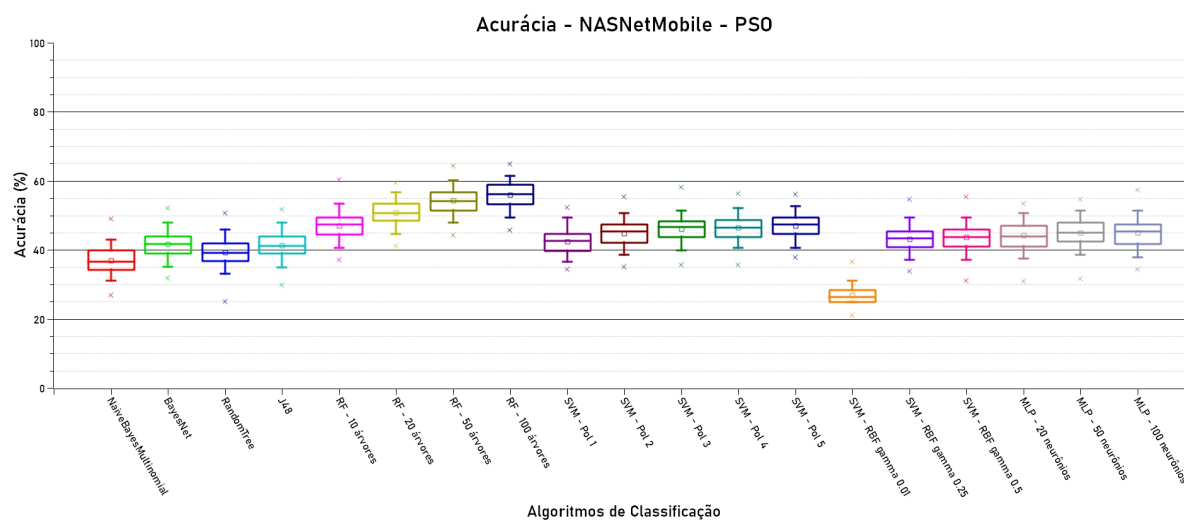
Os resultados para a *ResNet50* com PSO são mostrados nas Tabelas 24 e 25. Na Tabela 24, o *Random Forest* com 100 árvores apresentou o melhor desempenho para Acurácia (Figura 34) com 69,59% e desvio padrão de 3,56. Para o Índice Kappa (Figura 35), o SVM com kernel RBF de $\gamma = 0,5$ apresentou o melhor resultado, com taxa de 0,65

Tabela 21 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional LeNet e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas Área Sob Curva ROC e Tempo de Treinamento.

Rede Neural Convolutacional LeNet - PSO		
Classificador	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,59 ± 0,05	1,31 ± 0,97
Bayes Net	0,52 ± 0,05	0,04 ± 0,00
Random Tree	0,56 ± 0,04	0,02 ± 0,00
J48	0,57 ± 0,05	0,48 ± 0,00
RF - 10 árvores	0,64 ± 0,05	0,11 ± 0,01
RF - 20 árvores	0,69 ± 0,05	0,22 ± 0,01
RF - 50 árvores	0,74 ± 0,05	0,57 ± 0,01
RF - 100 árvores	0,77 ± 0,04	1032,18 ± 283,49
SVM - Polinômio 1º	0,59 ± 0,05	0,69 ± 0,02
SVM - Polinômio 2º	0,68 ± 0,05	16534,88 ± 6439,90
SVM - Polinômio 3º	0,68 ± 0,05	23153,90 ± 7923,57
SVM - Polinômio 4º	0,68 ± 0,05	30563,75 ± 12166,20
SVM - Polinômio 5º	0,68 ± 0,05	34214,03 ± 11617,26
SVM - RBF $\gamma = 0,01$	0,59 ± 0,05	1414,63 ± 427,35
SVM - RBF $\gamma = 0,25$	0,66 ± 0,05	1263,20 ± 464,82
SVM - RBF $\gamma = 0,5$	0,70 ± 0,05	1448,22 ± 515,74
MLP - 20 neurônios	0,65 ± 0,05	28012,71 ± 11282,91
MLP - 50 neurônios	0,69 ± 0,05	70650,87 ± 23133,66
MLP - 100 neurônios	0,70 ± 0,05	139373,19 ± 48341,23

Fonte: o autor (2022).

Figura 32 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN *NASNetMobile* e selecionados usando PSO.



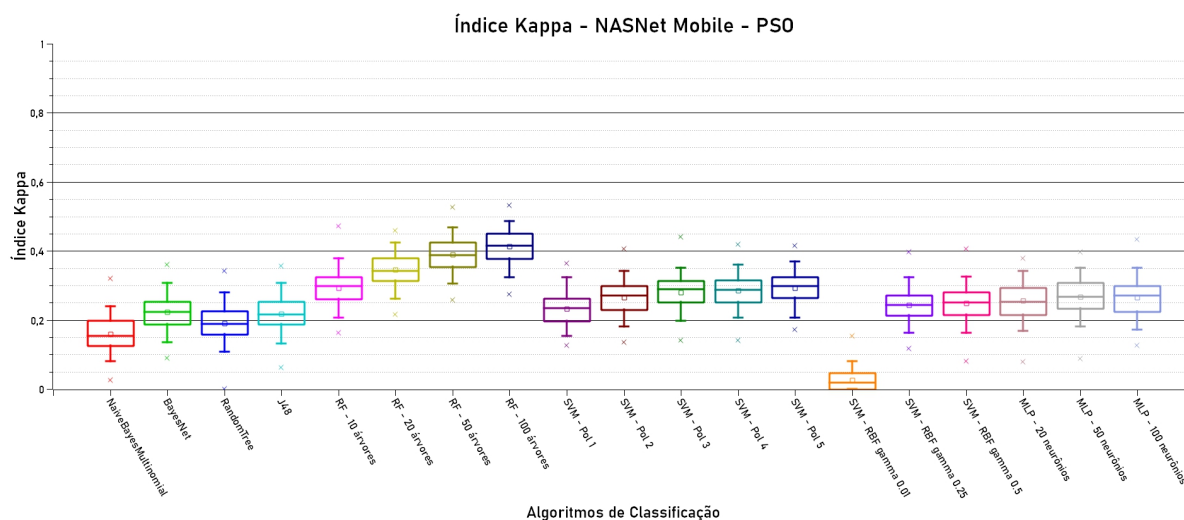
Fonte: o autor (2022).

Tabela 22 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *NASNetMobile* e selecionados com PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

Rede Neural Convolutacional NASNet Mobile - PSO				
Classificador	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	36,85 ± 3,76	0,16 ± 0,05	21,70 ± 6,19	83,83 ± 3,49
Bayes Net	41,66 ± 3,65	0,22 ± 0,05	24,43 ± 7,03	87,48 ± 3,39
Random Tree	39,25 ± 3,90	0,19 ± 0,05	31,66 ± 8,33	77,72 ± 4,05
J48	41,34 ± 3,85	0,22 ± 0,05	33,22 ± 7,59	78,72 ± 4,24
RF - 10 árvores	46,98 ± 3,92	0,29 ± 0,05	38,14 ± 7,45	77,54 ± 4,02
RF - 20 árvores	50,86 ± 3,55	0,34 ± 0,05	35,94 ± 8,01	82,44 ± 3,51
RF - 50 árvores	54,25 ± 3,72	0,39 ± 0,05	34,58 ± 7,67	85,84 ± 3,38
RF - 100 árvores	55,97 ± 3,71	0,41 ± 0,05	34,43 ± 7,88	87,64 ± 3,03
SVM - Polinômio 1º	42,41 ± 3,52	0,23 ± 0,05	18,96 ± 6,28	88,79 ± 2,93
SVM - Polinômio 2º	44,82 ± 3,67	0,26 ± 0,05	19,45 ± 6,25	91,66 ± 2,84
SVM - Polinômio 3º	46,01 ± 3,59	0,28 ± 0,05	20,10 ± 6,05	92,45 ± 2,80
SVM - Polinômio 4º	46,35 ± 3,44	0,28 ± 0,05	20,60 ± 6,07	92,77 ± 2,54
SVM - Polinômio 5º	46,91 ± 3,55	0,29 ± 0,05	21,34 ± 5,98	92,44 ± 2,57
SVM - RBF $\gamma = 0,01$	2,85 ± 2,38	0,03 ± 0,03	42,57 ± 41,13	58,88 ± 42,07
SVM - RBF $\gamma = 0,25$	43,16 ± 3,62	0,24 ± 0,05	15,36 ± 5,69	92,60 ± 2,48
SVM - RBF $\gamma = 0,5$	43,62 ± 3,68	0,25 ± 0,05	15,94 ± 5,37	92,70 ± 2,51
MLP - 20 neurônios	44,16 ± 4,12	0,25 ± 0,05	27,67 ± 8,15	86,72 ± 4,64
MLP - 50 neurônios	45,05 ± 3,94	0,27 ± 0,05	33,63 ± 8,91	83,65 ± 5,29
MLP - 100 neurônios	44,94 ± 4,01	0,26 ± 0,06	32,10 ± 9,25	83,26 ± 5,58

Fonte: o autor (2022).

Figura 33 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *NASNetMobile* e selecionados usando PSO.



Fonte: o autor (2022).

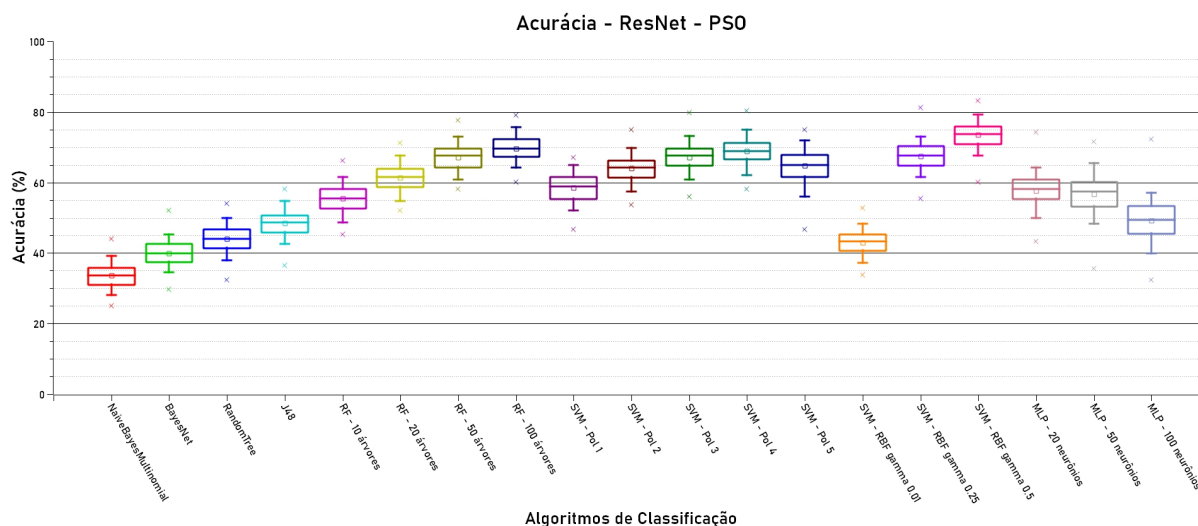
Tabela 23 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *NASNetMobile* e selecionados com PSO. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.

Rede Neural Convolutacional NASNet Mobile - PSO		
Classificador	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,54 ± 0,05	0,24 ± 0,43
Bayes Net	0,62 ± 0,05	6,29 ± 2,96
Random Tree	0,55 ± 0,04	8,86 ± 6,78
J48	0,57 ± 0,05	60,95 ± 21,61
RF - 10 árvores	0,61 ± 0,05	72,30 ± 20,59
RF - 20 árvores	0,63 ± 0,05	138,46 ± 45,21
RF - 50 árvores	0,66 ± 0,05	352,41 ± 136,61
RF - 100 árvores	0,67 ± 0,05	686,55 ± 253,83
SVM - Polinômio 1º	0,55 ± 0,05	59,48 ± 21,95
SVM - Polinômio 2º	0,57 ± 0,05	1166,06 ± 417,49
SVM - Polinômio 3º	0,58 ± 0,05	1451,80 ± 495,50
SVM - Polinômio 4º	0,58 ± 0,05	1868,62 ± 567,98
SVM - Polinômio 5º	0,58 ± 0,05	2453,07 ± 837,52
SVM - RBF $\gamma = 0,01$	0,51 ± 0,03	2217,76 ± 815,97
SVM - RBF $\gamma = 0,25$	0,56 ± 0,05	1106,37 ± 370,14
SVM - RBF $\gamma = 0,5$	0,56 ± 0,05	1075,82 ± 346,01
MLP - 20 neurônios	0,59 ± 0,05	5815,41 ± 2031,02
MLP - 50 neurônios	0,60 ± 0,05	14408,49 ± 5008,59
MLP - 100 neurônios	0,60 ± 0,05	29316,10 ± 9163,08

Fonte: o autor (2022).

e 0,05 de desvio padrão. Para Sensibilidade, o SVM com kernel RBF de $\gamma = 0,5$ teve o melhor resultado, obtendo 56,89% e 8,50 de desvio padrão. Para Especificidade, o *Bayes Net* apresentou o melhor desempenho com 99,42% e 0,84 de desvio padrão.

Figura 34 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN *ResNet50* e selecionados usando PSO.



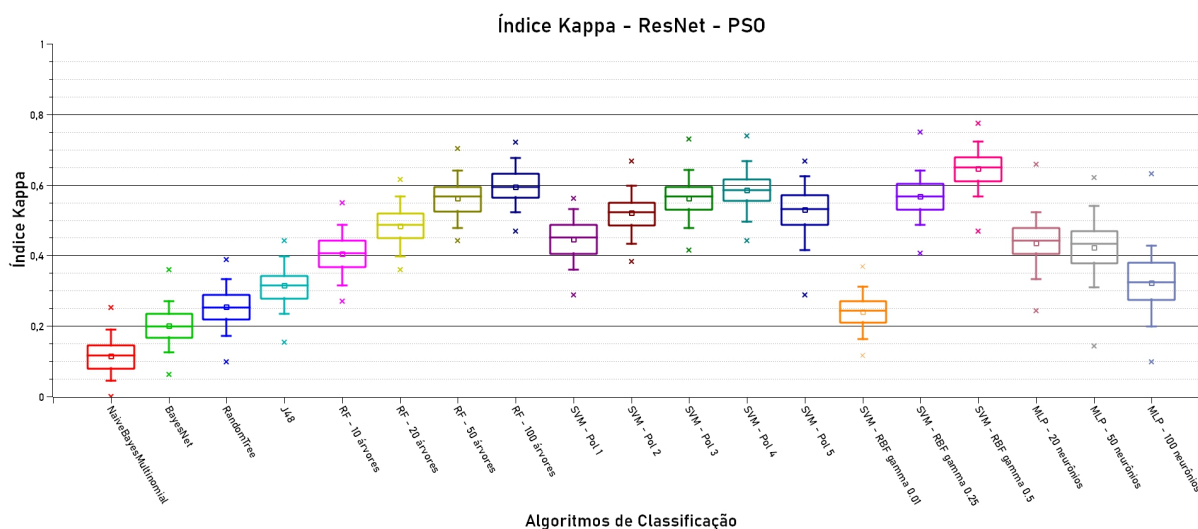
Fonte: o autor (2022).

Tabela 24 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *ResNet50* e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

Classificador	Rede Neural Convolutacional ResNet50 - PSO			
	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	33,57 ± 3,46	0,11 ± 0,05	2,95 ± 2,99	96,24 ± 2,20
Bayes Net	39,95 ± 3,59	0,20 ± 0,05	6,72 ± 3,54	99,42 ± 0,84
Random Tree	44,09 ± 3,72	0,25 ± 0,05	35,18 ± 7,39	79,02 ± 3,89
J48	48,55 ± 3,82	0,31 ± 0,05	39,82 ± 8,16	80,01 ± 3,75
RF - 10 árvores	55,38 ± 3,90	0,40 ± 0,05	43,52 ± 8,34	81,30 ± 3,50
RF - 20 árvores	61,28 ± 3,69	0,48 ± 0,05	42,12 ± 8,12	86,65 ± 3,23
RF - 50 árvores	67,12 ± 3,73	0,56 ± 0,05	40,37 ± 8,28	91,81 ± 2,59
RF - 100 árvores	69,59 ± 3,56	0,59 ± 0,05	39,68 ± 7,92	94,22 ± 2,33
SVM - Polinômio 1º	58,44 ± 3,92	0,45 ± 0,05	43,20 ± 7,70	85,53 ± 3,55
SVM - Polinômio 2º	63,99 ± 3,74	0,52 ± 0,05	54,62 ± 7,89	83,17 ± 3,37
SVM - Polinômio 3º	67,14 ± 3,70	0,56 ± 0,05	55,38 ± 7,24	85,53 ± 3,34
SVM - Polinômio 4º	68,85 ± 3,74	0,58 ± 0,05	56,03 ± 7,38	86,65 ± 3,26
SVM - Polinômio 5º	64,76 ± 4,83	0,53 ± 0,06	54,60 ± 9,79	83,44 ± 6,99
SVM - RBF $\gamma = 0,01$	42,92 ± 3,46	0,24 ± 0,05	1,03 ± 1,70	99,29 ± 0,99
SVM - RBF $\gamma = 0,25$	67,47 ± 3,64	0,57 ± 0,05	47,66 ± 8,21	90,75 ± 2,65
SVM - RBF $\gamma = 0,5$	73,47 ± 3,73	0,65 ± 0,05	56,89 ± 8,50	88,84 ± 3,02
MLP - 20 neurônios	57,65 ± 4,38	0,44 ± 0,05	44,05 ± 10,20	86,23 ± 5,70
MLP - 50 neurônios	56,72 ± 5,08	0,42 ± 0,06	41,01 ± 12,80	86,62 ± 8,25
MLP - 100 neurônios	49,12 ± 5,51	0,32 ± 0,07	37,66 ± 17,27	83,39 ± 15,53

Fonte: o autor (2022).

Figura 35 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *ResNet50* e selecionados usando a técnica PSO.



Fonte: o autor (2022).

Na Tabela 25 estão os resultados de Área Sob Curva ROC e Tempo de Treinamento (em milissegundos) para a *ResNet50*. O *Random Forest* para 100 árvores obteve o melhor resultado para Área Sob Curva ROC, com 0,80 e 0,05 de desvio padrão. O Tempo de Treinamento em milissegundos foi menor para o *Naive Bayes Multinomial*, com 3,65 e 1,34 de desvio padrão.

Tabela 25 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *ResNet50* e selecionados usando PSO. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.

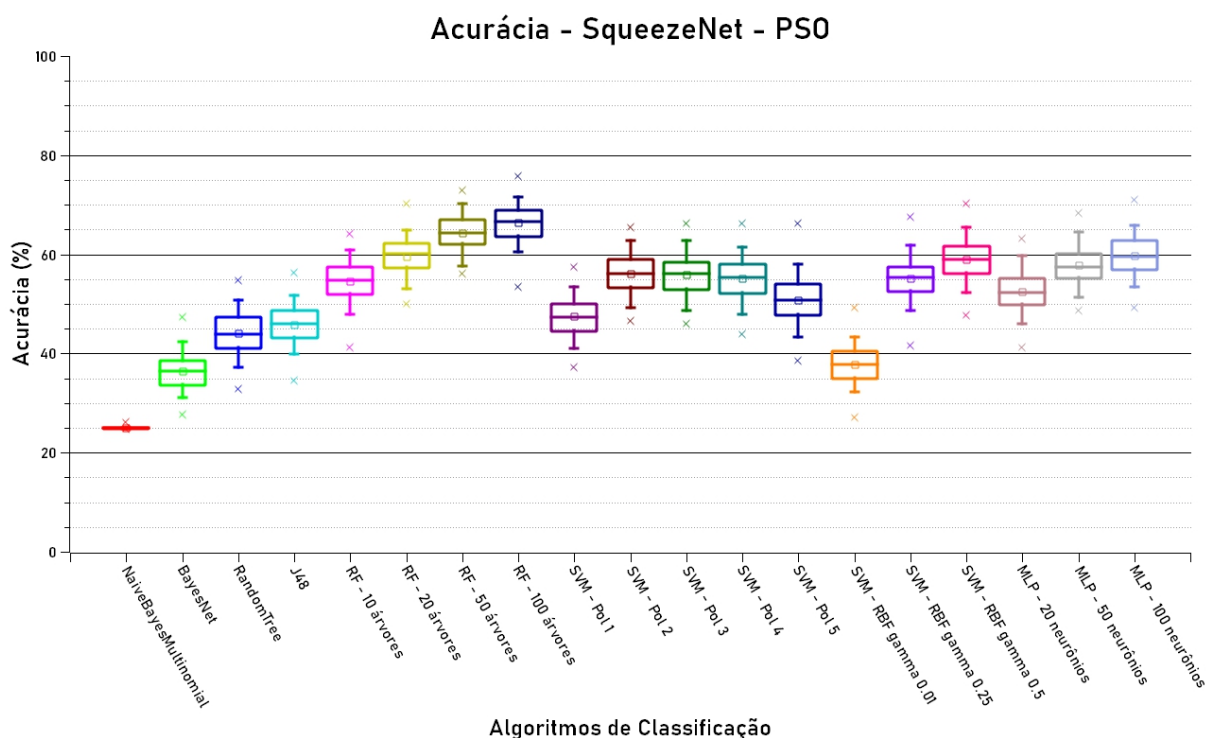
Rede Neural Convolutacional ResNet50 - PSO		
Classificador	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,53 ± 0,05	3,65 ± 1,34
Bayes Net	0,59 ± 0,05	143,60 ± 28,48
Random Tree	0,57 ± 0,04	27,55 ± 6,46
J48	0,60 ± 0,05	1419,75 ± 432,41
RF - 10 árvores	0,68 ± 0,05	169,36 ± 45,26
RF - 20 árvores	0,72 ± 0,05	345,25 ± 119,18
RF - 50 árvores	0,77 ± 0,05	951,85 ± 338,55
RF - 100 árvores	0,80 ± 0,05	2085,65 ± 663,85
SVM - Polinômio 1º	0,67 ± 0,05	2239,05 ± 767,19
SVM - Polinômio 2º	0,72 ± 0,04	21080,92 ± 8791,98
SVM - Polinômio 3º	0,73 ± 0,05	14826,37 ± 4868,56
SVM - Polinômio 4º	0,74 ± 0,05	10650,11 ± 4055,96
SVM - Polinômio 5º	0,72 ± 0,05	2266,50 ± 1035,92
SVM - RBF $\gamma = 0,01$	0,59 ± 0,05	2166,86 ± 680,77
SVM - RBF $\gamma = 0,25$	0,74 ± 0,05	2129,79 ± 680,77
SVM - RBF $\gamma = 0,5$	0,77 ± 0,05	2181,17 ± 702,60
MLP - 20 neurônios	0,69 ± 0,05	118296,49 ± 38000,12
MLP - 50 neurônios	0,69 ± 0,05	296532,87 ± 79614,98
MLP - 100 neurônios	0,65 ± 0,05	563503,09 ± 205731,81

Fonte: o autor (2022).

Nas Tabelas 26 e 27, são apresentados os resultados de classificação para a arquitetura CNN *SqueezeNet* com PSO. Pela Tabela 26, para a taxa de Acurácia (Figura 36) e Índice Kappa (Figura 37), o classificador *Random Forest* com 100 árvores apresentou os melhores resultados, com Acurácia de 66,32% e desvio padrão de 3,67, e Kappa de 0,55 e desvio padrão de 0,05. A taxa de Sensibilidade teve o melhor desempenho para o SVM com kernel polinomial de 2º, obtendo 55,30% e desvio padrão de 8,23. Para Especificidade, o SVM com kernel RBF $\gamma = 0,01$ apresentou o melhor resultado com 96,90% e 1,71 de desvio padrão.

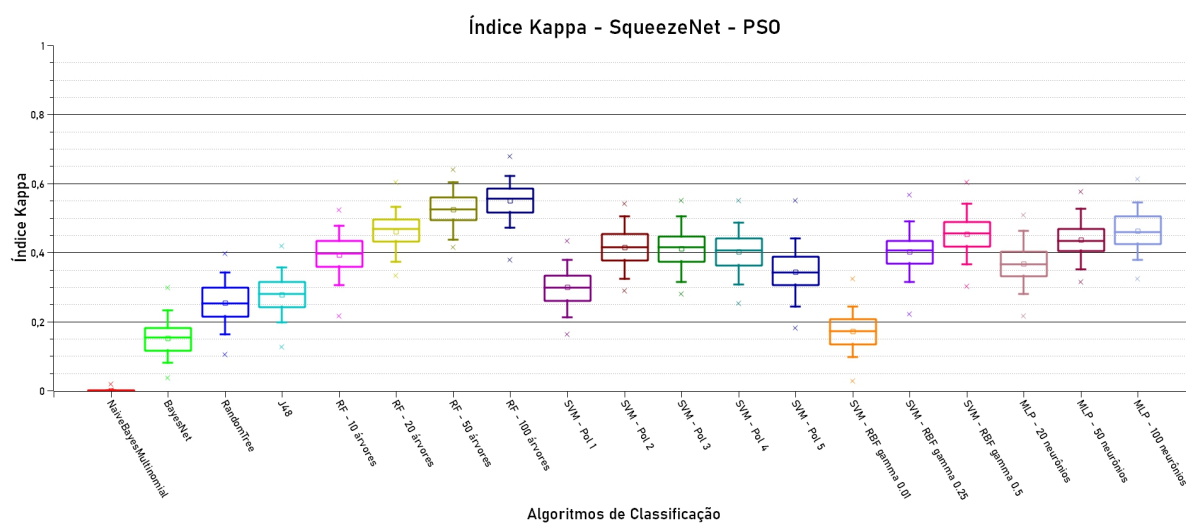
Para *SqueezeNet*, Tabela 12, a métrica de Área Sob Curva ROC teve 0,80 e desvio padrão de 0,04 como seu melhor resultado, obtido com o classificador *Random Forest* com

Figura 36 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN *SqueezeNet* e selecionados usando PSO.



Fonte: o autor (2022).

Figura 37 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *SqueezeNet* e selecionados usando PSO.



Fonte: o autor (2022).

Tabela 26 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *SqueezeNet* e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

Classificador	Rede Neural Convolutacional SqueezeNet - PSO			
	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	24,98 ± 0,11	0,00 ± 0,00	10,00 ± 30,05	90,01 ± 30,01
Bayes Net	36,43 ± 3,49	0,15 ± 0,05	0,25 ± 0,06	16,25 ± 5,19
Random Tree	44,06 ± 4,17	0,25 ± 0,05	38,45 ± 7,85	79,78 ± 3,79
J48	45,87 ± 3,68	0,28 ± 0,05	43,21 ± 7,78	80,10 ± 3,80
RF - 10 árvores	54,53 ± 3,96	0,39 ± 0,05	48,54 ± 8,72	81,13 ± 3,86
RF - 20 árvores	59,62 ± 3,63	0,46 ± 0,05	47,50 ± 8,29	85,81 ± 3,14
RF - 50 árvores	64,37 ± 3,66	0,52 ± 0,05	47,26 ± 8,06	89,51 ± 2,88
RF - 100 árvores	66,32 ± 3,67	0,55 ± 0,05	46,81 ± 7,96	90,94 ± 2,79
SVM - Polinômio 1º	47,48 ± 3,88	0,30 ± 0,05	42,62 ± 7,71	81,09 ± 3,97
SVM - Polinômio 2º	56,18 ± 4,11	0,42 ± 0,05	55,30 ± 8,23	78,60 ± 3,81
SVM - Polinômio 3º	55,92 ± 4,12	0,41 ± 0,05	54,68 ± 8,51	77,92 ± 3,93
SVM - Polinômio 4º	50,84 ± 4,62	0,40 ± 0,06	52,73 ± 9,37	75,11 ± 6,95
SVM - Polinômio 5º	50,84 ± 4,62	0,34 ± 0,06	52,73 ± 9,37	75,11 ± 6,95
SVM - RBF $\gamma = 0,01$	37,80 ± 3,61	0,17 ± 0,05	12,62 ± 5,27	96,90 ± 1,71
SVM - RBF $\gamma = 0,25$	55,18 ± 3,94	0,40 ± 0,05	45,20 ± 7,66	87,68 ± 3,14
SVM - RBF $\gamma = 0,5$	58,98 ± 3,92	0,45 ± 0,05	45,20 ± 7,66	86,95 ± 3,37
MLP - 20 neurônios	52,53 ± 4,10	0,37 ± 0,05	46,92 ± 8,58	82,44 ± 4,26
MLP - 50 neurônios	57,73 ± 3,85	0,44 ± 0,05	49,91 ± 7,50	83,68 ± 3,76
MLP - 100 neurônios	59,68 ± 3,83	0,46 ± 0,05	50,75 ± 7,57	84,92 ± 3,47

Fonte: o autor (2022).

100 árvores. Já o Tempo de Treinamento (em milissegundos) foi menor para o *Random Tree* com 0,07 ms e desvio padrão de 0,01.

A CNN *Inception v3*, obteve os resultados mostrados por meio das Tabelas 28 e 29. Pela Tabela 28, o SVM com kernel polinomial de grau 4 teve os melhores resultados para Acurácia (Figura 38) e Kappa (Figura 39), com Acurácia de 78,55% e desvio padrão de 3,04, e Kappa de 0,71 e 0,05 de desvio padrão. Para a Sensibilidade, o melhor resultado foi o SVM com kernel RBF $\gamma = 0,5$, com taxa de 99,49%. Já para a Especificidade, o melhor resultado foi para o *Random Forest* com 100 árvores: 93,54% e 2,16 de desvio padrão.

Os resultados das métricas de Área Sob Curva ROC e Tempo de Treinamento (em milissegundos), mostrados na Tabela 29 para a *Inception v3*, foram: SVM com kernel RBF de $\gamma = 0,25$ para Área Sob Curva ROC com taxa de 0,83 e desvio padrão de 0,03; e *Naive Bayes Multinomial* com o menor Tempo de Treinamento, com 4,85 ms e desvio padrão de 2,50.

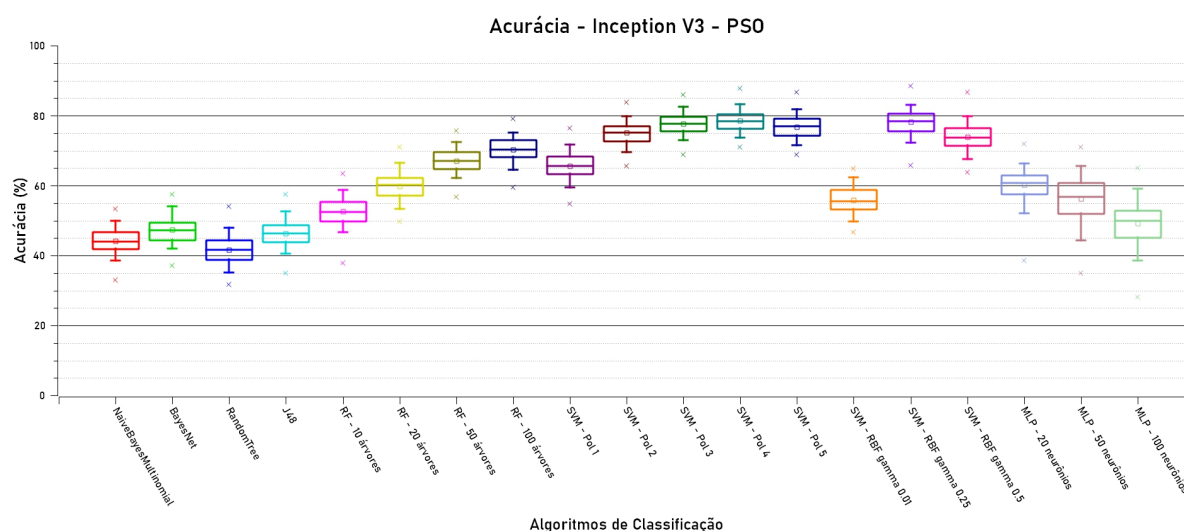
A Tabela 30 mostra a comparação entre os melhores classificadores por CNN

Tabela 27 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional *SqueezeNet* e selecionados usando PSO. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.

Rede Neural Convolutacional SqueezeNet- PSO		
Classificador	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,52 ± 0,05	1,63 ± 1,33
Bayes Net	0,59 ± 0,05	51,80 ± 17,73
Random Tree	0,59 ± 0,05	22,15 ± 12,20
J48	0,62 ± 0,05	580,61 ± 207,43
RF - 10 árvores	0,70 ± 0,05	148,51 ± 51,21
RF - 20 árvores	0,74 ± 0,05	300,01 ± 94,39
RF - 50 árvores	0,77 ± 0,05	717,99 ± 266,65
RF - 100 árvores	0,79 ± 0,05	1557,81 ± 477,74
SVM - Polinômio 1º	0,65 ± 0,05	1163,91 ± 342,14
SVM - Polinômio 2º	0,70 ± 0,04	53625,00 ± 20478,03
SVM - Polinômio 3º	0,69 ± 0,05	76970,54 ± 23947,52
SVM - Polinômio 4º	0,69 ± 0,05	66505,00 ± 19537,47
SVM - Polinômio 5º	0,68 ± 0,05	29640,90 ± 15886,91
SVM - RBF $\gamma = 0,01$	0,58 ± 0,05	1264,64 ± 402,50
SVM - RBF $\gamma = 0,25$	0,69 ± 0,05	1161,99 ± 452,53
SVM - RBF $\gamma = 0,5$	0,70 ± 0,05	1338,57 ± 380,68
MLP - 20 neurônios	0,70 ± 0,05	37360,12 ± 13143,53
MLP - 50 neurônios	0,73 ± 0,05	92298,51 ± 30223,81
MLP - 100 neurônios	0,74 ± 0,04	179696,22 ± 65342,26

Fonte: o autor (2022).

Figura 38 – Box plot para a taxa de Acurácia para cada classificador usando os atributos extraídos com a CNN *Inception v3* e selecionados usando PSO.



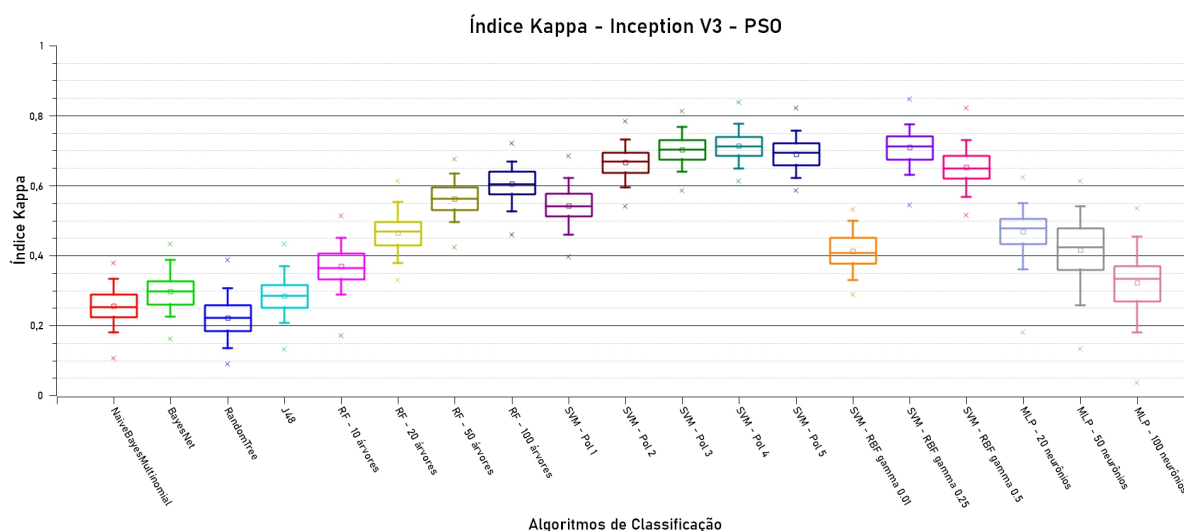
Fonte: o autor (2022).

Tabela 28 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional InceptionV3 e selecionados usando PSO. São mostradas nessa tabela as seguintes métricas estatísticas: Acurácia, Índice Kappa, Sensibilidade e Especificidade.

Classificador	Rede Neural Convolutacional InceptionV3 - PSO			
	Acurácia (%)	Índice Kappa	Sensibilidade (%)	Especificidade (%)
Naive Bayes Multinomial	44,19 ± 3,46	0,26 ± 0,05	15,29 ± 5,45	91,66 ± 2,59
Bayes Net	47,33 ± 3,65	0,30 ± 0,05	20,32 ± 6,14	92,99 ± 2,52
Random Tree	41,63 ± 3,97	0,22 ± 0,05	31,86 ± 7,34	78,88 ± 3,72
J48	46,37 ± 3,77	0,28 ± 0,05	37,77 ± 7,41	79,58 ± 3,52
RF - 10 árvores	52,64 ± 3,94	0,37 ± 0,05	42,17 ± 8,55	78,59 ± 4,11
RF - 20 árvores	59,84 ± 3,89	0,46 ± 0,05	41,20 ± 8,42	84,83 ± 3,47
RF - 50 árvores	67,10 ± 3,25	0,56 ± 0,05	40,14 ± 8,29	90,75 ± 2,57
RF - 100 árvores	70,35 ± 3,33	0,60 ± 0,05	39,14 ± 8,31	93,54 ± 2,16
SVM - Polinômio 1º	65,67 ± 3,62	0,54 ± 0,05	50,88 ± 7,39	83,18 ± 3,61
SVM - Polinômio 2º	75,06 ± 3,18	0,67 ± 0,04	54,90 ± 7,72	90,88 ± 2,60
SVM - Polinômio 3º	77,72 ± 3,05	0,70 ± 0,04	56,22 ± 7,70	92,67 ± 2,46
SVM - Polinômio 4º	78,55 ± 3,04	0,71 ± 0,04	56,32 ± 7,96	93,21 ± 2,36
SVM - Polinômio 5º	76,76 ± 3,15	0,69 ± 0,05	54,41 ± 9,08	92,24 ± 2,97
SVM - RBF $\gamma = 0,01$	55,92 ± 3,78	0,41 ± 0,05	25,88 ± 7,12	92,40 ± 2,54
SVM - RBF $\gamma = 0,25$	78,18 ± 3,51	0,71 ± 0,05	84,67 ± 6,39	77,80 ± 4,28
SVM - RBF $\gamma = 0,5$	73,87 ± 3,82	0,65 ± 0,05	99,49 ± 1,19	65,34 ± 5,05
MLP - 20 neurônios	60,13 ± 4,58	0,47 ± 0,05	45,66 ± 11,95	84,03 ± 6,95
MLP - 50 neurônios	56,15 ± 6,55	0,42 ± 0,09	38,92 ± 17,91	84,81 ± 12,52
MLP - 100 neurônios	49,16 ± 6,11	0,32 ± 0,08	33,48 ± 19,09	84,16 ± 15,77

Fonte: o autor (2022).

Figura 39 – Box plot do Índice Kappa para cada classificador usando os atributos extraídos com a CNN *Inception v3* e selecionados usando PSO.



Fonte: o autor (2022).

Tabela 29 – Resultados do treinamento e classificação para diferentes classificadores usando os atributos extraídos com a Rede Convolutacional InceptionV3 e selecionados usando PSO. São mostradas nessa tabela as métricas de Área Sob Curva ROC e o Tempo de Treinamento.

Rede Neural Convolutacional InceptionV3 - PSO		
Classificador	Área Sob Curva ROC	Tempo de Treinamento (ms)
Naive Bayes Multinomial	0,59 ± 0,05	4,85 ± 2,50
Bayes Net	0,67 ± 0,05	250,23 ± 98,86
Random Tree	0,55 ± 0,04	39,31 ± 13,62
J48	0,59 ± 0,05	4151,55 ± 13388,12
RF - 10 árvores	0,64 ± 0,05	273,63 ± 103,79
RF - 20 árvores	0,69 ± 0,05	560,87 ± 186,61
RF - 50 árvores	0,75 ± 0,05	1347,19 ± 481,04
RF - 100 árvores	0,79 ± 0,05	2542,37 ± 827,52
SVM - Polinômio 1º	0,69 ± 0,05	5572,29 ± 1813,91
SVM - Polinômio 2º	0,74 ± 0,04	5498,28 ± 1426,09
SVM - Polinômio 3º	0,76 ± 0,05	4992,19 ± 1600,89
SVM - Polinômio 4º	0,77 ± 0,04	3629,94 ± 1502,63
SVM - Polinômio 5º	0,76 ± 0,04	1978,87 ± 638,04
SVM - RBF $\gamma = 0,01$	0,65 ± 0,05	2232,08 ± 815,83
SVM - RBF $\gamma = 0,25$	0,83 ± 0,03	2354,31 ± 836,79
SVM - RBF $\gamma = 0,5$	0,82 ± 0,03	2270,53 ± 689,97
MLP - 20 neurônios	0,69 ± 0,05	147604,45 ± 60481,73
MLP - 50 neurônios	0,67 ± 0,06	376842,67 ± 113954,99
MLP - 100 neurônios	0,64 ± 0,06	736851,19 ± 251278,01

Fonte: o autor (2022).

para os resultados sobre os atributos selecionados usando PSO. Entre todas as Redes Convolucionais, a *Inception v3* teve a melhor taxa de acurácia com o classificador SVM de kernel polinomial de grau 4, alcançando 78,55% e desvio padrão de 3,04 de acurácia. O Índice Kappa, de acordo com a Tabela 31, obteve melhor desempenho para a *Inception v3* para o SVM com kernel polinomial de grau 4, com taxa de 0,71 e desvio padrão de 0,05. Para Sensibilidade, Tabela 32, a *Inception v3* apresentou o melhor desempenho para o SVM com kernel RBF $\gamma = 0,5$ obtendo 99,49% e desvio padrão de 1,19. Para Especificidade, Tabela 33, a *ResNet50* teve destaque com o classificador *Bayes Net* com uma taxa de 99,42 e desvio padrão de 0,84. O *Inception v3* para o SVM de kernel RBF de $\gamma = 0,25$ apresentou a melhor taxa de Área Sob Curva ROC, com 0,83 e desvio padrão de 0,03, conforme Tabela 34.

Tabela 30 – Comparação entre a taxa de acurácia dos melhores classificadores para cada Rede Neural Convolutacional com PSO.

Comparação da Taxa de Acurácia por CNN com PSO		
CNN	Classificador	Acurácia (%)
LeNet	RF - 100 árvores	65,37 ± 3,62
NASNetMobile	RF - 100 árvores	55,97 ± 3,71
ResNet50	SVM - RBF $\gamma = 0,5$	73,47 ± 3,73
SqueezeNet	RF - 100 árvores	66,32 ± 3,67
InceptionV3	SVM - Polinômio 4º	78,55 ± 3,04

Fonte: o autor (2022).

Tabela 31 – Comparação entre a taxa de índice kappa dos melhores classificadores para cada Rede Neural Convolutacional com PSO.

Comparação do Índice Kappa por CNN com PSO		
CNN	Classificador	Índice Kappa
LeNet	RF - 100 árvores	0,54 ± 0,05
NASNetMobile	RF - 100 árvores	0,41 ± 0,05
ResNet50	SVM - RBF $\gamma = 0,5$	0,65 ± 0,05
SqueezeNet	RF - 100 árvores	0,55 ± 0,05
InceptionV3	SVM - Polinômio 4º	0,71 ± 0,04

Fonte: o autor (2022).

Tabela 32 – Comparação entre a taxa de sensibilidade dos melhores classificadores para cada Rede Neural Convolutacional.

Comparação da Taxa de Sensibilidade por CNN com PSO		
CNN	Classificador	Sensibilidade (%)
LeNet	SVM - Polinômio 2º	47,16 ± 8,39
NASNetMobile	SVM - RBF $\gamma = 0,01$	42,57 ± 41,13
ResNet50	SVM - RBF $\gamma = 0,5$	56,89 ± 8,50
SqueezeNet	SVM - Polinômio 2º	55,30 ± 8,23
InceptionV3	SVM - RBF $\gamma = 0,5$	99,49 ± 1,19

Fonte: o autor (2022).

Tabela 33 – Comparação entre a taxa de especificidade dos melhores classificadores para cada Rede Neural Convolutacional com PSO.

Comparação da Taxa de Especificidade por CNN com PSO		
CNN	Classificador	Especificidade (%)
LeNet	SVM - RBF $\gamma = 0,01$	94,73 ± 2,35
NASNetMobile	SVM - Polinômio 4º	92,77 ± 2,54
ResNet50	Bayes Net	99,42 ± 0,84
SqueezeNet	SVM - RBF $\gamma = 0,01$	96,90 ± 1,71
InceptionV3	RF - 100 árvores	93,54 ± 2,16

Fonte: o autor (2022).

Tabela 34 – Comparação entre a taxa de área sob curva ROC dos melhores classificadores para cada Rede Neural Convolutacional com PSO.

Comparação da Área Sob Curva ROC por CNN com PSO		
CNN	Classificador	Área Sob Curva ROC
LeNet	RF - 100 árvores	0,77 ± 0,04
NASNetMobile	RF - 100 árvores	0,67 ± 0,05
ResNet50	RF - 100 árvores	0,80 ± 0,05
SqueezeNet	RF - 100 árvores	0,79 ± 0,05
InceptionV3	SVM - RBF $\gamma = 0,25$	0,83 ± 0,03

Fonte: o autor (2022).

5.3 Comparação dos resultados antes e após a seleção de atributos

Nessa sessão são comparados os resultados antes e depois de ser utilizado o PSO para seleção de atributos. A Tabela 35 mostra a comparação entre os resultados das melhores técnicas para Acurácia. A Acurácia foi um pouco maior para a abordagem sem seleção de atributos, porém o Tempo de Treinamento para essa abordagem ($7744,16 \pm 2350,95$ ms) foi duas vezes maior comparado ao tempo da abordagem com PSO ($3629,94 \pm 1502,63$ ms).

Tabela 35 – Comparação entre a Acurácia antes e depois da seleção de atributos usando PSO.

Comparação da Acurácia Antes e Após Seleção de Atributos com PSO				
Abordagens	CNN	Classificador	Tempo de Treinamento (ms)	Acurácia (%)
Sem PSO	InceptionV3	SVM - Polinômio 3º	$7744,16 \pm 2350,95$	$79,92 \pm 3,07$
Com PSO	InceptionV3	SVM - Polinômio 4º	$3629,94 \pm 1502,63$	$78,55 \pm 3,04$

Fonte: o autor (2022).

A Tabela 36 mostra a comparação entre os resultados para o Índice Kappa. Semelhante a Acurácia, a abordagem utilizando PSO teve um Tempo de Treinamento ($3629,94 \pm 1502,63$ ms) duas vezes menor comparada a outra abordagem ($7744,16 \pm 2350,95$ ms), porém esta última teve um Índice Kappa um pouco maior ($0,73 \pm 0,04$).

Tabela 36 – Comparação entre o índice kappa antes e depois da seleção de atributos usando PSO.

Comparação do Índice Kappa Antes e Após Seleção de Atributos com PSO				
Abordagens	CNN	Classificador	Tempo de Treinamento (ms)	Kappa
Sem PSO	InceptionV3	SVM - Polinômio 3º	$7744,16 \pm 2350,95$	$0,73 \pm 0,04$
Com PSO	InceptionV3	SVM - Polinômio 4º	$3629,94 \pm 1502,63$	$0,71 \pm 0,04$

Fonte: o autor (2022).

De forma semelhante, a Sensibilidade foi comparada pela Tabela 37. A diferença entre os resultados de Sensibilidade foi pequena, 100 % (sem PSO) e 99,49 % (com PSO).

Já o Tempo de Treinamento foi maior para a abordagem sem utilizar PSO ($5644,23 \pm 1777,81$ ms).

Tabela 37 – Comparação entre a Sensibilidade antes e depois da seleção de atributos usando PSO.

Comparação da Sensibilidade Antes e Após Seleção de Atributos com PSO				
Abordagens	CNN	Classificador	Tempo de Treinamento (ms)	Sensibilidade (%)
Sem PSO	InceptionV3	SVM - RBF $\gamma = 0,25$	$5644,23 \pm 1777,81$	$100,00 \pm 0,00$
Com PSO	InceptionV3	SVM - RBF $\gamma = 0,5$	$2270,53 \pm 689,97$	$99,49 \pm 1,19$

Fonte: o autor (2022).

Para Especificidade (Tabela 38), o maior resultado foi obtido com a aplicação do PSO ($99,42\% \pm 0,84$). O Tempo de Treinamento foi 4 vezes maior para a abordagem sem PSO ($692,69 \pm 236,09$ ms) comparada à abordagem com PSO ($143,60 \pm 28,48$ ms).

Tabela 38 – Comparação entre a Especificidade antes e depois da seleção de atributos usando PSO.

Comparação da Especificidade Antes e Após Seleção de Atributos com PSO				
Abordagens	CNN	Classificador	Tempo de Treinamento (ms)	Especificidade (%)
Sem PSO	ResNet50	Bayes Net	$692,69 \pm 236,09$	$99,03 \pm 0,99$
Com PSO	ResNet50	Bayes Net	$143,60 \pm 28,48$	$99,42 \pm 0,84$

Fonte: o autor (2022).

A Área Sob Curva ROC (Tabela 39) obteve um melhor resultado usando PSO ($0,83 \pm 0,03$), com um tempo semelhante a abordagem sem seleção de atributos.

Tabela 39 – Comparação entre a Área Sob Curva ROC antes e depois da seleção de atributos usando PSO.

Comparação da Área Sob Curva ROC Antes e Após Seleção de Atributos com PSO				
Abordagens	CNN	Classificador	Tempo de Treinamento (ms)	ROC
Sem PSO	NASNetMobile	SVM - RBF $\gamma = 0,5$	$2718,24 \pm 530,97$	$0,80 \pm 0,03$
	SqueezeNet	RF - 100 árvores	$3468,45 \pm 1185,11$	$0,80 \pm 0,04$
Com PSO	InceptionV3	SVM - RBF $\gamma = 0,25$	$2354,31 \pm 836,79$	$0,83 \pm 0,03$

Fonte: o autor (2022).

Sendo assim, a partir de uma análise geral sobre todas as métricas, é possível concluir que o uso da técnica de PSO reduz consideravelmente o Tempo de Treinamento ao mesmo tempo que a perda de desempenho pode ser considerada insignificante.

6 CONCLUSÃO

O uso da termografia para detecção do câncer de mama é uma técnica complementar que pode ser utilizada como exame de triagem para o diagnóstico desse tipo de câncer. A termografia destaca-se por ser um exame não invasivo, indolor, não emissor de radiação ionizante e apresenta um bom desempenho na detecção de lesões mamárias em mamas densas. Logo, a termografia é vista como uma técnica promissora que pode auxiliar o diagnóstico de lesões mamárias em estágios iniciais, contribuindo assim para a redução da taxa de mortalidade por câncer de mama.

Neste trabalho, foram realizadas duas abordagens: uma utilizando os vetores de atributos originais (extraídos com as CNNs) e outra utilizando os vetores de atributos reduzidos (extraídos com as CNNs e selecionados com PSO). Em cada abordagem, foram utilizadas redes neurais profundas para extrair os atributos da base de dados de imagens termográficas. Além disso, foram utilizados alguns classificadores para treinamento usando o respectivo conjunto de atributos. Já para a abordagem usando os vetores de atributos reduzidos, foi utilizada a técnica de PSO para seleção dos melhores atributos. Dessa maneira, utilizando os vetores de atributos completos, o melhor resultado da acurácia foi de 79,92% utilizando a CNN *Inception v3* combinada com o classificador SVM Polinomial de 3º. Já utilizando os vetores com os atributos selecionados com PSO, o melhor resultado da acurácia foi de 78,55%, também usando a CNN *Inception v3* combinada com SVM Polinomial de 4º. Portanto, foi possível observar que a abordagem utilizando seleção de atributos foi bastante favorável, porque foi possível reduzir a quantidade de atributos e o tempo de treinamento sem causar uma redução significativa na taxa de acurácia e nas outras métricas utilizadas.

Com base no que foi apresentado, os resultados mostraram-se interessantes e reforçaram a ideia de utilizar a termografia de mama como uma ferramenta de auxílio na detecção do câncer de mama, combinada a máquinas de aprendizado precisas o suficiente para apoiar o processo de tomada de decisão diagnóstica.

6.1 Dificuldades apresentadas

A principal dificuldade enfrentada foi com relação ao longo tempo de treinamento para abordagem sem PSO. Nessa abordagem, os experimentos demoraram várias semanas

para serem concluídos, o que dificultou a realização de uma análise mais célere dos resultados, bem como nos casos onde foi preciso fazer alguma alteração em parâmetros da rede ou quando houve intermitência no treinamento por motivos diversos, sendo necessário assim recomeçar os experimentos.

6.2 Contribuições e trabalhos futuros

Em virtude do que foi apresentado, esse trabalho contribuiu para análise de desempenho da combinação de diferentes técnicas de redes neurais profundas com diferentes classificadores. A parcela principal de contribuição ocorreu a partir da fundamentação de um modelo base para classificação e diagnóstico do câncer de mama usando termografia.

Para trabalhos futuros, sugere-se utilizar os modelos experimentados nesse trabalho para construção de um sistema web ou aplicativo móvel, de modo a ser utilizado como ferramenta de apoio ao diagnóstico do câncer de mama. A exploração da seleção de atributos, com o objetivo de reduzi-los, pode contribuir para construção de modelos inteligentes mais leves que podem ser utilizados em ambientes com restrição de recursos computacionais. Recomenda-se também, baseado na metodologia apresentada, a realização de experimentos com outras redes neurais profundas.

REFERÊNCIAS

- ANDRADE, F. Análise de Imagens de Termografia Dinâmica para Classificação de Alterações na Mama Usando Séries Temporais. *Conference on Graphics, Patterns and Images is an international conference annually promoted by the Brazilian Computer Society*, 2017. Disponível em: <http://sibgrapi.sid.inpe.br/col/sid.inpe.br/sibgrapi/2017/09.04.22.16/doc/CameraReady_Sibgrapi_termica.pdf>.
- AZEVEDO, R. do Socorro da S.; OLIVEIRA, C. S.; VASCONCELOS, P. F. da C. Risco do chikungunya para o brasil. *Revista de Saúde Pública*, SciELO, p. 49–51, 2015.
- BARBOSA, V. A. d. F.; GOMES, J. C.; SANTANA, M. A. de; LIMA, C. L. de; CALADO, R. B.; JUNIOR, C. R. B.; ALBUQUERQUE, J. E. d. A.; SOUZA, R. G. de; ARAÚJO, R. J. E. de; JUNIOR, L. A. R. M. et al. Covid-19 rapid test by combining a random forest-based web system and blood tests. *Journal of Biomolecular Structure and Dynamics*, Taylor & Francis, p. 1–20, 2021.
- BARBOSA, V. A. de F.; GOMES, J. C.; SANTANA, M. A. de; ALBUQUERQUE, J. E. d. A.; SOUZA, R. G. de; SOUZA, R. E. de; SANTOS, W. P. dos. Heg.ia: An intelligent system to support diagnosis of covid-19 based on blood tests. *Research on Biomedical Engineering*, Springer, p. 1–18, 2021.
- BARBOSA, V. A. de F.; SANTANA, M. A. de; ANDRADE, M. K. S.; LIMA, R. d. C. F. de; SANTOS, W. P. dos. Deep-wavelet neural networks for breast cancer early diagnosis using mammary termographies. In: *Deep Learning for Data Analytics*. [S.l.]: Elsevier, 2020. p. 99–124.
- BERNARDES, N. B.; SÁ, A. C. F. de; FACIOLI, L. de S.; FERREIRA, M. L.; SÁ, O. R. de; COSTA, R. de M. Câncer de mama x diagnóstico/breast cancer x diagnosis. *ID on line REVISTA DE PSICOLOGIA*, v. 13, n. 44, p. 877–885, 2019.
- BERRAR, D. Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, v. 1-3, n. January 2018, p. 542–545, 2018.
- BLAGUS, R.; LUSA, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, v. 14, 2013. ISSN 14712105.
- BOQUETE, L.; ORTEGA, S.; MIGUEL-JIMÉNEZ, J. M.; RODRÍGUEZ-ASCARIZ, J. M.; BLANCO, R. Automated detection of breast cancer in thermal infrared images, based on independent component analysis. *Journal of medical systems*, Springer, v. 36, n. 1, p. 103–111, 2012.
- CALDEIRA, A. Um olhar crítico sobre o autoexame das mamas. 06 2015.
- CAMPOS A. P. DE; CAMARGO, R. *Ultrassonografia, Mamografia e Densitometria Óssea*. São Paulo: Érica / Saraiva, 2015.
- CASTRO, F. A. D. Impacto do autoexame das mamas no diagnóstico de câncer de mama em países de média e baixa renda : uma revisão de literatura Impact of breast self-examination on breast cancer diagnosis in medium and low income countries : a literature review. p. 2973–2996, 2021.

CHAVES, E.; GONÇALVES, C. B.; ALBERTINI, M. K.; LEE, S.; JEON, G.; FERNANDES, H. C. Evaluation of transfer learning of pre-trained cnns applied to breast cancer detection on infrared images. *Applied Optics*, Optical Society of America, v. 59, n. 17, p. E23–E28, 2020.

CHEBBAH, N.; OUSLIM, M.; RYAD, T. A new approach for breast abnormality detection based on thermography. *Medical Technologies Journal*, v. 2, p. 245–254, 09 2018.

CHEBBAH, N. K.; OUSLIM, M.; BENABID, S. New computer aided diagnostic system using deep neural network and svm to detect breast cancer in thermography. *Quantitative InfraRed Thermography Journal*, Taylor & Francis, v. 0, n. 0, p. 1–16, 2022. Disponível em: <<https://doi.org/10.1080/17686733.2021.2025018>>.

CHEN, J.; RAN, X. Deep Learning With Edge Computing: A Review. *Proceedings of the IEEE*, p. 1–19, 2019. ISSN 15582256.

COMMOWICK, O.; ISTACE, A.; KAIN, M.; LAURENT, B.; LERAY, F.; SIMON, M.; POP, S. C.; GIRARD, P.; AMELI, R.; FERRÉ, J.-C. et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 1–17, 2018.

CORDEIRO, F. R.; BEZERRA, K. F. P.; SANTOS, W. P. D. Random walker with fuzzy initialization applied to segment masses in mammography images. In: IEEE. *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. [S.l.], 2017. p. 156–161.

CORDEIRO, F. R.; LIMA, S. M.; SILVA-FILHO, A. G.; SANTOS, W. Segmentation of mammography by applying extreme learning machine in tumor detection. In: SPRINGER. *International Conference on Intelligent Data Engineering and Automated Learning*. [S.l.], 2012. p. 92–100.

CORDEIRO, F. R.; SANTOS, W. P.; SILVA-FILHO, A. G. An adaptive semi-supervised fuzzy growcut algorithm to segment masses of regions of interest of mammographic images. *Applied Soft Computing*, Elsevier, v. 46, p. 613–628, 2016.

CORDEIRO, F. R.; SANTOS, W. P.; SILVA-FILHO, A. G. A semi-supervised fuzzy growcut algorithm to segment and classify regions of interest of mammographic images. *Expert Systems with Applications*, Elsevier, v. 65, p. 116–126, 2016.

CORDEIRO, F. R.; SANTOS, W. P. d.; SILVA-FILHO, A. G. Analysis of supervised and semi-supervised growcut applied to segmentation of masses in mammography images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Taylor & Francis, v. 5, n. 4, p. 297–315, 2017.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995.

CRUZ, T. N. D. SELEÇÃO DE ATRIBUTOS DE IMAGEM BASEADA EM ALGORITMO GENÉTICO E INTELIGÊNCIA COMPUTACIONAL PARA O DIAGNÓSTICO DO CÂNCER DE MAMA UTILIZANDO MAMOGRAFIAS DIGITAIS . Dissertação (Mestrado) - Universidade Federal de Pernambuco, 2017.

DALVI, R. D. F. Suporte a Detecção e Classificação de Câncer a partir de Mamografias Digitalizadas e Redes Neurais Convolucionais Suporte a Detecção e Classificação de Câncer a partir de Mamografias Digitalizadas e Redes Neurais Convolucionais. 2018.

Didática Tech. *O que é e como funciona o algoritmo Random Forest?* 1997. Disponível em: <https://didatica.tech/o-que-e-e-como-funciona-o-algoritmo-randomforest/>. Acesso em: 15/08/2021.

DIREITO, D. D. E.; POLI, L. C. Revista de Direitos Sociais, Seguridade e Previdência Social. p. 60–80, 2017.

Do Nascimento, F. B.; Da Rocha Pitta, M. G.; De Melo Rêgo, M. J. B. Análise dos principais métodos de diagnóstico de câncer de mama como propulsores no processo inovativo. *Arquivos de Medicina*, v. 29, n. 6, p. 153–159, 2015. ISSN 08713413.

EKICI, S.; JAWZAL, H. Breast cancer diagnosis using thermography and convolutional neural networks. *Medical Hypotheses*, v. 137, p. 109542, 2020. ISSN 0306-9877. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0306987719313659>.

ETEHADTAVAKOL, M.; NG, E. Breast thermography as a potential non-contact method in the early detection of cancer: A review. *Journal of Mechanics in Medicine and Biology*, v. 13, 04 2013.

FREITAS, R. C. de; NAIK, G. R.; VALENÇA, M. J. S.; BEZERRA, B. L. D.; SOUZA, R. E. de; SANTOS, W. P. dos. Surface electromyography classification using extreme learning machines and echo state networks. *Research on Biomedical Engineering*, Springer, p. 1–22, 2022.

GARDNER, M.; DORLING, S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, v. 32, n. 14, p. 2627–2636, 1998. ISSN 1352-2310. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1352231097004470>.

GOMES, J. C.; BARBOSA, V. A. d. F.; SANTANA, M. A.; BANDEIRA, J.; VALENÇA, M. J. S.; SOUZA, R. E. de; ISMAEL, A. M.; SANTOS, W. P. dos. Ikonos: an intelligent tool to support diagnosis of covid-19 by texture analysis of x-ray images. *Research on Biomedical Engineering*, Springer, p. 1–14, 2020.

GOMES, J. C.; MASOOD, A. I.; SILVA, L. H. d. S.; FERREIRA, J. R. B. da C.; JUNIOR, A. A. F.; ROCHA, A. L. d. S.; OLIVEIRA, L. C. P. de; SILVA, N. R. C. da; FERNANDES, B. J. T.; SANTOS, W. P. D. Covid-19 diagnosis by combining rt-pcr and pseudo-convolutional machines to characterize virus sequences. *Scientific Reports*, Nature Publishing Group, v. 11, n. 1, p. 1–28, 2021.

GONÇALVES, C. B. Detecção de câncer de mama utilizando imagens termográficas. Universidade Federal de Uberlândia, 2017.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2nd. ed. [S.I.]: Prentice Hall PTR, 1998.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, v. 2016-December, p. 770–778, 2016. ISSN 10636919.

HUSAINI, M. A. S. A.; HABAEBI, M. H.; HAMEED, S. A.; ISLAM, M. R.; GUNAWAN, T. S. A systematic review of breast cancer detection using thermography and neural networks. *IEEE Access*, IEEE, 2020.

LANDOLA, F. N.; HAN, S.; MOSKEWICZ, M. W.; ASHRAF, K.; DALLY, W. J.; KEUTZER, K. SQUEEZENET: ALEXNET-LEVEL ACCURACY WITH 50X FEWER PARAMETERS AND <0.5MB MODEL SIZE. 2016. Disponível em: <<https://github.com/DeepScale/SqueezeNet>>.

INSTITUTO NACIONAL DO CÂNCER. FALANDO SOBRE CÂNCER DE MAMA. 2002. Disponível em: <https://bvsms.saude.gov.br/bvs/publicacoes/falando_cancer_mama1.pdf>.

INSTITUTO NACIONAL DO CÂNCER. *Estatísticas de câncer*. 2020. Disponível em: <<https://www.inca.gov.br/numeros-de-cancer>>. Acesso em: 14/03/2021.

INSTITUTO NACIONAL DO CÂNCER. *Câncer de mama - versão para População*. 2022. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>>. Acesso em: 08/01/2022.

INSTITUTO NACIONAL DO CÂNCER. *Câncer de mama - versão para Profissionais da Saúde*. 2022. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama/profissional-de-saude>>. Acesso em: 08/01/2022.

INSTITUTO NACIONAL DO CÂNCER. *Fatores de risco*. 2022. Disponível em: <<https://www.inca.gov.br/controle-do-cancer-de-mama/fatores-de-risco>>. Acesso em: 08/01/2022.

INSTITUTO NACIONAL DO CÂNCER. *Tipos de câncer: Câncer de Mama*. 2022. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>>. Acesso em: 07/01/2022.

KANDLIKAR, S. G.; PEREZ-RAYA, I.; RAGHUPATHI, P. A.; GONZALEZ-HERNANDEZ, J.-L.; DABYDEEN, D.; MEDEIROS, L.; PHATAK, P. Infrared imaging technology for breast cancer detection – current status, protocols and new directions. *International Journal of Heat and Mass Transfer*, v. 108, p. 2303–2320, 2017. ISSN 0017-9310. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0017931016336031>>.

KENNEDY, J.; EBERHART, R. Particle swarm optimization. In: *Proceedings of ICNN'95 - International Conference on Neural Networks*. [S.l.: s.n.], 1995. v. 4, p. 1942–1948 vol.4.

KOVÁCS, B.; TINYA, F.; NÉMETH, C.; ÓDOR, P. Unfolding the effects of different forestry treatments on microclimate in oak forests: results of a 4-yr experiment. *Ecological Applications*, v. 30, n. 2, p. 321–357, 2020. ISSN 19395582.

LANG, S.; BRAVO-MARQUEZ, F.; BECKHAM, C.; HALL, M.; FRANK, E. Wekadeeplearning4j: A deep learning package for weka based on deeplearning4j. *Knowledge-Based Systems*, Elsevier, v. 178, p. 48 – 50, 2019. ISSN 0950-7051. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950705119301789>>.

Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998.

LEE, H. D. *Seleção de atributos importantes para a extração de conhecimento de bases de dados*. Tese (Doutorado) — Universidade de São Paulo, 2005.

LibreOffice. *Libre Office The Document Foundation*. 2021. Disponível em: <<https://pt-br.libreoffice.org/>>. Acesso em: 17/07/2021.

LIMA, S. M. de; SILVA-FILHO, A. G. da; SANTOS, W. P. dos. A methodology for classification of lesions in mammographies using zernike moments, elm and svm neural networks in a multi-kernel approach. In: IEEE. *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.], 2014. p. 988–991.

LIMA, S. M. de; SILVA-FILHO, A. G. da; SANTOS, W. P. dos. Detection and classification of masses in mammographic images in a multi-kernel approach. *Computer methods and programs in biomedicine*, Elsevier, v. 134, p. 11–29, 2016.

LOPES, B.; RAMOS, I. C. de O.; RIBEIRO, G.; CORREA, R.; VALBON, B. de F.; LUZ, A. C. da; SALOMÃO, M.; LYRA, J. M.; JUNIOR, R. A. Bioestatísticas: conceitos fundamentais e aplicações práticas biostatistics: fundamental concepts and practical applications artigo original. *Rev Bras Oftalmol*, v. 73, p. 16–22, 2014.

MACEDO, M.; SANTANA, M.; SANTOS, W. P. dos; MENEZES, R.; BASTOS-FILHO, C. Breast cancer diagnosis using thermal image analysis: A data-driven approach based on swarm intelligence and supervised learning for optimized feature selection. *Applied Soft Computing*, Elsevier, v. 109, p. 107533, 2021.

MACIEL, E.; DE, E. L. Uma Ferramenta para Extração de Padrões. 1997.

MADHU, H.; KAKILETI, S. T.; VENKATARAMANI, K.; JABBIREDDY, S. Extraction of medically interpretable features for classification of malignancy in breast thermography. In: IEEE. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.], 2016. p. 1062–1065.

MAMBOU, S. J.; MARESOVA, P.; KREJCAR, O.; SELAMAT, A.; KUCA, K. Breast cancer detection using infrared thermal imaging and a deep learning model. *Sensors (Switzerland)*, v. 18, n. 9, 2018. ISSN 14248220.

MENEZES, N. G. SÍNTESE TOPOLÓGICA EVOLUCIONÁRIA E OTIMIZAÇÃO NUMÉRICA DE PARÂMETROS EM ESTRUTURAS CMOS . Dissertação, Engenharia Elétrica, COPPE/UFRJ, Rio de Janeiro, 2007.

MINISTÉRIO DA SAÚDE Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) 6ª edição revista e atualizada. 2011. Disponível em: <www.inca.gov.br>.

MISHRA, S.; PRAKASH, A.; ROY, S. K.; SHARAN, P.; MATHUR, N. Breast cancer detection using thermal images and deep learning. In: *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*. [S.l.: s.n.], 2020. p. 211–216.

MOGHBEL, M.; MASHOHOR, S. A review of computer assisted detection/diagnosis (CAD) in breast thermography for breast cancer detection. *Artificial Intelligence Review*, Springer, 2013.

MONARD, M. C.; BARANAUSKAS, J. A. Capítulo 5: Indução de Regras e Árvores de Decisão. *Sistemas Inteligentes Fundamentos e Aplicações*, n. m, p. 57–74, 2003. Disponível em: <<http://labic.icmc.usp.br/publicacao/406>>.

MORAIS, K. C. C. Desenvolvimento de metodologia baseada na imagem infravermelha para triagem de lesão de mama. 2015.

MOREIRA, I. I. P.; MORAIS, E. C.; MORAES, R. I. d. R.; ZISSOU, A. d. J.; CAMPOS, P. S. d. S.; MARQUES, G. T.; CHASE, O. A.; De Almeida, J. F. S. Redes neurais artificiais para previsão de capacidade de carga em estacas do tipo hélice contínua / Artificial neural networks for load capacity prediction in continuous flight auger piles. *Brazilian Journal of Development*, v. 7, n. 12, p. 112577–112597, 2021.

NASTRI, C. O.; MARTINS, W. d. P.; LENHARTE, R. d. J. Ultrassonografia no rastreamento do câncer de mama. *Femina*, p. 97–102, 2011.

NETO, O. P. d. S. et al. Detecção automática de massas em imagens mamográficas usando particle swarm optimization (psa) e índice de diversidade funcional. Universidade Federal do Maranhão, 2016.

PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 10, p. 1345–1359, 2010.

PEREIRA, J.; SANTANA, M.; LIMA, N.; SOUSA, F.; LIMA, R.; SANTOS, W. Método para classificação do tipo da lesão na mama presentes nas imagens termográficas utilizando classificador elm. In: *Anais do I Simpósio de Inovação em Engenharia Biomédica-SABIO 2017*. [S.l.: s.n.], 2017. p. 1–5.

PEREIRA, J.; SANTANA, M. A.; GOMES, J. C.; BARBOSA, V. A. de F.; VALENÇA, M. J. S.; LIMA, S. M. L. de; SANTOS, W. P. dos. Feature selection based on dialectics to support breast cancer diagnosis using thermographic images. *Research on Biomedical Engineering*, Springer, v. 37, n. 3, p. 485–506, 2021.

PEREIRA, J. M. S. Seleção de atributos baseada em algoritmo dialético de otimização para classificação de lesões de mama em imagens termográficas Seleção de atributos baseada em algoritmo dialético de otimização para classificação de lesões de mama em imagens termográficas. Dissertação (Mestrado) - Universidade Federal de Pernambuco, 2020.

PEREIRA, J. M. S.; SANTANA, M.; LIMA, R.; LIMA, S.; SANTOS, W. Method for classification of breast lesions in thermographic images using elm classifiers. In: *Understanding a cancer diagnosis*. [S.l.]: Nova Science New York, 2020. p. 117–132.

PEREIRA, J. M. S.; SANTANA, M.; SILVA, W.; LIMA, R.; LIMA, S.; SANTOS, W. Dialectical optimization method as a feature selection tool for breast cancer diagnosis using thermographic images. In: *Understanding a cancer diagnosis*. [S.l.]: Nova Science New York, 2020. p. 95–118.

PEREIRA, J. M. S.; SANTANA, M. A. de; LIMA, C. L. de; LIMA, R. d. C. F. de; LIMA, S. M. L. de; SANTOS, W. P. dos. Feature selection based on dialectical optimization algorithm for breast lesion classification in thermographic images. In: *Biomedical Computing for Breast Cancer Detection and Diagnosis*. [S.l.]: IGI Global, 2021. p. 47–71.

RODRIGUES, A. L.; SANTANA, M. A. de; AZEVEDO, W. W.; BEZERRA, R. S.; BARBOSA, V. A.; LIMA, R. C. de; SANTOS, W. P. dos. Identification of mammary lesions in thermographic images: feature selection study using genetic algorithms and particle swarm optimization. *Research on Biomedical Engineering*, Springer, v. 35, n. 3, p. 213–222, 2019.

RODRIGUES, J. D.; CRUZ, M. S.; PAIXÃO, A. N. Uma análise da prevenção do câncer de mama no brasil. *Ciência & saúde coletiva*, SciELO Public Health, v. 20, p. 3163–3176, 2015.

Rosebrock, A. *Transfer Learning with Keras and Deep Learning* (2019). 2019. Disponível em: <<https://www.pyimagesearch.com/2019/05/20/transfer-learning-with-keras-and-deep-learning/>>. Acesso em: 15/08/2021.

ROSLIDAR, R.; SADDAMI, K.; ARNIA, F.; SYUKRI, M.; MUNADI, K. A study of fine-tuning cnn models based on thermal imaging for breast cancer classification. In: . [S.l.: s.n.], 2019. p. 77–81.

SANTANA, M.; PEREIRA, J. M. S.; LIMA, R.; SANTOS, W. Breast lesions classification in frontal thermographic images using intelligent systems and moments of haralick and zernike. In: *Understanding a cancer diagnosis*. [S.l.]: Nova Science New York, 2020. p. 65–80.

SANTANA, M. A. d. *Sistemas inteligentes para apoio ao diagnóstico do câncer de mama usando imagens mamográficas e termográficas*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2020.

SANTANA, M. A. d.; PEREIRA, J. M. S.; SILVA, F. L. d.; LIMA, N. M. d.; SOUSA, F. N. d.; ARRUDA, G. M. S. d.; LIMA, R. d. C. F. d.; SILVA, W. W. A. d.; SANTOS, W. P. d. Breast cancer diagnosis based on mammary thermography and extreme learning machines. *Research on Biomedical Engineering*, 2018.

SANTANA, M. A. de; GOMES, J. C.; BARBOSA, V. A. de F.; LIMA, C. L. de; BANDEIRA, J.; VALENÇA, M. J. S.; SOUZA, R. E. de; MASOOD, A. I.; SANTOS, W. P. dos. An intelligent tool to support diagnosis of covid-19 by texture analysis of computerized tomography x-ray images and machine learning. In: *Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis*. [S.l.]: Springer, 2022. p. 259–282.

SANTANA, M. A. de; PEREIRA, J. M. S.; LIMA, C. L. de; ALMEIDA, M. B. J. de; ANDRADE, J. F. S. de; SOUZA, T. K. S. de; LIMA, R. d. C. F. de; SANTOS, W. P. dos. Classification of breast lesions in frontal thermographic images using a diagnosis aid intelligent system. In: *Biomedical Computing for Breast Cancer Detection and Diagnosis*. [S.l.]: IGI Global, 2021. p. 28–46.

SANTANA, M. A. de; PEREIRA, J. M. S.; SILVA, F. L. da; LIMA, N. M. de; SOUSA, F. N. de; ARRUDA, G. M. S. de; LIMA, R. d. C. F. de; SILVA, W. W. A. de; SANTOS, W. P. dos. Breast cancer diagnosis based on mammary thermography and extreme learning machines. *Research on Biomedical Engineering*, v. 34, n. 1, p. 45–53, 2018. ISSN 24464740.

SANTANA, M. A. de; PEREIRA, J. M. S.; SILVA, W. W. A. da; SANTOS, W. P. dos. Breast cancer diagnosis in mammograms using wavelet analysis, haralick descriptors, and autoencoder. In: *AI Innovation in Medical Imaging Diagnostics*. [S.l.]: IGI Global, 2021. p. 76–91.

SANTOS, W. P. dos; SILVA, W. W. A. da; SANTANA, M. A. de. *Biomedical Computing for Breast Cancer Detection and Diagnosis*. [S.l.]: IGI Global, 2020.

SAXEN, F.; WERNER, P.; HANDRICH, S.; OTHMAN, E.; DINGES, L.; AL-HAMADI, A. Face Attribute Detection with MobileNetV2 and NasNet-Mobile. n. September, 2019.

Saúde bem estar. *Mamografia*. 2020. Disponível em: <<https://www.saudebemestar.pt/pt/exame/imagiologia/mamografia/>>. Acesso em: 09/01/2022.

SciDAVis. *SciDAVis*. 2021. Disponível em: <<http://scidavis.sourceforge.net/index.html>>. Acesso em: 17/07/2021.

SILVA, A.; SANTANA, M.; AZEVEDO, W.; BEZERRA, R.; Dos Santos, W.; LIMA, R. Seleção de Atributos para Apoio ao Diagnóstico do Câncer de Mama Usando Imagens Termográficas, Algoritmos Genéticos e Otimização por Enxame de Partículas. *II Simpósio De Inovação Em Engenharia Biomédica*, n. September, p. 1–6, 2018. Disponível em: <https://www.researchgate.net/publication/327384817_Selecao_de_Atributos_para_Apoio_ao_Diagnostico_do_Cancer_de_Mama_Usando_Imagens_Termograficas_Algoritmos_Geneticos_e_Otimizacao_por_Enxame_de_Partículas>.

SILVA, A. L. R. da; SANTANA, M. A. de; LIMA, C. L. de; ANDRADE, J. F. S. de; SOUZA, T. K. S. de; ALMEIDA, M. B. J. de; SILVA, W. W. A. da; LIMA, R. d. C. F. de; SANTOS, W. P. dos. Features selection study for breast cancer diagnosis using thermographic images, genetic algorithms, and particle swarm optimization. *International Journal of Artificial Intelligence and Machine Learning (IJAIML)*, IGI Global, v. 11, n. 2, p. 1–18, 2021.

SILVA, A. S. V. d. *Classificação e segmentação de termogramas de mama para triagem de pacientes residentes em regiões de poucos recursos médicos*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2015.

SILVA, I. R. R. da; SILVA, G. d. S. L. e; SOUZA, R. G. de; SANTANA, M. A. de; SILVA, W. W. A. da; LIMA, M. E. de; SOUZA, R. E. de; FAGUNDES, R.; SANTOS, W. P. dos. Deep learning for early diagnosis of alzheimer's disease: a contribution and a brief review. *Deep Learning for Data Analytics*, Elsevier, p. 63–78, 2020.

SILVA, P. A. d.; RIUL, S. d. S. Câncer de mama: fatores de risco e detecção precoce. *Revista Brasileira de Enfermagem*, SciELO Brasil, v. 64, n. 6, p. 1016–1021, 2011.

SILVA, W. W. A. da; SANTANA, M. A. de; FILHO, A. G. da S.; LIMA, S. M. L. de; SANTOS, W. P. dos. Morphological extreme learning machines applied to the detection and classification of mammary lesions. In: *Advanced machine vision paradigms for medical image analysis*. [S.l.]: Elsevier, 2021. p. 55–95.

SOUZA, R. G. de; SILVA, G. dos Santos Lucas e; SANTOS, W. P. dos; LIMA, M. E. de. Computer-aided diagnosis of alzheimer's disease by mri analysis and evolutionary computing. *Research on Biomedical Engineering*, Springer, v. 37, n. 3, p. 455–483, 2021.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S. E.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014. Disponível em: <<http://arxiv.org/abs/1409.4842>>.

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015. Disponível em: <<http://arxiv.org/abs/1512.00567>>.

THECKEDATH, D.; SEDAMKAR, R. R. Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks. *SN Computer Science*, Springer Singapore, v. 1, n. 2, p. 1–7, 2020. ISSN 2662-995X. Disponível em: <<https://doi.org/10.1007/s42979-020-0114-9>>.

TIWARI, A.; GUPTA, A.; KATARIA, H.; SINGAL, G. Analysis of Machine Learning and Deep Learning Classifiers to Detect and Classify Breast Cancer. v. 58, p. 471–480, 2021.

VASCONCELOS, J. de; SANTOS, W. D.; LIMA, R. d. C. F. de. Analysis of methods of classification of breast thermographic images to determine their viability in the early breast cancer detection. *IEEE Latin America Transactions*, IEEE, v. 16, n. 6, p. 1631–1637, 2018.

WITTEN, I. H.; FRANK, E. *Data Mining: Pratical Machine Learning Tools and Technique*. San Francisco, CA, USA: Morgan Kaufmann Publishers, 2005.

ZHANG, H. The optimality of Naive Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2004*, v. 2, p. 562–567, 2004.